PhD 09/2024-005

# Essays on Labour Economics and Industrial Organization

Mario Bernasconi

Netspar

Academic paper

# Essays on labour economics and industrial organization

Bernasconi, Mario

*DOI:*
10.26116/tisem.41560108

*Publication date:*
2024

*Document Version*
Publisher's PDF, also known as Version of record


# Essays on Labour Economics and Industrial Organization

MARIO BERNASCONI

TILBURG UNIVERSITY

# Essays on Labour Economics and Industrial Organization

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. W.B.H.J. van de Donk, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in zaal C 186 van de Universiteit op

woensdag 11 september 2024 om 13:30 uur

door

Mario Bernasconi

geboren te Gallarate, Italië.

| | |
|---|---|
| Promotor: | prof. dr. Arthur van Soest (Tilburg University) |
| Copromotores: | dr. Tunga Kantarcı (Rijksuniversiteit Groningen)<br>dr. Alexandros Theloudis (Tilburg University) |
| Leden promotiecommissie: | prof. dr. Jaap Abbring (Tilburg University)<br>prof. dr. Margherita Borella (Università di Torino)<br>dr. Jochem de Bresser (Tilburg University)<br>prof. dr. Eric French (University of Cambridge) |

Essays on Labour Economics and Industrial Organization

# Acknowledgments

This thesis marks the end of a journey that began six years ago, when I moved to Tilburg to pursue the Research Master in Economics. It has been a long, challenging, and immensely rewarding experience, made possible by the help and support of family, friends, and colleagues. I owe a great deal to the many people I have met along the way.

First and foremost, I want to express my deepest gratitude to my supervisors, Arthur van Soest, Alexandros Theloudis, and Tunga Kantarcı. I am profoundly thankful for their unwavering support, guidance and encouragement. They have always been generous with their time and advice, which went beyond work and research matters. Their mentorship transcends the scope of this thesis, as they have been exemplary role models as researchers, teachers, and colleagues. They truly made me love working on my PhD.

Additionally, I am grateful to Jaap Abbring and Jochem de Bresser for agreeing to be on my dissertation committee and for their invaluable feedback. Jaap and Jochem have been involved with my education and research over the years, significantly enhancing my work and helping me navigate the academic job market. Similarly, I also want to thank Margherita Borella and Eric French. Standing on the shoulders of the giants, this thesis is largely based on their research contributions. I learned a lot from them even before meeting them, and it is a privilege to now receive their comments on my work.

I owe thanks to many faculty members at the Econometrics & OR department for their feedback and support, even after having seen my presentations multiple times. Among others, Christoph Walsh, Tobias Klein, Martin Salm, Bas Werker, Bettina Sifflinger, Nikolaus Schweizer, Jeffrey Campbell, Denis Kojevnikov, Christoph Hambel, Pavel Cizek, Otilia Boldea. I would also like to thank my co-authors, Miguel Espinosa, Rocco Macchiavello, Carlos Suárez, and Jan-Maarten van Sonsbeek, from whom I've learnt immensely.

I would like to thank Gabriella, Jan, Jun-Hee, Lieke, Rik, Tinghan, Shobhit, and Leena for their friendship and for sharing this journey with me, and for making these years so memorable and enjoyable.

Finally, thank you to my parents, Laura and Daniele, for their unconditional support and devotion, and to my sisters, Paola, Carla, Elena, and Elisa. Thank you to my greatest fan and life companion, Giulia, for being the kindest person and always being there for me.

Mario Bernasconi
Antwerp, June 2024

# Contents

**Academic summary** **1**

**Academische samenvatting** **5**

**Chapter 1: The added worker effect: evidence from a disability insurance reform** **9**

1.1 Introduction . . . . . 10
1.2 Disability insurance in the Netherlands and the 2006 reform . . . . . 13
1.3 Data . . . . . 16
1.4 Time trends and other descriptive statistics . . . . . 17
1.5 Identification strategy . . . . . 22
1.6 The effect of the reform on labour participation of sick individuals and their spouses . . . . . 23
1.6.1 Baseline effects . . . . . 24
1.6.2 Dynamic effects . . . . . 24
1.6.3 Heterogeneous effects . . . . . 25
1.7 Comparing with the reform effects on sick individuals without a spouse . 34
1.8 Checking the identifying assumptions . . . . . 39
1.9 Conclusion . . . . . 42
1.10 Appendix . . . . . 44
1.10.1 Timeline of changes in the Dutch DI scheme . . . . . 44
1.10.2 Relation with other reforms . . . . . 44
1.10.3 Dynamic heterogenous effects . . . . . 46
1.10.4 Placebo test . . . . . 50
1.10.5 Effects of the transitional WAO reform . . . . . 51
1.10.6 Comparing with individuals without a spouse . . . . . 54
1.10.7 Regression Discontinuity instead of Difference-in-Differences . . . 56
1.10.8 Do individuals self-select into the old or new disability scheme? . 60
1.10.9 Do couples dissolve their cohabitation due to the reform? . . . . . 61

1.10.10 Can compositional differences account for the heterogenous reform effects? . . . 63
References . . . 66

**Chapter 2: Pension reforms and partial retirement** **71**
2.1 Introduction . . . 72
2.2 Institutional setting . . . 76
2.3 Data and sample . . . 80
2.4 Empirical evidence of the effect of the reforms . . . 81
2.5 Model . . . 90
2.5.1 Outline of the model . . . 90
2.5.2 Parametrization . . . 93
2.5.3 Model estimation . . . 97
2.6 Results . . . 100
2.7 Policy simulations . . . 104
2.7.1 The value of partial retirement . . . 104
2.7.2 Increasing the state pension age . . . 109
2.8 Conclusion . . . 112
2.9 Appendix . . . 113
2.9.1 Details on the model setup . . . 113
2.9.2 Details on the solution and estimation of the model . . . 121
2.9.3 Additional results . . . 126
References . . . 137

**Chapter 3: Present-biased preferences, retirement planning and demand for commitments** **141**
3.1 Introduction . . . 141
3.2 Model . . . 144
3.2.1 Model setup . . . 144
3.2.2 Consumption decision . . . 145
3.2.3 Retirement decision . . . 148
3.3 Commitment device . . . 153
3.3.1 Consumption decision . . . 153
3.3.2 Retirement decision . . . 154
3.3.3 Partially naïve agents . . . 157
3.4 The role of uncertainty . . . 159
3.5 Conclusion . . . 160
3.6 Appendix . . . 161
3.6.1 Case 1 in Section 3.3 for sophisticated agents . . . 161

3.6.2 Case 2 in Section 3.3 for sophisticated agents . . . 162
3.6.3 Case 3 in Section 3.3 for sophisticated agents . . . 164
References . . . 167

**Chapter 4: Relational collusion in the Colombian electricity market 169**
4.1 Introduction . . . 169
4.2 Institutional setting & background . . . 176
4.2.1 Electricity demand and generation . . . 176
4.2.2 Colombian wholesale electricity market . . . 176
4.2.3 Change in transparency policy . . . 178
4.3 Detecting collusive agreements . . . 179
4.3.1 The logic of the test for collusive behaviour . . . 180
4.3.2 The main fact & proxying for cartel membership . . . 181
4.3.3 Cartel membership & bidding behaviour around the transparency reform . . . 185
4.3.4 Announcement date and threats of enforcement . . . 187
4.3.5 Discussion . . . 192
4.3.6 Robustness to alternative definitions . . . 193
4.4 Incentives to collude & inner functioning of the cartel . . . 193
4.4.1 Incentives to collude . . . 194
4.4.2 Inner working of the cartel . . . 194
4.4.3 Suggestive evidence of communication . . . 201
4.4.4 Lower profits from positive reconciliations after the end of the cartel 202
4.5 Incentive to deviate and cost of the cartel . . . 203
4.5.1 Modelling choices . . . 204
4.5.2 Bidding strategy . . . 205
4.5.3 Dynamic enforcement constraints . . . 207
4.5.4 Cost of the cartel . . . 211
4.6 Policy implications and conclusions . . . 213
4.7 Appendix . . . 214
4.7.1 Data . . . 214
4.7.2 Robustness in the cartel definition . . . 215
4.7.3 Details on the cost of the cartel . . . 217
4.7.4 Calculation marginal costs . . . 217
4.7.5 Additional figures . . . 219
4.7.6 Additional tables . . . 236
References . . . 241

# Academic summary

This thesis is a collection of four self-contained papers on topics related to labour economics and industrial organization. In the first part of the thesis, I study how different components of a welfare state, namely disability insurance (DI) and pensions, affect individual and household labour supply decisions. I show how the decision to work – and how much to work – is influenced by the incentives embedded in these schemes, which provides guidance on how to design them to increase efficiency and people's well-being. In the second part of the thesis, instead, I study how the institutional features of a market can facilitate or hinder collusion between firms. I show that, in some cases, excessive market transparency makes collusion easier to sustain and ultimately lowers consumers' welfare.

In the Netherlands, the share of people receiving DI benefits in the insured population reached about 11% in 2002. To secure the financial sustainability of the DI system and promote work resumption, successive governments implemented several DI reforms. In the first chapter, "The added worker effect: evidence from a disability insurance reform", which is joint work with Tunga Kantarcı, Arthur van Soest, and Jan-Maarten van Sonsbeek, we study how household labour supply responds to a DI reform that made access to DI more difficult. In particular, we focus on the spouses of sick people: Since couples can pool income risk, spousal labour supply can be an important self-insurance mechanism to counterbalance the loss of income due to the reform. Based on a difference-in-differences identification strategy, we find clear evidence of an added worker effect for spouses of workers who report sick from a weak labour market position, from which work resumption is difficult. The effect on the spouse's labour participation is substantial, constituting approximately one-sixth and one-half of the reform's impact on disability benefit receipt and work resumption for the sick individuals themselves, respectively. On average, this added worker effect results in households not suffering any income loss due to the reform. The finding implies that for a complete evaluation of the DI reform and its effects on labour participation as well as adequacy of household income, it is important to consider spillover effects on spouses.

In the past decades, social security systems in most Western countries also faced grow-

ing financial pressure due to the increase in life expectancy and the decline in fertility rates, which have led to major pension reforms with the aim of stimulating old-age employment. In the second chapter, "Pension reforms and partial retirement", which is joint work with Tunga Kantarcı, we study how different pension schemes affect labour supply of older people. First, we exploit two pension reforms implemented in the Netherlands to show how different pension regimes affect the retirement age but also the decision to work part-time at old age. Second, we develop a structural model of retirement which accounts for assets and pension rights accumulation, the bunching of work hours at different levels, and the possibility to retire partially. We estimate and validate the model exploiting the exogenous variation stemming from two reforms. Finally, we use the model to show that the effect of partial retirement on labour supply is heterogeneous across pension regimes, but positive under the reformed Dutch one, and it increases total work hours by up to 2.5 percent at age 66. Workers with lower wealth, who cannot afford to work part-time otherwise, value partial retirement most. Moreover, the model suggests that marginal returns – in terms of labour supply – from increasing the state pension age are decreasing, raising the question to what extent labour supply at old age can be further increased with similar policies.

The third chapter also concerns retirement decisions, but with insights form behavioural economics. In this chapter, "Present-biased preferences, retirement planning and demand for commitments", I study how present-bias preferences affect savings and retirement decisions. Income adequacy in old age is crucial to ensure retirees' well-being, yet most people do not save enough for their retirement. Present-biased preferences can explain the mismatch between desired and realized savings, but they can also lead to time-inconsistent retirement choices. Through the lenses of a simple model, I show that illiquid assets can be used as a commitment device also to affect retirement decisions. However, when two interdependent choices have to be made, such as consumption and retirement, the demand for commitment is non-trivial. In this case, the effectiveness of one commitment device can be high or low depending on whether it can address both actions in the desired way at the same time. For partially naïve agents, who ignore their bias with respect to the retirement decision, the commitment can be costly while futile, i.e. it can negatively affect the retirement choice without improving the inter-temporal consumption allocation. Sophisticated agents, instead, can have high or low willingness to pay for the commitment device depending on whether it helps improving the consumption and retirement decisions at the same time or not.

The fourth chapter relates to topics in industrial organization. Despite the change in topics, the quasi-experimental methods used and the underlying theoretical considerations, in which dynamic incentives play a key role, are a recurrent theme of this thesis. In this chapter, "Relational collusion in the Colombian electricity market", which is joint

work with Miguel Espinosa, Rocco Macchiavello, and Carlos Suarez, we study how market transparency can facilitate collusion between firms. Under collusion, firms deviate from current profit maximization in anticipation of future rewards. As current profit maximization places little restrictions on firms' pricing behaviour, collusive conduct is hard to infer. We exploit a regulatory change in the Colombian wholesale electricity market, which increased the delay to disclose market information to participants, to infer the existence of a cartel. We show that cartel members – defined ex-ante as those more likely to gain from collusion – decreased bids after the announcement of the policy and before its implementation. After ruling out confounders, we provide evidence on the functioning of the cartel and on how firms may have communicated. Finally, we calibrate the dynamic incentive compatibility constraints of cartel members, which confirms that collusion was sustainable before, but not after, the reform. Our results suggest that regulators' actions can reduce collusive behaviour, and ultimately that dynamic incentive compatibility constraints can be taken seriously by empirical researchers and policy-makers fighting collusion.

# Academische samenvatting

Deze scriptie is een verzameling van vier zelfstandige artikelen over onderwerpen gerelateerd aan arbeidseconomie en industriële organisatie. In het eerste deel van de scriptie onderzoek ik hoe verschillende componenten van een verzorgingsstaat, namelijk arbeidsongeschiktheidsverzekering (AO) en pensioenen, de arbeidsaanbodbeslissingen van individuen en huishoudens beïnvloeden. Ik laat zien hoe de beslissing om te werken – en hoeveel te werken – wordt beïnvloed door de prikkels die in deze regelingen zijn ingebouwd, wat aanwijzingen geeft over hoe ze kunnen worden ontworpen om de efficiëntie en het welzijn van mensen te verhogen. In het tweede deel van de scriptie onderzoek ik hoe de institutionele kenmerken van een markt samenwerking tussen bedrijven kunnen vergemakkelijken of belemmeren. Ik laat zien dat, in sommige gevallen, overmatige markttransparantie samenwerking gemakkelijker kan maken en uiteindelijk het welzijn van de consumenten verlaagt.

In Nederland bereikte het aandeel mensen dat een AO-uitkering ontvangt in de verzekerde bevolking ongeveer 11% in 2002. Om de financile houdbaarheid van het AO-stelsel te waarborgen en de werkhervatting te bevorderen, hebben opeenvolgende regeringen verschillende AO-hervormingen doorgevoerd. In het eerste hoofdstuk, "Het toegevoegde arbeider-effect: bewijs uit een hervorming van de arbeidsongeschiktheidsverzekering", dat gezamenlijk werk is met Tunga Kantarcı, Arthur van Soest en Jan-Maarten van Sonsbeek, onderzoeken we hoe het arbeidsaanbod van huishoudens reageert op een AO-hervorming die de toegang tot AO moeilijker maakte. We richten ons met name op de echtgenoten van zieke mensen: aangezien echtparen inkomensrisico kunnen delen, kan het arbeidsaanbod van de echtgenoot een belangrijke zelfverzekeringsmechanisme zijn om het inkomensverlies als gevolg van de hervorming tegen te gaan. Op basis van een verschil-in-verschillen identificatiestrategie vinden we duidelijk bewijs van een toegevoegde arbeider-effect voor echtgenoten van werknemers die ziekmelden vanuit een zwakke arbeidsmarktpositie, waarvan werkhervatting moeilijk is. Het effect op de arbeidsparticipatie van de echtgenoot is aanzienlijk en bedraagt respectievelijk ongeveer een zesde en de helft van de impact van de hervorming op het ontvangen van arbeidsongeschiktheidsuitkeringen en werkhervatting van de zieke personen zelf. Gemiddeld resulteert dit toegevoegde arbeider-effect

erin dat huishoudens geen inkomensverlies lijden door de hervorming. De bevinding impliceert dat voor een volledige evaluatie van de AO-hervorming en de effecten ervan op arbeidsparticipatie en de toereikendheid van het huishoudinkomen, het belangrijk is om spillover-effecten op echtgenoten in overweging te nemen.

In de afgelopen decennia stonden de sociale zekerheidsstelsels in de meeste westerse landen ook onder toenemende financile druk als gevolg van de stijging van de levensverwachting en de daling van de vruchtbaarheidscijfers, wat heeft geleid tot grote pensioenhervormingen met als doel het stimuleren van ouderdomswerkgelegenheid. In het tweede hoofdstuk, "Pensioenhervormingen en gedeeltelijk pensioen", dat gezamenlijk werk is met Tunga Kantarcı, onderzoeken we hoe verschillende pensioenregelingen het arbeidsaanbod van oudere mensen beïnvloeden. Eerst benutten we twee pensioenhervormingen die in Nederland zijn doorgevoerd om te laten zien hoe verschillende pensioenregimes de pensioenleeftijd beïnvloeden, maar ook de beslissing om op oudere leeftijd deeltijd te werken. Ten tweede ontwikkelen we een structureel pensioenmodel dat rekening houdt met de opbouw van vermogen en pensioenrechten, de bundeling van werkuren op verschillende niveaus en de mogelijkheid om gedeeltelijk met pensioen te gaan. We schatten en valideren het model door gebruik te maken van de exogene variatie die voortkomt uit twee hervormingen. Tot slot gebruiken we het model om te laten zien dat het effect van gedeeltelijk pensioen op het arbeidsaanbod heterogeen is tussen pensioenregimes, maar positief onder het hervormde Nederlandse stelsel, en het verhoogt het totale aantal werkuren met maximaal 2,5 procent op 66-jarige leeftijd. Werknemers met een lager vermogen, die zich anders geen deeltijdwerk kunnen veroorloven, waarderen gedeeltelijk pensioen het meest. Bovendien suggereert het model dat de marginale opbrengsten in termen van arbeidsaanbod van het verhogen van de AOW-leeftijd afnemen, wat de vraag oproept in hoeverre het arbeidsaanbod op oudere leeftijd verder kan worden verhoogd met soortgelijke beleidsmaatregelen.

Het derde hoofdstuk betreft ook pensioenbeslissingen, maar met inzichten uit de gedragseconomie. In dit hoofdstuk, "Heden-vooringenomen voorkeuren, pensioenplanning en vraag naar verplichtingen", onderzoek ik hoe heden-vooringenomen voorkeuren sparen en pensioenbeslissingen beïnvloeden. Inkomensadequaatheid op oudere leeftijd is cruciaal om het welzijn van gepensioneerden te waarborgen, maar de meeste mensen sparen niet genoeg voor hun pensioen. Heden-vooringenomen voorkeuren kunnen de kloof tussen gewenste en gerealiseerde besparingen verklaren, maar ze kunnen ook leiden tot tijdsinconsistente pensioenbeslissingen. Door de lens van een eenvoudig model laat ik zien dat illiquide activa kunnen worden gebruikt als een verplichtingsmechanisme om ook pensioenbeslissingen te beïnvloeden. Echter, wanneer twee onderling afhankelijke keuzes moeten worden gemaakt, zoals consumptie en pensioen, is de vraag naar verplichting niet triviaal. In dit geval kan de effectiviteit van n verplichtingsmechanisme hoog of laag zijn,

afhankelijk van of het beide acties tegelijkertijd op de gewenste manier kan aanpakken. Voor gedeeltelijk naïeve agenten, die hun vooringenomenheid ten opzichte van de pensioenbeslissing negeren, kan de verplichting kostbaar maar zinloos zijn, dat wil zeggen dat het de pensioenbeslissing negatief kan beïnvloeden zonder de intertemporele consumptietoewijzing te verbeteren. Geavanceerde agenten kunnen daarentegen een hoge of lage bereidheid hebben om te betalen voor het verplichtingsmechanisme, afhankelijk van of het helpt om zowel de consumptie- als pensioenbeslissingen tegelijkertijd te verbeteren of niet.

Het vierde hoofdstuk heeft betrekking op onderwerpen in de industrile organisatie. Ondanks de verandering in onderwerpen, zijn de quasi-experimentele methoden die worden gebruikt en de onderliggende theoretische overwegingen, waarin dynamische prikkels een sleutelrol spelen, een terugkerend thema in deze scriptie. In dit hoofdstuk, "Relationele collusie in de Colombiaanse elektriciteitsmarkt", dat gezamenlijk werk is met Miguel Espinosa, Rocco Macchiavello en Carlos Suarez, onderzoeken we hoe markttransparantie samenwerking tussen bedrijven kan vergemakkelijken. Onder samenwerking wijken bedrijven af van de huidige winstmaximalisatie in afwachting van toekomstige beloningen. Aangezien huidige winstmaximalisatie weinig beperkingen oplegt aan het prijsbepalingsgedrag van bedrijven, is collusief gedrag moeilijk te achterhalen. We benutten een regelgevende verandering op de Colombiaanse groothandelsmarkt voor elektriciteit, die de vertraging om marktinformatie aan deelnemers bekend te maken, vergrootte, om het bestaan van een kartel af te leiden. We laten zien dat kartelleden – ex ante gedefinieerd als degenen die het meest waarschijnlijk van samenwerking profiteren – hun biedingen verlaagden na de aankondiging van het beleid en voor de implementatie ervan. Na het uitsluiten van verwarrende factoren, leveren we bewijs over de werking van het kartel en hoe bedrijven mogelijk hebben gecommuniceerd. Ten slotte kalibreren we de dynamische incentive-compatibiliteitsbeperkingen van kartelleden, wat bevestigt dat samenwerking voor de hervorming duurzaam was, maar niet daarna. Onze resultaten suggereren dat acties van regelgevers collusief gedrag kunnen verminderen, en uiteindelijk dat dynamische incentive-compatibiliteitsbeperkingen serieus kunnen worden genomen door empirische onderzoekers en beleidsmakers die strijd voeren tegen collusie.

# Chapter 1: The added worker effect: evidence from a disability insurance reform

Joint work with Tunga Kantarcı, Arthur van Soest, and Jan-Maarten van Sonsbeek.

**Abstract**

The Netherlands reformed its disability insurance (DI) scheme in 2006. Reintegration incentives for employers became stronger, access to DI benefits became more difficult, or benefits became less generous. Using administrative data on all individuals who fell sick shortly before and after the reform, we study the impact of the reform on labour participation of individuals who fell sick and their spouses. Difference-in-differences estimates show, among other things, that the reform led to an increase of labour participation of the individuals who fell sick only if these individuals had a permanent job, whereas spouses responded to the DI reform in other cases, where the individuals reporting sick had a temporary job or were unemployed. More generally, the spouses respond when the sick individual's labour market position is weak and the individual him- or herself has trouble finding or retaining employment. The effects are persistent during the ten years after the reform. The effect on the spouse can be seen as an "added worker effect," where additional earnings of the spouse compensate for the sick individual's income loss so that both partners share the burden of a more stringent DI scheme. Comparing individuals reporting sick with and without partner provides further support for the notion that the responses of couples to the reform are joint decisions of the two partners.[1]

[1]Published in the *Review of Economics of the Household* (2024). This research is supported by Netspar under grant number LMVP 2019.01. Its contents are the sole responsibility of the authors. We thank the Employee Insurance Agency (UWV), and in particular Lucien Rondagh, Willy van den Berk, Carla van Deursen, and Roel Ydema, for providing the disability data. We thank two anonymous referees, the participants of Netspar workshops in 2020 and 2021, ESPE 2021, SEHO 2022 and EALE 2022, and Jaap Abbring, Margherita Borella, Jochem de Bresser, and Eric French for their constructive comments on earlier versions of the paper. Results are based on calculations by the authors using non-public microdata from Statistics Netherlands. Under certain conditions, these microdata are accessible for statistical and scientific research. For further information: microdata@cbs.nl.

## 1.1 Introduction

A large and growing strand of the literature analyzes income complementarities in the household as an insurance mechanism. The "added worker effect" hypothesis suggests that married women respond to a negative shock on their husbands' earnings due to unemployment by increasing their hours of paid work (Lundberg, 1985). Most studies find no or a small added worker effect (Maloney, 1987, 1991; Spletzer, 1997; García-Gómez et al., 2012; Bredtmann et al., 2018; Halla et al., 2020; Cammeraat et al., 2023; Jolly and Theodoropoulos, 2023). One explanation is that the affected partner is insured through social insurance so that the spouse does not need to respond (Cullen and Gruber, 2000; Bentolila and Ichino, 2008). Couples may also self-insure through savings and run down their wealth in response to a negative income shock (Blundell et al., 2016). Similarly, the wife's response may be small if the husband's unemployment only leads to a transitory reduction in earnings (Cullen and Gruber; Bredtmann et al.) or if the husband's unemployment is anticipated by the household and the expected income loss already led to adjustments in household consumption and labour supply. In addition, the wife's response will depend on the magnitude of the expected loss in lifetime income (Cullen and Gruber; Stephens, 2002; Bredtmann et al.). An alternative explanation is that the wife's employment prospects may be affected by the factors causing the husband's unemployment (Cullen and Gruber).

Some recent studies, however, do find a notable added worker effect. Ayhan (2018) finds that the probability of a woman participating in the labour force increases by up to 28% in response to her husband's unemployment, although only for two quarters. Schøne and Strøm (2021) find that the rise in wives' labour supply annihilates around one third of the loss in husbands' earnings. Moreover, Blundell et al. show that of the total amount of consumption insured against permanent shocks to the husband's wage, about 63% comes from family labour supply.

In this paper we investigate the existence of an added worker effect in the context of a DI reform that limited DI eligibility. The DI context is interesting for several reasons. First, the number of DI recipients is large and growing in many countries, creating an important challenge for social security funding (OECD, 2018). Moreover, workers who lose income due to the reform have health problems limiting their possibilities to work and recover the income loss themselves – and the income loss is more likely to be permanent than in case of unemployment, which is often temporary. Finally, the reform weakens protection from social insurance, raising the need for self-insurance of the household, for example through a spousal response. Indeed, Bredtmann et al. (2018) show that added worker effects are larger in countries with less protection from social insurance schemes.

In the Netherlands, the share of people receiving DI benefits in the insured population

reached about 11%, with almost one million DI recipients in 2002 (Koning and Lindeboom, 2015). To reduce this number and promote work resumption, successive governments implemented several DI reforms. In 2006, the Work and Income According to Labour Capacity Act (WIA) came into effect, replacing the (transitional) Disability Insurance Act (WAO) as the final element of these reforms. WIA introduced major changes in both the DI scheme and the sickness insurance (SI) scheme preceding it, making it much more difficult to become eligible for and to stay on DI benefits. WIA introduced stricter entry criteria for DI and stronger incentives for work resumption, both for employees and employers.

Kantarcı et al. (2023) analyzed the effects of the WIA reform on labour participation and benefit receipt among long-term sick individuals (with and without partner) who report sick, i.e., are unable to perform their work because of occupational or nonoccupational illness or injury.[2] They found that the reform from transitional WAO to WIA substantially reduced the probability of DI receipt during the first ten years after the reform. They also found a rise in labour participation and in unemployment benefits receipt that adds up to almost half of the fall in DI receipt. The labour participation response was particularly strong for those who had a permanent contract when they fell sick and had more possibilities to go back to work than those who had a temporary contract or were unemployed. Since couples can pool income risk, spousal labour supply can be an insurance mechanism to compensate for the loss of DI benefits, particularly if the individual who fell sick does not manage to go back to work. Such a spousal response might be dampened if the spouse needs to provide care to the sick individual, not allowing her or him to do (more) paid work.

The current paper therefore focuses on whether spouses also responded to the reform – and how such a response varied depending on the labour market position of the individual who fell sick. Taking a difference-in-differences approach, we analyze the reform effects on both the individuals who fell sick and their spouses, focusing on heterogeneity of the effects: For individuals with a weaker initial labour market position, i.e., fewer opportunities to go back to work after recovery, there is a larger need for the spouse to compensate for the more stringent rules of the new DI system. Sick individuals who had a permanent work contract at the time of reporting sick increased labour participation, and indeed we find that their spouses did not respond. On the other hand, the fact that sick individuals who had a temporary work contract did not manage to increase labour participation, induced a substantial rise in their spouses labour participation. Similarly, if sick individuals had a weaker labour market position in the sense that they worked in a sector with a low vacancy rate or earned a low wage, the sick individuals themselves

[2]This concerns all types of health problems, such as virus infections, mental health issues, headaches, stomachaches, back pain, etc.

hardly responded but their spouses' labour participation rose substantially. These effects are persistent over a period of ten years after reporting sick. Findings for other outcomes (earnings, UI benefits) confirm that the spouse's response is larger in case of a weaker labour market position of the sick individual.

Finally, we compare with the reform effects of sick individuals who have a spouse with the effects on those who do not have a spouse. If there is no spouse who could compensate the loss of household income, the reform raises labour participation of sick individuals with a weak labour market position much more than if there is a spouse. This is in line with our main finding that in couples, the negative income effect of the DI reform is shared by the two partners – single sick individuals cannot rely on their partner to and make a greater effort themselves to resume work.

Our findings add to the limited evidence for the added worker, but also contribute to the literature on the impact of DI reforms. This literature analyses the effects of screening process and eligibility criteria (Karlström et al., 2008; De Jong et al., 2011; Staubli, 2011; Moore, 2015; Autor et al., 2016; Hullegie and Koning, 2018; Godard et al., 2022), benefit generosity (Gruber, 2000; Campolieti, 2004; Marie and Vall Castello, 2012; Mullen and Staubli, 2016; Deuchert and Eugster, 2019), and return-to-work incentives (Kostøl and Mogstad, 2014; Koning and van Sonsbeek, 2017; Ruh and Staubli, 2019; Zaresani, 2018, 2020). It also studies welfare effects (Low and Pistaferri, 2015; Deshpande, 2016; Fevang et al., 2017). None of these studies consider spillover effects on the spouse. Our findings suggest that for a complete evaluation of the DI reform, it is important to consider such spillover effects on both labour participation and the adequacy of household income.

At the intersection of the literature on the added worker effect and the impact of DI reforms are a few studies that analyze spousal labour supply responses when sick individuals receive DI benefits or eligibility rules for DI benefits change. Results of Duggan et al. (2010) suggest that reform in the US disability program for veterans increasing enrollment, somewhat reduced their wives' labour supply. Borghans et al. (2014) studied the impact of reassessing existing Dutch DI recipients and new applicants younger than 45 years based on new DI eligibility criteria. They found that affected individuals increased their earnings and social support income, but they found no significant effect on spousal earnings. Autor et al. (2019) analyzed the consequences of DI receipt for labour supply and consumption decisions in Norway. They showed that DI denial has little impact on income and consumption of married couples since spousal earnings and benefit substitution counteract the effect of denial of DI benefits. García-Mandicó et al. (2021) analyzed the impact of reassessment of earnings capacity under more stringent rules introduced in 2004. They found that earnings responses of the DI recipient and the spouse together almost fully compensate for the cut in DI benefits.

Our study differs from these studies in several respects. First, the DI reform in 2006

differs from the reforms studied earlier, restricting access to DI for a large group of workers, and therefore possibly leading to a stronger need for a spousal response. Moreover, due to the administrative nature of our data, we have enough statistical power to analyze heterogeneity in the response. This allows us to show that job security, employment opportunities, and earnings level of the sick spouse are important for the reform effects on both spouses. We also show that the added worker effect is evident for both wives and husbands whereas earlier studies focus on wives' responses to husbands' income shock. Moreover, unlike earlier studies, we also compare with singles, which helps to validating our finding that partners share the burden of the more stringent disability scheme after the reform.

This paper proceeds as follows. Section 1.2 explains the 2006 reform. Section 1.3 describes the data and the study sample. Section 1.4 gives descriptive evidence on the impact of the reform on spousal labour supply. Section 1.5 presents the empirical approach used to identify the effect of the reform. Section 1.6 discusses the results for couples and Section 1.7 compares with the reform effects on singles. Section 1.8 conducts some checks on the identifying assumptions. Section 1.9 concludes.

## 1.2 Disability insurance in the Netherlands and the 2006 reform

The current Dutch system of sickness and disability insurance (named WIA) protects against earnings loss due to incapacity for work and consists of a sickness scheme (SI) for the short term and a succeeding disability scheme for the long term. An individual working for an employer or receiving unemployment benefits (UI) who cannot work due to a health issue, reports sick and enters SI. While on SI, the development of the health condition and whether reintegration obligations are met are monitored by a certified company doctor from a private occupational health and safety firm. While the employer's responsibility lasts for the entire two-years period of sickness in case of a permanent contract, it lasts only until the contract ends in case of a temporary contract. Moreover, for workers employed through temporary work agencies, the contract ends as soon as the worker reports sick.

For workers with a permanent contract who report sick, the employer is obliged to pay at least 70% of the pre-sickness wage for a period of at most two years.[3] Workers with a temporary contract, those employed through a temporary work agency, and those who are entitled to UI have no employer to continue the wage payment and they are eligible to a sickness benefit of 70% of the pre-sickness wage from the Employment Insurance

[3]Most employers pay the full amount during the first year of sickness.

Agency. In fact, the Employment Insurance Agency takes over the role of the employer, both in paying benefits and facilitating reintegration.

Only after two years, a worker on SI can apply for the public DI benefit. Uniquely in the world, the Dutch DI scheme covers all causes of sickness, both occupational and non-occupational. During assessment for DI, a formal diagnosis is made and work limitations are determined. The loss of earnings is determined by comparing the pre-sickness wage to the potential wage accounting for the health condition of the sick worker defined as the median of the highest wages the sick worker could still earn in three jobs judged suitable. These jobs are selected from a representative sample of jobs in the Dutch labour market matching the job capabilities and limitations of the sick worker. The loss of earnings as a percentage of the pre-sickness wage is called the disability grade. A minimum disability grade of 35% is required to qualify for DI benefits; a disability grade of 80% or more defines full disability and implies the individual qualifies for full DI benefits. The public disability scheme is funded by employer's contributions, mostly flat rate, but partially also differentiated by DI risk ("experience rating").

Preceding the current scheme, the Disability Insurance Act (WAO) was introduced in 1967 to provide compulsory public insurance against loss of earnings due to long-term work incapacity, independent of the cause of the disability. It implied a period of at most one year on SI and a minimum disability grade of 15% for DI benefit eligibility. During the late 1970s and 1980s the number of DI beneficiaries rose rapidly to levels far beyond earlier expectations. Entry to the scheme was relatively easy because few reintegration incentives existed during sickness and DI applications were often accepted in case of doubt. Moreover, exit from the scheme was neither incentivized nor closely monitored. As a result, there was a substantial share of hidden unemployment in the DI scheme (Koning and van Vuuren, 2007).

Although major amendments were implemented in 1993, the WAO preserved its main features until 2006. The annual inflow rate into WAO rose to 1.5% of the insured working population in 2001, leading to further reforms. In April 2002 the "Gatekeeper Protocol" was introduced, in which clear and concrete mutual obligations of employers and sick employees for reintegration during the sickness period were specified. A "transitional WAO" scheme was introduced on 1 October 2004 for people who reported sick between 1 October 2003 and 1 January 2004, making entry criteria stricter. In particular, it adapted a broader definition of the work that the applicant could still do. For example, a sick part-time worker was now supposed to be able to accept a full-time job given the limitations of sickness, and a sick worker who did not speak Dutch was supposed to learn the language to qualify for jobs requiring understanding the Dutch language. As a result, the estimated wage loss due to disability was reduced, making it harder to reach the minimum disability grade to qualify for DI or to reach a higher disability grade (with a higher benefit).

The current Work and Income According to Labour Capacity Act (WIA) was introduced in 2006 for people who reported sick from 1 January 2004 onwards. It introduced major changes in both sickness and disability schemes, stimulating work resumption. It reduced the annual inflow rate into DI to 0.5% of the insured working population during the first six years after its introduction (Koning and Lindeboom, 2015).

WIA extended the duration of the sickness scheme from one to two years, implying an extension of two main incentives: First, the employer is obliged to compensate the employee for 70% of the wage loss during the additional year in the sickness scheme, creating a strong incentive for the employer to facilitate work resumption. Second, the Gatekeeper protocol was extended to a second year of sickness, strengthening the employers' sickness monitoring obligations (Hullegie and Koning, 2018).

WIA kept the stricter DI eligibility criteria of the transitional WAO scheme with the broader definition of what work can still be done. It introduced three other changes. First, the minimum disability grade for entering the scheme rose from 15 to 35 percent – workers with limited disability are expected to resume working (with adaptations of their work if necessary) or apply for UI. Second, it introduced a work resumption program providing strong financial incentives for partially disabled people to utilize their remaining work capacity. Third, experience rating for employers was extended from 5 to 10 years, implying that employers incurring disability costs are penalized with higher DI premiums for up to five additional years. At the same time experience rating was restricted to temporarily or partially disabled workers but abolished for permanently and fully disabled workers. Targeting the former group made experience rating more effective since the partially or temporary disabled have better prospects of reintegration. Experience rating was limited to permanent work contracts until 2013 and extended to temporary contracts afterwards. Figure 1.4 presents a timeline of detailed changes in the DI scheme starting from the introduction of the WAO in 1967 until the extension of experience rating to temporary contracts in 2013.

For the income of the sick individuals during sickness and disability periods, potential implications of the WIA reform are as follows. In the first year of sickness, wages are not affected by the reform. However, employers may already do more for reintegration in the first year of sickness if they anticipate the cost of the additional year of wage payments. These stronger employer incentives may induce sick individuals to return to work, especially in combination with the requirements of the Gatekeeper protocol. On the other hand, reintegration incentives for employees might have become weaker in the first year since employees are no longer subject to a DI assessment after one year of sickness.

In the second year of sickness, WIA requires that the employer replaces (at least) 70% of the former wage. In WAO, DI and UI benefits together replaced 70% of the former wage. From the third year onwards, a potential fall in income is due to lower or a complete

loss of DI benefits. As described above, this owes to the stricter eligibility criteria for DI, financial incentives for work resumption, and extended and more targeted reintegration incentives of experience rating. Note that these implications of the reform assume that the sick individual has a stable work contract with an employer. Employees with weak employer relationships or those who are unemployed will lack the reform incentives and may struggle to resume work and cope with the negative income shock of the reform. They may seek alternative welfare benefits, or rely on the income of their spouse.

## 1.3 Data

We use unique administrative data from the Employee Insurance Agency on all individuals who fell sick in the fourth quarter of 2003 or the first quarter of 2004, and therefore could become eligible to either the transitional WAO or the WIA scheme. We observe the beginning and ending dates of their sickness, their gender, date of birth, and sector of economic activity. They either earned a wage or receive UI at the time they report sick – other groups cannot enter the sickness scheme. For wage earners, we observe whether they had a permanent contract, a temporary contract, or a contract through a temporary work agency at the time they reported sick. We link these individuals to administrative data on themselves and their partners (married or cohabiting) from Statistics Netherlands (CBS), with monthly information on wages and benefits (including DI and UI) from January 1999 to February 2014.

The initial data set has 171,281 individuals reporting sick. To select the estimation sample, we drop individuals who participate in the special disability schemes for the self-employed or for young people, since the rules and incentives for them are quite different. We also drop individuals who already received DI when they reported sick. We drop individuals in same-sex partnerships and only keep couples if their cohabitation started before reporting sick. We drop individuals whose spouse also reported sick between October 2003 and March 2004. Finally, we only keep those who spent more than 90 days in sickness leave, since employers only have to report sickness cases if they last longer than 90 days.[4] We divide the sample into a "control group" of individuals (and their spouses) who fell sick in the fourth quarter of 2003 and were insured under the transitional WAO scheme and a "treatment group" of individuals (and their spouses) who reported sick in the first quarter of 2004 and were insured under the WIA scheme. We will not consider individuals who reported sick before October 1 2003 and fall under the old WAO scheme and refer to the transitional WAO group as WAO group from now on.

Based on the available data on wages and social security benefits, we define the following outcome variables: dummies that indicate labour participation and UI receipt,

[4] temporary work agencies have to report all sickness cases.

and the monthly amounts of wages and UI benefits. We transform earnings and benefit amounts as the natural logarithm of the amount plus 1, accounting for the skewed distribution and the zero values. During participation in the sickness scheme, the observed wage combines two types of payments: earnings (for the part of work capacity that is still used) and compensation for lost earnings due to sickness benefits paid by the employer. We do not observe the separate amounts. Since we measure labour participation as positive earnings, this implies that we cannot determine whether or not sick people are working when in the sickness scheme. We therefore discard the first two years after individuals reported sick in most of our analysis. After the first two years, SI expires for everyone and measuring labour force participation is no longer problematic.

## 1.4 Time trends and other descriptive statistics

Figure 1.1 shows the labour participation rates and fractions of DI and UI recipients in control and treatment groups over the observation period.[5] For the individuals in our data who all reported sick, DI benefit receipt increases sharply when they become eligible for DI benefits and continues to increase during the remaining years of the observation period. The WIA group is 3 pp less likely to receive DI benefits than the control group (13.5% versus 10.5%) and the difference between the two groups remains stable till the end of the observation period. This shows that the reform effectively limited access to DI benefits. For the spouses of sick people, DI receipt is stable and not affected by the reform (as expected).

For individuals who reported sick, the probability of working shows a strong time trend that is common to both groups. It increases until the date individuals report sick, reflecting that individuals can enter the sickness scheme only if they are working or receive UI. Before this, they can have another labour force status. The probability of working falls sharply during the first few years of sickness and continues to fall throughout the remaining years. The difference between WAO and WIA groups is small and insignificant before individuals fall sick, but notable and significant after that, suggesting that the reform increased labour participation of those who fell sick. For spouses, the probability of working shows a less pronounced decreasing pattern. The difference between groups is smaller before than after treatment, which would be in line with a positive spillover effect, but these differences are not significant.

For sick individuals in both groups, the use of UI falls sharply right after reporting sick, since those who are unemployed replace UI with sickness benefits. UI use rebounds and increases during the remaining months of the sickness scheme, since many individuals recover and replace their sickness benefit with UI. UI use peaks when individuals can apply

[5]Similar figures for wages and benefit amounts (not shown) reveal very similar patterns.

for DI, since rejected DI applicants turn to UI when the sickness period ends. UI use falls during the disability period because UI is temporary with a maximum of 38 months. The difference between the control and treatment groups is sizable and statistically significant during the disability period, suggesting that the DI reform increased UI use among those who reported sick. UI use among the spouses is fairly constant over time. The difference between control and treatment groups is insignificant, both pre- and post-treatment.

Table 1.1 presents sample means of some background characteristics when reporting sick for both groups, as well as outcomes before and after reporting sick. It also presents tests for equality of the means in control and treatment groups ("balancing tests"). Panel A shows that, in both groups, the average age is about 43 and there are more men than women. The majority held a permanent work contract when they fell sick; the others had a temporary contract or a contract through a temporary work agency, or were unemployed. Column 3 shows that there are small but significant differences between the treatment and control group. These possibly reflect labour market trends. Our identification strategy (difference-in-differences) accounts for such differences.

Columns 3 and 6 in panel B present mean differences in outcomes during the pre- and post-treatment periods for treatment and control group. The fraction of sick individuals receiving disability benefits falls due to the reform, as expected. In line with Figure 1.1, the difference is larger post- than pre-treatment for all outcomes, again suggesting that the reform has increased labour participation, average earnings, UI receipt, and the average amount of UI benefits.

Table 1.2 reproduces Table 1.1 for the spouses. Spouses in the treatment group are slightly older than ithe control group. Couples in the treatment group have cohabited somewhat longer pre-treatment but not post-treatment. Columns 3 and 6 in panel B show that the difference in labour participation between groups is larger post-treatment than pre-treatment, which, again, might suggest that the reform increased labour participation for the spouses. The mean differences in other outcomes are small and insignificant, both pre- and post-treatment.

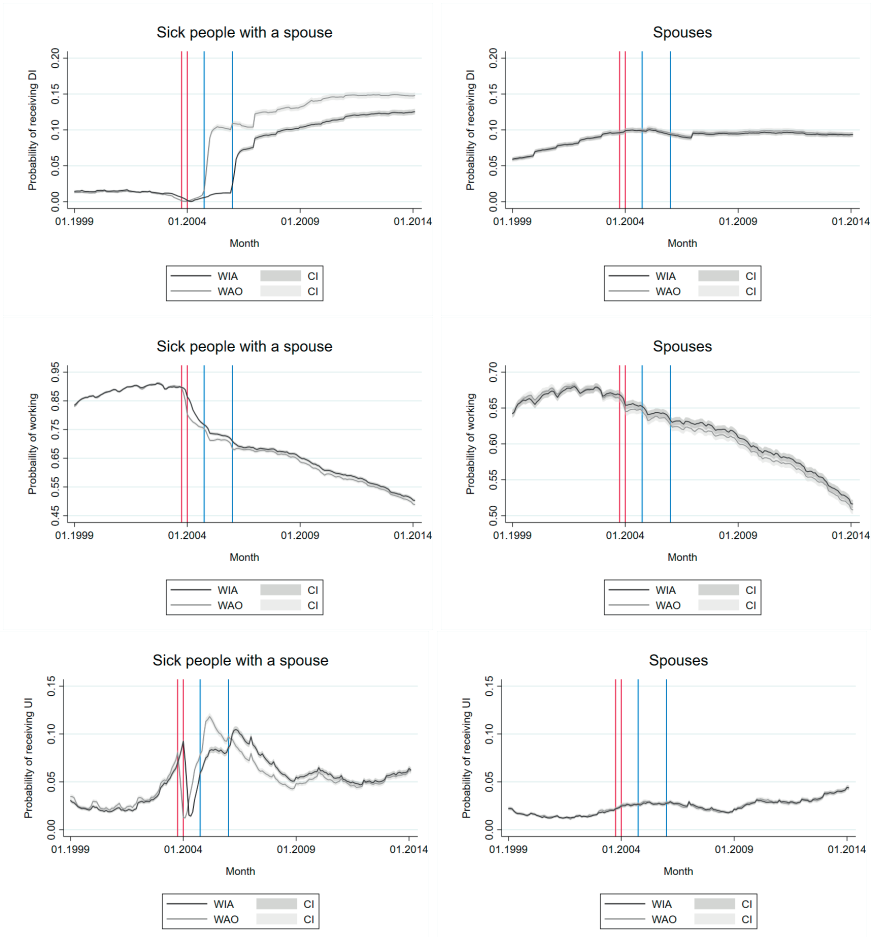



Figure 1.1: Probability of DI receipt, working, and UI receipt for control and treatment groups by calendar month; sick individuals (left) and their spouses (right). Vertical lines mark the first instance sick individuals can be entitled to sickness (red) and disability (blue) benefits.

Table 1.1: Sample means and balancing tests of background characteristics and outcome in control and treatment groups before and after sickness for sick individuals with a partner

| | Before | | | After | | |
|---|---|---|---|---|---|---|
| | WAO group | WIA group | Dif. WIA and WAO | WAO group | WIA group | Dif. WIA and WAO |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| A. Background characteristics | | | | | | |
| Age | 42.946 | 43.108 | 0.162* | | | |
| Female | 0.379 | 0.384 | 0.005 | | | |
| Permanent contract | 0.705 | 0.717 | 0.012*** | | | |
| Temporary contract | 0.123 | 0.106 | -0.017*** | | | |
| Unemployed | 0.171 | 0.176 | 0.005* | | | |
| B. Labour market outcomes | | | | | | |
| DI (possibly UI) receipt | | | | 0.135 | 0.106 | -0.029*** |
| Labour participation | 0.887 | 0.889 | 0.002 | 0.611 | 0.615 | 0.004 |
| UI (no DI) receipt | 0.033 | 0.033 | 0.000 | 0.057 | 0.063 | 0.006*** |
| DI (and possibly UI) per month | | | | 187.298 | 162.690 | -24.608*** |
| Wage per month | 1,999.228 | 2,012.334 | 13.106 | 1,713.940 | 1,738.323 | 24.383* |
| UI (excl. DI) per month | 39.938 | 40.128 | 0.190 | 89.747 | 92.873 | 3.126* |
| Observations | 1,589,880 | 1,716,480 | | 2,543,808 | 2,746,368 | |
| Individuals | 26,498 | 28,608 | | 26,498 | 28,608 | |

Notes: 1. "Before": period before individuals fall sick (January 1999 - October 2003 for individuals who fell sick in November 2003; January 1999 - January 2004 for individuals who fell sick in February 2004). "After": period after individuals fell sick excluding the first two years (November 2005 - January 2014 for individuals who fell sick in November 2003; February 2006 - January 2014 for individuals who fell sick in February 2004). 2. Age is at the time individuals fall sick. "Permanent contract", "temporary contract", and "unemployed" refer to labour market status of individuals when they fell sick. 3. Columns 1, 2, 4 and 5 present means in control (WAO) and treatment (WIA) group before and after start of sickness. Columns 3 and 6 present differences between individuals insured under WIA and WAO – the estimated coefficient from the regression of the characteristic or outcome as the dependent variable, and an indicator of participation in WIA as the explanatory variable. Standard errors clustered at the individual level.

Table 1.2: Sample means and balancing tests of background characteristics and the outcome in control and treatment before and after sickness for spouses

| | Before | | | After | | |
|---|---|---|---|---|---|---|
| | WAO group | WIA group | Dif. WIA and WAO | WAO group | WIA group | Dif. WIA and WAO |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| A. Background characteristics | | | | | | |
| Age | 42.333 | 42.535 | 0.205** | | | |
| Years of cohabitation | 7.031 | 7.173 | 0.142*** | 8.269 | 8.224 | -0.045 |
| B. Labour market outcomes | | | | | | |
| DI (possibly UI) receipt | | | | 0.096 | 0.094 | -0.002 |
| Labour participation | 0.670 | 0.669 | -0.001 | 0.586 | 0.590 | 0.004 |
| UI (no DI) receipt | 0.016 | 0.016 | 0.000 | 0.027 | 0.028 | 0.001 |
| DI (and possibly UI) received per month | | | | 115.472 | 113.790 | -1.682 |
| Wage per month | 1,299.714 | 1,312.522 | 12.808 | 1,552.520 | 1,561.632 | 9.112 |
| UI (excl. DI) per month | 18.356 | 19.220 | 0.864 | 38.291 | 38.761 | 0.470 |
| Observations | 1,589,880 | 1,716,480 | | 2,543,808 | 2,746,368 | |
| Individuals | 26,498 | 28,608 | | 26,498 | 28,608 | |

Notes: 1. "Before": period before individuals fall sick (January 1999 - October 2003 for individuals who fell sick in November 2003; January 1999 - January 2004 for individuals who fell sick in February 2004). "After": period after individuals fell sick excluding the first two years (November 2005 - January 2014 for individuals who fell sick in November 2003; February 2006 - January 2014 for individuals who fell sick in February 2004). 2. Age is at the time individuals fall sick. Years of cohabitation for the "Before" period indicates mean years of cohabitation by the time individuals fall sick. That for the "After" period indicates mean years of cohabitation during the period after individuals fell sick including the first two years. 3. Columns 1, 2, 4 and 5 present means in control and treatment before and after start of sickness. Columns 3 and 6 present differences between individuals insured under the WIA and WAO – the estimated coefficient from the regression of the characteristic or outcome as the dependent variable, and an indicator of participation in the WIA as the explanatory variable. Standard errors clustered at the individual level.

## 1.5 Identification strategy

We use a difference-in-differences approach to identify the causal effect of the WIA reform on each outcome variable $y_{it}$, either concerning the sick individual or the spouse. The first difference is across groups. Those who reported sick in the first quarter of 2004 (treatment or WIA group) face different eligibility criteria and incentives to work or claim benefits than individuals who reported sick in the fourth quarter of 2003 (control or WAO group). The second difference refers to event time: before and after reporting sick.

We start the DiD comparison using the following baseline regression model:

$$y_{it} = \alpha_i + \gamma \ (Treated_i \times Post_t) + \delta \ Post_t + \lambda_{s(i,t)} + \varepsilon_{it}. \tag{1.1}$$

Here $i$ indexes the sick individual or their spouse. $t$ indexes the month of event time: Values –57 to –1 indicate the months before reporting sick, 0 is the month when first reporting sick, and 1 to 119 are the months after reporting sick. (For some outcomes $y_{it}$, we do not use observations during the sickness period due to measurement issues; see Section 1.3). $\lambda_{s(i,t)}$ is a monthly calendar time effect – $s(i,t)$ indexes the calendar month (from January 1999 until February 2014; January 1999 is chosen as the base month) for individual $i$ at a given month of event time $t$. $\alpha_i$ is an individual-specific, time-invariant fixed effect that is potentially correlated with the control variables. $\varepsilon_{it}$ represents an idiosyncratic (unobserved) shock, assumed to be uncorrelated with all the explanatory variables.

$Treated_i$ is a dummy variable for the treatment (WIA) group.[6] $Post_t$ is an event time dummy with value 1 from the start of the sickness period. The individual effects capture differences between the two groups other than the reform effect. Under the identifying assumption that treatment and control group would have followed the same trend if there would not have been a reform, the coefficient $\gamma$ on the interaction term $Treated_i \times Post_t$ captures the effect of the reform, the parameter of interest.[7]

To disentangle the effect of the WIA reform in the short and long run, and test for the common trend assumption, we consider the following extended model:

$$y_{it} = \alpha_i + \sum_{l=-5}^{9} \gamma_l \ (Treated_i \times d_{lt}) + \sum_{l=-5}^{9} \delta_l d_{lt} + \lambda_{s(i,t)} + \varepsilon_{it}. \tag{1.2}$$

Instead of $Post_t$ which refers to the entire period after falling sick, this model has separate dummies for each year, after and before falling sick. $d_{lt}$ indicates the $l$-th year from the time the individual reports sick. Year $-1$ is chosen as the base year. The coefficients on

[6] Since this is time invariant, it is omitted in the fixed effects regression.

[7] We cannot separately identify the effects of the different components of the reform, i.e. the extension of the sickness period, changes in financial incentives, and stricter eligibility criteria.

the interaction terms of treatment and these year dummies are the estimated treatment effects.[8] For the years before reporting sick, they provide a test of the common trend assumption. For the period after reporting sick they reflect the dynamic effects of the reform. In this setup, treatment and control groups are compared over event time $t$, i.e., the months before and after the individual reported sick. The calendar time dummies $\lambda_s(i,t)$ on the other hand capture the (common) calendar time trend.

In Section 1.6.3, we allow for heterogeneous reform effects depending on the labour market status at the time the individuals reported sick. In particular, we hypothesize that the effects depend on how easy is for the sick individuals to go back to work (either to their old job or to a new one). In Section 1.7, we also apply the same model to individuals without a spouse who reported sick. This is to investigate whether the sick individuals behave differently if there is a spouse who can potentially respond to the reform by increasing labour supply and household income.

To control for observed differences between treatment and control individuals before reporting sick, we apply entropy balancing following Hainmueller (2012). In particular, individuals are weighted to adjust inequalities in representation with respect to the first moment of the covariate distributions. As covariates, we consider their gender and birth year, as well as all outcomes of the sick individuals and their spouses before the first group reported sick. Regressions of equations (1.1) and (1.2) are estimated based on the constructed weights.[9] The weights are regenerated in each subsample when analyzing heterogenous treatment effects. To check whether, after entropy balancing, the common trend assumption is satisfied pre-treatment and to analyze several other threats to our identification strategy, we perform additional analyses and robustness checks in Section 1.8.

## 1.6 The effect of the reform on labour participation of sick individuals and their spouses

We first present the effects for the whole post-treatment period (equation (1.1)), then analyze the short- and long-run effects of the reform (equation (1.2)), and finally check for heterogeneous effects. In the main text, we focus on the reform from transitional WAO (often referred to as WAO for convenience) and WIA. To further substantiate our main conclusion about the added worker effects, we repeated some of the analysis for the reform from (original) WAO to transitional WAO three months earlier. These results are

[8] Here we also include observations for $t = 0, \ldots, 23$.

[9] Following Imbens (2004) and using propensity scores to construct weights leads to almost identical estimates.

presented in Table 1.9 in the Appendix.

### 1.6.1 Baseline effects

Table 1.3 presents the baseline DiD estimates of the reform effects on labour participation and benefit receipt. For the sick individuals, the reform decreased the probability of DI receipt by 3.1 percentage points (pp) on average during the post-treatment period (excluding the first two years). It increased the probability of working by 1 pp and UI receipt by 0.8 pp.[10] The reform induced the spouses of the sick individuals to raise their labour participation by 0.5 pp, one sixth of the drop in DI receipt of the individuals who reported sick, but this effect is significant at the 10% level only.

The center panel of Table 1.3 shows the DiD estimates of the reform effects on monthly wages and benefits (in log of the amount plus 1). The reform reduced monthly DI benefits of individuals reporting sick by 20.3% and increased monthly earnings by 7.6% and monthly unemployment benefits by 5.6%. Moreover, it increased earnings of spouses by 3.6%, but this increase is not statistically significant.

The lower panel of Table 1.3 presents the DiD estimates of the reform effects for monthly total income (in log of the amount plus 1), pooling monthly wages and social security benefits (DI, UI and social assistance). It presents the total income of sick individuals and their spouses at the individual level, but also total income of the couple. On average, the reform did not significantly affect total income of sick individuals, their spouses, or their household. The first suggests that in most cases, sick individuals are able to compensate lost disability benefits by increasing earnings and income from UI. This may explain why the spouses' income increase is not that large (or significant) either – there is not much need for a spousal response. In the heterogeneity analysis in Section 1.6.3, however, we will find that this aggregate result is mainly due to groups of individuals reporting sick who relatively easily can go back to work and increase their earnings. The aggregate findings fail to show that for individuals reporting sick and who are less likely to compensate lost disability benefits by responding themselves, spousal earnings do respond.

### 1.6.2 Dynamic effects

Figure 1.2 presents the estimates of the reform effects separately for for ten years of the post-treatment period. For individuals reporting sick, the reform reduced DI receipt by about 3 pp from the third year after reporting sick, when both the treatment and

[10]This is qualitatively in line with what Kantarcı et al. (2023) found. The magnitude of the effect on DI receipt is different, mainly because we take everyone who has been sick for at least 90 days whereas Kantarcı et al. only consider those who have been sick for 180 days; see the sensitivity analysis in their Appendix B.

the control group can apply for DI. (Note that the large effect of the reform on labour participation during the first year of the sickness scheme is hard to interpret, due to the measurement issue explained in Section 1.3.) The effect on labour participation then falls to about 1 pp and remains fairly stable. For UI receipt, the large negative effect in the second year of the sickness scheme is due to the fact that individuals insured under WIA are still entitled to sickness wage payment if there is an employer, or the sickness benefit if there is no employer. From the third post-treatment year onwards, however, the reform has a positive effect on UI receipt. It falls over time and becomes insignificant from year 8 after reporting sick.[11] In each year after reporting sick, the effect of the reform on spousal labour participation is about 0.5 pp, but this is never significant. Moreover, in line with the exploratory analysis in Section 1.4, the reform has no significant effect on spouses' UI receipt.

### 1.6.3 Heterogeneous effects

Analyzing wives' labour supply response to an exogenous shock to husband's job earnings in Austria, Halla et al. (2020) find that the added worker effect for wives is almost negligible if computed for their full sample. To understand the reasons for this and to identify impediments to the intrahousehold insurance mechanism, they investigate heterogeneity in responses for different types of households and find significant added worker effects for several subgroups. Similarly, as shown in Table 1.3, we also found a small added worker effect for the sample as a whole. This finding, however, may mask interesting heterogenous effects.

Existing studies on the added worker effect discuss a variety of factors that could induce wives to respond to a negative shock on their husbands' earnings. Some consider the nature of the income shock and argue that spouses would respond if the income shock is permanent, unanticipated, or if its magnitude is large (Cullen and Gruber, 2000; Stephens, 2002; Blundell et al., 2016; Bredtmann et al., 2018; Fadlon and Nielsen, 2021). Others consider that lack of self-insurance through savings or formal insurance through social support programs, high earnings potential of the wife, and existence of job opportunities for wives may encourage wives to respond (Cullen and Gruber; Bentolila and Ichino, 2008; Blundell et al.; Halla et al., 2020).

Our heterogeneity analysis starts from the notion that the need for a spousal response in order to maintain family income depends on the extent to which the individual reporting sick is unable to respond him- or herself. We hypothesize that when the labour market position of a sick worker facing the reform is weak, the sick individual's response will be weaker and the spousal response will be stronger. The heterogeneity analyses we perform

[11] Again, these results are qualitatively in line with those of Kantarcı et al. (2023), cf. their Figure 4.

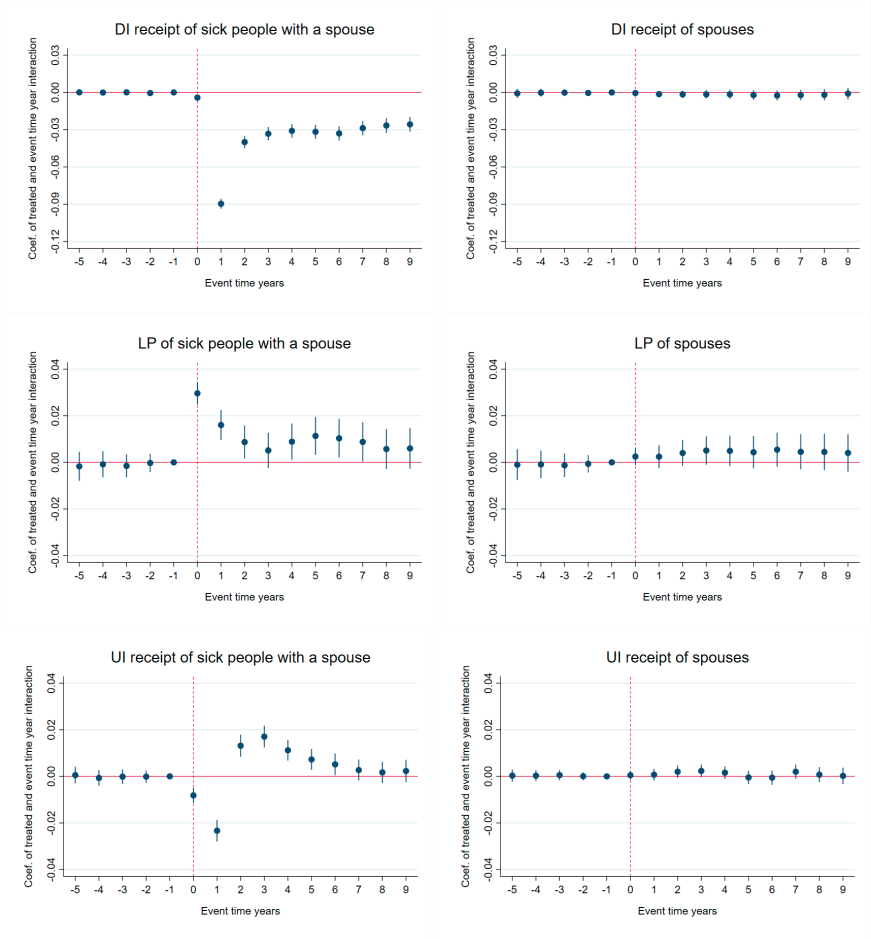



Figure 1.2: Estimated treatment effects in each of the five years before reporting sick and in each of the first ten years after reporting sick, with 95 percent confidence intervals. Observed differences between treatment and control individuals before reporting sick are controlled for using entropy balancing.

essentially distinguish in three different ways among groups of individuals reporting sick that differ in their chances to go back to work. We explore three indicators of a weak labour market position: employment status just before reporting sick (permanent job, temporary job, or unemployed), earnings level before reporting sick, and the sectoral vacancy rate in the year of reporting sick.[12] Figures 1.5 to 1.7 in the Appendix present the dynamic effects by employment status, vacancy rate, and earnings quartile for sick people with a spouse and their spouses, and also sick people without a spouse (discussed in Section 7). The pre-reform (left hand) parts of these figures show negligible and insignificant pre-reform effects, supporting the common trend assumption for each subgroup.

## Employment status just before reporting sick

If sick individuals have temporary work contracts, they may not be able to go back to their job after recovery or find another job. Similarly, workers who reported sick while unemployed may have trouble finding a job when their sickness benefit expires. Furthermore, stimulating employers to increase labour market participation has been a key element of Dutch labour market reforms throughout the years. The WIA reform in particular introduced strong reintegration incentives for employers (see Section 1.2), but these incentives do not apply uniformly to all workers: Employer incentives for temporary workers only last for the duration of the employment contract. Moreover, temporary work agencies do not face incentives for their sick employees during sickness, since their sickness benefits are paid by the Employee Insurance Agency. Unemployed individuals obviously do not benefit from positive effects of employer incentives either. On the other hand, employers of employees with a permanent contract are fully incentivized due to continued wage payments and experience rating, which applied only to permanent work contracts until 2013. Prinz and Ravesteijn (2020) found that the extension of experience rating to temporary workers reduced DI receipt by 12.7 pp and increased labour participation by 2.5 pp among workers with temporary contracts relative to those with permanent contracts. In summary, there are several reasons why those who had a temporary contract or were unemployed when falling sick have more problems to go back to paid work after recovery and more often have to cope with a negative income shock due the reform. They may then also more often have to rely on their spouse's income, and their spouse may respond more strongly if DI benefits are lost due to the reform.

We separately estimate the reform effects for individuals who were wage earners with a permanent contract, wage earners with a temporary contract, or unemployed just before

[12]The literature on the added worker effect typically focuses on the wife's response to shocks in the husband's income. In contrast to García-Mandicó et al. (2021), who analyzed the impact of the change in reassessment rules in 2004, we found hardly any differences between the spousal effects for men and women.

they started receiving sickness benefits. Since changing jobs takes time and involves adjustment costs (Zaresani, 2020), this makes selection into employment situation before falling sick due to the onset of the health problem very unlikely. On the other hand, as suggested by Koning and Lindeboom (2015), employers might respond to the reform by hiring fewer high-risk workers on a permanent basis. While this might play a role in the longer run, the data suggest that this issue is not relevant for the cohorts under study that fall sick just before and just after the reform. In fact, Table 1a shows that the post-reform WIA group has slightly more people with a permanent contract and fewer with a temporary contract than the pre-reform WAO group, the opposite of what selection would predict. Our results therefore are unlikely to be biased by different selection into temporary and permanent jobs for the pre- and post-reform groups.

The upper panel of Table 1.4 presents the results, which are largely in line with the hypotheses formulated above. Due to the reform, DI receipt fell substantially for all groups, and the largest fall is for the unemployed, who have no employer that can help them resume work. This increases their chances of remaining in the sickness scheme and facing the stricter requirements of WIA to enter DI.

The reform only increased labour participation among those who reported sick when they had a permanent work contract, even though the fall in DI receipt is larger for the other two groups. It suggests that the reform's work resumption incentives induced employers to reintegrate their permanent employees, but were not effective for temporary contracts or unemployed workers. For the unemployed, the longer sickness period may also lead to more human capital loss or a stronger scarring effect, reducing the prospects of finding a job (Arulampalam, 2001; Arulampalam et al., 2001). Moreover, their incentives to resume work quickly may be reduced by the additional year they can spend in the sickness scheme.

The reform increased UI receipt for all sick individuals, irrespective of their work status. The increase is largest for the unemployed, where UI is usually the primary source of income. The effect for those on a temporary contract is larger than for those on a permanent contract – the former have lower and less stable earnings, and seek additional income from UI if the reform blocks access to DI benefits.

Since sick individuals with a permanent work contract often resume work and increase earnings themselves, their spouses less often need to compensate, and indeed, the labour participation response of the spouses in this group is small and insignificant. On the other hand, the spouses of sick individuals on a temporary contract increased labour participation and earnings significantly and Figure 1.5 in the Appendix shows that this effect is persistent.[13] Since sick individuals with a temporary contract struggle to resume working, this confirms that their spouses increase labour participation and earnings to

[13]The righthand panel in this figure will be discussed in Section 7.

compensate for the lost disability benefits and lack of labour income. This added worker effect on labour participation is particularly large – 81% of the drop in DI receipt. The spouses of sick individuals who were unemployed also increase their labour participation by a notable amount of 1 pp, but this estimate is less precise and not significant.

The signs and significance levels of the estimated effects of the reform on earnings and benefit amounts in the upper panel of Table 1.5 are in line with the estimated effects on labour participation and benefit receipt. The upper right panel of Table 1.5 presents the effects of the reform on total income (in log of the amount plus 1) of sick individuals and their spouses at the individual and the household level. Individual income of sick individuals who had a temporary contract or were unemployed fell significantly due to the reform. However, due to the positive responses of their spouses, household income does not change significantly, confirming that spouses' responses help to smooth household income. This result is in line with the findings of Blundell et al. (2016) based on a structural family labour supply model where households self-insure through spousal labour supply in case of a negative income shock. For sick individuals with a permanent contract, total income did not change significantly at either the individual or the household level.

## Pre-sickness earnings

If sick individuals earn low wages (regardless of their sickness), spouses may respond more strongly for different reasons. Low wage earners tend to have smaller savings or wealth to draw on during sickness to smooth their consumption path. They also tend to work in jobs where prospects of recovery from ill health are limited. If as a result the income shock becomes permanent, spouses may exhibit stronger responses. Furthermore, the likelihood of receiving DI benefits depends on the wage earned before reporting sick (Section 1.2). For people with low earnings, the income decline due to disability is often limited. As a result, low-wage earners are much more likely to be denied DI benefits than higher wage earners (OCTAS, 2023). The workforce hit by the DI reform therefore includes many low-wage earners who are more likely to rely on a spousal response to maintain household income. The pre-sickness earnings measure we use is the average of the individual's earnings during the five years before they reported sick (where data is available). The center panel of Table 1.4 presents the estimation results by pre-sickness earnings quartile. Sick individuals in lower earnings quartiles increased their UI receipt somewhat more, in line with the argument that they struggle more to find suitable jobs where they can utilize their remaining work capacity and more often have to rely on income from UI. For the lowest two quartiles, we find that spouses notably increase their labour participation, a response which is more than half (52%) of the drop in DI receipt of the sick partner. For the higher quartiles, however, no spousal response is observed.

These results are in line with the results based on employment status in the preceding subsection – both suggest that spousal responses are stronger for sick individuals in a weaker labour market position. Figure 1.6 in the Appendix suggests that the pattern is persistent over time, but the precision of the estimates is limited.

