Zhenzhen Fan
Trading Strategies Based on Yield Curve Forecasting Models Using Macroeconomic Data

MSc Thesis 2011-073
Trading strategies based on yield curve forecasting models using macroeconomic data

by

Zhenzhen Fan (295896)
(Msc. Tilburg University 2011)

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Quantitative Finance and Actuarial Science

Faculty of Economics and Business Administration
Tilburg University

Supervisors
Dr. Roger Laeven
Dr. Rob van den Goorbergh
Dr. Roy Hoevenaars

Defended on: 26-08-2011
# Contents

1 Introduction 5

2 Interest rate models and macro data incorporation methods 8
   2.1 Three Pass Regression Filter: A Primer 8
   2.2 AR model 9
   2.3 VAR model 11
   2.4 Dynamic Nelson Siegel model 11
   2.5 Affine Term Structure model 13

3 Data 15
   3.1 Yield data 15
   3.2 Macro economic data 15
   3.3 Futures price data 17

4 Estimation 19
   4.1 Monte Carlo Simulation Results 19
   4.2 Estimation techniques 20
   4.3 Estimation results 21
      4.3.1 AR and VAR model 21
      4.3.2 DNS and ATSM 22

5 Forecasting 26
   5.1 Forecast procedure 26
   5.2 Forecast evaluation 26
   5.3 Forecast combination 28
   5.4 Forecast results 29

6 Evaluation of interest rate models and macro data incorporation methods 36
   6.1 Interest rate models 36
   6.2 Macro data incorporation methods 37
      6.2.1 A comparison of PCA, Kalman Filter and 3PRF factor
            extraction methods 38
      6.2.2 A comparison between 3PRFRF and PCAF 44

7 Trading Strategies 47
   7.1 Adjusting for risk exposure 47
   7.2 Trading strategies based on individual model forecasts 47
   7.3 Trading strategy based on the Model Confidence Set selection 48
7.4 Performance of trading strategies .................................................. 49

8 Analysis of interest rate models in the period of financial crisis 54

9 Conclusion ......................................................................................... 60

A Monte Carlo simulations ..................................................................... 62

B PC Augmented Forecasting approach of incorporating macro factors ........................................................................... 64

C Kalman Filter estimation of macro factors ........................................ 66

D List of macro series in the macroeconomic dataset .......................... 67
Acknowledgments

I'd love to express my sincere gratitude to my University supervisor, Dr. Roger Laeven, for his perfectionism down to the last detail and sharp critics that fostered a scientific and academic attitude inside of me. I am heartily thankful to my company supervisors, Dr. Rob van den Goorbergh and Dr. Roy Hoevenaars, who patiently and earnestly guided me with their abundant experience and expertise from the initial to the final level through regular weekly meetings, crash meetings and immediate consultations. I am also grateful to Dr. P.J. van der Sluis who introduced the techniques of Ssf package, which has been intensively used in my research. I owe my deep gratitude to Professor Hans Schumacher, without whom I wouldn’t even have started this research project. This thesis would not have been possible without the sufficient financial and technical support provided by APG. In particular, I would like to thank Louis Chailllet who kindly shared with me his programming and helped me with the many software problems I came across; Ruben Laros, who organized the intermediate presentations and reports which kept me on the timeline. I would also like to show my gratitude to the senior supervisors in APG who provided useful comments during presentations. Furthermore, I am indebted to my many friends who encouraged and motivated me all along and gave their hands whenever and wherever. Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the project.
Abstract

This paper aims at constructing profitable trading strategies based on yields forecasts from interest rates models with macroeconomic data. I first construct and estimate AR, VAR, Dynamic Nelson Siegel (DNS) and Affine Term Structure Model (ATSM). The Monte Carlo simulation results together with the estimation results suggest that little benefit is achieved by adding macro factors to the DNS and ATSM in the estimation stage. In this regard, I propose a new way of exploiting predictability contained in macro factors, the 3PRFRF approach, to extract efficient macro information. I evaluate the forecast performance of each model with and without macro factors and combine forecasts using the Model Confidence Set (MCS) to arrive at an optimal forecast that I show to be able to out-perform the Random Walk. I give a comprehensive evaluation of the interest rate models as well as different macro factor incorporation approaches based on the estimation and forecasting results. I show that the 3PRF residual forecasting approach performs better than conventional principal components augmented forecasting approach. Furthermore, I construct trading strategies using individual model forecasts and use Model Confidence Set to yield combined trading strategies. I show that macro factors play a crucial role in the profitability of trading strategies. Lastly, I apply the forecast model to the period of recent financial crisis and find that neither no-arbitrage assumptions nor macro fundamentals significantly improve the forecast or the trading strategies performance as they do in the pre-crisis period.
Chapter 1

Introduction

There are a large number of models for interest rate forecasting. Starting from the time series models, there are Autoregressive (AR) and Vector Autoregressive (VAR) models. AR explains the current yield through its long term mean and its own persistency. If the yield process is perfectly persistent, then it is said to have a unit root and behaves like a Random Walk. Random walk suggests that the current level of yields is the best predictor for the future, implying unpredictability of future interest rate. The VAR model goes one step further than AR by stacking yields of various tenors, which is another name for maturities, into a vector and regressing each one, not only on its own, but also on the other variables’ lags. VAR allows for interactions between the cross section of interest rates so that shocks to one of the tenors influence other tenors as well as itself.

Turning to more structured models, Nelson and Siegel (1987) proposed a three-factor parsimonious model that produces a well in-sample fit of a large range of yield curves. In order to facilitate forecasts with the Nelson Siegel model, Diebold and Li (2006) show how to extend the Nelson Siegel model to a Dynamic Nelson Siegel by treating the latent factors as time varying. Similar to principal components, which are found to convey the level, slope and curvature of the yield curve (Littleman and Scheinkman 1991), Diebold and Li (2006) recognize that the three latent factors capture the same elements. The last model considered in this paper is the Affine Term Structure Model proposed in Cox, Ingersoll, and Ross (1985)’s milestone paper. Unlike previous models which are only based on statistical properties, the affine models are derived from economic theory. The no-arbitrage opportunity assumption gives rise to cross section restrictions. Researchers have been working on sensible and stable ATSM specifications. For example, Chen and Scott (1993) explicitly derive the state variables by assuming some of the yield tenors to be measured without error. Later the specification of the affine model family is comprehensively analyzed by Dai and Singleton (2000). Duffee (2002) further relaxes the risk premium specification and come up with “essentially affine” models. Ait-Sahalia and Kimmel (2005) show how to derive the closed-form likelihood expansion to stabilize the estimation procedure.

Macroeconomic data have been documented to contain forecast power for interest rates (Stock and Watson 2002b). Since the number of variables contained in the macro data is typically large, including macro data in the regression model
either involves picking up the most useful indicators based on prior knowledge or summarizing the information into a few manageable factors. Examples of the first approach are Cox, Ingersoll, and Ross (1985), Ang and Piazzesi (2003), Diebold, Rudebusch, and Aruoba (2006), and Moench (2008). Ludvigson and Ng (2009) and De Pooter, Ravazzolo, and van Dijk (2010) demonstrate the application of the second approach. Principal Component Analysis (PCA) is so far the most widely used approach in extracting factors. Forecasting using PCA is extensively discussed by Stock and Watson (2002a, 2002b, 2005). Joslin, Le, and Singleton (2011) estimate latent factors using the Kalman Filter that allows for missing data and different sampling frequencies. Stock and Watson (2002b) show that the expectation maximization algorithm is also competent in this task. A systematic comparison of methods for forecasting using many predictors can be found in Stock and Watson (2004).

The basis of the thesis is De Pooter, Ravazzolo, and van Dijk (2010). They compare the forecast performance of different interest rate models and find that the principal components of macro data improve interest rates forecasts. They also show that combining individual model’s forecasts mitigates model uncertainty and leads to consistent prediction improvement. Based on De Pooter, Ravazzolo, and van Dijk (2010)’s findings, I improve and extend the framework in a number of ways. I start with the construction, estimation and forecast of generic AR, VAR, dynamic Nelson Siegel and the affine models. In a second stage, I include macro factors in the generic models by applying the Three Pass Regression Filter proposed by Kelly and Pruitt (2011) to extract information which is useful in prediction but not overlapping with information embedded in the yields. Thirdly, I combine individual forecasts from models to improve forecast performance. Two schemes are used in the forecast combinations, which are inverted mean squared prediction error weighted scheme and Model Confidence Set (Hansen, Lunde, and Nason 2005). I evaluate individual forecasts as well as forecast combinations with several error metrics, such as mean square prediction error, the hit ratio and model survival ratio. I show that the forecast combinations out-perform Random Walk with respect to root mean squared prediction error and that macro factors improve the direction forecast performance significantly. Fourthly, I design long and short trading strategies on 2-year, 5-year and 10-year futures contracts using individual interest rate forecasts. The Model Confidence Set is once again applied to produce an combined trading strategies out of individual trading strategies. The trading strategies are evaluated with information ratios and cumulated profits and loss in which I show that macroeconomic information contributes to the profitability of trading strategies. Finally, I extend the framework to the recent financial crisis and find that structured models which used to deliver better forecasts underperform simple models in that period. In addition, the improvement resulting from macro factors, although still noticeable, is not as significant as in the pre-crisis period.

