Supplementing expenditure surveys with unlinkable administrative data

An application to household consumption behaviour in The Netherlands[1][2]

Pim Kastelein

University of Amsterdam, Tinbergen Institute

29th October 2020

[1] Netspar Online Pension Day 2020

[2] Part of NWO project 'Pension funding, housing wealth and macroeconomic demand'

Introduction

- Heterogeneity in household consumption behaviour central to modern macromodels (Kaplan & Violante, 2018)
- Marginal propensity to consume (MPC) out of wealth/income
- Attention to balance sheet effects on consumption for amplification of shocks (e.g. liquidity, leverage)
- Empirically, data availability is a concern:
  - Expenditure surveys lack information on income and wealth
  - Administrative data lack information on consumption
- Imputation required, but which way and do the techniques work?

This paper

- Combine all information on expenditures, income and wealth available at Statistics Netherlands (CBS)
- Try one novel and one widely used imputation technique to fill data gaps:
  - Supplement expenditure survey with unlinkable administrative data on income and wealth using household matching
  - Impute consumption in administrative data using household budget constraint: $c = y - \Delta w$
- Evaluate fit using perfect link in 2015
- Cautiously estimate relation between c, y and h (liquidity, leverage, pension pressure)

Main findings

1. Many higher earning households have:
   - High mortgage indebtedness
   - Low levels of liquidity
   - High mandatory pension fund contributions
2. Household matching works pretty well
3. Consumption imputation implies too many hand-to-mouth households
4. Both imputation techniques identify higher MPC's for constrained households

Literature

- Comparison expenditure survey with consumption imputation:
  Koijen et al. (2014), Kreiner et al. (2014), Abildgren et al. (2018)
- Consumption imputation using budget constraint:
  Browning & Leth-Petersen (2003), Baker et al. (2018), Eika et al. (2020), Fagereng & Halvorsen (2017)
- Role of leverage and liquidity for household consumption:
  Campbell & Cocco (2007), Bunn & Rostom (2015), Mian & Sufi (2011), Cooper (2013)
- Similar studies with data from Statistics Netherlands:
  Bijlsma & Mocking (2017), Ji et al. (2019), Zhang (2019)

Constraints in universe of Dutch households (2015)

Mortgage debt overhang and borrowing constraints imply deviations from permanent income hypothesis

Exacerbated by pension system:

- Mandatory pension fund contributions scale with income
- Household primarily invested in illiquid housing and pensions
- Less liquidity and ability to smooth out shocks

Constraints in universe of Dutch households (2015)

Variable	*Quintile*	*Unconstrained*	*Liquidity*	*Leverage*	*Pension pressure*	*Jointly*
Disposable income	1^{st}	60.2%	38.3%	2.4%	0.2%	0.0%
	2^{nd}	65.4%	30.2%	6.4%	1.8%	0.1%
	3^{rd}	55.8%	31.1%	19.7%	5.7%	0.9%
	4^{th}	46.8%	28.3%	31.7%	14.8%	2.6%
	5^{th}	44.6%	20.7%	33.9%	25.1%	3.5%
Net wealth	1^{st}	8.4%	59.1%	63.9%	17.2%	6.6%
	2^{nd}	36.7%	56.7%	6.7%	4.1%	0.2%
	3^{rd}	69.0%	13.2%	12.9%	10.6%	0.1%
	4^{th}	73.8%	13.2%	6.2%	10.3%	0.2%
	5^{th}	84.9%	6.6%	4.4%	5.5%	0.0%

Liq. $= 1$ if $\frac{\text{bank account}}{\text{income}} < \frac{1}{12}$, Lev. $= 1$ if $\frac{\text{mortgage}}{\text{house value}} > 1$, Pen. press. $= 1$ if $\frac{\text{pf contrib.}}{\text{income}} > 75^{\text{th}}$ pctile

Data availability

Administrative data:

- Available from 2011-2018
- Universe of Dutch households
- Constructed from tax returns and municipality registers
- Linkable datasets on household characteristics, income, assets, liabilities, residential location
- No consumption

Expenditure survey:

- Conducted in 2012, 2013 (unlinkable) and 2015 (linkable)
- 6000, 5000 and 14000 Dutch households
- Households record purchases in diaries and fill in additional questionnaires
- No detailed information on income and wealth

Income and wealth imputation

- Match ES households to AD households using overlapping variables:
  - Categorical variables exactly *(e.g. composition, age cohort, municipality code, primary income source)*
  - Continuous variables with bandwidths *(e.g. income, value of house, paid interest on mortgage)*
- Overlap 81%-99% between two data sources
- 80% of ES households matched
- 65% correct identification
- Median of two AD households per ES household
- No sample selection

Imputation errors

Median absolute errors small, but some large outliers

Variable	*Mean*	*Median*	*p95*
Disposable income	1,618	325	7,100
Bank account balance	17,758	3,554	76,833
Outstanding mortgage	11,608	0	58,607
Pension fund contributions	216	0	1,139

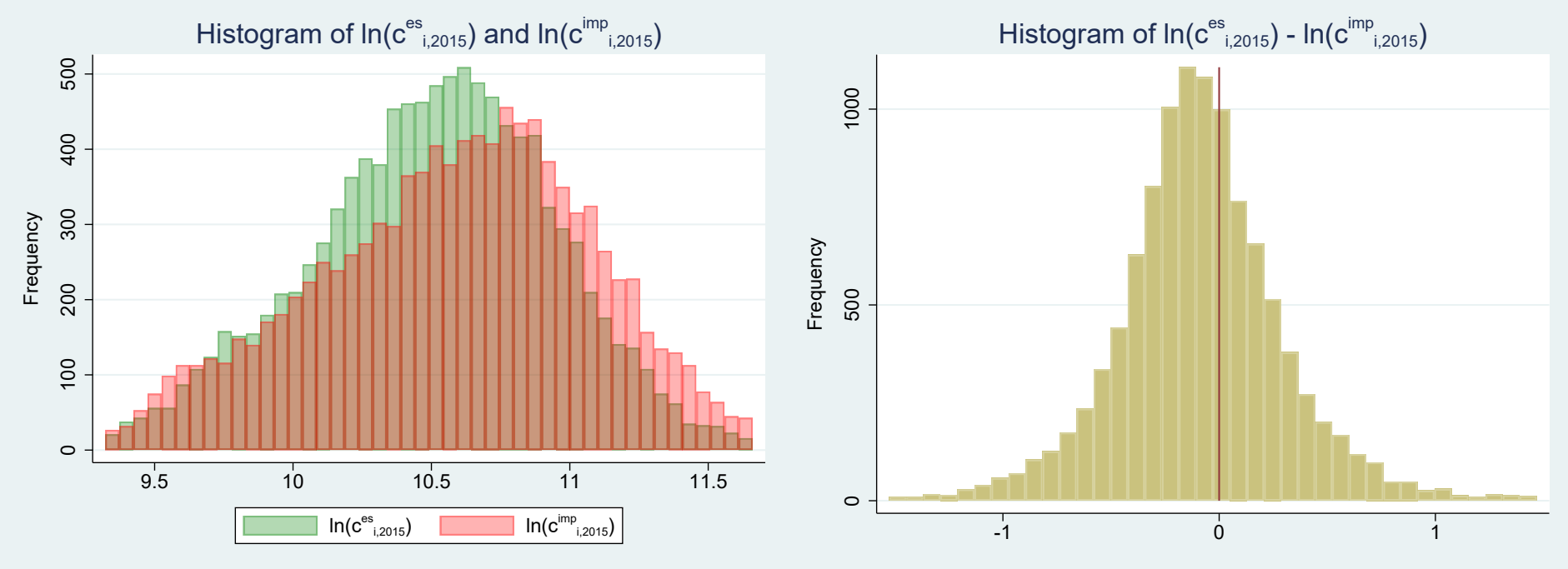
Consumption imputation

- Household budget constraint:

$$c_{i,t} = y_{i,t} - \Delta a_{i,t} + \Delta l_{i,t} - t^{g}_{i,t} + t^{r}_{i,t}$$