## Vacancies in the sector

For sick individuals who have limited employment opportunities and hence a higher risk of unemployment, responding to the work incentives of the DI reform can be more difficult and spousal labour supply responses can be stronger. We consider the sectoral vacancy rate (the number of open vacancies per one thousand jobs) as an indicator of employment opportunities.[14] We distinguish two groups: individuals who at the time of reporting sick worked in sectors with vacancy rates below (e.g., construction, manufacturing, transport, public sector) or above the average vacancy rate (e.g., agriculture, trade, financial services, catering).

The lower panels of Tables 1.4 and 1.5 present the results. If sick individuals work in a sector with a vacancy rate below the average, their spouses increase labour participation by 0.9 pp, a sizable extensive margin added worker effect of 23% of the drop in DI receipt by the sick partner who has limited employment opportunities him or herself. In contrast, if sick individuals work in a sector where the vacancy rate is above the average, the reform raises their own labour participation by 1 pp while their spouses do not respond significantly. Again, these results confirm that the added worker effect is a more powerful insurance mechanism when the labour market position of the sick individual affected by the reform is weak.

[14]The sector where sick individuals are or were employed is available in the sickness data, and we determine the vacancy rate in each sector prior to and in the year of reporting sick using data from Statistics Netherlands.

Table 1.3: Estimated effects of the WIA reform: Sick individuals and their spouses

| | Sick individual | Spouse |
|---|---|---|
| DI receipt | -0.031*** | -0.001 |
| | (0.002) | (0.002) |
| Labour participation | 0.010*** | 0.005* |
| | (0.003) | (0.003) |
| UI receipt | 0.008*** | 0.000 |
| | (0.001) | (0.001) |
| ln DI | -0.203*** | -0.011 |
| | (0.018) | (0.012) |
| ln Wage | 0.076*** | 0.036 |
| | (0.027) | (0.023) |
| ln UI | 0.056*** | 0.003 |
| | (0.009) | (0.006) |
| ln Total individual income | -0.010 | 0.016 |
| | (0.023) | (0.022) |
| ln Total household income | 0.009 | |
| | (0.018) | |
| Observations | 8,431,218 | |
| Individuals | 55,106 | |

Notes: ***, **, * denote statistical significance at 1, 5 and 10 percent, respectively. Standard errors (in parentheses) account for heteroskedasticity and clustering at the individual level. All specifications control for individual and calendar month fixed effects. The regressions use data available for the whole pre-treatment period but exclude data for the first two years of the post-treatment period.

Table 1.4: Estimated effects of the WIA reform on DI receipt, labour participation and UI receipt of sick people and their spouses by labour market status, quartiles of average earnings before reporting sick, and sectoral vacancy rate

| | DI receipt | | Labour participation | | UI receipt | |
|---|---|---|---|---|---|---|
| | Sick individual | Spouse | Sick individual | Spouse | Sick individual | Spouse |
| On permanent contract | -0.027*** | -0.002 | 0.016*** | 0.001 | 0.004*** | 0.002** |
| | (0.002) | (0.002) | (0.004) | (0.004) | (0.001) | (0.001) |
| On temporary contract | -0.031*** | -0.008 | -0.015 | 0.025*** | 0.013*** | -0.002 |
| | (0.009) | (0.006) | (0.012) | (0.009) | (0.004) | (0.003) |
| Unemployed | -0.041*** | 0.003 | -0.014 | 0.010 | 0.021*** | -0.004 |
| | (0.008) | (0.005) | (0.009) | (0.008) | (0.005) | (0.003) |
| Earnings in 4th quartile | -0.037*** | -0.001 | 0.013* | -0.004 | 0.007*** | -0.000 |
| | (0.004) | (0.003) | (0.007) | (0.006) | (0.002) | (0.002) |
| Earnings in 3rd quartile | -0.036*** | -0.001 | 0.004 | -0.002 | 0.007*** | 0.002 |
| | (0.005) | (0.004) | (0.006) | (0.006) | (0.002) | (0.002) |
| Earnings in 2nd quartile | -0.024*** | -0.001 | 0.010 | 0.012* | 0.008*** | 0.002 |
| | (0.005) | (0.004) | (0.007) | (0.006) | (0.002) | (0.002) |
| Earnings in 1st quartile | -0.025*** | -0.000 | 0.006 | 0.013** | 0.011*** | -0.001 |
| | (0.006) | (0.004) | (0.007) | (0.006) | (0.003) | (0.002) |
| Vacancy rate above the mean | -0.022*** | -0.002 | 0.010** | 0.002 | 0.008*** | 0.001 |
| | (0.003) | (0.002) | (0.005) | (0.004) | (0.002) | (0.001) |
| Vacancy rate below the mean | -0.039*** | -0.000 | 0.008* | 0.009** | 0.008*** | 0.000 |
| | (0.004) | (0.003) | (0.005) | (0.004) | (0.002) | (0.001) |

Notes: ***, **, * denote statistical significance at 1, 5 and 10 percent, respectively. Standard errors (in parentheses) account for heteroskedasticity and clustering at the individual level. All specifications control for individual and calendar month fixed effects. The regressions use data available for the whole pre-treatment period but exclude data for the first two years of the post-treatment period.

Table 1.5: Estimated effects of the WIA reform on DI benefits, wages, UI benefits and total income of sick people and their spouses by labour market status, quartiles of average earnings before reporting sick, and sectoral vacancy rate

| | ln DI | | ln Wage | | ln UI | | ln Total individual income | | ln Total household income |
|---|---|---|---|---|---|---|---|---|---|
| | Sick individual | Spouse | Sick individual | Spouse | Sick individual | Spouse | Sick individual | Spouse | |
| On permanent contract | -0.183*** | -0.014 | 0.130*** | 0.004 | 0.027*** | 0.015** | 0.029 | -0.006 | 0.012 |
| | (0.018) | (0.014) | (0.031) | (0.027) | (0.009) | (0.007) | (0.027) | (0.026) | (0.021) |
| On temporary contract | -0.196*** | -0.057 | -0.126 | 0.182** | 0.091*** | -0.016 | -0.180** | 0.090 | -0.017 |
| | (0.067) | (0.041) | (0.088) | (0.007) | (0.030) | (0.019) | (0.075) | (0.071) | (0.054) |
| Unemployed | -0.265*** | 0.015 | -0.126* | 0.076 | 0.148*** | -0.030 | -0.145** | 0.070 | -0.040 |
| | (0.059) | (0.035) | (0.070) | (0.059) | (0.033) | (0.018) | (0.062) | (0.059) | (0.049) |
| Earnings in 4th quartile | -0.260*** | -0.006 | 0.121** | -0.036 | 0.051*** | -0.000 | 0.015 | -0.029 | -0.036 |
| | (0.033) | (0.022) | (0.057) | (0.047) | (0.016) | (0.012) | (0.050) | (0.045) | (0.041) |
| Earnings in 3rd quartile | -0.232*** | -0.004 | 0.030 | -0.008 | 0.048*** | 0.014 | -0.031 | 0.003 | -0.010 |
| | (0.035) | (0.024) | (0.052) | (0.044) | (0.015) | (0.012) | (0.043) | (0.042) | (0.034) |
| Earnings in 2nd quartile | -0.152*** | -0.013 | 0.076 | 0.071 | 0.057*** | 0.010 | 0.017 | 0.019 | 0.025 |
| | (0.036) | (0.026) | (0.052) | (0.047) | (0.017) | (0.013) | (0.041) | (0.046) | (0.032) |
| Earnings in 1st quartile | -0.164*** | -0.004 | 0.039 | 0.098** | 0.073*** | -0.009 | -0.056 | 0.080* | 0.032 |
| | 0.041 | (0.028) | (0.053) | (0.049) | (0.023) | (0.014) | (0.051) | (0.048) | (0.039) |
| Vacancy rate above the mean | -0.145*** | -0.018 | 0.084** | 0.010 | 0.058*** | 0.005 | 0.037 | -0.009 | 0.025 |
| | (0.025) | (0.017) | (0.038) | (0.034) | (0.013) | (0.009) | (0.033) | (0.033) | (0.026) |
| Vacancy rate below the mean | -0.258*** | -0.002 | 0.064* | 0.063* | 0.060*** | 0.000 | -0.048 | 0.046 | -0.001 |
| | (0.026) | (0.018) | (0.039) | (0.033) | (0.013) | (0.009) | (0.034) | (0.031) | (0.027) |

Notes: ***, **, * denote statistical significance at 1, 5 and 10 percent, respectively. Standard errors (in parentheses) account for heteroskedasticity and clustering at the individual level. All specifications control for individual and calendar month fixed effects. The regressions use data available for the whole pre-treatment period but exclude data for the first two years of the post-treatment period.

## 1.7 Comparing with the reform effects on sick individuals without a spouse

The results in the preceding section suggest that spouses increased their labour participation to compensate for lost disability benefits of their sick partners, particularly if the sick individuals cannot increase their own earnings. Here we compare the reform effects on labour participation of sick individuals with and without a spouse. Since only sick individuals with a spouse can compensate the loss of household income through spousal labour supply, they might less often make the effort to go back to work than singles do, particularly if their labour market position is weak. Figure 1.10 compares labour participation and benefit receipt of sick people with and without a spouse of control and treatment groups. It suggests that indeed, the positive reform effect on labour participation is substantially larger for the sick individuals without a spouse. Figures for wages and benefit amounts lead to the same conclusions (not shown).

Table 1.10 in presents summary statistics for sick people without a spouse. As before, the (small) differences between the treatment and control groups will be taken into account by our empirical strategy. Comparing with Table 1.1, singles tend to be younger and more often female than sick people with a partner. They are also less likely to have a permanent contract at the time of reporting sick, which suggests it might be important to allow for heterogeneity by labour market conditions.

Table 1.6 presents the DiD estimates of the reform effects for sick people without a spouse, and reproduces the baseline estimates for sick people with a spouse from Table 1.3.[15] The estimates confirm that the reform increases the probability of working post-treatment by 1.2 pp more among sick people without a spouse than for sick individuals in couples. Compared to the effects on DI receipt, the effects on labour participation are almost twice as large for singles than for partnered individuals: 61% vs. 32%. Together with the earlier finding that spouses increase their labour participation in response to the reform (Table 1.3), this suggests that in couples, the response to the disability reform is shared by both partners: Spousal labour supply is a substitute for sick individuals' own labour supply. The reform effects for earnings and benefits are in line with this. For example, sick people without a spouse increase their earnings by 15.8% in response to the reform, whereas for sick people with a spouse the increase in earnings is only 7.6%.

As in Section 1.6, we consider the possibility that the reform effects depend on the time since the individual fell sick, see Figure 1.8 in the Appendix. The time patterns of the effects on labour participation are similar for sick people with and without a spouse,

[15] Figure 1.8 in the Appendix presents the estimates of pre-treatment effects for sick individuals without a spouse for all outcomes, supporting the common trend assumption.

but substantially larger for sick people without a spouse, also in the long run. The time patterns of the effects on DI and UI receipt are also similar to those with and without a spouse – persistent and significant throughout the entire post-treatment period. Similar time patterns are found for the effects on wage and benefit amounts (not shown).

In Section 1.6.3 we analyzed heterogeneity in labour supply responses to the DI reform to better understand how couples make joint labour supply decisions. We conducted a heterogeneity analysis for singles and compare with individuals in couples in Table 1.7.[16] Like before, we focus on heterogeneity in terms of labour market position when falling sick, characterized by type of contract (top panel), earnings level (middle panel), or vacancy rate in the sector (bottom panel). First, the effects of the reform on the chances to receive disability benefits after the sickness period are negative and similar for individuals in couples and singles in all cases. The point estimates tend to be somewhat larger for singles, but the differences with partnered individuals are not significant, even though individuals in couples and singles may also differ in other characteristics. The most interesting part in the table is the middle column. Individuals with a relatively strong labour market position (permanent contract, high earnings, or low vacancy rate sector) respond themselves, irrespective of whether there is a spouse or not. In particular, the responses for singles and non-singles with a permanent contract are remarkably similar, even though the two groups may differ in many other characteristics. This group has relatively good chances to go back to work and does not need to rely on a partner (if there is one). On the other hand, the sick individuals that more often struggle to go back to work (temporary workers, for example) respond much more if they are single than if they have a partner. It suggests that in these cases, it is easier for couples if the partner of the sick individual responds, whereas for singles this option does not exist, and the individual makes a larger effort to go back to work. The effects on UI receipt never differ significantly between sick individuals with and without a spouse.

Similar results are obtained for monthly earnings and benefits (Table 1.8). The main difference between sick individuals with and without a spouse is the response in earnings for those on a temporary contract or in the lower pre-sickness earnings groups: it is much larger for singles, who cannot compensate the loss in household income through spousal earnings. There is not much difference between the reform effects on labour participation in the sectors with lower and higher vacancy rates, suggesting the vacancy rate is not a strong indicator of the opportunities to go back to work.

Figures 1.5 to 1.7 in the Appendix present the dynamic effects by employment status, vacancy rate, and earnings quartile for all groups: sick people with a spouse, their spouses, and sick people without a spouse. They are in line with the main findings. In groups

[16] As for the partnered sick individuals, we find similar effects for male and female singles (results not presented).

where spouses respond, sick individuals without spouse also respond, confirming that in couples both partners share the burden of a more stringent DI scheme. The estimated effects at individual event years are not always significant at the 5 percent level, however.

Table 1.6: Estimated effects of the WIA reform: Individuals with and without spouse

| | Sick individual with a spouse | Sick individual without a spouse |
|---|---|---|
| DI receipt | -0.031*** | -0.036*** |
| | (0.002) | (0.004) |
| Labour participation | 0.010*** | 0.022*** |
| | (0.003) | (0.005) |
| UI receipt | 0.008*** | 0.010*** |
| | (0.001) | (0.002) |
| ln DI | -0.203*** | -0.246*** |
| | (0.018) | (0.027) |
| ln Wage | 0.076*** | 0.158*** |
| | (0.027) | (0.038) |
| ln UI | 0.056*** | 0.072*** |
| | (0.009) | (0.012) |
| ln Total individual income | -0.010 | -0.008 |
| | (0.023) | (0.032) |
| Observations | 8,431,218 | 4,425,372 |
| Individuals | 55,106 | 28,924 |

Notes: ***, **, * denote statistical significance at 1, 5 and 10 percent, respectively. Standard errors (in parentheses) account for heteroskedasticity and clustering at the individual level. All specifications control for individual and calendar month fixed effects. The regressions use data available for the whole pre-treatment period and exclude data for the first two years of the post-treatment period.

Table 1.7: Estimated effects of the WIA reform on DI receipt, labour participation and UI receipt of sick people with and without a spouse by labour market status, quartiles of average earnings before reporting sick, and sectoral vacancy rate

| | DI receipt | | Labour participation | | UI receipt | |
|---|---|---|---|---|---|---|
| | Sick people with a spouse | Sick people without a spouse | Sick people with a spouse | Sick people without a spouse | Sick people with a spouse | Sick people without a spouse |
| On permanent contract | -0.027*** | -0.022*** | 0.016*** | 0.015** | 0.004*** | 0.007*** |
| | (0.002) | (0.004) | (0.004) | (0.006) | (0.001) | (0.002) |
| On temporary contract | -0.031*** | -0.036*** | -0.015 | 0.022* | 0.013*** | 0.015*** |
| | (0.009) | (0.009) | (0.012) | (0.011) | (0.004) | (0.004) |
| Unemployed | -0.041*** | -0.055*** | -0.014 | 0.000 | 0.021*** | 0.016*** |
| | (0.008) | (0.010) | (0.009) | (0.011) | (0.005) | (0.005) |
| Earnings in 4th quartile | -0.037*** | -0.037*** | 0.013* | -0.000 | 0.007*** | 0.012*** |
| | (0.004) | (0.007) | (0.007) | (0.009) | (0.002) | (0.003) |
| Earnings in 3rd quartile | -0.036*** | -0.034*** | 0.004 | 0.022** | 0.007*** | 0.006* |
| | (0.005) | (0.007) | (0.006) | (0.009) | (0.002) | (0.003) |
| Earnings in 2nd quartile | -0.024*** | -0.040*** | 0.010 | 0.031*** | 0.008*** | 0.015*** |
| | (0.005) | (0.008) | (0.007) | (0.009) | (0.002) | (0.004) |
| Earnings in 1st quartile | -0.025*** | -0.033*** | 0.006 | 0.028*** | 0.011*** | 0.007* |
| | (0.006) | (0.009) | (0.007) | (0.010) | (0.003) | (0.004) |
| Vacancy rate above the mean | -0.022*** | -0.031*** | 0.010** | 0.020*** | 0.008*** | 0.011*** |
| | (0.003) | (0.005) | (0.005) | (0.007) | (0.002) | (0.002) |
| Vacancy rate below the mean | -0.039*** | -0.040*** | 0.008* | 0.023*** | 0.008*** | 0.010*** |
| | (0.004) | (0.006) | (0.005) | (0.007) | (0.002) | (0.002) |

Notes: ***, **, * denote statistical significance at 1, 5 and 10 percent, respectively. Standard errors (in parentheses) account for heteroskedasticity and clustering at the individual level. All specifications control for individual and calendar month fixed effects. The regressions use data available for the whole pre-treatment period but exclude data for the first two years of the post-treatment period.

Table 1.8: Estimated effects of the WIA reform on DI benefits, wages, UI benefits and total income of sick people with and without a spouse by labour market status, quartiles of average earnings before reporting sick, and sectoral vacancy rate

| | ln DI | | ln Wage | | ln UI | | ln Total individual income | |
|---|---|---|---|---|---|---|---|---|
| | Sick people with a spouse | Sick people without a spouse | Sick people with a spouse | Sick people without a spouse | Sick people with a spouse | Sick people without a spouse | Sick people with a spouse | Sick people without a spouse |
| On permanent contract | -0.183*** | -0.142*** | 0.130*** | 0.112** | 0.027*** | 0.050*** | 0.029 | 0.060 |
| | (0.018) | (0.030) | (0.031) | (0.050) | (0.009) | (0.014) | (0.027) | (0.041) |
| On temporary contract | -0.196*** | -0.248*** | -0.126 | 0.182** | 0.092*** | 0.105*** | -0.180** | -0.046 |
| | (0.067) | (0.065) | (0.089) | (0.090) | (0.030) | (0.026) | (0.075) | (0.075) |
| Unemployed | -0.265*** | -0.378*** | -0.126* | -0.015 | 0.148* | 0.118*** | -0.145** | -0.212*** |
| | (0.059) | (0.073) | (0.070) | (0.083) | (0.033) | (0.037) | (0.062) | (0.069) |
| Earnings in 4th quartile | -0.260*** | -0.241*** | 0.121** | 0.011 | 0.051*** | 0.086*** | 0.015 | -0.026 |
| | (0.033) | (0.050) | (0.057) | (0.074) | (0.016) | (0.023) | (0.050) | (0.061) |
| Earnings in 3rd quartile | -0.232*** | -0.226*** | 0.030 | 0.176** | 0.048*** | 0.043* | -0.031 | 0.040 |
| | (0.035) | (0.053) | (0.052) | (0.071) | (0.015) | (0.024) | (0.043) | (0.054) |
| Earnings in 2nd quartile | -0.152*** | -0.275*** | 0.076 | 0.241*** | 0.057*** | 0.109*** | 0.017 | 0.064 |
| | (0.036) | (0.056) | (0.052) | (0.072) | (0.017) | (0.026) | (0.041) | (0.056) |
| Earnings in 1st quartile | -0.164*** | -0.227*** | 0.039 | 0.220** | 0.073*** | 0.050* | -0.056 | -0.051 |
| | 0.041 | (0.061) | (0.053) | (0.074) | (0.024) | (0.023) | (0.051) | (0.069) |
| Vacancy rate above the mean | -0.145*** | -0.209*** | 0.084** | 0.146*** | 0.058*** | 0.076*** | 0.037 | -0.007 |
| | (0.025) | (0.040) | (0.040) | (0.053) | (0.013) | (0.017) | (0.033) | (0.044) |
| Vacancy rate below the mean | -0.258*** | -0.269*** | 0.064* | 0.169*** | 0.060*** | 0.072*** | -0.048 | 0.016 |
| | (0.026) | (0.040) | (0.039) | (0.056) | (0.013) | (0.019) | (0.034) | (0.046) |

Notes: ***, **, * denote statistical significance at 1, 5 and 10 percent, respectively. Standard errors (in parentheses) account for heteroskedasticity and clustering at the individual level. All specifications control for individual and calendar month fixed effects. The regressions use data available for the whole pre-treatment period but exclude data for the first two years of the post-treatment period.

## 1.8 Checking the identifying assumptions

### Is the pre-treatment time trend common to control and treatment groups?

Our main identifying assumption is that, conditional on observables, control and treatment groups share the same time trend in the potential outcome variables before and after individuals report sick and face the reform incentives or not. The assumption is testable during the pre-treatment period. Figure 1.1 already suggested that control and treatment groups, both for sick individuals and their spouses, share very similar time trends until individuals fall sick, supporting this identifying assumption. For a formal test, we use equation (1.2). Statistically insignificant estimates on the treatment and annual dummy interactions during the pre-treatment period provide evidence supporting the assumption. Year $-1$ is chosen as the base for comparison. Figure 1.2 plots the estimates for sick individuals (left hand panel) and their spouses (right hand panel). For both groups and all outcomes, the estimates are insignificant throughout the pre-treatment period. They are also jointly insignificant, with p-values of at least 70%. The estimates are also (individually and jointly) insignificant for wages and benefit amounts, and in all sub-group analyses of heterogenous treatment effects - see the Appendix.

### Placebo test: Is a treatment effect absent in a non-reform year?

The effects we find could be not due to the reform but due to, for example, some seasonal effect that leads to different changes in labour market position for those who fell sick before and after January 1 2004 (the control and treatment groups, respectively). To confirm that we effects we find are indeed due to the reform, we performed the same DiD estimation (Equation 1.1) comparing the groups who reported sick one year later (last quarter of 2004 and first quarter of 2005). Both groups fall under the new WIA regime so there should not be any reform effects. Table 1.8 in the Appendix shows that, indeed, for both the sick individuals and their spouses, estimated treatment effects are close to 0 for all outcomes, and insignificant at (at least) the 10 percent level.

### Are the results robust to a regression discontinuity approach?

An alternative identification strategy is a regression discontinuity (RD) approach, using the date of falling sick as the running variable (since the reform applies to those who reported sick as of January 1, 2004). Figure 1.11 and Table 1.11, in the Appendix, present the results. Both identification strategies lead to the same qualitative conclusions for all outcomes and to similar relative sizes of the effects across sick individuals with

and without a spouse and for spouses. On the other hand, the RD estimates are typically much larger than the DiD estimates. A possible explanation is that individuals who report sick just before and just after January 1 are different, due to the Christmas holidays. For example, workers in specific sectors or professions may continue working during the last weeks of the calendar year, whereas others do not. If the difference affects levels but not trends, this is accounted for in the DiD estimates but not in the RD estimates.

## Do individuals self-select into the old or new disability scheme?

Reporting sick before or after January 1 2004 determines eligibility for either WAO or WIA, implying that individuals with adverse health shocks in 2003 might select themselves into the WAO or WIA scheme from the time the reform is announced. In particular, the government presented a sketch of its reform plans on 15 September 2003, announcing that the sickness period would be extended from one to two years and that a stricter DI law would be introduced for individuals reporting sick as of 1 January 2004. The transitional WAO reform was announced on 12 March 2004, and details of the WIA reform were announced on 18 August 2004. Following the first announcement in September 2003, individuals could report sick during the last quarter of 2003 instead of after 1 January 2004 to enter the more lenient WAO scheme instead of WIA. In principle, they also might want to postpone their sickness claim until January 2004, to get an additional year of sickness benefits. This seems unlikely since the sickness benefit falls from 100% of the former wage in the first year to 70% in the second year, generally making income while on sickness benefits lower than if on DI or UI (cf. Section 1.2). If individuals strategically choose the disability regime, our results could be biased.

We argue that such self-selection is unlikely. Figure 1.3 shows how many individuals reported sick in the last quarter of 2003 and first quarter of 2004. The distribution is fairly uniform and does not suggest any particular pattern. It certainly does not suggest that many individuals report sick in the last quarter of 2003 instead of early 2004. On the contrary, if anything, there are more sick reports in January 2004, after the stricter WIA scheme was introduced. The relatively low number of workers reporting sick in December 2003 is probably due to a seasonal employment pattern in absence from work, implying that few people report sick during the Christmas and New Year holidays. This is confirmed by the numbers reporting sick one year after the reform, also presented in Figure 1.3. This distribution is very similar to that the year before when the reform was introduced.

In addition, self-selection would be plausible among people with mild impairments only, who would be able to manipulate the timing of their sick reporting. However, both pre- and post-reform, the same Gatekeeper protocol was in place, according to which

after 6 weeks of sickness a first reintegration plan has to be submitted to the Employee Insurance Agency by the employer. Due to this formal screening, mild sickness cases tend to be denied sickness benefits already at this stage.

Finally, if some individuals manage to select themselves into one of the two DI schemes, they would probably do this around 1 January 2004 when the WIA reform came into effect. If we exclude individuals who reported sick within two weeks before or after this date, our DiD results in the heterogeneity analysis remain very similar – see Table 1.12 in the Appendix.

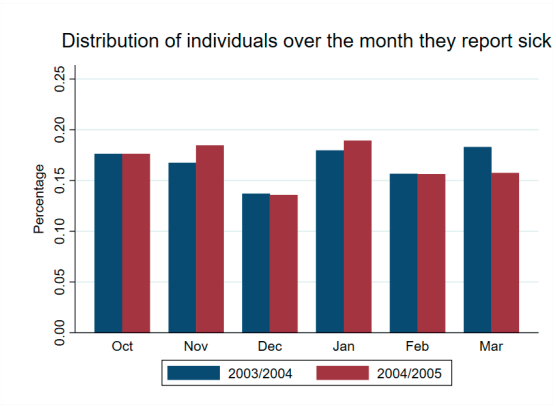



Figure 1.3: Distribution of the number of individuals reporting sick, among those who reported sick in the last quarter of 2003 and first quarter of 2004 and participated, respectively, in the WAO and WIA, and those who reported sick in the last quarter of 2004 and first quarter of 2005 and participated in the WIA.

## Do couples separate due to the reform?

We study the labour supply responses of couples to the DI reform who started cohabiting before reporting sick. Cohabitation can end during the post-treatment period due to the reform or other reasons. In this case the estimated treatment effect may not only reflect the labour supply responses to the DI reform. In the sample, we find no statistical difference between the fractions of couples whose cohabitation ends in the treatment and control groups during the post-treatment period, suggesting that the reform has no effect on cohabitation status – see Figure 1.12 in the Appendix.

### Can compositional differences drive heterogeneity in the reform effects?

In Section 1.6.3 we documented clear reform effects of spousal labour supply responses for individuals reporting sick with a weak labour market position. A threat to our identification strategy might be that the two groups, e.g. permanently and temporarily employed, could differ in other observable and unobservable characteristics, correlated with employment status. For example, individuals with a temporary contract tend to be younger than individuals with a permanent contract. For younger individuals, sickness may represent a larger or more unexpected shock, they might have limited eligibility for UI, and they might have accumulated less wealth to smooth the negative income shock, all of which may lead to a stronger spousal response among younger couples. To address this, we consider gender and age as main variables driving compositional differences across labour market groups. We weigh individuals across labour market groups to have similar distributions of gender and age of the spouse across these groups. We then estimate equation (1.1) within each labour market group using the re-weighted sample of that group. In other words, we control for compositional differences between control and treatment group, but also between the groups with different labour market status when reporting sick. In Table 1.14 in the Appendix we show that the estimated treatment effects in heterogenous labour market groups are robust to these compositional differences across these groups, suggesting that spousal responses indeed stem from weak labour market conditions.

## 1.9 Conclusion

We have analyzed the labour supply and earnings responses of individuals who reported sick and their spouses to a major reform of the Dutch disability insurance (DI) system that introduced stricter eligibility criteria for DI and stronger employer and employee incentives for work resumption. An advantage compared to earlier studies is that we use unique administrative data that include everyone who spent more than three months on sickness benefits, not only the individuals who entered disability after the sickness benefit period expired. We focus on the added worker effects on spouses: Since couples can pool income risk, spousal labour supply can be an important self-insurance mechanism to counterbalance the loss of income due to the reform. Based on a difference-in-differences identification strategy, we find clear evidence of an added worker effect for spouses of workers who report sick from a weak labour market position where work resumption is difficult. Compared to the reform effect on disability benefit receipt and work resumption of the sick individuals themselves, the effect on the spouse's labour participation is substantial (about one sixth and one half, respectively). This finding is notable given that

an earlier major DI reform (implemented in 1993) had no significant effect on spousal labour supply (Borghans et al., 2014). It implies that for a complete evaluation of the DI reform and its effects on labour participation as well as adequacy of household income, it is important to consider spillover effects on spouses.

The effect of the reform on spousal labour supply depends on the type of the employment contract of the sick individual when falling sick. People who had a permanent contract at the time they fell sick increased labour market participation by 1.6 pp due to the reform, while their spouses did not respond. On the other hand, people who had a temporary contract when they fell sick did not increase labour participation because of the reform, but their spouses increased labour participation by 2.5 pp. Furthermore, the spousal response is persistent during the ten years following the start of sickness. Overall, the response at the couple level is sizable regardless in all cases, driven by either the response the sick partners, or the spouses. The effect of the reform also depends on the vacancy rate in the sector where the sick individual was working: if this vacancy rate was above the mean, they increased labour participation by 1 pp themselves while their spouses did not respond, but if the sectoral vacancy rate was below the mean, only the spouses increased labour participation, by a significant 0.9 pp. Finally, spouses increased labour participation more often if their sick partners were low wage earners. All these findings support the hypothesis that partners substitute for each other's labour force participation and spouses respond more often if the labour market position of the sick individual is weaker. Comparing individuals reporting sick with and without partner provide additional evidence for this hypothesis.

Most of the earlier estimates of the added worker effect are small. Our findings add to the few recent studies that find economically meaningful added worker effects (Section 1.1). On average, the extensive margin added worker effect is 16%, as spouses' labour participation increases by 0.5 pp in response to the 3.1 pp drop in DI receipt for sick partners due to the reform. The extensive margin added worker effect attains 81% for sick parters with temporary contracts and it attains 52% and 23% respectively for sick partners with low earnings and for sick partners working in a low vacancy sector. There are several reasons why the 2006 Dutch DI reform did lead to a substantial added worker effect. First, the reform led to a permanent reduction of the income of the affected individual. In line with this, we find persistent responses of both the sick individuals and their spouses in the ten years following sickness (Figure 1.2). Second, the reform could not be anticipated so that couples could not adjust their consumption and labour supply before the reform took place. Third, as the DI reform limited DI entitlement, social protection has become weaker and the need for households' self-insurance increased. These arguments apply particularly if the chances to resume work for the individual who fell sick are small, e.g. since the sick individual had a temporary work contract or was unemployed.

In this paper, we have focused on the WIA reform, replacing the final version of WAO (transitional WAO) by WIA. Using the same source of data and control and treatment groups that reported sick three months earlier, we checked our main findings by performing the same analysis for the reform that introduced the transitional WAO to replace the WAO system preceding this. In Appendix E, we reproduce Table 1.3 and Figure 1.2 for the impact of this reform, comparing sick people who participated in the transitional WAO scheme to those who participated in the WAO scheme preceding this. We find a large participation response of the individuals reporting sick, but no added worker effect on the spouse. This is in line with our main findings since if the sick individuals can respond themselves, there is no need for a spousal response in order to maintain household income.

## 1.10 Appendix

### 1.10.1 Timeline of changes in the Dutch DI scheme

| | | |
|---|---|---|
| 1 July 1967 | Disability Insurance Act (WAO) | • Minimum disability grade for DI eligibility: 15%.<br>• Duration of SI: 1 year. |
| 1 August 1993 | Reduction in DI Benefit Use Act (TBA) | • Major amendments, such as stricter entitlement criteria, financial incentive to resume working, reexaminations. |
| 1 March 1996 | Wage Compensation during SI (Wulbz) | • Employers are obliged to compensate at least 70% of pre-sickness wage during SI. |
| 1 January 1998 | Experience Rating Act (Pemba) | • DI is financed by premiums that are experience rated for the last 5 years.<br>• Applied only to workers with permanent work contracts. |
| 1 April 2002 | Gatekeeper Protocol (Wvp) | • New reintegration obligations for employers and employees are introduced in the SI scheme. |
| 1 October 2004 | Transitional WAO (aSB) | • A broader definition of what work the DI applicant can still do is adapted. |
| 1 January 2006 | Work and Income According to Labour Capacity Act (WIA) | • Applies to workers who reported sick since 1 January 2004.<br>• Minimum disability grade for DI eligibility: 35%.<br>• Duration of SI: 2 years.<br>• Wage compensation and Gatekeeper Protocol are exercised for an additional year in SI compared to the (transitional) WAO.<br>• Experience Rating Act is extended from 5 to 10 years and restricted to temporarily or partially disabled workers and abolished for permanently and fully disabled workers.<br>• A work resumption program with financial incentives to utilize remaining work capacity is introduced for the partially disabled.<br>• A generous full benefit scheme is introduced for the permanently and fully disabled. |
| 1 January 2013 | Extended Experience Rating Act (Bezava) | • Experience Rating Act is extended to workers with temporary work contracts. |

Figure 1.4: Timeline of changes in the Dutch DI scheme.

### 1.10.2 Relation with other reforms

The early retirement reform discussed in Chapter 2 was implemented in 2006 and only applies to people born as of January 1950, for which the early pension claiming age is increased from 55 to 60 and pension benefits become less generous. The disability insurance (DI) reform discussed in Chapter 1, instead, applies to all sickness cases reported as of January 2004, for which access to disability insurance becomes more difficult. Because

one reform discriminates based on the date of birth and the other based on the time of falling sick, we can separately identify the effects of the two reforms. However, it is plausible that the reforms interact. Consider four possible individuals, falling into the four treatment and control groups defined by the two reforms as described in Table 1.9 below:

- Person A is born on 31/12/1949 and falls sick on 31/12/2003.
- Person B is born on 31/12/1949 and falls sick on 01/01/2004.
- Person C is born on 01/01/1950 and falls sick on 31/12/2003.
- Person D is born on 01/01/1950 and falls sick on 01/01/2004.

| | | DI reform | |
|---|---|---|---|
| | | Control | Treatment |
| Pension reform | Control | A | B |
| | Treatment | C | D |

Table 1.9: Interaction of the DI and pension reforms.

The average effect of the DI reform is given by comparing B and D (treated) to A and C (control). However, we might expect the DI reform effect to be larger under the new pension regime, i.e. the difference between D and C might be larger than the difference between B and A. That is because when B and D do not get DI due to the reform, B can still benefit from generous early retirement rules while D cannot.

Similarly, the average effect of the pension reform is given by comparing C and D (treated) to A and B (control). However, we might expect the pension reform effect to be larger under the new DI regime, i.e. the difference between D and B might be larger than the difference between C and A. That is because both C and D do not have access to the old early retirement scheme, but C is more likely to qualify for DI benefits and might thus work less than D.

On average, we find that the DI reform increases labor participation by 0.5-2 percentage points among a selected sample of sick people and their spouses (Chapter 1, Table 1.3 and 1.6), while the early retirement reform increased labor participation by up to 30 percentage points in a less selected sample (Chapter 2, Figure 2.3). Therefore, I would expect the combined effect of the two reforms – comparing D to A, in our example – to be close to the effect of the early retirement reform, but slightly larger.

### 1.10.3 Dynamic heterogenous effects

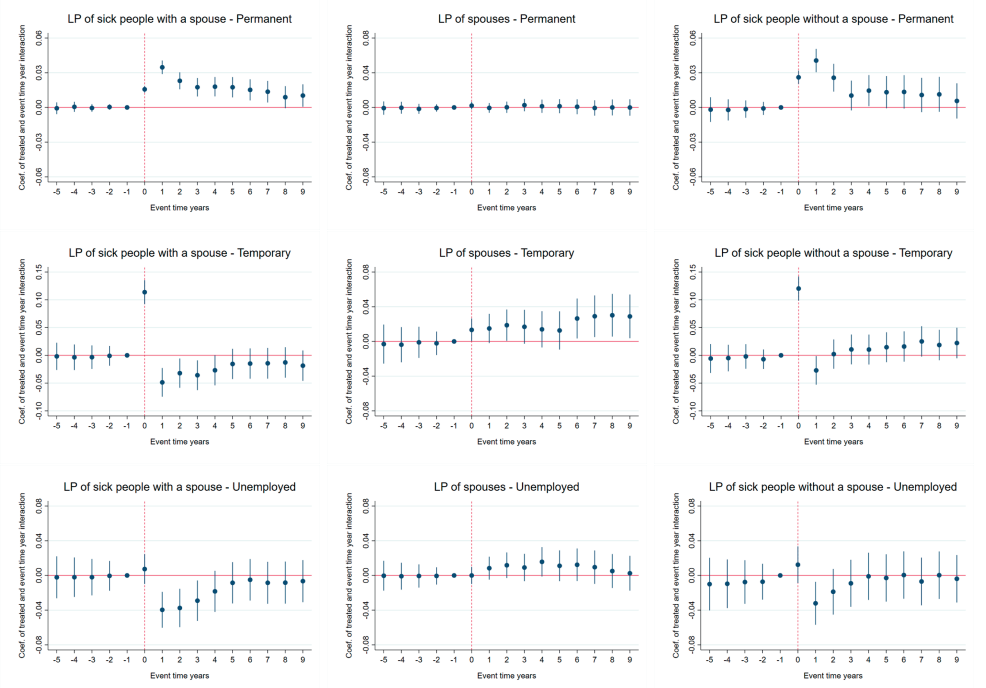



Figure 1.5: Estimated treatment effects in each of the five years before reporting sick and in each of the first ten years after reporting sick, with 95 percent confidence intervals for sick individuals, their spouses, and sick individuals without spouses, by labour market status when reporting sick. Observed differences between treatment and control individuals before reporting sick are controlled for using entropy balancing.

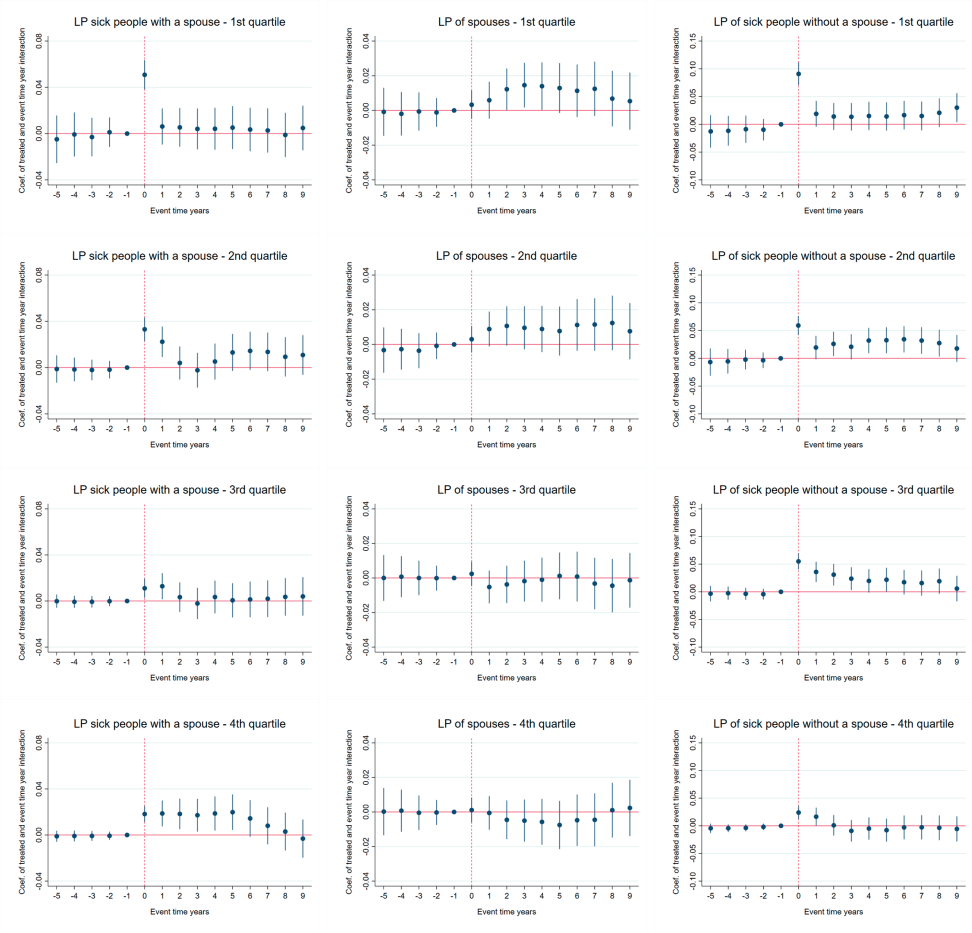



Figure 1.6: Estimated treatment effects in each of the five years before reporting sick and in each of the first ten years after reporting sick, with 95 percent confidence intervals for sick individuals, their spouses, and sick individuals without spouses, by earnings quartile. Observed differences between treatment and control individuals before reporting sick are controlled for using entropy balancing.

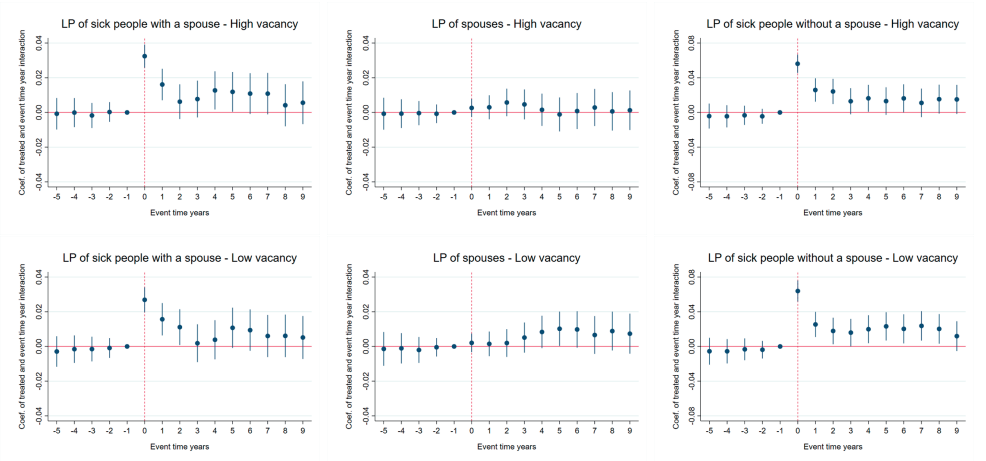



Figure 1.7: Estimated treatment effects in each of the five years before reporting sick and in each of the first ten years after reporting sick, with 95 percent confidence intervals for sick individuals, their spouses, and sick individuals without spouses, by sectoral vacancy rate above (high) or below (low) the average vacancy rate in all sectors. Observed differences between treatment and control individuals before reporting sick are controlled for using entropy balancing.

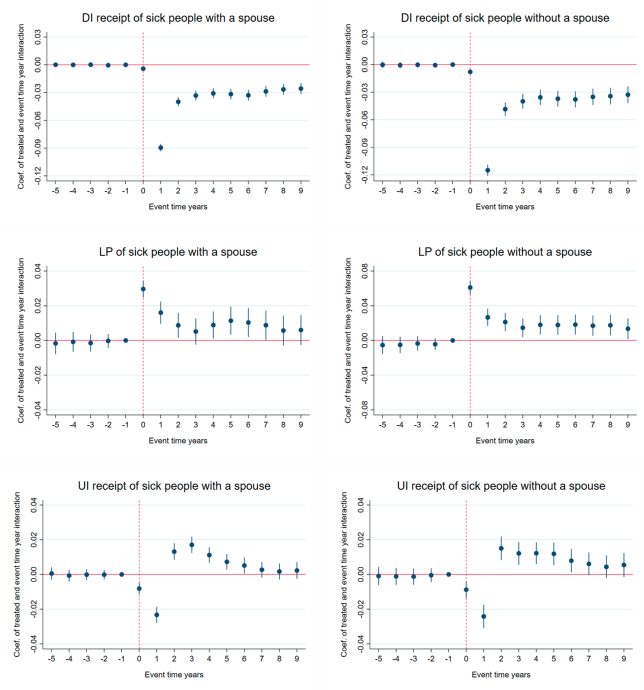



Figure 1.8: Estimated treatment effects in each of the first ten years after falling sick, with 95 percent confidence intervals. Observed differences between treatment and control individuals before reporting sick are controlled for using entropy balancing.

### 1.10.4 Placebo test

Table 1.8: Estimated effects in a non-reform year: Sick individuals and their spouses

| | Sick individual | Spouse |
|---|---|---|
| DI receipt | 0.000 | 0.000 |
| | (0.002) | (0.002) |
| Labour participation | 0.003 | 0.002 |
| | (0.004) | (0.003) |
| UI receipt | 0.010 | -0.000 |
| | (0.001) | (0.001) |
| ln DI | 0.006 | 0.003 |
| | (0.018) | (0.013) |
| ln Wage | 0.033 | 0.021 |
| | (0.029) | (0.025) |
| ln UI | 0.006 | -0.005 |
| | (0.010) | (0.007) |
| ln Total individual income | 0.042 | 0.015 |
| | (0.026) | (0.024) |
| ln Total household income | 0.031 | |
| | (0.021) | |
| Observations | 7,480,935 | |
| Individuals | 48,895 | |

Notes: $^{***}$, $^{**}$, $^{*}$ denote statistical significance at 1, 5 and 10 percent, respectively. Standard errors (in parentheses) account for heteroskedasticity and clustering at the individual level. All specifications control for individual and calendar month fixed effects. The regressions use data available for the whole pre-treatment period but exclude data for the first two years of the post-treatment period.

### 1.10.5 Effects of the transitional WAO reform

We analyzed the impact of changing from the transitional WAO scheme to the stricter WIA scheme for the sick individuals and their spouses. In this appendix, using the same identification strategy, we analyze the impact of changing from the WAO scheme to the transitional WAO scheme. Compared to the WAO, the transitional WAO broadened the definition of what work can still be done by the DI applicant. This means that the Employer Insurance Agency considered a larger number of jobs to match the remaining work capabilities of the sick individuals, implying tighter eligibility criteria for DI (see Section 1.2).

Table 1.9 presents the estimated effects of this reform. The effect on DI awards is very similar to that of the WIA reform, but the effect on labour participation is much larger (cf. Table 1.3). Given that the WIA reform introduced multiple incentives to limit inflow and stimulate outflow, the incentive created by the transitional WAO reform to limit inflow appears to have been very effective. A plausible explanation is that individuals who are affected by the transitional WAO reform have fewer health issues and larger remaining work capabilities than those affected by the WIA reform. Moreover, those who are deprived of DI in the WIA spent an additional year in SI which might have reduced their labour market attachment and employability more than those in the transitional WAO. Compared to the WIA reform, the transitional WAO reform has a much smaller effect on UI receipt, in line with the larger labour participation effect – as more of the sick individuals resumed work, fewer of them made UI claims. Figure 1.9 presents the dynamic effects of the reform, suggesting that the effects on DI receipt and work resumption of the individuals reporting sick were long-lasting. (Moreover, the figure suggests that the common trend assumption is satisfied pre-treatment.)

Unlike for the WIA reform, we find hardly any spousal responses to the transitional WAO reform. This is in line with the large response of the individuals reporting sick: As we saw, the sick individuals were often able to increase labour participation themselves, so there was less need for the spouse to compensate for the lost DI benefits.

Table 1.9: Estimated effects of the transitional WAO reform: Sick individuals and their spouses

| | Sick individual | Spouse |
|---|---|---|
| DI receipt | -0.028*** | -0.003 |
| | (0.003) | (0.002) |
| Labour participation | 0.018*** | -0.001 |
| | (0.004) | (0.003) |
| UI receipt | 0.003** | -0.002* |
| | (0.001) | (0.001) |
| ln DI | -0.192*** | -0.018 |
| | (0.021) | (0.013) |
| ln Wage | 0.145*** | -0.007 |
| | (0.029) | (0.025) |
| ln UI | 0.024** | -0.012* |
| | (0.010) | (0.007) |
| ln Total individual income | 0.015 | -0.030 |
| | (0.025) | (0.024) |
| ln Total household income | -0.009 | |
| | (0.019) | |
| Observations | 7,332,000 | |
| Individuals | 48,800 | |

Notes: ***, **, * denote statistical significance at 1, 5 and 10 percent, respectively. Standard errors (in parentheses) account for heteroskedasticity and clustering at the individual level. All specifications control for individual and calendar month fixed effects. The regressions use data available for the whole pre-treatment period but exclude data for the first two years of the post-treatment period.

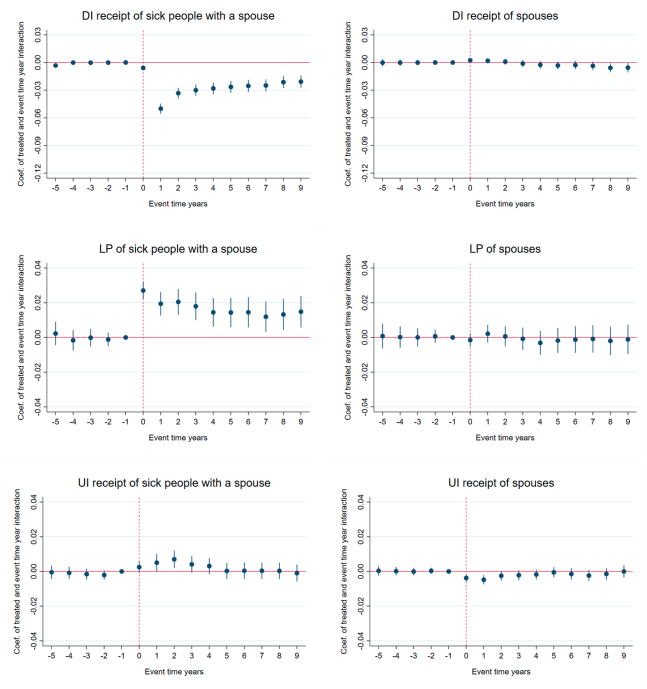



Figure 1.9: Estimated treatment effects of the transitional WAO reform in each of the five years before reporting sick and in each of the first ten years after reporting sick, with 95 percent confidence intervals. Observed differences between treatment and control individuals before reporting sick are controlled for using entropy balancing.

### 1.10.6 Comparing with individuals without a spouse

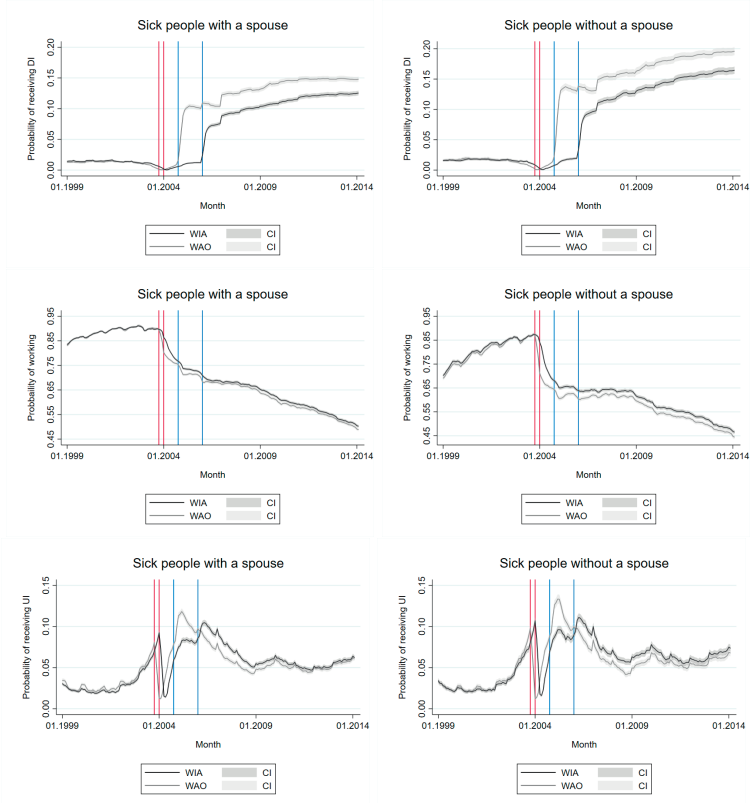



Figure 1.10: Probability of DI receipt, working, and UI receipt for control and treatment groups over calendar months: for sick individuals with a spouse (left panel reproducing the left panel of Figure 1.1) and without a spouse (right panel). Vertical lines mark the first instance sick partners could become entitled to the sickness and disability benefits in the WAO and WIA schemes. Red lines correspond to 1 October 2003 and 1 January 2004 for the WAO and WIA groups, respectively. Blue lines correspond to 1 October 2004 and 1 January 2006 for the WAO and WIA groups, respectively.

Table 1.10: Sample means and balancing tests of background characteristics and outcome in control and treatment groups before and after sickness for sick individuals without a partner

| | Before | | | After | | |
|---|---|---|---|---|---|---|
| | WAO group | WIA group | Dif. WIA and WAO | WAO group | WIA group | Dif. WIA and WAO |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| A. Background characteristics | | | | | | |
| Age | 36.855 | 37.365 | 0.501*** | | | |
| Female | 0.450 | 0.454 | 0.004 | | | |
| Permanent contract | 0.547 | 0.596 | 0.049*** | | | |
| Temporary contract | 0.220 | 0.184 | -0.036*** | | | |
| Unemployed | 0.233 | 0.220 | -0.013*** | | | |
| B. Labour market outcomes | | | | | | |
| DI (possibly UI) receipt | | | | 0.170 | 0.136 | -0.034*** |
| Labour participation | 0.807 | 0.822 | 0.015*** | 0.561 | 0.579 | 0.018*** |
| UI (no DI) receipt | 0.034 | 0.035 | 0.001 | 0.061 | 0.070 | 0.009*** |
| DI (and possibly UI) per month | | | | 224.508 | 195.771 | -28.737*** |
| Wage per month | 1,462.196 | 1,530.015 | 67.819*** | 1,391.216 | 1,454.990 | 63.774*** |
| UI (excl. DI) per month | 41.925 | 44.290 | -3.022* | 77.395 | 70.084 | -7.311*** |
| Observations | 849,300 | 900,060 | | 1,358,880 | 1,440,096 | |
| Individuals | 14,155 | 15,001 | | 14,155 | 15,001 | |

Notes: 1. "Before": period before individuals fall sick (January 1999 - October 2003 for individuals who fell sick in November 2003; January 1999 - January 2004 for individuals who fell sick in February 2004). "After": period after individuals fell sick excluding the first two years (November 2005 - January 2014 for individuals who fell sick in November 2003; February 2006 - January 2014 for individuals who fell sick in February 2004). 2. Age is at the time individuals fall sick. "Permanent contract", "temporary contract", and "unemployed" refer to labour market status of individuals when they fell sick. 3. Columns 1, 2, 4 and 5 present means in control (WAO) and treatment (WIA) group before and after start of sickness. Columns 3 and 6 present differences between individuals insured under WIA and WAO – the estimated coefficient from the regression of the characteristic or outcome as the dependent variable, and an indicator of participation in WIA as the explanatory variable. Standard errors clustered at the individual level.

### 1.10.7 Regression Discontinuity instead of Difference-in-Differences

Our DiD estimates of the effect of the WIA reform rely on the assumption that trends of the outcome variable over event time would have been the same for the treatment and control groups had the reform not been implemented. Although it is not possible to directly test this assumption, we provided evidence that trends are parallel in the pre-treatment period. Here we argue that the fact that we find a significant effect of the reform does not depend on the specific identifying assumption we made. We consider an alternative identification strategy that relies on different identifying assumptions, and the results confirm the results based on the DiD method.

We exploit the date at which the WIA reform came into effect as a source of exogenous variation in treatment status. The assignment to the treatment or control group is a deterministic step-function of the date at which people reported sick – people who reported sick right before 1 January 2004 are insured under the WAO scheme, while people who reported sick right after this "cut-off" date are insured under WIA. We rely on a sharp regression discontinuity (RD) design to estimate the effect of the reform. In particular, the discontinuous jump at the cut-off identifies the treatment effect of interest which can be formalized as

$$\lim_{x \downarrow c} \mathbb{E}[Y_i \mid X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i \mid X_i = x] \tag{1.3}$$

where $X_i$ is the date at which people report sick and $c$ is the cut-off point of 1 January 2004. The treatment effect is estimated using a triangular kernel and a MSE-optimal bandwidth selector (see Calonico et al., 2014). We use a robust variance estimator clustered at the individual level in order to account for the correlation of the error terms across calendar months for the same individual. We consider the same time horizon as with the DiD estimates – the period after treatment but excluding the first two years. We pool all monthly observations of the post-treatment period excluding the first 24 months, implying that we have 96 observations for each individual. We do not account for individual fixed effects but this should not result in biased estimates since the distance from the cut-off date is assumed to be random for individuals who report sick close to 1 January 2004.

The sharp RD design relies on two main assumptions (Imbens and Lemieux, 2008). The first assumption requires a sharp discontinuity in treatment. This assumption holds in our setting by design of the reform, since all individuals $i$ for which $X_i \geq c$ are in the treatment group (WIA regime) and all individuals $i$ for which $X_i < c$ are in the control group (WAO regime).

The second assumption requires continuity in potential outcomes as a function of the assignment variable around the cut-off point. This implies that had the reform not been implemented, the outcome variables should not discontinuously jump at the cut-off point. In other words, "all other factors" driving the outcome variables must be continuous

at the cut-off point (see, e.g., Hahn et al., 2001). Although this assumption cannot be tested directly, relevant variables can be checked for whether they change significantly at the cut-off. We consider contract type at the time of reporting sick as a most relevant variable. We consider dummies for having a permanent contract, temporary contract and being unemployed as outcome variables, and check if they exhibit discontinuity at the cut-off. For sick people with a spouse, we find no significant change at the cut-off in any of the three outcomes. For sick people without a spouse, however, the RD estimate of the treatment effect on being unemployed is -0.088 with a standard error of 0.028. Therefore, we treat the RD estimates of the reform effects as suggestive rather than conclusive, at least for sick people without a spouse.

Figure 1.11 provides graphical evidence for labour participation and benefit receipt. In the figure we distinguish among sick people with a spouse, spouses of sick individuals, and sick individuals without a spouse. For each sample, the figure shows local linear fits for outcome with symmetric bandwidth thirty days around the cut-off date. The figure shows clear discontinuities at the cut-off point, in the expected direction. Furthermore, the relative size of the jumps are in line with the DiD estimates presented in Tables 1.3 and 1.6. For example, for labour participation, sick people without a spouse show the largest effect, followed by sick people with a spouse and by the spouses of sick individuals.

Table 1.11 presents estimated average treatment effects at the cut-off. Both the RD and DiD estimators provide evidence of a positive and significant effect of the reform on the employment probability of spouses and sick individuals with or without spouse. The RD estimates, however, are larger than the DiD estimates. A possible explanation is the fact that RD only identifies the average treatment effect at the cut-off point, that is, for a specific group of people who report sick around 1 January. These people might differ from those who report sick in other months of the year. Overall, both identification strategies provide evidence that sick people in couples rely on the labour supply of their spouses to counterbalance the effect of the DI reform. This is confirmed by the finding that, due to the reform, sick individuals without a spouse increase their labour participation more as they are not able to compensate through spousal labour supply.

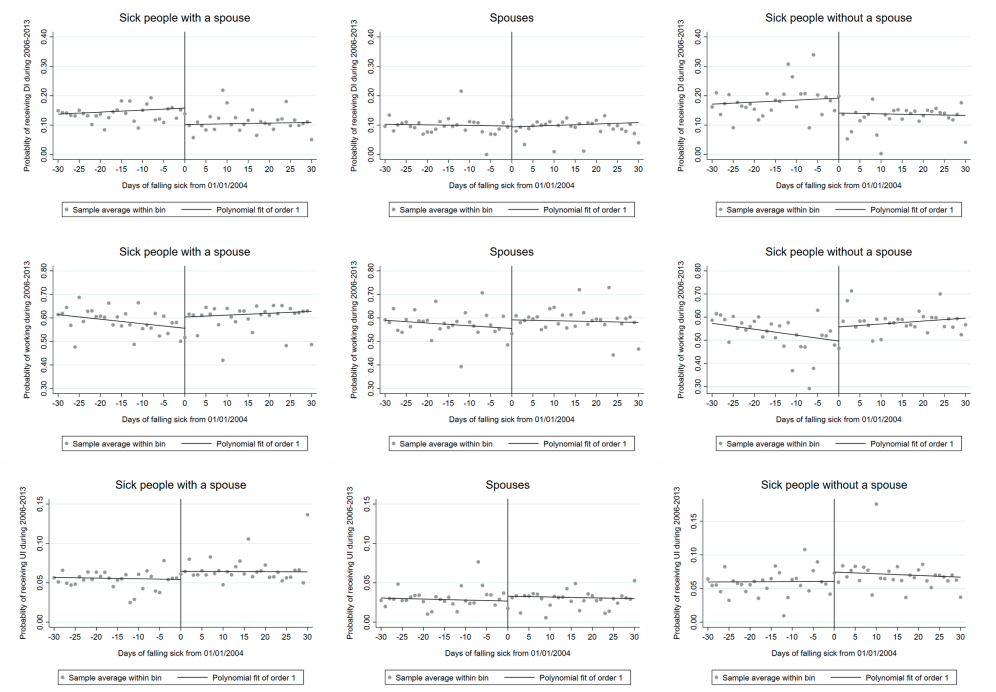



Figure 1.11: Local linear fit on the two sides of the cut-off. Standard errors are clustered at the individual level. All subfigures exclude data for the first two years after reporting sick.

Table 1.11: Sharp RD estimate of the effect of the reform on the labour participation of sick individuals and their spouses and of sick individuals without spouse

| | Sick individual with a spouse | Spouse | Sick individual without a spouse |
|---|---|---|---|
| DI receipt | -0.054*** | -0.001 | -0.051*** |
| | (0.012) | (0.012) | (0.017) |
| Labour participation | 0.048*** | 0.038** | 0.059*** |
| | (0.017) | (0.019) | (0.022) |
| UI receipt | 0.010** | 0.004 | 0.014*** |
| | (0.004) | (0.003) | (0.006) |
| ln DI | -0.382*** | -0.013 | -0.374*** |
| | (0.089) | (0.085) | (0.117) |
| ln Wage | 0.375*** | 0.316** | 0.434** |
| | (0.131) | (0.156) | (0.176) |
| ln UI | 0.076** | 0.025 | 0.103** |
| | (0.030) | (0.021) | (0.043) |
| ln Total individual income | 0.123 | 0.327*** | 0.170 |
| | (0.096) | (0.139) | (0.135) |
| ln Total household income | 0.110 | | |
| | (0.089) | | |

Notes: ***, **, * denote statistical significance at 1, 5, and 10 percent, respectively. The estimates are obtained using a triangular Kernel and an MSE-optimal bandwidth selector. Standard errors are clustered at the individual level. The regressions are based on post-treatment data excluding the first two years. Effective number of observations and individuals used in the estimations depend on the bandwidth. For example, 1,354,272, 1,274,784 and 835,392 observations for 14,107, 13,279 and 8,702 individuals are used when the bandwidths (days) are 25.9, 24.3 and 29.0 in the regressions of labour participation of sick people with a spouse, spouses and sick people without a spouse, respectively.

### 1.10.8 Do individuals self-select into the old or new disability scheme?

Table 1.12: Estimated effects of the WIA reform on labour participation excluding individuals who reported sick less than two weeks before or after the reform date: Sick individuals and their spouses

| | Sick individual | Spouse |
|---|---|---|
| On permanent contract | 0.016*** | 0.002 |
| | (0.004) | (0.004) |
| On temporary contract | -0.015 | 0.024** |
| | (0.012) | (0.010) |
| Unemployed | -0.017* | 0.006 |
| | (0.010) | (0.008) |
| Earnings in 4th quartile | 0.014* | -0.004 |
| | (0.007) | (0.007) |
| Earnings in 3rd quartile | 0.003 | -0.003 |
| | (0.007) | (0.006) |
| Earnings in 2nd quartile | 0.008 | 0.010 |
| | (0.007) | (0.007) |
| Earnings in 1st quartile | 0.005 | 0.013* |
| | (0.008) | (0.007) |
| Vacancy rate above the mean | 0.010* | -0.001 |
| | (0.005) | (0.005) |
| Vacancy rate below the mean | 0.007 | 0.010** |
| | (0.005) | (0.005) |

Notes: ***, **, * denote statistical significance at 1, 5 and 10 percent, respectively. Standard errors (in parentheses) account for heteroskedasticity and clustering at the individual level. All specifications control for individual and calendar month fixed effects. The regressions use data available for the whole pre-treatment period but exclude data for the first two years of the post-treatment period.

### 1.10.9 Do couples dissolve their cohabitation due to the reform?

We studied the labour supply responses of couples to the DI reform who started cohabiting before reporting sick. Couples can dissolve their cohabitation during post-treatment due to the reform or other reasons. This may confound the estimated reform effects. Here we check to which extent the reform affected cohabitation status during post-treatment.

The left panel of Figure 1.12 presents the probability that couples end their cohabitation during post-treatment. The probability is small and shows a decreasing time trend that is common to control and treatment groups. The confidence intervals for the two groups overlap which could suggest that the reform has no statistically significant effect on cohabitation status. These figures are in line with Table 1.2 which showed that couples in both the treatment and control groups cohabit for about 8 years on average during the the 10-year period of post-treatment.

To test whether the reform affected cohabitation status, we rely on a sharp RD design as in Appendix E. In particular, we exploit the date at which the reform came into effect as a source of exogenous variation in treatment status, and analyze whether sick-listed workers insured under the WAO and WIA differ in their cohabitation status during the ten years after reporting sick. The right panel of Figure 1.12 provides graphical evidence. It shows local linear fits for the probability that cohabitation ends with symmetric bandwidth thirty days around the cut-off date. The figure shows no discontinuity at the cut-off. The RD estimate (standard error in parenthesis) of the reform effect is 0.000 (0.000) and is statistically insignificant at the 10 percent level.[17] This shows that the reform did not cause couples to dissolve their cohabitation.

[17]The RD estimation uses the MSE-optimal bandwidth and data for all available years after reporting sick which includes 1,398,000 observations for 11,650 couples.

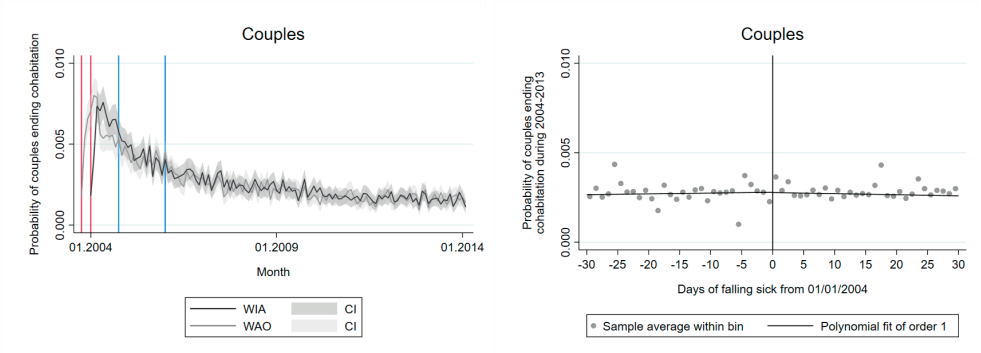



Figure 1.12: Top panel: Probability of couples ending cohabitation after one spouse falls sick. Vertical lines mark the first instance sick partners could become entitled to the sickness and disability benefits in the WAO and WIA schemes. Red lines correspond to 1 October 2003 and 1 January 2004 for the WAO and WIA groups, respectively. Blue lines correspond to 1 October 2004 and 1 January 2006 for the WAO and WIA groups, respectively. Bottom panel: Local linear fit on the two sides of the cut-off. Standard errors are clustered at the individual level. The figure uses 1,996,200 observations for 16,635 couples. The figure uses data for all available years after reporting sick.

### 1.10.10 Can compositional differences account for the heterogenous reform effects?

In Section 1.6.3 we showed that the added worker effect is a more powerful insurance mechanism when the labour market position of the sick individual affected by the reform is weak. In Section 1.7, our analysis of the heterogenous reform effects among the sick individuals without spouses confirmed that in couples both partners share the burden of a more stringent DI scheme. Our estimated treatment effects across the labour market groups could in principle be driven by compositional differences across the groups. We consider gender and age as main variables driving compositional differences across labour market groups. We weight individuals across labour market groups to have similar distributions of gender and age of the spouse across these groups. With respect to employment status, for example, we take the mean age of spouses of sick individuals who have a temporary contract as reference, and weight spouses of sick individuals who are with a permanent contract or unemployed so that the three groups share similar distributions of age. We then estimate equation (1.1) within each labour market group using the re-weighted sample of that group.

The left most columns of Table 1.13 presents the fraction of women and average age among spouses of sick individuals in heterogeneous labour market groups in our study sample. Across the labour market groups, while differences in the average age of spouses are small, there is considerable variation in the fraction of female spouses. The right most two columns of Table 1.13 show the mean gender and age after re-weighting. Table 1.14 presents estimation results for labour participation of spouses. The left column reproduces estimates from Table 1.4 and the right column presents new results based on re-weighted samples. The table shows minor differences suggesting that differences in gender and age are not main drivers of the observed differences in responses across labour market groups.

Table 1.13: Sample means of gender and age for spouses by labour market status, quartiles of average earnings before reporting sick, and sectoral vacancy rate

| | Original sample | | Re-weighted sample | |
|---|---|---|---|---|
| | Female | Age | Female | Age |
| On permanent contract | 0.671 | 42.879 | 0.537 | 39.069 |
| On temporary contract | 0.542 | 39.131 | 0.540 | 39.301 |
| Unemployed | 0.451 | 42.801 | 0.542 | 39.091 |
| Earnings in 4th quartile | 0.893 | 45.144 | 0.504 | 40.253 |
| Earnings in 3rd quartile | 0.817 | 42.141 | 0.509 | 40.243 |
| Earnings in 2nd quartile | 0.507 | 40.186 | 0.503 | 40.206 |
| Earnings in 1st quartile | 0.255 | 42.273 | 0.499 | 40.220 |
| Vacancy rate above the mean | 0.516 | 42.390 | 0.507 | 42.307 |
| Vacancy rate below the mean | 0.720 | 42.422 | 0.521 | 42.437 |

Notes: Female is indicator of the gender of spouse. Age is the average age of the spouse at the time the partner is reported sick.