The thesis contributes to the existing literature in four aspects. First, in addition to the forecast performance of interest rate models, I also address the concern of parameter identification and estimates’ efficiency. I employ Monte Carlo simulation to recover the data generating parameters of DNS and ATSM. The Monte Carlo simulation results show that the numerical optimization process gets highly unstable with the additional parameters brought by macro factors. In this regard, I impose necessary restrictions to reduce the number of parameters to be estimated and exclude the macro factors from the estimation of DNS
and ATSM. Moreover, I find that the two step estimation in the affine model suggested by Ang and Piazzesi (2003) leads to large standard errors due to the correction for first step estimates, therefore I turn to the one step Kalman Filter estimation for efficiency consideration. The one-step Kalman Filter estimation approach also circumvents the necessity to impose additional assumptions on the state variables. It turns out that the latent states of DNS and ATSM happen to have significant co-movement, which suggests that the no-arbitrage assumption is embedded in the DNS to a reasonable approximation.

Second, I incorporate macro factors in a different way, namely, the 3PRF residual forecast approach. The idea is to extract information from macro data that is not only relevant but also orthogonal to existing regressors. I adopt the Three Pass Regression Filter (3PRF) proposed by Kelly and Pruitt (2011) to ensure relevance and use the regression residuals to ensure orthogonality. Furthermore, motivated by the findings that the generic DNS and ATSM already produce almost perfect in-sample fit, I postpone the incorporation of macro factors to the forecast stage from the estimation stage since there is essentially little to be improved by adding the macro factors in the estimation stage. Moreover, excluding macro factors from estimation to limit the number of parameters is also in line with the concern of parameter identification. In addition, I use direction forecasts, which suggests to find explanatory power of the current yields in the h-step lagged value, instead of iterated forecasts, which is to construct next period forecast on the basis of the prediction of last period, to avoid having a time series model to forecast the macro factors. The direct forecast approach does not require to transform the macro data to ensure stationarity, the process of which, might result in a substantial loss of information. I show the 3PRF residual forecast approach performs better than the PC augmented forecast approach used by De Pooter, Ravazzolo, and van Dijk (2010).

Thirdly, the estimation and forecast results from different models with and without macro factors, together with the models’ performance in the period of financial crisis enable me to draw insightful conclusions about the mechanism as well as the pros and cons of interest rate models.

Lastly, I apply the forecasts to design the trading strategies. The evaluation of the performance of trading strategies provides a different dimension of evaluating forecasts by emphasizing more on the direction accuracy than the absolute prediction errors. I also show techniques to combine individual strategies given the historical performance of each trading strategy.

The rest of the thesis is organized as follows. Chapter 2 gives an introduction to the Three Pass Regression Filter method and the specifications of the models used for estimation and forecasting. Chapter 3 describes the data used in the empirical studies. Chapter 4 gives details about the Monte Carlo simulation results, the estimation techniques and estimation results. Chapter 5 looks into the forecast procedure. Chapter 6 evaluates the interest rate models and macro factors extraction methods given the estimation and forecast results in the previous two chapters. Chapter 7 discusses the trading strategy design and presents the performance of trading strategies. Chapter 8 studies the interest rate forecast in the period of financial crisis and evaluate the model forecast performance during that abnormal period. Finally Chapter 9 concludes the thesis.
Chapter 2

Interest rate models and macro data incorporation methods

The interest rates models studied in this paper are AR, VAR, Dynamic Nelson Siegel (DNS) and Affine Term Structure Model (ATSM). Their corresponding macro counterparts are labeled as AR-X, VARX, DNS-X, and ATSM-X. The specifications of models for estimation mostly follow De Pooter, Ravazzolo, and van Dijk (2010). However, instead of following De Pooter, Ravazzolo, and van Dijk (2010)'s forecast approach, which I label as a PC Augmented Forecasting approach (PCAF) (see Appendix B for details), I apply direct forecasts for models without macro factors and develop a Three Pass Regression Filter Residual Forecasting approach (3PRFRF) for forecasts with macro factors. The core idea of 3PRFRF, the Three Pass Regression Filter is briefly explained in Section 2.1. Without going into too much theoretical backgrounds, the rest of the chapter gives an overview of how estimation and forecasting are conducted for each interest rate model with and without macro factors.

2.1 Three Pass Regression Filter: A Primer

The Three Pass Regression Filter is proposed by Kelly and Pruitt (2011). Similar to PCA, 3PRF aims to summarize a wide range of variables with a small number of fundamental factors. While PCA computes factors via the variance matrix of the data X itself, 3PRF also takes into account a target variable y which is to be explained by the X. It stems from the idea that the series that have a large influence on the variance structure of X might turn out to be irrelevant to y, whereas seemingly negligible series may in fact contain substantial explanatory power for y. The 3PRF is particularly useful in extracting relevant information in a large dimension of data when the pervasive predictors are only a subset. In this section only the algorithm of 3PRF is presented, please refer to the original paper by Kelly and Pruitt (2011) for details.

The 3PRF approach requires a proxy Z (T × K) that is only dependent on target-relevant factors. Z serves as a bridge which joins the predictors to the target variable. Kelly and Pruitt (2011) show that no priori assumption
is needed for the existence of a proxy and there always exits a set of readily available proxies. Assume there is a target variable with its historical values contained in a $T \times 1$ vector $y$ and a $T \times N$ matrix of predictors $X$ where $N$ is typically very large. The panel data $X$ has the structure $X = (x_1', x_2', \ldots, x_T') = (x_1, x_2, \ldots, x_N)$, where $x_i'$ denotes the time series representation and $x_n$ is the cross section representation. The 3PRF derives its name in the three rounds of regressions in its algorithm:

1. Run time series regression of $x_i$ on $Z$ for all $i \leq N$,
   \[ x_{i,t} = \phi_{0,i} + z_t' \phi_{1,i} + \varepsilon_{i,t}, \]  
   and retain slope estimate $\hat{\phi}_{1,i}$;

2. Run cross section regression of $x_t$ on $\hat{\phi}_{1,i}$ for all $t \leq T$,
   \[ x_{i,t} = \gamma_t + \hat{\phi}_{1,i}' F_t + \tilde{\varepsilon}_{i,t}, \]  
   and retain slope estimate $\hat{F}_t$;

3. Run time series regression of $y_{t+h}$ on predictive factors $\hat{F}_t$,
   \[ y_t = \beta_0 + \hat{F}_t' \hat{\beta}_t + \hat{\varepsilon}_t \]  
   \[ (2.3) \]

All regressions are OLS regressions and the estimates can be derived in closed form solutions.

The first pass regression consists of $N$ separate time series regression equations. Each column of $X$ is a dependent variable. The proxy $Z$ is the regressor. The idea of the first pass is to filter out the information that is not related to the proxy. The second pass regression consists of $T$ separate cross section regression equations. Each row of $X$ is a dependent variable. The coefficient estimator in the first pass is used as a regressor. The intuition of the second pass regression is to extract the relevant information from the predictors. The factors are constructed in the first two passes of regression and are carried on to the third round to perform forecasting. The third pass regression is a single time series regression. The fitted value $\hat{\beta}_0 + \hat{F}_t' \hat{\beta}_t$ is the 3PRF $h$ ahead forecast at time $t$.

### 2.2 AR model

The simplest interest rate model I consider is the univariate autoregressive model. In this section, the estimation and forecasting procedure of AR without macro factors, the extraction of macro factors as well as the estimation and the forecasting with macro factors are discussed in detail.

Denote $y^\tau_t$ as the yield of tenor $\tau$ at time $t$, the AR(1) model is formulated as
\[ y^\tau_t = c^\tau_1 + \gamma^\tau_1 y^\tau_{t-1} + \varepsilon_{1,t}, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2) \]  
\[ (2.4) \]
The one month ahead forecast given the yield data at time $t$ is given by:
\[ \hat{y}^\tau_{t+1} = \hat{c}^\tau_1 + \hat{\gamma}^\tau_1 y^\tau_t, \]
\[ (2.5) \]
where $\hat{c}_1^\tau$, $\hat{\gamma}_1^\tau$ are parameter estimates. The three month ahead iterated forecasts would be computed iteratively as

\[
y_{t+2}^\tau = \hat{c}_1^\tau + \hat{\gamma}_1^\tau y_{t+1}^\tau \\
y_{t+3}^\tau = \hat{c}_1^\tau + \hat{\gamma}_1^\tau y_{t+2}^\tau
\]

(2.6)

(2.7)

Alternatively, the three month ahead forecasts can also be constructed in a direct manner. The direct forecasting approach requires a separate estimation of a restricted AR(3) process in which the coefficients of $L(y)$ and $L^2(y)$ are restricted to be 0.

\[
y_t^\tau = c_3^\tau + \gamma_3^\tau y_{t-3}^\tau + \varepsilon_{3,t}
\]

(2.8)

and the three month ahead direct forecast is given by

\[
y_{t+3}^\tau = \hat{c}_3^\tau + \hat{\gamma}_3^\tau y_t^\tau
\]

(2.9)

I use the 3PRF approach described in Section 2.1 to extract macro factors. I use the residual from the AR regression (see Equation (2.4) in one month forecast horizon case and Equation (2.8) in three month forecast horizon case) as the proxy $Z$ for the $i^{th}$ tenor, $h$ period ahead forecast and define

\[
z_{t,h}^\tau = \varepsilon_{h,t} = y_t^\tau - (c_h^\tau + \gamma_h^\tau y_{t-h}^\tau)
\]

(2.10)

Denote the macro data matrix as $M_t$, the extracted macro factors for the yield of tenor $\tau$ at $h$ forecast horizon as $F_{t,h}^\tau$, the derivation of $F_{h,t}^\tau$ from $M_t$ follows the algorithm in Section 2.1. The first pass regression application of Equation (2.1) is given by

\[
M_{t-1} = \lambda_{t,h}^\tau z_{t,h}^\tau + \xi_{t,h}^\tau
\]

(2.11)

The second pass regression which corresponds to Equation (2.2) is

\[
M_t' = F_{t,h}^\tau \lambda_{h}^\tau + \eta_{t,h}^\tau
\]

(2.12)

Since the proxy is a vector, the AR Equation (2.4) and (2.8) are augmented by only one macro factor, which is distinct for each tenor and forecast horizon. The choice of proxy is motivated by the desired properties of the macro factors:

1. Relevance: The macro factors should have predictive power for future yields.

2. Orthogonality: The information contained in the macro factor should not overlap with the information contained in the lagged yields, or more generally, what has already been used as explanatory variables.