- Difficulty: is $\Delta a_{i,t}$ and $\Delta l_{i,t}$ due to change in price or quantity?
- Price $\neq$ consumption, quantity $=$ consumption
  - Bank account, stocks, bonds: apply national mutation
  - Housing wealth: only if move
  - Other asset categories: abs. growth rate $> 15\%$ is quantity
  - No price correction for liabilities
- Shortcoming: $c_{i,t}$ is total consumption

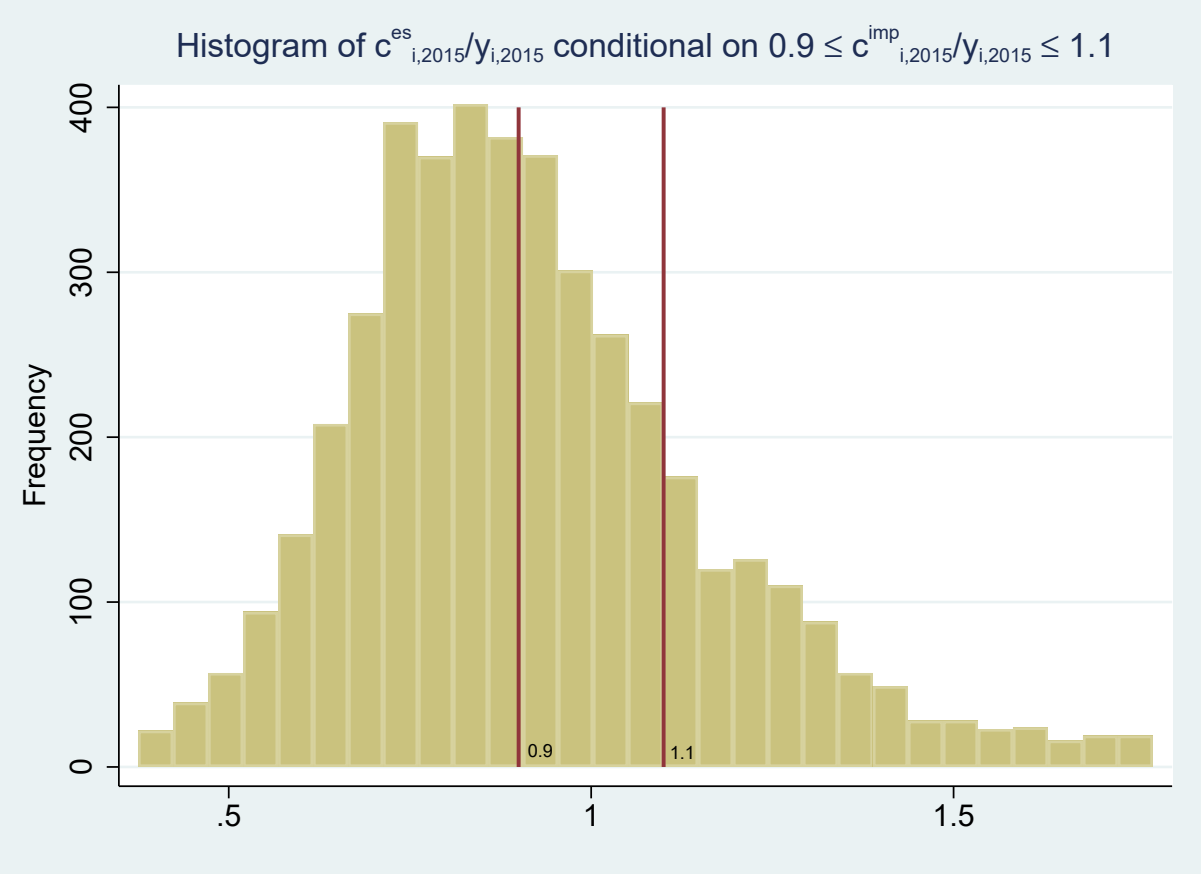
Imputation errors



- Regressing $\ln c^{imp}_{i,2015}$ on $\ln c^{es}_{i,2015}$: $\beta = 0.83$
- Imputation error correlated with income, age, net wealth, home ownership status, family size (even after controlling for income)

Implied consumption rates

4500 households hand-to-mouth consumers with $0.9 \leq \frac{c}{y} \leq 1.1$ according to consumption imputation, while 1150 according to ES:



Regression framework

$$\Delta c_{i,t} = \alpha_i + \beta_1 \Delta y_{i,t} + \beta_2 \Delta h_{i,t} + \beta_3 Z_{i,t} + \varepsilon_{i,t}$$

- Structural form from Campbell & Cocco (2007)
- $\Delta y_{i,t}$ and $\Delta h_{i,t}$ interacted with indicators of home ownership status, liquidity, leverage, pension pressure
- True panel of $\sim$1 million households with AD
- Pseudo-panel with three cross-sections of ES (Deaton, 1985)
- Fixed effects with robust SE's

Regression results with expenditure survey

Independent variable	*(i)*	*(ii)*	*(iii)*	*(iv)*	*(v)*
$\Delta y_{i,t}$	0.218***	0.219***	0.198**	0.204**	0.198**
	(0.059)	(0.057)	(0.066)	(0.062)	(0.061)
$\Delta y_{i,t} \times$ Liq. Constr.			0.102		