Table 1.14: Estimated effects of the WIA reform on labour participation of spouses by labour market status, quartiles of average earnings before reporting sick, and sectoral vacancy rate in the original sample and when the sample is re-weighted

| | Labour participation of spouse | |
|---|---|---|
| | Original sample | Re-weighted sample |
| On permanent contract | 0.001 | -0.001 |
| | (0.004) | (0.004) |
| On temporary contract | 0.025*** | 0.025*** |
| | (0.009) | (0.010) |
| Unemployed | 0.010 | 0.011 |
| | (0.008) | (0.008) |
| Earnings in 4th quartile | -0.004 | -0.017 |
| | (0.006) | (0.010) |
| Earnings in 3rd quartile | -0.002 | -0.002 |
| | (0.006) | (0.007) |
| Earnings in 2nd quartile | 0.012* | 0.019* |
| | (0.006) | (0.006) |
| Earnings in 1st quartile | 0.013** | 0.013* |
| | (0.006) | (0.007) |
| Vacancy rate above the mean | 0.002 | 0.002 |
| | (0.004) | (0.004) |
| Vacancy rate below the mean | 0.009** | 0.011** |
| | (0.004) | (0.005) |

Notes: $^{***}$, $^{**}$, $^{*}$ denote statistical significance at 1, 5 and 10 percent, respectively. Standard errors (in parentheses) account for heteroskedasticity and clustering at the individual level. All specifications control for individual and calendar month fixed effects. The regressions use data available for the whole pre-treatment period but exclude data for the first two years of the post-treatment period.

# References

Arulampalam, W. (2001). Is unemployment really scarring? Effects of unemployment experiences on wages. *The Economic Journal*, 111(475):F585–606.

Arulampalam, W., Gregg, P., and Gregory, M. (2001). Introduction: unemployment scarring. *The Economic Journal*, 111(475):F577–584.

Autor, D., Kostøl, A., Mogstad, M., and Setzler, B. (2019). Disability benefits, consumption insurance, and household labor supply. *American Economic Review*, 109(7):2613–54.

Autor, D. H., Duggan, M., Greenberg, K., and Lyle, D. S. (2016). The impact of disability benefits on labor supply: Evidence from the VA's disability compensation program. *American Economic Journal: Applied Economics*, 8(3):31–68.

Ayhan, S. H. (2018). Married womens added worker effect during the 2008 economic crisis–The case of Turkey. *Review of Economics of the Household*, 16:767–790.

Bentolila, S. and Ichino, A. (2008). Unemployment and consumption near and far away from the Mediterranean. *Journal of Population Economics*, 21(2):255–280.

Blundell, R., Pistaferri, L., and Saporta-Eksten, I. (2016). Consumption inequality and family labor supply. *American Economic Review*, 106(2):387–435.

Borghans, L., Gielen, A. C., and Luttmer, E. F. P. (2014). Social support substitution and the earnings rebound: evidence from a regression discontinuity in disability insurance reform. *American Economic Journal: Economic Policy*, 6(4):34–70.

Bredtmann, J., Otten, S., and Rulff, C. (2018). Husbands unemployment and wife's labor supply: the added worker effect across Europe. *ILR Review*, 71(5):1201–1231.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal*, 14(4):909–946.

Cammeraat, E., Jongen, E., and Koning, P. (2023). The added-worker effect in the Netherlands before and during the great recession. *Review of Economics of the Household*, 21:217–243.

Campolieti, M. (2004). Disability insurance benefits and labor supply: some additional evidence. *Journal of Labor Economics*, 22(4):863–889.

Cullen, J. B. and Gruber, J. (2000). Does unemployment insurance crowd out spousal labor supply? *Journal of Labor Economics*, 18(3):546–572.

De Jong, P., Lindeboom, M., and van der Klaauw, B. (2011). Screening disability insurance applications. *Journal of the European Economic Association*, 9(1):106–129.

Deshpande, M. (2016). The effect of disability payments on household earnings and income: Evidence from the SSI children's program. *Review of Economics and Statistics*, 98(4):638–654.

Deuchert, E. and Eugster, B. (2019). Income and substitution effects of a disability insurance reform. *Journal of Public Economics*, 170:1–14.

Duggan, M., Rosenheck, R., and Singleton, P. (2010). Federal policy and the rise in disability enrollment: Evidence for the veterans affairs disability compensation program. *The Journal of Law and Economics*, 53(2):379–398.

Fadlon, I. and Nielsen, T. H. (2021). Family labor supply responses to severe health shocks: evidence from Danish administrative records. *American Economic Journal: Applied Economics*, 13(3):1–30.

Fevang, E., Hardoy, I., and Red, K. (2017). Temporary disability and economic incentives. *The Economic Journal*, 127(603):1410–1432.

García-Gómez, P., van Kippersluis, H., O'Donnell, O., and van Doorslaer, E. (2012). Long-term and spillover effects of health shocks on employment and income. *The Journal of Human Resources*, 48(4):873–909.

García-Mandicó, S., García-Gómez, P., Gielen, A., and ODonnell, O. (2021). The impact of social insurance on spousal labor supply: Evidence from cuts to disability benefits in the Netherlands. Mimeo, Erasmus University Rotterdam.

Godard, M., Koning, P., and Lindeboom, M. (2022). Application and award responses to stricter screening in disability insurance. *The Journal of Human Resources*, 57(3).

Gruber, J. (2000). Disability insurance benefits and labor supply. *Journal of Political Economy*, 108(6):1162–1183.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.

Hainmueller, J. (2012). Entropy balancing for causal effects: a multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25–46.

Halla, M., Schmieder, J., and Weber, A. (2020). Job displacement, family dynamics, and spousal labor supply. *American Economic Journal: Applied Economics*, 12(4):253–287.

Hullegie, P. and Koning, P. (2018). How disability insurance reforms change the consequences of health shocks on income and employment. *Journal of Health Economics*, 62:134–146.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86(1):4–29.

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615 – 635.

Jolly, N. A. and Theodoropoulos, N. (2023). Health shocks and spousal labor supply: An international perspective. *Journal of Population Economics*, 36:973–1004.

Kantarcı, T., van Sonsbeek, J.-M., and Zhang, Y. (2023). The heterogeneous impact of stricter criteria for disability insurance. *Health Economics*, 32(9):1898–1920.

Karlström, A., Palme, M., and Svensson, I. (2008). The employment effect of stricter rules for eligibility for di: Evidence from a natural experiment in sweden. *Journal of Public Economics*, 92(10-11):2071–2082.

Koning, P. and Lindeboom, M. (2015). The rise and fall of disability insurance enrollment in the Netherlands. *Journal of Economic Perspectives*, 29(2):151–172.

Koning, P. and van Sonsbeek, J.-M. (2017). Making disability work? The effects of financial incentives on partially disabled workers. *Labour Economics*, 47:202–215.

Koning, P. and van Vuuren, D. (2007). Hidden unemployment in disability insurance. *LABOUR*, 21(4-5):611–636.

Kostøl, A. R. and Mogstad, M. (2014). How financial incentives induce disability insurance recipients to return to work. *American Economic Review*, 104(2):624–655.

Low, H. and Pistaferri, L. (2015). Disability insurance and the dynamics of the incentive insurance trade-off. *American Economic Review*, 105(10):2986–3029.

Lundberg, S. (1985). The added worker effect. *Journal of Labor Economics*, 3(1):11–37.

Maloney, T. (1987). Employment constraints and the labor supply of married women: a reexamination of the added worker effect. *The Journal of Human Resources*, 22(1):51–61.

Maloney, T. (1991). Unobserved variables and the elusive added worker effect. *Economica*, 58(230):173–187.

Marie, O. and Vall Castello, J. (2012). Measuring the (income) effect of disability insurance generosity on labour market participation. *Journal of Public Economics*, 96(1-2):198–210.

Moore, T. J. (2015). The employment effects of terminating disability benefits. *Journal of Public Economics*, 124:30–43.

Mullen, K. J. and Staubli, S. (2016). Disability benefit generosity and labor force withdrawal. *Journal of Public Economics*, 143:49–63.

OCTAS (2023). Beoordeling van het arbeidsongeschiktheidsstelsel. Technical report.

OECD (2018). Public spending on incapacity. *OECD Publishing, Paris*.

Prinz, D. and Ravesteijn, B. (2020). Employer responsibility in disability insurance: Evidence from the Netherlands. Mimeo, Harvard University.

Ruh, P. and Staubli, S. (2019). Financial incentives and earnings of disability insurance recipients: evidence from a notch design. *American Economic Journal: Economic Policy*, 11(2):269–300.

Schøne, P. and Strøm, M. (2021). International labor market competition and wives labor supply responses. *Labour Economics*, 70(101983).

Spletzer, J. R. (1997). Reexamining the added worker effect. *Economic Inquiry*, 35(2):417–427.

Staubli, S. (2011). The impact of stricter criteria for disability insurance on labor force participation. *Journal of Public Economics*, 95(9-10):1223–1235.

Stephens, M. J. (2002). Worker displacement and the added worker effect. *Journal of Labor Economics*, 20(3):504–537.

Zaresani, A. (2018). Return-to-work policies and labor supply in disability insurance programs. *AEA Papers and Proceedings*, 108:272–276.

Zaresani, A. (2020). Adjustment cost and incentives to work: Evidence from a disability insurance program. *Journal of Public Economics*, 188(104223).

# Chapter 2: Pension reforms and partial retirement

Joint work with Tunga Kantarcı.

**Abstract**

Most Western countries reformed their pension systems to foster employment at old age, but many people are unwilling to work full-time until the statutory retirement age. In this paper, we study the implications of allowing people to retire partially, that is to combine part-time work with partial pension income, for labour supply at old age. To do so we first study two reforms that (i) abolished a generous early retirement scheme and (ii) increased the statutory retirement age in the Netherlands. We find that they have opposite effects on the incidence of part-time work at old age and that those part-time workers often claim pension benefits at the same time. Second, we develop a structural model of retirement and combine it with the two reforms for estimation and validation. Third, the model shows that the effect of partial retirement on labour supply is heterogeneous across pension regimes, but positive under the reformed Dutch one, and it increases total work hours by up to 2.5 percent at age 66. Workers with lower wealth, who cannot afford to work part-time otherwise, value partial retirement most. Partial retirement increases the taxes paid by workers and lowers the net expenditures of the pension fund.[18]

[18]We thank ABP, and in particular Alexander Paulis, Jo Speck and Erwin Blezer for facilitating this research and providing the data. We thank Arthur van Soest, Alexandros Theloudis, Jaap Abbring, Margherita Borella, Jochem de Bresser, Eric French, Rocco Macchiavello, Luc Bissonnette, Rob Alessie, Sander Muns, Rik Rozendaal, Ruben van den Akker and conference and seminar audiences at the Structural Econometrics Group at Tilburg University, Netherlands Econometrics Study Group 2023, Netspar International Pension Workshop 2023, AIEL 2023, EALE 2023, Meeting of Young Economists 2023, Nederlandse Economendag 2023, International Pension Research Association (IPRA) Doctoral Tuturial. Results are based on calculations by the authors using non-public microdata from Statistics Netherlands. Under certain conditions, these microdata are accessible for statistical and scientific research. For further information: microdata@cbs.nl.

## 2.1 Introduction

Population ageing and declining fertility rates have led to major reforms of social security systems in most Western countries (OECD, 2021). Classical reforms abolish generous early retirement plans or increase the eligibility age to claim pension benefits. These reforms aim to stimulate employment among older workers for a longer period of tax and social security contributions, and a shorter period of pension claims. Both types of reforms have been found successful in increasing the average retirement age (Lindeboom and Montizaan, 2020; Lalive et al., 2023; Atav et al., 2023). However, many people are unable or unwilling to work (full-time) until the statutory retirement age. In this paper, we study whether allowing people to retire partially, that is to combine part-time work with partial pension income, increases labour supply at old age and retirees' well-being.

While the most common retirement transition is from a full-time job into full retirement, many workers take a more gradual path reducing the number of work hours or taking 'bridge jobs' with less demanding tasks before they fully retire (around 30% in the US, Rogerson and Wallenius, 2013). The reasons for working part-time at old age may vary and be related to preferences over consumption and leisure, to the financial incentives embedded in the pension or tax system, or the external constraints due to deteriorating health and declining productivity and wage (Ameriks et al., 2020; Hudomiet et al., 2021; Maestas et al., 2023).

From a policy perspective, incentives for partial retirement have been proposed as a potential instrument to stimulate later retirement (Kantarcı and van Soest, 2013; van Soest and Vonkova, 2014; Berg et al., 2020). The rationale is that some people who otherwise would fully retire may be willing to work part-time if they are given the opportunity to top up part-time wages with partial pensions to finance their consumption. With this motivation, over half of EU member states introduced partial retirement schemes in the past decades, even though with different rules (Eurofound, 2016). Partial retirement schemes could be welfare-improving if they foster labour supply while increasing retirement flexibility at the same time. However, they could reduce labour supply if people use them to replace full-time work (Börsch-Supan et al., 2018; Elsayed et al., 2018). The net effect on labour supply is the sum of a negative effect at the intensive margin and a positive effect at the extensive margin, such that the result is theoretically ambiguous.

Despite its potential to increase labour supply and retirees' well-being, we know little about the effects of partial retirement. This is mainly because reforms that only introduce partial retirement are scarce. In this paper, we combine a structural model of retirement with two pension reforms that (also) changed financial incentives to partially retire in order to study the effect of partial retirement on labour supply, how it varies across pension regimes, and its effect on workers' welfare. We proceed in three steps.

First, we study how the likelihood of retiring partially varies across pension regimes. We exploit two reforms of the Dutch pension system to study their effect on retirement behaviour, with a focus on partial retirement. Similar reforms were conducted in many Western countries (such as the USA, UK, Sweden).[19] To this end, we combine administrative data from the pension fund of public sector employees, the largest in the country, and from Statistics Netherlands. The first reform, implemented in 2006, abolished the early retirement scheme offered by occupational pension funds (the 'early retirement reform' from now on). Pension benefits became less generous and the minimum claiming age was raised from 55 to 60. The new rules applied only to people born in and after 1950. The second reform, implemented in 2011, increased the eligibility age for the state pension (the 'state pension reform' from now on). Birth cohorts of years 1948 to 1960 are subjected to a progressively increasing retirement eligibility age, from 65 to 67, while the yearly benefit amount did not change. Partial retirement was possible before and after both reforms.

Both reforms made pension provisions less generous and as a result they stimulated labour participation. In fact, both reforms increased the average retirement age (Lindeboom and Montizaan, 2020; Atav et al., 2023). However, we find that the reforms had different effects on the incidence of part-time work. The early retirement reform decreased the share of people in part-time work in the ten years before the state pension age, which is the period when people can partially retire. This is because the old regime incentivized people to move from full-time to part-time work to be able to claim early retirement benefits, which would otherwise be lost. The state pension reform, on the other hand, increased the share of people working part-time after age 60. In particular, due to the reform people work for more years, but on average they also work fewer hours each year, possibly to smooth leisure on their path to full retirement. We also find that, depending on the pension rules, up to 40 and 60% of part-time workers claim pension at the same time between age 60 and 65, i.e. they partially retire. Pension reforms, therefore, can have opposite effects on part-time work at old age, suggesting that partial retirement might have different implications depending on the considered pension regime.

Second, we develop a structural model of retirement to study the effects of partial retirement on labour supply and welfare. In order to capture the complexity of retirement decisions, our life-cycle model embeds several key features: Savings and pension rights accumulation, a continuous consumption/savings choice, a discrete choice for the number of work hours, a binary pension claiming choice, wage, health, and survival uncertainty. The combination of the work and claiming choices allows for two partial retirement options with different work hours. We use the state pension reform to estimate the model by targeting retirement behaviour across birth cohorts subjected to different state pension

[19] See OECD (2021) for the most recent pension reforms in OECD countries.

ages. We then use the early retirement reform, which induced very different life-cycle profiles, for out-of-sample validation. The model replicates the large and negative effect of the early retirement reform on the incidence of partial retirement, and the smaller and positive effect of the state pension reform. It also replicates the positive (negative) correlation of wages (wealth) with the use of partial retirement and retirement ages.

Third, using the validated model estimates, we conduct two main counterfactual policy experiments. In the first experiment, we analyse the implications of partial retirement on various outcomes. In our counterfactual analysis this means eliminating the partial retirement option. We find that labour participation is about 3 percentage points higher between age 62 and 66 when people are offered the partial retirement option. This positive effect at the extensive margin hides two different underlying effects: Some people that work part-time through partial retirement but would otherwise work full-time, and some who would otherwise not work at all. To quantify which effect is stronger we estimate the change in the total number of hours worked. We find that the net effect is heterogeneous across pension regimes, but positive after the abolishment of the early retirement scheme. In this case, partial retirement increases total work hours by up to 2.5 percent at age 66. Our welfare analysis shows that poorer workers, i.e. those that cannot finance gradual retirement with private savings, benefit most from the additional flexibility provided by the partial retirement option. For people in the bottom decile of the wealth distribution, partial retirement is as valuable as 9% of their wealth. We also quantify the broader implications on the government's budget. We find that each partial retiree pays around 4,600 EUR more in taxes and social contributions compared to the case without the partial retirement option, which is comparable to what the government spends for six months of state pension benefits per person.

In the second policy experiment, we simulate the effect of increasing the state pension age by an additional year, similar to the planned increase by the Dutch law.[20] In line with the causal evidence on the impact of the state pension reform, we find that raising the state pension age increases labour participation, but it also leads to more part-time work already before the old state pension age. The marginal return – in terms of labour supply – from increasing the state pension age decreases, and gains from a further increase above age 68 might be limited.

Our contribution to the existing literature on retirement and labour supply of older workers is three-fold. First, we contribute to the literature evaluating pension reforms. We show that classical reforms can have large effects on labour supply not only at the extensive

[20]The people in our estimation sample with the highest state pension age were born in 1953 and reached their state pension age of 66 years and 4 months in 2019. People born after 30 September 1961 have a state pension age of around one year higher (67 years and 3 months), meaning that only in 2028 we will be able to judge the effect of such increase.

but also at the intensive margin. In this respect, we expand on Lindeboom and Montizaan (2020) and Atav et al. (2023), who analyse the impact of the same two reforms discussed here but abstract from part-time work and partial retirement choices. In particular, while both reforms decreased life-time pension income, they had opposite effects on the incidence of part-time work at old age. The early retirement reform decreased part-time work before full retirement, while the state pension reform increased it. We also investigate the effect of a further planned, but not yet implemented, increase of the state pension age in the Netherlands.

Second, we contribute to the literature on structural modelling of retirement (Gustman and Steinmeier, 1986; Rust and Phelan, 1997; Heyma, 2004; French, 2005; van der Klaauw and Wolpin, 2008; de Bresser, 2023). First, we introduce partial retirement in a life-cycle model, which we find to be important to understand both the increase of part-time work at old age and the timing of retirement. This means that we model labour supply and pension claiming as two separate decisions, as well as the dynamic implications of partial retirement decisions on current and future pension benefits. Second, unlike earlier studies, we draw on quasi-experimental variation for both estimation (by targeting the effects of a reform) and validation (with out-of-sample predictions) of the structural model. While combining policy reforms with a structural model is becoming a popular approach (Todd and Wolpin, 2006; Attanasio et al., 2011; Kaboski and Townsend, 2011; Voena, 2015; Blundell et al., 2016), its use in the retirement literature has been limited so far. Exceptions are French and Jones (2011), de Bresser (2023) and Iskhakov and Keane (2021), who exploit changes in pension rules only for validation (although the latter reform did not lead to any effect). In this study, we exploit two policy reforms which made pension provisions less generous, but which had opposite effects on the incidence of part-time work and partial retirement at old age. We show that our model is able to capture these different effects of pension rules on work and pension claiming decisions, and thus can be used for counterfactual policy analysis.

Third, we contribute to the literature on partial retirement. Earlier studies provide mixed evidence on the effect of partial retirement on labour supply. Berg et al. (2020) find that incentives for partial retirement increase labour supply in Germany. However, Börsch-Supan et al. (2018) find a negative effect using aggregated data among various European countries. Another strand of the literature relies on stated preference (van Soest and Vonkova, 2014; Elsayed et al., 2018; Kantarcı et al., 2023), and also provides mixed results. The advantage of the stated preference approach is that it allows to study choice opportunities that are not available to workers. In particular, gradual retirement arrangements are often based on informal agreements negotiated between employees and employers (Hutchens, 2010). To address this challenge, our study focuses on retirement behaviour among Dutch public sector workers who face fewer restrictions if they want to

partially retire.[21] The pension fund of public sector employees has been offering a partial retirement plan for decades, and part-time work is much more common in the Netherlands compared to many other European countries. Our results suggest that the effect of partial retirement on labour supply strongly depends on the other features of the pension system, such as penalties in case of early claiming, which can help reconcile the existing mixed evidence.

The rest of the paper is organized as follows. Section 2 describes the Dutch pension system and its reforms. Section 3 presents the data and sample selection. Section 4 presents empirical evidence on the effects of the two reforms. Section 5 presents the model, the solution and the estimation approach. Section 6 discusses the model estimates and model fit. Section 7 presents counterfactual policy experiments. Section 8 concludes.

## 2.2 Institutional setting

**The pension system** Retirement income in the Netherlands mainly stands on two pillars: The state pension and the occupational pension.[22] The General Old-Age Pensions Act (AOW) is the state pension scheme, paying a flat-rate benefit when people reach the state pension age, independent of the individual work history. It provides a subsistence-level income to individuals older than the state pension age. The scheme is unfunded and based on the pay-as-you-go principle: Current state pensions are financed from the current premiums paid by workers through income taxes. The state pension age, originally set at 65 years, is gradually increasing since it was reformed in 2011. Employment contracts are terminated at the state pension age and, as a result, few people work beyond it (Atav et al., 2023). Access to most welfare programs (e.g. disability and unemployment insurance) expires at the same age.

Participation in the occupational pension scheme is mandatory for all employees. Participants accrue pension rights which are paid from the age of claiming. Accrual is based on the number of contribution years, the full-time wage, and the number of work hours expressed as full-time equivalent.[23] The minimum claiming age was set at 55 until 2006, when it was increased to 60 and benefits became less generous. Employees can choose, but are actuarially penalized for claiming early and rewarded for claiming later. They can also partially retire: They can claim part of their pension rights while working

[21] There are no other substantial differences between private and public sector employees with respect to retirement. In particular, they are subject to the same early retirement age and similar rules to compute benefits. Nowadays, most pension funds in the Netherlands offer the possibility to retire partially (e.g. PFZW, KPN, BPL, PHENC).

[22] A third pillar is private pension savings and its share in retirement income is limited.

[23] Until 2004, only the last wage applied for the final calculation. As of 2004, instead, the period-specific wage applied for the amount accrued in that year.

part-time and delay claiming of the remaining part. The scheme is fully funded, meaning that pensions are financed from the premiums paid by participants (and their employers) and from the returns on the invested premiums.

**The early retirement reform** The early retirement scheme of public sector workers (FPU) was introduced on 1 April 1997 by ABP, the pension fund of public sector employees. It allowed workers to retire before the state pension age, as early as of 55, with a generous pension benefit. Therefore, early retirement was the norm. Effectively, these employees were subject to two different schemes: The early retirement scheme, paying benefits before the state pension age, and the normal occupational pension scheme, paying benefits after the state pension age.

The early retirement pension benefit incorporated a flat-rate component, independent of the work history, and an accrued component, which depended on the individual work history. The final amount was equal to the sum of the two components multiplied by a retirement age specific factor to penalize retirement before the pivotal age of 62 or to reward working beyond the pivotal age. The early retirement benefit could be claimed between ages 55 and 65, until the state pension age, and would be lost otherwise. As of the state pension age, employees became eligible for the normal occupational pension and the state pension. In case of early retirement, only the benefits received before the state pension age were actuarially adjusted. Under this scheme, people could retire at age 62 and receive an early pension equal to around 70% of their gross wages until age 65. As of 65, the total pension (including the state pension) would decrease to around 65% of the gross wage.[24] The institutional setting thus provided strong incentives to retire early, as we empirically show later.[25]

After an initial announcement on 5 July 2005, the early retirement scheme was abolished on 1 January 2006 for workers born in and after 1950.[26] While the reform was not unexpected due to the ongoing public discussion at that time, the speed at which it was implemented and the differential treatment of workers born before and after 1 January 1950 came as a surprise when the reform was first announced.[27]

[24]Calculations based on the examples in Lindeboom and Montizaan (2020).

[25]Figure 2.2 shows that more than 90% of the people claimed early retirement benefits under this scheme (Group 1).

[26]The scheme was also abolished for those born before 1950 but who had not worked continuously in the public sector since 1 April 1997 (i.e. from age 48 for someone born in 1949). Before the reform, people didn't need to be employed at age 54 to qualify for the early retirement. The two requirements (date of birth and continuous employment since 1997) cannot be manipulated when the reform is announced in 2005 and cannot lead to any anticipation effect. Using a broader sample and a longer panel from Statistics Netherlands, we indeed could not find any sign of anticipation (results not reported here).

[27]After the announcement of the reform, the pension fund ABP launched a campaign to inform its clients about the new system. In a special newsletter, unions, employers, and ABP explained the new

For the reform cohort, the early retirement scheme was abolished and they were eligible to participate only in the normal occupational pension scheme. In this scheme, the minimum age individuals can claim pension rights is 60, such that the reform implied an increase of the minimum claiming age from 55 to 60. Upon claiming, the pension benefit is equal to the sum of accrued rights and the state pension multiplied by an actuarial factor. While the government only starts to pay the state pension as of the state pension age, employees covered by ABP automatically benefit from an "AOW-bridge" (AOW-overbrugging).[28] This means that, in case of (full) early retirement, the total gross pension income remains the same before and after reaching the state pension age, because ABP provides a top-up in the early retirement years. With partial retirement, a partial "bridge" applies for the part that is withdrawn early. In case of early retirement, an actuarial adjustment implies a reduction of benefits for claiming before the state pension age. In this case, early claiming impacts future benefits at all ages via the actuarial penalties. The 1950 cohort could retire at age 62 and receive a pension equal to around 64% of their gross wages for all subsequent years.

**The state pension reform** In 2011, the Dutch government introduced a reform to gradually increase the state pension age from age 65 to above age 67.[29] Figure 2.1 shows the increase in the state pension age for different birth cohorts.[30] Here we focus on the individuals born between November 1949 and July 1953 (the shaded area in Figure 2.1). For the former cohort, the state pension age was increased form 65 to 65 and 3 months, and for the latter group it was increased from 65 to 66 and 4 months.

The major reforms of the early retirement scheme and the state pension age implied very different pension rules across birth cohorts. People born in November and December 1949 could retire with generous early retirement benefits as of age 55 and receive the state pension as of age 65 and 3 months (group 1 in Table 2.1). People born shortly after, however, face the same state pension age but could only claim occupational pension rights as of age 60 and with less generous provisions (group 2). Furthermore, people born 3 years later (group 6) face a significantly higher state pension age (1 year and 1 month higher). The pension system therefore became progressively less generous, with the

pension scheme. Furthermore, ABP clients and their employers received a personalized letter about the core characteristics of the new scheme, along with a complete electronic service package.

[28] Also the other major Dutch pension funds provide the possibility of using an "AOW-bridge", but that may have to be requested rather than being automatic.

[29] The social partners approved the final draft of the pension agreement on 9 June 2011.

[30] As of 1 January 2023, the state pension age is 67 years and 3 months for individuals born between 1 January 1960 and 30 September 1961. As the retirement age is now linked to life expectancy, for individuals born after 30 September 1961, the exact retirement age is not yet known (at least 67 years and 3 months). The final retirement age will be fixed 5 years in advance.

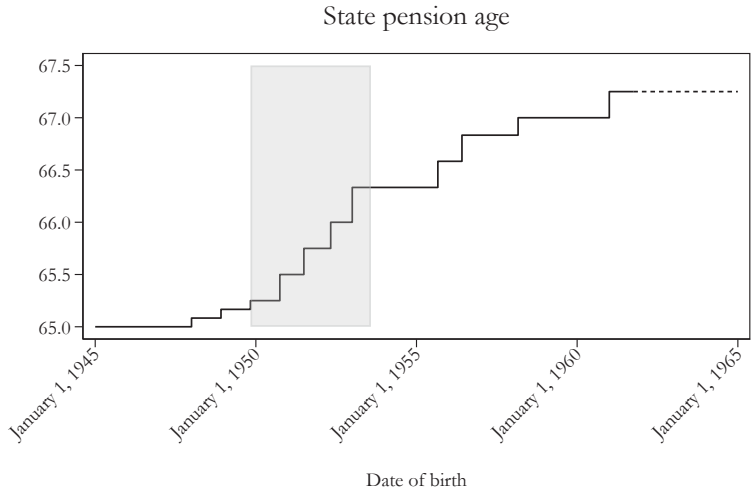



Figure 2.1: Reform of the state pension.

Note: The shaded area refers to the cohorts used in our analysis, born between November 1949 and July 1953. Information about the state pension age can be found at svb.nl.

aim of stimulating employment during the otherwise retirement years. No other reform differentially affected these birth cohorts.[31]

| Group | Birth cohort | Early retirement | State pension age |
|---|---|---|---|
| 1 | Nov/1949 - Dec/1949 | from 55 with minor penalties | 65+3 months |
| 2 | Jan/1950 - Sept/1950 | from 60 with large penalties | 65+3 months |
| 3 | Oct/1950 - Jun/1951 | ” | 65+6 months |
| 4 | Jul/1951 - Mar/1952 | ” | 65+9 months |
| 5 | April/1952 - Dec/1952 | ” | 66 |
| 6 | Jan/1953 - Jul/1953 | ” | 66+4 months |

Table 2.1: Pension rules by date of birth.

[31] A new disability insurance (DI) scheme, which potentially affects labour supply at old age, came into effect in January 2006 and applied to all sickness cases reported as of January 2004 (admission to the DI scheme comes after 2 years of sickness leave). This means that, at the time when the early retirement reform was implemented, all sickness case were insured based on the same DI rules (regardless of the date of birth). Hence, during our study period, the rules of the current DI scheme apply. See Chapter 1 for a detailed discussion of the DI system (Section 1.2 and Section 1.10.2).

## 2.3 Data and sample

**Data** We use unique administrative data from ABP, the occupational pension fund of workers in the government and education sectors.[32] ABP is the largest pension fund in the Netherlands (among the largest in the world) and insures about 15% of the Dutch population. Our initial sample includes all people who are participants of ABP as of December 2019 and were born before 1957. We observe, between January 2005 and December 2019, their wage, full-time equivalent (FTE, the number of work hours divided by the number of hours in a full-time schedule), accrued and paid pension rights, as well as background characteristics including date of birth, gender and marital status. Exact dates are observed when the value of a variable changes. Based on this information, we construct a panel dataset of individuals with monthly observations.

**Sample selection** We impose several restrictions on the data. First, we select individuals based on their date of birth such that we cover different pension rules, but also such that we observe everyone over a relevant period of life regarding retirement choices. That is, we select people born between November 1949 and July 1953 who have been subjected to different pension rules in the occupational and state pension schemes (as summarized in Table 2.1). We do not select cohorts born before November 1949 because they are subjected to a different state pension age (see Figure 2.1), and because for the older cohorts we have fewer observations at younger ages. Similarly, we limit our sample to people born in July 1953 (or before) because we last observe their employment status at age 66 and 5 months in December 2019, right after reaching the state pension age of 66 and 4 months, while for younger cohorts we would progressively miss observations before their state pension age. We observe everyone in our final sample between age 55 years and 2 months and 66 years and 5 months.

Second, we limit our analysis to men because the majority of women work part-time throughout their career in the Netherlands, and thus gradually retire rarely (see Figure 2.11 in the Appendix).

Third, our data includes all individuals who worked for the government and in the education sector at any point in time in their career. This includes people who have been long retired or those who worked in the public sector for short periods of time at young ages. Therefore, some individuals might not have been affected by the pension reforms. We thus keep individuals who are observed working at the age of 55 and two months, which is the first available age for our oldest cohort (people born in November 1949 are

[32]The government sector includes the sectors of central government, provinces, municipalities, water boards, army, police, judiciaries, and all civil servants. The education sector includes all school levels as well as universities, public research institutes and university medical centres.

first observed on January 2005). Finally, we drop individuals who moved to a sector not covered by ABP after age 55 and hence accrued pension rights in a different fund. This implies excluding only a small number of individuals and our empirical results remain virtually identical if we do not impose this last restriction.

These restrictions lead to a sample of 62,402 individuals born in a period of four years and working in the public sectors. We classify them into 6 groups. Table 2.2 shows, for each group, the average age at which individuals stop working ('Retirement age'), the average age at which they start claiming occupational pension rights ('Claiming age'), and the sample size. The claiming age can be lower than the retirement age if a person works part-time while claiming partial pension, i.e. partial retirement, but it can also be higher. Both ages gradually increase across birth cohorts (the differences in averages between two subsequent groups is always significant at the 0.1% level). Overall, the average retirement and pension claiming ages increase by about two years across the selected birth cohorts.

| Group | Birth cohort | Retirement age | Claiming age | Individuals |
|---|---|---|---|---|
| 1 | Nov/1949 - Dec/1949 | 62.71 | 62.54 | 2,456 |
| 2 | Jan/1950 - Sept/1950 | 63.72 | 64.37 | 11,554 |
| 3 | Oct/1950 - Jun/1951 | 63.88 | 64.42 | 12,358 |
| 4 | Jul/1951 - Mar/1952 | 64.03 | 64.53 | 12,562 |
| 5 | April/1952 - Dec/1952 | 64.27 | 64.76 | 13,566 |
| 6 | Jan/1953 - Jul/1953 | 64.34 | 64.82 | 9,906 |
| Total | Nov/1949 - Jul/1953 | 63.99 | 64.49 | 62,402 |

Table 2.2: Main sample.

Note: Groups are defined in Table 2.1. The retirement age is defined as the last age (in months) at which a person is observed working. The claiming age is defined as the first age (in months) at which a person is observed claiming pension benefits.

## 2.4 Empirical evidence of the effect of the reforms

**Trends in employment and pension claiming** We analyse the work and pension claiming decisions across the different pension regimes presented above. We observe labour supply in terms of FTE and define part-time work as working less than 0.875 FTE (e.g. less than 35 hours compared to a weekly full-time schedule of 40 hours).

In Figure 2.2, panel a) presents the share of individuals working at different ages for the 6 groups. First, the figure shows that retirement is an active choice: The majority of the sample retires before reaching their state pension age. Also, people rarely work beyond the state pension age, because work contracts are terminated (Atav et al., 2023).

Second, the average retirement age increases as the pension system becomes less generous. Comparing group 1 with the other groups suggests that the early retirement reform had a large effect on the employment rate, especially from age 62. Comparing groups 2 to 6 suggests that the gradual increase of the state pension age made people work longer.

Panel b) presents the share of individuals who claim pension. While the work and pension claiming choices are strictly related to each other, the figure shows that they are two different choices. For group 1, the increase of the pension claiming rate at age 56 is larger than the corresponding drop in employment, meaning that some people claim their pensions while still working. Instead, for groups 2 to 6, the claiming rate is zero until age 60 (the early claiming age), but the employment rate decreases from age 55 to 60. That is, some people stop working even though they cannot claim their pensions yet, and possibly rely on savings. Therefore, we will model the work and pension claiming choices separately as well as (dis-)saving decisions.

Panel c) reports the share of people in part-time work. At age 55, the part-time rate is close to 10% in all groups. For group 1, the share of part-time workers increases at age 56, when people start claiming pension benefits, and remains stable until age 62. The increase could then be driven by individuals who partially retire to claim pension benefits while still working. For the other groups, the share of part-time workers increases from age 62 and decreases after 64. This hump-shaped pattern is more pronounced for individuals who face a higher state pension age. Setting aside differences by group, the increasing incidence of part-time work with age could be driven by deteriorating health. Appendix 2.9.3.2 shows that, even after taking into account objective health limitations, the part-time rate among healthy people increases notably for groups 2 to 6 as of age 60.

Because differences in the part-time rate could be mechanically driven by differences in employment across groups, in panel d) we condition the outcome variable on being employed. The conditional rate is notably larger for group 1, which shows that individuals in this group are not only less likely to work at every age, but those employed are also more likely to work fewer hours. For groups 2 to 6, instead, the share of part-timers among workers remains stable until about age 60 and increases afterwards, when they become eligible to claim pension.

These trends suggest that the early retirement reform increased labour supply along both the extensive and intensive margins: Individuals work for more years but also more hours in every year (group 1 vs 2). On the other hand, the state pension reform increased the employment rate, but also the part-time rate already before the state pension age, as suggested by the comparison of groups 2 to 6. While the net effect of the reform on labour supply seems positive, it is worth noting that the effect at the intensive margin is negative: People work for more years but also fewer hours in every year.

Panels e) and f) of Figure 2.2 present the share of people in partial retirement (that is

working part-time and claiming pension at the same time).[33] The figure shows that partial retirement is a fairly popular choice in the early retirement scheme. In group 1, 7% of the individuals participate in partial retirement at age 63, which corresponds to about 14% of the employed people and 70% of the people working part-time. In the reformed pension scheme, however, partial retirement is less attractive. The share of people participating in partial retirement shows a steep increase from age 60 to 65 in groups 2 to 6, with around 30-40% of the individuals working part-time at ages 64 and 65 being in partial retirement. Across all groups, the take up of partial retirement seems important to explain part-time work trends at old ages.

[33]We only have data on pension benefits paid under the early retirement scheme from 2006 onwards. Therefore, for group 1, we can compute the share of people in partial retirement only as of age 56. For the other groups the share is zero until age 60, the earliest age at which occupational pension rights can be claimed after the early retirement reform.

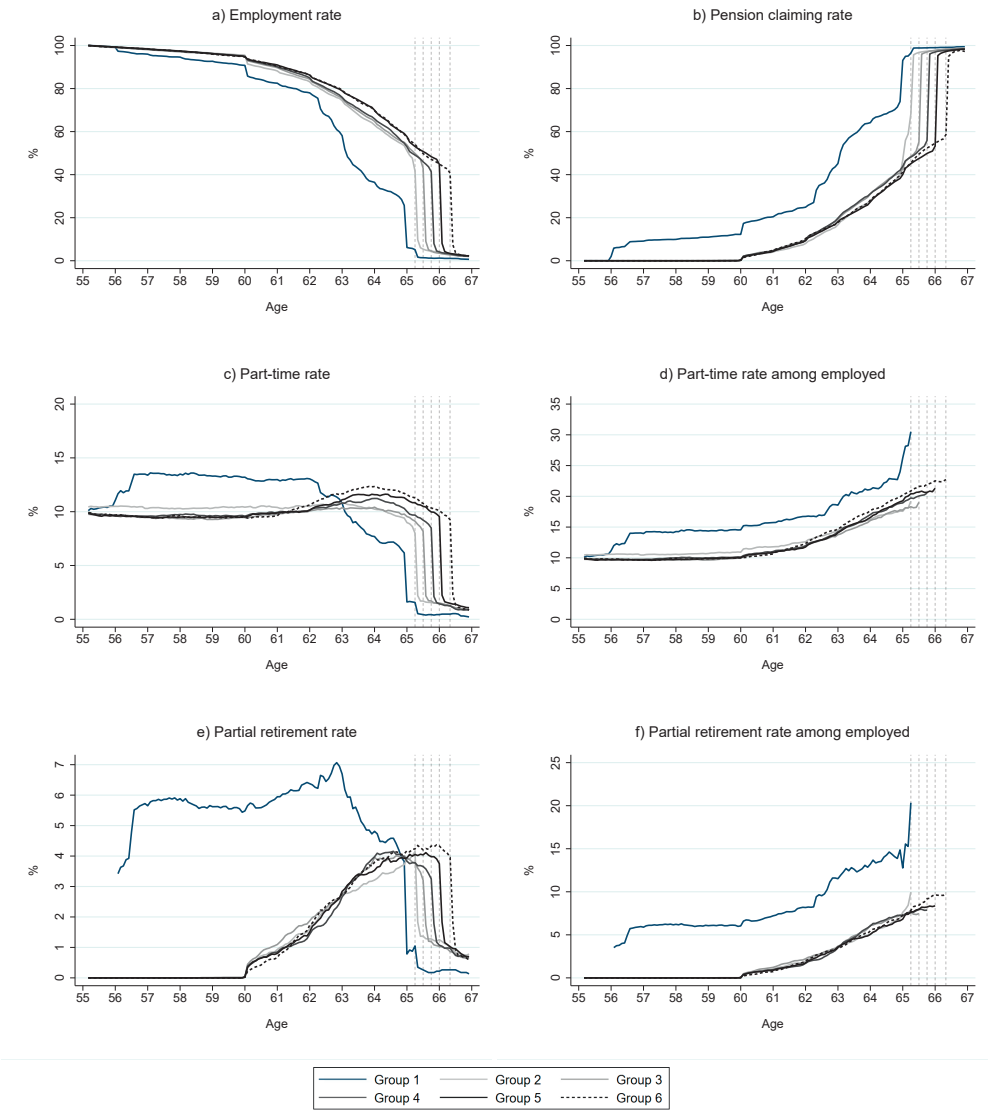



Figure 2.2: Labour supply and pension claiming over age.

Note: Panel a): Share of people working. Panel b): Share of people claiming pension benefits. Panel c): Share of people working part-time (i.e. less than 0.875 FTE). Panel d): Share of people working part-time among those that work. Panel e): Share of people in partial retirement (i.e. working part-time and claiming pension). Panel f): Share of people in partial retirement among those that work. Vertical lines refer to the different state pension ages.

**Difference-in-Differences estimates** Before using the two reforms presented above to estimate and validate our structural model, we check whether they had economically meaningful and statistically significant effects on retirement behaviour.[34] To estimate causal effects of the reforms, we could rely on a Regression Discontinuity approach as in Lindeboom and Montizaan (2020) and Atav et al. (2023). However, for the state pension reform, this allows to measure treatment effects of increases in the state pension age by three to four months. Since we model *annual* decisions in our structural analysis, an increase by few months would not be captured.[35] We therefore rely on a Difference-in-Differences (DiD) approach and compare individuals from group 1, 2 and 6 over time.[36]

We compare people born in November and December 1949 (group 1) to people born in January and February 1950 (part of group 2) to study the effect of the early retirement reform. The reform was announced on 5 July 2005. The January-February 1950 cohort learns at age 55 and a half that they cannot use the early retirement scheme to retire, while nothing changes for the November-December 1949 cohort. Before age 55 and a half, both groups know that a reform will likely be implemented due to ongoing public discussion, but the differential treatment of workers born around January 1, 1950 came as a surprise. We consider the two birth cohorts as treatment and control groups, respectively, and compare their decisions before and after age 55 and a half in a DiD framework. The two groups are affected in the same way by the state pension reform, which implies a small change in rules for these workers: At age 61 and a half they learn that the state pension age is increased from 65 to 65 and 3 months. In fact, the reason why we do not consider people born earlier (e.g. in October 1949) is that they face a different state pension age. We select the treatment group to consist of people born in January and February 1950, and exclude those born in March or later, to keep individuals with approximately the same age of the control group when they are informed about the reforms.

To analyse the impact of the state pension reform, we compare people born in June and July 1953 (part of group 6) to people born in January and February 1950 (part of group 2).[37] In this setting, the former cohort represents the treatment group and the latter the control group. This reform was signed on 9 June 2011 and the June-July 1953 cohort learns at age 58 that their state pension age is notably increased from 65 to 66

[34]The existence of a non-zero effect is not strictly necessary, because changes in rules that do not lead to changes in behaviour are still informative for the model estimates, as in Iskhakov and Keane (2021).

[35]We still provide evidence in the spirit of a Regression Discontinuity approach in Appendix 2.9.3.4

[36]This is similar to Li et al. (2016)'s work on the early retirement reform.

[37]Results are similar if we use a larger or different sample. Our preferred specification is to select a small sample based on the month of birth such that individuals have approximately the same age when they are informed about the reform. In particular, the June and July 1953 cohort learn about the reform around their 58th birthday on 9 June 2011. Age after 58 would then be the 'post-period' in our DiD estimates.

and 4 months. For the January-February 1950 cohort the state pension age was almost unchanged, as it is increased from 65 to 65 and 3 months when they were 61. The fact that the control group received the announcement of the reform at a different age complicates the interpretation of the results, but given the minor increase of the state pension age of just three months the reform is likely to have a limited effect on their behaviour anyway. If anything, our estimates would provide a lower bound for the effect of increasing the state pension age from 65 to 66 and 4 months, because the control group is also partially treated. The treatment and control groups were also affected similarly by the 2006 reform as both groups do not have access to the early retirement scheme, but they learn this at different ages, i.e. 52 and 55 and six months. This means that one group had more time to prepare by, for example, saving more. Any difference in retirement behaviour between the two groups could therefore reflect such differences, on top of the effect of the 2011 reform. While the DiD estimates do not control for this, our structural model does since it explicitly models existing differences in savings and work histories via accumulated assets and pension rights. Therefore, while the DiD estimates could reflect (minor) differences in exposure time to the first reform to some extent, the second reform still provides differential treatment over cohorts that can be used to estimate the structural model, to which we return later.

We estimate the following linear probability model using monthly age observations

$$y_{it} = \alpha_i + \gamma_{s(t)} + \sum_k \beta_k (I\{s(t) = k\} \times T_i) + \varepsilon_{it}$$

where $t$ is age in months $(55+2, 55+3, 55+4, \ldots, 66+11)$, $s$ is age in semesters ($s = 1$ for $t \in [55; 55.5)$, 2 for $t \in [55.5; 56)$, etc.), $\alpha_i$ and $\gamma_{s(t)}$ are individual and age fixed effects, $I$ is the indicator function and $T_i$ is a dummy equal to one for the treatment group. We group observations in semesters to increase the precision of our estimates. $\beta_k$ represents the DiD effect at semester $k$ with respect to the baseline semester (age between 55 and 55.5). $y_{it}$ is alternatively a dummy for work, for part-time work or for partial retirement. Standard errors are clustered at the individual level. Figures 2.3 and 2.4 presents the estimated effects of the reforms: The left panels refer to the early retirement reform, while the right panels refer to the state pension reform.

Panel a) of Figure 2.3 shows a clear effect of the early retirement reform on the employment rate. In particular, it shows a large effect between age 62 and 65, when most people retire under the generous early retirement scheme. The effect disappears when everyone stops working after age 66. On the other hand, panel c) shows that the effect on the probability of working part-time is negative between ages 56 and 62. That is, people are more likely to work full-time due to the reform. The effect turns positive between 62 and 65, when most people in the control group retire. Panel e) shows the effect on the probability of working part-time conditional on working. It shows how the

reform affected the composition of the workforce in terms of full-time versus part-time employment. Among the employed people, the reform increased the probability of working full-time rather than part-time by 3 to 6 percentage points at ages 56 to 64. The reason is that under the early retirement scheme many people switch from full-time work to part-time work to be able to claim partial pension benefits, which would otherwise be lost. As a result, the reform made partial retirement less popular. This interpretation is confirmed by the results for partial retirement presented in panels a) and c) of Figure 2.4.

Panel b) of Figure 2.3 shows no effect of the state pension reform on employment until age 60, the minimum pension claiming age. A significant difference between the two groups opens after age 60 and is largest in the period between the old and the new state pension ages and falls again afterwards as everyone retires. Similarly, panel d) shows that the effect on the probability of working part-time is zero until 60 and positive afterwards. Part of this effect is mechanically driven by the higher propensity to work of the treated group. However, panel f) shows that even after conditioning on working, the part-time rate among employed people is significantly higher for the treated group. In fact, employees are significantly more likely to work part-time already at 62, well before the new state pension age. Panels b) and d) of Figure 2.4 suggest, again, that part of this increase in part-time work is driven by a higher propensity of taking partial retirement.

Overall, our estimates suggest that, while both reforms made the pension system less generous, they had opposite effects on the incidence of part-time work among older workers. They also suggest that, to some extent, these effects on part-time work are driven by a different propensity to retire partially under different pension schemes. However, they have little to say about the effect of offering a partial retirement scheme on labour supply and how much workers value it, for which we turn to the structural analysis.

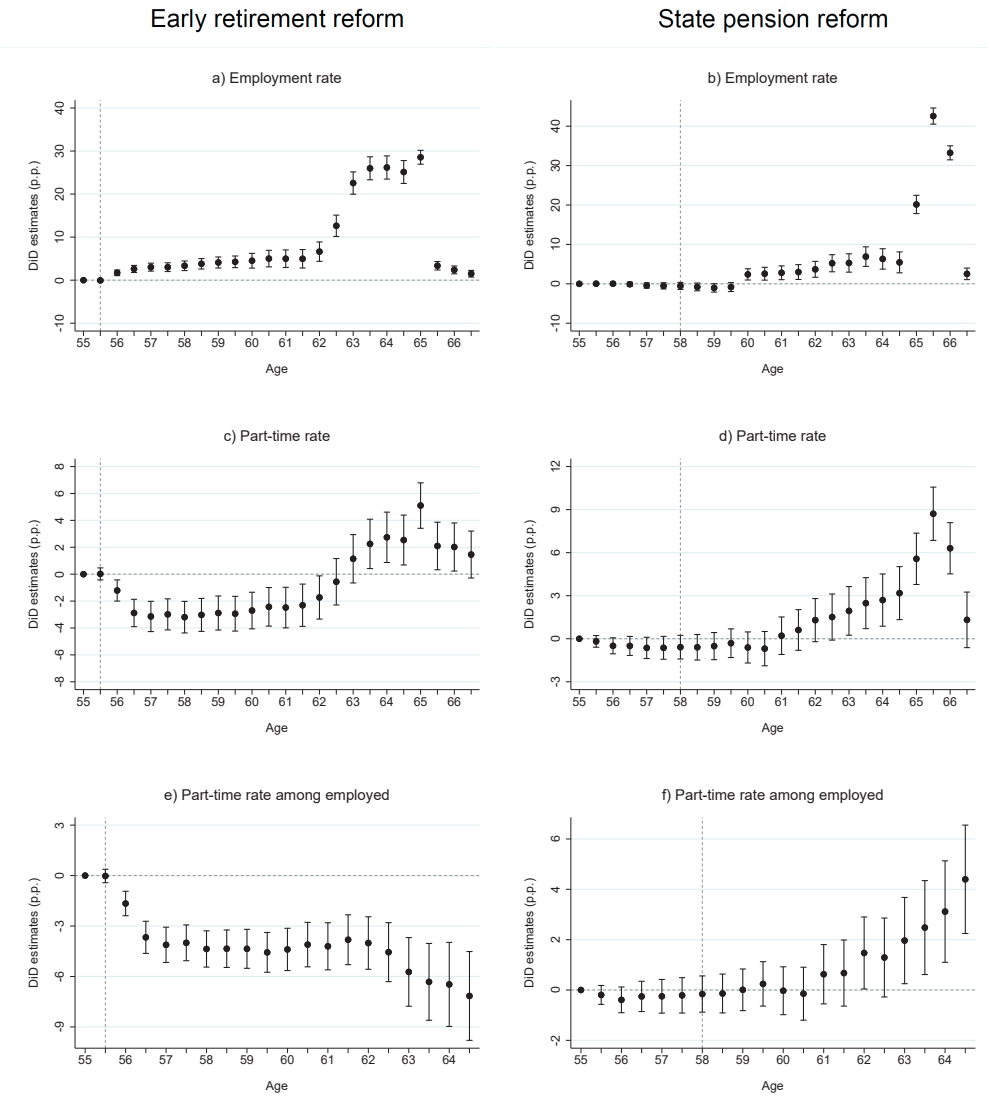



Figure 2.3: DiD estimates of the effects of the reforms on the probability of working, of working part-time, and of working part-time conditional on working.

Note: To study the occupational pension reform (left panels) we use cohort January-February 1950 as the treated group and cohort November-December 1949 as the control group. To study the state pension reform (right panels) we use cohort June-July 1953 as the treated group and cohort January-February 1950 as the control group. The semester from age 55 to 55.5 is used as baseline in all regressions. The vertical lines indicate the age at which the treatment groups receive information about the reforms. Standard errors are clustered at the individual level.

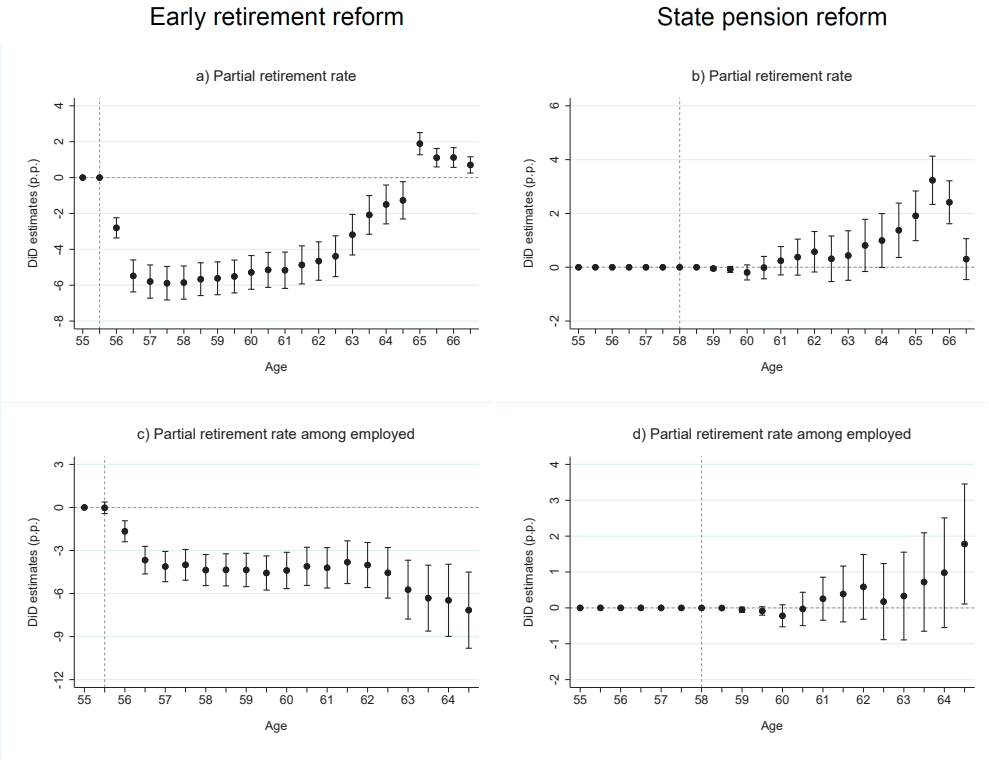



Figure 2.4: DiD estimates of the effects of the reforms on the probability of being in partial retirement, unconditionally and conditional on working.

Note: To study the occupational pension reform (left panels) we use cohort January-February 1950 as the treated group and cohort November-December 1949 as the control group. To study the state pension reform (right panels) we use cohort June-July 1953 as the treated group and cohort January-February 1950 as the control group. The semester from age 55 to 55.5 is used as baseline in all regressions. The vertical lines indicate the age at which the treatment groups receive information about the reforms. Standard errors are clustered at the individual level.

## 2.5 Model

The empirical analysis presented above establishes the responsiveness of part-time work decision, and in particular partial retirement choices, to changes in pension rules. However, it has little to say about the mechanisms underlying those choices. The model we develop below allows us to understand the effects of pension rules on behaviour and on welfare, carry out counterfactual analysis, and ultimately address policy questions regarding the effects of partial retirement and increasing state pension ages.

### 2.5.1 Outline of the model

We model individuals' annual consumption, labour supply and pension claiming choices as of age 56.[38] The labour supply choice set includes (voluntary) non-employment, full-time employment, and two different part-time work levels (0.5 and 0.8 FTE) which cover the two most popular choices of part-time contracts in the sample (see Figure 2.13).[39] Retirement arises endogenously from the labour supply decision, and individuals cannot work beyond the state pension age because employment contracts are terminated.[40] We restrict labour supply choices such that the number of hours worked cannot increase with age. Since it rarely happens in our sample, we make this simplifying assumption to ease the computation.[41] This implies that the current labour supply decision affects the future choice set.

People can claim pension benefits only as of the early retirement age (cohort-specific) and have to claim as of the state pension age (also cohort-specific). Once they claimed pension, they cannot stop, as required by the pension regulation. Therefore, again, the current claiming decision affects the future choice set.

The combination of the discrete labour supply choice and the binary pension claiming choice allows two different partial retirement options: 0.5 FTE work and half pension, and 0.8 FTE work and one fifth of the pension benefit.[42] Full-time work while claiming

[38] The model starts at 56 because this is the first age at which we observe wealth for the older cohorts, and because policy uncertainty largely resolves at 56.

[39] We also express the labour decision in terms of annual hours of work: 0 for not working, 1,000 hours for 0.5 FTE work, 1,600 for 0.8 FTE work and 2,000 for full-time work.

[40] AWVN, the largest employer organization in the Netherlands, confirmed that indeed a challenge to increase labour participation is that most collective labour agreements prevent working after the state pension age.

[41] Given the estimated preferences reported later, the life-cycle decisions simulated by the model are mostly unchanged if we allow people to increase the number of work hours. The results of the main policy experiment (regarding the effect of partial retirement on labour supply) are also essentially unchanged.

[42] ABP regulation states that at most a share equal to $(1 - FTE)\%$ of pension rights can be claimed while working. In the model, we assume that people claim the maximum amount possible if they decide

pension is not allowed, as per regulation.

Agents are thus free to reduce the number of working hours and start claiming pension whenever they want to. This is consistent with the institutional Dutch context, where workers have the right to request a reduction in work hours to which employers must respond in writing. Requests may only be refused if they violate the interests of the firm or the service. Furthermore, part-time work is much more common in the Netherlands compared to similar countries, meaning that workers are less likely to face any stigma for working part-time. While not working or working part-time, employees can freely decide when to start claiming their occupational pension benefits.

Labour supply and pension claiming decisions are affected by (potential) earnings. Wages, which also affect future pension income, evolve exogenously according to a stochastic autoregressive model. This also reflects the (limited) role of human capital accumulation in this institutional context, where the evolution of wages mostly depend on seniority scales fixed in collective labour agreements.

In every period, an individual may die with a probability that increases with age. People die with certainty if they reach the age of 100. Differences in labour supply and claiming decisions can also arise due to differences in health status, which can change in every period. Future sickness status depends on age and current status. For computational and data reasons, we assume that health status can only be good or bad. Having only two possible values for health reduces the state space, but it still allows to account for severe and objective sickness cases that lead to very different behaviour, as shown in Figure 2.12. We model sick people as qualifying as disabled and receiving disability insurance (DI) benefits. That is, we do not model disability insurance claiming choices or implicitly assume that eligible people would always claim. This is because we use administrative data on DI recipiency to infer health status – alternative data, such as health care expenditure, is only available from 2009.[43] While receiving DI benefits, sick people can either work part-time or claim pension benefits, but not both.[44] The model does not feature involuntary unemployment and thus no unemployment insurance as only 1% of

to claim, which is consistent with observed behaviour in our data.

[43]Essentially, the model is consistent with a world in which everyone who is eligible for DI applies to it, and all applicants are granted DI benefits. While this is a strong assumption, several DI reforms were implemented in the Netherlands to improve the screening process, such that it became more likely that only actually disabled people (as defined by the law) get access to DI benefits, and for these people it is typically financially convenient to apply for DI.

[44]We assume that all people in bad health have the capacity to work part-time, because most of the people on DI in our sample work before age 60 (see Appendix 2.9.3.2). To some extent this is likely due to selecting a sample of people who are employed at age 55. However, even at older ages, the share of people on DI who work is relatively high. This might be due to the fact that people who falls sick and starts receiving DI afterwards are subject to the "WIA" DI regime, which strongly incentives people to use their renaming work capacity (if they are not fully disable).

the people in our sample make use of it.

In the model, observed ex-ante heterogeneity is further driven by accumulated savings and pension rights as of age 56, both of which reflect individual specific work histories. As described in Section 2.2, people build pension rights proportionally to their wages and labour supply, which determine pension income in retirement. While the model does not include any uncertainty with respect to pension rules, initial conditions in assets and pension rights also reflect past uncertainty. Consistently, we estimate the model with data on people for whom pension rules uncertainty is largely resolved before entering the model. For cohorts 1949 and 1950 uncertainty essentially resolves with the early retirement reform at age 56. The state pension reform, announced when they were 61, implies an increase of just three months in the state pension age which we abstract from since (i) it is a minor change and (ii) its effect would not be relevant for this exercise as we model *annual* choices. The 1953 cohort is 53 years old at the time of the early retirement reform and 58 at the time of the state pension reform, when uncertainty resolves. We assume they already know their final state pension age as of 56 and avoid modelling uncertainty for the first two years.

All choices are affected by the tax and welfare system, which define disposable income under each employment and retirement option. We model taxes, pension rights accumulation and benefits, pension and other types of contributions according to the cohort-specific regulations.

Table 2.3 summarizes the main features of the model. As explained above, we treat retirement as an absorbing state, meaning that people cannot move from retirement to employment, as it rarely happens in our data. Similarly, they cannot stop claiming pension, as required by the pension regulation. Therefore, the model includes an endogenous state variable 'Retirement status' which affects the future choice sets for the work and claiming decisions, and it is affected by current choices. The exogenous state variable 'Birth cohort' is time-invariant and determines the pension regime which each individual is subject to based on their date of birth. It thus affects the way pension rights and benefits are computed, but also the choice sets because different birth cohorts can early retire at different ages. Table 2.8 in the Appendix summarizes the timeline of the model and provides additional details.

| | | |
|---|---|---|
| Choice variables | 1. Consumption/savings (continuous) | $c/a$ |
| | 2. Hours of work per year (0, 1000, 1600, 2000) | $h$ |
| | 3. Occupational pension claiming (0,1) | $op$ |
| State variables | 1. Age | $t$ |
| | 2. Assets | $a$ |
| | 3. Full-time wage | $W$ |
| | 4. Pension rights | $PR$ |
| | 5. Health status (good or bad) | $health$ |
| | 6. Retirement status | $ret$ |
| | 7. Birth cohort/pension regime | $cohort$ |
| Uncertainty | 1. Survival | |
| | 2. Wage | |
| | 3. Health | |

Table 2.3: Model overview

### 2.5.2 Parametrization

**Preferences** We assume people derive utility from consumption and leisure according to the following specification[45]

$$u(c_t, l_t) = \frac{\lambda}{\lambda - 1} c_t^{\frac{\lambda-1}{\lambda}} + \psi \frac{\gamma}{\gamma - 1} l_t^{\frac{\gamma-1}{\gamma}} \tag{2.1}$$

where $\lambda$ and $\gamma$ represent intertemporal elasticities of substitution, and leisure $l_t \in [0, 1]$ is given by

$$l_t = (4,000 - h_t - \delta I\{health_t = bad\})/4,000 \tag{2.2}$$

$h_t$ being the annual number of hours worked, and $\delta$ is the (time) cost of bad health.[46] We expect $\delta$ to be positive, that is sick people have higher marginal utility of leisure compared

[45]The separable utility function assumed is similar to that in Gustman and Steinmeier (2005) and Keane and Wasi (2016). An earlier version of the paper used a non-separable utility, similar to French (2005), French and Jones (2011) and de Bresser (2023). The two specifications give similar results and match the data equally well until the retirement age. After retirement, the non-separable specification predicts a drop in consumption and an increase in savings. Both these trends for consumption and savings do not match the Dutch data. For that reason, we decided to use the separable specification.

[46]The time endowment is based on de Bresser (2023): $(24-7)\times 5\times 47 = 3995$ hours per year (minimum of 7 hours per day for sleep and does not count weekends and 5 weeks of vacation). It's also in line with estimates from French (2005). Fixed costs of work are typically included in models of labour supply and retirement (e.g. French, 2005; French and Jones, 2011; de Bresser, 2023) because they can explain the fact that most people either do not work or work full-time, while part-time work is not so common in the data. The introduction of fixed costs of work are thus used to match this regularity in the data. Our

to healthy people. Following De Nardi (2004), people derive utility from leaving a bequest as specified by (2.3), where $b_1$ captures the relative weight of the bequest motive and $b_2$ determines its curvature

$$B(a_{t+1}) = b_1 \frac{\lambda}{\lambda - 1}(b_2 + a_{t+1})^{\frac{\lambda-1}{\lambda}}. \tag{2.3}$$

**Wage and health** The logarithm of gross full-time wage evolves exogenously according to an AR(1) process. While the initial conditions allow for cross-sectional dependence between individual tenure and wage, the dynamics of wages do not depend on individual work experience. This is in line with the institutional context of the Netherlands, where the relationship between tenure and salary is fixed according to collective labour agreements (de Bresser, 2023). Furthermore, most workers have reached the end of their wage scale when entering the model. All these considerations are particularly true for the public employees covered by our data. Earnings depend on the number of hours worked and on a wage penalty for part-time work $(1 - \eta)$, as documented by Russo and Hassink (2008) for the Netherlands:

$$\ln(W_t) = (1 - \rho)\mu + \rho \ln(W_{t-1}) + \xi_t \tag{2.4}$$

$$Earnings_t = FTE_t \times W_t \times \exp[\log(\eta) I\{FTE_t < 1\}] \tag{2.5}$$

$$FTE_t = \begin{cases} 0.0 & \text{if } \ h_t = 0 \\ 0.5 & \text{if } \ h_t = 1,000 \\ 0.8 & \text{if } \ h_t = 1,600 \\ 1.0 & \text{if } \ h_t = 2,000. \end{cases}$$

We assume that the idiosyncratic errors term are normally distributed and *iid*, $\xi_t \sim N(0, \sigma_\xi^2)$.

Health status can only take two values (good or bad). The probability of being healthy or unhealthy next year depends on age and this year's status

$$\Pr(health_{t+1} = bad | health_t, t) = \frac{\exp[\pi_0 + \pi_1 t + \pi_2 I\{health_t = bad\}]}{1 + \exp[\pi_0 + \pi_1 t + \pi_2 I\{health_t = bad\}]}. \tag{2.6}$$

We assume that people in bad health are eligible for DI, which in the Dutch context implies having at most 65% of the work capacity left. We assume that everyone who is eligible for DI claims it but can also work part-time (0.5 FTE) at the same time. The assumption that people in bad health have some remaining work capacity is based on the fact the most people on DI in our sample are observed to be working, although they tend to retire early (see Appendix 2.9.3.2).

model does not feature fixed costs of work but it includes a part-time wage penalty, which also makes part-time work less attractive compared to not working and full-time work. With a relatively small choice set for working hours, the part-time penalty and the fixed cost of work act similarly.

**Budget constraint** In every period, the agent receives income and pays an income tax, the pension premium for occupational pension, and other social contributions. Pension contributions allow workers to build up pension rights. Pension rights ($PR$) when entering the model accumulated differently for the different cohorts.[47] Taking $PR$ as given at 56, they then accumulate according to (2.7). The increase in $PR$ is proportional to the number of work hours ($FTE$) and the full-time wage ($W$), and also a function of the accrual rate ($AR$) and the state pension offset ($SPO$), which implies that people do not accrue pension rights on the entire wage. There is, however, no earning test in the Dutch pension system. In that sense, the system does not provide an incentive to switch from full-time to part-time work. Pension rights (and ultimately pension benefits) are proportional to the number of hours worked throughout ones career. A factor $f(\cdot)$ takes into account the age- and cohort-specific actuarial adjustments in case of early claiming. The pension rights translate into benefits $b_t$ as in (2.8), which can be claimed as of 61 (56) for cohorts 1950 and 1953 (1949). Those benefits can be claimed fully if not working or partially if working part-time. Equations (2.7) and (2.8) are the exact formulas used by the occupational pension fund to compute pension rights and benefits. The only assumption we make is that – conditional on claiming – the share of pension claimed is always equal to the maximum, that is one minus the full-time equivalent implied by the work choice. The assumption is consistent with observed claiming behaviour in our data. The benefit is then constant afterwards, unless the number of hours worked changes. If the number of work hours changes, the share of pension claimed changes but also an actuarial adjustment is applied to the additional share claimed. The benefit also changes at age 66 for the 1949 cohort as they move from the early retirement scheme to the normal pension scheme. Therefore, differences across birth cohorts, given by the two reforms, enter the model via the budget constraint and in particular through the different actuarial adjustment factors, accrual rate, state pension offset, minimum claiming age and early pension benefits calculation. More details are presented in Appendix 2.9.1.3.

$$PR_{t+1} = [PR_t + FTE_t \times AR_{cohort} \times (W_t - SPO_{cohort})] \times f(op_t, h_t, ret_t, cohort) \quad (2.7)$$

$$AR = \begin{cases} 1.75\% \\ 2.05\% \end{cases}, \; SPO = \begin{cases} 15,500 & \text{if cohort } = 1949 \\ 9,600 & \text{if cohort } = 1950, 1953 \end{cases}$$

$$b_t = \begin{cases} [\textit{early retirement rights}_t] \times (1 - FTE_t) \times f(\cdot) & \text{if cohort } = 1949 \;\&\; t < 66 \\ [PR_t + \textit{state pension}] \times (1 - FTE_t) \times f(\cdot) & \text{otherwise} \end{cases} \quad (2.8)$$

[47] That is mainly because cohorts 1950 and 1953 only accumulate rights for the old age pension as of 2006, but also receive a compensation for not being able to retire via the old early retirement scheme (see Appendix 2.9.1.3).