To forecast with macro factors, first run the regression

\[
y_t^\tau = c_h^\tau + \gamma_h^\tau y_{t-h}^\tau + \theta_h^\tau F_{t-h,h}^\tau + \varepsilon_{h,t}
\]

(2.13)

The forecasts of yields using macro factors are given by

\[
y_{t+h}^\tau = \hat{c}_h^\tau + \hat{\gamma}_h^\tau y_t^\tau + \hat{\theta}_h^\tau F_{t,h}^\tau
\]

(2.14)

Notice that $c_h^\tau$ and $\gamma_h^\tau$ in Equation (2.13) are the same as those in equations (2.4) and (2.8) due to the orthogonality of $F_{h,t}^\tau$ to $y_t^\tau$. Same applies to $\hat{c}_h^\tau$ and $\hat{\gamma}_h^\tau$. 

10
2.3 VAR model

Vector autoregressive (VAR) model takes into account the history of all maturities in addition to the target maturity itself. The unrestricted VAR(1) has the form

\[ Y_t = C + \Phi Y_{t-1} + H \varepsilon_t \]  

(2.15)

where \( Y_t \) is an \( N \times 1 \) vector of interest rates of different tenors at time \( t \), \( C \) is an \( N \times 1 \) constant vector, \( \Phi \) is an \( N \times N \) full coefficient matrix and \( H \) is a Choleski decomposition matrix of the covariance of residuals. The system could be estimated equation by equation using OLS. However, given the highly correlated interest rates of different tenors, \( Y_{t-1} \) is close to be a singular matrix, or more precisely, far from being an perpendicular matrix, which gives highly inefficient estimates. \(^1\)

Therefore, same as De Pooter, Ravazzolo, and van Dijk (2010), I replace \( Y_{t-1} \) with the first three PCs of the yields panel \( y_{t-1} \). The cost of reducing the number of regressors is trivial in this case, since the first three principal components account for over 99% of the variation. The model is reformulated as

\[ Y_t = C_h + \Phi PC_{t-h} + H_h \varepsilon_t \]  

(2.16)

Unlike De Pooter, Ravazzolo, and van Dijk (2010), who forecast the principal components and yields in a reciprocal and recursive way, I am able to predict the \( h \) ahead forecast of \( Y \) without forecasting PCs using the direct forecast approach.

The incorporation of macro factors is similar to the AR case. Stack all column vectors \( F^{\tau}_t \) to form a \( N \times N \) matrix \( F_t \). The estimation model is given by

\[ Y_t = C_h + \Phi PC_{t-h} + \Psi F_{t-h,t} + H_h \varepsilon_t \]  

(2.17)

and correspondingly the forecast is constructed as

\[ \hat{Y}_{t+h} = \hat{C}_h + \hat{\Phi} PC_{t} + \hat{\Psi} F_{t,h} \]  

(2.18)

As the each tenor has its own residual vector and hence its own macro factor, the regression Equation (2.17) is a restricted VAR where \( \Psi \) is restricted to be a diagonal matrix, so that each regression equation is only augmented by one more explanatory variable.

2.4 Dynamic Nelson Siegel model

Unlike AR and VAR, which capture the dynamics of interest rates, the Nelson Siegel model focuses on the contemporaneous term structure of yields. The static Nelson Siegel model is able to represent a wide range of shapes of the yield curve.

To enable forecast capability of the traditional Nelson Siegel model, I follow Diebold and Li (2006) and allow factors to be time varying. The Dynamic

---

\(^1\) Indeed, my first trial is to run a full unrestricted VAR and I end up with insignificant off diagonal entries in the coefficient matrix and the stationarity assumption is sometimes violated.
Nelson Siegel without macro factors is formulated as:

\[ y_t^\tau = \beta_{1,t} + \beta_{2,t} \left[ \frac{1 - \exp \left( -\frac{\tau}{\lambda} \right)}{\tau/\lambda} \right] + \beta_{3,t} \left[ \frac{1 - \exp \left( -\frac{\tau}{\lambda} \right)}{\tau/\lambda} - \exp \left( -\frac{\tau}{\lambda} \right) \right] + H \varepsilon_t, \]

(2.19)

\[ \beta_t = a + \Gamma \beta_{t-1} + Qu_t, \]

(2.20)

where \( y_t^\tau \) is the yield of maturity \( \tau \) at time \( t \), \( \lambda \) is the scale factor and the vector \( \beta_t = (\beta_{1,t}, \beta_{2,t}, \beta_{3,t})' \), contains the latent factors of level, slope and curvature at time \( t \).

Diebold, Rudebusch, and Aruoba (2006) show how to map the Dynamic Nelson Siegel to the state space model, which is given by

\[
\begin{pmatrix}
\beta_{t+1} \\
Y_t
\end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix} + \begin{pmatrix} \Gamma \\ X \end{pmatrix} \beta_t + \begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix}
\]

(2.21)

\[
\begin{pmatrix}
\varepsilon_t \\
u_t
\end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0_{3 \times 1} \\ 0_{N \times 1} \end{pmatrix}, \begin{pmatrix} QQ' & 0 \\ 0 & HH' \end{pmatrix} \right)
\]

where \( X \) is an \( N \times 3 \) time-invariant matrix whose \( i^{th} \) row is:

\[
\begin{pmatrix}
1 & \frac{1 - \exp \left( -\frac{\tau_i}{\lambda} \right)}{\tau_i/\lambda} & \frac{1 - \exp \left( -\frac{\tau_i}{\lambda} \right)}{\tau_i/\lambda} - \exp \left( -\frac{\tau_i}{\lambda} \right)
\end{pmatrix}
\]

The state space model is estimated using the Kalman Filter. When performing forecasts, extend the yields panel with a \( h \times N \) matrix of missing values. The Kalman Filter fills up the missing values with the \( h \) period forecasts of the yields. For a detailed algorithm, please refer to Durbin and Koopman (2001).

The incorporation of macro factors in DNS is different from that in the AR and VAR cases. I introduce the Three Pass Regression Filter Residual Forecast (3PRFRF) approach. The idea is to keep the state space model Equation (2.4) as it is and use 3PRF factors to forecast the latent states, \( \hat{\beta}_{t+h} \), which ultimately leads to the yield forecast \( \hat{y}_{t+h} \). Given the parameter estimates of the state space, first use the Kalman smoother to get smoothed state dynamics, a \( T \times 3 \) matrix \( (\beta_1, \beta_2, \beta_3) \), as well as the conditional variance of \( \beta_t \), denoted by \( P_t \)

\[ P_t \equiv Var(\beta_t | Y_{t-1}) \] (2.22)

Then use the residual of the restricted VAR(\( h \)) process of the states as proxy to get the 3PRF factors. The procedure is similar to the AR case (see Equation (2.10), (2.11) and (2.12)) except that the target variable is \( \beta \) rather than \( y \). The three latent factors \( \beta_1, \beta_2 \) and \( \beta_3 \) give rise to a \( T \times 3 \) matrix \( (\beta_1, \beta_2, \beta_3) \), as well as the conditional variance of \( \beta_t \), denoted by \( P_t \)

\[
\beta_t = \delta_0 + \delta_1 \beta_{t-h} + \delta_3 F_{h,t-h} + G \varepsilon_t 
\]

(2.23)

The forecast is constructed as

\[
\hat{\beta}_{t+h} = \hat{\delta}_0 + \hat{\delta}_1 \beta_t + \hat{\delta}_3 F_{t,h}, 
\]

(2.24)

where \( \hat{\delta}_0, \hat{\delta}_1 \) and \( \hat{\delta}_3 \) are estimates of Equation (2.23). With the forecast of states, the forecast of yields is only one-step away

\[
\hat{Y}_{t+h} = X \hat{\beta}_{t+h}
\]

(2.25)
The standard error of the forecast $\hat{Y}_{t+h}$ needs to be adjusted for the two-step error of $\hat{\beta}_{t+h}$. The variance matrix of the predicted $\hat{Y}_{t+h}$, $F_{t+h}$ is given by

$$F_{t+h} = XP_{t+h}X' + HH', \quad (2.26)$$

where $Q$ is the variance matrix of the measurement equation from Equation (2.19).

$$P_{t+h} = \delta_1 P_t \delta'_1 + GG', \quad (2.27)$$

where $P_t$ is given in Equation (2.22) and $GG'$ is the variance matrix of residuals from Equation (2.23).