			(0.131)		
$\Delta y_{i,t} \times$ Lev. Constr.				0.136	
				(0.131)	
$\Delta y_{i,t} \times$ Pen. Constr.					0.181
					(0.139)
$\Delta h_{i,t} \times$ Own	0.276***				
	(0.052)				
$\Delta h_{i,t} \times$ Rent	0.075				
	(0.055)				
$\Delta h_{i,t}$		0.185***	0.203***	0.167***	0.140***
		(0.039)	(0.047)	(0.043)	(0.038)
$\Delta h_{i,t} \times$ Liq. Constr.			-0.059		
			(0.080)		
$\Delta h_{i,t} \times$ Lev. Constr.				0.098	
				(0.094)	
$\Delta h_{i,t} \times$ Pen. Constr.					0.190
					(0.111)
R^2	0.161	0.175	0.163	0.167	0.180
N	1,521	1,521	1,521	1,521	1,521

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, control variables omitted

Regression results with administrative data

Independent variable	*(i)*	*(ii)*	*(iii)*	*(iv)*	*(v)*
$\Delta y_{i,t}$	0.550***	0.484***	0.573***	0.556***	0.510***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
$\Delta y_{i,t} \times$ Liq. Constr.		0.176***			0.177***
		(0.004)			(0.004)
$\Delta y_{i,t} \times$ Lev. Constr.			-0.123***		-0.123***
			(0.006)		(0.006)
$\Delta y_{i,t} \times$ Pen. Constr.				-0.067***	-0.029***
				(0.008)	(0.008)
$\Delta h_{i,t}$	0.014**	-0.076***	-0.007	0.006	-0.076***
	(0.005)	(0.006)	(0.0061)	(0.006)	(0.007)
$\Delta h_{i,t} \times$ Liq. Constr.		0.380***			0.386***
		(0.01)			(0.011)
$\Delta h_{i,t} \times$ Lev. Constr.			0.080***		-0.022*
			(0.009)		(0.009)
$\Delta h_{i,t} \times$ Pen. Constr.				0.048***	0.034**
				(0.011)	(0.012)
R^2	3,364,090	3,364,090	3,364,090	3,364,090	3,364,090
N	0.062	0.064	0.062	0.062	0.064