In the model, apart from wages and pensions, the other source of income is the DI benefit. If health status is 'bad', the agent receives a DI benefit and cannot work full-time. The yearly DI benefit ($DI_t$) is equal to 70% of the current full-time wage realization $W_t$ times the disability grade.[48] We assume that the disability grade is always 50%.[49] The agent continues to accrue pension rights in proportion to the DI benefit.

$$DI_t = I\{health = bad\} \times W_t \times 0.5 \times 0.7 \tag{2.9}$$

To a given set of choices $(c_t, l_t, op_t)$ and state realization $(X_t)$ corresponds a net income as computed by $\tau$, which takes into account the different income sources as well as the tax paid on income and the contributions to social security schemes (details in Appendix 2.9.1.2).[50] Assets, which include all types of savings, pay a constant return $r$ and accumulate according to equation (2.10).

$$a_{t+1} = (a_t + \tau(c_t, l_t, op_t, X_t) - c_t)(1 + r) \tag{2.10}$$

**Recursive formulation** Equation (2.11) shows the recursive value function representing the agent's problem. The state variables are jointly denoted as $X_t$. $p_t$ is the probability to survive period $t$ conditional on surviving period $t-1$, but eventually no one survives age 100. Expectation is taken with respect to future wage and health, conditional on current state and choices. As uncertainty is essentially resolved for people in our sample before entering the model, there is no uncertainty with respect to future pension rules. The maximum is taken with respect to current consumption $(c_t)$, leisure time $(l_t)$, and the occupational pension claiming choice $(op_t)$. The choice set for consumption depends on the realized state variables but also on the discrete choices because both affect disposable income and savings. Similarly, the choice sets for the discrete decisions depend on the realized state variables, for example because claiming is only possible from certain ages and retirement is an absorbing state, and it's also not possible to work full-time while claiming pension. The model has no closed form solution and therefore we rely on a

[48] In reality, the benefit amount depends on the last wage before sickness. We make this simplifying assumption because we do not keep track of the wage history and because wages follow a very persistent AR(1) process.

[49] To ease computation the model does not feature heterogeneity in the disability grade. People qualifying as disabled, in the Netherlands, have at most 65% of the working capacity left. Since we want to allow sick people to work part-time in the model, as observed in the data, the only available work option corresponds to 0.5 FTE making it a natural choice to assume 50% disability.

[50] French and Jones (2011) argue that it's relevant to include spousal income in their model as it can insurance against uncertain medical expenses. Our model does not feature medical expenditure because they were of limited importance in the Netherlands during the period studied in this paper and consisted mostly of monthly premiums for mandatory health insurance. We thus abstract from spousal income and also from medical expenses.

numerical solution via value function iteration. More details on the model solution are presented in Appendix 2.9.2.1.

$$V_t(X_t) = \max_{c_t, l_t, op_t} \{u(c_t, l_t) + p_t \beta \mathbb{E}_t[V_{t+1}(X_{t+1})|X_t, c_t, l_t, op_t] + (1 - p_t)B(a_{t+1})\} \quad (2.11)$$
$$\text{s.t. (2.1) to (2.10)}$$

### 2.5.3 Model estimation

As in French (2005), we follow a three-step procedure to estimate the model. In the first step, we externally set some parameters to values from the literature. In particular, we set the interest rate $r$ to 1%. Our wealth data covers the period from 2006 to 2021, during which interests rates were particularly low, if not negative, and the average Euro Interbank Offered Rate (Euribor) was 0.9%.[51] We set the time discount factor $\beta$ to 0.99, similar to estimates from de Bresser (2023) and close to $1/1 + r$. We set the part-time penalty to 13% $(1 - \eta)$, similarly to Keane and Wasi (2016). We finally take survival probabilities from the life tables published by the Dutch Royal Actuarial Society. In the second step, we estimate the parameters that govern the exogenous evolution of wage and health using a regression approach (results and details are presented in Appendix 2.9.2.2). Finally, we estimate preferences using the Method of Simulated Moments (MSM).

The goal of the MSM estimator is to find the preference vector that yields simulated life-cycle decision profiles that 'best match' (as measured by a GMM criterion function) the profiles from the data. Due to the poor small–sample properties of the optimal weighting matrix, we use a diagonal weighting matrix that contains only inverses of the estimated variances of sample moments on the diagonal (Altonji and Segal, 1996). Further details on the MSM estimator are presented in Appendix 2.9.2.3, and targeted moments are discussed below. We use initial conditions from cohorts born in January 1950 and June 1953 to construct the simulated profiles (summary statistics in Table 2.4), as well as random realizations for wages, survival and disability status according to the externally estimated process in the second step.

The minimization of the MSM objective function is complicated because the objective function is not uniformly differentiable and has multiple local minima. Similarly to Theloudis (2018), we combine a global with a local optimizer. We start with a simulated annealing algorithm as in Goffe et al. (1994). We then use the best guesses of the global

[51] While this is somewhat lower compared to what is typically used in the literature for similar models, previous work mainly used data on different time periods. One of the latest works, De Nardi et al. (2024), sets $r$ slightly higher to 2% but uses British data for the period 1991-2008. We haven't estimated the model with a different interest rate. With the current estimates and an interest rate of 2%, the model predicts slightly less labour supply and slightly more early claiming, but savings clearly grow much faster over time.

optimizer as starting values for the subplex algorithm Rowan (1990).

We only use the birth cohorts 1950 and 1953 to construct the initial conditions, from which we simulate decisions, and the targeted moments. Our estimation, therefore, exploits only the exogenous variation in pension rules generated by the state pension reform. This means that our estimates are designed to match the retirement decisions of these two cohorts, and therefore also the difference between them. Still, it does not mean that we would necessarily achieve a good fit if the model is miss-specified. Instead, we do not target moments from the 1949 cohort, which have access to generous early retirement benefits. This means that we can use the 1949 cohort to validate the model estimates. The reason to exploit the early retirement reform for validation and the state pension reform for estimation, and not the other way around, is that the former reform had stronger effects on retirement choices as the changes in pension rules were more drastic. Therefore, it provides a more demanding test to check whether our model can replicate well out-of-sample behaviour. We present results when using the early retirement reform for estimation in Appendix 2.9.3.6, i.e. when we switch the role of the reforms used for estimation and validation, and show that results are very similar.

| Variables | 1949 | 1950 | 1953 |
|---|---|---|---|
| Savings (2006 euros) | 145,865 | 144,003 | 136,062 |
| | (101,621) | (110,195) | (108,582) |
| Full-time wage (2006 euros) | 47,535 | 47,871 | 49,831 |
| | (16,498) | (16,529) | (17,171) |
| Accrued years of pension | 29.73 | 29.77 | 29.81 |
| | (7.63) | (7.09) | (6.88) |
| Part-time work (%) | 0.09 | 0.08 | 0.06 |
| | (0.29) | (0.27) | (0.23) |
| Disability (%) | 0.03 | 0.04 | 0.02 |
| | (0.17) | (0.21) | (0.15) |
| Individuals | 838 | 893 | 880 |

Table 2.4: Average initial conditions by cohort.

Note: Standard deviations in parentheses. In the model, the state variable is accrued pension rights and not accrued years. We report accrued years for ease of comparison across cohorts because differences in accrued rights do not necessarily reflect differences in final benefits due to different pension regimes. Savings are adjusted as discussed in Appendix 2.9.2.4.

**Targeted moments** For each pension regime (i.e. for the 1950 and 1953 cohorts) and each age at which choices are active, we target average wealth, the pension claiming rate, the full-time and the part-time rate for healthy people, the employment rate for sick

people, and the partial retirement rate. In total, we target 112 moments, of which 55 for cohort 1950 and 57 for cohort 1953. We access sickness and wealth data by linking our sample to administrative information from Statistics Netherlands. As we model annual decisions, we match annual life-cycle profiles. Labour supply and claiming profiles are measured annually on the birthday. Wealth, however, is only measured on January 1st and therefore, for each individual, we use the closest measure to their birthday to compute average savings at the respective ages.

For labour participation and benefit claiming, we simply target the average level over age for the estimation sample. For wealth, however, we are concerned that inflation and business cycle fluctuations might bias our model estimates because we do not model macroeconomic trends. Therefore, we first deflate wealth using a Consumer Price Index. Second, we use a regression approach to net out year effects. In practice, we use for this task a larger sample (cohorts from 1919 to 1956) and regress wealth on age and calendar year dummies. We then subtract the estimated coefficients for calendar years from observed wealth in the corresponding years, and subsequently compute the average at each age for our estimation sample. For both adjustments we use 2006 as the baseline year (when cohorts 1949 and 1950 enter the model). Details in Appendix 2.9.2.4.

**Identification and simulation exercise** Since all parameters are affected by all moments in complex non-linear models, it is difficult to judge how the identification of each parameter relates to a specific profile. Heuristically, the leisure cost of poor health is driven by the difference in labour supply between people with good and bad health conditions, and the weight of leisure is driven by the combination of the employment and the part-time rate. It is more difficult to relate the remaining parameters to specific groups of moments since they all have profound effects on labour supply, benefit claiming and assets accumulation. We verified that the moments identify these parameters by fixing key parameters at different levels and estimating the remaining parameters or by checking how the value of the MSM objective function changes.

We also conduct a simulation exercise to confirm that our model is estimable. In particular, we want to check if we are able to recover the underlying parameters when we know the true data generating process. Given the model solution corresponding to an arbitrary preference vector, we simulate the life-cycle profiles starting from the initial conditions. The goal of this exercise is to estimate preferences using these simulated profiles as if they were the 'real' data. This exercise confirms that the model is estimable (results not presented).

## 2.6 Results

This section presents the results of the model estimation and fit when using cohorts 1950 and 1953, that is exploiting the exogenous variation induced by the state pension reform. Estimates for the preference parameters are shown in Table 2.5 (computation of the standard errors is discussed in Appendix 2.9.2.3). While context-specific, our estimates are largely in line with those in earlier studies. We estimate $\lambda$, the intertemporal elasticity of substitution of consumption, to be around 0.74, very close to Gustman and Steinmeier (2005)'s estimate of 0.79. The time cost of bad health, $\delta$, is estimated to be about 643 hours per year, which is somewhat larger but comparable to estimates from French (2005), de Bresser (2023) and French and Jones (2011) – between 130 and 500. However, our definition of bad health only includes severe cases, explaining the higher estimate. The estimate suggests that the marginal utility of leisure is significantly higher for sick people, and that sickness implies a reduction of around 16% in disposable time. The relative weight of the bequest motive and its curvature, $b_1$ and $b_2$, are estimated to be around 49 and 270 thousand euros and fall within the (large) interval provided by past studies (French and Jones, 2011; de Bresser, 2023). It is more difficult to compare estimates of preferences over leisure, $\gamma$ and $\psi$, due to differences in functional form assumptions.

The first and second columns of Figures 2.5 and 2.6 show that the model performs well in replicating the targeted moments of labour supply, pension claiming, and saving decisions. In particular, the model replicates well the share of people in partial retirement for both cohorts. Notably, the increase in partial retirement at age 62 corresponds to a similar increase in the part-time work rate, suggesting that it reflects mainly people moving from a full-time work position into partial retirement. Figure 2.15 in the Appendix further reports the average full-time equivalent and simulated consumption levels (not targeted).

| Parameter | Estimate | Std.Err. |
|---|---|---|
| $\lambda$ | 0.736 | 0.010 |
| $\gamma$ | 0.870 | 0.701 |
| $\psi$ | 0.040 | 0.007 |
| $\delta$ | 643.974 | 150.910 |
| $b_1$ | 49.617 | 23.811 |
| $b_2$ | 272,556.868 | 10,461.200 |
| $u(c_t, l_t) = \frac{\lambda}{\lambda-1} c_t^{\frac{\lambda-1}{\lambda}} + \psi \frac{\gamma}{\gamma-1} l_t^{\frac{\gamma-1}{\gamma}}$ | | |
| $l_t = (4,000 - h_t - \delta I\{health_t = bad\})/4{,}000$ | | |
| $B(a_{t+1}) = b_1 \frac{\lambda}{\lambda-1}(b_2 + a_{t+1})^{\frac{\lambda-1}{\lambda}}$ | | |

Table 2.5: MSM estimates and asymptotic standard errors.

**Retirement decisions by wage and wealth** Before turning to the model validation using the early retirement reform, we investigate heterogeneity in retirement decisions and whether the model is able to replicate them. We define two outcome variables: A dummy which equals one if an individual ever participates in partial retirement and a variable representing the individuals' retirement age. We do this separately using the administrative data and our model simulations, for a total of four outcomes. We then regress them on full-time wage and wealth measured when entering the model. Table 2.6 presents the results.

First, the data suggest that wage is positively (and significantly) correlated with both the probability of retiring partially and the average retirement age. Our model replicates both the sign and the magnitude of these correlations. These correlations suggest that a higher cost of leisure leads people with a higher wage to retire later, but also that only people with higher wages might afford to retire partially, potentially because they do not have to reduce their consumption when moving from full-time to part-time work. On the other hand, wealth is negatively correlated with both outcomes: People with higher wealth can afford to retire earlier and partially retire less often, which means that if they reduce work hours as they age they do not need to rely on partial pension income. The model replicates well that the correlation with partial retirement is negative but close to zero, but it overstates the magnitude of the negative correlation with the retirement age. Overall, the results suggest that the model captures fairly well retirement heterogeneity driven by wage and wealth, even though these margins are not targeted by the estimation.

| | | Data | | Model |
|---|---|---|---|---|
| | | Estimate | 95% CI | Estimate |
| Partial retirement | Wage (10,000) | 0.0086** | [0.0007;0.0164] | 0.0186 |
| | Wealth (10,000) | -0.0012* | [-0.0024;0.0000] | -0.0002 |
| Retirement age | Wage (10,000) | 0.1921*** | [0.1383;0.2459] | 0.1441 |
| | Wealth (10,000) | -0.0213*** | [-0.0311;-0.0116] | -0.1096 |

Table 2.6: Correlations between wage and wealth and retirement decisions.

**Validation** We further validate our model estimates by simulating the life-cycle profiles of the 1949 cohort and comparing them to the data. This implies adjusting the model to incorporate the rules of the pension regime that applies to the 1949 cohort, which are different from the ones used for estimation. It also implies simulating the model starting from the initial conditions (assets, wage, health, etc.) observed in the data for this cohort.[52]

[52]Because the 1949 cohort can make use of effectively two different pension schemes, the early retirement and the old age pension scheme, we need to add a state variable in the model with respect to the one used

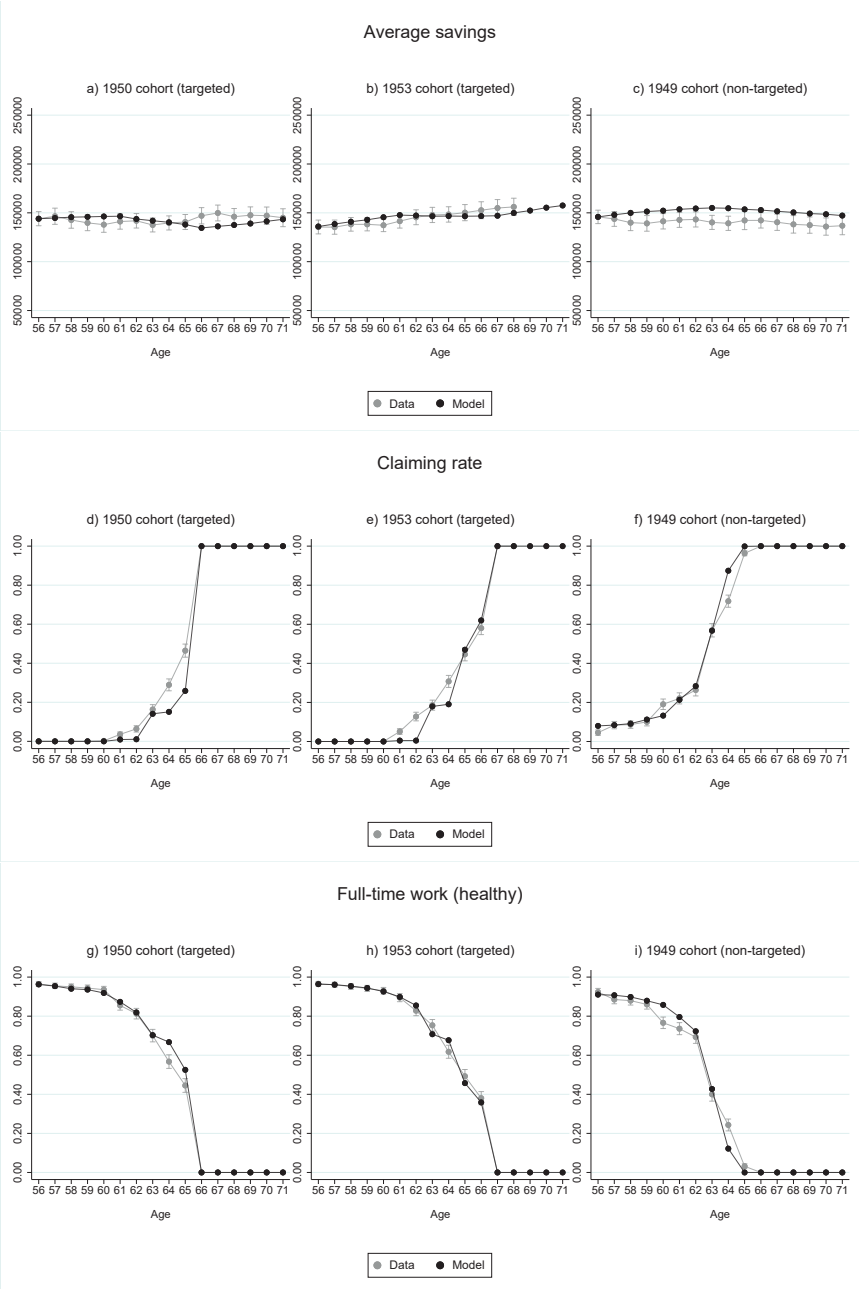



Figure 2.5: Model fit for targeted and non-targeted cohorts.

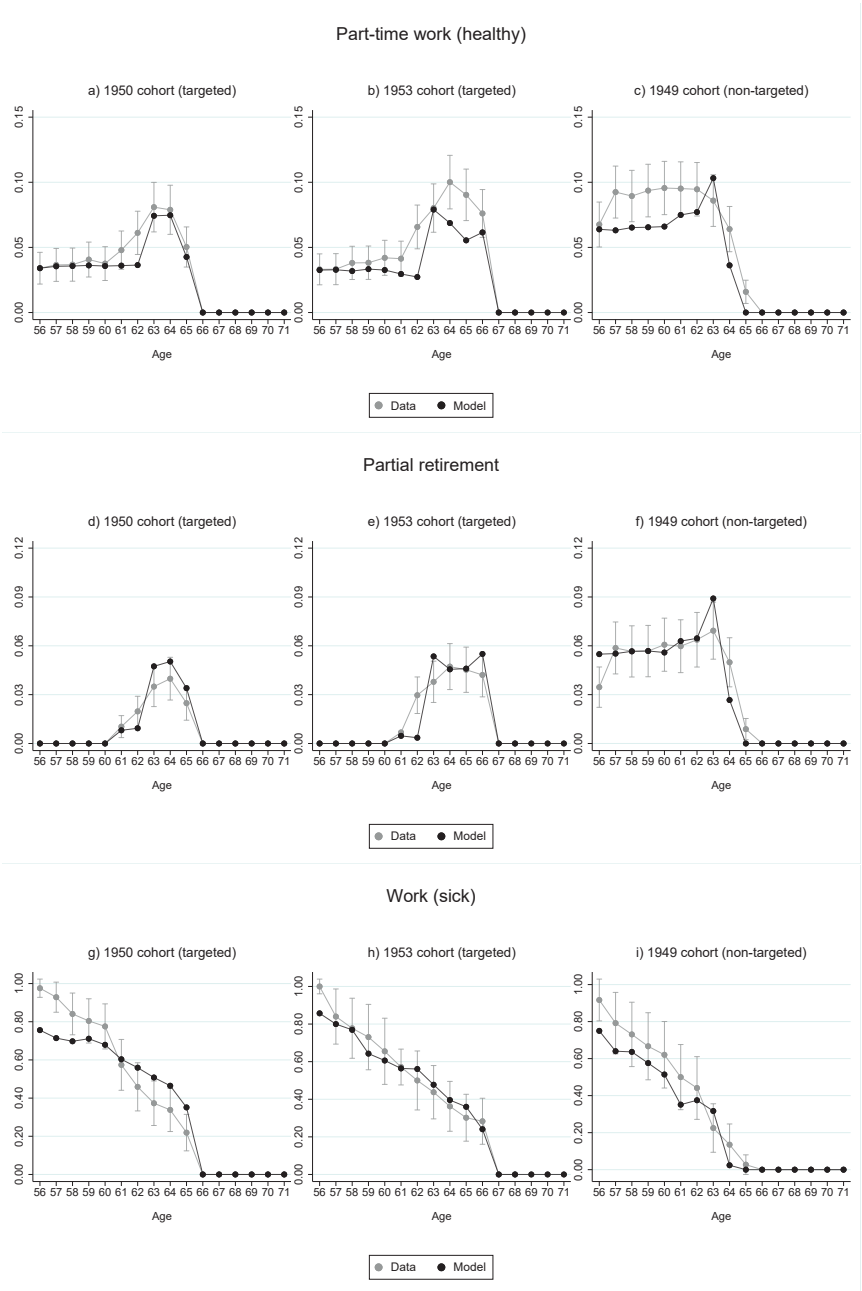



Figure 2.6: Model fit for targeted and non-targeted cohorts.

The results are presented in the third column of Figures 2.5 and 2.6. The figures show that the model replicates pension claiming and full-time work decisions very well. In particular, the model is able to capture the differences across the pension regimes. At age 63, 60% of people claim pension benefits in the 1949 cohort while this is only 20% for the 1950 and 1953 cohorts. Similarly, among healthy people aged 63, the full-time work rate is 40% for the 1949 cohort and around 70% for the other two cohorts. The model also performs well with respect to the labour supply choices of sick people (who can only work part-time). The fit is somewhat less good for the part-time work choices of healthy people: The model slightly under-predicts the part-time rate between age 57 and 60. However, the model fit is still good for partial retirement decisions, which is the main object of interest. In particular, the three cohorts markedly differ with respect to the evolution of the partial retirement rate over ages and the model is able to mimic these differences. To highlight these differences, Figure 2.7 reports the difference in the partial retirement rate between birth cohorts 1950 and 1949 in panel a), and between birth cohorts 1953 and 1950 in panel b). It shows that the model is able to capture well the direction and the magnitude of the effect of both reforms. This suggests that our estimates reflect policy-invariant preference parameters which can be used for counterfactual analysis.

## 2.7 Policy simulations

In this section, we present the results of two counterfactual policy simulations. First, we investigate the effect of allowing people to retire partially on their labour supply, their well-being, and on the budgets of the government and of the occupational pension fund. Second, we study the effect of increasing the state pension age by one additional year, as planned by the Dutch law. Our simulations are conditional on the observed state variables when entering the model, meaning that we study the effects of unexpected and permanent policy changes implemented when individuals are 55 years old.

### 2.7.1 The value of partial retirement

In this section, we quantify the effect of partial retirement on labour supply under the three considered pension regimes. We simulate choices for the three cohorts, with and without the partial retirement option, for a total of six different policy scenarios. With

for cohorts 1950 and 1953. This additional state variable allows to model the evolution of pension rights and benefits under both schemes. In practice, this simply means that the way how pension income is computed is more complicated. This additional state variable does not bring in any additional parameter, meaning that it is essentially switched off for cohorts 1950/1953. Appendix 2.9.3.6 also verifies that this change is innocuous, because using (also) the 1949 cohort and model for estimation does not change the results.

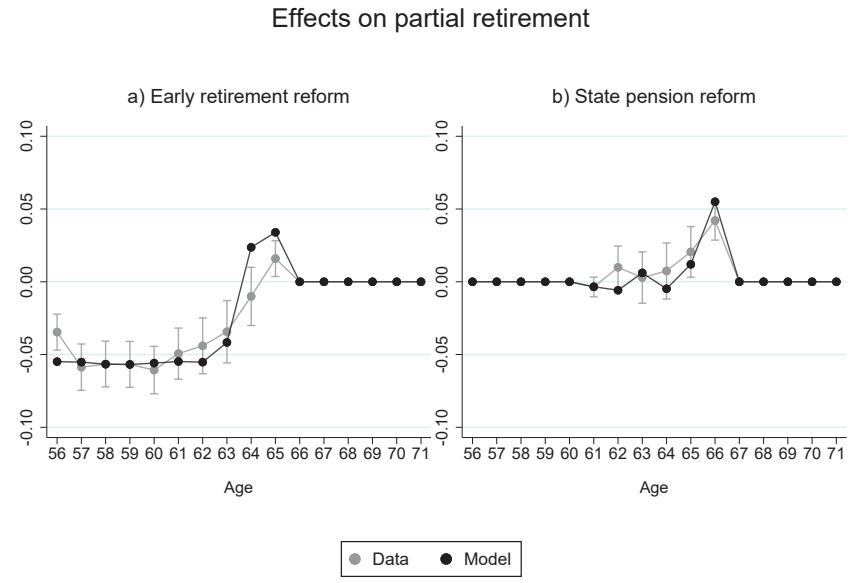



Figure 2.7: Model fit for the effects of the two reforms on partial retirement.

respect to the baseline scenario, we restrict choices from the early retirement age until the state pension age by excluding the two partial retirement options.

The panels in the first row of Figure 2.8 present the differences in the share of people working, working full-time, and working part-time, comparing the scenario that allows partial retirement with the scenario that does not. First, as expected, we find a positive effect of partial retirement at the extensive margin with people being around 3 percentage points more likely to work between age 62 and the state pension age when the partial retirement option is available for cohorts 1950 and 1953. The effect is smaller for cohort 1949. Second, we confirm that this increase in the employment rate comes (partially) at the expense of full-time employment, which decreases by around 1 or 2 percentage points at the same ages. Third, we find that the part-time rate increases somewhat more than the employment rate. That is, some of the people that work part-time through partial retirement would otherwise work full-time, while others would not work at all.

To quantify which effect is stronger, in panel d) we report the change in the total number of work hours between the same two scenarios, divided by the number of people (alive) at each age times 100. A value of +0.5 FTE/100 persons can be interpreted as

follows: Among every 100 people, partial retirement makes one individual switch from not working to half part-time work. The figure shows that the effect is positive at all ages for cohorts 1950 and 1953. Instead, the effect is negative at most ages for cohort 1949. Over the entire period from age 56 to age 66 (i.e. summing up the bars in the left panel), the net effect is negative but close to zero for cohort 1949, but positive for cohorts 1950 and 1953. In panel e) we report the percentage changes in the total number of hours worked due to the availability of partial retirement, which suggest again that the effects are sizeable from age 63, when most people retire. Overall, these results suggest that the effect of partial retirement is heterogeneous across pension regimes, but it's positive under the current one (i.e. cohort 1953). In particular, the abolished early retirement scheme provides an incentive to claim pension benefits before the state pension age, which would otherwise be lost. Because benefits cannot be claimed while working full-time, the early retirement scheme provides an incentive to stop working but also to move from full-time work to partial retirement. Once this incentive is abolished, partial retirement has a positive effect of labour supply, meaning that it is more often used as a substitute for full retirement.

We also quantify how much partial retirees value the flexibility given by partial retirement following the approach in De Nardi et al. (2016). We define the value function of individual $i$ when entering the model ($t = 1$) under the two scenarios as

$$V_{i1}(X_{i1}, PR = 0)$$
$$V_{i1}(X_{i1}, PR = 1)$$

where $PR$ equals one if the partial retirement option is available and zero otherwise, and $X_{i1}$ refers to the initial value of the state variables. We define the equivalent variation ($EV$) as the monetary amount the agent is willing to accept – when entering the model – in lieu of the partial retirement option, and the compensating variation ($CV$) as the monetary amount the agent is willing to give up to have the same option. That is, the $EV_i$ ($CV_i$) is the individual-specific monetary amounts which is added to (subtracted from) the savings available when entering the model such that equation 2.12 (2.13) holds.

$$V_{i1}(X_{i1}, PR = 0, EV_i) = V_{i1}(X_{i1}, PR = 1) \tag{2.12}$$
$$V_{i1}(X_{i1}, PR = 0) = V_{i1}(X_{i1}, PR = 1, CV_i) \tag{2.13}$$

In general, the monetary amounts of the equivalent and compensating variation can differ because of the different policy regimes at which compensation is assumed to occur in these two measures of welfare change. Focusing on people in the cohorts 1950 and 1953 who would make use of partial retirement if it was available, we find an average $EV$ and $CV$ of around 400 euros, but with valuations ranging from zero to 5,000 euros with a long right tail in the distribution. The average monetary valuation is not large per se, but

should be interpreted carefully. In the counterfactual scenario without partial retirement, individuals can still work part-time. In fact, they can still retire gradually and use private savings to smooth consumption and top up labour income. However, our result suggests that the average worker (in this selected group) would be willing to pay around 400 euros to be able to flexibly allocate pension income over time while working more hours in the years preceding full retirement (as shown by Figure 2.8). Essentially, partial retirement allows to move private resources across time and therefore improves workers' welfare, despite the fact that people decide to work more. The average valuation of 400 euros can be seen as the price people are willing to pay to borrow from their 'future self' pension. While partial retirement cannot decrease workers' welfare (at least in a partial equilibrium model), we expect its benefit to be larger for those with lower assets, that is people who cannot finance a gradual retirement path out of private savings. We thus regress the equivalent and compensating variation on assets, wages, and pension rights as observed when entering the model. In fact, for both outcomes, we find a negative and significant correlation between the monetary valuation of partial retirement and assets, but not with wages and pension rights (see Table 2.13 in Appendix 2.9.3.7). For people in the bottom decile of the wealth distribution, the partial retirement option is as valuable as 9% of their wealth.

On top of its effect on labour supply and well-being, partial retirement has broader implications for the government budget and for that of the occupational pension fund. For the government, the expenses with or without the partial retirement option are the same, because partial retirement can only be done with the occupational pension, while the state pension is always paid at the state pension age. However, revenues will be different. If partial retirement changes labour supply decisions, then this possibly affects (i) the income taxes and social contributions collected by the government while people work, (ii) the income taxes paid on occupational pension, which changes if people have longer careers, (iii) the wealth taxes if people save differently when working more or less due to partial retirement. We thus compute the taxes and social contributions paid under the two scenarios, with and without the partial retirement option, over the life-cycle. Similarly, for the occupational pension fund, we need to compute the difference between the pension premia collected by the fund and the pension benefits paid by it. Table 2.7 summarizes the results. On average, each partial retiree pays 4,644 EUR more in taxes and social contributions compared to the scenario without partial retirement. The additional revenues for the government per partial retiree are sizeable, as they correspond to around six months of state pension benefits. The net revenues for the pension fund are even larger, and amount to around 6,654 EUR per partial retiree. However, since around 8% of the sample retire partially, the additional revenues per person are about 350 and 500 EUR for the government and the pension fund, respectively.

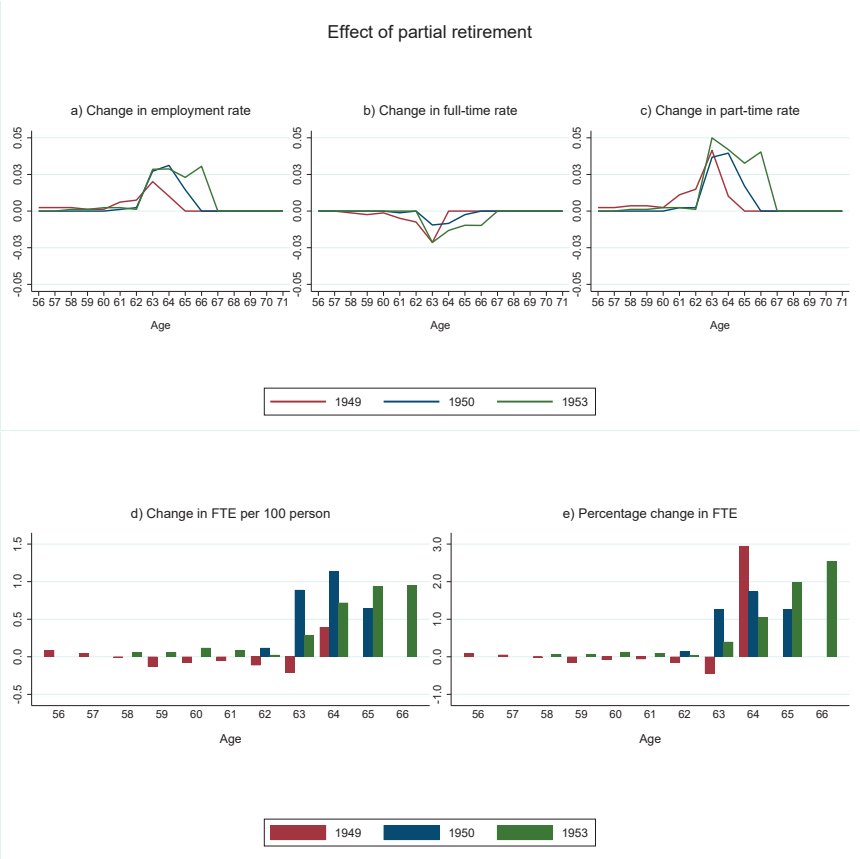



Figure 2.8: The effect of partial retirement on labour supply.

| Budgetary savings (EUR) | Government | Pension fund |
|---|---|---|
| Per person | 351 | 502 |
| Per partial retiree | 4,644 | 6,654 |

Table 2.7: Implications of partial retirement for the budgets.

### 2.7.2 Increasing the state pension age

In this section, we investigate the effects of a further increase in the state pension age. This exercise reflects the increase planned by the Dutch pension law, as explained in Figure 2.1.[53] The increase of the state pension age implies two changes. First, the state pension will be received one year later. The actuarial factors applied in case of early claiming are adjusted with respect to the new state pension age, while the accrual rate does not change. This means that, keeping the work history constant, the corresponding total pension amount will be lower. Second, employees can keep working one year longer because the automatic job termination is also postponed.

We conduct this simulation exercise focusing on the 1953 cohort, for which the state pension age is 66 years and 4 months and therefore can work in the model until age 66 but not until 67 (baseline scenario). We adjust the baseline scenario by incorporating the changes described above. In particular, in the counterfactual scenario we assume that people can work until age 67, but not age 68 (see Table 2.10 for the actuarial factors). We take initial conditions from cohort 1953 and simulate behaviour under both scenarios using the same random shocks to survival, wage, and health.

Figure 2.9 presents the results. Average savings are essentially unchanged compared to the baseline. The claiming rates are fairly similar until age 64 but are markedly different at later ages as people postpone pension claiming due to the higher state pension age. Similarly, the full-time employment rate is higher at the same ages. Our model also predicts that the part-time rate will be higher already as of age 61, when part-time work can be combined with partial pension recipiency, in line with patterns from Figure 2.2. The partial retirement rate shows a trend that is very similar to that of part-time work. The share of people who choose partial retirement at some point increases from around 9 to 14% of the sample. This highlights the attractiveness of partial retirement when the state pension age increases. The employment rate for sick people shows, again, that employees would be more likely to work in the few years before the higher state pension age due to the hypothetical reform. However, the panels in the last row show that while the reform increases the average number of hours worked yearly due to people being more likely to work, there is a negative effect at the intensive margin. In fact, the reform increases the share of people working part-time among those employed, in line with the evidence presented in Section 2.4.

Overall, this exercise suggests that the planned increase in the state pension age will increase labour supply among older workers, although it will also increase the share of people working part-time. Findings are in line with those presented in Section 2.4 for the

[53]The first people who will reach the new state pension age of 67 years and 3 months are those born in 1961, meaning that only in 2028 we will be able to judge the effect of such increase.

state pension reform. At the same time, the labour supply just before the state pension age is reduced by almost half. In the baseline scenario, the average person works for around 0.4 FTE at age 66. In the counterfactual scenario, however, this drops to 0.2 at age 67. This suggests that marginal returns from increasing the state pension age are decreasing, raising the question to what extent labour supply at old age can be further increased with similar policies.

It's worth noting that 4% of the sample is better off under the counterfactual scenario, meaning that their value function when entering the model is higher under the counterfactual case. That's because, for some people, the gain from postponing the automatic job termination by one year more than compensates for the negative wealth effect induced by the hypothetical reform. As we explore in Appendix 2.9.3.8, increasing flexibility along this dimension might also increase labour supply and workers' welfare at the same time.

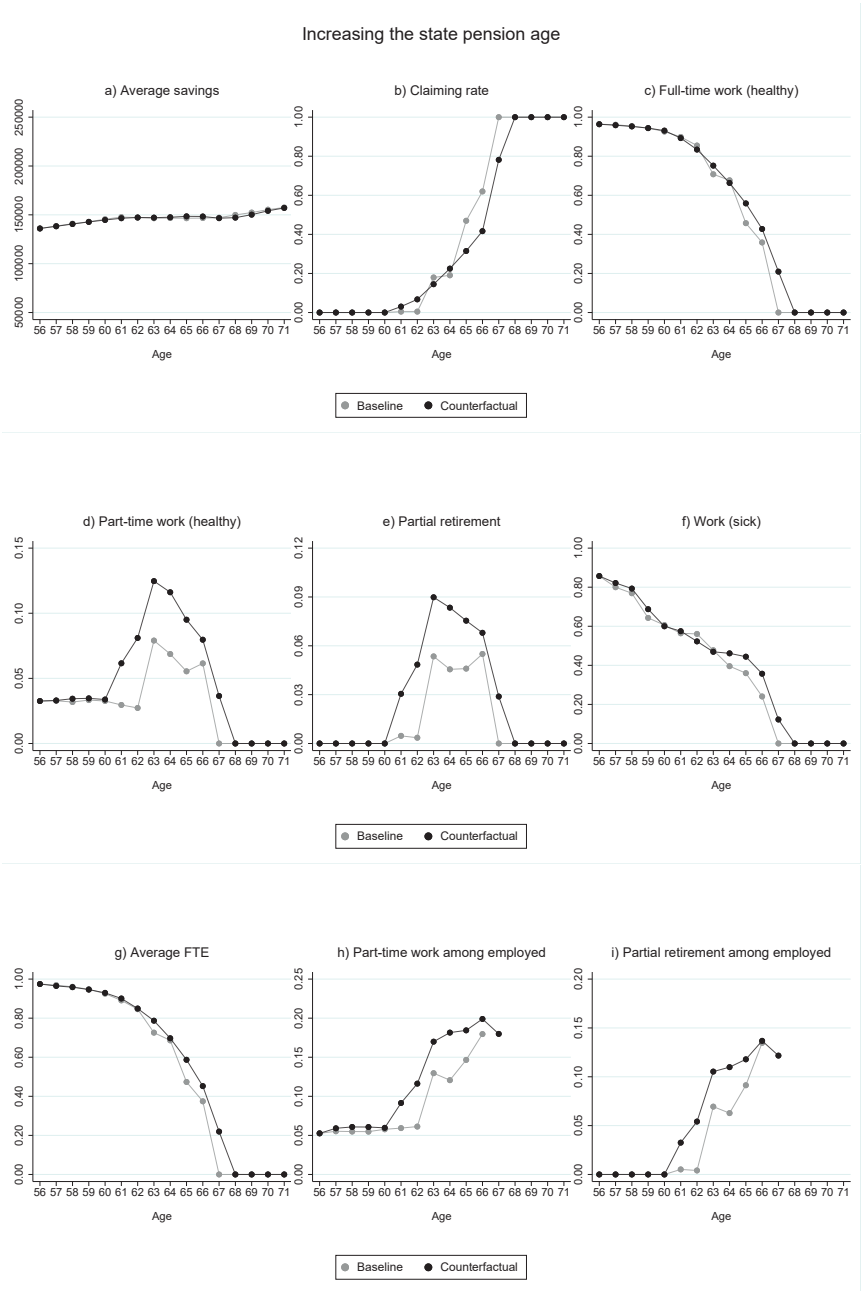



Figure 2.9: The effects of further increasing the state pension age.

## 2.8 Conclusion

This paper studies the effect of partial retirement on labour supply at old age and retirees' well-being. We start by exploiting two pension reforms implemented in the Netherlands to show how different pension regimes affect labour supply decisions at old age also due to their effects on part-time work choices. First, we document that part-time work at old age is often combined with partial pension income, that is partial retirement. Second, we show that the 2006 reform, which abolished a generous early retirement scheme, led people to work longer but also to work full-time more often before retirement. This is because the early retirement scheme provided a financial incentive to claim early retirement benefits before the state pension age, which would otherwise be lost. Because pension benefits cannot be claimed while working full-time, the scheme induced people to stop working but also to move from full-time work to partial retirement. Third, we show that the 2011 reform, which increased the state pension age, resulted in a higher share of people working part-time already a few years before the old state pension age. In other words, on average, people retire later but also work fewer hours in the years preceding full retirement.

Based on these findings, we develop a structural model of retirement which accounts for assets and pension rights accumulation, the bunching of work hours at four discrete levels, and the possibility to retire partially. We estimate the model exploiting the exogenous variation stemming from the 2011 state pension reform, while we use the 2006 early retirement reform, which greatly changed retirement behaviour, to validate our model estimates. Given that the model is able to replicate the effects of non-targeted policy changes well, we use it for counterfactual policy simulations.

In a first policy experiment, we find that the net effect of partial retirement on labour supply is heterogeneous across pension regimes, but positive under the current reformed scheme in the Netherlands. In this case, the positive effect on total work hours is substantial and up to 2.5 percent at age 66. We thus show that partial retirement increases labour supply and workers' well-being at the same time, with poorer workers benefiting most. For people in the bottom decile of the wealth distribution, we find that the partial retirement option is as valuable as 9% of their wealth. We further show that the positive effect on labour supply translates into higher revenues for the government, which comes form taxes and social contributions, and lower net expenditures for the pension fund.

A second policy experiment confirms that further increasing the state pension age increases the average retirement age, but it also stimulates part-time work before retirement. Moreover, it shows that marginal returns – in terms of labour supply – from increasing the state pension age are decreasing, raising the question to what extent labour supply at old age can be further increased with similar policies. Our analysis, however, is limited to public sector employees and external validity of our results for other sectors remains

to be examined, although the public sector covers a large share of the workforce.

# 2.9 Appendix

## 2.9.1 Details on the model setup

### 2.9.1.1 Timeline of the model

Table 2.8 presents the timeline of the model for a healthy person born in 1950. A sick person would face a different choice set for $h_t$ as he/she cannot work full-time nor 0.8 FTE. A person born in 1949, instead, could claim pension benefits already from 56, meaning that the choice set for $op_t$ would be unrestricted from the beginning of the model. A person born in 1953 could still work at age 66.

The state pension age for birth cohorts 1949 and 1950 is 65 and 3 months. Since we model annual choices and match them to actual behaviour as observed on the birthday, we assume that people from these cohorts can decide to work at age 65 but cannot work at age 66. The state pension age is therefore 66 in our model for cohorts 1949 and 1950 and 67 for cohort 1953. Similarly, the early retirement age is 61 for birth cohorts 1950 and 1953 and 56 for birth cohort 1949.

| Period | Age | Choice set $h_t$ | Choice set $op_t$ | Possible status $ret_t$ | Notes |
|---|---|---|---|---|---|
| 1 | 56 | 0,0.5,0.8,1 | 0 | 1,3,5,7 | |
| 2 | 57 | 0,0.5,0.8,1 | 0 | 1,3,5,7 | |
| 3 | 58 | 0,0.5,0.8,1 | 0 | 1,3,5,7 | |
| 4 | 59 | 0,0.5,0.8,1 | 0 | 1,3,5,7 | |
| 5 | 60 | 0,0.5,0.8,1 | 0 | 1,3,5,7 | |
| 6 | 61 | 0,0.5,0.8,1 | 0,1 | 1,3,5,7 | Can start claiming |
| 7 | 62 | 0,0.5,0.8,1 | 0,1 | 1,...,8 | |
| 8 | 63 | 0,0.5,0.8,1 | 0,1 | 1,...,8 | |
| 9 | 64 | 0,0.5,0.8,1 | 0,1 | 1,...,8 | |
| 10 | 65 | 0,0.5,0.8,1 | 0,1 | 1,...,8 | Last period to work |
| 11 | 66 | 0 | 1 | 1,...,8 | Mandatory retirement |
| 12 | 67 | 0 | 1 | 8 | |
| ... | ... | ... | ... | ... | |
| 45 | 100 | 0 | 1 | 8 | Last period (if alive) |

Table 2.8: Timeline of the model for a healthy person born in 1950

Note: The labour supply choice $h_t$ is expressed in terms of full-time equivalent. The claiming choice $op_t$ takes value 1 when claiming occupational pension and 0 otherwise. The retirement status is the status at the beginning of $t$, before making any choice and reflecting choices in $t-1$, as explained in Table 2.9.

| $ret_{t+1}$ | | $op_t$ | |
|---|---|---|---|
| | | 0 | 1 |
| $h_t$ | 1.0 | 1 | 2 |
| | 0.8 | 3 | 4 |
| | 0.5 | 5 | 6 |
| | 0.0 | 7 | 8 |

Table 2.9: Retirement status evolution

Note: The table shows how work and claiming choices this year, $h_t$ and $op_t$, affect retirement status next year, $ret_{t+1}$. $h_t$ is expressed here in terms of FTE, and $op_t$ as a binary claiming (1)/not claiming (0) variable. Status 2 cannot realize and $ret_{t+1} \geq ret_t$ in the model.

#### 2.9.1.2 Taxes and contributions

This tax function is based on OECD (2004), with all nominal amounts expressed in 2006 Euros using the CPI published by Statistics Netherlands. Taxes are levied at the individual level in the Netherlands. The starting point is gross income, which is defined as the sum of earnings, DI benefits, and pension benefits

$$Grossinc_t = Earnings_t + DI_t + b_t$$

Workers pay unemployment contributions and contributions to public health based on their gross income, which are then deducted from gross income to arrive at taxable income

$$UIcontr_t = \begin{cases} 0 & Grossinc_t < 15,562 \\ 0.058 \times (Grossinc_t - 15,562) & 15,562 \leq Grossinc_t < 44,800 \\ 0.058 \times (44,800 - 15,562) & Grossinc_t \geq 44,800 \end{cases}$$

$$Pubmed_t = \begin{cases} 0.0125 \times Grossinc_t + 399 & Grossinc_t < 30,320 \\ 0.0125 \times 30,320 + 399 & 30,320 \leq Grossinc_t < 33,514 \\ 0 & Grossinc_t \geq 33,514 \end{cases}$$

$$Taxable_t = \max\{0;\, Grossinc_t - UIcontr_t - Pubmed_t\}$$

Social insurance contributions and income taxes are levied on taxable income. Social insurance contributions only pertain to the first two taxable income brackets and discriminate by age

$$Soc_t = \begin{cases} 0.324 \times Taxable_t & Taxable_t < 30,371 \text{ \& until state pension age} \\ 0.324 \times 30,371 & Taxable_t \geq 30,371 \text{ \& until state pension age} \\ 0.145 \times Taxable_t & Taxable_t < 30,371 \text{ \& from state pension age} \\ 0.145 \times 30,371 & Taxable_t \geq 30,371 \text{ \& from state pension age} \end{cases}$$

$$IncTax_t = \begin{cases} 0.01 \times Taxable_t & Taxable_t < 16,721 \\ 0.01 \times 16,721 & \\ \quad +0.0795 \times (Taxable_t - 16,721) & 16,721 \leq Taxable_t < 30,371 \\ 0.01 \times 16,721 & \\ \quad +0.0795 \times (30,371 - 16,721) & \\ \quad +0.42 \times (Taxable_t - 30,371) & 30,371 \leq Taxable_t < 52,072 \\ 0.01 \times 16,721 & \\ \quad +0.0795 \times (30,371 - 16,721) & \\ \quad +0.42 \times (52,072 - 30,371) & \\ \quad +0.52 \times (Taxable_t - 52,072) & Taxable_t \geq 52,072 \end{cases}$$

A general tax credit of 1,876 euro applies and is deducted from income tax. Also, a work credit is deducted from income tax which depends on earnings

$$WorkCredit_t = \begin{cases} 0.01753 \times Earnings_t & Earnings_t < 8,328 \\ 0.01753 \times 8,328 & \\ \quad +0.11213 \times (Earnings_t - 8,328) & 8,328 \leq Earnings_t < 18,147 \\ 0.01753 \times 8,328 & \\ \quad +0.11213 \times (18,147 - 8,328) & Earnings_t \geq 18,147 \end{cases}$$

Wealth is taxed at a rate of 1.2% above the threshold of 39,568 euro

$$WealthTax_t = \begin{cases} 0 & a_t < 39,568 \\ 0.012 \times (a_t - 39,568) & a_t \geq 39,568 \end{cases}$$

Workers also pay a pension premium to the occupational pension fund at a rate of 6.36% for wages above a state pension offset of 9,839 2006 euro

$$ABPpremium_t = \begin{cases} 0 & W_t < 9,839 \\ 0.0639 \times FTE_t \times (W_t - 9,839) & W_t \geq 9,839 \end{cases}$$

The net income is then given by

$$\begin{aligned} NetIncome_t = Grossinc_t - UIcontr_t - Pubmed_t - Soc_t - IncTax_t \\ + 1,876 + WorkCredit_t - WealthTax_t - ABPpremium_t \end{aligned}$$

#### 2.9.1.3 Pension rights and benefits

Since we model retirement decisions in different pension schemes that are specific to different birth cohorts, here we explain how we calculate pension rights across the different schemes. We closely follow the information reported in the FPU regulation (version January 1, 2014) and the pension regulation (version January 1, 2015) published by ABP.[54]

**Cohort 1949** For each individual, pension rights when entering the model are calculated as

$$PR_1 = Years \times 0.0175 \times (W - 15,000)$$

The accrued component depends on the number of accrued years (provided by ABP for each individual), an accrual rate of 1.75%, a state pension offset of 15,000 euros and the last earned wage. The number of accrued years already takes into account the work history in terms of full-time or part-time work (the number of years is the sum of the full-time equivalent worked over the years).

Given the initial condition as described above, in the the model pension rights continue to accumulate as follows

$$PR_{t+1} = PR_t + FTE_t \times 0.0175 \times (W_t - 15,000)$$

Pension rights are further adjusted to take into account that actuarial penalties apply for early claiming, but only to the share of pension rights that is claimed (that is 100% for full retirement and $(1 - FTE_t)$% for partial retirement). Furthermore, when claimed early, pension benefits are increased by a basic component of 15,000 euros (to which the actuarial adjustment also applies). As of the state pension age, instead, pension benefits include the state pension of 9,600. Importantly, old age pension benefits are not penalized due to early claiming.

$$b_t = \begin{cases} [early\ retirement\ rights_t] \times (1 - FTE_t) \times Act.Adj._{\cdot t,cohort} & \text{if } t < 66 \\ [PR_t + state\ pension] \times (1 - FTE_t) & \text{otherwise} \end{cases}$$

**Cohort 1950** For each individual, pension rights when entering the model are computed as the sum of two components: The accrued rights as of age 56 and the compensation for the abolishment of the early retirement scheme

$$PR_1 = Accrued + Compensation$$

[54] Available at https://abppensioen.nl/wp-content/uploads/2018/05/FPU-reglement-2014-1.pdf and https://abppensioen.nl/wp-content/uploads/2018/05/ABP-Pensioenreglement-2015.pdf, respectively.

The accrued component depends on the number of accrued years, an accrual rate of 1.75%, a state pension offset of 15,000 euros and the last earned wage. Note that the accrual rate and the state pension offset are those which applied to rights accrued before the occupational pension reform of 2006, because cohort 1950 enters the model at age 56. In theory, until 2004 only the last wage earned is used to compute pension rights, while after 2004 the per-period accrual is proportional to the period-specific wage. Since the largest share of pension rights are accrued before 2004, and since wages – only observed from 2005 – follow a very persistent process, we only use the last earned wage.

$$Accrued = Years \times 0.0175 \times (W - 15,000)$$

The compensation is equal to 22.5% of the rights accrued under the early retirement scheme, which effectively means that

$$Compensation = 0.225 \times Accrued$$

Given the initial condition as described above, in the the model pension rights then accumulate as follows. The second component is the part accrued when working in period $t$, where now the accrual rate and the state pension offset are as defined by the 2006 reform.

$$PR_{t+1} = PR_t + FTE_t \times 0.0205 \times (W_t - 9,600)$$

Pension benefits equal pension rights plus the state pension times an actuarial adjustment that depends on the first age of claiming times the share that is claimed.

$$b_t = (PR_t + state\ pension) \times Act.Adj._t \times (1 - FTE_t)$$

**Cohort 1953** Pension rights for the 1953 cohort are computed similarly as to the 1950 cohort. There are, however, two main differences. The 1953 cohort enters the model at 56 in year 2009, which means they already faced the new accrual rate and state pension offset for three years. Also, the compensation is lower compared to the 1950 cohort as they had less time to build rights under the early retirement regime.

$$\begin{aligned} PR_1 = {} & Accrued\ until\ 2006 + Accrued\ after\ 2006 \\ & + Compensation \end{aligned}$$

where

$$\begin{aligned} Accrued\ until\ 2006 &= Years\ until\ 2006 \times 0.0175 \times (W - 15,000) \\ Accrued\ after\ 2006 &= Years\ after\ 2006 \times 0.0205 \times (W - 9,600) \end{aligned}$$

and

$$Compensation = 0.225 \times Accrued\ until\ 2006$$

The state pension is again equal to 9,600. Pension rights accumulate as for the 1950 cohort, and benefits are computed in the same way (using the cohort-specific actuarial adjustments).

$$PR_{t+1} = PR_t + FTE_t \times 0.0205 \times (W_t - 9,600)$$
$$b_t = (PR_t + state\ pension) \times Act.Adj._t \times (1 - FTE_t)$$

**Actuarial factors** Table 2.10 reports the actuarial factors we used for the model simulations. The actuarial factors for cohort 1949 apply when claiming early retirement benefit under the FPU regime. We compute the actuarial factors as described in Annex A of the FPU regulation, version January 1, 2014, as published by ABP. That is, we divide the actuarial factors for each age by that corresponding to the pivotal age of 62 years and three months. We report an actuarial factor of one at age 66 meaning that no actuarial factor applies when claiming benefits under the old age pension (as opposed to the early retirement FPU benefits).

The actuarial factors for cohort 1950 and 1953 apply to the old age pension and are reported in the corresponding columns. We compute them using the factors reported in the ABP regulation, version January 1, 2015. We use the factors from the regulation and adjust them to the relevant state pension ages of 65 and 3 months and of 66 and 4 months, respectively.

The last column reports the actuarial factors used in the first counterfactual exercise, reported in Section 2.7.2, when we increase the state pension age. These actuarial factors are computed as they would be for someone born in 1963, that is for people with a state pension age of 67 and 3 months (the last step of the currently planned increase by the Dutch regulation).

| Age/Cohort | 1949 | 1950 | 1953 | C. 1 |
|---|---|---|---|---|
| 56 | 0.25 | / | / | / |
| 57 | 0.28 | / | / | / |
| 58 | 0.32 | / | / | / |
| 59 | 0.37 | / | / | / |
| 60 | 0.44 | / | / | / |
| 61 | 0.54 | 0.74 | 0.73 | 0.71 |
| 62 | 0.69 | 0.79 | 0.78 | 0.76 |
| 63 | 0.94 | 0.85 | 0.84 | 0.81 |
| 64 | 1.44 | 0.89 | 0.88 | 0.86 |
| 65 | 2.94 | 0.95 | 0.94 | 0.91 |
| 66 | 1.00 | 1.00 | 0.98 | 0.96 |
| 67 | / | / | 1.04 | 1.01 |
| 68 | / | / | / | 1.07 |

Table 2.10: Actuarial factors by age and cohort and for the counterfactual exercise.

## 2.9.2 Details on the solution and estimation of the model

### 2.9.2.1 Computational details on the solution of the model

In order to estimate and simulate the model we first need to solve it. Since there is no analytical solution to the maximization problem, we approximate numerically the policy functions for labour supply, consumption, and pension claiming choices conditionally on the information at each age (the state variables, jointly denoted by $X$ in Section 2.5.2). We solve the model using backward recursion, starting from the end of life (age 100). A key feature of our work is that we jointly model the consumption, labour supply and pension claiming decisions over the life-cycle, where the former is a continuous choice while the latter are discrete choices. The numerical solution of problems with simultaneous discrete and continuous choices is considerably harder than that of problems with only continuous or only discrete choices, which explains the scarce literature considering such models. Studies related to ours opted for different approaches, such as French (2005) or Iskhakov and Keane (2021). We follow a procedure similar to that in French (2005).

The main difficulty in solving dynamic problems that combine discrete and continuous choices is that the smoothness and concavity of the value function – which is typical of continuous problems and ensures the existence and uniqueness of a solution that is itself continuous and, if interior, is the root of the optimality condition – does not hold in a problem with a discrete choice variable. The addition of a discrete choice makes the value function piecewise concave, with kinks falling at the points where the agent is indifferent between any two possible alternatives along the discrete choice domain; these then translate into discontinuities in the optimal choice of the continuous variable (consumption or savings).

As discussed in previous work (e.g. Blundell et al., 2016), kinks can be eliminated and the expected continuation value can be 'concavified' by introducing uncertainty in the model. In our model, kinks in the value function occur at the level of assets where the agent is indifferent between the different labour supply options or between claiming or not claiming pension, or at points of indifference with respect to the same decisions but in the future.

At the same time, we also deal with both continuous (savings, wage, pension rights) and discrete state variables (health, retirement status, pension regime). To address this, we discretize the continuous variables over a predetermined grid and solve the model over a finite number of grid points.[55] We then use linear interpolation to approximate the value function at points for which a solution was not computed as we move backward in

[55]We use a grid with 10 points for assets, 5 points for full-time wage, and 5 points for pension rights. We use the method described in Tauchen (1986) to compute the transition matrix for discretized wages.

time.[56]

The recursive formulation of the problem presented in Section 2.5.2 can be rewritten by combining the work and pension claiming decisions in a unique discrete choice $d_t$ with (at most) seven possible options. Note that the options are seven and not eight because full-time work while claiming pension is not an option, and the available choice set $\mathcal{D}$ is a function of past choices which are reflected in the current state variables $X_t$.

$$V_t(X_t) = \max_{c_t \in \mathcal{C}(X_t,d_t), d_t \in \mathcal{D}(X_t)} \{u(c_t, d_t) + p_t \beta \mathbb{E}_t[V_{t+1}(X_{t+1})|X_t, c_t, d_t] + (1 - p_t)B(a_{t+1})\}$$

When we solve the model numerically, in a first step we look for the optimal consumption level conditional on each of the $k$ (available) discrete options $\mathsf{d_k}$

$$V_t(X_t|d_t = \mathsf{d_k}) = \max_{c_t \in \mathcal{C}(X_t,\mathsf{d_k})} \{u(c_t, \mathsf{d_k}) + p_t \beta \mathbb{E}_t[V_{t+1}(X_{t+1})|X_t, c_t, \mathsf{d_k}] + (1 - p_t)B(a_{t+1})\}$$

In particular, if the expected value function is smooth and concave we can simply rely on 'golden section' search (we also code the problem in terms of optimal saving level rather than optimal consumption). We then select the discrete choice associated with the highest value in the second step.

$$V_t(X_t) = \max_{d_t \in \mathcal{D}(X_t)} \{V_t(X_t|d_t = \mathsf{d_1}), V_t(X_t|d_t = \mathsf{d_2}), \dots\}$$

We finally compute the expected value function before moving to period $t - 1$ and verify that it is smooth and concave over assets.

#### 2.9.2.2 Second-step estimation: Wage and health processes

**Wage process** We assume that the logarithm of full-time gross wage evolves according to an AR(1) process. As explained above, this is in line with the rules governing wage determination in the Netherlands in which wages are set at the national level based on wage scales which determine wage levels and increases. Furthermore, we focus on workers who most likely have reached the end of their wage scale (people older than 55) and face little uncertainty.

$$\ln(W_{it}) = (1 - \rho)\mu + \rho \ln(W_{it-1}) + \xi_{it} \tag{2.14}$$

$$\ln(W_{it}) = \alpha + \rho \ln(W_{it-1}) + \xi_{it} \tag{2.15}$$

Since there are no particular differences across birth cohorts with respect to wage setting rules, we estimate the model pooling the different cohorts and therefore use the

[56] We also need to rely on interpolation and extrapolation when simulating behaviour at points for which the optimal choices were not computed. We thus rely on linear interpolation for consumption/saving choices, and on nearest neighbour interpolation for work and claiming choices.

same estimates when solving and simulating the structural model. We also do not make any adjustment to take into account that full-time wages are observed only for working people, because the same wage scale applies to everyone. We therefore estimate (2.14) with ordinary least squares and cluster standard errors at the individual level. Before estimating the model, we deflate monetary amounts using the Consumer Price Index published by Statistics Netherlands taking 2006 as the baseline year.

Results are presented in Table 2.11. The autoregressive parameter $\rho$ is very close to unity, in line with the estimate from de Bresser (2023). We estimate $\mu = \alpha/(1-\rho) = 10.98$, which represents the expected value of the logarithm of wages. The variance of the error term is estimated to be equal to 0.005.

**Health process** Health status only takes two values: good or bad. The probability of being healthy or unhealthy next year depends on age and the health status in the current year.

$$\Pr(health_{it} = bad | health_{it-1}, t) = \frac{\exp[\pi_0 + \pi_1 t + \pi_2 I\{health_{it-1} = bad\}]}{1 + \exp[\pi_0 + \pi_1 t + \pi_2 I\{health_{it-1} = bad\}]} \quad (2.16)$$

We define bad health as being eligible for DI, which implies having at most 65% of the work capacity left in the Dutch DI scheme. As we adopt an institutional definition for health, we argue that our health measure is an objective one. Since we use administrative data, we lack alternative (subjective) measures of the health status. An alternative measure could be based on medical expenditures available from Statistics Netherlands, but this data is only available from 2009 and therefore does not fully span our study period.

Given these considerations, we use information from Statistics Netherlands about DI recipiency. We assume that eligible people would always claim DI, and therefore treat DI recipiency as being independent from people's choices (i.e. being a measure of exogenous health shocks). Since people are insured under DI only until the state pension age, we use observations until age 65 or 66 to estimate the health process. We estimate the parameters in equation (2.16) with maximum likelihood by regressing health on age and the lag of health, and we assume that the error terms have a logistic distribution. Since there are no particular differences across the birth cohorts, we estimate the model pooling the different cohorts and therefore use the same probabilities when solving and simulating the structural model. Results are presented in Table 2.11.

#### 2.9.2.3 MSM estimates distribution

Our discussion of the Method of Simulated Moments (MSM) estimator is based on French and Jones (2011). The objective of MSM estimation is to find the preference vector that yields simulated life-cycle decision profiles that "best match" (as measured by a GMM

| | Wage | | Health |
|---|---|---|---|
| $\alpha$ | 0.30*** | $\pi_0$ | -5.45*** |
| | (0.04) | | (0.16) |
| $\rho$ | 0.97*** | $\pi_1$ | -0.01 |
| | (0.00) | | (0.02) |
| | | $\pi_2$ | 10.36*** |
| | | | (0.27) |
| N | 33,698 | N | 45,531 |

Table 2.11: Estimates of the auxiliary processes.

Note: Standard errors are clustered at the individual level.

criterion function) the profiles from the data. Formally, the estimator is given by

$$\hat{\theta} = \arg\min_{\theta} \varphi(\theta, X)^T \widehat{W} \varphi(\theta, X)$$

where $\theta$ is an $l \times 1$ vector of unknown parameters; $\varphi(\cdot)$ is the $k \times 1$ vector of moment conditions, whose $k$-th entry is given by $m_k^d(X) - m_k^s(\theta)$, where $m_k^d$ is the $k$-th moment from the data and $m_k^s$ the corresponding moment from the model simulation. $\widehat{W}$ is a $k \times k$ weighting matrix. Even though the optimal weighting matrix is asymptotically efficient, it can be severely biased in small samples (see, e.g., Altonji and Segal, 1996). We therefore use a diagonal weighting matrix that contains only inverses of the estimated variances of the data on the diagonal, e.g. the first entry is $\widehat{W}_{1,1} = [\widehat{Var}(m_1^d)]^{-1}$.

Suppose we have a data set of $n$ independent individuals who are each observed for $T$ periods. Under the regularity conditions stated in Pakes and Pollard (1989) and Duffie and Singleton (1993), the MSM estimator $\hat{\theta}$ is both consistent and asymptotically normally distributed:

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(0, V)$$

with the variance–covariance matrix $V$ given by

$$V = \left(1 + \frac{1}{N_{sim}}\right) (D^T W D)^{-1} D^T W S W D (D^T W D)^{-1}$$

where $N_{sim} = 1$ is the number of times we simulate each individual in the estimation procedure.[57] $S$ is the $k \times k$ variance–covariance matrix of the data, and $D$ is the $k \times l$ Jacobian matrix of the moment vector evaluated at the MSM estimate $\hat{\theta}$

$$D = \frac{\partial \varphi(\theta, X)}{\partial \theta^T}\Big|_{\theta=\hat{\theta}}$$

[57]Since the estimator is consistent for a fixed number of simulation (Adda and Cooper, 2003), and because we are not particularly interested in making inference on the model estimates, we only simulate each individual once to save time in the estimation process.

In our baseline setting, when using cohorts 1950 and 1953 for estimation, we have $k = 112$ moments and $l = 6$ parameters. We compute $D$ by numerical approximation and estimate $S$ using clustered bootstrap, to account for the fact that the same people are repeatedly observed over the years. We use 500 bootstrap samples with replacements, as in de Bresser, 2023.

#### 2.9.2.4 Moment construction

As discussed in Section 2.5, we use the Method of Simulated Moments to estimate preferences. In particular, we target the evolution of work, pension claiming and savings decisions at each age when the choices are active in the model. We use administrative data from Statistics Nederland on wealth to construct the savings moments. Wealth is measured at the household level on the 1st of January of every year and we use data from 2006 to 2021. Wealth includes financial assets, bank savings, real estate, debts and mortgages. We consider the sum of all components. We make several adjustments to the raw data. First, as wealth is measured at the household level and we focus on married men, we divide reported wealth by the squared root of two, which is the OECD equivalence scale measure. Second, we deflate wealth using the Consumer Price Index published by Statistics Nederland and express monetary amounts in terms of 2006 Euros. Third, we are concerned that business cycle and financial market fluctuations, which are not captured by our model, could differentially affect savings across cohorts.

We use a regression approach to net out year effects. For this, we use a larger panel including all available cohorts from 1919 to 1956, and regress wealth on a set of dummies for age and calendar year. We then subtract the estimated coefficients for the corresponding years form observed wealth. Finally, we aggregate the data by averaging over age and birth cohorts for our main sample to construct the targeted moments. For this adjustment we use, again, 2006 as base year (which corresponds to age 56 for cohorts 1949 and 1950, but not for cohort 1953).

Figure 2.10 reports average wealth, before and after we net out year fixed-effects, for the sample used in the reduced-form analysis and in the structural model. The left panel shows that the average wealth is substantially affected by macroeconomic trends, with a decrease from 2008 to 2015 and an increase afterwards.[58] In the right panel, the version adjusted for year fixed effects does not show any particular trend over age and also no large differences across cohorts. It is consistent with the fact that neither net income nor consumption change substantially at retirement age in the Netherlands (Knoef et al., 2017; Been and Goudswaard, 2023).

[58] See the evolution of Dutch Gross Domestic Product for a comparison over the same years (link).

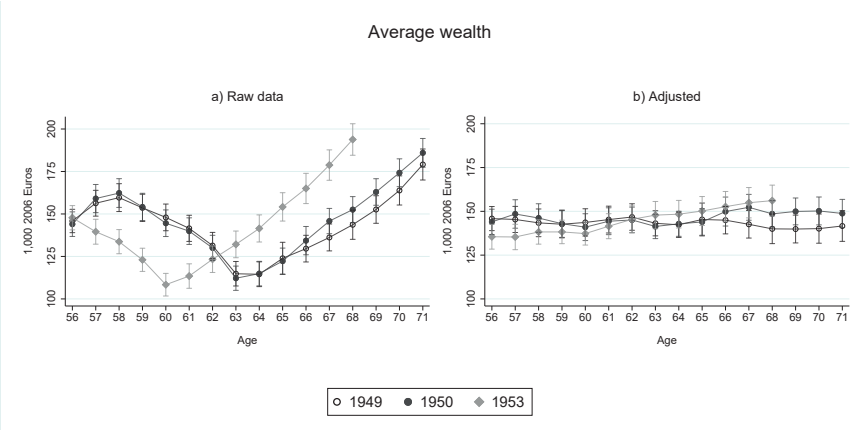



Figure 2.10: Average wealth across cohorts before and after removing year fixed-effects.

## 2.9.3 Additional results

### 2.9.3.1 Trends for women

We limit our analysis to men because the majority of women work part-time throughout their career in the Netherlands, and thus gradually retire rarely. In fact, panel a) in Figure 2.11 shows that around 60% of employed women work part-time as of age 55. Furthermore, this share is constant over age and does not greatly vary across the six treatment groups. These patterns are markedly different from those observed for men (panel d) in Figure 2.2), and they suggest that, as opposed to men, (i) part-time work does not become more popular as women age, and (ii) the reforms didn't largely affected the probability of working part-time at old age.

Panel b) in Figure 2.11 further shows that partial retirement is not so popular among women. For example, at age 63 around 20% of the women who work part-time in Group 1 are also claiming pension benefits at the same time. Panels c) and e) in Figure 2.2 suggests that this is around 80% for men in the same Group at the same age.


Figure 2.11: Part-time and partial retirement among women.

Note: Left: Share of women working part-time among those working. Right: Share of women in partial retirement among those working part-time. Vertical lines refers to the different state pension ages that apply to the different groups.

#### 2.9.3.2 Part-time employment by health status

Figure 2.2 shows that the part-time work patterns differ notably by health status. We define the health status based on DI receipt as discussed above. DI recipients cannot work full-time (eligibility for disability insurance requires having at most 65% of work capacity left), which means that the part-time rate is the same as the employment rate for this group.

Panel a) of Figure 2.12 shows part-time work patterns similar to those presented in Figure 2.2 for the whole sample. The main differences are that (i) the part-time rate is lower among healthy people compared to the whole sample, as expected; (ii) the increase in part-time work after age 60 for groups 2 to 6 is even more pronounced among healthy people, because sick people tend to stop working earlier – as shown by panel b). Panel b) also suggests that people in different groups behave differently only from age 62, when the employment rate becomes lower for group 1, and around the cohort-specific state pension ages, with younger cohorts working slightly longer.