### 2.5 Affine Term Structure model

The Affine Term Structure model (ATSM) is made popular by Cox, Ingersoll, and Ross (1985) from a microeconomics perspective. The no-arbitrage opportunity assumption imposes restrictions on the measurement of the cross section of yields. The set up of the affine model is similar to that of Ang and Piazzesi (2003), except that I have imposed more restrictions to arrive at a more parsimonious model (see Section 4.2). The set-up of the affine model without macro factors first consists of a time series model on the dynamics of the state variables

$$Z_t = \mu + \Psi Z_{t-1} + Q\varepsilon_t \quad (2.28)$$

Second, the short rate is affine in state variables

$$r_t = \delta_0 + \delta'_1 Z_t + \eta_t \quad (2.29)$$

Third, the market price of risk is also affine in state variables

$$\lambda_t = \lambda_0 + \lambda_1 Z_t \quad (2.30)$$

According to the first Fundamental Theory of Asset pricing, the no-arbitrage opportunity assumption implies the existence of a stochastic discount factor

$$m_{t+1} = \exp \left( -\frac{1}{2} \lambda'_t \lambda_t - \delta_0 - \delta'_1 Z_t - \lambda'_t \varepsilon_{t+1} \right) \quad (2.31)$$

Further assume the zero-coupon bond price is an exponentially affine function of the state variables, together with the no-arbitrage opportunity assumptions

$$P_t^\tau = \exp A^\tau + B^\tau Z_t \quad (2.32)$$

$$P_t^\tau = E_t[m_{t+1}P_{t+1}^\tau] \quad (2.33)$$

From Equation (2.32) and Equation (2.33), the coefficients $A^\tau$ and $B^\tau$ can be derived in a recursive fashion

$$A^\tau + 1 = A^\tau + B^\tau (\mu - \Sigma \lambda_0) + \frac{1}{2} B^\tau \Sigma \Sigma' B^\tau - \delta_0 \quad (2.34)$$

$$B^\tau + 1 = B^\tau (\Psi - \Sigma \lambda_1) - \delta'_1 \quad (2.35)$$

The initial values are $A^0 = 0$ and $B^0 = 0$. Correspondingly, the yields are affine in the state variables

$$y^\tau_t = a^\tau + b^\tau Z_t + H\varepsilon_t, \quad (2.36)$$
with \( a^\tau = -A^\tau / \tau \) and \( b^\tau = -B^\tau / \tau \). Therefore we arrive at the state space model for yields

\[
\begin{pmatrix}
Z_{t+1} \\
Y_t
\end{pmatrix} = \begin{pmatrix}
\mu \\
a
\end{pmatrix} + \begin{pmatrix}
\Psi \\
b
\end{pmatrix} Z_t + \begin{pmatrix}
\varepsilon_t \\
u_t
\end{pmatrix}
\]  

\( (2.37) \)

\[
\begin{pmatrix}
\varepsilon_t \\
u_t
\end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix}
0_{3 \times 1} \\
0_{1 \times 1}
\end{pmatrix}, \begin{pmatrix}
QQ' & 0 \\
0 & HH'
\end{pmatrix}\right)
\]

The short rate process in Equation (2.29) plays double roles in the state space regression Equation (2.5). The parameters \( \delta_0 \) and \( \delta_1 \) constitute the stochastic discount factor in Equation (2.31) and the short rate process itself is modeled in Equation (2.36) by setting \( \tau = 1 \).

The two-step estimator, favored by Ang and Piazzesi (2003), Moench (2008), De Pooter, Ravazzolo, and van Dijk (2010), first estimates the state dynamic and then the affine regression of yields. It requires observable state variables \( Z_t \) in Equation (2.28), which can be achieved by either constructing state variables from the yields or by direct observations. Examples of the former approach can be found in Chen and Scott (1993), Ang and Piazzesi (2003), De Pooter, Ravazzolo, and van Dijk (2010) and the latter approach is practiced by Moench (2008). Nevertheless, to avoid imposing any assumptions on the state variables and explicitly modeling their dynamics, I estimate the affine model in a one-step fashion using Kalman Filter with three latent state variables.

Ang and Piazzesi (2003) add macro factors directly into the state vector and expand all the other parameter matrices accordingly. This is by no means infeasible as long as the model correctly identifies all the parameters. However, I maintain the current structure of the affine setting and only allow macro factors to come into being at the forecast stage, applying once again the 3PRFRF approach described in Section 2.4.
Chapter 3

Data

3.1 Yield data

The yield dataset dates from July 1981 to May 2011, consisting of end of month, annualized yields levels on U.S. zero-coupon bonds with maturity horizons of 3 months, 1 to 10 years, 15 years, 20 years and 30 years. I take logarithm of the yields levels in practice. I choose the zero-coupon yields data to study interest rate models for three reasons. First, zeros-coupon yields are measured with little error. Second, zero-coupon yields are released daily and long historical data is available. Third, zero-coupon yields serve as the basis of all other interest rates and are most relevant with trading strategy design.

The primary data source is Bloomberg, which constructs the zero-coupon rates from average bid-ask price quotes on U.S. Treasuries from the CRSP government bond files. Unfortunately the Bloomberg dataset only dates back to 1995. Data before 1995 is complemented with data from other sources. For the 3-month short rate and 30-year yields, the data before 1995 is filled with 3-month T-bill rates and 30-year constant maturity bond yields from FRED (Federal Reserve Economic Data). Wright data (Gurkaynak, Sack, and Wringer 2007) is used in the same manner for the other maturities. While the zero-coupon rates might come out of different timing and construction methods, the differences among data in the overlapping period are negligible. Figure 3.1 depicts the time series graphs for all 14 maturities for the full sample period and summary statistics are reported in Table 3.1.

3.2 Macro economic data

All macroeconomic indicators come from Bloomberg. The data set consists of 78 series, selected out of the macro universe used by Stock and Watson (2002b). However, series that are released less frequently and those which do not go back as far as 1981 are deleted and I end up with 40 macroeconomic indicators. The macro indicators can be roughly categorized into six groups, namely, output, employment, consumer sentiment, housing, inflation and miscellaneous. The miscellaneous group mainly consists of Federal Reserve policy and budget indicators. A complete list of macroeconomic series can be found in Tables D.1 and D.2 in the Appendix together with their relevant information. The macroeconomic data is dated according to the their release dates for real time analysis.
Figure 3.1: U.S. zero-coupon yields

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>stdev</th>
<th>skew</th>
<th>kurt</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-month</td>
<td>4.74</td>
<td>2.80</td>
<td>0.35</td>
<td>3.25</td>
<td>0.02</td>
<td>14.43</td>
</tr>
<tr>
<td>1-year</td>
<td>5.39</td>
<td>3.20</td>
<td>0.53</td>
<td>3.38</td>
<td>0.19</td>
<td>16.11</td>
</tr>
<tr>
<td>2-year</td>
<td>5.70</td>
<td>3.18</td>
<td>0.55</td>
<td>3.22</td>
<td>0.37</td>
<td>15.78</td>
</tr>
<tr>
<td>3-year</td>
<td>5.95</td>
<td>3.11</td>
<td>0.59</td>
<td>3.14</td>
<td>0.55</td>
<td>15.57</td>
</tr>
<tr>
<td>4-year</td>
<td>6.16</td>
<td>3.03</td>
<td>0.62</td>
<td>3.09</td>
<td>0.84</td>
<td>15.35</td>
</tr>
<tr>
<td>5-year</td>
<td>6.32</td>
<td>2.96</td>
<td>0.68</td>
<td>3.05</td>
<td>1.18</td>
<td>15.18</td>
</tr>
<tr>
<td>6-year</td>
<td>6.46</td>
<td>2.89</td>
<td>0.71</td>
<td>3.03</td>
<td>1.56</td>
<td>15.06</td>
</tr>
<tr>
<td>7-year</td>
<td>6.59</td>
<td>2.82</td>
<td>0.74</td>
<td>3.03</td>
<td>1.63</td>
<td>14.99</td>
</tr>
<tr>
<td>8-year</td>
<td>6.70</td>
<td>2.77</td>
<td>0.78</td>
<td>3.02</td>
<td>2.16</td>
<td>14.94</td>
</tr>
<tr>
<td>9-year</td>
<td>6.77</td>
<td>2.74</td>
<td>0.79</td>
<td>3.00</td>
<td>2.22</td>
<td>14.91</td>
</tr>
<tr>
<td>10-year</td>
<td>6.81</td>
<td>2.73</td>
<td>0.80</td>
<td>2.96</td>
<td>2.28</td>
<td>14.89</td>
</tr>
<tr>
<td>15-year</td>
<td>7.20</td>
<td>2.52</td>
<td>0.88</td>
<td>3.09</td>
<td>3.28</td>
<td>14.86</td>
</tr>
<tr>
<td>20-year</td>
<td>7.35</td>
<td>2.48</td>
<td>0.89</td>
<td>3.09</td>
<td>3.23</td>
<td>14.84</td>
</tr>
<tr>
<td>30-year</td>
<td>7.05</td>
<td>2.40</td>
<td>0.77</td>
<td>2.85</td>
<td>2.57</td>
<td>14.14</td>
</tr>
</tbody>
</table>

Table 3.1: Summary statistics
The dates in the raw dataset are the “accrued dates”. The release dates, however, are often lagged from a week to a month’s period. Taking the Industrial Production (IP) indicator as an example, the index is updated on a monthly frequency and is subject to a one-month lag. In other words, the IP for May 2011 is released in June 2011 and is documented under May 2011, which is, as a matter of fact, not yet available. Therefore, to make sure the forecast does not involve future information, all series are manually lagged to their release date.

Another issue about the macro variables is that they are typically measured with error and are subject to multiple revisions months after they are released. Since the actual release is not available as far back as 1981, I use the revised data, which necessarily involves the future information. Nevertheless, usually the more timeliness an indicator is released, the more substantial revision it undergoes. The selection of indicators usually involves a trade off between timeliness and accuracy.