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, control variables omitted

Wrapping up

Conclusions:

- Proof of concept for household matching
- Potentially useful elsewhere (e.g. some waves of Danish ES)
- Not enough waves of Dutch ES to apply advanced techniques

To do:

- Asymmetric responses to $\Delta y_{i,t}$ and $\Delta h_{i,t}$
- Interactions with age and year dummies
- Decompose $\Delta y_{i,t}$ and $\Delta h_{i,t}$ into expected and unexpected parts (Browning et al., 2013)
- Structural approach to estimate MPC using variance-covariance restrictions (Blundell et al., 2008)

References I

Abildgren, K., Kuchler, A., Rasmussen, A. S. L., & Sørensen, H. S. (2018). *Consistency between household-level consumption data from registers and surveys*. Danmarks Nationalbank Working Papers 131, Danmarks Nationalbank.

Baker, S. R., Kueng, L., Meyer, S., & Pagel, M. (2018). *Measurement Error in Imputed Consumption*. Working Paper 25078, National Bureau of Economic Research.

Bijlsma, M. & Mocking, R. (2017). The impact of house price shocks on the savings of dutch homeowners and renters. *CPB Discussion Paper 346*.

Blundell, R., Pistaferri, L., & Preston, I. (2008). Consumption inequality and partial insurance. *The American Economic Review*, 98(5), 1887–1921.

Browning, M., Gørtz, M., & Leth-Petersen, S. (2013). Housing wealth and consumption: A micro panel study. *The Economic Journal*, 123(568), 401–428.

References II

Browning, M. & Leth-Petersen, S. (2003). Imputing consumption from income and wealth information. *The Economic Journal*, 113(488), F282–F301.

Bunn, P. & Rostom, M. (2015). *Household debt and spending in the United Kingdom*. Bank of England working papers 554, Bank of England.

Campbell, J. Y. & Cocco, J. F. (2007). How do house prices affect consumption? Evidence from micro data. *Journal of Monetary Economics*, 54(3), 591–621.

Cooper, D. (2013). House price fluctuations: the role of housing wealth as borrowing collateral. *The Review of Economics and Statistics*, 95(4), 1183–1197.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30(1), 109–126.

Eika, L., Mogstad, M., & Vestad, O. L. (2020). What can we learn about household consumption expenditure from data on income and assets? *Journal of Public Economics*, (pp. 104–163).

References III

Fagereng, A. & Halvorsen, E. (2017). Imputing Consumption from Norwegian Income and Wealth Registry Data. *Journal of Economic and Social Measurement*, 42(1), 67–100.

Ji, K., Teulings, R., & Wouterse, B. (2019). Disentangling the effect of household debt on consumption. *CPB Discussion Paper 395*.

Kaplan, G. & Violante, G. L. (2018). Microeconomic heterogeneity and macroeconomic shocks. *Journal of Economic Perspectives*, 32(3), 167–94.

Koijen, R., Van Nieuwerburgh, S., & Vestman, R. (2014). Judging the quality of survey data by comparison with "truth" as measured by administrative records: Evidence from sweden. In *Improving the Measurement of Consumer Expenditures* (pp. 308–346). National Bureau of Economic Research, Inc.

References IV

Kreiner, C., Lassen, D., & Leth-Petersen, S. (2014). Measuring the accuracy of survey responses using administrative register data: Evidence from denmark. In *Improving the Measurement of Consumer Expenditures* (pp. 289–307). National Bureau of Economic Research, Inc.

Mian, A. & Sufi, A. (2011). House prices, home equity-based borrowing, and the us household leverage crisis. *American Economic Review*, 101(5), 2132–56.

Zhang, L. (2019). Do house prices matter for household consumption? *CPB Discussion Paper 396*.