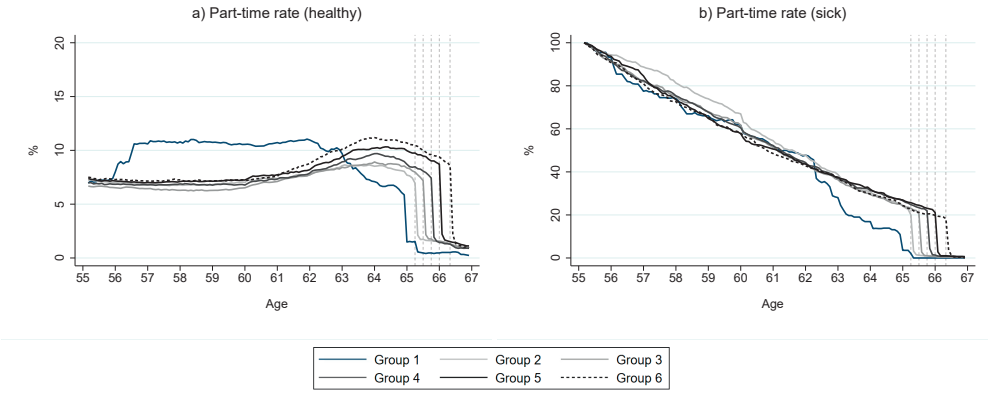



Figure 2.12: Part-time employment by health status over age.

Note: Left: Share of people working part-time among people who do not receive disability insurance benefits. Right: Share of people working part-time among people who receive disability insurance benefits. Vertical lines refers to the different state pension ages that apply to the different groups.

#### 2.9.3.3 Part-time work contracts

Figure 2.13 shows the distribution of the full-time equivalent in the sample for people working part-time (less than 0.875 FTE) on the left, and for people in partial retirement (working part-time and claiming pension rights) on the right. Both figures show bunching corresponding to 0.50 and 0.80 FTE, which are the two most popular levels of part-time work. For computational reasons and to take into account constraints form the demand side of the labour market, we only allow these two levels of part-time work in our model.

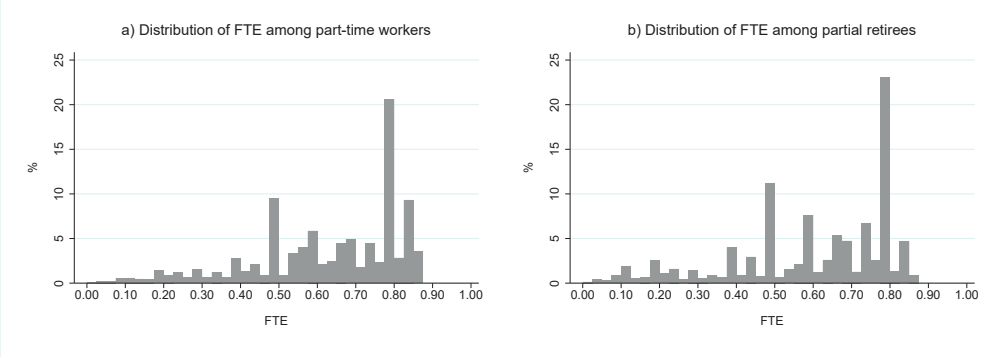



Figure 2.13: Distribution of FTE for part-time workers and partial retirees.

Note: Part-time work is defined as working less than 0.875 FTE (30 hours/week compared to a full-time contract of 40).

#### 2.9.3.4 Regression Discontinuity Approach

Here we present evidence about the causal effects of the two reforms taking a Regression Discontinuity (RD) approach. Figure 2.14 presents RD plots, with the mean value of selected variables plotted against date of birth. In particular, we compute the share of people working, working part-time, and in partial retirement at age 65 (left panels) and at age 66 (right panels). Vertical lines mark the date of birth at which pension rules change, as reported in Table 2.1. We also report linear fits using observations for people subjected to the same pension regime.

The panels in the top row show clear effects of both reforms on the retirement age. The left panel shows how the abolishment of the early retirement scheme had a large effect – about 40 percentage points – on the probability of working at age 65. Similarly, the right panel shows how increasing the state pension age from 65 years and 9 months to 66 years increases the share of people working at 66 by around 40 percentage points.

The panels in the middle row show results for the probability of working part-time. The left panel shows that the abolishment of the early retirement increased the share of people working part-time at 65 (mainly because people postponed retirement). It also suggests that the increase of the state pension age increased the probability of working part-time, as this probability discontinuously jumps at the third and fourth cut-offs. The right panel shows a similar behaviour for the probability of working and working part-time at 66.

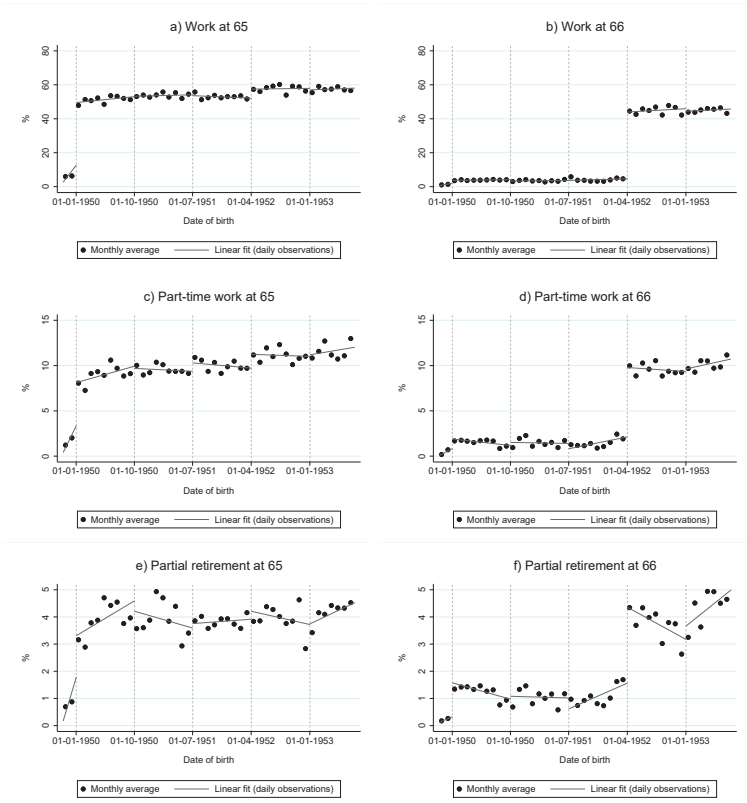



Figure 2.14: Regression Discontinuity plots.

Note: The dots in the figures represent the mean value of the outcome variable computed at age 65 or 66 for people born in the same month. The lines represent the linear fit using for a given pension regime. Vertical lines define the different policy regimes depending on the date of birth.

#### 2.9.3.5 Additional life-cycle profiles

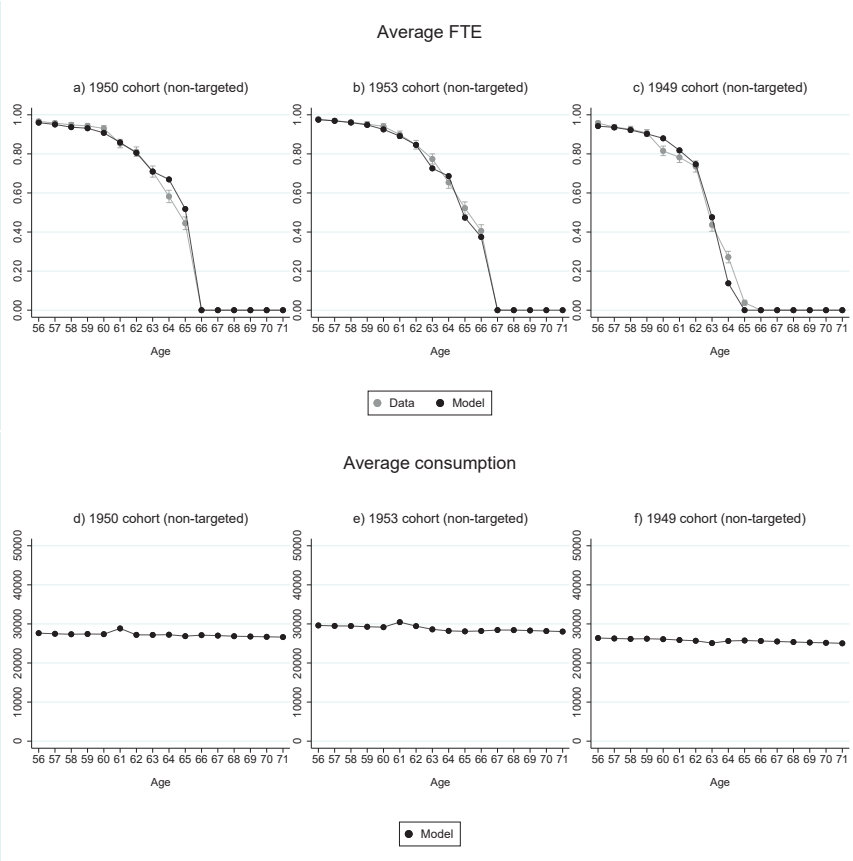



Figure 2.15: Model fit for non-targeted moments.

We do not target average FTE (i.e. the number of hours worked) because this is mainly driven by the full-time work rate. In fact, the data and model profiles for the average FTE closely resembles those of full-time work, shown respectively in Figures 2.15 and 2.5. We also do not target consumption patterns since data is not available. However, since we target savings and model realistically the budget constraint, we expect to be able to replicate consumption behaviour. The lower panels of Figure 2.15 present the model predictions for average consumption showing a smooth age profile which does not exhibit any drop around retirement, in line with existing evidence for the Netherlands (Been and

Goudswaard, 2023).

#### 2.9.3.6 Robustness check: Switching the role of the reforms used for estimation and validation

As explained in Section 2.5.3, in our baseline approach we exploit the 2011 state pension reform to estimate the structural model parameters, and the 2006 early retirement reform to validate our estimates. That is, our Method of Simulated Moments estimates are computed by matching the life-cycle choices of the 1950 and 1953 cohorts, whose pension rules differ because of the 2011 state pension age reform. Instead, for out-of-sample validation, we use the 1949 birth cohort whose regime differs because of the 2006 early retirement reform. The reason behind this choice is that the 2006 reform had larger effects on retirement choices as the changes in pension rules were more drastic due to this reform. Therefore, this reform provides a more demanding test to check whether our model can replicate well non-targeted pension regimes.

We analyse how the model estimates change if we switch the role of the two reforms, that is if we target the 1949 and 1950 cohorts for estimation.[59] Table 2.12 reports the result of this exercise (column 'Robustness'), along with the baseline estimates presented in Table 2.5. The two sets of estimates are very similar and consequently the model fit does not change compared to the baseline (for all three cohorts, not shown). This is not surprising because, as long as the parameters are identified by the two (partially) different set of moments, we would not expect the results to largely differ. The results, therefore, also suggest that the model is not (too) misspecified.

#### 2.9.3.7 Valuation of partial retirement and initial conditions

Table 2.13 presents the estimates from regressing the valuation of partial retirement, as measured by the Equivalent and Compensating variation presented in Section 2.7.1, and the state variables when entering the model. The sample used includes people from the 1950 and 1953 cohorts who retire partially. The estimates suggests that, as expected, partial retirement is more valuable for people with lower savings, who cannot rely on private savings to smooth consumption when reducing the number of hours worked if pension income is not available. The estimates for wage and pension rights, instead, are

[59] The number of moments targeted is now 120, instead of 112. We have 65 moments for cohort 1949, 55 for 1950, and 57 for 1953. That's because (i) savings for cohort 1949 are observed from age 56 to 71, while for cohort 1953 only from 56 to 68 (+3 moments); (ii) cohort 1949 can start claiming five years earlier compared to cohort 1953 (which means five more moments for claiming and also for partial retirement: +10 moments); (iii) cohort 1949 can work one less year compared to cohort 1953 (which means one less moments for full-time work, for part-time work, for work of sick people, for claiming, for partial retirement: -5 moments).

| Parameter | Baseline | Robustness |
|---|---|---|
| $\lambda$ | 0.736 | 0.736 |
| $\gamma$ | 0.870 | 0.877 |
| $\psi$ | 0.040 | 0.040 |
| $\delta$ | 643.974 | 595.008 |
| $b_1$ | 49.617 | 44.540 |
| $b_2$ | 272,556.868 | 246,604.023 |

$u(c_t, l_t) = \frac{1}{\lambda} c_t^{\lambda} + \psi \frac{1}{\gamma} l_t^{\gamma}$

$l_t = (4,000 - h_t - \delta I\{health_t = bad\})/4{,}000$

$B(a_{t+1}) = b_1 \frac{1}{\lambda} (b_2 + a_{t+1})^{\lambda}$

Table 2.12: MSM estimates.

Note: 'Baseline' estimates are MSM estimates obtained targeting cohorts 1950 and 1953, i.e. the 2011 reform; 'Robustness' estimates are MSM estimates obtained targeting cohorts 1949 and 1950, i.e. the 2006 reform.

not significant and very imprecise.

| | (1) | (2) |
|---|---|---|
| VARIABLES | EV | CV |
| Assets (10,000) | -30.32*** | 28.91*** |
| | (11.06) | (10.35) |
| Wage (10,000) | 40.04 | -39.66 |
| | (85.48) | (86.26) |
| Pension rights (10,000) | -96.77 | 87.14 |
| | (96.28) | (106.65) |
| Constant | 960.65 | -902.00 |
| | (420.87) | (371.52) |
| Observations | 134 | 134 |
| R-squared | 0.09 | 0.09 |

Robust standard errors in parentheses

Table 2.13: Valuation of partial retirement and initial conditions.

Note: The table presents the estimates from regressing the valuation of partial retirement, as measured by the Equivalent (EV) and Compensating variation (CV) presented in Section 2.7.1, on the state variables when entering the model. EV only takes positive values while CV only takes negative values, and EV is approximately equal to minus CV.

#### 2.9.3.8 Policy simulation: Relaxing automatic job termination

By comparing employees to self-employed workers, Atav et al. (2023) find that automatic job termination, rather than financial incentives and social norms, is the main driver of the observed bunching of retirement at the state pension age in the Netherlands. Job protection is strong for permanent work contracts but it only lasts until the state pension age, when contracts are terminated. A new contract has to be negotiated with the employer if an employee wants to work beyond the state pension age, or a new job has to be found. From a labour demand perspective, the bunching could be explained by the fact that employers are finally able to lay off expensive workers with declining productivity. Wage rigidity and default effect could be alternative explanations. It is therefore interesting to study how many of the people employed right before the state pension age would continue working if their contracts were not automatically terminated.

While our model explicitly incorporates financial incentives, it does not capture social norms – which should have a minor role according to Atav et al., 2023 – and demand side constraints. Therefore, we provide an upper bound of the effect of relaxing the automatic termination policy. In particular, as explained in Section 2.2, working beyond the state pension age is attractive because employers and employees are exempted from social insurance contributions. Moreover, it is financially attractive to postpone claiming (part) of accrued pension rights as they are actuarially increased for delayed claiming. Both of these incentives to work longer are effectively shut down by automatic job termination.

In this simulation exercise, we focus on cohort 1953 and simulate behaviour given its initial conditions and pension rules, but we postpone automatic job termination to age 70. We then compare this simulation with the baseline results and the first counterfactual simulation exercise, where we increase the state pension age by one year (Section 2.7.2). Figure 2.16 presents the results. The figure shows that relaxing automatic job termination ('Counterfactual 3') could substantially increase labour supply at old age. In particular, the employment rate would be higher already in the years preceding the state pension age, because future rewards keep forward looking workers employed. Moreover, the effect would be larger compared to that due to an increase of the state pension age by one year ('Counterfactual 1'). While the figure provides an upper bound for the treatment effect, it suggests that there are potentially large gains from the supply side: Employees would work longer and their welfare would increase due to the additional flexibility and higher income.

Regarding welfare gains, we define, as in Section 2.7.1, the equivalent and compensating variation such that

$$V_{i1}(X_{i1}, JT = 1, EV_i) = V_{i1}(X_{i1}, JT = 0)$$
$$V_{i1}(X_{i1}, JT = 1) = V_{i1}(X_{i1}, JT = 0, CV_i)$$

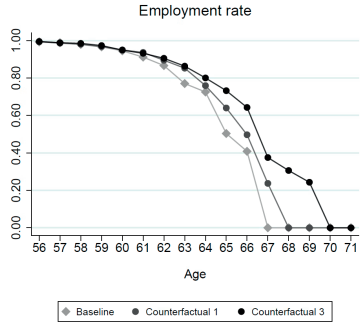



Figure 2.16: Result of the third counterfactual analysis: Relaxing automatic job termination.

Note: The figure compare simulations under the baseline scenario with the 'Counterfactual 1', where we increase the state pension age by one year, and with the 'Counterfactual 3', where we relax automatic job termination at the state pension age, which stays the same as in the baseline.

where $JT$ is a dummy equal to one if jobs are terminated at the state pension age and zero otherwise. We find that, on average, the $EV$ and $CV$ are close to 40,000 euros. The amounts refer to the average worker – not just those that would work anyway until the state pension age – and are fairly substantial for two reasons. First, abolishing automatic job termination gives additional flexibility to workers. The choice to work longer can help to insure against an unexpected drop in earnings, for example. Second, life-time earnings increase because labour supply is much higher under the counterfactual. The valuation of 40,000 euros is then comparable to the additional income due to an extra year of work compared with respect to the baseline (the average wage when entering the model is around 55,000 euros).

# References

Adda, J. and Cooper, R. (2003). *Dynamic Economics: Quantitative Methods and Applications*. MIT Press, Cambridge, US.

Altonji, J. G. and Segal, L. M. (1996). Small-Sample Bias in GMM Estimation of Covariance Structures. *Journal of Business & Economic Statistics*, 14(3):353–366.

Ameriks, J., Briggs, J., Caplin, A., Lee, M., Shapiro, M. D., and Tonetti, C. (2020). Older Americans Would Work Longer If Jobs Were Flexible. *American Economic Journal: Macroeconomics*, 12(1):174–209.

Atav, T., Rabaté, S., and Jongen, E. (2023). Increasing the Retirement Age: Policy Effects and Underlying Mechanisms. Forthcoming. *American Economic Journal: Economic Policy*.

Attanasio, O. P., Meghir, C., and Santiago, A. (2011). Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA. *The Review of Economic Studies*, 79(1):37–66.

Been, J. and Goudswaard, K. (2023). Intertemporal and Intratemporal Consumption Smoothing at Retirement: Micro Evidence from Detailed Spending and Time Use Data. *Journal of Pension Economics and Finance*, 22(1):1–22.

Berg, P., Hamman, M. K., Piszczek, M., and Ruhm, C. J. (2020). Can Policy Facilitate Partial Retirement? Evidence from a Natural Experiment in Germany. *ILR Review*, 73(5):1226–1251.

Blundell, R., Costa Dias, M., Meghir, C., and Shaw, J. (2016). Female Labor Supply, Human Capital, and Welfare Reform. *Econometrica*, 84(5):1705–1753.

Börsch-Supan, A., Bucher-Koenen, T., Kutlu-Koc, V., and Goll, N. (2018). Dangerous Flexibility - Retirement Reforms Reconsidered. *Economic Policy*, 33(94):315–355.

de Bresser, J. (2023). Evaluating the Accuracy of Counterfactuals Heterogeneous Survival Expectations in a Life Cycle Model. Forthcoming. *The Review of Economic Studies*.

De Nardi, M. (2004). Wealth Inequality and Intergenerational Links. *The Review of Economic Studies*, 71(3):743–768.

De Nardi, M., Fella, G., and Paz-Pardo, G. (2024). Wage Risk and Government and Spousal Insurance. Forthcoming. *The Review of Economic Studies*.

De Nardi, M., French, E., and Jones, J. B. (2016). Medicaid Insurance in Old Age. *American Economic Review*, 106(11):3480–3520.

Duffie, D. and Singleton, K. J. (1993). Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica*, 61(4):929–952.

Elsayed, A., de Grip, A., Fouarge, D., and Montizaan, R. (2018). Gradual Retirement, Financial Incentives, and Labour Supply of Older Workers: Evidence from a Stated Preference Analysis. *Journal of Economic Behavior & Organization*, 150:277–294.

Eurofound (2016). *Extending Working Lives Through Flexible Retirement Schemes – Partial Retirement*. Publications Office of the European Union.

French, E. (2005). The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour. *The Review of Economic Studies*, 72(2):395–427.

French, E. and Jones, J. B. (2011). The Eeffects of Health Insurance and Self-Insurance on Retirement Behavior. *Econometrica*, 79(3):693–732.

Goffe, W. L., Ferrier, G. D., and Rogers, J. (1994). Global Optimization of Statistical Functions with Simulated Annealing. *Journal of Econometrics*, 60(1):65–99.

Gustman, A. L. and Steinmeier, T. L. (1986). A Structural Retirement Model. *Econometrica*, 54(3):555–584.

Gustman, A. L. and Steinmeier, T. L. (2005). The Social Security Early Entitlement Age in a Structural Model of Retirement and Wealth. *Journal of Public Economics*, 89(2):441–463.

Heyma, A. (2004). A Structural Dynamic Analysis of Retirement Behaviour in the Netherlands. *Journal of Applied Econometrics*, 19(6):739–759.

Hudomiet, P., Hurd, M. D., Parker, A. M., and Rohwedder, S. (2021). The Effects of Job Characteristics on Retirement. *Journal of Pension Economics and Finance*, 20(3):357373.

Hutchens, R. (2010). Worker Characteristics, Job Characteristics, and Opportunities for Phased Retirement. *Labour Economics*, 17(6):1010–1021.

Iskhakov, F. and Keane, M. (2021). Effects of Taxes and Safety Net Pensions on Life-Cycle Labor Supply, Savings and Human Capital: The Case of Australia. *Journal of Econometrics*, 223(2):401–432. Annals issue: Implementation of Structural Dynamic Models.

Kaboski, J. P. and Townsend, R. M. (2011). A Structural Evaluation of a Large-Scale Quasi-Experimental Microfinance Initiative. *Econometrica*, 79(5):1357–1406.

Kantarcı T. and van Soest, A. (2013). Full or Partial Retirement? Effects of the Pension Incentives and Increasing Retirement Age in the United States and the Netherlands. *Netspar Discussion Paper No. 10/2013-038*.

Kantarcı T., van Soest, A., van Vuuren, D., and Been, J. (2023). Partial Retirement Opportunities and the Labor Supply of Older Individuals. *Netspar Discussion Paper No. 08/2023-039*.

Keane, M. P. and Wasi, N. (2016). Labour Supply: The Roles of Human Capital and The Extensive Margin. *The Economic Journal*, 126(592):578–617.

Knoef, M., Been, J., Caminada, K., Goudswaard, K., and Rhuggenaath, J. (2017). De Ttoereikendheid van Pensioenopbouw na de Crisis en Pensioenhervormingen. *Netspar Design Paper*, 68.

Lalive, R., Magesan, A., and Staubli, S. (2023). How Social Security Reform Affects Retirement and Pension Claiming. *American Economic Journal: Economic Policy*, 15(3):115–50.

Li, Y., Mastrogiacomo, M., Hochguertel, S., and Bloemen, H. (2016). The Role of Wealth in the Start-Up Decision of New Self-Employed: Evidence from a Pension Policy Reform. *Labour Economics*, 41:280–290. SOLE/EALE conference issue 2015.

Lindeboom, M. and Montizaan, R. (2020). Disentangling Retirement and Savings Responses. *Journal of Public Economics*, 192:104297.

Maestas, N., Mullen, K. J., Powell, D., von Wachter, T., and Wenger, J. B. (2023). The Value of Working Conditions in the United States and the Implications for the Structure of Wages. *American Economic Review*, 113(7):2007–47.

OECD (2004). Taxing Wages 2004. Technical report.

OECD (2021). *Pensions at a Glance 2021*.

Pakes, A. and Pollard, D. (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, 57(5):1027–1057.

Rogerson, R. and Wallenius, J. (2013). Nonconvexities, Retirement, and the Elasticity of Labor Supply. *American Economic Review*, 103(4):1445–62.

Rowan, T. (1990). Functional Stability Analysis of Numerical Algorithms. Ph.D. thesis, Department of Computer Sciences, University of Texas at Austin.

Russo, G. and Hassink, W. (2008). The Part-Time Wage Gap: a Career Perspective. *De Economist*, 156(2):145–174.

Rust, J. and Phelan, C. (1997). How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets. *Econometrica*, 65(4):781–832.

Tauchen, G. (1986). Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions. *Economics Letters*, 20(2):177–181.

Theloudis, A. (2018). Wages and Family Time Allocation. Working Paper 2018-06, LISER, Luxembourg.

Todd, P. E. and Wolpin, K. I. (2006). Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility. *American Economic Review*, 96(5):1384–1417.

van der Klaauw, W. and Wolpin, K. I. (2008). Social Security and the Retirement and Savings Behavior of Low-Income Households. *Journal of Econometrics*, 145(1):21–42.

van Soest, A. and Vonkova, H. (2014). How Sensitive are Retirement Decisions to Financial Incentives? A Stated Preference Analysis. *Journal of Applied Econometrics*, 29(2):246–264.

Voena, A. (2015). Yours, Mine, and Ours: Do Divorce Laws Affect the Intertemporal Behavior of Married Couples? *American Economic Review*, 105(8):2295–2332.

# Chapter 3: Present-biased preferences, retirement planning and demand for commitments

**Abstract**

Present-biased preferences induce time-inconsistent consumption and retirement choices, creating scope for commitment products to increase people's well-being. Illiquid assets could be used to commit to a consumption plan by constraining the future budget set, but the liquidity of savings can also affect retirement decisions. For partially naïve agents, who ignore their bias with respect to the retirement decision, the commitment can be costly while futile, i.e. it can negatively affect the retirement choice without improving the inter-temporal consumption allocation. Sophisticated agents, instead, can have high or low willingness to pay for the commitment device depending on whether it helps improving the consumption and retirement decisions at the same time or not.[60]

## 3.1 Introduction

Income adequacy in old age is crucial to ensure retirees' well-being. Yet, many people think that they save less for retirement than they should (Bernheim, 1995). Even in countries with universal public pensions and (almost) universal occupational pensions, such as the Netherlands, a fifth of the people cannot afford their minimal self-reported retirement expenditures (de Bresser and Knoef, 2015). Quasi-hyperbolic time discounting, or present-biased preferences, can explain the mismatch between desired and realized

[60]I thank the participants of the Netspar Pension Day 2020, the Netspar International Pension Workshop 2021, the Microlab at UAB, the seminar of the Graduate Students' Society of Tilburg. In particular, Jaap Abbring, Margherita Borella, Jochem de Bresser, Eric French, Jorgo Goossens, Max Groneck, Fabien Ize, Gabriella Massenz and Eduard Suari-Andreu provided helpful comments and constructive remarks on an earlier version of this chapter.

savings. Laibson (1997) shows how present-bias induces dynamically inconsistent preferences, implying a motive for people to constrain their future choice set. In particular, in this setting, illiquid assets can serve as a commitment technology. In this paper, we show that in a more realistic model where people take two interdependent decisions at the same time, and in particular a continuous and a discrete one, that is consumption and retirement, illiquid assets do not necessarily allow effective commitment and can lead to unintended consequences.

We extend Diamond and Kőszegi (2003)'s stylized model where present-biased agents decide how much to consume and when to retire by introducing a commitment device. Without commitment, naïve agents exhibit time-inconsistent choices, which can result in retiring earlier than originally planned. Sophisticated agents, instead, adjust savings to influence their future retirement date. This can result in different consumption paths for naïve and sophisticated agents, differently from what Phelps and Pollak (1968) find with exogenous retirement date (and logarithmic utility). Sophisticated agents can also commit to the optimal consumption plan using an illiquid asset as a commitment device when the retirement date is set exogenously. However, when retirement is endogenous, illiquid assets can have both a positive and a negative spillover effect on the retirement decision, resulting in different willingness to pay for the commitment. Partially naïve agents, who ignore their bias with respect to the retirement decision, could even demand too much of the illiquid asset to improve savings, which ultimately results in postponing the retirement date while not constraining consumption. This makes the commitment futile while costly, and thus partially naïve agents can be worse off compared to naïve ones, similarly to the results of Heidhues and Kőszegi (2009).

First, we contribute to the literature that studies savings and retirement choices of present-biased agents. Laibson (1997)'s seminal paper introduces present-biased preferences in a life-cycle model and emphasizes the use of illiquid assets as a commitment technology. In his model, illiquid assets can be used to constrain agents' future consumption and savings and thus to limit time-inconsistent choices. This model assumes that labour is exogenously set. Diamond and Kőszegi (2003) propose a stylized three-period model for the analysis of the effect of endogenous retirement decisions on savings behaviour in a quasi-hyperbolic discounting context. In particular, they show how present-biased preferences can lead naïve agents to retire earlier than planned. Merkle et al. (2024)'s online experiment confirms this theoretical prediction by showing that time-inconsistent participants decrease their planned retirement age as they grow older. Zhang (2013) shows that quasi-hyperbolic discounting can lead to a coexistence of under-saving and early retirement. This holds for naïve as well as for sophisticated agents. Findley and Caliendo (2015), in contrast, show that if financial planning is enriched to include the choice of when to retire, then naïve quasi-hyperbolic discounters may borrow far less and

start saving for retirement significantly earlier than exponential discounters. No previous work investigated the demand for commitments in such a theoretical framework.

Second, we contribute to the literature that studies the demand for and the effectiveness of commitment devices. In particular, the evidence on self-commitment is spread over different fields and it is mixed. For example, Giné et al. (2010) find that only a minority of smokers are willing to take up the voluntary commitment product the authors designed to help them quit smoking. Ashraf et al. (2006) and Beshears et al. (2015) provide evidence from field experiments that people use illiquid saving accounts as a commitment device to increase savings. Kaur et al. (2010) find evidence of self-control problems and commitment effectiveness in the workplace context. Carrera et al. (2019) suggest that 'perhaps the puzzle is why we see so much take-up in our experiments' compared to modest real-world evidence. They show that, with uncertainty, time-inconsistent preferences generate demand for commitment only in special cases. It is also not clear whether people are willing to pay for commitment devices. Augenblick et al. (2015) find that commitment is popular in their student sample at a zero price but not at a strictly positive price. Laibson (2015)'s theoretical model originates from the observation that few people express a willingness to pay a significant price to have their choice-set reduced. In his model, a small price of commitment can tip the scales against commitment. On the other hand, Schilbach (2019) provides evidence from a field experiment that participants exhibited significant demand for commitment to sobriety, even at the cost of giving up considerable payments. The author also finds evidence that the reduction of day-drinking increased participants' savings. This suggests that with one commitment device the participants of the experiment were able to achieve two interdependent goals and thus the marginal utility deriving from committing was particularly high, implying the willingness to pay a positive price.

Motivated by these mixed findings on the willingness to use commitment devices and to pay for them, we suggest that in order to better understand the empirical evidence on the demand for commitments, we should consider that individuals often take decisions that are interdependent. If one commitment device is able to address multiple actions in a favourable way at the same time, it is more appealing. On the other hand, if the commitment pushes one decision in an unintended direction while trying to address another decision, it is less appealing.

The remainder of the paper is organized as follows. In sections 2, we first solve the model for the cases of exogenous retirement and review the results of Phelps and Pollak (1968) in our stylized model (Results 1 and 2). We then solve the model with exogenous retirement to review the findings of Diamond and Kőszegi (2003) (Results 3 and 4) and present a new one (Result 5). In section 3, we introduce an illiquid asset. We first review the finding of Laibson (1997) for the case of exogenous retirement (Result 6) and then

present new ones (Results 7, 8 and 9). Section 4 discusses the implications of uncertainty for our results. Section 5 concludes.

## 3.2 Model

### 3.2.1 Model setup

Our model builds on the that of Diamond and Kőszegi (2003). The inter-temporal choices are affected by a quasi-hyperbolic discounting function characterized by two parameters $(\beta, \delta)$. We assume that $\beta \in (0, 1)$ and $\delta = 1$, as in Laibson (2015) and Fahn and Seibel (2022). While results do not qualitatively differ for $\delta \in (0, 1)$, assuming $\delta = 1$ simplifies the exposition. The agent is then characterized by a constant relative risk aversion CRRA) instantaneous utility function. We further assume unitary relative risk aversion, i.e. $u(c) = \ln(c)$.[61] The theoretical implication of the first assumption (i.e. CRRA) is that the consumer's decisions are unaffected by scale. That is, the fraction of wealth optimally consumed in one period is independent of the level of initial wealth. Consider a two-period allocation problem given wealth $W$ (and no return on savings) as in (3.1):

$$\max_{c_1, c_2} u(c_1) + \beta u(c_2) \text{ s.t. } c_1 + c_2 = W \tag{3.1}$$

which yields optimal consumption

$$c_1^* = \frac{1}{1+\beta} W = \lambda W, \tag{3.2}$$

$$c_2^* = \frac{\beta}{1+\beta} W = (1-\lambda) W. \tag{3.3}$$

Optimal consumption in the first period (3.2) is a fixed share of wealth equal to $\lambda$, which only depends on $\beta$, and is larger than optimal consumption in the second period (3.3) since $\lambda > 1 - \lambda$. The logarithmic utility assumption is made to simplify the exposition

[61] We use a logarithmic utility because it leads to two simplifications in the exposition. First, the optimal consumption and saving levels of Self 1 for period 1 are the same regardless of the consumption allocation between periods 2 and 3, that is regardless of whether the consumption in periods 2 and 3 is optimal from the perspective of Self 1. This makes it more intuitive to study the implications of illiquid assets, as their introduction only affects future consumption in periods 2 and 3. Second, the retirement decision does not depend on the consumption allocation between periods 2 and 3 (or on who is deciding the consumption allocation), allowing to separate the two dimensions of the problem in the baseline case without illiquid assets. However, the main ingredients of the paper hold for a more general CRRA utility specification – i.e. the results in Laibson (1997) regarding the value of illiquid assets (with exogenous retirement) and in Diamond and Kőszegi (2003) about endogenous retirement (without illiquid assets). We also focus on a logarithmic utility to show that the result in Phelps and Pollak (1968), which is specific of logarithmic utility, doesn't hold with exogenous retirement.

of the model. The same is true for assuming a quasi-hyperbolic discounting function, instead of a psychologically more accurate hyperbolic discounting function, and for the assumption of no return on savings.

In the remainder of the paper, we focus on a three-period model. In period 1, the agent works and earns a salary $w_1 > 0$, decides consumption $c_1$ and saves $s_1 = w_1 - c_1$. In period 2, he can decide to work or to retire. If he works, he earns a salary $w_2 > 0$ and faces an additive disutility due to his effort $e > 0$. If he doesn't work, he has at his disposal only the savings $s_1$. He consumes $c_2$ and decides how much to leave for the next period. In the third period, he cannot work and consumes what is left. Similarly to Laibson (1997), we assume that $c_2 > w_2$ on the equilibrium path.[62] There is no uncertainty.

With time consistent preferences (such as exponential discounting) the consumer would choose an optimal consumption path $\{c_1^*, c_2^*, c_3^*\}$ in period $t = 1$ and would later stick to that path because in every period he would find it optimal. Behaviour is different with time-inconsistent preferences. If the consumer is naïve, implying he is not aware of his time-inconsistent preferences, he would choose an optimal consumption path in the initial period but will deviate from it in subsequent periods. If the consumer is sophisticated, i.e. he is aware of his time-inconsistent preferences, he anticipates his future inconsistency and tries to affect his future behaviour. In particular, if commitment devices are available, he may use them to constrain his future behaviour.

When making welfare considerations we follow the literature (see DellaVigna and Malmendier, 2004; Heidhues and Kőszegi, 2009; Fahn and Seibel, 2022). The literature evaluates welfare by assigning weights $1, \delta, \delta^2, \dots$ to periods $1, 2, 3, \dots$, such that the ratio of the weights of two subsequent periods is always $1/\delta$. This is equivalent to using the weights from period-1 preferences $(1, \beta\delta, \beta\delta^2, \dots)$ but with $\beta = 1$, i.e. treating the bias a mistake. This is also equivalent – in terms of relative wrights – to using the weights from period-0 preferences $(\beta\delta, \beta\delta^2, \beta\delta^3, \dots)$, or the 'long-run' point of view.

### 3.2.2 Consumption decision

The consumer's choice can be modelled as an equilibrium in a sequential game played by different selves (self 1 in period 1, self 2 in period 2, etc.). The game can then be

[62] As will become clear later, relaxing this assumption implies that the optimal liquid saving level for self 1 might be negative, i.e. self 1 leaves a debt that has to be paid by self 2. Laibson (1997) explains that a negative level of savings "would be interpreted as a contract with an outside agent requiring the consumer to transfer funds to the outside agent, which the outside agent would then deposit in an illiquid account of the consumer". He also notes that these contracts could be renegotiated by future selves, shifting the burden of repaying the debt from self 2 to self 3, for example. If such renegotiation is not allowed, then the assumption is redundant and all results would still hold as they are. In practice, we either need to assume that $c_2 > w_2$ or that renegotiation is not allowed.

solved using backward induction starting from the last period. In the model, the agent understands perfectly the consequences of his actions, and acts optimally within the constraints imposed by his discount function. We start by ignoring the retirement decision to demonstrate time inconsistency regarding the consumption decision. The consumption decision (with exogenous retirement date) has already been studied extensively and the current study does not contribute to this literature, apart from studying it in a stylized model. As in Phelps and Pollak (1968), with logarithmic utility the consumption path of a naïve agent is identical to that of a sophisticated agent if retirement is set exogenously. However, in the next section we show that this property does not hold when the retirement date is endogenous.

Consider the consumption decision in a stylized version of the model of Laibson (1997). The agent only works in period 1, earns wage $w_1$ and the only choice he can make concerns consumption. For the rest of the paper, we adopt the following notation: We use subscripts to indicate periods, e.g. $c_2$ is consumption in period 2, and we use asterisks to indicate the self who is deciding, e.g. $c_2^{**}$ is the optimal consumption according to self 2's preferences, while $c_2^*$ is the optimal consumption from the perspective of self 1.

#### 3.2.2.1 Naïve Agent

Consider a naïve agent who in period $t = 1$ has to decide on his optimal consumption path by solving:

$$\max_{c_1,c_2,c_3} u(c_1) + \beta u(c_2) + \beta u(c_3) \text{ s.t. } c_1 + c_2 + c_3 = w_1$$

He knows that in the last period he would just consume what is left. In the second period he would consume a fixed share of his wealth, as we showed earlier, and leave the rest for the third period. Since he is naïve, he thinks that in period 2 he will have the same time preference as he has now (in period 1). Therefore he concludes that in period 2 he would decide based on the following maximization problem, where $s_1$ are the savings from period 1: $s_1 = w_1 - c_1$:

$$\max_{c_2,c_3} \beta u(c_2) + \beta u(c_3) = \max_{c_2,c_3} u(c_2) + u(c_3) \text{ s.t. } c_2 + c_3 = s_1 \tag{3.4}$$

which yields the following optimal quantities, where we define $\lambda_1 = \frac{1}{2}$:

$$c_2^* = \frac{1}{2}s_1 = \lambda_1 s_1,$$
$$c_3^* = \frac{1}{2}s_1 = (1 - \lambda_1)s_1,$$

The choice of $c_1/s_1$ is made in period 1 by solving

$$\max_{s_1} u(w_1 - s_1) + \beta u\left(\frac{1}{2}s_1\right) + \beta u\left(\frac{1}{2}s_1\right) \tag{3.5}$$

which yields

$$s_1^* = \frac{2\beta}{1+2\beta} w_1,$$
$$c_1^* = \frac{1}{1+2\beta} w_1.$$

When self 1 is naïve, his optimal consumption plan is $\{c_1^*, c_2^*, c_3^*\}$. However, because of quasi-hyperbolic discounting, he will not stick to this plan. Self 1 will indeed consume $c_1 = c_1^*$ and save what is left, so that self 2 will inherit wealth $s_1 = s_1^*$. However, self 2 will decide how much to consume based on the following problem:

$$\max_{c_2, c_3} u(c_2) + \beta u(c_3) \text{ s.t. } c_2 + c_3 = s_1$$

which yields the following, where we define $\lambda_2 = \frac{1}{1+\beta}$:

$$c_2^{**} = \frac{1}{1+\beta} s_1 = \lambda_2 s_1, \tag{3.6}$$
$$c_3^{**} = \frac{\beta}{1+\beta} s_1 = (1-\lambda_2) s_1, \tag{3.7}$$

with $c_2^* < c_2^{**}$ and $c_3^* > c_3^{**}$ because $\beta \in (0,1)$. That is, naïve self 1 plans to consume a certain quantity in period 2 but ends up consuming more because of his present-bias. The opposite is true for period 3. By definition, this results in time-inconsistency.

**Result 1.** *Quasi-hyperbolic time discounting leads to time-inconsistent consumption for a naïve agent. In particular, $c_2^* < c_2^{**}$ and $c_3^* > c_3^{**}$.*

#### 3.2.2.2 Sophisticated Agent

A sophisticated self 1 knows that self 2 would ultimately consume $\{c_2^{**}, c_3^{**}\}$, defined in (3.6)-(3.7), regardless of his initial planning. His choice is then the solution of the problem in (3.8). With instantaneous logarithmic utility, problem (3.5) and problem (3.8) give the same solution, i.e. naïve self 1 makes the same choice of sophisticated self 1. However, for a sophisticated agent the planned consumption path is equal to the realized one, because already in period 1 he knows that in periods 2 and 3 he will consume the optimal quantity chosen by self 2, which is different from self 1's first best.

$$\max_{s_1} u(w_1 - s_1) + \beta u\left(\frac{1}{1+\beta} s_1\right) + \beta u\left(\frac{\beta}{1+\beta} s_1\right) \tag{3.8}$$

which yields again

$$s_1^* = \frac{2\beta}{1+2\beta} w_1,$$
$$c_1^* = \frac{1}{1+2\beta} w_1.$$

**Result 2.** *With quasi-hyperbolic time discounting, logarithmic utility, and exogenous retirement date, the realized consumption path of a naïve agent is identical to that of a sophisticated agent. That is,* $\{c_1^*, c_2^{**}, c_3^{**}\}$ *is the same for naïve and sophisticated agents.*[63]

This result highlights that it might not be possible to infer sophistication from observed choices when the retirement date is set exogenously. We will show that this is not necessarily the case when the retirement date is endogenous.

This property of logarithmic utilities is useful to simplify the analysis of the model when retirement is made endogenous, and it highlights that the optimal consumption and saving levels of self 1 for period 1 are the same regardless of the consumption allocation between periods 2 and 3. In particular, this follows from the fact that with logarithmic utility the change in life-time utility due to a reallocation of consumption between periods 2 and 3 does not depend on the saving/consumption level in period 1. Consider any two different allocations characterised by $\lambda_i$ and $\lambda_j$. From the perspective of self 1, changing from one allocation to another simply shifts life-time utility upwards or downwards, because $\Delta U$ does not depend on $s_1$:

$$\begin{aligned}\Delta U &= \beta \ln(\lambda_i s_1) - \beta \ln(\lambda_j s_1) + \beta \ln((1-\lambda_i)s_1) - \beta \ln((1-\lambda_j)s_1) \\ &= \beta \ln\left[\frac{\lambda_i(1-\lambda_i)}{\lambda_j(1-\lambda_j)}\right] \end{aligned} \tag{3.9}$$

With $\lambda_i = \lambda_1$ and $\lambda_j = \lambda_2$, $\Delta U$ is positive by construction as $\lambda_1$ is chosen to maximize (3.4). This holds also if we assume that self 2 has to work in period 2.

### 3.2.3 Retirement decision

The retirement decision was incorporated into a model with quasi-hyperbolic discounting by Diamond and Kőszegi (2003). Their focus was primarily on comparative statics in a stylized three-period model and on observational equivalence between quasi-hyperbolic and exponential discounting. They showed that for certain levels of savings, an agent could exhibit time-inconsistency regarding retirement choices. We solve the model and characterize the behaviour and expectations of naïve and sophisticated agents, and derive the conditions under which retirement plans are time-inconsistent. We review the findings of Diamond and Kőszegi (2003) to motivate the introduction of a commitment device in the next section, and also to show that the consumption path of a naïve agent may not be identical to that of a sophisticated agent with logarithmic utilities, differently from what Phelps and Pollak (1968) find with an exogenous retirement date.

[63] As in Phelps and Pollak (1968).

#### 3.2.3.1 Naïve Agent

We now endogenize the retirement decision in period 2.[64] Self 2 inherits $s_1$, which he takes as given. He can decide to work and thus he would earn $w_2 > 0$ and incur an additive constant disutility $e > 0$, or he can retire. Self 2 decides the consumption allocation, thus he consume a fraction of wealth equal to $\lambda_2 = \frac{1}{1+\beta}$ in period 2. If self 2 does not work, the wealth he can spend is just $s_1$ and his utility from periods 2 and 3 is given by:

$$u(\lambda_2 s_1) + \beta u((1-\lambda_2)s_1)$$

If he does work, instead, he gets:

$$u(\lambda_2(s_1 + w_2)) - e + \beta u((1-\lambda_2)(s_1 + w_2))$$

Self 2 then works if[65]

$$u(\lambda_2(s_1 + w_2)) - u(\lambda_2 s_1) + \beta[u((1-\lambda_2)(s_1 + w_2)) - u((1-\lambda_2)s_1)] \geq e \tag{3.10}$$

Similarly to the notation for consumption, we write $l_2^{**} = 1$ if self 2 works in period 2, and $l_2^{**} = 0$ otherwise. Since $u(\cdot)$ is a concave function, condition (3.10) holds for every $s_1$ smaller or equal to a certain threshold. We define this threshold as $\bar{k}_2$, that is the value of $s_1$ for which self 2 is indifferent between working or not (left side of (3.10) equal to the right side). Self 2 works if $s_1 \leq \bar{k}_2$. This is intuitive: if the savings are small, the agent prefers to work when he is old, whereas if he saved enough he prefers to retire earlier.

$$\bar{k}_2 = \frac{w_2}{\exp(\frac{e}{1+\beta}) - 1} \tag{3.11}$$

Equation (3.11) shows how self 2 is more likely to work for large values of wage ($w_2$) and of the discount factor (large $\beta$, that is he cares more about the future). The opposite is true if the disutility from working is high ($e$).

Consider now self 1's point of view regarding the decision in period 2. Similarly to the notation for consumption, we write $l_2^* = 1$ if self 1 prefers work in period 2, and $l_2^* = 0$ otherwise. Since self 1 is naïve, he thinks that self 2 will stick to the allocation that is optimal for self 1 ($\lambda_1 = \frac{1}{2}$) and that self 2 will decide based on the following comparison:

$$\beta[u(\lambda_1(s_1 + w_2)) - u(\lambda_1 s_1)] + \beta[u((1-\lambda_1)(s_1 + w_2)) - u((1-\lambda_1)s_1)] \geq \beta e$$

[64] The model only features a discrete labor supply decision. Having a continuous labour supply decision would make time-inconsistency even more salient. Similarly to the case of consumption, with a continuous labour choice, self 2 would always like to work a bit less than self 1 (as long as self 1 doesn't want to work zero hours or full-time).

[65] We assume that the agent works when indifferent.

or

$$u(\lambda_1(s_1+w_2))-u(\lambda_1 s_1)+[u((1-\lambda_1)(s_1+w_2))-u((1-\lambda_1)s_1)]\geq e \tag{3.12}$$

The difference between (3.10) and (3.12) is just the factor $\beta\in(0,1)$, which generates time-inconsistent preferences, and the different $\lambda$'s are in fact innocuous because they cancel out with logarithmic utility. This implies that with logarithmic utility the retirement decision does not depend on the consumption allocation between periods 2 and 3 (or on who is deciding the consumption allocation). The left-hand side of (3.12) is always greater than the left-hand side of (3.10). Consequently, there is a range of values of $s_1$ for which self 2 would not work, but self 1 thinks he will work. Naïve self 1 thinks that the threshold that self 2 will use to decide is given by:

$$\bar{k}_1=\frac{w_2}{\exp(\frac{e}{2})-1} \tag{3.13}$$

with $\bar{k}_1>\bar{k}_2$. In particular, if $\bar{k}_2<s_1\leq\bar{k}_1$ self 2 will retire, but naïve self 1 thinks that self 2 will work, that is $l_2^*>l_2^{**}$.

#### 3.2.3.2 Possibility of Time-Inconsistent Behaviour

The previous subsection showed how a certain range of savings would lead to unplanned early retirement. In particular, in order to have time inconsistency, savings should be low enough so that self 1 thinks that self 2 will work, but high enough so that self 2 prefers not to work. However, the saving level is not exogenous, but it comes from the choice of self 1. We would need $s_1^*=s_1^w\in(\bar{k}_2,\bar{k}_1]$, where is $s_1^w$ is the optimal saving level chosen by self 1 when working is optimal according to his preferences. In particular,

$$s_1^w=\frac{2\beta w_1-w_2}{1+2\beta} \tag{3.14}$$

With (3.11), (3.13) and (3.14), the condition $s_1^w\in(\bar{k}_2,\bar{k}_1]$ can be written as

$$\frac{w_2}{\exp(\frac{e}{1+\beta})-1}<\frac{2\beta w_1-w_2}{1+2\beta}\leq\frac{w_2}{\exp(\frac{e}{2})-1} \tag{3.15}$$

because both the optimal saving level and the two thresholds depend on $w_2$ and $\beta$, it is not immediately clear whether condition (3.15) generally holds. Rewriting it as follows,

$$\frac{1}{2\beta}\left[\frac{1+2\beta}{\exp(\frac{e}{1+\beta})-1}+1\right]<\frac{w_1}{w_2}\leq\frac{1}{2\beta}\left[\frac{1+2\beta}{\exp(\frac{e}{2})-1}+1\right] \tag{3.16}$$

shows that the left-hand side of (3.16) is indeed lower than its right-hand side, and both are positive. Therefore, there exists values $\frac{w_1}{w_2}\in\mathbb{R}^+$ that satisfy the condition.

**Result 3.** *If condition (3.16) holds, quasi-hyperbolic time discounting leads a naïve agent to retire earlier than planned, that is* $l_2^*>l_2^{**}$

#### 3.2.3.3 Sophisticated Agent

So far, we showed how quasi-hyperbolic discounting induces time-inconsistent behaviour regarding both consumption and retirement of a naïve agent. But quasi-hyperbolic discounting also affects self 1's behaviour if he is sophisticated because there is a conflict between self 1's optimal choice and self 2's optimal choice, and self 1 might try to affect the decision of self 2.

A sophisticated self 1 knows that self 2 chooses the consumption allocation between period 2 and 3. Therefore, self 1 compares two expressions like (3.10) and (3.12) where $\lambda = \lambda_2$ in both cases. However, with logarithmic utility the thresholds are still given by the same $\bar{k}_1$ and $\bar{k}_2$ we presented earlier.

Self 1 is sophisticated and knows that self 2 will work if $s_1 \leq \bar{k}_2$, and will retire otherwise. This means that self 1 can influence the time of retirement by choosing the level of savings that self 2 will inherit. Using backward induction, self 1 knows how self 2 will behave and maximizes his discounted lifetime utility only with respect to $s_1$ (or $c_1$, that is equivalent). We can formalize this as:

$$\max_{s_1} U(s_1) \text{ s.t. budget constraint \& optimal response of self 2}$$

where $U(s_1)$ is the lifetime utility from periods 1, 2 and 3. For $s_1 \leq \bar{k}_2$, $U^w(s_1)$ would be the lifetime utility deriving from working in period 2. Vice-versa, for $s_1 > \bar{k}_2$, $U^r(s_1)$ would correspond to retirement in period 2. Suppose now that self 2 is forced to work in period 2. Then, self 1 would choose the optimal savings level $s_1^w$. If self 2 is forced to retire in period 2, instead, self 1 would choose $s_1^r$. It is intuitive that $s_1^r > s_1^w$, because an agent who is forced to retire earlier saves more than someone who is forced to retire later. In particular, the savings have the following functional form:[66]

$$s_1^w = \frac{2\beta w_1 - w_2}{1 + 2\beta} \tag{3.17}$$

$$s_1^r = \frac{2\beta w_1}{1 + 2\beta} \tag{3.18}$$

Imagine now that self 1 can somehow fully commit to the retirement decision (but not the consumption decision) in period 1, and that self 2 cannot change it. Then self 1 would simply choose to work in period 2 if $s_1^w$ leads to a higher life-time utility than $s_1^r$. In reality, self 2 is not forced to work or to retire. This means that, because of the concavity of $u(\cdot)$, self 1's optimal choice is one of the following: $s_1^r, s_1^w$ or $\bar{k}_2$. This can be seen from Figure 3.1, which shows the lifetime utility of self 1 as a function of $s_1$ for the two cases

[66] Note that these saving levels are the same for a naïve agent because of logarithmic utility, and $s_1^w > 0$ since $c_2 > w_2$ on the equilibrium path.

when self 2 works ($U^w$, red line) and when he retires ($U^r$, blue line). In practice, self 1 would only consider the part of $U^w$ to the left of the threshold $\bar{k}_2$ and the part of $U^r$ to its right.

If $\bar{k}_2 < s_1^w < s_1^r$, then self 1 would have to choose between $\bar{k}_2$ and $s_1^r$ even when $s_1^w$ gives a higher utility. This is the case in the figure, where self 1's first best choice is to save $s_1^w$ and to work, but self 2 would retire for this level of savings. If $\bar{k}_2$ leads to the highest discounted utility, then it means that self 1 is under-saving in order to induce self 2 to work since he cannot reach his first best choice $s_1^w$. On the other hand, if $s_1^r$ gives a higher utility than $\bar{k}_2$, even if lower than $s_1^w$, self 1 is over-saving to compensate for the early retirement decision of self 2.

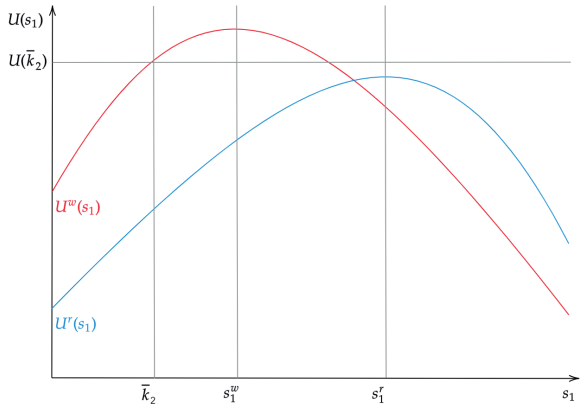



Figure 3.1: Life-time utility of self 1 when self 2 decides on the consumption allocation between periods 2 and 3.

**Result 4.** *If $\bar{k}_2 < s_1^w < s_1^r$ and $U(s_1^r) < U(\bar{k}_2) < U(s_1^w)$, quasi-hyperbolic time discounting leads a sophisticated agent to under-save in order to induce his future self to work. If $\bar{k}_2 < s_1^w < s_1^r$ and $U(\bar{k}_2) < U(s_1^r) < U(s_1^w)$, quasi-hyperbolic time discounting leads a sophisticated agent to over-save in order to compensate for the early retirement decision of his future self.*

The previous result, first shown by Diamond and Kőszegi (2003), lead to a second novel result. If a sophisticated agent with logarithmic utility over or under-save compared to his first best choice, then his consumption path is different from that of a naïve agent.

**Result 5.** *With logarithmic utility and endogenous retirement date, if $\bar{k}_2 < s_1^w < s_1^r$ and $U(s_1^r) < U(s_1^w)$ and $U(\bar{k}_2) < U(s_1^w)$, the consumption path of a naïve agent is different from that of a sophisticated agent.*

## 3.3 Commitment device

### 3.3.1 Consumption decision

As in the previous section, we start discussing the consumption choice assuming that retirement is mandatory in period 2 and subsequently add the retirement decision in the model. Assume that an illiquid asset is made available in the assets market. In this section we assume that the agent is sophisticated since a naïve agent would not realize that it might be useful to constrain his future behaviour. The asset is illiquid in the sense that if bought in period 1 it yields a return in period 3, and not in period 2 as the liquid asset. If the agent decides to sell the illiquid asset in period 2, he would get paid in period 3. Furthermore, if a consumer applies for a loan at time 2, the associated cash flow will not be available for consumption until time period 3.[67] These assumptions are the same as in Laibson (1997). In period 1, self 1 can decide to invest a share $\alpha$ of his savings in the liquid asset and $(1-\alpha)$ in the illiquid one, which also does not yield any return as the liquid one.[68] In period 2, self 2 will have at his disposal only the part of savings that has been invested in the liquid asset: $\alpha s_1$. In period 3, self 3 will consume what is left, that is $\alpha s_1 - c_2 + (1-\alpha)s_1$.[69]

The maximization problem of self 2 is now slightly different because he cannot consume his preferred quantity $c_2$ if this is larger than the liquid wealth he can spend. This is formalized as follow, with a liquidity (LC) on top of the budget constraint (BC):

$$\begin{aligned} &\max_{c_2,c_3} u(c_2) + \beta u(c_3) \\ \text{s.t. } &c_2 \leq \alpha s_1 && \text{(LC)} \\ &c_3 = \alpha s_1 - c_2 + (1-\alpha)s_1 && \text{(BC)} \end{aligned}$$

which yields

$$c_2^{**} = \begin{cases} \frac{1}{1+\beta} s_1 & \text{if } \frac{1}{1+\beta} s_1 \leq \alpha s_1 \\ \alpha s_1 & \text{otherwise} \end{cases}, \tag{3.19}$$

[67] If we assume that consumers can instantaneously borrow against their illiquid asset, then this economy would be equivalent to one in which there is no illiquid asset.

[68] The qualitative results do not depend on the identical returns assumption.

[69] Alternatively, we could think of a commitment device such that the agent at $t = 1$ can decide that a certain share of current and future wage income goes into illiquid assets. If that share cannot be changed in $t = 2$, then this is a stronger commitment technology than the one proposed in the paper. Therefore, it would be better for a sophisticated agent (but possibly worst for a partially naïve one who could commit to an even more excessive extent). However, if we were to change the current model by allowing Self 1 to leave a debt that has to be paid in period 2 - and which cannot be paid out of illiquid assets –, the two devices would be equivalent.

$$c_3^{**} = \begin{cases} \frac{\beta}{1+\beta} s_1 & \text{if } \frac{1}{1+\beta} s_1 \leq \alpha s_1 \\ (1-\alpha) s_1 & \text{otherwise} \end{cases}. \tag{3.20}$$

The solution in (3.19) formalizes the intuition that self 2 can consume his preferred quantity if he has enough liquid wealth to buy it; if the liquid wealth at his disposal is not enough, he would just consume all of it to get as close as possible to his preferred quantity (this directly follows from the concavity of instantaneous utility form consumption).

This also shows how self 1 can constrain the choice of self 2, $\{c_2^{**}, c_3^{**}\}$, through the parameter $\alpha$. We have shown before how self 1 would like to consume $c_2^* = \frac{1}{2} s_1$ in period 2, and his optimal savings $s_1^r$ are given by (3.18). With logarithmic utilities, the choice of $c_1/s_1$ is not affected by the allocation of consumption between periods 2 and 3. This means that if self 1 can influence the consumption choice of self 2, he still chooses the same $c_1/s_1$ compared to when he cannot influence self 2's choice. Therefore, what the illiquid asset does is to shift upward the life-time utility of the agent, because the consumption allocation between periods 2 and 3 is improved and the optimal $c_1/s_1$ does not change. Self 1 sets $\alpha^* = \frac{1}{2} < \frac{1}{1+\beta}$ and $s_1 = s_1^r$. In this way, self 2 will be forced to consume all the liquid asset in period 2 and all the illiquid asset in period 3, where these consumption levels are exactly the first best solution of self 1, $\{c_2^*, c_3^*\}$. In this setting, a sophisticated agent can perfectly commit his future selves by simply using an illiquid asset. The same reasoning holds if we assume that the agent has to work in period 2, with a positive $\alpha$ as we assume that $c_2 > w_2$ on the equilibrium path. The following result simply mimics Laibson (1997)'s finding in a different model.

**Result 6.** *A sophisticated agent can fully commit his future consumption behaviour using an illiquid asset as a commitment device, that is* $c_2^* = c_2^{**}$ *and* $c_3^* = c_3^{**}$.

### 3.3.2 Retirement decision

We now introduce the retirement decision in period 2. Self 2's consumption allocation if he retires is the same as in the case where he is forced to retire and thus the solution is still the one in (3.19). We refer to this allocation as $(c_2^{r**}, c_3^{r**})$. Self 2's allocation problem if he decides to work is given by the following, with the liquidity constraint on consumption in period 2:

$$\begin{aligned} &\max_{c_2, c_3} u(c_2) - e + \beta u(c_3) \\ \text{s.t. } & c_2 \leq \alpha s_1 + w_2 && \text{(LC)} \\ & c_3 = \alpha s_1 + w_2 - c_2 + (1-\alpha) s_1 && \text{(BC)} \end{aligned}$$

which yields

$$c_2^{w**} = \begin{cases} \frac{1}{1+\beta}(s_1 + w_2) & \text{if } \frac{1}{1+\beta}(s_1 + w_2) \leq \alpha s_1 + w_2 \\ \alpha s_1 + w_2 & \text{otherwise} \end{cases}, \tag{3.21}$$

$$c_3^{w**} = \begin{cases} \frac{\beta}{1+\beta}(s_1 + w_2) & \text{if } \frac{1}{1+\beta}(s_1 + w_2) \leq \alpha s_1 + w_2 \\ (1-\alpha)s_1 & \text{otherwise} \end{cases}. \tag{3.22}$$

For given values of $s_1$ and $\alpha$ set by self 1, self 2 will compute $(c_2^{r**}(s_1, \alpha), c_3^{r**}(s_1, \alpha))$ and $(c_2^{w**}(s_1, \alpha), c_3^{w**}(s_1, \alpha))$. Self 2 will then decide whether to work or not by comparing the difference in the utility from the two different consumption plans with the disutility from working. In this framework, the choice of self 1 regarding his savings $(s_1, \alpha)$ will influence both the consumption in periods 2 and 3 and the retirement decision. Self 1's problem is now more complicated: Simply setting $\alpha$ and $s_1$ to reach self 1's first best consumption choice would not necessarily work, because this could change the retirement decision of self 2 in an unintended way.

To gain intuition, we analyse the three possible different cases. As shown earlier, in some cases self 2 would not work while self 1 would like him to work, but never the other way around. Furthermore, self 1 would like to influence the allocation of consumption between periods 2 and 3. In particular, changing the allocation of consumption between periods 2 and 3 shifts the life-time utility of self 1 upward or downward, without changing the saving level $s_1$ that maximizes the life-time utility, as shown in Section 3.2.2. This means that self 1 would ideally stick to his first best savings level and use the illiquid asset to reallocate consumption from period 2 to period 3 and to also influence the retirement decision if necessary.

Case 1: $s_1^w < \bar{k}_2 < \bar{k}_1$ (Self 1 and self 2 agree that it is optimal to work in period 2 given self 1's first best saving $s_1^w$ and no commitment.)
Self 1 does not need to influence self 2's retirement decision. Self 1 sticks to $s_1^w$ and set $\alpha$ to influence the consumption decision of self 2. In particular, self 1 sets $\alpha$ such that the wealth available to self 2 is equal to self 1's optimal consumption choice for period 2, call it $\alpha^w$:

$$\frac{1}{2}(s_1^w + w_2) = \alpha^w s_1^w + w_2 \implies \alpha^w = \frac{\frac{1}{2}(s_1^w + w_2) - w_2}{s_1^w} \in (0,1)^{70} \tag{3.23}$$

Self 2 would still prefer to work given $\alpha^w$ and $s_1^w$ (proof in the Appendix 3.6.1.1). Intuitively, if self 2 has less resources at his disposal, because of the liquidity constraint, he is even more motivated to work and not to retire. In this case, the

[70] As we assume that $c_2 > w_2$ on the equilibrium path.

agent can fully commit. The maximum willingness to pay to have access to the commitment is equal to the extra-utility that the commitment buys, deriving from the reallocation of consumption (call it $\Delta U^*$, see (3.9) for the exact expression).

Case 2: $\bar{k}_2 < \bar{k}_1 < s_1^r$ (Self 1 and self 2 agree that it is optimal to retire in period 2 given self 1's first best saving $s_1^r$ and no commitment.)
In order to reach his first best solution, self 1 should stick to $s_1^r$ and he would use $\alpha$ to reallocate consumption between periods 2 and 3, that is:

$$\frac{1}{2}s_1^r = \alpha^r s_1^r \implies \alpha^r = \frac{1}{2} \in (0,1) \tag{3.24}$$

Given the lower liquid wealth at his disposal, however, self 2 might change his decision and be induced to work (proof in the Appendix 3.6.2.1). In particular, if working allows self 2 to break the liquidity constraint it is more appealing. If $\alpha^r$ together with $s_1^r$ does not induce working, self 1 can reach his first best allocation and retirement decision. The illiquid asset is a perfect commitment device. The maximum willingness to pay to have access to the commitment is the same as in Case 1, as we have shown that the extra-utility deriving from the reallocation of consumption ($\Delta U^*$) does not depend on the saving level, if the saving level is the same with and without commitment. But $\alpha^r$ together with $s_1^r$ can induce working. If that is the case, self 1's optimal choice is given by a combination of $s_1$ and $\alpha$ that can either induce working or not (proof in the Appendix 3.6.2.2). That is, even tough self 1 and self 2 agree to retire without the commitment, if the illiquid asset is available self 1 might decide to induce working. If $\alpha^r$ together with $s_1^r$ induce working, the willingness to pay to have access to the commitment is lower than $\Delta U^*$.[71]

Case 3: $\bar{k}_2 < s_1^w < \bar{k}_1$ (Self 1 and self 2 disagree on the retirement decision given self 1's first best saving $s_1^w$ and no commitment.)
When there is time inconsistency regarding both consumption and retirement, self 1 is in a better position to constrain his future self. In fact, in order to force self 2 to work, self 1 has to set a low $\alpha$. Similarly, in order to favourably affect the consumption allocation, he has to set a low $\alpha$. Self 1 could set $s_1 = s_1^w$, which would actually induce retirement if no commitment is provided. His optimal consumption allocation is induced by the same level $\alpha^w$ as in case 1. If this value is low enough to induce working, then full commitment can be achieved. Regardless of the optimal

[71] The maximum extra utility achievable derives from reaching the first best consumption allocation, since the retirement choice is already optimal without the commitment, and is equal to $\Delta U^*$. Since full commitment cannot be achieved in this case the maximum willingness to pay is lower. See also Appendix 3.6.2.2.

choice of self 1 without commitment, i.e. over or under-save - see Result 4, the maximum willingness to pay is higher than $\Delta U^*$, as the commitment improves the consumption allocation but also the retirement date (see Appendix 3.6.3.1 for a graphical explanation). If $\alpha^w$ is not low enough to induce working, self 1's optimal choice is given by a combination of $s_1$ and $\alpha$ that can either induce working or not (similar to case 2). Also, the maximum willingness to pay to have access to the commitment can be higher or lower than $\Delta U^*$.

The possibility of buying an illiquid asset increases self 1's life-time utility if he is sophisticated in a setting where he has to decide how much to consume and when to retire. However, the illiquid asset may not be sufficient to reach his first best choices regarding the two decisions. If we were to ignore that the retirement date is endogenous, we would expect the agent to be able to reach his first best consumption allocation and to be willing to pay $\Delta U^*$ to access the commitment. Our expectation would be wrong, as self 1 is not necessarily able to reach his first best consumption allocation and the agent would be willing to pay more if he can address both the consumption and the retirement allocation, or less if the commitment pushes the retirement date in an unintended direction.

**Result 7.** *A sophisticated agent may not be able to commit his future consumption and retirement behaviour using an illiquid asset. In particular, if* $s_1^w < \bar{k}_2 < \bar{k}_1$*, then the first-best can be reached:* $c_2^* = c_2^{**}$ *and* $l_2^* = l_2^{**}$*. If* $\bar{k}_2 < \bar{k}_1 < s_1^r$*, instead, that is not always the case:* $c_2^* \leq c_2^{**}$ *and* $l_2^* \leq l_2^{**}$*. Similarly, if* $\bar{k}_2 < s_1^w < \bar{k}_1$*, then* $c_2^* \leq c_2^{**}$ *and* $l_2^* \geq l_2^{**}$*.*

**Result 8.** *The willingness to pay for the commitment depends on whether it helps addressing both decisions at the same time or not. In particular, if* $s_1^w < \bar{k}_2 < \bar{k}_1$*, the (maximum) willingness to pay is equal to* $\Delta U^*$*. If* $\bar{k}_2 < \bar{k}_1 < s_1^r$*, the willingness to pay is lower or equal to* $\Delta U^*$*. If* $\bar{k}_2 < s_1^w < \bar{k}_1$*, it can be higher, equal or lower than* $\Delta U^*$*.*

The use of the illiquid asset alone might thus not be sufficient to reach the first best consumption and work decision (cases 2 and 3 in the chapter). Potentially, self 1 could sign a work contract in period 1 with (i) harsh penalties in case of early resignation from the job, (ii) a fixed end which is difficult to re-negotiate. Both these attributes, however, could be quite undesirable if uncertainty is added to the model. In practice, reaching the first best might thus be difficult even for sophisticated people.

### 3.3.3 Partially naïve agents

So far, we have assumed in the discussion concerning self-commitment devices that the agent is sophisticated. However, it is unclear to what extent people are sophisticated or in which circumstances. It seems natural that people are more prone to recognize their

self-control problems when they concern actions that are repeated frequently over time, such as smoking, alcohol consumption or gym attendance.

In our setting, it seems reasonable to assume that individuals are aware of their present-biased preferences over consumption choices, and thus that they save too little (Bernheim, 1995). Since this is an action that we repeat daily, we tend to be aware of our time preferences over consumption. However, since the retirement decision is typically taken only once, there is much less scope for learning. We refer to partially naïve (or partially sophisticated) agents as those who are aware of their present-biased preferences over consumption but who ignore their bias over the retirement decision, as in Diamond and Kőszegi (2003).

Naïve agents would clearly not exhibit a demand for commitment devices. The result is that, regardless of their retirement decision in period 2, they would end up with a sub-optimal consumption allocation between periods 2 and 3, and possibly also with a sub-optimal retirement decision, from the perspective of self 1.

A partially naïve agent would instead always demand some amount of the illiquid asset to reallocate consumption from period 2 to period 3, regardless of his retirement plan. We consider the three possible cases outlined before. As for sophisticated agents, case 2 is the most interesting.