I extend the Stock and Watson (2002b) dataset to cover more aspects of the macro economy. Additional indicators on monetary policy, federal reserve policy and government budgets are included and grouped to Miscellaneous category. What’s more, I delete multiple indicators describing the same aspect of the economy. Regional and segmented indicators are replaced with aggregate or total values.

Since no assumptions are imposed on the evolution of the macro series and no time series model is implemented, stationarity is not required in the macro series. No transformation is taken except for detecting outliers and standardization. I use the same technique for outlier detection as Stock and Watson (2002a). Observations outside the interquartile range are replaced by the first median value of the previous five observations.

### 3.3 Futures price data

The trading strategies are targeted at generic futures contracts on 2-year, 5-year and 10-year interest rates. The price data of the futures contracts with delivery date less than 3 months is also from Bloomberg. The prices are the end-of-the-month settlement prices, rolled forward at the delivery date.

Theoretically, the trading strategy should be based on the prediction of the forward rates, which entails the estimation of the whole yield curve for each period. Nonetheless, the return on the futures contracts has presented significant correlation with change of the spot rate both in 1 month and 3 month evaluation horizons (see Table 3.2), which implies that when, say, the 27-month forward rate falls, the price of the 2-year futures rises, and in 96% of the cases the 2 year spot rate falls as well. Therefore, the forecast is focused on the spot rates instead of the forward rates, eliminating the trouble to estimate every point on the yield curve. The cost of the convenience is that in the remaining of the 4% circumstances, the right prediction of 2 year rates of time $t + 1$ leads to a negative return on the 2-year futures, which can be attributed to the change duration patterns or peculiar shapes of the yield curve.
Table 3.2: Correlation between futures return and spot rate for the period 1999:12-2011:5. $\Delta(y^\tau)$ denotes the change of interest rate of tenor $\tau$ and $R^\tau$ stands for return on futures written on interest rate tenor $\tau$. Correlations of corresponding returns and spot rates are denoted in bold.

<table>
<thead>
<tr>
<th></th>
<th>$\Delta(y^2)$</th>
<th>$\Delta(y^5)$</th>
<th>$\Delta(y^{10})$</th>
<th>$R^2$</th>
<th>$R^5$</th>
<th>$R^{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation ($\Delta(y^\tau), R^\tau$), 1 month horizon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta(y^2)$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta(y^5)$</td>
<td>0.90</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta(y^{10})$</td>
<td>0.73</td>
<td>0.92</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>-0.96</td>
<td>-0.91</td>
<td>-0.78</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^5$</td>
<td>-0.86</td>
<td>-0.97</td>
<td>-0.90</td>
<td>0.92</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$R^{10}$</td>
<td>-0.77</td>
<td>-0.94</td>
<td>-0.95</td>
<td>0.83</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>Panel B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation ($\Delta(y^\tau), R^\tau$), 3 month horizon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta(y^2)$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta(y^5)$</td>
<td>0.90</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta(y^{10})$</td>
<td>0.74</td>
<td>0.94</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>-0.94</td>
<td>-0.90</td>
<td>-0.77</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^5$</td>
<td>-0.84</td>
<td>-0.95</td>
<td>-0.91</td>
<td>0.92</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$R^{10}$</td>
<td>-0.76</td>
<td>-0.94</td>
<td>-0.95</td>
<td>0.83</td>
<td>0.97</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Chapter 4

Estimation

The full sample is divided into four parts. The first part is from the beginning, July 1981, to December 1999, consisting of 222 observations. The second part starts from January 2000 to December 2003, including 48 observations. The third part has 53 observations, lasting from January 2004 to May 2008. And the remaining 37 observations go to the fourth part. Estimation is done recursively using an expanding window 1, starting from the initial sample, July 1981 to December 1999. From the second period onwards, forecasts and trading strategies are constructed. The combination of forecasts and trading strategies starts from the third period using the second period results as historical performance. The last period is reserved for a separate financial crisis study.

4.1 Monte Carlo Simulation Results

Monte Carlo simulations are used to recover data generating in DNS and ATSM specifications. I specify two sets of plausible parameter values for DNS and ATSM respectively as data generating process and simulate 222 observations (same sample size as the initial window size in the recursive estimation). Such simulation is repeated 1000 times using the same set of data generating parameters. For either DNS or ATSM, I arrive at $222 \times 1000$ simulated observations and they are carried into the estimation procedures to recover the true parameters. The simulation results show that the specification of DNS model in section 2.4 gives back the true parameters. However, when macro factors augment the state variable $\beta$, the coefficients on macro factors are not recovered effectively. As for ATSM, the short rate process and the market price of risk are difficult to identify. Even more problematic is the recovery process when macro factors further enlarge the dimension of the short rate and the risk premium parameter matrices. In order to make sure the parameters are recovered, I impose restrictions on ATSM to reduce the number of parameters that enter the numerical optimization and exclude macro factors from the estimation of the two models. Models like DNS and ATSM get more unstable when the dimension of parameter estimates increases. As it turns out later, parsimony does not necessarily come as a cost of in-sample fit in the DNS and ATSM cases. Section 4.2 and

---

1A rolling window estimation is also conducted. The forecasts constructed out of the expanding window estimates are marginally better than those from a rolling window, therefore only expanding window case is documented
4.3 are based on the optimal specification that survives the Monte Carlo Simulation. Corresponding graphs of the results from Monte Carlo simulations can be found in Appendix A.

### 4.2 Estimation techniques

The AR and VAR model is estimated equation by equation using OLS. While ATSM and DNS are estimated via maximum likelihood. The numerical optimization is time consuming and it's difficult to have strong convergence.

In DNS, yields are scaled by 12 to convert to monthly yields. The measurement Equation (2.19) is linear given \( \lambda \). It does little harm to assume a static \( \lambda \) whose calibration could be done in a separate step before employing the maximum likelihood to estimate the time dynamic of \( \beta \). First treat the panel of yields as a pooled cross sectional time series data of dimension \( 14 \times T \), where each column is a separate sample. Conditional on \( \lambda \), the \( \beta \) at each time period has the OLS closed form solution given by

\[
\beta_t = E(xx'|\lambda)^{-1}E(xy_t|\lambda)
\]

where the vector \( x \) is calculated through Equation (2.4). Repeating the procedure for every time period, I end up with a \( 14 \times T \) residual matrix \( R \) with each entry \( r_{ij}(\lambda) \) being a function of \( \lambda \). The optimal \( \lambda^* \) is defined as the one that minimizes the sum of the squared residual in the residual matrix

\[
\lambda^* = \arg \min_{\lambda} \sum_{i=1}^{14} \sum_{j=1}^{T} r_{ij}(\lambda)
\]

Set \( T \) to 222, which is the size of the initial sample and calibrate \( \lambda \) using the above algorithm. The optimal \( \lambda^* \) turns out to be 19.378. In the later stage, this value is plugged into the state space model of DNS and DNSX as an exogenous constant.\(^2\)

The estimation of the Affine model entails modeling the short rate, typically the one month yield in this case, which is not available in the dataset. Therefore all yields are scaled by 4 so that yields are measured in 3 months’ unit. The maturity horizons become \((1, 4, 8, \ldots, 40, 60, 80, 120)\). In this way the 3 month yield can be used as the short rate. The yields forecasts as well as their standard errors are re-scaled back to have the annualized nature. Even after the initial scaling, the estimation is still troublesome due to the large dimension of parameter space as well as the highly nonlinear constructions. To reduce the number of parameters, some restrictions are imposed. Since the states are latent, they are normalized to have zero mean and diagonal variance matrix. In Equation (2.28), \( \mu \) becomes zeros vector and drops out; \( \Psi \) is structured to be diagonal; the variance of the residuals is derived by \( Q = Var(Z) - \Psi Var(Z)\Psi' \).

In addition, according to Equation (2.36), it holds that \( E(y_t^2) = a^2 + b^2 E(Z_t) \). Under the condition \( E(Z_t) = 0 \), I have \( a^2 = E(y_t^2) \), which means \( \delta_0 = E(r) \) and the recursive construction for \( A^\tau \) described in Equation (2.34) is no longer

\(^2\)A drawback of this approach is that \( \lambda^* \) does not get updated as the estimation window expands. Of course \( \lambda^* \) can be re-calibrated each time when a new observation comes in. Nevertheless since the static \( \lambda^* \) already does a good job at in-sample fit I do not bother to allow a time-variant \( \lambda \).
needed and $\lambda_0$ drops out of the parameter space. Moreover, $\Sigma \lambda_1$ is restricted to be a lower diagonal matrix and is optimized as a whole for stability considerations. As a result, the number of parameters that enter the optimization for ATSM is reduced to 29. I use a one-step Kalman Filter estimator and leave the system to come up with the optimal latent states by itself. While the one-step estimation may be more time-consuming, there is no need to derive observable latent states.

In both Dynamic Nelson Siegel and Affine cases, stationarity is imposed as additional restriction in the optimization and sometimes results in the situation that the optimization search gets stuck in the corner with an unavailable score vector. If the corner solution takes place, the starting value for parameters are changed to facilitate a new route for numerical search.

### 4.3 Estimation results

#### 4.3.1 AR and VAR model

The parameter estimates reported in Table 4.1 and 4.2 use the full sample from July 1981 to May 2011 (358 observations in total) for AR-X(1) and VARX(1) in regression Equation (2.13) and (2.17) of $h = 1$ case. AR(1) and VAR(1) share the same estimates of constants and autoregressive coefficients with AR-X(1) and VARX(1) which can be find in the first 2 columns of Table 4.1 and 4.2.

<table>
<thead>
<tr>
<th></th>
<th>$c^*$</th>
<th>$\gamma^*$</th>
<th>$\theta^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3m</td>
<td>3.14</td>
<td>(0.84)</td>
<td>0.98 (0.006)</td>
</tr>
<tr>
<td>1y</td>
<td>3.44</td>
<td>(1.06)</td>
<td>0.98 (0.006)</td>
</tr>
<tr>
<td>2y</td>
<td>3.64</td>
<td>(1.21)</td>
<td>0.98 (0.006)</td>
</tr>
<tr>
<td>3y</td>
<td>3.88</td>
<td>(1.26)</td>
<td>0.98 (0.006)</td>
</tr>
<tr>
<td>4y</td>
<td>4.25</td>
<td>(1.19)</td>
<td>0.98 (0.007)</td>
</tr>
<tr>
<td>5y</td>
<td>4.55</td>
<td>(1.13)</td>
<td>0.98 (0.007)</td>
</tr>
<tr>
<td>6y</td>
<td>4.83</td>
<td>(1.07)</td>
<td>0.98 (0.007)</td>
</tr>
<tr>
<td>7y</td>
<td>5.14</td>
<td>(0.97)</td>
<td>0.98 (0.007)</td>
</tr>
<tr>
<td>8y</td>
<td>5.29</td>
<td>(0.95)</td>
<td>0.98 (0.007)</td>
</tr>
<tr>
<td>9y</td>
<td>5.41</td>
<td>(0.92)</td>
<td>0.98 (0.007)</td>
</tr>
<tr>
<td>10y</td>
<td>5.51</td>
<td>(0.90)</td>
<td>0.98 (0.007)</td>
</tr>
<tr>
<td>15y</td>
<td>6.24</td>
<td>(0.69)</td>
<td>0.97 (0.008)</td>
</tr>
<tr>
<td>20y</td>
<td>6.64</td>
<td>(0.51)</td>
<td>0.97 (0.009)</td>
</tr>
<tr>
<td>30y</td>
<td>6.36</td>
<td>(0.52)</td>
<td>0.97 (0.009)</td>
</tr>
</tbody>
</table>

Table 4.1: AR-X parameter estimates with standard error in parentheses. Refer to Equation (2.13) for notation details. The second column reports the estimated long term mean; the third column reports the autoregressive coefficients and the last column lists the coefficient on macro factors. All estimates are significant at the 90% confidence level.
4.3.2 DNS and ATSM

Figure 4.1 pictures the dynamics of the latent factor $\beta$ of DNS state space model against the latent factor $Z$ from ATSM and the first three PCs from the yield panel. (Diebold and Li 2006) recognize that $\beta_1$ relates to PC1, $\beta_2$ to PC2 and $\beta_3$ to PC3. Figure 4.1 shows that Z shares common features with $\beta_1$. As it turns out, $Z_2$ happens to be the level factor, $Z_3$ the slope factor and $Z_1$ the curvature factor. They are plotted correspondingly with similar factors from DNS and PCs. The correlation between the factors can be read from Table 4.3. All elements in the lower diagonal $9 \times 9$ correlation matrix are reported. Figures in bold are the correlation between states that account for the same factor. Stronger correlations are observed among level and slope factors than that among the curvature factors. PCs are orthogonal to each other by construction but latent states are typically correlated.

The PCs are labeled as level, slope and curvature as early as the work of Littleman and Scheinkman (1991). The latent factor $\beta$ is recognized as in the same way by Diebold and Li (2006). Figure 4.1 implies that the latent states which drive evolution of yields under the no-arbitrage affine assumption are also level, slope and curvature, since both DNS and ATSM explain yields from the interaction between time-variant state variables and tenor-dependent coefficient matrices. The dynamics of the state variables are expected to capture the time series part of the yields while the tenor specific coefficients account for the cross-section relations.

Table 4.4 reports the $R^2$ of all tenors for the DNS and ATSM models. Of course, $R^2$ is far from a desired statistics to evaluate the performance of models, nevertheless it proves that the specification of $Y_t = \Theta X_t$, where $\Theta$ denotes a

<table>
<thead>
<tr>
<th></th>
<th>$C$</th>
<th>$\Phi$</th>
<th>diag($\Psi$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3m</td>
<td>4.70 (0.018)</td>
<td>-2.57 (0.018)</td>
<td>-0.909 (0.019)</td>
</tr>
<tr>
<td>1y</td>
<td>5.35 (0.020)</td>
<td>-3.08 (0.022)</td>
<td>-0.816 (0.022)</td>
</tr>
<tr>
<td>2y</td>
<td>5.66 (0.020)</td>
<td>-3.07 (0.021)</td>
<td>-0.497 (0.021)</td>
</tr>
<tr>
<td>3y</td>
<td>5.91 (0.020)</td>
<td>-3.04 (0.020)</td>
<td>-0.346 (0.024)</td>
</tr>
<tr>
<td>4y</td>
<td>6.12 (0.019)</td>
<td>-2.96 (0.020)</td>
<td>-0.158 (0.020)</td>
</tr>
<tr>
<td>5y</td>
<td>6.28 (0.019)</td>
<td>-2.91 (0.019)</td>
<td>-0.064 (0.023)</td>
</tr>
<tr>
<td>6y</td>
<td>6.43 (0.019)</td>
<td>-2.84 (0.019)</td>
<td>0.058 (0.019)</td>
</tr>
<tr>
<td>7y</td>
<td>6.56 (0.019)</td>
<td>-2.78 (0.019)</td>
<td>0.100 (0.023)</td>
</tr>
<tr>
<td>8y</td>
<td>6.67 (0.018)</td>
<td>-2.72 (0.019)</td>
<td>0.196 (0.019)</td>
</tr>
<tr>
<td>9y</td>
<td>6.74 (0.018)</td>
<td>-2.68 (0.018)</td>
<td>0.209 (0.023)</td>
</tr>
<tr>
<td>10y</td>
<td>6.78 (0.018)</td>
<td>-2.67 (0.018)</td>
<td>0.296 (0.018)</td>
</tr>
<tr>
<td>15y</td>
<td>7.17 (0.018)</td>
<td>-2.43 (0.019)</td>
<td>0.379 (0.018)</td>
</tr>
<tr>
<td>20y</td>
<td>7.32 (0.019)</td>
<td>-2.38 (0.020)</td>
<td>0.496 (0.020)</td>
</tr>
<tr>
<td>30y</td>
<td>7.03 (0.016)</td>
<td>-2.28 (0.017)</td>
<td>0.484 (0.016)</td>
</tr>
</tbody>
</table>

Table 4.2: VARX parameter estimates with standard error in parentheses. Refer to Equation (2.17) for notation details. The second column refers to the constant term, which is also the long term mean of each tenor; the third column reports the autoregressive coefficient matrix and the last column lists the coefficient on macro factors. All estimates are significant at the 90% confidence level.
Figure 4.1: Dynamics of level factor. Series are multiplied by -1 when necessary to facilitate comparison.
tenor-variant coefficient matrix and $X_t$ denotes the state vector at time $t$, fits the data very well.

Give a closer look at how DNS and ATSM come up with the tenor specific coefficient matrix $\Theta$, which is $X$ in DNS and $a^\tau$ and $b^\tau$ in ATSM. In DNS, there’s a closed form expression for the coefficient $X$, see Equation (2.4) where the only free parameter is $\lambda$, whereas in ATSM specification, the coefficient matrix $a^\tau$ and $b^\tau$ are derived in a recursively following the first fundamental theory of asset pricing, see (2.34) and (2.35) in which the short rate and risk premium process are the internal drivers. The analogy between $\beta$ and $Z$ suggests that $b^\tau$ is analogous to $X$, which further implies that $\lambda$ is in some way related to the no-arbitrage opportunity assumption. Table 4.5 lists selected parameter estimates with their standard errors. Only the first and last entries of the risk premium matrix are significant.  

3 In fact, Table 4.4 explains the results in Table 4.5 in some way: if one free parameter $\lambda$ suffices to capture the cross sectional bindings of yields, one better not expect to identify more than 1 factor that drives the market risk premium.

\begin{table}[h]
\centering
\begin{tabular}{cccccccc}
\hline
& $\beta_1$ & $\beta_2$ & $\beta_3$ & $Z_1$ & $Z_2$ & $Z_3$ & $PC_1$ & $PC_2$ & $PC_3$ \\
$\beta_1$ & 1 & & & & & & & & \\
$\beta_2$ & -0.02 & 1 & & & & & & & \\
$\beta_3$ & 0.58 & 0.44 & 1 & & & & & & \\
$Z_1$ & 0.54 & 0.44 & 0.99 & 1 & & & & & \\
$Z_2$ & -0.99 & 0.02 & -0.58 & -0.53 & 1 & & & & \\
$Z_3$ & 0.61 & 0.78 & 0.73 & 0.71 & -0.60 & 1 & & & \\
$PC_1$ & -0.95 & -0.27 & -0.76 & -0.72 & 0.95 & -0.81 & 1 & & \\
$PC_2$ & 0.30 & -0.94 & -0.34 & -0.35 & -0.30 & -0.57 & 0.00 & 1 & \\
$PC_3$ & 0.04 & 0.08 & -0.35 & -0.40 & -0.04 & 0.08 & 0.00 & 0.00 & 1 \\
\hline
\end{tabular}
\caption{Correlation among latent factors 1981:7-2011:5} 
\end{table} 

Table 4.3: Lower diagonal correlation matrix among latent states and PCs of the yields panel. Correlation between corresponding factors are denoted in bold.