Case 1: $s_1^w < \bar{k}_2 < \bar{k}_1$ (Self 1 and self 2 agree that it is optimal to work in period 2 given self 1's first best saving $s_1^w$ and no commitment.)
Self 1 wants to address the consumption decision. Self 1 sticks to $s_1^w$ and sets $\alpha$ such that the wealth available to self 2 is equal to self 1's optimal consumption choice for period 2 ($\alpha^w$). Self 1's maximum willingness to pay is $\Delta U^*$. Self 2 would still prefer to work given $\alpha^w$ and $s_1^w$ (same proof as before in the Appendix 3.6.1.1). Self 1 can induce his optimal consumption allocation.

Case 2: $\bar{k}_2 < \bar{k}_1 < s_1^r$ (Self 1 and self 2 agree that it is optimal to retire in period 2 given self 1's first best saving $s_1^r$ and no commitment.)
Self 1 wants to address the consumption decision. Self 1 sticks to $s_1^r$ and and sets $\alpha$ such that the wealth available to self 2 is equal to self 1's optimal consumption choice for period 2 ($\alpha^r$). Self 1's maximum willingness to pay is $\Delta U^*$. Since self 1 still prefers self 2 to work when using the commitment, self 1 believes that self 2 will work (he is naïve about the retirement decision). However, self 2 might work (same proof as before in the Appendix 3.6.2.1). In particular, self 2 finds working more appealing when it allows to break the liquidity constraint. This would make the commitment costly while futile: The welfare in the case of commitment is lower than the welfare without commitment because working is not optimal and the consumption allocation is not improved.

Case 3: $\bar{k}_2 < s_1^w < \bar{k}_1$ (Self 1 and self 2 disagree on the retirement decision given self 1's first best saving $s_1^w$ and no commitment.)
Self 1 thinks that self 2 also prefer to work and thus wants to address the consumption decision. Self 1 sticks to $s_1^w$ and sets $\alpha$ such that the wealth available to self 2 is equal to self 1's optimal consumption choice for period 2 ($\alpha^w$). Self 1's maximum willingness to pay is $\Delta U^*$. If $\alpha^w$ is low enough to induce working, then self 1 can induce his optimal consumption allocation (voluntarily) and his optimal retirement choice (involuntarily). The welfare is larger compared to the no-commitment case. If $\alpha^w$ is not low enough to induce working, it still affects the consumption allocation. The consumption allocation induced by $\alpha^w$ together with $s_1^w$ when self 2 retires is not the optimal one from the perspective of self 1, but it can be better (or worse) compared to the one self 2 would have chosen without the commitment. Thus, the welfare can be larger (or smaller) compared to the no-commitment case.

**Result 9.** *For partially naïve agents, the commitment can be costly while futile. If* $\bar{k}_2 < \bar{k}_1 < s_1^r$ *or* $\bar{k}_2 < s_1^w < \bar{k}_1$*, the long-run utility of a naïve agent can be higher than that of a sophisticated agent.*

## 3.4 The role of uncertainty

When talking about savings and retirement, there are typically three sources of uncertainties that are interesting to consider: Wage, survival, and health (including medical expenses) uncertainty. While also Laibson (1997) and Diamond and Kőszegi (2003) seminal works don't feature any uncertainty, it is worth discussing the expected implications of these types of uncertainty for the presented results.[72]

First, introducing wage uncertainty would decrease the demand for illiquid assets of Self 1. That is because the precautionary motive for holding liquid assets – a risk adverse agent values insurance against a bad wage draw –, which is now absent, would decrease the commitment value of illiquid assets.

Second, introducing survival uncertainty would decrease the weight of future periods in the (expected) value function of Self 1. Because commitment devices increase future, but not current, instantaneous utility, their marginal utility would be lower. Demand would thus decrease if the commitment has a positive price, but not if it has a zero price, like in the model.

Third, introducing uncertain health in the model, with the health status affecting the marginal utility of consumption and/or leisure, would change the optimal consumption

[72] Laibson (2015) consider the role of uncertainty but in a very different model in which a non-divisible task needs to be done and the agent decides when to do the task. In this setting, higher uncertainty of the effort cost leads to lower commitment demand.

and saving levels in ways that depend on the assumed instantaneous utility function. More interestingly, introducing uncertain future medical expenses would decrease the demand for illiquid assets of Self 1, in the same spirit of wage uncertainty. Again, the precautionary motive for holding liquid assets (as insurance against medical expenses) would decrease the commitment value of illiquid assets.

The addition of wage and medical expenses uncertainty would thus decrease the demand for commitment due to the rise of precautionary motives for holding liquid assets, but that would similarly impact sophisticated and partially naïve agents. When the value of committing the consumption/saving decision outweighs the precautionary motive, we would still expect the presented differences between sophisticated and partially naïve agents.

## 3.5 Conclusion

This paper presents a model that is able to generate time-inconsistent consumption and retirement decisions. The model is simple and tractable, yet it provides intuition for the underlying mechanisms. It extends Diamond and Kőszegi (2003)'s model, who show that naïve agents with quasi-hyperbolic discounting functions could exhibit time-inconsistent behaviour regarding both retirement and consumption planning. Besides, sophisticated agents could be induced to over- or under-save, with respect to their first best solution, to affect their future behaviour.

First, we provide evidence that illiquid assets can be used as a commitment device also to affect retirement decisions. This generalizes the intuition derived from Laibson (1997) that present-biased agents can use illiquid assets to affect their savings level, and also extends the model of Diamond and Kőszegi (2003) by introducing the possibility of buying a commitment device.

Second, we show that when two interdependent choices have to be made, such as consumption and retirement, then demand for commitment is non-trivial. In this case, the effectiveness of one commitment device can be high or low depending on whether it can address both actions in the desired way at the same time.

Third, we show that welfare and willingness to pay for commitment are not strictly monotone in the sophistication degree. Partially naïve agents could be willing to pay more to commit than sophisticated agents because they do not realize that the illiquid asset could push their retirement decision in an unintended direction. Moreover, while sophistication is a desirable attribute, partial naïveté may not be desirable compared to naïveté.

# 3.6 Appendix

## 3.6.1 Case 1 in Section 3.3 for sophisticated agents

### 3.6.1.1 The commitment does not change self 2's retirement choice

Self 1 and self 2 agree to work given self 1's first best saving level $s_1^w$ and no commitment. That is, $U(s_1^w) \geq U(s_1^r)$ and $s_1^w \leq \bar{k}_2 < \bar{k}_1$. Self 1 sets $\alpha$ such that

$$\frac{1}{2}(s_1^w + w_2) = \alpha s_1^w + w_2 \tag{3.25}$$

$$\alpha^w = \frac{\frac{1}{2}(s_1^w + w_2) - w_2}{s_1^w} \in (0,1) \tag{3.26}$$

If given $s_1^w$ and $\alpha^w$ self 2 works, $s_1^w$ and $\alpha^w$ leads to the highest level of life-time utility achievable by self 1, because he enforces his first bests consumption allocation and retirement decision. We prove with the following steps that, given $s_1^w$ and $\alpha^w$, self 2 works.

1. By assumption for this case it holds that, given $s_1^w$ and no commitment, self 2 works. As proven earlier, this choice is independent from who decides the consumption allocation between periods 2 and 3 - see section 3.2.3.

$$\ln(\lambda_i(s_1^w + w_2)) - \ln(\lambda_i s_1^w) + \beta[\ln((1-\lambda_i)(s_1^w + w_2)) - \ln((1-\lambda_i)s_1^w)] \geq e \tag{3.27}$$

2. The constraint in (3.21) holds by construction:

$$\frac{1}{1+\beta}(s_1^w + w_2) > \alpha^w s_1^w + w_2 = \frac{1}{2}(s_1^w w_2) \tag{3.28}$$

This means that if self 2 works he decides to consume $\lambda_1(s_1^w + w_2)$ in period 2 and $(1-\lambda_1)(s_1^w + w_2)$ in period 3.

Because the constraint in (3.21) holds, the constraint in (3.19) also holds:

$$\frac{1}{1+\beta}(s_1^w + w_2) > \alpha^w s_1^w + w_2 \tag{3.29}$$

$$\frac{1}{1+\beta}(s_1^w + w_2) - w_2 > \alpha^w s_1^w + w_2 - w_2 \tag{3.30}$$

$$\frac{1}{1+\beta}s_1^w - \frac{\beta}{1+\beta}w_2 > \alpha^w s_1^w \tag{3.31}$$

$$\frac{1}{1+\beta}s_1^w > \alpha^w s_1^w \tag{3.32}$$

This means that if self 2 does not work he decides to consume $\alpha^w s_1^w$ in period 2 and $(1-\alpha^w)s_1^w$ in period 3.

3. From point (2) it follows that, given $s_1^w$ and $\alpha^w$, self 2 works if

$$\ln(\lambda_1(s_1^w + w_2)) - \ln(\alpha^w s_1^w) + \beta[\ln((1-\lambda_1)(s_1^w + w_2)) - \ln((1-\alpha^w)s_1^w)] \geq e \tag{3.33}$$

4. Suppose $\alpha^w = \lambda_1$. Then (3.33) holds because of point (1).

5. Since $s_1^w > 0$ ($c_2 > w_2$ on the equilibrium path), then $\alpha^w < \lambda_1 < \lambda_2$. Since self 2 discounted utility is concave in $\lambda$ and $\lambda_2$ is the maximum, (3.33) holds.

### 3.6.2 Case 2 in Section 3.3 for sophisticated agents

#### 3.6.2.1 $s_1^r$ and $\alpha^r$ can induce working

Self 1 and self 2 agree that it is optimal to retire in period 2 given self 1's first best saving $s_1^r$ and no commitment. In order to reach his first best solution, self 1 should stick to $s_1^r$ and he would use $\alpha$ to reallocate consumption between periods 2 and 3, that is:

$$\frac{1}{2}s_1^r = \alpha^r s_1^r \implies \alpha^r = \frac{1}{2} \in (0,1) \tag{3.34}$$

Given the lower liquid wealth at his disposal, however, self 2 might change his decision and be induced to work. The proof is as follows:

1. Given $s_1^r$ and no commitment, self 2 prefers to retire, which means that

$$u(\lambda_2(s_1^r + w_2)) - u(\lambda_2 s_1^r + \beta[u((1-\lambda_2)(s_1^r + w_2)) - u((1-\lambda_2)s_1^r)] < e \tag{3.35}$$

2. Given $s_1^r$ and $\alpha^r$, self 2 liquidity constraint when retiring holds by construction. However, self 2 liquidity constraint when working does not necessarily hold (we showed that (3.29) implies (3.32), but not other way around). If it does not hold, i.e. working allows to break the liquidity constraint and (3.29) does not hold, the comparison is

$$u(\lambda_2(s_1^r + w_2)) - u(\alpha^r s_1^r) + \beta[u((1-\lambda_2)(s_1^r + w_2)) - u((1-\alpha^r)s_1^r)] \lesseqgtr e \tag{3.36}$$

   The LHS of (3.36) is larger than the LHS of (3.35) because $\alpha^r = \lambda_1 < \lambda_2$. This means that working is more appealing compared to the no-constraint scenario. Depending on the value of $e$, self 2 can prefer to work.

#### 3.6.2.2 Optimal choice

Assume that $s_1^r$ and $\alpha^r$ induce working. We highlight that the optimal choice can be to induce working, despite self 1 and self 2 agree to retire without commitment. Consider the following example where self 1 and self 2 agree to retire without the illiquid asset, that is $U(s_1^w) < U(s_1^r)$ and $\bar{k}_2 < s_1^r$.

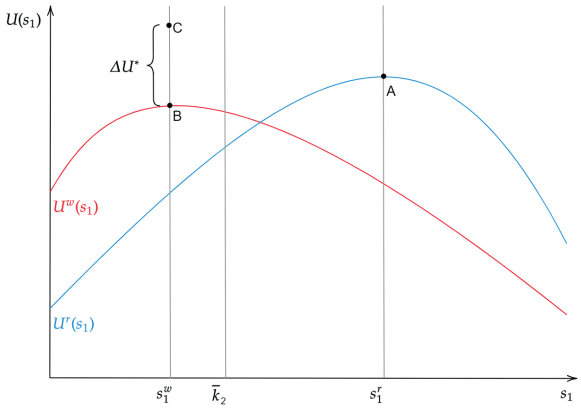



Figure 3.2: Life-time utility of self 1 when self 2 decides on the consumption allocation between periods 2 and 3.

If $\alpha^r = \frac{1}{2}$ together with $s_1^r$ induces working, $\alpha^r$ and $s_1^r$ cannot be the optimal choice of self 1. Consider the choice of $s_1^w$ and $\alpha^w = \frac{\frac{1}{2}(s_1^w + w_2) - w_2}{s_1^w}$. As in Case 1, this choice induces work in period 2 and by definition $U^w(s_1^w) > U^w(s_1^r)$. It also induces self 1's optimal consumption allocation between periods 2 and 3. Therefore $s_1^w$ and $\alpha^w$ lead to the highest utility achievable when the agent works in period 2, and it is strictly preferred to $(\alpha^r, s_1^r)$. This shifts the life-time utility of self 1 upwards by $\Delta U^*$ in correspondence of $s_1^w$ in Figure 3.2, but the gain compared to the no commitment case (point A: $\alpha = 1, s_1^r$) is lower than $\Delta U^*$ since $U(s_1^w) < U(s_1^r)$.

Instead, if self 1 decides to induce retirement, he cannot achieve his first best consumption plan, thus in any case the maximum willingness to pay is lower than $\Delta U^*$. The optimal option is either to stick to $s_1^r$ and chose a lower level of illiquid assets not to induce working and slightly improve the consumption allocation, or he can chose a different saving level $s_1 > s_1^r$ and use a higher level of the illiquid assets.

### 3.6.3 Case 3 in Section 3.3 for sophisticated agents

#### 3.6.3.1 If full commitment is possible, the maximum willingness to pay is larger than $\Delta U^*$

Without commitment, the optimal choice of self 1 is either to under or over-save (points A or B in Figure 3.3, respectively) as shown in Result 4. If full commitment is achievable, self 1 can reach the utility level corresponding to point D. The extra utility deriving from committing is thus always larger than $\Delta U^*$ if $\alpha^w$ is low enough to induce self 2 to work.

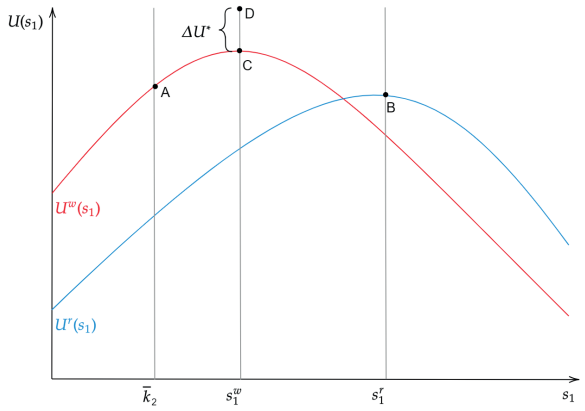



Figure 3.3: Life-time utility of self 1 when self 2 decides on the consumption allocation between periods 2 and 3.

#### 3.6.3.2 If full commitment is not possible, the maximum willingness to pay can be higher or lower than $\Delta U^*$

Consider the case in which, without the commitment, self 1 would under-save and chose a level of savings $\bar{k}_2$, as in Figure 3.4. If full commitment is not possible, the utility level corresponding to point D cannot be reached. Self 1 can stick to $\bar{k}_2$ and use the commitment to reach the optimal consumption allocation between periods 2 and 3. The commitment would not induce self 2 to retire, as in subsection 3.6.1.1. The extra utility from committing would be $\Delta U^*$ (point B). For any other saving level below $\bar{k}_2$, self 1 could use the illiquid asset but would reach a utility level below point B, thus this cannot be optimal. There might exists saving levels between $\bar{k}_2$ and $s_1^w$ which, combined with $\alpha$, induce working. In that case the extra utility from committing is larger than $\Delta U^*$. Thus, the maximum willingness to pay is at least $\Delta U^*$.

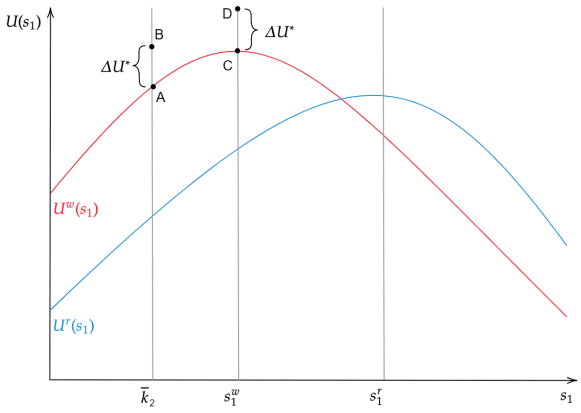



Figure 3.4: Life-time utility of self 1 when self 2 decides on the consumption allocation between periods 2 and 3.

Let's consider the case in which, without the commitment, self 1 would under-save and chose a level of savings $s_1^r$, as in Figure 3.5. In that case, the maximum willingness to pay can be higher or lower than $\Delta U^*$. Again, the utility level corresponding to point D cannot be reached in Figure 3.5. The highest utility level achievable when self 2 retires is given by point B, which can be reached if $\alpha = \frac{1}{2}$ does not induce working and implies an extra utility of $\Delta U^*$ compared to A. If this level of liquid assets induce working, the extra utility achievable when self 2 retires is below $\Delta U^*$. Still, there might exists saving levels between $\bar{k}_2$ and $s_1^w$ which, combined with $\alpha$, induce working and can lead to a higher utility compared to B, and thus to an extra utility larger than $\Delta U^*$ compared to A.

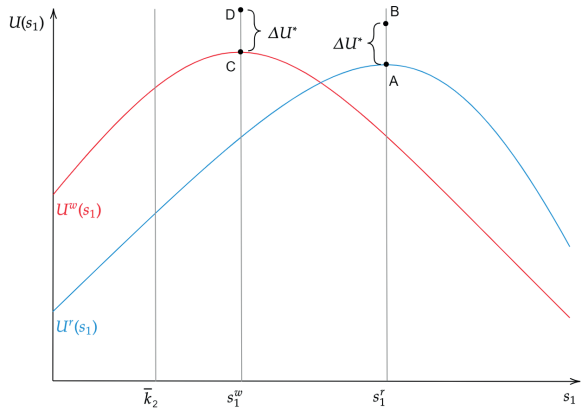



Figure 3.5: Life-time utility of self 1 when self 2 decides on the consumption allocation between periods 2 and 3.

# References

Ashraf, N., Karlan, D., and Yin, W. (2006). Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines. *The Quarterly Journal of Economics*, 121(2):635–672.

Augenblick, N., Niederle, M., and Sprenger, C. (2015). Working over Time: Dynamic Inconsistency in Real Effort Tasks. *The Quarterly Journal of Economics*, 130(3):1067–1115.

Bernheim, B. D. (1995). Do Households Appreciate Their Financial Vulnerabilities? An Analysis of Actions, Perceptions, and Public Policy. Tax Policy and Economic Growth, Washington, DC: American Council for Capital Formation.

Beshears, J., Choi, J. J., Harris, C., Laibson, D., Madrian, B. C., and Sakong, J. (2015). Self Control and Commitment: Can Decreasing the Liquidity of a Savings Account Increase Deposits? NBER Working Papers 21474, National Bureau of Economic Research, Inc.

Carrera, M., Royer, H., Stehr, M., Sydnor, J., and Taubinsky, D. (2019). How are Preferences For Commitment Revealed? NBER Working Papers 26161, National Bureau of Economic Research, Inc.

de Bresser, J. and Knoef, M. (2015). Can the dutch meet their own retirement expenditure goals? *Labour Economics*, 34:100–117. European Association of Labour Economists 26th Annual Conference.

DellaVigna, S. and Malmendier, U. (2004). Contract Design and Self-Control: Theory and Evidence. *The Quarterly Journal of Economics*, 119(2):353–402.

Diamond, P. and Kőszegi, B. (2003). Quasi-hyperbolic Discounting and Retirement. *Journal of Public Economics*, 87(9-10):1839–1872.

Fahn, M. and Seibel, R. (2022). Present Bias in the Labor Market – When it Pays to be Naive. *Games and Economic Behavior*, 135:144–167.

Findley, T. S. and Caliendo, F. N. (2015). Time Inconsistency and Retirement Choice. *Economics Letters*, 129:4–8.

Giné, X., Karlan, D., and Zinman, J. (2010). Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation. *American Economic Journal: Applied Economics*, 2(4):213–35.

Heidhues, P. and Kőszegi, B. (2009). Futile Attempts at Self-Control. *Journal of the European Economic Association*, 7(2-3):423–434.

Kaur, S., Kremer, M., and Mullainathan, S. (2010). Self-Control and the Development of Work Arrangements. *American Economic Review Papers and Proceedings*, 100(2):624–28.

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *The Quarterly Journal of Economics*, 112(2):443–478.

Laibson, D. (2015). Why Don't Present-biased Agents Make Commitments? *American Economic Review Papers and Proceedings*, 105(5):267–272.

Merkle, C., Schreiber, P., and Weber, M. (2024). Inconsistent retirement timing. *Journal of Human Resources*, 59(3):929–974.

Phelps, E. S. and Pollak, R. A. (1968). On Second-Best National Saving and Game-Equilibrium Growth. *The Review of Economic Studies*, 35(2):185–199.

Schilbach, F. (2019). Alcohol and Self-Control: A Field Experiment in India. *American Economic Review*, 109(4):1290–1322.

Zhang, L. (2013). Saving and Retirement Behavior Under Quasi-hyperbolic Discounting. *Journal of Economics*, 109:57–71.

# Chapter 4: Relational collusion in the Colombian electricity market

Joint work with Miguel Espinosa, Rocco Macchiavello, and Carlos Suárez.

**Abstract**

Under collusion, firms deviate from current profit maximization in anticipation of future rewards. As current profit maximization places little restrictions on firms' pricing behaviour, collusive conduct is hard to infer. We show that bids from certain firms in the Colombian wholesale electricity market collapsed immediately after the *announcement*, and before the *implementation*, of a reform that potentially made collusion harder to sustain. After ruling out confounders, we uncover how the cartel functioned and how firms may have communicated. Calibrating the dynamic enforcement constraint confirms that collusion was sustainable before, but not after, the reform. The conclusions discuss policy implications.[73]

## 4.1 Introduction

Parties often rely on informal arrangements sustained by the value of future interactions to cooperate when contracts are unenforceable (Macchiavello, 2022). These arrangements

[73]We thank Jaap Abbring, Margherita Borella, Jochem de Bresser, Eric French, Estelle Cantillon, Sylvain Chassang, Paola Conconi, Francesco Decarolis, Michele Fioretti, Guido Friebel, Bob Gibbons, Jonas Hjort, Mitsuru Igami, Matti Liski, Mar Reguant, Tristan Reed, Alvaro Riascos, Ksenia Shakhgildyan, Otto Toivanen, John Van Reenen and conference and seminar audiences at 2020 SIOE, 6th Workshop on Relational Contracts, UCL-IFS, Tilburg, LSE, Helsinki, LEAP Frankfurt, ECARES and NBER Org. Economics Fall 2002 for many comments and suggestions. Jairo Galvis provided exceptional research assistance. We thank comments and insights from experts in different Colombian Regulatory entities: XM, SSPD and CREG. An earlier version of the paper circulated as CEPR Press Discussion Paper No. 18056. Also, Chapter 4 in the PhD thesis of Carlos Suárez (*Essays on Regulation, Liberalization and Privatization in Energy Markets*, 2019) studies the same regulatory change. However, the differences between the two papers, as explained in the text, are substantial.

benefit their participants but may harm the market as a whole. Horizontally competing firms colluding to raise prices – cartels – offer perhaps the most prominent example. Such cartels might be particularly relevant in developing countries, where entry barriers protect colluding incumbents (Djankov et al., 2002), competition authorities are weaker − if at all existent (World Bank, 2016; Besley et al., 2020) and markets thinner and more concentrated (Mitton, 2008; Leone et al., 2022).

Despite the policy relevance, evidence on cartels in low-income countries and how they function remains scarce. Most empirical studies focus on cartels investigated by competition authorities. As those are weaker in developing countries, fewer documented examples exist.[74] Furthermore, collusive behaviour is notoriously difficult to identify (Chassang and Ortner, 2023). In models of collusive behaviour, firms deviate from current profit maximization in anticipation of future rewards. Profit maximization, however, places little restrictions on firms' behaviour making these models hard to test and collusive conduct hard to infer from pricing behaviour alone.

This paper uncovers collusion in the Colombian energy sector. Access to adequate sources of reliable and cheap energy is a critical engine for development (Greenstone, 2014). Besides its intrinsic relevance, the context enables us to develop a novel test of collusive behaviour supported by future rewards. First, the Colombia wholesale electricity market is regulated and therefore detailed data from its uniform price multi-unit daily auctions are available. Second, we take advantage of a regulatory change. In 2008 the market witnessed a significant increase in prices. During a meeting on January $6^{th}, 2009$ (the *announcement* date), the regulator, concerned with the price increase, invited Professor Peter Cramton to advise on market rules. At that time, the operator disclosed all information to all market participants with two days delay. Such transparency increases market efficiency and simplifies monitoring and implementation. Cramton, however, had previously advised regulators on how transparency also facilitates collusion and was thus expected to recommend a tightening of the transparency policy. Indeed, Cramton recommended increasing to 90 days the delay to disclose information to market participants during a presentation delivered on January $24^{th}, 2009$. The regulator adopted the recommended change on January $30^{th}, 2009$, with effect from February $6^{th}, 2009$ onward (the *implementation* date).

A subset of the firms in the market lowered bids by between 47% and 30% immediately after the *announcement* − and well before the actual *implementation* − of the regulatory change. Consistently with the key implications of models of collusion, (at least some) members of the cartel reacted to the announcement in an anticipatory way, leading to its

[74]For example, the Private International Cartel database (Connor, 2020) reveals that only 5% of proven cartels are in Africa (72% of those were proven in South Africa alone), 7% in Latin America and 11% in Asia.

instantaneous unravelling. The logic of the test doesn't rely on cartel members perfectly foreseeing the actual change in transparency enacted by regulators and how it would make the cartel unsustainable. It simply requires some of them to become sufficiently pessimistic about their *future* ability to sustain a cartel. Inter alia, our strategy allows us to rule out several confounders, including the fact that changes in market transparency itself could alter firms' bidding behaviour. We also investigate the effect of unannounced inspections both before and after the announcement to explore whether participants might have also updated beliefs about the likelihood of enforcement and find little support for such a mechanism. The evidence shows that dynamic enforcement considerations underpin collusive behaviour – an observation with policy implications that we discuss in the conclusions.[75]

Green and Porter (1984) discusses the sustainability of noncooperative collusion under imperfect price information. They show that, if the information which firms use to monitor whether the cartel is in a collusive or reversionary state is only imperfectly correlated with firms' conduct, then both competitive and collusive behaviour will be observed. Otherwise, for perfect correlation, reversion would never occur. A main implication from their work is that recurrent episodes in which price ("price wars") and profit levels sharply decrease are consistent with the existence of collusion in the market (as opposed to the interpretation that "performance of this type indicates an industry where firms are engaging in a sequence of abortive attempts to form a cartel"). In our setting, before the policy reform, firms can directly and immediately (that is in t+1) observe the bids of their competitors, meaning that – if the cartel is sustainable – reversions to competition would never occur. In fact, that would make it difficult to infer competitive or collusive conduct from observed prices. The fact that we observe a drop in bids *before* the reform implementation, i.e. when the monitoring information is still perfectly correlated with conduct, is exactly what implies that the cartel became unsustainable.

It is important to clarify what our evidence is *not* meant to prove. The evidence isn't sufficient to pin down a particular equilibrium concept of collusive conduct. For example, while subgame perfect equilibrium (SPE) relies on players reacting to information instantaneously (which is consistent with our evidence), it also assumes correct beliefs about future payoffs and play both on- and off- equilibrium paths (about which, instead, our evidence is essentially mute). Our test for collusion does not assume – and the evidence certainly does not prove – that firms in the cartel were playing SPE. Similarly, the evidence is consistent with – but doesn't prove that – limited feedback about auction

[75]We complement our main findings with a forensic analysis that uncovers the mechanisms through which firms colluded as well as suggestive evidence of communication. We also calibrate the dynamic enforcement constraints required to sustain the cartel and confirm that they were satisfied before, but not after, the reform.

outcomes leading to the cartel's collapse.[76]

Section 4.2 provides background information on the Colombian wholesale electricity market, the regulatory change, and the data. We describe the ideal dispatch (i.e., the production allocation resulting from submitted bids) and the real dispatch, in which the market regulator allocates production taking into account shocks to the transmission network. This is done through a process of positive and negative reconciliations which, as we later clarify, plays a critical role in our analysis.

Section 4.3 presents the main evidence and rules out confounders. Chassang and Ortner (2023) elucidate the challenges involved in identifying collusive conducts in the data: e.g., non-competitive behaviour is not necessarily collusive (e.g., firms might make mistakes); in dynamic environments, pricing behaviour can deviate from static profit maximization without implying collusive conduct. Our test identifies an *instantaneous* response in anticipation to future changes in market conditions that is the central implication of reward-punishment schemes at the heart of collusive equilibria and arguably overcomes most of these challenges.

Unlike studies that rely on proven cartels (see, e.g., Porter and Zona, 1993; Asker, 2010; Igami and Sugaya, 2021), we do not know the identity of the firms participating in the collusive arrangement – if one existed. We construct several proxies for cartel membership to sharpen our test. In our baseline definition, we conjecture that thermal units in the Atlantic region had the incentives and ability to form a cartel.[77] This classification isolates a group of 14 units − henceforth, the *cartel*. Using both DID and more flexible event-study specifications, we show that the average bid for cartel units falls after the announcement, and before the implementation of the regulatory change.[78]

Section 4.4 conducts forensic analysis to uncover the incentives, and strategies used, to collude. When awarded a positive reconciliation, a unit is paid a price proportional to its bid rather than the lower market clearing price emerging from the ideal dispatch. Using an instrumental variable strategy, we thus begin by confirming that firms submit higher bids when they expect to be awarded a positive reconciliation. We then show that units in the cartel coordinated their bids. Specifically, some units increased bids particularly so at times in which other units in the cartel bid low prices, win the ideal

[76]Moreover, there are possibly other ways in which the cartel operated, which we are not able to identify or rule out (for example, collusion on the prices of forward contracts).

[77]The rationale for this choice is that thermal units have higher costs and can't make profits in the ideal dispatch. We thus hypothesize, and later confirm, that thermal units profit from colluding on the positive reconciliations market. Because positive reconciliations occur when there are disruptions to transmission or generation, units are more likely to compete for positive reconciliation with nearby units. This justifies the regional focus of the cartel despite the unique national market.

[78]Results are robust to alternative definitions of the cartel, several confounders and placebo specifications.

dispatch, and subsequently declare unavailable thereby generating positive reconciliations for the high bidders in the cartel. This coordinated behaviour shows up only for cartel units and ceases after the reform.

Courts require evidence of the express agreement and overt communication to declare collusive behaviour illegal (Chassang and Ortner, 2023). We do not observe whether members of the collusive arrangements explicitly communicated and/or whether they used transfers to share the spoils. However, we look into both issues. First, we use data from the minutes of the meetings of the Association of Generating Units (CNO in Spanish). The association holds regular meetings to discuss engineering problems related to technical difficulties and constraints on the network and prohibits discussions about bidding behaviour. We downloaded the minutes of all the meetings in 2008 (during the cartel) and 2009 (after the announcement date). Within a DID framework, we find that after the reform, units in the cartel stopped sending employees involved in setting bids to the meetings. The meetings of the association might have thus been used to discuss bids and collude. Second, if the forensic analysis is correct, we expect that profits – particularly from positive reconciliations – fall relatively more for cartel units after the end of the cartel. We confirm this to be the case in the data.[79]

Section 4.5 quantifies the incentive to collude and the cost of the cartel for consumers. First, we calibrate the optimal static bidding strategy and show that, before − but not after − the announcement date, cartel units could increase static profits by submitting lower bids. In contrast, bids from similar non-cartel units are in line with profit maximization. We then embed such deviations into a dynamic incentive compatibility constraint. For reasonable parametrizations of the discount factor, such deviations are not incentive compatible under the old transparency rule but become so under the new rule. As noted above, our evidence is not meant to establish that the change in transparency led to the demise of the cartel. The estimates, nevertheless, provide a sanity check that such an interpretation would be consistent with economic magnitudes in our context.

Counterfactuals allow us to provide a lower bound of the excessive costs paid by consumers for electricity during the cartel. The cartel increased by 12% the price paid for positive reconciliations. Positive reconciliations account for approximately 10% of the electricity procured by the regulator, but since they are paid above the spot price, the overall increase in costs was 2.5-3%.

Section 4.6 discusses the policy implications of our results.

This paper contributes to three branches of the literature on firms in developing countries: on collusion, on energy markets, and on relational contracts.

We contribute to the empirical literature on collusion (see Asker and Nocke, 2021, for

[79]Across units in the cartel, however, profits fell for *all* units – regardless of their costs and role in the cartel. Transfers may *not* have been needed to sustain this cartel.

a recent review). A first branch of the literature studies known cartels to gain insights into their functioning and quantify associated efficiency losses (see, e.g., Porter and Zona, 1993, 1999; Asker, 2010). Igami and Sugaya (2021) calibrate the dynamic incentive compatibility constraint of the collusive arrangement in the international vitamin C cartel to perform counterfactuals and is particularly related to our paper.

A second branch designs empirical tests to detect anti-competitive behavior when a cartel has not been proven. Porter (2005) and Harrington (2008) provide overviews of the literature. Chassang and Ortner (2019) study of procurement in Japan derives a test from the dynamic enforcement constraint and is particularly related to our paper. They note that higher minimum prices can make punishment less effective and lead to lower winning bids. Instead, we exploit the fact that the announcement of a *future* change in market transparency leads to the instantaneous demise of the cartel. Chassang and Ortner (2023) discuss the processes involved in regulating collusion, including the information required not just to mark collusive behaviour as illegal, but even to hear a case and begin an investigation. The logic of our empirical test and the combination of forensic approaches are applicable to other contexts and might help meet the informational hurdle, at least in some cases.[80]

There is a general perception, but limited evidence, that cartels are particularly common in developing countries (World Bank, 2016).[81] The critical role of electricity for the development process is increasingly appreciated (Rud, 2012; Lipscomb et al., 2013; Greenstone, 2014; Allcott et al., 2016). A recent review (Greenstone et al., 2021) notes that "rigorous evidence from developing countries on market design is lacking" (see also World Bank, 2019).[82] Intrinsic features of electricity markets make them prone to abuse of market power and even collusion − evidence on which policies improve market efficiency is thus particularly valuable. For example, through counterfactual simulations, Ryan (2021) finds that a more integrated grid would increase surplus by 22% in the Indian market.[83]

[80]Indirectly, we also contribute to ongoing debates on collusion and market transparency. Conventional wisdom holds that transparency facilitates collusion (see, e.g., Whinston, 2008; Perloff and Carlton, 1999). A number of notable contributions, e.g., Genesove and Mullin (2001) study of the sugar cartel in the U.S. and Albæk et al. (1997) analysis of the Danish antitrust authority's decision to publish firm-specific transactions prices of ready-mixed concrete in three regions, support this view. The evidence and theoretical literature on the matter, however, is less conclusive. Sugaya and Wolitzky (2018) argue that transparency can hinder cartels by helping firms devise more profitable deviations and discuss examples in which that appears to have been the case.

[81]Asker and Nocke (2021) review cites only two studies on collusion in developing countries (Bergquist and Dinerstein, 2020, on Kenya maize and Barkley, 2023, on Mexican insulin).

[82]The literature on energy markets in advanced econmices is vast (see Kellogg and Reguant, 2021 for a review). Fabra and Toro (2005) test for collusion in the Spanish market.

[83]A few papers study the Colombian electricity market, albeit with a different focus (Camelo et al., 2018, on centralized unit commitment, Fioretti et al., 2024, on substitution between fossil fuels and

Finally, markets in developing countries are characterized by weaker formal contract enforcement and governance, making the study of relational contracts particularly important (see, e.g., Macchiavello, 2022, for a review). We here use the term 'relational collusion' to underline the similarity between the underlying functioning of cartels and of relational contracts. The key difficulty in testing models of relational contracting is that neither the *future* value of the relationship nor the *current* temptations to deviate are typically observed. Macchiavello and Morjaria (2015) tests the implications of a relational contracting model exploiting information on temptations to deviate and an exogenous supply shock in the Kenya flower sector. Blouin and Macchiavello (2019) uses unanticipated increases in temptations to deviate to test for, and quantify, the extent of opportunistic behaviour in the international coffee market. We contribute a test for relational contracting that relies on changes in *current* behaviour in anticipation of changes in the *future* value of the relationship − a central implication of relational contracting models.[84]

For the sake of clarity, it's worth to highlight that Chapter 4 in the PhD thesis of Carlos Suárez (*Essays on Regulation, Liberalization and Privatization in Energy Markets*, 2019) also studies the same regulatory change. The differences between the two papers, however, are substantial in terms of research question, empirical approach, and (partially) also for the data used. In particular, Carlos' thesis focused on the different reaction of private versus public firms to the reform. In this chapter, instead, we hypothesize and test the idea that a cartel existed only among a selected group of units. Those units are characterized by their production technology and location, rather than their ownership. We then study how this group of cartel units reacted differently to the announcement of the reform compared to other units, but also provide additional evidence on the existence and functioning of a cartel through novel exercises. Those are, mainly: (i) the evidence that units increase prices when they are more likely to expect positive reconciliation, (ii) the evidence on prices coordination across units, (iii) the suggestive evidence on communication across firms, (iv) the evidence on the fall of profits for cartel units, (v) the evidence on cartel units deviating from static-profit maximization, (vi) the calibration of the incentive to deviate from the collusive agreement, (vii) the quantification of the cost of the cartel for consumers.

hydropower, and Suarez, 2023, 2022, on the interaction of market power and public ownership).

[84]Ghani and Reed (2022) study how relationships evolved in response to an increase in supply in the Sierra Leone market for ice, Macchiavello and Morjaria (2021) finds that higher competition inhibits relational contracting in the Rwanda coffee chain.

## 4.2 Institutional setting & background

This section describes the Colombian wholesale electricity market and the timeline of events used in Section 4.3 to detect the collusive arrangement.

### 4.2.1 Electricity demand and generation

The average daily generated electricity in Colombia was 149.81 GWh in 2009.[85] In 2008/2009 electricity was produced by 47 generation units. Among these units, 32 units owned by 11 private firms produce about 70% of the market output, and the rest is produced by publicly owned units. The market was a moderately concentrated oligopoly in 2008/2009 with a Herfindahl-Hirschman index of installed capacity around 1306 (see CREG, 2009a). The 4 largest firms accounted for 65% of installed capacity.[86] Data are described in Appendix 4.7.1.

Bids to supply electricity in the wholesale market are submitted by individual units. Most of our analysis, therefore, considers units, rather than firms, as the relevant decision-makers. However, we use information on firms' ownership of units for robustness checks and to gain further insights into the functioning of the collusive agreement.

Electricity was generated using different technologies: 66.7% hydro-power, 32.9% thermal generation (20.4% gas-fired, 7.3% coal-fired, and 5.2% other fuels). Thermal generation mainly relies on gas and coal. In 2009, 82% of gas consumption for electricity generation came from the basin Guajira, located on the northern coast of the country. Colombia was the fourth largest exporter of coal in 2009. Most coal-fired units are located close to large coal mines. Coal is usually transacted through long-term contracts with negotiated prices.

### 4.2.2 Colombian wholesale electricity market

Electricity markets are characterized by volatile demand, prohibitively high storage costs, and economies of scale. To improve efficiency, encourage participation, and minimize expected payments to generators, many countries trade electricity through auction mechanisms. The Colombian wholesale electricity spot market works as a uniform price multi-unit procurement auction.[87] Once a day, each unit submits its hourly availability and a

[85] For comparison, it was: 1277,15 in Brazil, 340,82 in Argentina, 260,93 in Pakistan, 54.18 in Nigeria, 24.54 in Ghana, 937,02 in the UK, and 10822,82 in the US.

[86] These values are typical of other developing countries, for instance, the HHI index was 3.500 in Kenya, 2,300 in Peru, and 677 in Pakistan World Bank (2016).

[87] Uniform price multi-unit auctions electricity markets include Spain (Fabra and Toro, 2005), Texas (Hortacsu and Puller, 2008) and U.K. (Crawford et al., 2007).

unique bidding price for the next day. Although only one bidding price is allowed for each unit per day, the Colombian wholesale electricity market clears every hour. There are no intra-day balancing markets and the same spot price is paid in all the regions.

Once the units have submitted their bids, *XM*, the system operator, minimizes the cost of fulfilling the demand for each hour, by arranging in increasing order the submitted bids. For each hour, the price that clears the market, the *spot* price, is the bidding price of the marginal unit necessary to fulfill the demand. This process, which does not consider transmission network restrictions, gives rise to the *ideal dispatch*. It establishes for each unit how much and at which hour it should supply energy to the system. Throughout the sample period, hydro-power units tended to have significantly lower costs than thermal units and were the marginal bidder around three-quarters of the time.

Once the ideal dispatch has been determined, contingencies such as transmission constraints may arise and make unfeasible the initially planned allocation.[88] As a consequence, *XM* proposes a different set of production assignments, called the *real dispatch*. Units that were initially called upon to produce but cannot supply electricity to the network do not do it, while units that were not called upon may be called in. To compensate the generators for the differences between the *ideal* and *real dispatches*, the market operator has a scheme called positive and negative reconciliations.

A unit receives a *positive reconciliation* when the real dispatch allocation is greater than the ideal dispatch. In that case, the system compensates each energy unit at a price equal to the minimum between a cost-based regulated price and the generation unit's bidding price.[89] In case two or more units are eligible to be called for positive reconciliations, the system regulator selects the one with the lowest bidding price. A *negative reconciliation* arises when the real dispatch generation is less than the ideal dispatch generation. The system compensated these units at a price equal to the average between the spot price and the unit's bid.

[88]The actual availability can be lower than the expected availability due to exogenous reasons (e.g., production shocks to the unit) or to strategic decisions (the unit decides to produce less than declared to the regulator ahead of the auction –i.e., declare unavailable some or all the initial production capacity). In the data, we are unable to distinguish between these two motives. Regulators investigate units that declare unavailabilities frequently.

[89]Unfortunately, we do not have data on the cost-based regulated price or information on how it is determined. However, profit-maximizing units should always submit bids below the regulated price cap. In the six months before the announcement, the average cartel bid was 1,213 COP and the average production cost 113 COP (Table 4.1). This suggests that the regulated price was at least ten times larger than costs. Still, it is difficult to say whether this price cap is "too high" or not. In fact, even for competitive units, the submitted bids were six times larger than costs (362 COP vs. 61 COP, Table 4.1). Fixed costs likely play a prominent role in this industry, which might justify a relatively large price cap.

### 4.2.3 Change in transparency policy

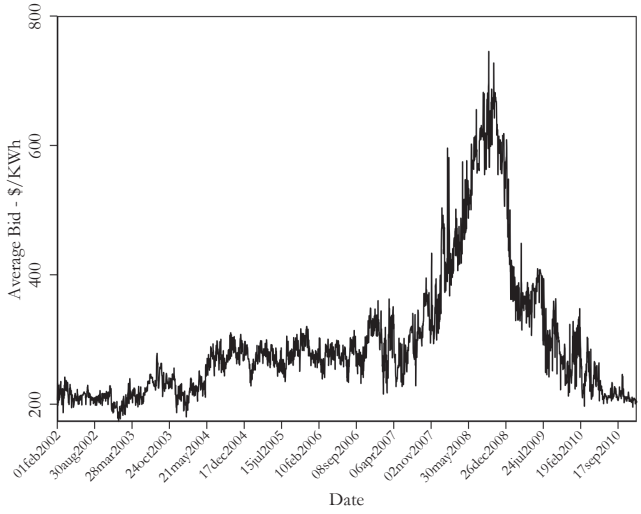



Figure 4.1: Average bid time series.

Note: Daily time series of the average bid from 2002 to 2010 in the Colombian wholesale electricity market.

The average bid in the wholesale electricity market markedly increased during 2007/2008 (see Figure 4.1). The electricity market regulator began to suspect that, among the potential reasons to explain the sharp increase, anti-competitive practices might have been at play.[90]

Figure 4.2 summarizes the timeline of events leading to the policy change. To deal with the price increase, the authorities held a meeting on January 6th, 2009, a date that we label *announcement date*, to discuss measures to deal with the increases in bids. During this meeting, it was decided to hire Professor Peter Cramton as a consultant for the case to advise on potential changes to market design, including its transparency policy.

Cramton advised several governments on auction design before advising Colombian regulators. In particular, and possibly known by Colombian market participants, Cramton consistently mentioned the importance of considering the relative costs and benefits of market transparency. On the one hand, transparency might improve efficiency, but on the other hand, it might facilitate collusion (Cramton and Wilson, 1998). In those cases where the market is expected to suffer from collusion, Cramton argued against a fully-transparent policy (Cramton and Wilson, 1998; Cramton and Schwartz, 1998a,b).

[90]See (Superintendencia de Servicios Publicos, 2008) and CREG (2009a), page 74.

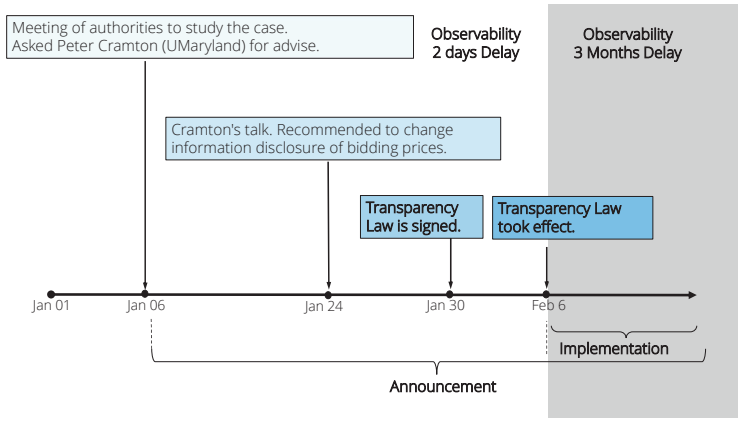



Figure 4.2: Timeline.

Note: Timeline of the announcement and implementation of the transparency policy.

On January 24th, 2009, Cramton recommended changing the bidding disclosure policy. Before the implementation of the policy, production schedules (ideal and real dispatches) and bidding prices at date $t$ were released as public information *two* days after (in $t+2$). Cramton recommended revealing all bids at $t+90$, only 90 days after the auction took place.

Following his advice, regulators approved Resolution 006 on January 30, 2009 with effect on February 6th, 2009 (CREG, 2009a). The law mandated that from then onward day $t$ production schedules and bidding prices would become public information only ninety days after (in $t+90$). The spot price for each hour of the day $t$ was still available to everyone, two days after. Privately though, each generation unit was informed whether or not they won in the multi-unit auction or they had any type of reconciliations. The measure also mandated that the generating units kept their bidding programs' information secret from other units. The law established that failure to comply with the disclosure policy would be sanctioned.

## 4.3 Detecting collusive agreements

This Section provides evidence that a cartel was likely operating in the Colombia wholesale electricity market. We begin explaining the logic of our empirical test and present the

sudden decrease in bids around the time of the reform. Then, we introduce a proxy for cartel membership and present DID and event-study specifications. Finally, we perform several robustness checks and rule out confounding explanations.

### 4.3.1 The logic of the test for collusive behaviour

The argument has two parts. First, we argue that a central implication of repeated-games models of cartel behaviour is that shocks to parties' *future* ability to detect deviations should lead to *instantaneous* changes in behavior. Second, we review theoretical arguments regarding the role of information in auction markets and its role in facilitating collusion. Although our test does not rely on potential cartel members having correctly anticipated the ensuing regulatory change at the time of the announcement, we still describe how the reform likely reduced parties *future* ability to punish deviations leading to an *instantaneous* unraveling of the cartel. The test allows us to check the existence of a cartel and whether it was dissolved by the regulator's actions. It is however important to stress that the test does not allow us to definitively conclude that it was the anticipation of a less transparent market regime that induced the cartel's demise.

In models of collusive behaviour firms deviate from current profit maximization in anticipation of future rewards. As current profit maximization places little restrictions on firms' pricing behaviour, these models are hard to test and collusive conduct is hard to infer from pricing behaviour alone (Ortner et al., 2022). Repeated-game models of collusive behaviour share a central insight with models of relational contracting: the *future* value of the relationship – the discounted (expected) difference in the payoffs from cooperation and defection – deters *current* temptations to deviate – the difference in payoff between deviating from the agreement and sticking to it (Baker et al., 2002). The key difficulty in testing these models is that neither the *future* value of the relationship nor the *current* temptations to deviate are typically observed (Macchiavello, 2022). The former depends on discount rates, on beliefs about other players' future behaviors on- and off-the-equilibrium-path. The latter on the off-the-equilibrium-path payoffs associated with defection. Discount rates are difficult to estimate, and beliefs and off-the-equilibrium-path actions are not observed in the data.

A key implication of these models, however, is that *anticipated* changes to *future* relationship value should lead to *instantaneous* changes in behaviour. To the extent that the announcement date induced at least some of the members of the potential cartel to become sufficiently pessimistic about their ability to sustain a collusive arrangement in the future, the ideal test can be implemented in our context exploiting the difference between the *announcement* and *implementation* dates. Of course, we cannot prove that any of the firms anticipated the exact reform, nor that they were able to work out its implication

for the equilibrium of the collusive arrangement, if one was indeed being played at all. *A fortiori*, the test does not allow us to infer the exact equilibrium (e.g., a subgame perfect one) played by firms. We return to a discussion of this issue further below.

Transparency can potentially increase efficiency and simplify implementation (see, e.g., Cramton and Wilson, 1998).[91] Transparency, however, can also facilitate collusion (see, Perloff and Carlton, 1999; Whinston, 2008, and Cramton and Schwartz, 1998a,b). For instance, if the regulator reveals the bids of all bidders, a cartel faces a much easier problem in policing its agreement. A reduction in market transparency worsens parties' ability to detect, and increases payoffs from, defection. Of course, changes in market transparency could influence behaviour through different channels.[92] In our case, the difference between the *announcement* and the *implementation* dates allows us to rule out such confounders: holding constant current temptations to deviate, the anticipation of less transparency in the future makes it harder to satisfy current dynamic incentive constraints and immediately increases the likelihood of defection.

**Test for Collusion:** *At least some of the units that belong to a cartel sustained by a relational arrangement lower their bids after the announcement, and before the implementation, of the regulatory change.*

### 4.3.2 The main fact & proxying for cartel membership

We do not know the identity of the firms participating in the collusive arrangement. Yet, such information, or a proxy for cartel participation, would allow us to sharpen our empirical test and investigate mechanisms. We thus construct a proxy for cartel membership. To define a baseline proxy, we put forward two characteristics of the units that we believe, on a priori grounds, to be correlated with units' incentives to enter, and ability to sustain, a collusive arrangement. Specifically, in our baseline definition, we hypothesize that cartel units are those *thermal units located in the Atlantic region.* All but one of the 15 units in the Atlantic region are thermal. The baseline definition yields 14 units in the cartel (9 private and 5 public) belonging to 5 firms.[93]

It is worth describing the rationale for our baseline choice. First, thermal units have larger marginal costs than hydro units (Knittel and Stango, 2003). Due to higher costs,

[91]For example, better knowledge of the residual demand allows hydro units to improve the intertemporal allocation of scarce water resources. Also under transparency, the regulator does not need to worry about her employees or generation units sharing information with other market participants.

[92]For example, more information provides firms with more precise estimates of their residual demand curve potentially altering bidding behaviour.

[93]Barranquilla 3 and 4, Guajira 1 and 2 and Tebsab from firm GECELCA; Cartagena 1, 2 and 3 from firm EMGESA; Flores 2 and 3 from firm COLINVERSIONES; Proelectrica 1 and 2 from firm PROELECTRICA and finally, Termocandelaria 1 and 2 from firm TERMOCANDELARIA.

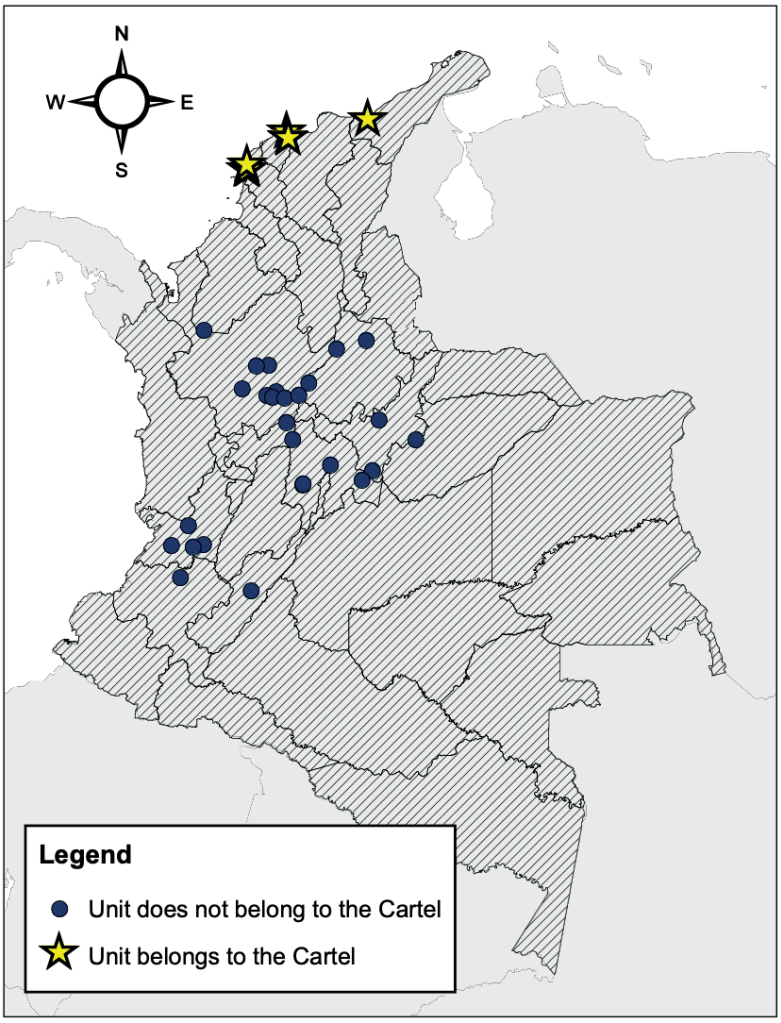



Figure 4.3: Geographical location of cartel units.

Note: The figure presents Colombia's map and the location of the electricity generation units participating in the wholesale electricity market in 2008/2009. The dark-shaded area represents Colombian territory. The black thick line represents the division of the country in political units called "departamentos". The star-shaped yellow shows cartel generation units and the circle-shaped blue shows non-cartel units.

thermal units do not win on the ideal dispatch even when they bid at marginal costs. We hypothesize, and later confirm, that thermal units might profit from colluding on high

bids through the process of positive reconciliations. As explained above, conditional on receiving a positive reconciliation, a unit receives a price that is tightly linked to its bid – potentially strengthening the incentives to collude.

At the same time, because positive reconciliations occur when units that won the ideal dispatch face disruptions (e.g., due to network transmission reasons), units nearby are more likely to compete with each other for positive reconciliations. Units within the Atlantic region are not just near each other, but they are also far from the rest (see Figure 4.3). This geographical situation creates, in practice, a separate market when disruptions isolate the Atlantic coast from the rest of the network. This isolation has three distinct effects. First, it increases the importance of reconciliations relative to the ideal dispatch. Second, it softens competition by reducing the number of direct competitors (especially excluding the more competitive hydro units). Third, it reduces the number of players that would need to coordinate to sustain a cartel. All these mechanisms thus create incentives to collude relative to other parts of the country.[94]

Figure 4.1 shows a large drop in the average bidding price around the policy change described in Figure 4.2. Zooming in around the regulatory change and splitting units into two groups − cartel units and the rest −, Figure 4.4 shows a sharp decrease in the average bidding price right after the *announcement* date *only* for cartel units. The average price for these firms falls by about 43% − the price for other units barely moves.[95] Figure 4.13 present the same information over a longer period of time, showing how cartel bids progressively increased as of October 2007.

This parsimoniously constructed proxy might be imprecise and/or *ad hoc*. We thus explore robustness and alternative definitions along several dimensions. First, we note that the baseline definition *does not* rely on the unit's bidding behaviour around the time of the policy change − there is thus no mechanical correlation between variation used to proxy for cartel membership and bidding behaviour around the time of the reform. We exploit, however, information on units' bidding behaviour in a battery of robustness checks. Second, our proxy might suffer from both type-I and type-II errors. Provided our proxy is moderately positively correlated with actual membership in the cartel, miss-classification of units into (and out of) the cartel leads to attenuation bias, making it harder for us to

[94]Besides the local market created by network constraints, collusion might be more likely to occur under weak institutions: The Atlantic region has the worst governance and highest corruption in Colombia (Duque, 2014).

[95]To counter the negative effects of El Nio, the 90-days disclosure rule was eliminated on December 3, 2009 and the market reverted back to the older two-days disclosure policy (CREG, 2009b). Figure 4.14 does not show any different behavior around this policy change for cartel vs non-cartel units. Given the difficulty of building a collusive arrangement and its fragility (Byrne and De Roos, 2019), we do not expect the change in policy to necessarily result in a new cartel or the old cartel reverting back to its old behavior.

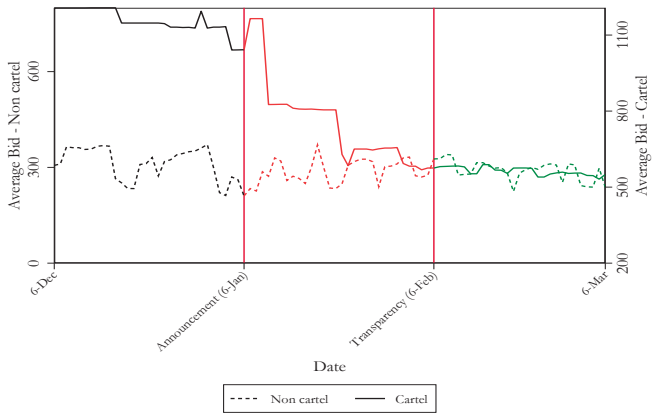



Figure 4.4: The Main Fact: Cartel and non cartel groups bids.

Note: Time series of the average bid of the cartel (solid line, right axis) and non-cartel groups (dashed line, left axis) around the dates of announcement and implementation of the transparency policy. The vertical lines show the announcement and implementation dates.

detect collusive behaviour (see Mirenda et al., 2022, for a similar argument). Nevertheless, we explore alternative definitions of the cartel proxy, relying both on variation in the characteristics considered for classification, on additional information, and on placebos.

Table 4.1 presents descriptive statistics for units classified inside and outside the cartel before and after the announcement date. Besides providing descriptive statistics, the top panel reveals patterns consistent with intuition.[96] Bids from cartel units are about 4 times larger than bids from non-cartel units. This contrasts with cost differences that are only about 2 times larger for the cartel group. Cartel units have a lower probability of receiving positive reconciliations, and conditional on receiving them, revenue from these reconciliations is larger for cartel units than non-cartel ones. Note that units receive positive reconciliation infrequently. It is thus almost impossible to infer deviations of other cartel members solely relying on one's own reconciliations information. Market transparency might be needed for a cartel to function.[97]

[96] As expected, units classified in the cartel are more likely to be privately owned (64% vs. 48%). Private units maximize profits while public firms might maximize profits as well as other objective functions (Barros and Modesto, 1999).

[97] Cartel units have a lower fraction of forward contracts, relative to other units. As firms serve forward contract obligations independently of the level of the spot price, firms that have committed a large portion of their capacity in forward sale contracts have a lower incentive to increase prices (Wolak, 2007) and collude.

Turning to the comparison between the top and bottom panels, we see that the bids of units classified in the cartel decreased after the announcement of the policy significantly more than the bids of other units. Furthermore, the amount of positive reconciliations and associated revenues decrease, while availability increases more for cartel units than non-cartel ones. Of course, patterns in Table 4.1 are only suggestive – we now subject our hypothesis to more rigorous testing.

| | | Before | 06/01/2009 | | | | |
|---|---|---|---|---|---|---|---|
| | Cartel | | | No Cartel | | | |
| Variable(Unit) | Obs | Mean | SD | Obs | Mean | SD | T-Test |
| Bid(COP/KWh) | 2212 | 1213.57 | 714.17 | 5214 | 362.06 | 557.36 | 49.99 |
| Ratio forward contracts/availability(Percentage) | 2212 | 0.27 | 0.25 | 5046 | 0.67 | 1.17 | -23.11 |
| Probability positive reconciliation(probability) | 2212 | 0.13 | 0.31 | 5214 | 0.24 | 0.34 | -13.50 |
| Average Positive reconciliation(KWh) | 2212 | 22702.29 | 76145.57 | 5214 | 10127.97 | 29856.41 | 7.53 |
| Revenue from Positive reconciliation(Millions COP) | 2212 | 107.76 | 347.30 | 5214 | 17.87 | 53.33 | 12.11 |
| Average Availability(KW) | 2212 | 126946.42 | 164209.28 | 5214 | 282285.07 | 299716.71 | -28.64 |
| Estimated Marginal Cost(COP/KWh) | 2212 | 113.22 | 19.07 | 5214 | 60.55 | 63.09 | 54.69 |
| | | After | 06/01/2009 | | | | |
| | Cartel | | | No Cartel | | | |
| Variable(Unit) | Obs | Difference | T-Test | Obs | Difference | T-Test | T-Test |
| Bid(COP/KWh) | 2898 | -631.84 | 35.70 | 6831 | -73.00 | 7.65 | 27.38 |
| Ratio forward contracts/availability(Percentage) | 2898 | 0.01 | -1.35 | 6799 | 0.26 | -9.43 | -29.03 |
| Probability positive reconciliation(probability) | 2898 | 0.01 | -1.39 | 6831 | 0.04 | -5.47 | -18.23 |
| Average Positive reconciliation(KWh) | 2898 | -3157.89 | 1.51 | 6831 | 2263.29 | -3.64 | 5.12 |
| Revenue from Positive reconciliation(Millions COP) | 2898 | -36.15 | 4.10 | 6831 | 5.94 | -5.17 | 9.75 |
| Average Availability(KW) | 2898 | 8199.63 | -1.71 | 6831 | 731.54 | -0.13 | -30.20 |
| Estimated Marginal Cost(COP/KWh) | 2898 | -27.14 | 50.93 | 6831 | -7.95 | 7.22 | 44.44 |

Table 4.1: Descriptive Statistics

Note: The table presents the descriptive statistics of the cartel and non-cartel groups for two different periods before and after the announcement of the policy. Columns 2 to 4 present information on the cartel group while columns 5 to 7 present information on the non-cartel group. The top panel presents information for the period 1st August of 2008 until 6th January of 2009. The bottom panel starts on the 6th January of 2009 and ends 31st July 2009.

### 4.3.3 Cartel membership & bidding behaviour around the transparency reform

We use a difference-in-differences specification to quantify the differential change in bidding behaviour across the two groups around the time of the reform. We distinguish the announcement and the implementation of the policy, controlling for time-invariant heterogeneity across units and heterogeneous time effects. The baseline specification is given by:

$$ln(b_{it}) = \beta_1 \mathbb{1}\{Cartel\}_i \times \mathbb{1}\{Announ\}_t + \beta_2 \mathbb{1}\{Cartel\}_i \times \mathbb{1}\{Trnsp\}_t + \lambda_i + \mu_t + \epsilon_{it} \quad (4.1)$$

Where $ln(b_{it})$ is the logarithm of the bidding price of unit $i$ at date $t$, the dummy variable $\mathbb{1}\{Cartel\}_i$ takes the value of one if $i$ is a unit is classified to be in the cartel and zero

otherwise. The dummy variable $\mathbb{1}\{Announ\}_t$ takes the value of one if $t$ is a date after the announcement date (January 6th, 2009) and zero otherwise, the dummy variable $\mathbb{1}\{Trnsp\}_t$ takes the value of one if $t$ is a date after the implementation of the transparency policy (February 6th, 2009) and zero otherwise. $\lambda_i$ are unit fixed effects and $\mu_t$ are date fixed effects, which control for common market conditions (such as demand and input prices). We also explore specifications in which date fixed effects $\mu_t$ vary either by technology type or by region, as different technologies (e.g., thermal vs. hydro), or different regions, might be exposed to different daily shocks. Standard errors are two-way clustered by date and generation unit.

Table 4.2 presents the results. Across a variety of specifications, we find a statistically significant decrease in bidding prices of cartel units after the announcement of the policy. Depending on the specification, the estimates range between a drop of 47% and 30%.[98] Column (1) reports results without including any fixed effect. Column (2) controls for unit and date fixed effects and finds identical results. Columns (3) and (4) control for forward contracts.[99] Column (3) allows for the interaction of date-fixed effects with technology-fixed effects. Column (4) instead controls for the interaction of date-fixed effects with regional dummies.[100] Overall, we find a significant and negative coefficient for cartel announcement.[101] Interestingly, the coefficient for $\mathbb{1}\{Cartel\}_i \times \mathbb{1}\{Trnsp\}_t$ turns out to be small and statistically insignificant in specifications that more adequately control for potential confounders in columns (3) and (4): Market transparency did not further change bidding behaviour differently between units in the cartel (which had already collapsed) and units outside, once we account for the differential role of shocks (e.g., gas prices and

[98] These estimates are quite sizeable. Connor and Bolotova (2006) provides a meta-analysis of cartel overcharges and finds, in a sample of 395 documented cartels, a median (average) overcharge of 19% (29%).

[99] Incentives to collude depend on the fixed-price forward contracts signed by the unit. Figure 4.15 plots the daily average ratio of forward contracts over total availability for cartel and non-cartel units. Two patterns emerge. First, given the differences in levels of contract commitments, cartel firms have more incentives to collude than the rest of the units. Second, the drop in bidding prices of the collusive units is unrelated to a sudden change in the profile of forward contracting around the dates of the transparency policy.

[100] We cannot include the interaction date fixed effects with *both* regional dummies and technology type since there is only one non-thermal unit in the Atlantic region. However, in further robustness checks in which we use additional criteria to define the cartel, we include both sets of interactions simultaneously and obtain similar results.

[101] The differential drop in bids is not explained by a change in production costs for cartel thermal units. As a matter of fact, Table 4.1 shows that, if anything, the decrease in marginal costs before and after the reform for cartel units was stronger than the decrease for non-cartel units. Figure 4.16 shows an abrupt fall in the margin ($Bid - Mg.Cost$) for units in the collusive agreement but not for the rest of the units, on the dates after the announcement. Using margins instead of bidding price provides qualitatively similar results (see Table 4.7).

rainfall patterns) across technologies in column (3).

Figure 4.5 reports estimates from a more flexible event-study specification. We extend the baseline specification – defined in equation 4.1– including interactions between weekly dummies for leads and lags relative to the announcement date and the $Cartel_i$ dummy. First, the specification rules out differences in pre-trends in bidding behaviour between units assigned and not assigned to the cartel. Second, the differential drop in bids right after the announcement remains persistent throughout the rest of the sample period.[102]

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VARIABLES | LnBid | LnBid | LnBid | LnBid |
| | | | | |
| Cartel Announcement | -0.54 | -0.54 | -0.36 | -0.63 |
| | (0.13) | (0.14) | (0.13) | (0.12) |
| Cartel Implementation | -0.18 | -0.18 | -0.03 | -0.08 |
| | (0.08) | (0.10) | (0.12) | (0.05) |
| Announcement | -0.01 | | | |
| | (0.06) | | | |
| Implementation | -0.12 | | | |
| | (0.08) | | | |
| | | | | |
| Observations | 17,155 | 17,155 | 16,955 | 16,955 |
| R-squared | 0.29 | 0.82 | 0.83 | 0.84 |
| Unit FE | NO | YES | YES | YES |
| Date FE | NO | YES | N/A | N/A |
| Date x Technology FE | NO | NO | YES | NO |
| Date x Region FE | NO | NO | NO | YES |
| Forward Contracts | NO | NO | YES | YES |

Robust standard errors in parentheses

Table 4.2: Difference-in-difference estimates

Note: The table presents the estimation results of the difference in difference model proposed in equation 4.1 in sub-section 4.3.3 using the logarithm of the bid as the dependent variable. In columns 3-4 we further control for forward contracts over total capacity and alternatively for Date × Technology FE or for Date × Region FE. Regions are Atlantic, North-West, Central, and South-West. Robust s.e. clustered by unit and date in parenthesis.

### 4.3.4 Announcement date and threats of enforcement

The sudden relative decline in bids for units assigned to the cartel immediately after the *announcement* date is thus consistent with a shock to members' perceptions about their ability to sustain collusive behaviour in the future. As noted above, it is not essential

[102]Similar results are found when we use alternative specifications from Table 4.2 or when we use margins as a dependent variable (see Figure 4.17).

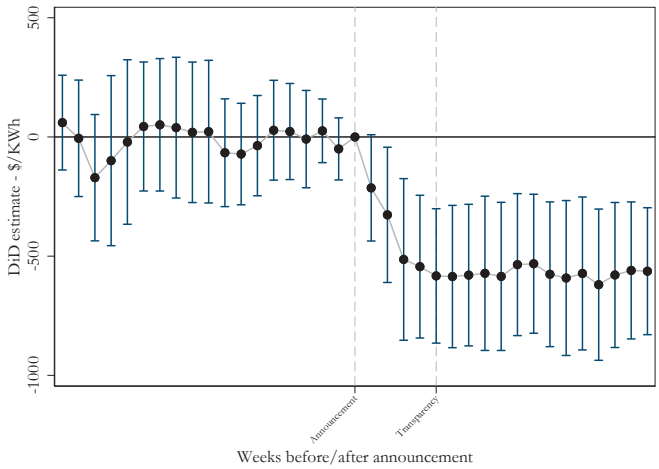



Figure 4.5: Event study representation of the differences-in-differences model.