\[\text{Note it is different from Cochrane and Piazzesi (2005) who report only the first entry being significant, but since I use a different specification of ATSM and different estimation methods, the results might not be comparable.}\]
<table>
<thead>
<tr>
<th>DNS</th>
<th>ATSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>0.9778</td>
</tr>
<tr>
<td>1 year</td>
<td>0.9991</td>
</tr>
<tr>
<td>2 year</td>
<td>0.9999</td>
</tr>
<tr>
<td>3 year</td>
<td>0.9999</td>
</tr>
<tr>
<td>4 year</td>
<td>0.9999</td>
</tr>
<tr>
<td>5 year</td>
<td>0.9999</td>
</tr>
<tr>
<td>6 year</td>
<td>0.9999</td>
</tr>
<tr>
<td>7 year</td>
<td>0.9998</td>
</tr>
<tr>
<td>8 year</td>
<td>0.9999</td>
</tr>
<tr>
<td>9 year</td>
<td>0.9999</td>
</tr>
<tr>
<td>10 year</td>
<td>0.9994</td>
</tr>
<tr>
<td>15 year</td>
<td>0.9968</td>
</tr>
<tr>
<td>20 year</td>
<td>0.9901</td>
</tr>
<tr>
<td>30 year</td>
<td>0.9848</td>
</tr>
</tbody>
</table>

Table 4.4: Goodness of fit of DNS and ATSM. The statistics measure how well the variation in $\hat{y}_\tau^t$ relates to variation in $y_\tau^t$. The figures are very close to 1, indicating high quality of approximation.

<table>
<thead>
<tr>
<th>ATSM estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_1$</td>
</tr>
<tr>
<td>-0.639</td>
</tr>
<tr>
<td>0.336</td>
</tr>
<tr>
<td>$\Sigma \lambda_1$</td>
</tr>
<tr>
<td>0.090*</td>
</tr>
<tr>
<td>(0.018)</td>
</tr>
<tr>
<td>-9.51e-005</td>
</tr>
<tr>
<td>(0.005)</td>
</tr>
<tr>
<td>-0.573</td>
</tr>
<tr>
<td>(2.320)</td>
</tr>
</tbody>
</table>

Table 4.5: ATSM parameters that determines the recursive formula of the tenor dependent coefficient matrix $\delta^\tau$. * denotes significance at 90% confidence level.
Chapter 5

Forecasting

5.1 Forecast procedure

After each estimation, 1 and 3 months’ forecasts are constructed for each model using the latest parameter estimates. Both point forecasts and their standard errors are recorded. The forecast combination is constructed in the third part of the sample, from January 2004 to May 2008, based on the individual model’s forecast performance in the previous sample period, January 2000 to December 2003.

5.2 Forecast evaluation

I evaluate the forecast performance of each model against three dimensions: the magnitude of the prediction error, the accuracy of direction prediction and the consistency of forecast performance. The three criteria correspond to Root Mean Squared Prediction Error (RMSPE), Hit Ratio and Model Survival Ratio (MSR), respectively. The RMSPE is given by

\[
RMSPE^m_{\tau} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{y}_{t+h|i,t,m}^{\tau} - y_{t+h}^{\tau})^2}
\]  

(5.1)

where \(\tau\) stands for the \(i_{th}\) maturity; \(T\) is the number of out-of-sample forecasts; \(m\) represents a particular model and \(h\) is the forecast horizon. The average of the RMSPE across all maturities is also computed for each model and each forecast horizon. The RMSPE of the Random Walk is used as a benchmark, i.e. the RMSPE for all models are divided by the RMSPE of the Random Walk model. A value larger than 1 indicates that the particular model under-performs the Random Walk whereas a value smaller than 1 implies outperforming. The RMSPE gives a clear picture of how large the forecast error is. To have a idea how errors are built up across time, one could turn to Cumulated Squared Prediction Error (CSPE). Unlike RMSPE which is a point statistics, CSPE is vector that contains squared prediction error at every time point. I calculate

\[\text{CSPE} = \sum_{t=1}^{T} (\hat{y}_{t+h|i,t,m}^{\tau} - y_{t+h}^{\tau})^2 \]

16 and 12 month forecasts are also constructed but are not reported in the paper since they present large prediction errors and are irrelevant to the trading strategy.
the CSPE relevant to Random Walk. The formula for relevant CSPE of model $m$ is given by

$$CSPE^m = \sum_{t=1}^{T} \sum_{\tau} (\hat{y}_{t+\tau} - y_{t+\tau}^m)^2 - (\hat{y}_{t+\tau|m} - y_{t+\tau}^m)^2$$  \hspace{1cm} (5.2)$$

However, both RMSPE and CSPE fail to identify the direction of the error and they not immune to occasional extreme values. In other words, if the model has an outrageously large error at a certain point but acceptable forecasts in all other cases, the final statistics might be biased towards the large error.

The Hit ratio serves as a complementary evaluation of SPE. It ignores the magnitude of the error but focuses on whether the prediction gets the direction right. The Hit ratio is calculated in three steps. First, define an indicator function of the direction change of actual interest rates which has a value of 1 when the rate goes up and 0 when it goes down.

$$v_{t,\tau, h} \equiv 1_{y^\tau_{t+\tau} - y^\tau_t > 0}$$  \hspace{1cm} (5.3)$$

Second, define a similar indicator function for model forecast change which has value 1 when the model forecasts a rise in the yield and 0 otherwise.

$$\hat{v}_{t,\tau, h}^m \equiv 1_{\hat{y}^\tau_{t+\tau|m} - y^\tau_t > 0}$$  \hspace{1cm} (5.4)$$

Third, compare the results of the aforementioned two functions and compute how many times they agree with each other.

$$H_{m, \tau, h}^m = \frac{1}{T} \sum_{t=1}^{T} (1_{v_{t,\tau, h} = \hat{v}_{t,\tau, h}^m})$$  \hspace{1cm} (5.5)$$

where $H$ denotes the Hit ratio. A Random Walk model would have a Hit ratio of 0.5 by construction. The larger the Hit ratio is, the more powerful the model is at forecasting the direction of change. Since the "hit" follows a binomial distribution with Hit ratio as the estimated mean, it has a normal approximation of

$$H \sim N(\hat{H}, \frac{\hat{H}(1-\hat{H})}{T})$$  \hspace{1cm} (5.6)$$

which enables a test of whether the Hit ratio is significantly different from 0.5.

The last evaluation statistics is the Model Survival Ratio (MSR). It stems from the Model Confidence Set (MCS) approach proposed by Hansen, Lunde, and Nason (2005). The idea is to kick out the worst models according to a predefined confidence interval $\alpha$ so that the surviving models form a MCS. The MSR is the average times that a model stays in the MCS. The null hypothesis $H_0$ is that all models have the same performance according to a loss function $L_m$. MCS method involves the design of an equivalence test, $\delta_M$ and an elimination rule, $e_M$. With the initial setting of MCS containing all models, the null hypothesis is tested. If $H_0$ is rejected, apply the elimination rule to eliminate the worst model and re-test $H_0$. This process is repeated until $H_0$ is no longer rejected and MCS contains desirable number of models, the survivors. To elaborate, the derivation of MCS for maturity $\tau$, forecast horizon $h$ at time $T$ follows the following procedure Define the loss function

$$L_{m, t, \tau} = (\hat{y}^\tau_{t+\tau|m} - y_{t+\tau}^\tau)^2$$  \hspace{1cm} (5.7)$$

27
\[ L_{t,\tau}^{m} = \frac{1}{|\text{MCS}|} \sum_{m \in \text{MCS}} L_{t,\tau}^{m}, \quad (5.8) \]

where \(|\text{MCS}|\) denotes the number of models in MCS. Calculate \(D\), defined as the difference between each loss function and the average loss, and the average of \(D\) to get the \(t\) statistics for each model:

\[ D_{t,\tau}^{m} = L_{t,\tau}^{m} - \bar{L}_{t,\tau}, \quad (5.9) \]

\[ \bar{D}_{T,\tau}^{m} = \frac{1}{T - t_0} \sum_{t=t_0}^{t=T} D_{t,\tau}^{m}, \quad (5.10) \]

\[ t_{T,\tau}^{m} = \frac{D_{T,\tau}^{m}}{\sqrt{\text{var}(\bar{D}_{T,\tau}^{m})}}, \quad (5.11) \]

where \(t_0\) is the starting point of evaluation history window, in which case I use an expanding window. The \(t\) statistics \(t_{T,\tau}^{m}\) depends on the nuisance parameter \(\text{var}(\bar{D}_{T,\tau}^{m})\), which I approximates using \(\frac{\text{var}(\bar{D}_{T,\tau}^{m})}{T-t_0}\). Hansen, Lunde, and Nason (2005) have shown that

\[ F \equiv \sum_{m \in \text{MCS}} t_{T,\tau}^{m}, F \sim \chi^2_{|\text{MCS}|}, \quad (5.12) \]

Repeat this procedure is for each time period, each maturity and each forecast horizon to get \(\text{MCS}_{t,m,h}\). The MSR is computed as

\[ \text{MSR}_{h,\tau}^{m} = \frac{1}{T} \sum_{t=1}^{T} 1_{m \in \text{MCS}_{t,m,h}} \sum_{t=1}^{T} 1_{m \in \text{MCS}_{t,m,h}}, \quad (5.13) \]

### 5.3 Forecast combination

Since no single model consistently outperforms all other models in all evaluation criteria, one might expects to have an optimal forecast which combine the models’ specialties at each period for each tenor. The combinations are carried out among the forecasts results from the 8 models, half of which have incorporated macro factors. Two schemes are adopted to combine forecasts. The first scheme uses the inverse mean squared prediction error (MSPE) as weights and the second scheme uses the Model Confidence Set (MCS) discussed previously to select the acceptable individual forecasts for each time period.