Note: The figure presents event study estimates using bid as the dependent variable. We performed a two-way fixed effects model including a specific treatment effect for each week of the period studied. Robust s.e. are clustered by unit and date. The x-axis represents weeks around the policy announcement. The y-axis reports the estimates using the week of the announcement as baseline. Dots and bars represent point estimates and 95% confidence intervals. The dotted line labeled as "Announcement" represents the week of the announcement of the transparency policy. The dotted line labeled as "Transparency" represents the week of the implementation of the transparency policy.

for the logic of the test that (all) members exactly anticipated the regulatory change eventually put in place. For example, the *announcement* date could have signaled to market participants a future tightening of enforcement or regulators' willingness to act to uncover and prosecute collusive behaviour. Evidence from two sets of inspections –before and after the announcement– however suggests that the threat of enforcement is unlikely to explain the differential reaction to the announcement.

On January 20th i.e., after the announcement, the Supervisory Authority of Public Services (SSPD) conducted unannounced *in-situ* inspections to the four biggest electricity generation companies: EMGESA, ISAGEN, EPM and EPSA. The inspections aimed to find information related to potential collusive practices. Figure 4.6 extends the event-study specification in Figure 4.5 adding the interactions between dummies for leads and lags relative to the *inspection* date for inspected firms. Two patterns emerge. First, the results for the cartel units are virtually unchanged. Furthermore, the bulk of the differential drop in bids for cartel units happens *before* the inspection date. Second, after the inspection, inspected firms do not change much their bids. The point estimates are negative but small and not statistically different from zero, suggesting that a tightening of enforcement is unlikely to explain the differential drop in bids.

A potential concern in interpreting results from inspections that occurred *after* the announcement date is that the announcement might have already signaled an increase in the likelihood of tightening enforcement and that, once the cartel had collapsed, no further reaction should be expected. We can however use a separate episode of inspections that occurred *before* the announcement date to gain further insights into whether the threat of enforcement is likely to be driving the reaction that followed the announcement. On 5th December 2008, SSPD conducted a separate surveillance episode.[103] This surveillance action included three firms with units classified in the cartel. We thus replicate the event study including an event interaction for this surveillance action, split between cartel and non-cartel units. Figure 4.7 shows the results. After including the surveillance actions of the SSPD as control variables the effect of the announcement of the transparency policy remains economic and statistically significant. Furthermore, neither cartel nor non-cartel firms seem to have modified their bidding behavior following the December surveillance action. This suggests that firms might have not perceived enforcement to be a significant threat.

[103]The SSPD called in to its headquarters a number of firms (MERILECTRICA, TERMOEMCALI, TERMOTASAJERO, TERMOFLORES, TERMOCANDELARIA, GENSA) to discuss high bidding prices and other firms (EMGESA, EPSA, EPM, GECELCA, and ISAGEN) for bidding behavior and frequent stops in the operation of their units.

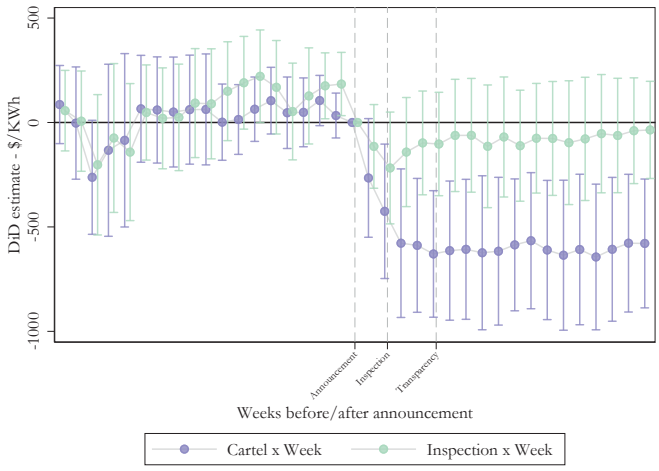



Figure 4.6: Event study representation of the differences-in-differences model.

Note: We investigate whether the threat of enforcement can explain the differential drop in bids using the inspection of January 20th (after the announcement). The figure presents the estimates using bid as the dependent variable and the event study of the inspection sites conducted on 20th January 2009 (inspected firms were EMGESA, ISAGEN, EPM and EPSA). We performed a two-way fixed effects model including a specific treatment effect for each week of the period studied. Robust s.e. are clustered by unit and date. The x-axis represents weeks around the policy announcement and inspection. The y-axis reports the estimates using the week of the announcement as baseline. Dots and bars represent point estimates and 95% confidence intervals. The dotted line labeled as "Announcement" represents the week of the announcement of the transparency policy. The dotted line labeled as "Transparency" represents the week of the implementation of the transparency policy. Finally, the dotted line labeled as "Inspection" represents the week of the inspection (20th January 2009).

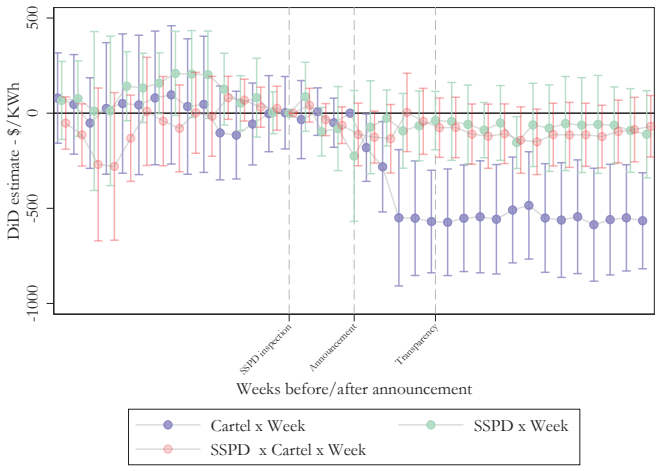



Figure 4.7: Event study representation of the differences-in-differences model SSPD Inspection.

Note: We investigate whether the threat of enforcement can explain the differential drop in bids using the inspection of December 5th (before the announcement). The figure presents the estimates using bid as the dependent variable and the event study of an inquiry action performed by the SSPD on 5th December 2008. We performed a two-way fixed effects model including a specific treatment effect for each week of the period studied. Robust s.e. are clustered by unit and date. The x-axis represents weeks around the policy announcement inspections. The y-axis reports the estimates using the week of the announcement or of the SSPD inspection as baseline. Dots and bars represent point estimates and 95% confidence intervals. The dotted line labeled as "Announcement" represents the week of the announcement of the transparency policy. The dotted line labeled as "Transparency" represents the week of the implementation of the transparency policy. Finally, the dotted line labeled as "SSPD inspection" represents the week of the inquiry action (5th December 2008).

### 4.3.5 Discussion

We have shown that cartel units decreased bids after the announcement of the policy and before the actual implementation of it. This not only provides suggestive evidence that there was a Cartel in the Colombian electricity market but that regulators' actions can reduce collusive behavior. Furthermore, Figures 4.6 and 4.7 suggest that this reaction was unlikely to stem entirely from anticipated threats of oversight and enforcement. It thus appears plausible that bidding behavior changed at least in part in anticipation of a transparency regime that would have made collusion harder to enforce.[104]

Collusive arrangements are complex: even when members can explicitly communicate, successful collusion requires a mutual understanding of many elements of the agreement (Harrington, 2008; Byrne and De Roos, 2019). It is thus highly implausible that, following the announcement, all units in the cartel immediately reacted in an *anticipatory* way to the uncertain prospect of a less transparent market regime in the future. More likely, many (if not most) units might have reacted in an *adaptive* way to the (unexpected) behaviour of other units. Consistent with this interpretation, Figure 4.4 shows that following the announcement, units reduced their bids in different waves: some units reacted before others. Interestingly, the first units decreasing prices were the Cartagena units belonging to EMGESA – the largest firm among the collusive firms. This is potentially consistent with evidence from other contexts in which larger firms are more sophisticated bidders (Hortacsu et al., 2019) and/or tend to take on the role of leaders that coordinate pricing (see Byrne and De Roos, 2019, for an example), as in basing points pricing systems common in, e.g., the cement industry.

In sum, the test of collusion does not assume – and the evidence certainly does not imply – that all of the units in the cartel were fully anticipatory. Instead, some of them (the ones that reacted first) can have anticipatory reactions and the rest adaptive reactions. While we are not aware of empirical analyses that try to test for cartel behaviour that distinguish between these two different types of behaviour, the experimental literature testing repeated game models in the lab, see., e.g., Dal Bo (2005) and Dal Bo and Frechette (2018), has found evidence for both. While observed sophistication in the lab is generally lower than assumed in models that rely on notions of subgame perfect equilibrium, some subjects do show the kind of sophistication consistent with anticipatory behaviour.[105]

[104] In theory, cartel members could device other ways to share information to police the cartel (McMillan, 1991). Leaving aside the fact that the new regulation explicitly forbids such information sharing, in the next Section we argue that the cartel colluded on the market for positive reconciliations. As positive reconciliations are relatively infrequent, devicing new information-sharing strategies was likely complicated.

[105] Bigoni et al. (2019) finds that participants in a lab experiment are sufficiently sophisticated to understand the impact of imperfect monitoring and the frequency of interaction on the sustainability of

### 4.3.6 Robustness to alternative definitions

Before providing more direct evidence of how firms in the cartel colluded, we investigate the robustness of our results.

Our cartel definition classifies units and not firms. We consider two alternative definitions based on firms' ownership of units: (1) we exclude from the cartel units that belong to firms that own other units not classified in the baseline cartel (*refined definition*), (2) we include all other units that belong to firms that have at least one unit in the baseline definition of the cartel (*extended definition*). Details are presented in Appendix 4.7.2.1, together with a placebo exercise that randomly determines which units belong to the cartel. The results are robust to the refined and extended definitions. The placebo exercise reveals that our findings are unlikely to be the result of chance.

Our baseline cartel definition is based on geographic location (Atlantic region) and production technology (thermal units). We consider additional criteria to classify units: private (vs public) ownership, forward contract positions, and bidding behaviour in 2008. We refine our baseline definition including these additional criteria progressively, building on our baseline definition. Appendix 4.7.2.2 presents the details of this exercise and finds results closely in line with our baseline findings.

## 4.4 Incentives to collude & inner functioning of the cartel

This Section provides evidence on cartel units' incentives to enter the agreement and on the inner working of the cartel. Subsection 4.4.1 shows that cartel units had incentives to enter a cartel to increase payments in the positive reconciliation market. We show that cartel units have costs high enough that they would not be able to earn the right to supply electricity through the ideal dispatch if they were to bid competitively. Given this, they maximize profits through the positive reconciliation market. We show that revenues and profits in the positive reconciliations market are inverted-U shaped in bids. Units thus benefit from a coordinated increase in bids.

Subsection 4.4.2 provides forensic evidence on the functioning of the cartel. We show that cartel units coordinated bids increase, particularly so when the probability of being called for positive reconciliation increases. Concretely, the cartel worked as follows: certain low-cost units win in the ideal dispatch auction and, from time to time, declare unavailability and generate positive reconciliations for other units. Given network restrictions, these positive reconciliations are disproportionately awarded to other cartel units

collusion.

that coordinated increases in bids to maximize revenue from these positive reconciliations. We complement this analysis using data from the minutes of the meetings of the Association of Generating Units (CNO in Spanish). We find that prior to the reform, but not after, cartel units were sending more staff involved in setting bids (instead of personnel dealing with engineering problems) to these meetings relative to non-cartel firms. This strategic behaviour hints at the possibility that these meetings might have been used to communicate about bidding strategies. Finally, we confirm that, after the announcement of the regulatory change, profits from the reconciliation market (but not from the ideal dispatch) decreased relatively more for cartel units than for other units.

### 4.4.1 Incentives to collude

Figure 4.8 compares the distribution of calculated marginal costs and average spot price for cartel units and other units separately. The average marginal cost of the units in the collusive agreement is larger than the average spot price. This contrasts with the units that are not in the collusive agreement. Given their higher marginal costs, cartel units try to make profits in the only remaining possible way: the positive reconciliations market.[106]

We check that the positive reconciliation market features the usual price-setting trade-off: higher prices increase margins, but reduce quantity (in this case, the likelihood of being called for a positive reconciliation). In such cases, firms benefit from coordinated price increases. Using the same controls of Table 4.2, Table 4.3 confirms that the revenue and profits from positive reconciliations are indeed inverted-U shaped in the submitted bid: If the bid is very low, the unit is not called for positive reconciliations as it would be allocated the right to produce in the ideal dispatch. However, when the bid increases, the potential payment and likelihood of being called for positive reconciliation increases. This is however true only up to a certain point. When the bid is too high, the unit is unlikely to be called in for positive reconciliation. This descriptive evidence should be interpreted cautiously: Bidding behaviour is endogenous and – as we shall see momentarily discuss – responds to the anticipation of positive reconciliations.

### 4.4.2 Inner working of the cartel

We first check that units strategically increase bids when they anticipate a higher likelihood of being called for a positive reconciliation. In the positive reconciliation market, the price paid to the unit is equal to the submitted bid (at least up to a certain maximum allowed price). Note that this incentive applies to both cartel and non-cartel units.

[106] As a sanity check, Table 4.8 shows that under different scenarios cartel units obtain higher profits in the positive reconciliation market than in a counterfactual case in which they bid their marginal costs to win in the ideal dispatch.

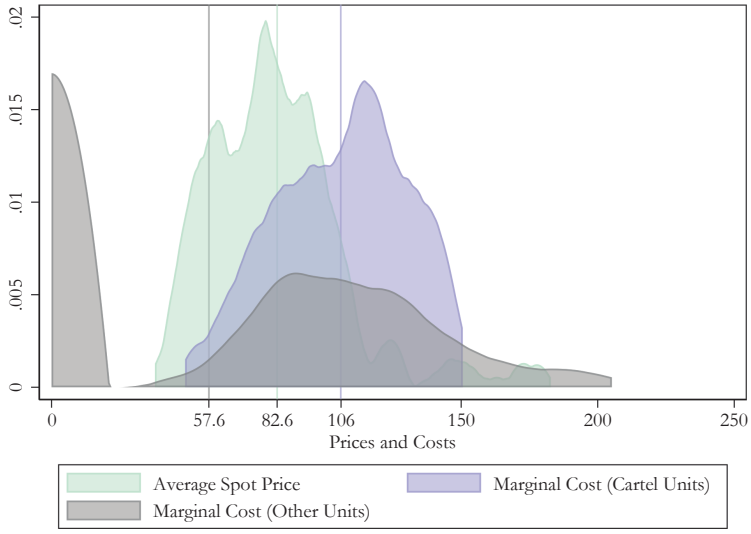



Figure 4.8: Competitiveness of the cartel and non-cartel units.

Note: The figure presents the kernel densities of the average daily spot price (green shaded density), marginal cost of non-cartel units (grey shaded density) and cartel units (purple shaded density) for the second semester in the year 2008. The grey vertical line indicates the mean marginal cost for non-cartel units, the green vertical line denotes the average daily spot price and finally, the purple vertical line denotes the average marginal cost for cartel units.

We investigate the relationship between the bid submitted at time $t-1$ for the auction of day $t$ with the expected (in $t-1$) likelihood of being awarded a positive reconciliation in $t$. This likelihood is not directly observed and must be proxied with actual positive reconciliations (in $t-1$). However, those are endogenous to bidding behaviour. We, therefore, need an instrument for the probability of a positive reconciliation for unit $i$ at date $t-1$.

We use *security contingencies* as an instrument. Security contingencies provide us with an observable, unit-day level varying measure of exogenous shocks to the transmission network that increases the likelihood of positive reconciliations. Specifically, when contingent restrictions to the network occur, certain units might be asked to produce security contingencies – small amounts of electricity to help the transmission system recover stability and compensate for overcharges. Security contingencies are exclusively based on engineering criteria: units are called in depending on exogenous shocks to the transmission network and independently of their bids and outcomes in the ideal dispatch. The exclusion restriction is thus likely to be satisfied.

| VARIABLES | (1) LnRevenues | (2) LnRevenues | (3) LnRevenues | (4) LnProfits | (5) LnProfits | (6) LnProfits |
|---|---|---|---|---|---|---|
| Ln(Bid) | 13.41 | 13.16 | 14.09 | 16.68 | 16.52 | 18.36 |
| | (2.96) | (2.95) | (2.57) | (3.73) | (3.67) | (2.76) |
| Ln(Bid) 2 | -1.06 | -1.05 | -1.13 | -1.29 | -1.28 | -1.43 |
| | (0.25) | (0.25) | (0.22) | (0.31) | (0.30) | (0.24) |
| Observations | 991 | 991 | 788 | 907 | 907 | 700 |
| R-squared | 0.67 | 0.68 | 0.72 | 0.68 | 0.69 | 0.72 |
| Unit FE | YES | YES | YES | YES | YES | YES |
| Date FE | YES | N/A | N/A | YES | N/A | N/A |
| Date x Technology FE | NO | YES | NO | NO | YES | NO |
| Date x Region FE | NO | NO | YES | NO | NO | YES |
| Forward Contracts | NO | YES | YES | NO | YES | YES |

Robust standard errors in parentheses

Table 4.3: Revenues and Profits inverted-U shaped in the submitted bid

Note: All the columns control for unit fixed effects. Columns (1) and (4) include additional controls for Date Fixed Effects. Columns (2) and (5) control for Date x Technology Fixed Effects and the level of forward contracts. Columns (3) and (6) control for Date x Region Fixed Effects and level of forward contracts. Robust s.e. clustered by unit and date in parenthesis.

Table 4.4 shows that units increase their bids when they have a positive reconciliation in the previous period. Column 1 presents the OLS estimate which is negative but not significant. The OLS estimate could be either upward or downward biased as a higher bid can either increase (the unit is less likely to win the ideal dispatch) or decrease (the unit, if eligible, is less likely to be called in) the likelihood of being awarded a positive reconciliation. Column 2 reports a strong first stage (F-stat 25.37): Conditional on unit and date fixed effects, shocks to the infrastructure significantly increase the probability that the unit is awarded a positive reconciliation. Column 3 reports the second stage and finds a large, and statistically significant, increase in bids for units that anticipate being more likely to be awarded positive reconciliations. We can also directly regress our outcome variable on the instrument to estimate a sort of Intention To Treat effect (what is sometimes referred to as the 'the reduced form' regression in IV settings). Indeed, after controlling for unit and time fixed-effects, we still find a positive and significant effect if we regress bids on security contingencies (column 4).[107]

We now show that cartel units coordinated to increase bids precisely when other cartel units generated positive reconciliations by winning the ideal dispatch and then declaring unavailable. For this coordination to happen, two conditions are necessary. First, it must be the case that the cartel comprises (specialized) units that win in the ideal dispatch,

[107]Moreover, we do not need the exclusion restriction to hold in order to give a causal interpretation to the reduced form regression (as long as the instrument is as good as randomly assigned), which might make the ITT estimate easier to interpret.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Ln(Bid) | Probability Pos. Rec. (t-1) | Ln(Bid) | Ln(Bid) |
| Probability Pos. Rec. $(t-1)$ | -0.199 | | 0.620 | |
| | (0.130) | | (0.168) | |
| Security Contingencies $(t-1)$ | | 0.113 | | 0.070 |
| | | (0.0225) | | (0.018) |
| Observations | 17,087 | 17,087 | 17,087 | 17,087 |
| R-squared | 0.838 | 0.539 | -0.135 | 0.839 |
| Unit F.E. | YES | YES | YES | YES |
| Date F.E. | YES | YES | YES | YES |
| Sample | 2008 | 2008 | 2008 | 2008 |
| Estimation | OLS | First Stage | Second Stage | Reduced form |
| Kleibergen-Paap F | - | 25.369 | - | - |

Table 4.4: Positive Reconciliations and Electric Network Contingencies

Note: The table presents the instrumental variables regression of the logarithm of the bid price on the first lag of the probability of positive reconciliation using observations from the year 2008. The first column presents the results of the OLS estimates. The second column presents the first stage of the IV estimation. We use the security contingencies in the transmission system as instruments of the lag of the probability of positive reconciliation. The coefficient estimate of this column is multiplied by 10.000 to facilitate interpretation. The last column presents the second stage of the IV estimation. All the columns control by Unit and Date fixed effects. The probability of positive reconciliation in day $t$ for unit $i$ is computed as the mean across the 24 hourly dummies that equal one if unit $i$ got a positive reconciliation in hour $h$ in day $t$. We then use its lagged value as this is known at the time of submitting bids. Robust s.e. clustered by unit in parenthesis.

at least sometimes. The top-left panel in Figure 4.9 shows that during the collusive period, some of the units in the cartel submit relatively low bids, and are sometimes awarded production in the ideal dispatch. Importantly, the average bid dropped after the announcement of the policy both for high and low-price cartel units, but the decrease in percentage terms was higher for high-price units. The top-right panel reports the likelihood that a unit declares unavailable upon winning in the ideal dispatch. This is larger for cartel units than non-cartel units. Finally, the bottom of the figure shows that the probability that high-price cartel units receive positive reconciliations when low-price cartel units win is much larger than when low-price no cartel units win. In sum, this provides suggestive evidence of coordination.

We now show that units are indeed able to coordinate. To conduct our test, we would ideally know network restrictions that make it more likely that a given unit $i$ receives a positive reconciliation when unit $j$ declares unavailable. This would allow us to test whether unit $i$ increases bids precisely when unit $j$ ends up declaring unavailable. Unfortunately, we do not observe the underlying electricity grid and we thus proxy these

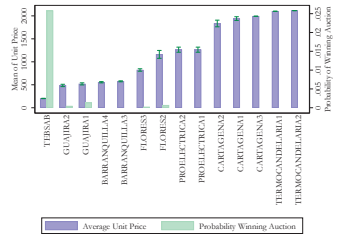
relationships between units relying on observed behaviour. For each unit $i$ we identify "friends", i.e., units that are more likely to get a positive reconciliation when unit $i$ has a negative reconciliation. For each unit $i$, we rank "friends" by the probability of receiving positive reconciliations when unit $i$ declares unavailable. We focus on observations 6 months before and 6 months after the announcement date.

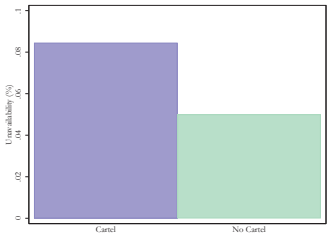
We test whether the average bid of $i$'s friends increases when unit $i$ declares (at least partially) unavailable. While we are unable to separate negative reconciliations that arise from strategic considerations from those that arise due to exogenous shocks (either to production or to transmission), a striking pattern emerges. Figure 4.10 shows that cartel units before the reform coordinated higher bids with declared unavailabilities of their "friends". Interestingly, this coordination only appears for cartel units and ceases once the cartel unravels. This provides suggestive evidence that this coordination was part of how the cartel functioned.[108,109]

While the cartel might have also adopted additional collusive practices, our analysis suggests that one way through which the cartel functioned was to coordinate (strategically) declared unavailability and bids so as to increase profits for its members in the positive reconciliations market. Of course, units in the cartel did not engage in this behaviour too frequently, presumably to avoid detection from the regulator (to declare unavailability, units need to submit a report and, if they do it too often, they risk being investigated). Thus, we do not interpret this mechanism as the central one for the profitability of the cartel, but rather as an occasional phenomenon to increase profits.

[108]Results are robust to changes in the number of friends considered, the baseline period used to define friends, and the definition of the explanatory variable. See Figure 4.18. While the geographic clustering of cartel units correlates with the set of identified friends, this, *per se*, does not explain the time pattern in observed coordination. We nevertheless consider whether units clustered in the South-West part of Colombia –which as the Atlantic units are relatively isolated (see Figure 4.3) – display a similar coordinated behaviour as a placebo. The estimates for the South-West area are zero throughout the years.

[109]We also explore an alternative exercise in which "friends" are units that belong to the same firm. Under this definition, we find no evidence of coordination between unavailability and bids for both cartel and non-cartel units. This suggests that bid coordination likely happened across and not within firms.



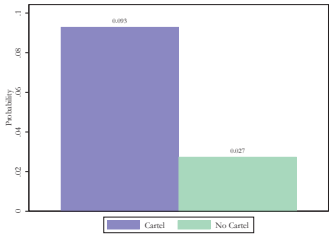




Figure 4.9: Inner Working of the Cartel

Note: The top left figure shows the average unit prices for the cartel units and their probability of winning the auction. The top right figure shows the fraction of unavailabilities over the total number of times that they have won in the auction for high and low-bid cartel and non-cartel firms. High-bid cartel units are those for which their average bid in the second semester of 2008 was above the median of all of the average bids. Low bids are those below the median. The bottom figure shows the probability that high-price cartel units receive positive reconciliations when low-price cartel units win, or low price no cartel units win. All of the graphs only use data for the second semester of 2008.

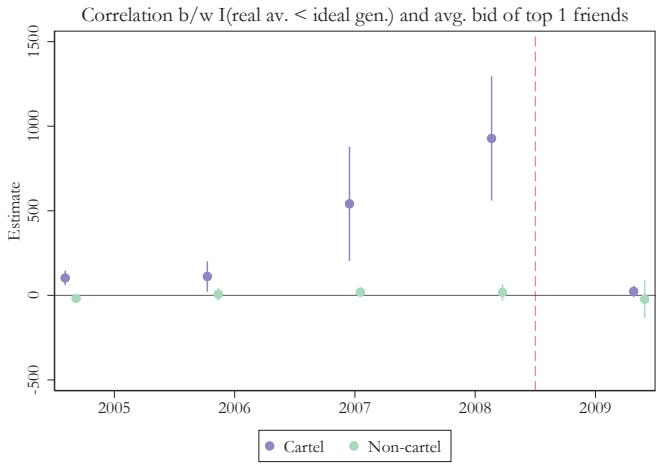



Figure 4.10: Baseline coordinated bids analysis.

Note: We investigate whether cartel units coordinated their bids before the policy change. The figure presents estimates from regressions where the outcome variable is the average bid of the friends of unit $i$ and the explanatory variables is an indicator for unit $i$ declaring a level of real availability below the ideal generation quantity it was awarded. We only include in the explanatory dummy the 75% cases where the difference between real availability and ideal generation is the largest. We run separate regressions for the two groups (cartel, non-cartel) and repeat for the years 2005 to 2009. The estimates for 2009 need to be interpreted cautiously. Data on real availability is missing for 63% of cartel observations and for 6% of non-cartel observations in 2009.

### 4.4.3 Suggestive evidence of communication

We complement our analysis using hand-collected data from the minutes of the meetings of the Association of Generating Units (CNO in Spanish) (see Appendix 4.7.1 for details). This association holds meetings to solve technical difficulties and constraints to the system. The association's explicit rule was that agents from the commercial area (i.e., likely involved in setting bids) cannot attend the meetings. However, as we shall see, the rule was not enforced.

We downloaded the minutes of all the meetings in the second semester of 2008 and the first semester of 2009. Meetings report attendees and the firm that they belong to. Within a DID framework, we test if there was any differential change in attendance between the cartel and non-cartel units before and after the policy change. Since firms send only one attendee per meeting (if any at all), we focus on two dependent variables: a dummy that takes the value of 1 if the firm sends someone to the meeting and a dummy that takes the value of 1 if the firm sends someone from the commercial area to the meeting. We also explore the composition of attendees conditional on sending someone to the meeting. We categorize participants as working in the commercial area if, at the time of the meeting, their CV (accessed through websites such as LinkedIn, newspapers and industry publications) reports that the attendee worked in the commercial area, proxied with job titles mentioning the words *commercial* or *marketing*.[110]

There are 97 attendees in 18 different meetings for a total of 435 attendee-meeting observations. We were able to assign a job title to 63% of these 435 observations. The data shows that before the reform, for cartel firms, the probability that and attendee is from the commercial area conditional on the firm sending someone to the meeting is 71%. That probability is only 19% for the other firms. We explore DID specifications that control for firm and meeting fixed effects focusing on the interaction between meetings in 2009 (i.e., after the reform) and firms in the cartel.[111]

Table 4.5 reports the results. First, column (1) shows that, due to the reform, cartel firms didn't change attendance behaviour. However, column (2) shows that the composition of the attendees changed: after the reform, firms in the cartel are *less* likely to send someone from the commercial area.[112] Finally, column (3) confirms that, conditional on

[110]Results are robust if we drop workers with job titles related to marketing.

[111]Note that attendees can only be assigned to firms, not units. For this exercise, our definition of cartel must be at the firm level. A firm belongs to the cartel if at least one unit belongs to the baseline definition of the cartel. Unfortunately, it is not possible to obtain results using the refined definition in which a firm is classified in the cartel if all its units are. Firms that are so classified in the cartel seldom send attendees to the meetings and we match the occupation of only one attendee in 2008.

[112]The share of delegates from the commercial area dropped from 17 to 0% for cartel firms, but in fact it increased from 6 to 18% for non-cartel ones. The latter increase is mainly due to two firms with one hydro unit each. Especially for small (i.e. with less resources) or hydro (i.e. facing larger uncertainty)

sending someone to the meetings, the probability of sending someone from the commercial area decreased for cartel firms relative to the others.

In sum, while this does not prove that cartel units explicitly communicated to coordinate bidding behaviour around the timing of the meetings, the evidence points to strategic behaviour in attendance. Similar evidence could presumably be used to evaluate the possibility of prosecution in other cases.

| | (1) | (2) | (3) |
|---|---|---|---|
| VARIABLES | Someone | Someone Commercial | Cond. Probability Commercial |
| Cartel x 2009 | -0.001 | -0.293 | -0.817 |
| | (0.220) | (0.128) | (0.068) |
| Observations | 480 | 480 | 170 |
| R-squared | 0.519 | 0.425 | 0.818 |
| Firm FE | YES | YES | YES |
| Meeting FE | YES | YES | YES |

Robust standard errors in parentheses

Table 4.5: Meetings Minutes Evidence

Note: The table presents the relationship between having any worker or someone from the commercial area on an interaction term Cartel×2009. Cartel takes the value of 1 for the units classified in the baseline collusive agreement and 0 otherwise. A firm belongs to the cartel if at least one unit belongs to the baseline definition of the cartel. There are three dependent variables: 1. Sending someone to the meetings, 2. Sending someone from the commercial area and 3. Sending someone from the commercial area conditional to sending someone to the meetings. Robust s.e. clustered by unit and date in parenthesis.

### 4.4.4 Lower profits from positive reconciliations after the end of the cartel

Finally, the hypothesis of an implicit agreement in the bidding scheme implies that after the break of the agreement, the profits of the cartel units should decrease. As a sanity check, we, therefore, revisit our baseline specifications from Section 4.3.3 and consider as dependent variables a dummy for receiving positive reconciliations, profits from positive reconciliations, and total profits.

Table 4.6 shows that, while the likelihood of receiving positive reconciliations was unaffected, the profits from positive reconciliations as well as the total profits sharply

units, the announcement might lead to sending more people for two reasons. First, because they want to collect more information about what the policy change will effectively be. Second, because the 90-day transparency rule might make it more difficult to price correctly, such that these firms might try to collect more information about pricing at those meetings. This is consistent with the observation that the two aforementioned firms did not send anyone from the commercial area in the 19 meetings before the policy change, but suddenly started sending people afterwards.

decreased for the collusive group after the announcement date. (Despite the fall in profits, no firm exited the market after the collapse of the cartel).

We further explore whether profits were differently affected by the announcement depending on the costs, or the role, of units in the cartel. The underlying idea is that some units might have been worse off colluding rather than competing, and therefore transfers within the cartel could have been necessary to sustain it. We classify units according to proxies for their ability to compete or for their role in the cartel. Table 4.9 shows that total profits fell for all units, and slightly more for high costs units that would unlikely be able to increase profits in the ideal generation market. Instead, profits from positive reconciliations fell more for low-cost units which again suggests that they are now focusing on the ideal generation market. Transfers might thus *not* have been needed to sustain the cartel in this case, as all units were better off colluding.[113]

| | (1) | (2) | (3) |
|---|---|---|---|
| VARIABLES | Dummy for PR | Profits from PR | Total profits |
| Cartel Post | 0.02 | -135.88 | -74.29 |
| | (0.05) | (62.03) | (21.80) |
| Observations | 17,155 | 6,725 | 17,155 |
| R-squared | 0.43 | 0.68 | 0.79 |
| Unit FE | YES | YES | YES |
| Date FE | YES | YES | YES |

Robust standard errors in parentheses

Table 4.6: Effects of announcement on profits

Note: The table presents differences in differences estimates for various outcomes controlling for unit and time-fixed effects, where the Post period refers to the period after the policy announcement. Column 1 presents the estimates for the probability of receiving positive reconciliations. Column 2 presents the estimates for the profits from positive reconciliations, conditional on receiving some positive reconciliations. Column 3 presents the estimates for the total profits (unconditional). Robust s.e. clustered by unit and date in parenthesis.

## 4.5 Incentive to deviate and cost of the cartel

In this section, we first show that cartel units could increase short-run profits by lowering bids and deviating from the collusive agreement. Second, we show that such deviations are not profitable in the long-run under the old transparency rule, but they are profitable

[113]In columns 1-2, 'high' units are those with average marginal cost in the second half of 2008 above the median, and 'low' otherwise. In columns 3-4, 'high' units are those with average bids above the median, and 'low' otherwise. In columns 5-6, 'high' units are those with an average amount of negative reconciliations *below* the median, and 'low' otherwise.

under the new 90 days transparency rule. Although, as noted in Section 4.4, our test does not require nor proves that the cartel unravelled because firms anticipated that collusion would become unsustainable following the change in transparency rules, these estimates show that such an interpretation is at least potentially consistent with the economic environment under consideration. Finally, we provide a back-of-the-envelope estimate of the cost of the cartel for consumers.

### 4.5.1 Modelling choices

Before presenting the results described above, we introduce here the modelling choices behind them. First, we model how the positive reconciliations that a unit gets depend on the submitted bid, as our evidence suggests that the cartel was operating in the market for positive reconciliations. Together with the engineering measure of production costs, this allows to investigate counterfactual profits that result from alternative bids. In turn, this allows to check whether observed bids maximize static profit and, if not, what would be the optimal bid for a unit that wants to unilaterally deviate from the collusive agreement.[114] Second, in order to study the sustainability of the cartel after the policy change, we need to construct what would have been the collusive bids for the counterfactual scenario where the cartel did not collapse.

When a transmission constraint makes the allocation of the ideal dispatch unfeasible, eligible units might be called for positive reconciliations. In case two or more units are eligible, the system regulator selects the one with the lowest bid. Therefore, we model the amount of positive reconciliations that a unit receive as a function of (i) time effects that are common across units, which reflect aggregate demand and daily transmission constraints, (ii) unit time-invariant characteristics, which capture the unit's location in the transmission network and influence eligibility, (iii) the rank of the submitted bid, which determines who produces when more units are eligible (in the spirit of Porter and Zona, 1993). To flexibly accommodate the relation between these covariates and quantities we use a two-stage model. First, to model the probability of receiving positive reconciliations, we regress a dummy for having a positive reconciliation in day $t$ on the rank of the bid for the same day with respect to its competitors ($Rank_{it}$), its squared value ($Rank_{it}^2$), unit and time fixed effects ($\gamma_i, \delta_t$). For example, $Rank_{it} = 1$ if unit $i$ submitted the lowest bid for day $t$. Since our goal is to make predictions, we estimate the model with maximum likelihood assuming the errors follow a logistic distribution, which

[114] We think of the game for thermal units as a repeated static game, in which today's action do not affect tomorrow's state variable. On the other hand, the game is inherently dynamic for hydro units, which need to decide how much water to consume today at the expense of the water available tomorrow. The reason to focus on unit-level maximization is that (i) each unit submit its own bid and (ii) in a competitive environment the result should be the same.

implies modelling probabilities as in Equation (4.2). Second, to model quantities, we regress the natural logarithm of the awarded (positive) amount of positive reconciliation on the same covariates as in the first step, as shown in Equation (4.3). We estimate the second model with OLS. Given these two sets of estimates, we can predict the expected amount of positive reconciliations corresponding to any possible bid submitted by a given unit in a given day.

$$\Pr[Pos.Rec._{.it} > 0] = \frac{\exp\left(\beta_1 Rank_{it} + \beta_2 Rank_{it}^2 + \gamma_i + \delta_t\right)}{1 + \exp\left(\beta_1 Rank_{it} + \beta_2 Rank_{it}^2 + \gamma_i + \delta_t\right)} \tag{4.2}$$

$$\ln\left(Pos.Rec._{.it}\right) = \tilde{\beta}_1 Rank_{it} + \tilde{\beta}_2 Rank_{it}^2 + \tilde{\gamma}_i + \tilde{\delta}_t + \varepsilon_{it} \text{ if } Pos.Rec._{.it} > 0 \tag{4.3}$$

While competitive bids should be set to maximize (static) profits, we are agnostic with respect to the strategy used to determine collusive bids (e.g. whether cartel units maximize joint profits or not). Therefore, we simply model bids as a function of market fundamentals, similarly to Pesendorfer (2000): We regress observed bids submitted in $t-1$ for the auction of day $t$, $b_{it}$, on unit fixed effects, $\bar{\gamma}_i$, production costs for day $t$, $C_{it}$, the logarithm of the total amount of positive reconciliation in the previous day, $\ln\left(Pos.Rec._{.t-1}\right)$, and the logarithm of the total amount of ideal generation in $t$, $\ln\left(Ideal_t\right)$. We use aggregate quantities as these are exogenous with respect to bids, while individual quantities are endogenous. As for the timing, we use the most recent available information at the time of submitting bids.

$$b_{it} = \bar{\beta}_1 C_{it} + \bar{\beta}_2 \ln\left(Pos.Rec._{.t-1}\right) + \bar{\beta}_3 \ln\left(Ideal_t\right) + \bar{\gamma}_i + \nu_{it} \tag{4.4}$$

The simple mapping we use to model positive reconciliations directly reflects the institutional setting and the rules used to award positive reconciliations. In turn, the simple model for bids reflects the relevant economic state variables that (thermal) units face, but it avoids taking a stand with respect to the objective function (and constraints) that cartel units maximize when colluding. We evaluate to goodness of fit of these modelling choices in the next subsections.

### 4.5.2 Bidding strategy

Following the game-theoretic framework of Chassang and Ortner (2023), the existence of a cartel involves departures from a static Nash equilibrium for its members, which implies a short-run incentive to deviate from the cartel. The sustainability of a cartel depends on whether such unilateral deviations are incentive compatible or not, that is whether the gain from short-run deviations compensates for the loss of future profits from collusion.

Our hypotheses are that cartel units set bids to maximize their individual static profits after the policy change but not before, while comparable non-cartel units always maximize static profits. In particular, cartel bids should be larger than static profits-maximizing bids before the policy change. We test these hypotheses with the following three-step procedure

In the first step, we identify a suitable comparison group of non-cartel units using the following criteria. (i) Thermal units, because all cartel units are thermal. Furthermore, the assumption that units maximize static profits is realistic for thermal but not for hydro units (Fioretti et al., 2024). (ii) Private units, because publicly owned units might not be profit maximizers (Barros and Modesto, 1999). (iii) Units that are not owned by a firm that also owns cartel units, because we want to limit the possibility of considering units that are actually in the cartel or pursuing different goals, as in the robustness exercise on the cartel definition presented in Section 4.3.3. These criteria leave us with a comparison group consisting of five non-cartel units.[115] We focus on a one year period: six months before the policy change and the six months after the policy change. For these two periods, we compare observed bids with an estimate of static profits maximizing bids.

Second, we estimate how positive reconciliations depend on submitted bids using the models in Equations (4.2) and (4.3). Given that the geographical position of a unit is crucial in determining who gets a positive reconciliation (due to transmission network constraints) and since cartel and non-cartel units are located in different regions, we estimate the two models separately for cartel and non-cartel units. In particular, for cartel units the rank of bids is computed with respect to all the other cartel units, while for non-cartel units the rank of bids is computed with respect to all the other non-cartel units. Figure 4.22 presents in-sample predictions from this estimation procedure versus observed quantities. We further compare the distribution of our predictions and of observed quantities in Figure 4.23. Both diagnostic figures suggest that our model is able to replicate the amount of positive reconciliations awarded.

In the last step, given the estimates obtained above, we study whether there exist profitable unilateral deviations. We simulate alternative bids for each unit in each day and compute the corresponding quantities of positive reconciliation conditional on other units' observed bids. We then compute counterfactual profits – using our engineering measure of production costs – and select the bid yielding the highest profits. At the unit-day level, we compute the ratio between the observed bid and the profit-maximizing bid and plot the density of this ratio separately for the two groups of units and the two periods of time. Figure 4.11 presents the results. Before the policy change, the distribution for

[115]The main results are robust to relaxing the second restriction (by also including public units in the control group) and/or the third restriction (by also including units owned by firms that also owns cartel units in the control group).

cartel units is bimodal and displays a peak at around four (significantly larger than one): cartel units could often increase profits by lowering bids. For non-cartel units, instead, the distribution is single-peaked with most of its mass closely around a ratio equal to one. After the policy change, instead, for both cartel and non-cartel units the density is centered around one, suggesting that both groups are now bidding competitively. A Kolmogorov-Smirnov test for the equality of the distributions of the ratio for cartel and non-cartel units rejects the null hypothesis of equality pre-reform (p-value = 0.00), but not post-reform (p-value = 0.62). In sum, cartel units appear to systematically deviate from static profit maximization before, but not after, the reform.

### 4.5.3 Dynamic enforcement constraints

Cartel units could increase short-run profits by deviating from the collusive bidding strategy. However, cartel sustainability relies on the fact that deviations are not profitable due to the future value of the collusive agreement. Similarly to Igami and Sugaya (2021), which however rely on a proven cartel, we check that the incentive to collude is positive for all cartel units under the old transparency rules, but negative for at least one unit under the new ones.

We assume that deviation of a unit from the collusive agreement triggers static Nash competition as soon as past bids are made public. Under the old transparency rule, this implies that a unit can unilaterally deviate for two days and undercut other cartel's bids to increase static profits. But, from the third day onward, cartel units would bid competitively. However, under the new transparency rule, a unit can unilaterally deviate for 90 days. We thus define daily profits for unit $i$ at time $t$ as $\pi_{it}^{j}$ under three alternative scenarios: $j$ = C (collusion), N (competition), D (optimal deviation from the collusive agreement). We assume that units hold static expectations and denote with $\beta$ the (common) daily discount factor. We define the Dynamic Enforcement Constraint (DEC) under the old (4.5) and new (4.6) transparency rule as follows:

$$\frac{1}{1-\beta}\pi_{it}^{C} - \frac{1-\beta^2}{1-\beta}\pi_{it}^{D} - \frac{\beta^2}{1-\beta}\pi_{it}^{N} > 0 \qquad \text{For all units} \tag{4.5}$$

$$\frac{1}{1-\beta}\pi_{it}^{C} - \frac{1-\beta^{90}}{1-\beta}\pi_{it}^{D} - \frac{\beta^{90}}{1-\beta}\pi_{it}^{N} < 0 \qquad \text{For at least one unit} \tag{4.6}$$

Our hypothesis is that (4.5) is satisfied for *all* cartel units for all periods $t$ before the policy change, while (4.6) is not satisfied for *at least* one cartel unit after the policy change. We empirically test this hypothesis by focusing on cartel units and constructing counterfactual bids and quantities for a one-year period around the policy change.[116] From August 2008

[116]We do not consider the unit Proelectrica 2 because it was never awarded positive reconciliations in the considered period of time. We thus focus on 13 of the 14 cartel units.

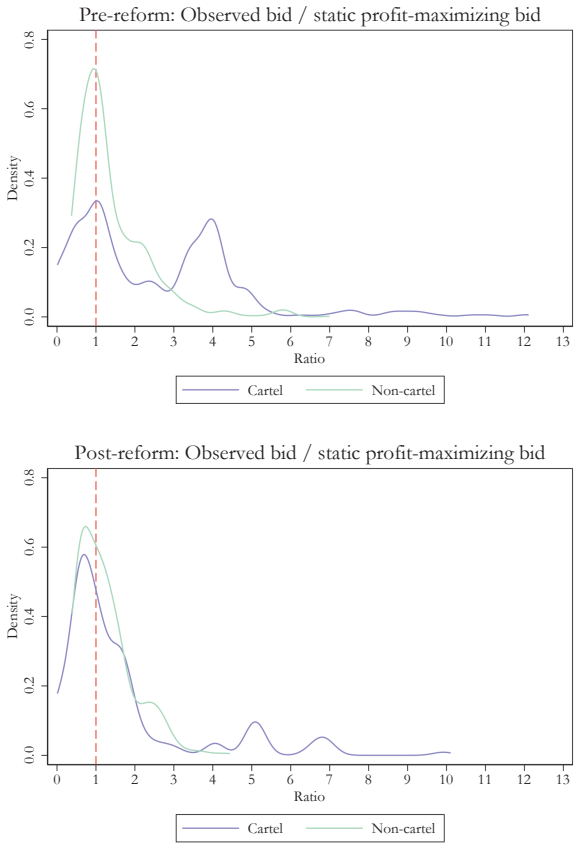



Figure 4.11: Density of the ratio between unit-daily observed and static profits maximizing bids for the cartel and non-cartel units over two different six-month periods.

Note: For cartel and non-cartel units, we simulate counterfactual bids and the corresponding amount of positive reconciliations and select the static profit-maximizing bids. We plot the density of the ratio between the observed bid and the profit-maximizing bid. The top (bottom) figure presents the density using data from the six months before (after) the policy change.

until the announcement of the new policy we observe bids and quantities under collusion and need to compute bids and quantities under competition and deviation from collusion, while from the policy implementation until June 2009 we observe bids and quantities under competition and need to compute bids and quantities under collusion and deviation from collusion. For the period between the announcement and the implementation we remain agnostic about the moment in which each unit moved from collusive to competitive bids, and we construct counterfactual variables for all the three scenarios.

First, we model how units set bids to construct counterfactual bids. Second, we model counterfactual quantities of positive reconciliations, which depend on bids. Third, given these bids and quantities and production costs, we compute profits for the counterfactual scenario of competition (collusion) before (after) the policy change. Fourth, we simulate alternative bids for each unit and the corresponding profits, assuming other units stick to the collusive bids, in order to define the optimal unilateral deviation (as we did in the previous Section). Finally, we compute the DEC for each unit.

First, we need counterfactual competitive bids $b_{it}^N$ for the pre-reform period and counterfactual collusive bids $b_{it}^C$ for the post-reform period. We model bids a function of market fundamentals as in Equation (4.4). We estimate this regression separately for the pre and post-reform periods and use these estimates to simulate counterfactual bids. Table 4.12 presents the results. As argued by Porter and Zona (1993) and Ishii (2009), our estimates suggest that the levels of bids do not necessarily reflect the underlying market fundamentals when units are colluding, but they do when units are competing. In order to assess the goodness of our method, in Figure 4.24 we plot in-sample predictions versus observed bids for the collusive and competitive period separately, and the respective distributions in Figure 4.25. Our model seems to replicate well how cartel units set bids under both scenarios.

Second, we need counterfactual competitive quantities $q_{it}^N$ for the pre-reform period and counterfactual collusive quantities $q_{it}^C$ for the post-reform period. However, to correct for differences in units' availability and transmission problems, we also simulate $q_{it}^C$ for the pre-reform period and $q_{it}^N$ for the post-reform period instead of relying on observed ones. In practice, we thus simulate both counterfactual quantities - collusion and competition - for all days and units. We model positive reconciliations according to Equations (4.2) and (4.3) (estimates in Table 4.13). We then predict $q_{it}^C$ using observed bids for the pre-reform period and using simulated bids for the post-reform period. We do the opposite for $q_{it}^N$. Figures 4.26 and 4.27 show the goodness of in-sample predictions. Again, the model seems to replicate fairly well how cartel bids translates into positive reconciliations under both scenarios. The comparison between Figure 4.25 and 4.27 also suggests that the collapse of the cartel changed the distribution of bids but not so much the distribution of awarded positive reconciliations, as expected.

Given simulated and observed bids, estimated costs, and simulated quantities, we predict profits $\pi_{it}^C$ and $\pi_{it}^N$ for all days and units. Finally, as in the previous Section, for each unit and day separately, we simulate profits for different possible values of a deviation bid $b_{it}^D$ (above production cost and below the collusive bid) and select the one yielding the highest profits $\pi_{it}^D$. Figure 4.28 presents the resulting average profits under the three different scenarios. By construction, for each unit, deviation yield the largest average profits, competition the lowest with the collusion payoff in the middle. The figure reveals significant variation in how much units stand to gain from collusion relative to competition. We average daily profits within a month to compute the incentive to collude in each month as described in equations (4.5) and (4.6). Results are presented in the top panel of Figure 4.12.

The solid lines in the top panel of Figure 4.12 report the *smallest* incentive to collude across cartel units over time, assuming a daily discount factor $\beta = 0.9996$.[117] The solid lines suggest that all cartel units were better off colluding until January 2009, but that afterwards the cartel became unsustainable as it was more profitable for at least one unit to deviate from the collusive agreement. As it happens, our estimates reveal that the DEC was unlikely to hold for two units after the reform (Termocandelaria 1 and 2). If we further assume that these two units optimally deviate and compute the incentive to collude for the remaining cartel units, our model indicates that four additional units would prefer to deviate (Cartagena 1 and 3, Flores 2 and 3), potentially starting a chain effect that would lead to the collapse of the cartel.[118]

The dotted lines in the top panel of Figure 4.12 further show that the data are consistent with a drop in the incentive to collude in January 2009 following the policy change, and not as a result of other differences between the pre- and post-reform periods.[119] Figure 4.29 further plots the smallest incentive to collude across cartel units for different values of the discount factors. In particular, it shows that when a unit can deviate for two days before being caught, collusion can be sustained for any reasonable discount factor.

[117]This corresponds to an annual rate $\beta^{365} = 0.86$. The lending interest rates in Colombia in 2008 and 2009 were respectively 17.2% and 13.0% according to the IMF, which correspond to discount rates of 0.85 (1/1.172) and 0.88 (1/1.130). As an additional robustness exercise, we repeat the calculations for slightly higher and lower values of the daily discount factor (0.9995 and 0.9997), corresponding to an interest rate of 20% and 11.6%. The shaded area in Figure 4.12 presents the lower and upper bounds of the smallest incentive to deviate using the different discount factors.

[118]In the data, we observe that the first units decreasing prices after the reform were the Cartagena units. Note that our calibration exercise only speaks for the economic incentives of a unit, conditional on everyone else sticking to the collusive agreement. Observed behaviour, however, depends on both first and second order believes about incentives (i.e. Cartagena units might deviate first even if their constraints are still satisfied if they believe that other units' constraints are not).

[119]This exercise suggests that a minimum disclosure delay of around 60 days would have been necessary to trigger a unilateral deviation for at least one cartel unit.

However, if a unit can deviate for 90 days before retaliation happens, collusion can be only sustained for values of the discount factor above 0.9.

### 4.5.4 Cost of the cartel

Our counterfactual estimates of bids and quantities allows to provide a back-of-the-envelope quantification of the additional cost consumers paid due to the high cost imposed by the cartel in the reconciliations market. We focus on the second semester of 2008 and compare the total price paid for positive reconciliations with the total price that would have been paid if cartel firms had behaved competitively. The former quantity is observed, while the latter is deduced from the counterfactual analysis.[120]

The cartel generated at least an additional cost of around 11 billion COP per month, which corresponds to an increase of around 12% with respect to the competitive scenario (see the bottom panel of Figure 4.12). Positive reconciliations account for approximately 10% of the electricity procured by the regulator, but since they are paid above the spot price this lead to an increase in overall costs of about 2.5%. Around 10 million households lived in Colombia in 2008. If all the energy allocated via positive reconciliations is bought by households, and assuming a full pass-through of the cost increase to consumers, the average household paid 1,100 COP in excess per month in the second semester of 2008 due to the collusive agreement (with many household living with less than a minimum wage of 461.500 COP).[121]

[120]The assumptions underlying the quantification of the cost of the cartel are discussed in Appendix 4.7.3.

[121]Ideally, we would explore the reduced form quantification of the costs of the cartel to downstream sectors, for instance in manufacturing. However, contextual confounders and data limitations prevent such analysis. In particular, the Colombian manufacturing Census is yearly, and in the second semester of 2009, El Nio adversely affected the production capacity of hydro-power units, increasing equilibrium prices.

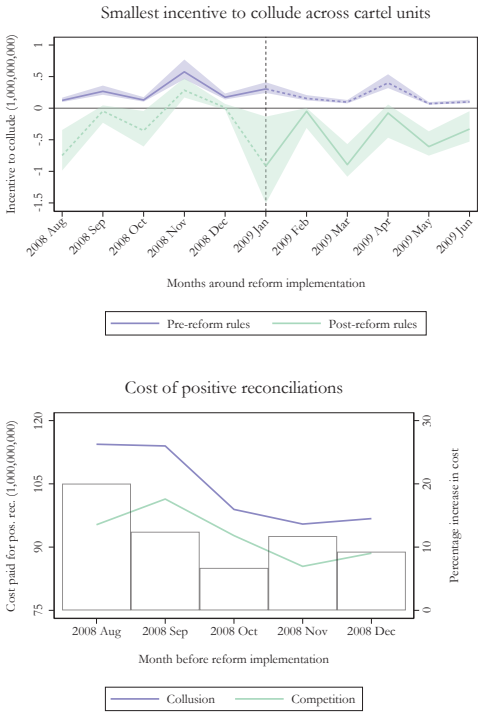



Figure 4.12: Smallest incentive to collude across cartel firms and cost of the cartel.

Note: The top figure presents the smallest incentive to collude across cartel firms. For each cartel unit, we compute the incentive to collude each day from August 2008 to June 2009 and then aggregate it into months. The purple line shows the smallest incentive to collude across cartel units assuming that a unit can unilaterally deviate for two days before triggering competition; for the the green line we assume that a unit can deviate for 90 days. Under the pre-reform rules we use a solid line in the pre-reform period and a dashed line in the post-reform; the opposite is true for post-reform rules. The incentive to collude is computed assuming a daily discount factor $\beta = 0.9996$ ($0.9996^{365} = 0.86$). The shaded area presents the boundaries of the result when the calculation is based on a daily discount factor of 0.9995 or 0.9997. The bottom figure presents the cost of the cartel for consumers. We multiply bids and amounts at the unit-day level and then sum over units. We then aggregate costs at the monthly level. The purple line (scale on the left axis) represents the total cost paid to cartel and non-cartel units for positive reconciliations in every month between August and December 2008. The green line (scale on the left axis) represents the counterfactual cost assuming cartel units were competing rather than colluding. The bars (scale on the right axis) present the percentage increase in the cost paid for positive reconciliations with respect to the competitive scenario.

## 4.6 Policy implications and conclusions

This paper identified collusion among a subset of firms in the Colombia wholesale energy market. Our test uncovers sudden changes in bidding behaviour after the announcement, but before the actual implementation, of a regulatory reform that reduced market transparency. Actions from the regulator, therefore, curbed collusion. Our evidence suggests that this reaction was unlikely to stem from anticipated threats of oversight and enforcement. It thus appears plausible that some firms changed their bidding behavior in anticipation of a transparency regime that would have made collusion harder to enforce. While this anticipatory response doesn't imply that subgame perfect equilibrium provides a tight description of this cartel, it does suggest that dynamic incentive compatibility constraints can be taken seriously by empirical researchers and policy-makers fighting collusion.

Our analysis has policy implications for market design – including energy markets – in developing countries. Distortions due to collusive practices in upstream sectors that provide inputs to many other sectors, such as energy, are particularly detrimental to aggregate welfare (Liu, 2019). The Colombia case provides a particularly interesting example. The country's energy sector was successfully reformed in the nineties and is generally considered one of the best-designed and regulated markets among developing countries (World Bank, 2019). We suspect collusive behaviour would be even more likely and create larger distortions in less well-designed energy markets.

In our context, the Colombian regulator lacked sufficient evidence to open targeted investigations and attempt prosecution. This induced the regulator to instead alter the market design in the hope of hindering (potential) collusive practices. Changes in market design, however, can be costly. For example, in our context, market transparency facilitates the efficient inter-temporal allocation of scarce water resources. The fact that at least some cartel members reacted in an anticipatory way (i.e., dynamic incentive compatibility constraints underpin collusive behaviour) raises the possibility that regulators might be able to strategically use *announcements* to induce behavioural responses and acquire sufficient evidence to open investigations and attempt prosecution (see Chassang and Ortner, 2023, for a discussion regarding the importance of acquiring minimum evidence to start investigations). A careful investigation of this possibility merits further theoretical and empirical scrutiny.

Our analysis hints at how market transparency affects firms' conduct and how a (particular) policy that limited public information might have reduced anti-competitive behavior. In our context, the policy had an effect because cartel members likely did not have other ways to credibly share information and police the agreement. The impact of market transparency on collusion in other contexts – including public procurement,

e-commerce, and agricultural markets – deserves further scrutiny. Digital technologies, for example, have the potential to increase sellers' visibility among buyers, reduce search costs and increase competition (Bai et al., 2020; Baldwin et al., 2021; Bergquist et al., 2023). Our evidence introduces a word of caution: increased transparency could backfire if it allows firms to detect and punish deviations from collusive agreements. More research is needed to evaluate the impact of market transparency in other contexts.

Finally, our work presents a new case-study of a collusive agreement which exploited the reconciliation systems. Future research (or scrutiny by regulators) could determine whether this institutional feature lead to widespread collusion in similar settings.

## 4.7 Appendix

### 4.7.1 Data

In this paper we use three main sources of data. The first one, available from the webpage of *XM*, contains detailed information on market variables of the Colombian wholesale electricity market from August 2008 to July 2009. The database has the universe of submitted bidding programs, the forward contracts hourly sales of each firm, the hourly demand and spot price, the daily water intakes of the reservoirs for each hydro unit, the quantities and revenues from positive and negative reconciliations as well as the contingencies of the transmission infrastructure.

The second dataset provides time-varying marginal costs for each generation unit. To construct them, we follow a standard engineering methodology (Green and Newbery, 1992; Wolfram, 1998, 1999; Wolak, 2000; Fabra and Reguant, 2014) that uses technical specifications of each generation unit (i.e. heat rate), fuel prices and transportation costs (see Appendix 4.7.4 for details about calculations and data sources).

Finally, we hand-collected data from the minutes of the meetings of the Association of Generating Units (CNO in Spanish).[122] We first download the minutes and type the name of each attendee in an excel file. Then, we give the excel file to two different RAs to complete the occupation. They searched for the CV of the attendees of these meetings through Linkedin and other web sources. We were particularly interested to know if attendees had a job position in the commercial area, and therefore were likely to be directly involved in setting bids at the time of the meeting. The great majority of information collected was uniform across RAs. In case of discrepancies, the authors took a decision. The rule we follow is that unless there is clear evidence of the occupation, we will leave it as a missing value.

[122]For more information, see https://www.cno.org.co/content/quienes-somos and the report from the regulators (Superintendencia Delegada para Energa y Gas, 2008).

## 4.7.2 Robustness in the cartel definition

### 4.7.2.1 Robustness in the cartel definition: Firms' ownership

Our cartel definition has classified units and not firms. To know the extent of which this can biases our results, Figure 4.19 reports estimated coefficients of the interaction between the dummies for announcement, $\mathbb{1}\{Announc\}_t$, and implementation, $\mathbb{1}\{Trnsp\}_t$, with four alternative definitions of cartel membership: Baseline, Refined, Extended and Placebo units.

The first group comprises the 14 units from the baseline definition. The second group includes only 9 units that belong to firms for which *all* their units were initially classified in the baseline Cartel. The *extended* units group, with 22 units, includes the baseline units plus other units that also belong to the firms that have at least one unit in the baseline definition of the cartel. Finally, to conduct the placebo exercise, we randomly allocate some of the units to the placebo cartel and the rest to the control group. In doing so, we keep the same proportion of cartel and non-cartel units as is in our baseline definition (14/47). We repeat this procedure 1,000 times and report the mean of the effect across repetitions along with confidence intervals constructed with the standard deviation across repetitions.

Figure 4.19 presents the results and shows two main patterns. First, for the refined and extended units groups, both the announcement and implementation coefficients are significantly lower than zero. The coefficient of the interaction term of the announcement is lower than the coefficient of the interaction term of the implementation for both groups. Second, the previous pattern is different for the placebo exercise. Units randomly allocated to the cartel group sometimes have an increase and sometimes a decrease in bidding prices after the announcement or the implementation period, which results in a zero average effect. Importantly, the standard deviation of the estimates from the bootstrap exercise suggests that our baseline estimates are unlikely to be the result of chance.

### 4.7.2.2 Robustness in the cartel definition: Alternative criteria

So far, we have assumed that the cartel was formed by Thermal Atlantic units and have explored robustness using firms' ownership of units. In this subsection, we pursue a different approach in which we consider additional criteria to define our proxy for cartel membership. Specifically, we consider the role of (1) private (vs public) ownership, (2) forward contract positions, and (3) bidding behaviour in 2008, i.e., *before* the announcement date. We refine our baseline definition including these additional criteria progressively building on our baseline definition. In particular, we use factor analysis to define cartel membership based on different sets of variables. Given a set of explanatory variables,

we define the cartel as being composed by those units to which the factor analysis assigns positive factors. Changing the variables used in the factor analysis leads to four alternative definitions of cartel:

1. **Cartel 2**: *Three dummies: Atlantic, Thermal, and Private.* The logic of this definition is to question the extent that private ownership matters for our results (in our baseline cartel, 36% of units are public). For instance, Barros and Modesto (1999) argue that private units maximize profits while public firms maximize welfare or other objective functions.

2. **Cartel 3**: *Two dummies: Atlantic and Thermal, and one continuous variable: Forward Contracts.* We include forward contracts to capture the incentive to modify short-term market aggregates. Since forward contracts are defined at the firm level, we include in the factor analysis the share of a firm's capacity that is not covered by forward contracts.

3. **Cartel 4**: *Three dummies: Atlantic, Thermal and Private, and one continuous variable: Trend in Bidding Behaviour in the Pre-Period.* We construct a proxy for the bidding behavior of each unit in all of the period of 2008 by regressing the logarithm of bids on unit fixed effects interacted with a linear time trend during 2008. We then include in the factor analysis the average estimated fixed effect for each unit. This exercise yields a parsimonious estimate of how a given unit changed its bidding behaviour during 2008.

4. **Cartel 5**: *Three dummies: Atlantic, Thermal, Private , and two continuous variable: Forward Contracts, and Bidding Behaviour in the Pre-Period.* Finally, we include in the factor analysis all the considered variables: A dummy for being located in the Atlantic coast, a dummy for Thermal production technology, a dummy for private ownership, our continuous measure for Forward Contract coverage, and our proxy for Bidding Behavior in 2008.

Table 4.10 shows the correlation matrix for the different definitions. Although the correlation is always positive and significant –at 1%–, it ranges from moderate (0.45) to high (0.95).

Table 4.11 shows the difference in difference estimates for these four alternative definitions. The coefficient of Cartel Announcement is always negative and significant and ranges from -0.27 to -0.73, suggesting that the effect of the policy change could be larger than that captured by our baseline definition. The coefficient of Cartel Implementation is not significant at conventional levels.[123]

[123]Unreported result are robust to the contemporaneous inclusion of the interaction between date and

Figure 4.20 shows the event study for these four definitions. For all of them, the level of the coefficients after the announcement is lower than before the announcement. In particular, for all definitions, there is a sharp and discontinuous drop in the coefficients right after the announcement date.

Figure 4.21 shows that when we refine or extend the Cartel definitions as well as when we conduct a similar placebo exercise as proposed above, the coefficient estimated for the announcement interaction is always negative and larger in magnitude than the coefficient estimated for the implementation interaction.

While our baseline definition of the cartel focuses a priori on Thermal units, the alternative definitions do not. In fact, Cartel 3 to 5 include one hydro unit each (not always the same) and suggest the main finding is robust to their inclusion.

### 4.7.3 Details on the cost of the cartel

We assume that the total amount of positive reconciliations produced by the cartel is independent of its members colluding or competing. That is, (i) units cannot strategically create positive reconciliations; (ii) the collusive behavior only changes the particular allocation of production of energy within cartel units. Our measure thus provides a lower bound estimate of the benefit of competition. The rationale of why this is the case is that if (i) does not hold, competition would imply that a share of positive reconciliation is awarded via the ideal dispatch and paid at the lower spot market price. Similarly, if (ii) does not hold, lower cartel members' bid could increase the market share of these units in the positive reconciliation market if their bids are lower than non-cartel units. If that is the case, we ignore the lower cost consumers would pay on the additional market share.

In practice, we multiply the bids and amounts constructed to test the DEC at the unit-day level and then sum over units. As for the previous exercises, we aggregate costs at the monthly level and present the results in the bottom panel of Figure 4.12.

### 4.7.4 Calculation marginal costs

As previous studies in the literature on market power in electricity markets (Green and Newbery, 1992; Wolfram, 1998, 1999; Wolak, 2000; Fabra and Reguant, 2014), we use information about the fuel burned, the thermal efficiency, and the price and transportation cost of the corresponding fuel to compute an estimate of the unit cost per kilowatt hour of each generation plant.

We calculated marginal costs of thermal plants using the heat rate, fuel costs and fuel transportation costs with the following formula:

technology fixed effects as well as date and region fixed effects. The additional criteria introduce variation within our baseline characterization that enables us to include this more exhaustive set of controls.

$$\underbrace{Exchange\ R._t}_{\frac{COP\$}{US\$}} \times \Big[ \underbrace{Heat\ R._i}_{\frac{MBTU}{KWh}} \times \underbrace{(Transp.\ fuel\ cost_i + Fuel\ cost_t)}_{\frac{US\$}{MBTU}} \Big] = \underbrace{Marginal\ Cost_{it}}_{\frac{COP\$}{kWh}}$$

Where *COP* are Colombian pesos, *MBTU* are one thousand of the British thermal unit, *US* are United States dollars and *KWh* is one kilowatt per hour. The heat rate is a measure of the thermal efficiency of the generation unit. It represents the quantity of fuel measured in *MBTU* necessary to generate one kilowatt per hour. As previous studies, we obtained heat rates from statistical reports issued by public entities (Green and Newbery, 1992; Wolfram, 1998, 1999). The parameters of the heat rate of thermal electricity generation Colombian units were extracted from the website of the market operator (XM).[124]

Regarding fuel prices, for non-internationally tradable inputs, we used a reference price of the contracts as in Wolfram (1999) and for tradable inputs, we used public information on prices in international energy markets as in Fabra and Reguant (2014).

In 2008 and 2009 natural gas was a non-tradable input in Colombia, given that it did not have import regasification facilities nor it was connected to an international gas hub. We use as a reference of the price of the natural gas contracts the price of the basin Guajira which is the most important gas supply source for Colombian thermal generation. From September 1995 Until August 2013, the Colombian Government regulated the prices of the sales contracts of this gas source. The regulation consist in imposing a maximum sale price of gas. This maximum price at period $t$, $p_t$, is given by the formula $p_{t-1}[index_{t-1}/index_{t-2}]$ where $index_{t-1}$ is the average of the last semester of the New York Harbor Residual Fuel Oil 1.0 % Sulfur LP Spot Price according to the series that was published by the Energy Information Administration of the United States. A period $t$ is defined as semester and it changes 1st February and 1st August of each year.[125] This price is given in *US dollars/MBTU*.

We calculated the Guajira regulated price applying the formula presented above and converting the resulting price *(US dollars/MBTU)* to *Colombian pesos/KWh*. The exchange rate data was obtained from the Colombian central bank (Banco de la Repblica)[126].

As the previous studies of Green and Newbery (1992) and Wolfram (1999) we included the transportation cost in the marginal cost computation.

Consequently with the fuel cost reference, for gas fired units, we take as transportation costs the sum of the fees for the use of each segment of the gas transmission network necessary to take the gas from Guajira well to the respective generation units. These fees are regulated by the CREG and are published in regulatory acts (CREG, 2003a,b).

[124]See: http://paratec.xm.com.co/paratec/SitePages/generacion.aspx?q=capacidad.
[125]The formula was established in Resolution 119/2005 of CREG (CREG, 2005)
[126]See:https://www.banrep.gov.co/es/estadisticas/trm

### 4.7.5 Additional figures

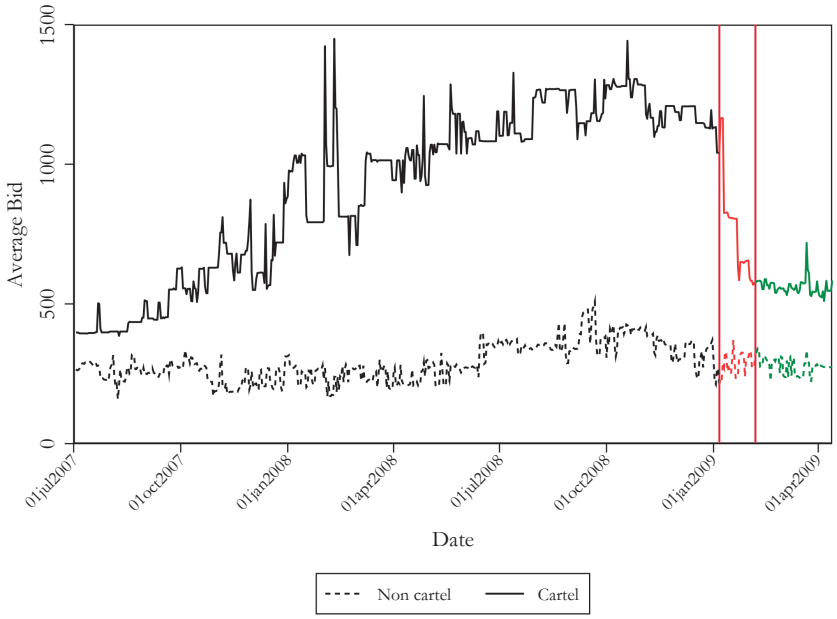



Figure 4.13: The Main Fact: Cartel and non cartel groups bids (over a longer period of time).

Note: Time series of the average bid of the cartel (solid line) and non-cartel groups (dashed line) around the dates of announcement and implementation of the transparency policy. The vertical lines show the announcement and implementation dates.

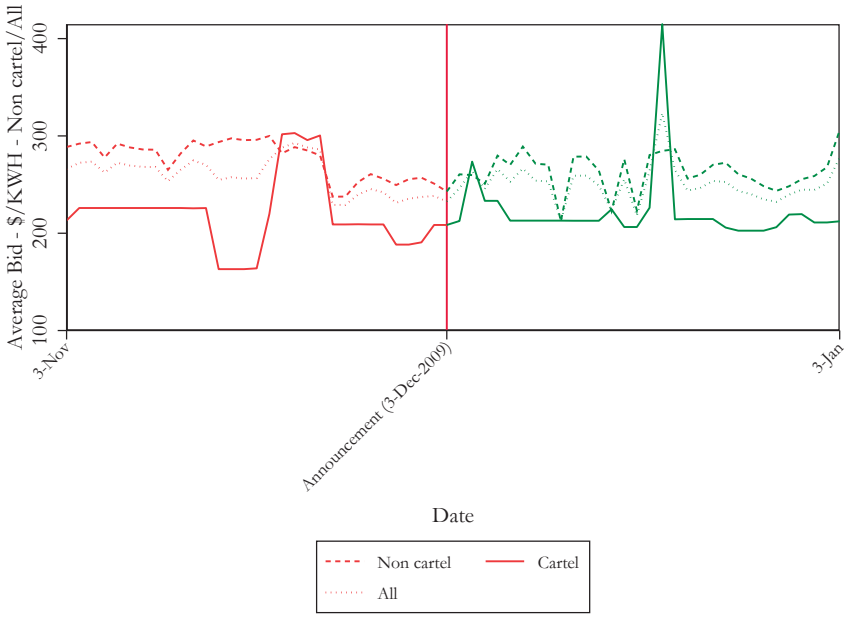



Figure 4.14: Cartel and non-cartel bids around the policy change of December 2009

Note: Time series of the average bid of the cartel (solid line) and non-cartel groups (dashed line) around the announcement date related to the (second) transparency policy change in December 2009. We also report the overall average (dotted line). The vertical line points to the announcement date which is the same implementation date.

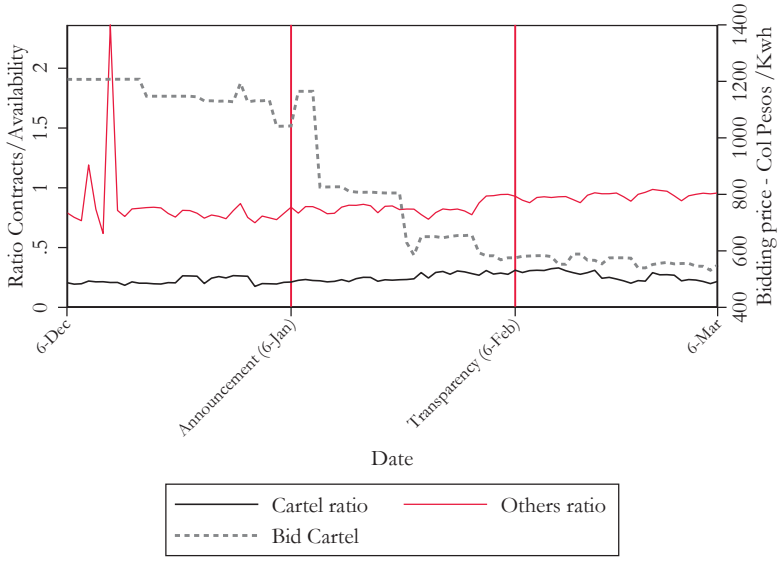



Figure 4.15: Time Series of Forward Contracts for Cartel and Non-Cartel Members

Note: The figure presents the time series of the portion of capacity sold through forward contracts of the cartel and non-cartel groups around the dates of announcement and implementation of the transparency policy. The black line represents the time series of the fraction of the forward contracts for the cartel group, while the red line represents the non-cartel group time series. The dotted line represents the average bidding price for the cartel group. As the availability and contract variables are set for each hour, we simply sum across hours to have a daily measure.

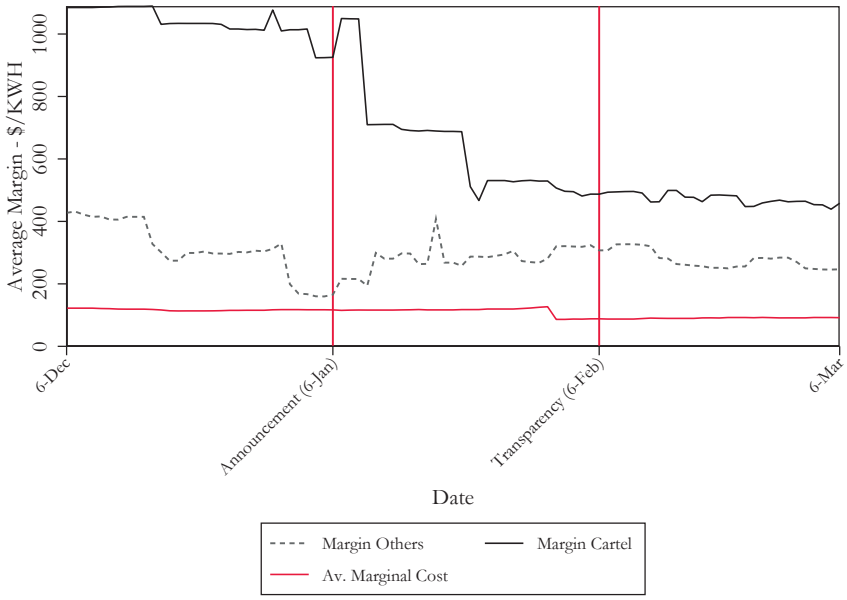



Figure 4.16: Average margin time series.

Note: The figure presents the time series of the average margin of the cartel (solid black line) and non-cartel (dotted grey line) groups around the dates of announcement and implementation of the transparency policy (From November 2008 to April 2009). The margin is computed as the difference between the bid minus the marginal cost. It also presents the time series of the average marginal cost of the cartel units (solid red line).

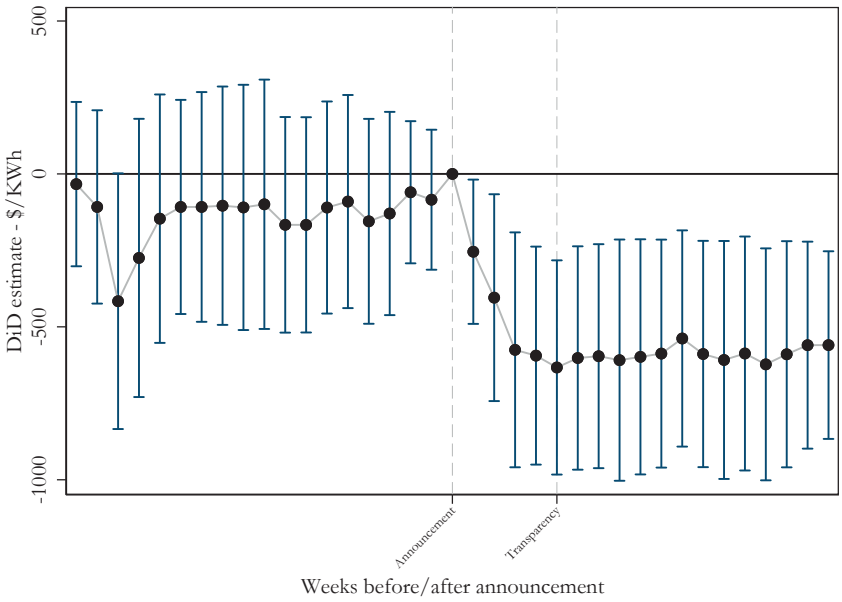



Figure 4.17: Event study representation using margin as the dependent variable

Note: The figure presents the event study representation of the difference-in-difference model using margin as the dependent variable, computed as bid minus marginal cost. We performed a two-way fixed effects model including a specific treatment effect for each week of the period studied. Robust s.e. are clustered by unit and date. The x-axis represents the weeks around the policy announcement. The y-axis reports the estimates using the week of the announcement as baseline. Dots and bars represent point estimates and 95% confidence intervals. The dotted line labeled as "Announcement" represents the week of the announcement of the transparency policy. The dotted line labeled as "Transparency" represents the week of the implementation of the transparency policy.

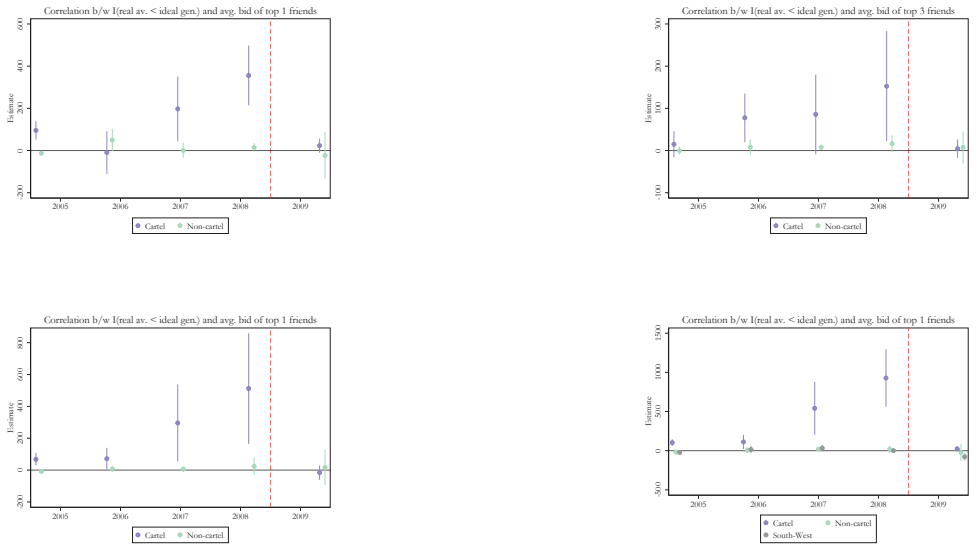



Figure 4.18: Robustness for coordinated bids analysis.

Note: Estimates from regressions where the outcome variable is the average bid of the friends of unit $i$ and the explanatory variables is an indicator for unit $i$ declaring a level of real availability below the ideal generation quantity it was awarded. We run separate regressions for the two groups (cartel, non-cartel) and repeat for years 2005 to 2009. Compared to the baseline analysis in Figure 4.10, we perform four robustness exercises. (i) In the top left panel, we still consider 'top 1' friends from the same period as in the baseline, but we include in the explanatory dummy **all** cases where the real availability is smaller than ideal generation (differently from the baseline, where we consider the 75% cases where the difference between real availability and ideal generation is the largest). (ii) In the top right panel, we consider the same period and same cases as in the baseline, but use the 'top 3' friends. (iii) In the bottom left panel, we consider 'top 1' friends and the same cases as in the baseline, but we construct 'friends' using observations from a longer period (2005-2008) compared to the baseline. (iv) In the bottom right panel, we repeat the same analysis as in the baseline but also report separately the estimates for the units clustered in the South-West part of Colombia. The estimates for 2009 needs to be interpreted cautiously. Data on real availability is missing for 63% of cartel observations and for 6% of non-cartel observations in 2009.

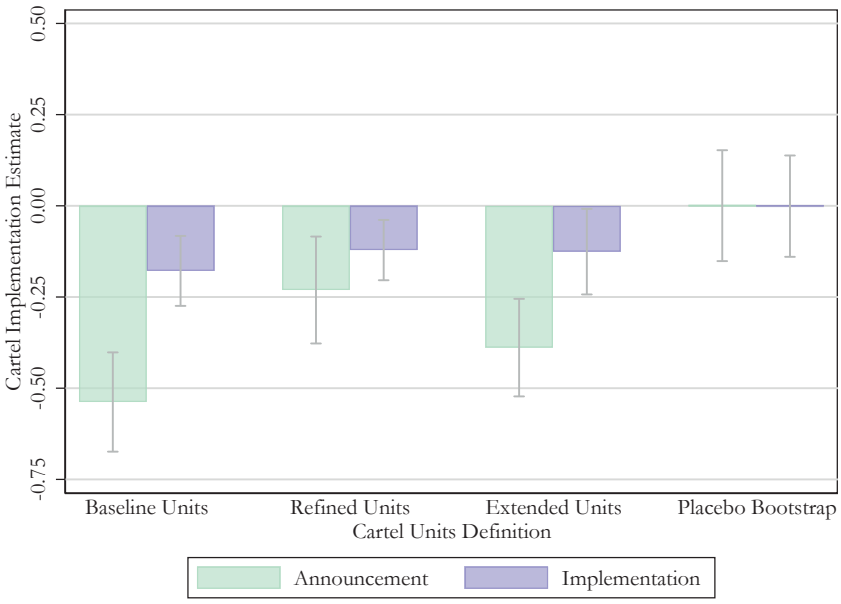



Figure 4.19: Robustness Exercises.

Note: The figure shows estimates of the 'announcement' and 'implementation' parameters from 4 different DiD estimations. 'Baseline units' reports estimates for our baseline cartel definition (14 units). 'Refined units' reports estimates when we include in the cartel group only units (9 units) that belong to firms that have all their units in the baseline cartel definition. 'Extended units' reports estimates when we include in the cartel group all the units (22 units) of firms for which at least one unit belong to the baseline cartel definition. For the placebo exercise, we randomly allocate some of the units to the placebo cartel and the rest to the control group. In doing so, we keep the same proportion of cartel and non-cartel units as is in our baseline definition. We repeat this procedure 1000 times. All estimates control for unit and date fixed effects and robust s.e. are clustered by unit and date.

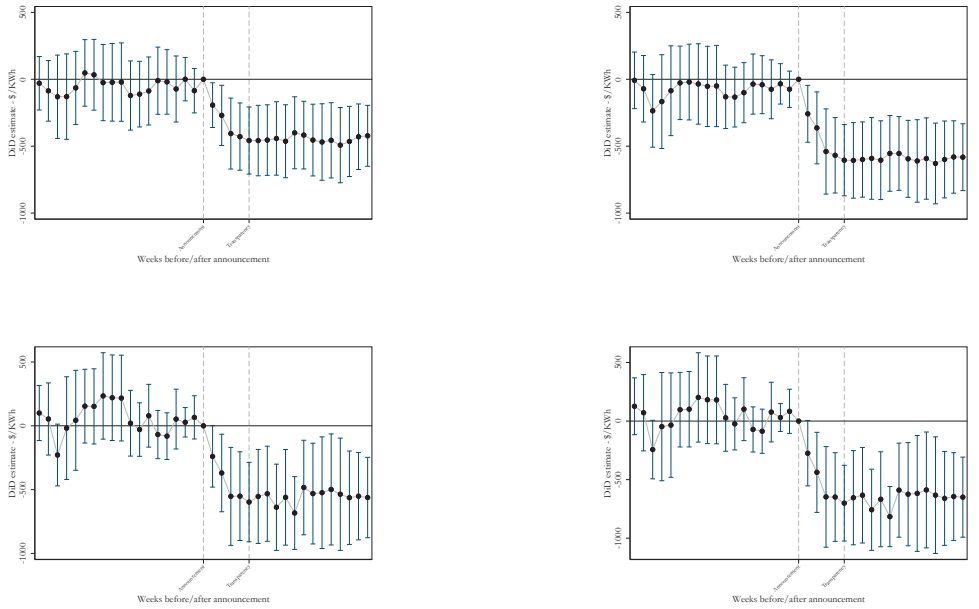



Figure 4.20: Event study representation for alternative cartel definitions

Note: The figure presents the event study representation for bids. The top left figure shows the event study for cartel 2 (PCA on Atlantic, Thermal, and Private) definition. The top right figure shows the event study for cartel 3 (PCA on Atlantic, Thermal, and Forward Contracts) definition. The bottom left figure shows the event study for cartel 4 (PCA on Atlantic, Thermal, Private, and Bid slope) definition. The bottom right figure shows the event study for cartel 5 (PCA on Atlantic, Thermal, Forward Contracts, Private and Bid slope) definition.

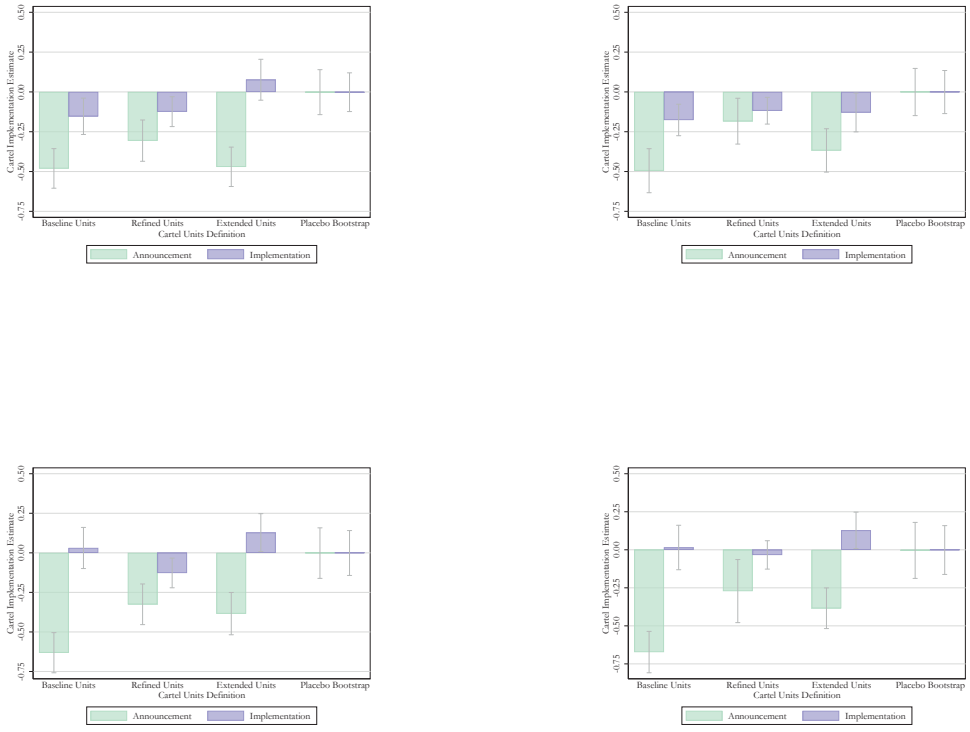



Figure 4.21: Refined and extended units from four cartel definitions

Note: Each sub-figure shows the estimates for 'announcement' and 'implementation' parameters from 4 different DiD regressions. The top left figure shows the result for cartel 2 (PCA on Atlantic, Thermal, and Private) definition. The top right figure shows the result for cartel 3 (PCA on Atlantic, Thermal, and Forward Contracts) definition. The bottom left figure shows the result for cartel 4 (PCA on Atlantic, Thermal, Private and Bid slope) definition. The bottom right figure shows the result for cartel 5 (PCA on Atlantic, Thermal, Forward Contracts, Private and Bid slope) definition. In each sub-figure, baseline units refers to each of the corresponding cartel definition (2, 3, 4 or 5). The refined units group only includes the cartel units that belong to firms that have all their units in the baseline cartel. The extended units group includes in the cartel all of the units of firms for which at least one unit belongs to the baseline cartel definition. For the placebo exercise, we randomly allocate some of the units to the placebo cartel and the rest to the control group. In doing so, we keep the same proportion of cartel and non-cartel units as is in the baseline definition. We repeat this procedure 1000 times. All estimates control for unit and date fixed effects and robust s.e. are clustered by unit and date.

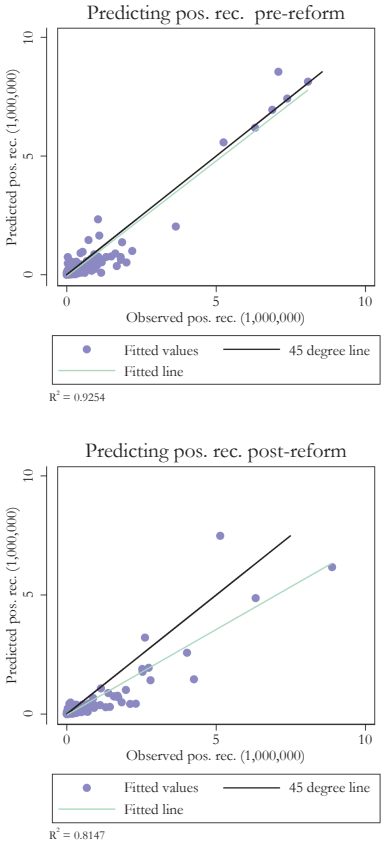



Figure 4.22: Comparing observed quantities of positive reconciliations with in-sample predictions for the cartel and non-cartel units in the pre and post-reform periods.

Note: We estimate how the quantity of positive reconciliation awarded to a unit depends on the rank of its bid. We use cartel and non-cartel units in this exercise (while Figure 4.26 refers to cartel units only). We use the estimates to make in-sample predictions for positive reconciliations at the day-unit level based on units' bids. In the figure, we compare the average predicted quantity (y-axis) with the average observed one (x-axis). The top (bottom) figure refers to observations from the six months before (after) the reform.

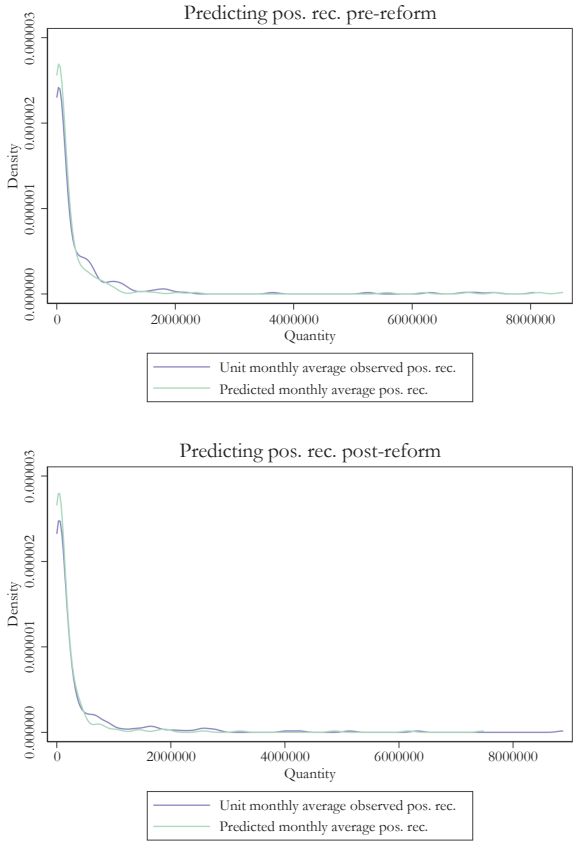



Figure 4.23: Comparing the distribution of observed quantities of positive reconciliations with in-sample predictions for the cartel and non-cartel units in the pre and post-reform periods.

Note: We estimate how the quantity of positive reconciliation awarded to a unit depends on the rank of its bid. We use cartel and non-cartel units in this exercise (while Figure 4.27 refers to cartel units only). We use the estimates to make in-sample prediction for positive reconciliations at the day-unit level based on units' bids. In the figure, we compare the density of the average predicted quantity (green line) with the density of the average observed one (purple line). The top (bottom) figure refers to observations from the six months before (after) the reform.

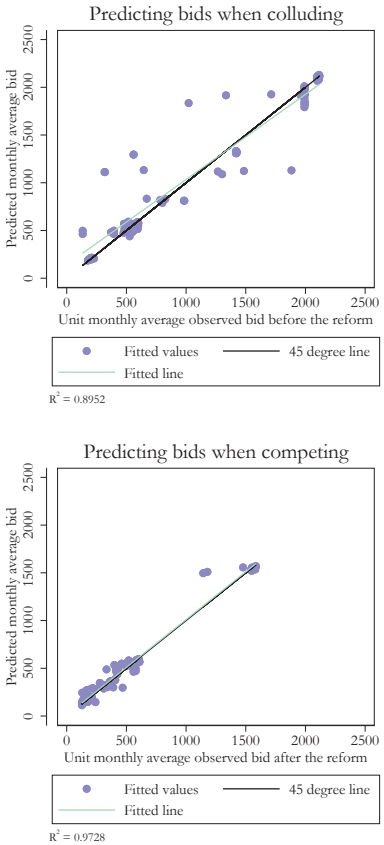



Figure 4.24: Comparing observed bids with in-sample predictions.

Note: We estimate how cartel units set bids by regressing bids on costs, the lagged logarithm of the total amount of positive reconciliations, and the logarithm of the ideal generation quantity. We use the resulting estimates to make in-sample predictions and average at the monthly level for each unit. In the figure, we compare the average predicted bid (y-axis) with the average observed one (x-axis). The top (bottom) figure refers to observations from the six months before (after) the reform.

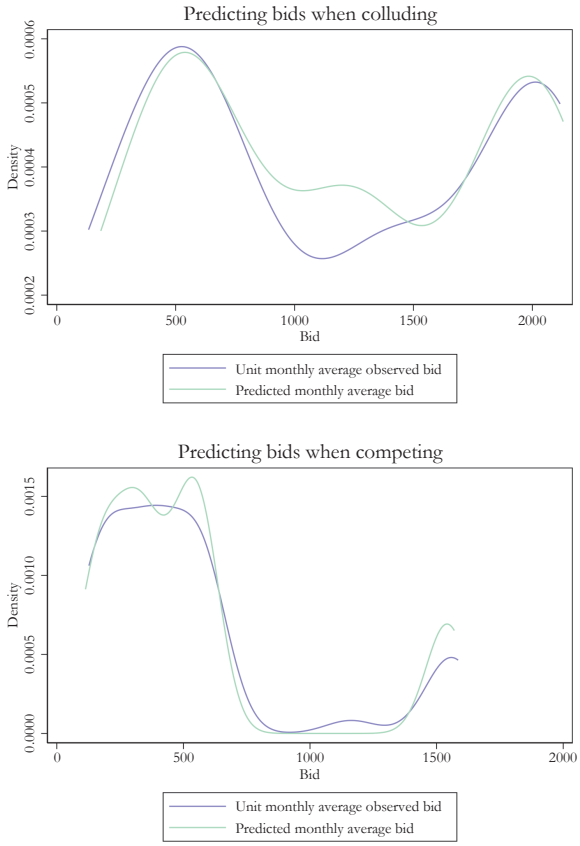



Figure 4.25: Comparing the distribution of observed bids with in-sample predictions.

Note: We estimate how cartel units set bids by regressing bids on costs, the lagged logarithm of the total amount of positive reconciliations, and the logarithm of the ideal generation quantity. We use the resulting estimates to make in-sample prediction and average at the monthly level for each unit. In the figure, we compare the density of the average predicted bid (green line) with the density of the average observed one (purple line). The top (bottom) figure refers to observations from the six months before (after) the reform.

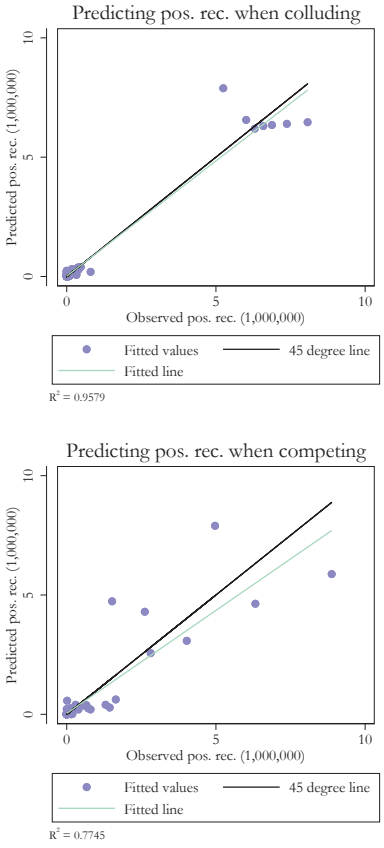



Figure 4.26: Comparing observed quantities of positive reconciliations with in-sample predictions for cartel units in the pre and post-reform periods.

Note: We estimate how the quantity of positive reconciliation awarded to a unit depends on the rank of its bid. We focus on cartel units in this exercise (while Figure 4.22 refers to all units only). We use the estimates to make in-sample predictions for positive reconciliations at the day-unit level based on units' bids. We then average at the monthly level for each unit. In the figure, we compare the average predicted quantity (y-axis) with the average observed one (x-axis). The top (bottom) figure refers to observations from the six months before (after) the reform.

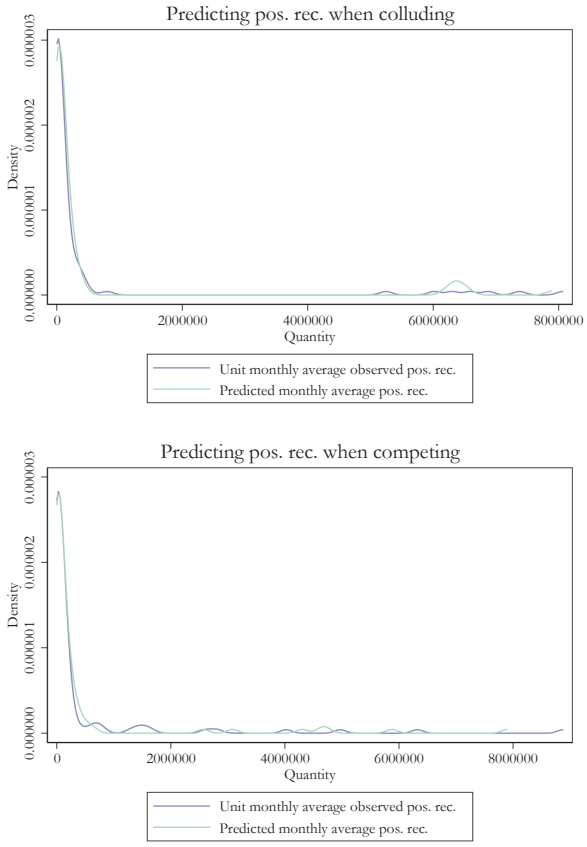



Figure 4.27: Comparing the distribution of observed quantities of positive reconciliations with in-sample predictions for cartel units in the pre and post-reform periods.

Note: We estimate how the quantity of positive reconciliation awarded to a unit depends on the rank of its bid. We focus on cartel units in this exercise (while Figure 4.23 refers to all units only). We use the resulting estimates to make in-sample predictions for positive reconciliations at the day-unit level based on units' bids. We then average at the monthly level for each unit. In the figure, we compare the density of the average predicted quantity (green line) with the density of the average observed one (purple line). The top (bottom) figure refers to observations from the six months before (after) the reform.

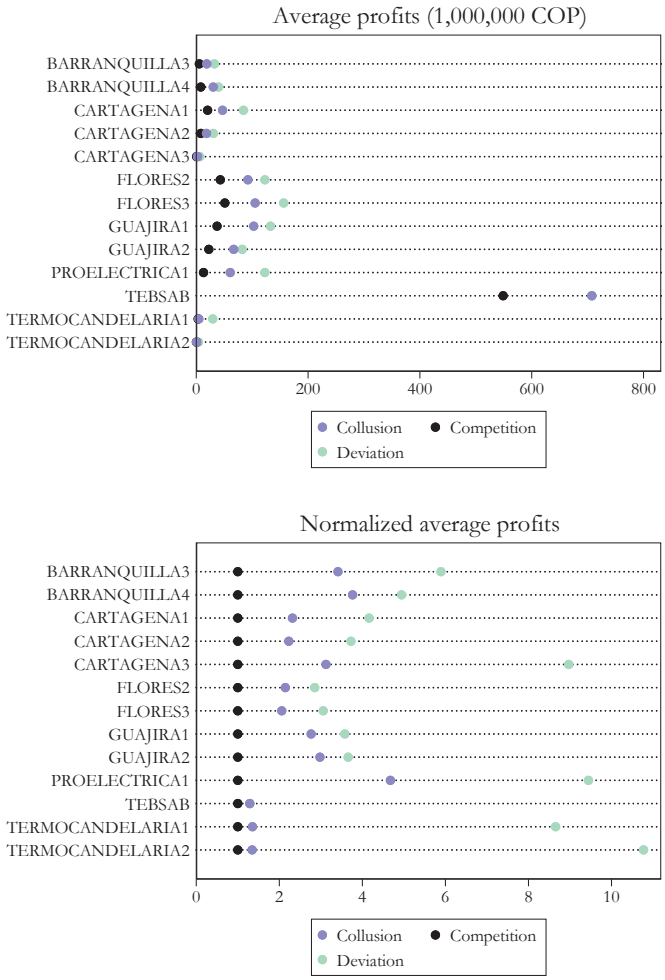



Figure 4.28: Average profits of cartel units under competition, collusion, and optimal deviation.

Note: We construct counterfactual bids and quantities under three alternative scenarios: Collusion, competition, and optimal deviation from collusion. Based on these variables we construct profits for each unit under the three scenarios and average over a one-year period around the reform. The top figure reports the level of the average profits (for some units, profits under different scenarios overlap in the figure), while the bottom one reports the ratio with respect to competitive profits.

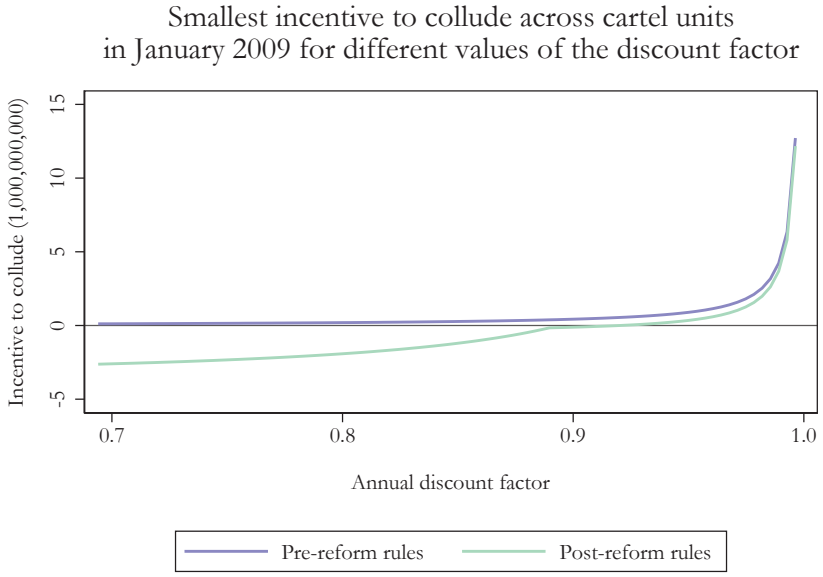



Figure 4.29: Cartel sustainability for different discount factors.

### 4.7.6 Additional tables

| VARIABLES | (1) Margin | (2) Margin | (3) Margin | (4) Margin |
|---|---|---|---|---|
| Cartel Announcement | -320.52 | -320.52 | -308.62 | -454.11 |
| | (125.83) | (130.48) | (130.39) | (116.92) |
| Cartel Implementation | -146.57 | -146.57 | -146.40 | -145.18 |
| | (47.66) | (57.51) | (56.76) | (35.37) |
| Announcement | -130.83 | | | |
| | (41.96) | | | |
| Implementation | -33.78 | | | |
| | (24.93) | | | |
| Observations | 11,315 | 11,315 | 16,955 | 16,955 |
| R-squared | 0.23 | 0.82 | 0.81 | 0.82 |
| Unit FE | NO | YES | YES | YES |
| Date FE | NO | YES | N/A | N/A |
| Date x Technology FE | NO | NO | YES | NO |
| Date x Region FE | NO | NO | NO | YES |
| Forward Contracts | NO | NO | YES | YES |

*** p<0.01, ** p<0.05, * p<0.1

Table 4.7: Difference-in-difference estimates - Margin

Note: The table presents the estimation results of the differences in differences model using margin as the dependent variable, computed as bid minus marginal cost. Only thermal units are included in the sample. In columns 3 and 4 we further control for forward contracts over total capacity and alternatively for Date × Technology FE or for Date × Region FE. Regions are Atlantic, North-West, Central and South-West. Robust s.e. clustered by unit and date in parenthesis.

| Scenario | Variable | Obs | Mean | Std. Dev. | Min | Max | Freq. Negative | T-Test |
|---|---|---|---|---|---|---|---|---|
| 1 | Actual profit | 2142 | 61.81 | 201.43 | -29.84 | 1275.53 | 0.0014 | 13.89 |
| | Counterf. spot market profit | 2142 | 2.79 | 15.57 | 0.00 | 259.88 | 0.0000 | |
| 2 | Actual profit | 1064 | 62.87 | 208.94 | 0.00 | 1139.43 | 0.0000 | 7.63 |
| | Counterf. spot market profit | 1064 | 20.68 | 45.41 | 0.00 | 341.46 | 0.0000 | |
| 3 | Actual profit | 1064 | 62.87 | 208.94 | 0.00 | 1139.43 | 0.0000 | 9.82 |
| | Counterf. spot market profit | 1064 | 0.01 | 0.16 | 0.00 | 4.70 | 0.0000 | |
| 4 | Actual profit | 1078 | 60.77 | 193.82 | -29.84 | 1275.53 | 0.0028 | 5.58 |
| | Counterf. spot market profit | 1078 | 34.57 | 68.54 | 0.00 | 672.81 | 0.0000 | |
| 5 | Actual profit | 1078 | 60.77 | 193.82 | -29.84 | 1275.53 | 0.0028 | 9.98 |
| | Counterf. spot market profit | 1078 | 2.42 | 16.84 | 0.00 | 239.49 | 0.0000 | |

Table 4.8: Comparison of profits form positive reconciliation and counterfactual competition

Note: We performed a comparison between the profits that the units of the cartel group obtained from the positive reconciliation collusive agreement and the profits that those units would obtain in the counterfactual case in which they bid their marginal costs and try to win in the ideal dispatch. First, we computed the profits from the positive reconciliation collusive agreement as the value of income from positive reconciliations minus the cost of generating the energy. Second, for computing the counterfactual of the profits if the units bid as a competitive firm, we assumed that if cartel firms would bid competitively it had the same probability of being in merit as the competitive units. Hence, we computed the probability of being on merit of the no cartel units. We computed the counterfactual profits for cartel units as the product of the probability of being on merit (if the unit is competitive) multiplied by the profit obtained by the unit if it would sell its energy at the spot price and would generate its declared availability. We allow the possibility of inaction of the unit. Hence if the profit above is negative we replace it with zero. For this computation, we only consider thermal units. We use the data for the second semester of2008. We compute five counterfactual scenarios. **Scenario 1**: Average spot price and average hydro resources condition. The spot price used for computation is the average spot price. All the days in the sample are considered. **Scenario 2**: High spot price and high hydro resources condition. The spot price used for computation is the spot price in the higher demand hour (7 p.m.). Only the days with hydro resources higher than the median are considered. **Scenario 3**: Low spot price and high hydro resources condition. The spot price used for computation is the spot price in the lower demand hour (3 a.m.). Only the days with hydro resources higher than the median are considered. **Scenario 4**: High spot price and low hydro resources condition. The spot price used for computation is the spot price in the higher demand hour (7 p.m.). Only the days with hydro resources lower than the median are considered. **Scenario 5**: Low spot price and low hydro resources condition. The spot price used for computation is the spot price in the lower demand hour (3 a.m.). Only the days with hydro resources lower than the median are considered.

| VARIABLES | (1) Profits from PR | (2) Total profits | (3) Profits from PR | (4) Total profits | (5) Profits from PR | (6) Total profits |
|---|---|---|---|---|---|---|
| Cartel High Post | -66.25 | -89.11 | -67.85 | -83.41 | -91.02 | -85.83 |
| | (27.57) | (21.05) | (43.89) | (21.21) | (37.14) | (21.20) |
| Cartel Low Post | -153.99 | -59.47 | -146.14 | -65.17 | -141.79 | -62.75 |
| | (70.58) | (22.92) | (67.54) | (23.66) | (68.03) | (23.35) |
| Observations | 6,725 | 17,155 | 6,725 | 17,155 | 6,725 | 17,155 |
| R-squared | 0.69 | 0.79 | 0.68 | 0.79 | 0.68 | 0.79 |
| Unit FE | YES | YES | YES | YES | YES | YES |
| Date FE | YES | YES | YES | YES | YES | YES |
| Split high/low | Mg. Cost | Mg. Cost | Bid | Bid | Neg. Rec. | Neg. Rec. |

Robust standard errors in parentheses

Table 4.9: Effects of announcement on profits

Note: The table presents difference in differences estimates controlling for unit and time fixed effects, where the Post period refers to the period after the policy announcement, for two outcome variables: The profits from positive reconciliations, conditional on receiving some positive reconciliations, and the total profits (unconditional). We split the cartel group in two using different measures. In columns 1-2, 'high' units are those with average marginal cost in the second half of 2008 above the median, and 'low' otherwise. In columns 3-4, 'high' units are those with average bids in the second half of 2008 above the median, and 'low' otherwise. In columns, 5-6, 'high' units are those with an average amount of negative reconciliations *below* the median in the second half of 2008, and 'low' otherwise. Robust s.e. clustered by unit and date in parenthesis.

| | Cartel 1 | Cartel 2 | Cartel 3 | Cartel 4 | Cartel 5 |
|---|---|---|---|---|---|
| Cartel 1 | 1.000 | 0.694 | 0.951 | 0.579 | 0.684 |
| Cartel 2 | 0.694 | 1.000 | 0.638 | 0.526 | 0.450 |
| Cartel 3 | 0.951 | 0.638 | 1.000 | 0.541 | 0.648 |
| Cartel 4 | 0.579 | 0.526 | 0.541 | 1.000 | 0.888 |
| Cartel 5 | 0.684 | 0.450 | 0.648 | 0.888 | 1.000 |

Table 4.10: Correlation Table of Alternative Cartel Definitions

Note: The table shows the correlation between the different cartel definitions. All the correlations are significant at 1% level.

| VARIABLES | (1) LnBid | (2) LnBid | (3) LnBid | (4) LnBid | (5) LnBid | (6) LnBid | (7) LnBid | (8) LnBid | (9) LnBid | (10) LnBid |
|---|---|---|---|---|---|---|---|---|---|---|
| Cartel Announcement | -0.54 | -0.36 | -0.48 | -0.27 | -0.49 | -0.33 | -0.63 | -0.50 | -0.67 | -0.54 |
| | (0.14) | (0.13) | (0.12) | (0.14) | (0.14) | (0.13) | (0.13) | (0.14) | (0.14) | (0.14) |
| Cartel Implementation | -0.18 | -0.03 | -0.15 | 0.09 | -0.18 | -0.06 | 0.03 | 0.16 | 0.02 | 0.12 |
| | (0.10) | (0.12) | (0.11) | (0.21) | (0.10) | (0.12) | (0.13) | (0.13) | (0.15) | (0.14) |
| Observations | 17,155 | 16,955 | 17,155 | 16,955 | 17,155 | 16,955 | 17,155 | 16,955 | 17,155 | 16,955 |
| R-squared | 0.82 | 0.83 | 0.82 | 0.83 | 0.82 | 0.83 | 0.81 | 0.83 | 0.81 | 0.84 |
| Unit FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Date FE | YES | N/A | YES | N/A | YES | N/A | YES | N/A | YES | N/A |
| Date x Technology FE | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES |
| Forward Contracts | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES |
| Cartel Definition | 1 | 1 | 2 | 2 | 3 | 3 | 4 | 4 | 5 | 5 |

Robust standard errors in parentheses

Table 4.11: Difference in Difference Estimations for Alternative Cartel Definitions

Note: The table presents the estimation results of the DiD model using the logarithm of the bid as the dependent variable. Column 1 controls for unit and date Fixed effects. Column 2 controls for Date x Technology and unit Fixed Effects as well as forward contracts. The next columns have similar patterns. We repeat the same estimation for different cartel definitions as reported in the bottom row. Cartel 1 is the baseline. Cartel 2 comes from using PCA to Atlantic, Thermal, and Private. Cartel 3 comes from using PCA to Atlantic, Thermal, and Forward Contracts. Cartel 4 comes from using PCA to Atlantic, Thermal, and Bid slope. And Cartel 5 comes from using PCA to Atlantic, Thermal, Forward Contracts, and Bid slope. Robust s.e. clustered by unit and date in parenthesis.

| VARIABLES | (1) Pre-reform | (2) Post-reform |
|---|---|---|
| Marginal cost *(t)* | 1.065 | 1.763 |
| | (1.245) | (1.464) |
| (log) total amount of positive reconciliations *(t-1)* | 51.43 | -22.72 |
| | (77.33) | (10.72) |
| (log) total ideal generation *(t)* | 20.99 | -95.14 |
| | (93.33) | (27.74) |
| Observations | 2,506 | 2,534 |
| R-squared | 0.859 | 0.940 |
| Unit FE | YES | YES |

Robust standard errors in parentheses

Table 4.12: Estimation of bids on market fundamentals

Note: The table presents the estimates of the model used to predict the bids of cartel units. We regress bids on costs, the lagged value of the logarithm of the total amount of positive reconciliations, and the logarithm of the total amount of ideal generation. We use observations from cartel units from a one-year period around the reform (six months pre and six months post-reform in columns 1 and 2 respectively). Robust s.e. clustered by unit in parenthesis.

| VARIABLES | (1) Logit | (2) OLS |
|---|---|---|
| Rank | -1.089 | 0.189 |
| | (0.267) | (0.233) |
| Rank 2 | 0.0468 | -0.0156 |
| | (0.0266) | (0.0150) |
| Observations | 4,211 | 1,033 |
| R-squared | | 0.648 |
| Unit FE | YES | YES |
| Date FE | YES | YES |

Robust standard errors in parentheses

Table 4.13: Estimation of the positive reconciliation quantities model

Note: The table presents the estimates of the models used to predict the expected quantity of positive reconciliations for cartel units. In the first column, we present the logit estimates of a binary model where we regress a dummy for receiving positive reconciliations in a day on the rank of the bid, its squared value, and unit and date fixed effects. The second column presents the OLS estimates of a linear model where we regress the logarithm of the amount of positive reconciliations in a day on the same covariates as above, using only observations with some amount of positive reconciliations. We use observations from a one-year period around the reform. Robust s.e. clustered by unit in parenthesis.

# References

Albæk, S., Møllgaard, P., and Overgaard, P. B. (1997). Government-assisted Oligopoly Coordination? A Concrete Case. *The Journal of Industrial Economics*, 45(4):429–443.

Allcott, H., Collard-Wexler, A., and O'Connell, S. D. (2016). How Do Electricity Shortages Affect Industry? Evidence from India. *American Economic Review*, 106(3):587–624.

Asker, J. (2010). A Study of the Internal Organization of a Bidding Cartel. *American Economic Review*, 100(3):724–62.

Asker, J. and Nocke, V. (2021). Collusion, Mergers, and Related Antitrust Issues. In Ho, K., Hortacsu, A., and Lizzeri, A., editors, *Handbook of Industrial Organization*, volume 5 of *Handbook of Industrial Organization*, pages 177–279. Elsevier.

Bai, J., Chen, M., Liu, J., Mu, X., and Xu, D. Y. (2020). Search and Information Frictions on Global E-Commerce Platforms: Evidence from AliExpress. Working Paper 28100, National Bureau of Economic Research.

Baker, G., Gibbons, R., and Murphy, K. J. (2002). Relational Contracts and the Theory of the Firm. *The Quarterly Journal of Economics*, 117(1):39–84.

Baldwin, R. E., Chiarotti, E., and Taglioni, D. (2021). Trading Through Platforms: Evidence from AliExpress. Working paper.

Barkley, A. (2023). The Human Cost of Collusion: Health Effects of a Mexican Insulin Cartel. *Journal of the European Economic Association*, 21(5):1865–1904.

Barros, F. and Modesto, L. (1999). Portuguese Banking Sector: a Mixed Oligopoly? *International Journal of Industrial Organization*, 17(6):869–886.

Bergquist, L. F. and Dinerstein, M. (2020). Competition and Entry in Agricultural Markets: Experimental Evidence from Kenya. *American Economic Review*, 110(12):3705–47.

Bergquist, L. F., McIntosh, C., and Startz, M. (2023). Search Cost, Intermediation, and Trade: Experimental Evidence from Ugandan Agricultural Markets. Working paper.

Besley, T., Fontana, N., and Limodio, N. (2020). Antitrust Policies and Profitability in Non-Tradable Sectors. *American Economic Review: Insights*, 3(2):251–65.

Bigoni, M., Potters, J., and Spagnolo, G. (2019). Frequency of Interaction, Communication and Collusion: an Experiment. *Economic Theory*, 68:827844.

Blouin, A. and Macchiavello, R. (2019). Strategic Default in the International Coffee Market. *The Quarterly Journal of Economics*, 134(2):895–951.

Byrne, D. P. and De Roos, N. (2019). Learning to Coordinate: A Study in Retail Gasoline. *American Economic Review*, 109(2):591–619.

Camelo, S., Papavasiliou, A., de Castro, L., Riascos, Á., and Oren, S. (2018). A Structural Model to Evaluate the Transition from Self-commitment to Centralized Unit Commitment. *Energy Economics*, 75:560–572.

Chassang, S. and Ortner, J. (2019). Collusion in Auctions with Constrained Bids: Theory and Evidence from Public Procurement. *Journal of Political Economy*, 127(5):2269–2300.

Chassang, S. and Ortner, J. (2023). Regulating Collusion. *Annual Review of Economics*, 15:177–204.

Connor, J. M. (2020). Private International Cartels Full Data 2019 edition. Purdue University Research Repository.

Connor, J. M. and Bolotova, Y. (2006). Cartel Overcharges: Survey and Meta-analysis. *International Journal of Industrial Organization*, 24(6):1109–1137.

Cramton, P. and Schwartz, J. (1998a). Collusive Bidding in the FCC Spectrum Auctions. Technical report, University of Maryland.

Cramton, P. and Schwartz, J. (1998b). Collusive Bidding: Lessons from the FCC Spectrum Auctions. Technical report, University of Maryland.

Cramton, P. and Wilson, R. (1998). A Review of ISO New England's Proposed Market Rules. Technical report, Market Design Inc.

Crawford, G. S., Crespo, J., and Tauchen, H. (2007). Bidding Asymmetries in Multiunit Auctions: Implications of Bid Function Equilibria in the British Spot Market for Electricity. *International Journal of Industrial Organization*, 25(6):1233–1268.

CREG (2003a). Decisón sobre la solicitud de revisión de Cargos Regulados del Sistema de Transporte de PROMIGAS S.A. E.S.P. Technical report, Comisión de Regulación de Energía y Gas, Resolución 70.

CREG (2003b). Por la cual se resuelven los Recursos de Reposición interpuestos contra la Resolución CREG-013 de 2003. Technical report, Comisión de Regulación de Energía y Gas, Resolución 125.

CREG (2005). Sustitución del articulo 3 de la Resolución CREG 023 de 2000. Technical report, Comisión de Regulación de Energía y Gas, Resolución 119.

CREG (2009a). Manejo de la Información en el Mercado Mayorista. Technical report, Comisión de Regulación de Energía y Gas, Número 005.

CREG (2009b). Por la cual se modifica la Resolucin CREG-127 de 2009. Technical report, Comisión de Regulación de Energía y Gas, Resolución 159.

Dal Bo, P. (2005). Cooperation Under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games. *American Economic Review*, 95(5):1591–1604.

Dal Bo, P. and Frechette, G. R. (2018). On the Determinants of Cooperation in Infinitely Repeated Games: A Survey. *Journal of Economic Literature*, 56(1):60–114.

Djankov, S., La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2002). The Regulation of Entry. *The Quarterly Journal of Economics*, 117(1):1–37.

Duque, J. (2014). Gobernadores y Corrupcin en la Costa Atlántica. Clanes Políticos, Polticos de Negocios y Paramilitarismo. *Ciudad Paz-ando*, 7(2):174–200.

Fabra, N. and Reguant, M. (2014). Pass-through of Emissions Costs in Electricity Markets. *American Economic Review*, 104(9):2872–99.

Fabra, N. and Toro, J. (2005). Price Wars and Collusion in the Spanish Electricity Market. *International Journal of Industrial Organization*, 23(3-4):155–181.

Fioretti, M., He, J., and Tamayo, J. (2024). Saving for a Dry Day: Coal, Dams and the Energy Transition. Working paper.

Genesove, D. and Mullin, W. P. (2001). Rules, Communication, and Collusion: Narrative Evidence from the Sugar Institute Case. *American Economic Review*, 91(3):379–398.

Ghani, T. and Reed, T. (2022). Relationships on the rocks: Contract evolution in a market for ice. *American Economic Journal: Microeconomics*, 14(1):330–65.

Green, E. J. and Porter, R. H. (1984). Noncooperative Collusion under Imperfect Price Information. *Econometrica*, 52(1):84–100.

Green, R. and Newbery, D. M. (1992). Competition in the British Electricity Spot Market. *Journal of Political Economy*, 100(5):929–53.

Greenstone, M. (2014). Energy, growth and development. Evidence paper, International Growth Center.

Greenstone, M., Reguant, M., Ryan, N., and Dobermann, T. (2021). Energy & environment. Evidence paper, International Growth Center.

Harrington, J. (2008). Detecting Cartels. In Buccirossi, P., editor, *Handbook of Antitrust Economics*. MIT Press, Cambridge.

Hortacsu, A., Luco, F., Puller, S. L., and Zhu, D. (2019). Does Strategic Ability Affect Efficiency? Evidence from Electricity Markets. *American Economic Review*, 109(12):4302–42.

Hortacsu, A. and Puller, S. L. (2008). Understanding Strategic Bidding in Multi-unit Auctions: a Case Study of the Texas Electricity Spot Market. *The RAND Journal of Economics*, 39(1):86–114.

Igami, M. and Sugaya, T. (2021). Measuring the Incentive to Collude: The Vitamin Cartels, 199099. *The Review of Economic Studies*, 89(3):1460–1494.

Ishii, R. (2009). Favor exchange in collusion: Empirical study of repeated procurement auctions in Japan. *International Journal of Industrial Organization*, 27(2):137–144.

Kellogg, R. and Reguant, M. (2021). Energy and Environmental Markets, Industrial Organization, and Regulation. Working Paper 29235, National Bureau of Economic Research.

Knittel, C. and Stango, V. (2003). Price Ceilings as Focal Points for Tacit Collusion: Evidence from Credit Cards. *American Economic Review*, 93(5):1703–1729.

Leone, F., Macchiavello, R., and Reed, T. (2022). Market Size, Markups and International Price Dispersion in the Cement Industry. Technical report, Centre for Economic Performance.

Lipscomb, M., Mobarak, A. M., and Barham, T. (2013). Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil. *American Economic Journal: Applied Economics*, 5(2):200–231.

Liu, E. (2019). Industrial Policies in Production Networks. *The Quarterly Journal of Economics*, 134(4):1883–1948.

Macchiavello, R. (2022). Relational Contracts and Development. *Annual Review of Economics*, 14:337–362.

Macchiavello, R. and Morjaria, A. (2015). The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports. *American Economic Review*, 105(9):2911–45.

Macchiavello, R. and Morjaria, A. (2021). Competition and relational contracts in the rwanda coffee chain. *The Quarterly Journal of Economics*, 136(2):1089–1143.

McMillan, J. (1991). DANGO: Japan's Price-Fixing Conspiracies. *Economics & Politics*, 3(3):201–218.

Mirenda, L., Mocetti, S., and Rizzica, L. (2022). The Economic Effects of Mafia: firm Level Evidence. *American Economic Review*, 112(8):2748–73.

Mitton, T. (2008). Institutions and Concentration. *Journal of Development Economics*, 86(2):367–394.

Ortner, J. M., Chassang, S., Kawai, K., and Nakabayashi, J. (2022). Screening Adaptive Cartels. Working Paper 30219, National Bureau of Economic Research.

Perloff, J. M. and Carlton, D. W. (1999). *Modern industrial organization*. Addison-Wesley: Massachusetts.

Pesendorfer, M. (2000). A Study of Collusion in First-Price Auctions. *The Review of Economic Studies*, 67(3):381–411.

Porter, R. and Zona, J. (1993). Detection of Bid Rigging in Procurement Auctions. *Journal of Political Economy*, 101(3):518–38.

Porter, R. H. (2005). Detecting Collusion. *Review of Industrial Organization*, 26(2):147–167.

Porter, R. H. and Zona, J. D. (1999). Ohio School Milk Markets: An Analysis of Bidding. *The RAND Journal of Economics*, 30(2):263–288.

Rud, J. P. (2012). Electricity Provision and Industrial Development: Evidence from India. *Journal of Development Economics*, 97(2):352–367.

Ryan, N. (2021). The Competitive Effects of Transmission Infrastructure in the Indian Electricity Market. *American Economic Journal: Microeconomics*, 13(2):202–42.

Suarez, C. (2022). Private Management and Strategic Bidding Behavior in Electricity Markets: Evidence from Colombia. *Energy Economics*, 111:106058.

Suarez, C. (2023). Mixed Oligopoly and Market Power Mitigation: Evidence from the Colombian Wholesale Electricity Market. *The Journal of Industrial Economics*, 71(2):354–406.

Sugaya, T. and Wolitzky, A. (2018). Maintaining Privacy in Cartels. *Journal of Political Economy*, 126(6):2569–2607.

Superintendencia de Servicios Publicos (2008). Actos Relevantes del Mercado de Energía Mayorista. Technical report, Superintendencia de Servicios Publicos.

Superintendencia Delegada para Energa y Gas (2008). Resumen Acciones MEM Diciembre de 2008. Technical report, Superintendencia de Servicios Publicos.

Whinston, M. D. (2008). *Lectures on Antitrust Economics*. MIT Press, Cambridge, US.

Wolak, F. (2000). An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market. *International Economic Journal*, 14(2):1–39.

Wolak, F. (2007). Quantifying the Supplyside Benefits from Forward Contracting in Wholesale Electricity Markets. *Journal of Applied Econometrics*, 22:1179–1209.

Wolfram, C. D. (1998). Strategic Bidding in a Multiunit Auction: An Empirical Analysis of Bids to Supply Electricity in England and Wales. *RAND Journal of Economics*, 29(4):703–725.

Wolfram, C. D. (1999). Measuring Duopoly Power in the British Electricity Spot Market. *American Economic Review*, 89(4):805–826.

World Bank (2016). Breaking Down Barriers: Unlocking Africa's Potential through Vigorous Competition Policy. Technical report, World Bank.

World Bank (2019). Rethinking power sector reform in the developing world. Technical report, World Bank.

**CENTER DISSERTATION SERIES**

CentER for Economic Research, Tilburg University, the Netherlands

| No. | Author | Title | ISBN | Published |
|---|---|---|---|---|
| 672 | Joobin Ordoobody | The Interplay of Structural and Individual Characteristics | 978 90 5668 674 1 | February 2022 |
| 673 | Lucas Avezum | Essays on Bank Regulation and Supervision | 978 90 5668 675 8 | March 2022 |
| 674 | Oliver Wichert | Unit-Root Tests in High-Dimensional Panels | 978 90 5668 676 5 | April 2022 |
| 675 | Martijn de Vries | Theoretical Asset Pricing under Behavioral Decision Making | 978 90 5668 677 2 | June 2022 |
| 676 | Hanan Ahmed | Extreme Value Statistics using Related Variables | 978 90 5668 678 9 | June 2022 |
| 677 | Jan Paulick | Financial Market Information Infrastructures: Essays on Liquidity, Participant Behavior, and Information Extraction | 978 90 5668 679 6 | June 2022 |
| 678 | Freek van Gils | Essays on Social Media and Democracy | 978 90 5668 680 2 | June 2022 |
| 679 | Suzanne Bies | Examining the Effectiveness of Activation Techniques on Consumer Behavior in Temporary Loyalty Programs | 978 90 5668 681 9 | July 2022 |
| 680 | Qinnan Ruan | Management Control Systems and Ethical Decision Making | 978 90 5668 682 6 | June 2022 |
| 681 | Lingbo Shen | Essays on Behavioral Finance and Corporate Finance | 978 90 5668 683 3 | August 2022 |
| 682 | Joshua Eckblad | Mind the Gales: An Attention-Based View of Startup Investment Arms | 978 90 5668 684 0 | August 2022 |
| 683 | Rafael Greminger | Essays on Consumer Search | 978 90 5668 685 7 | August 2022 |
| 684 | Suraj Upadhyay | Essay on policies to curb rising healthcare expenditures | 978 90 5668 686 4 | September 2022 |

| No. | Author | Title | ISBN | Published |
|---|---|---|---|---|
| 685 | Bert-Jan Butijn | From Legal Contracts to Smart Contracts and Back Again: An Automated Approach | 978 90 5668 687 1 | September 2022 |
| 686 | Sytse Duiverman | Four essays on the quality of auditing: Causes and consequences | 978 90 5668 688 8 | October 2022 |
| 687 | Lucas Slot | Asymptotic Analysis of Semidefinite Bounds for Polynomial Optimization and Independent Sets in Geometric Hypergraphs | 978 90 5668 689 5 | September 2022 |
| 688 | Daniel Brosch | Symmetry reduction in convex optimization with applications in combinatorics | 978 90 5668 690 1 | October 2022 |
| 689 | Emil Uduwalage | Essays on Corporate Governance in Sri Lanka | 978 90 5668 691 8 | October 2022 |
| 690 | Mingjia Xie | Essays on Education and Health Economics | 978 90 5668 692 5 | October 2022 |
| 691 | Peerawat Samranchit | Competition in Digital Markets | 978 90 5668 693 2 | October 2022 |
| 692 | Jop Schouten | Cooperation, allocation and strategy in interactive decision-making | 978 90 5668 694 9 | December 2022 |
| 693 | Pepijn Wissing | Spectral Characterizations of Complex Unit Gain Graphs | 978 90 5668 695 6 | November 2022 |
| 694 | Joris Berns | CEO attention, emotion, and communication in corporate financial distress | 978 90 5668 696 3 | November 2022 |
| 695 | Tom Aben | The (long) road towards smart management and maintenance: Organising the digital transformation of critical infrastructures | 978 90 5668 697 0 | December 2022 |
| 696 | Gülbike Mirzaoğlu | Essays in Economics of Crime Prevention and Behavior Under Uncertainty | 978 90 5668 698 7 | February 2023 |
| 697 | Suwei An | Essays on incentive contracts, M&As, and firm risk | 978 90 5668 699 4 | February 2023 |
| 698 | Jorgo Goossens | Non-standard Preferences in Asset Pricing and Household Finance | 978 90 5668 700 7 | February 2023 |

| No. | Author | Title | ISBN | Published |
|---|---|---|---|---|
| 699 | Santiago Bohorquez Correa | Risk and rewards of residential energy efficiency | 978 90 5668 701 4 | April 2023 |
| 700 | Gleb Gertsman | Behavioral Preferences and Beliefs in Asset Pricing | 978 90 5668 702 1 | May 2023 |
| 701 | Gabriella Massenz | On the Behavioral Effects of Tax Policy | 978 90 5668 703 8 | May 2023 |
| 702 | Yeqiu Zheng | The Effect of Language and Temporal Focus on Cognition, Economic Behaviour, and Well-Being | 978 90 5668 704 5 | May 2023 |
| 703 | Michela Bonani | Essays on Innovation, Cooperation, and Competition Under Standardization | 978 90 5668 705 2 | June 2023 |
| 704 | Fabien Ize | The Role of Transparency in Fairness and Reciprocity Issues in Manager-Employee Relationships | 978 90 5668 706 9 | June 2023 |
| 705 | Kristel de Nobrega | Cyber Defensive Capacity and Capability: A Perspective from the Financial Sector of a Small State | 978 90 5668 707 6 | July 2023 |
| 706 | Christian Peters | The Microfoundations of Audit Quality | 978 90 5668 708 3 | June 2023 |
| 707 | Felix Kirschner | Conic Optimization with Applications in Finance and Approximation Theory | 978 90 5668 709 0 | July 2023 |
| 708 | Zili Su | Essays on Equity Incentive and Share Pledging in China | 978 90 5668 710 6 | September 2023 |
| 709 | Rafael Escamilla | Managing the Nanostore Supply Chain: Base-of-the-Pyramid Retail in Emerging Markets | 978 90 5668 711 3 | September 2023 |
| 710 | Tomas Jankauskas | Essays in Empirical Finance | 978 90 5668 712 0 | August 2023 |
| 711 | Tung Nguyen Huy | Fostering Sustainable Land Management in Sub-Saharan Africa: Evidence from Ghana and Burkina Faso | 978 90 5668 713 7 | September 2023 |
| 712 | Daniel Karpati | Essays in Finance & Health | 978 90 5668 714 4 | September 2023 |
| 713 | Mylène Struijk | IT Governance in the Digital Era: Insights from Meta-Organizations | 978 90 5668 715 1 | September 2023 |

| No. | Author | Title | ISBN | Published |
|---|---|---|---|---|
| 714 | Albert Rutten | Essays on Work and Retirement | 978 90 5668 716 8 | November 2023 |
| 715 | Yan Liu | Essays on Credit Rating Agencies in China | 978 90 5668 717 5 | October 2023 |
| 716 | Xiaoyue Zhang | Distortions and industrial upgrading in China | 978 90 5668 718 2 | September 2023 |
| 717 | Andries van Beek | Solutions in multi-actor projects with collaboration and strategic incentives | 978 90 5668 719 9 | October 2023 |
| 718 | Andries Steenkamp | Polynomial Optimization: Matrix Factorization Ranks, Portfolio Selection, and Queueing Theory | 978 90 5668 720 5 | October 2023 |
| 719 | Luis Vargas | Sum-of-Squares Representations for Copositive Matrices and Independent Sets in Graphs | 978 90 5668 721 2 | November 2023 |
| 720 | Zhenshu Wu | Essays in Corporate Finance and ESG | 978 90 5668 722 9 | November 2023 |
| 721 | Frank de Meijer | Integrality and Cutting Planes in Semidefinite Programming Approaches for Combinatorial Optimization | 978 90 5668 723 6 | November 2023 |
| 722 | Chayanin Wipusanawan | Standard-Essential Patents, Innovation, and Competition | 978 90 5668 724 3 | November 2023 |
| 723 | Ke Wang | Essays in Corporate Risk | 978 90 5668 725 0 | November 2023 |
| 724 | Mustafa Ahci | Essays on corporate disclosures, innovation, and investments | 978 90 5668 726 7 | November 2023 |
| 725 | Frederique van Leeuwen | Motifs: From Theory to Practice | 978 90 5668 727 4 | December 2023 |
| 726 | Lkhagvaa Erdenesuren | Expectation, Anticipation, and Identification: Essays on Subjective Expectations and Economic Decision-Making | 978 90 5668 728 1 | December 2023 |
| 727 | Wouter van Eekelen | Distributionally Robust Views on Queues and Related Stochastic Models | 978 90 5668 729 8 | December 2023 |
| 728 | Jia Bi | Essays on Extreme Returns and Investor Behavior | 978 90 5668 730 4 | January 2024 |

| No. | Author | Title | ISBN | Published |
|---|---|---|---|---|
| 729 | Jesse van der Geest | Economic Effects of Tax Avoidance and Compliance | 978 90 5668 731 1 | January 2024 |
| 730 | Zeng Yu | Essays on Incentive Contract and Corporate Finance | 978 90 5668 732 8 | January 2024 |
| 731 | Joost de Kruijff | Commitment-Based Smart Contracts using CommitRuleML | 978 90 5668 733 5 | February 2024 |
| 732 | Simona Hannon | Essays On Consumer Finance | 978 90 5668 734 2 | January 2024 |
| 733 | Bart Dees | Individualized Pension Contracts: Risks, Welfare Losses and Investment Choices | 978 90 5668 735 9 | April 2024 |
| 734 | José Gabriel Carreño Bustos | Three Essays on Wage Compensation and Flexible Contracts | 978 90 5668 736 6 | March 2024 |
| 735 | Lennart Dekker | Essays on Asset Liquidity and Investment Funds | 978 90 5668 737 3 | April 2024 |
| 736 | Valentijn Stienen | Analytics for Humanitarian Networks in Uncertain and Data-Scarce Environments | 978 90 5668 738 0 | May 2024 |
| 737 | Constanza Martinez Ventura | Essays on Liquidity Provision in Wholesale Funding Markets and on Financial Fragility | 978 90 5668 739 7 | May 2024 |
| 738 | Kartik Chawla | Guarding the Digital Cookie Jar: An Interdisciplinary Study of Automated Privacy Preference Negotiation, Monitoring, and Enforcement | 978 90 5668 740 3 | May 2024 |
| 739 | Koen Peters | Operationalizing analytics to improve food security | 978 90 5668 741 0 | June 2024 |
| 740 | Joyce Kox | Navigating Crises in Family Businesses: Unraveling the Role of the Family | 978 90 5668 742 7 | June 2024 |
| 741 | Hazal Sezer | Historical and Contemporary Perspectives in Labor | 978 90 5668 743 4 | June 2024 |
| 742 | Hadi Abbaszadehpeivasti | Performance Analysis of Optimization Methods for Machine Learning | 978 90 5668 744 1 | October 2024 |
| 743 | Jeroen Verbouw | Essays in Entrepreneurial Finance | 978 90 5668 745 8 | September 2024 |
| 744 | Meike Reusken | Optimization under uncertainty for Food Security | 978 90 5668 746 5 | Augustus 2024 |

| No. | Author | Title | ISBN | Published |
|---|---|---|---|---|
| 745 | Martijn Ketelaars | Clearing in Financial Networks and Dynamic Investment under Uncertainty | 978 90 5668 747 2 | Augustus 2024 |
| 746 | Farzan Faninam | Essays on Real Options: Triopoly Dynamics, Disconnected Investment Regions, and Multiple Lumpy Investment | 978 90 5668 748 9 | September 2024 |
| 747 | Mario Bernasconi | Essays on Labour Economics and Industrial Organization | 978 90 5668 749 6 | September 2024 |

This thesis is a collection of four self-contained papers on topics related to labour economics and industrial organization. In the first part of the thesis, I study how different components of a welfare state, namely disability insurance and pensions, affect individual and household labour supply decisions. I show how the decision to work – and how much to work – is influenced by the incentives embedded in these schemes, which provides guidance on how to design them to increase efficiency and people's well-being. In the second part of the thesis, instead, I study how the institutional features of a market can facilitate or hinder collusion between firms. I show that, in some cases, excessive market transparency makes collusion easier to sustain and ultimately lowers consumers' welfare.

Mario Bernasconi (born in Gallarate, 1994) obtained his Bachelor's degree in Economics and Management from Università degli Studi dell'Insubria in 2012 and his Master's degree in Economic and Social Sciences from Università Bocconi in 2015. After working in consulting for a year, he moved to Tilburg in 2016. Upon completing the Research Master in Economics at Tilburg University in 2018, Mario joined the Department of Econometrics and Operations Research as a Ph.D. candidate under the supervision of Arthur van Soest, Tunga Kantarcı, and Alexandros Theloudis.

ISBN: 978 90 5668 749 6
DOI: 10.26116/tisem.41560108

TILBURG SCHOOL OF ECONOMICS
AND MANAGEMENT

CentER Dissertation Series