For the inverted MSPE weighted forecasts, the MSPE at time \(T\) is calculated as

\[ \text{MSPE}_{T,h,\tau}^{m} = \frac{1}{T - t_0} \sum_{t=t_0}^{t=T} (y_{t+h,\tau|t} - y_{t+h,\tau})^2, \quad (5.14) \]

where \(t_0\) has the same value as that in Equation (5.11). The weight assigned to each individual forecast at time \(T + 1\) is based on the MSPE up to time \(T\) and is given by

\[ w_{T+h,\tau|T}^{m} = \frac{1/\text{MSPE}_{T,h,\tau}^{m}}{\sum_{m} 1/\text{MSPE}_{T,h,\tau}^{m}}, \quad (5.15) \]
The combined forecast at time $T + h$ is a weighted average of all the individual forecast, computed as

$$\hat{y}_{T+h,\tau|T} = \sum m w_{T,h,\tau|T} \hat{y}_{m,T,h,\tau}$$

(5.16)

The second scheme involves the application of MCS described in section 5.2. Unlike the first scheme where all models contribute to the forecast combination, this approach eliminates the worst models before assigning weights so that only selected models are involved in the forecast combination every time. The algorithm follows a two-step fashion, first calculate MCS at 90% confidence level and in the second step use MSPE as weights to calculate the weighted average of the surviving models to form the combined forecast.

5.4 Forecast results

Figure 5.1 and 5.3 plot the 1 month and 3 month generic models’ forecasts against the actual market data together with the Random Walk forecasts. Their macro counterparts are shown in Figure 5.2 and 5.4. Generally speaking, the models’ forecasts are close to the Random Walk forecasts and 3 months’ forecasts have more heterogeneity than the 1 month’s forecasts.

![Figure 5.1: 1 month forecast without macro factors](image)

The value of incorporating macro factors can be quantitatively evaluated by comparing SPE and Hit ratio of generic models and macro models. Figure 5.5 and 5.6 compares the CSPE of models without and with macro factors. A positive value at time $t$ implies that the specific model out performs the Random Walk model with regard to CSPE up to time $t$. From the figures one can tell none of the models consistently give smaller prediction errors than Random Walk. For the one-month ahead forecast, AR-X and VARX perform marginally
Figure 5.2: 1 month forecast with macro factors

Figure 5.3: 3 month forecast without macro factors
better than AR and VAR, whereas at a certain point DNSX and ATSM-X seem to give very wrong forecasts, resulting in larger CSPE than those of DNS and ATSM ever after. Tracing back to the 3 month forecast in Figure 5.1 and 5.2 where DNS-X and ATSM-X give extremely volatile forecasts around the beginning of 2002, which is the first period of financial crisis observed in the sample.

Table 5.1 reports the Root Mean Squared Prediction Error (RMSPE) in 1 and 3 month forecast horizons for 8 models as well as the Forecast Combinations constructed from two schemes (FC1, FC2) for the period 2004:01 to 2008:06. The table shows that it is difficult to outperform the Random Walk model in 1 month forecast horizon. Although forecast combinations are marginally better than RW on average, the gain mainly comes from the short end (3 month) while in the other maturities RW still gives best forecasts with respect to RMSPE. With regard to 3 month forecast horizon, although still no single model consistently outperforms RW, the forecast combinations are able to give the best forecasts for all reported tenors.

Table 5.2 gives the Hit ratio for all models at both horizons. While quite a few forecasts have Hit ratios significantly higher than 0.5, there’s also a couple of Hit ratios which are significantly lower than 0.5. The Hit ratios for the 2 year forecasts both in 1 month and 3 month horizons are worse compared to other tenors and the direction prediction for the short rate is the most accurate. However, recall the $R^2$ statistics listed in Table 4.4, the in-sample fit for 2 year tenor turns out to be the best of the 14 maturities while the 3 month tenor is the worst, which is the other way round. This observation indicates that good in-sample fit does not necessarily lead to accurate out-of-sample forecast.

Another conclusion drawn from Table 5.2 is that macro factors do add value to the direction prediction. The Hit ratios of forecasts with macro factors are

![Figure 5.4: 3 month forecast with macro factors](image-url)
Figure 5.5: CSPE, 1 month horizon

Figure 5.6: CSPE, 3 month horizon
Table 5.1: Root mean squared prediction error of all interest rate models and forecast combinations. The RMSPE of random walk is used as a benchmark and all others are relevant values. A figure larger than 1 denotes under-performance while a figure smaller than 1 stands for out-performance.
almost always higher than those without. The benefit of macro factor on Hit ratio is especially prominent for DNS-X model.

| Panel A: Hit Ratio 2004:1-2008:6, 1 month horizon |
|---------------------------------|------|------|------|------|------|
|                                 | 3 month | 2 year | 5 year | 10 year | Avg |
| AR                              | 0.65   | 0.52   | 0.46   | 0.43   | 0.51 |
| VAR                             | 0.80   | 0.41   | 0.46   | 0.54   | 0.55 |
| DNS                             | 0.52   | 0.37   | 0.48   | 0.50   | 0.47 |
| ATSM                            | 0.67   | 0.46   | 0.46   | 0.48   | 0.52 |
| AR-X                            | 0.70   | 0.56   | 0.54   | 0.46   | 0.56 |
| VAR-X                           | 0.80   | 0.48   | 0.44   | 0.52   | 0.56 |
| DNS-X                           | 0.57   | 0.54   | 0.56   | 0.57   | 0.56 |
| ATSM-X                          | 0.69   | 0.50   | 0.56   | 0.61   | 0.59 |
| FC1                             | 0.83   | 0.52   | 0.59   | 0.52   | 0.62 |
| FC2                             | 0.83   | 0.46   | 0.59   | 0.52   | 0.60 |

| Panel B: Hit Ratio 2004:3-2008:8, 3 month horizon |
|---------------------------------|------|------|------|------|------|
|                                 | 3 month | 2 year | 5 year | 10 year | Avg |
| AR                              | 0.70   | 0.59   | 0.59   | 0.54   | 0.61 |
| VAR                             | 0.78   | 0.43   | 0.43   | 0.59   | 0.56 |
| DNS                             | 0.48   | 0.41   | 0.44   | 0.44   | 0.44 |
| ATSM                            | 0.70   | 0.56   | 0.59   | 0.46   | 0.58 |
| AR-X                            | 0.81   | 0.67   | 0.61   | 0.56   | 0.66 |
| VAR-X                           | 0.80   | 0.52   | 0.57   | 0.52   | 0.60 |
| DNS-X                           | 0.76   | 0.57   | 0.67   | 0.56   | 0.64 |
| ATSM-X                          | 0.54   | 0.57   | 0.61   | 0.54   | 0.56 |
| FC1                             | 0.80   | 0.57   | 0.69   | 0.63   | 0.67 |
| FC2                             | 0.81   | 0.56   | 0.69   | 0.57   | 0.66 |

Table 5.2: Hit ratio of model forecasts. Figures significantly different from 0.5 at the 90% confidence level are denoted in bold.

Table 5.3 calculates the surviving ratio for each models at all horizons in the MCS elimination procedure at 70% confidence level. Remarks are summarized as follows. First, the figures in Table 5.3 are either close to 1 or close to 0 which means that the performance of each model(maybe not for AR-X in the 3 month forecast horizon case) on a typical tenor is fairly consistent over time (using MSPE as measurement). Second, Models with macro factors usually have a larger chance of survival, although macro models do not necessarily deserve a higher weights. Third, comparing panel A and panel B in Table 5.3, when the forecast horizon is expanded from 1 month to 3 month, FC2 starts to prefer more structured models to simple autoregressive models. The AR-X and VAR-X models’ survival ratios decrease while the DNS-X and ATSM-X models’ rise. A reason might be as forecast horizon extends, the momentum of interest rates become less dominant and results in less accurate predictions from sheer time series models such as AR and VAR. More structure should be imposed on the cross section of yields and their join dynamics, which explains the improving performance of more structured models, such as DNS and ATSM.
### Table 5.3: Model surviving ratios in MCS during the construction of FC2.

A figure of 1 means that the MCS always contains the model m in the construction of $FC^{\tau,h}$. Correspondingly, a figure 0 denotes a never survived. The confidence level is set at 70% level.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3 month</td>
<td>2 year</td>
</tr>
<tr>
<td>AR</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VAR</td>
<td>1.00</td>
<td>0.24</td>
</tr>
<tr>
<td>DNS</td>
<td>0.00</td>
<td>0.96</td>
</tr>
<tr>
<td>ATSM</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>AR-X</td>
<td>0.89</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Chapter 6

Evaluation of interest rate models and macro data incorporation methods

The forecasting results in Section 5.4 show that there are variations of forecasting performance across models, across time and even within a model for different tenors. This chapter looks into the mechanisms of interest rate models that result in the observed variations. The second part of the chapter compares different methods of macro factor extraction, namely, PCA, Kalman Filter and 3PRF. The insights of interest rate models and macro extraction methods gives sound reasonings of using 3PRFRF approach.

6.1 Interest rate models

Five factors contribute to the change of interest rates: business cycle, inflation pressure, shape of the yield curve, momentum and risk aversion of the investors. The decomposition of models’ mechanisms makes it clear to see how each model works and why some models perform better than the other ones in particular circumstances, see Table 6.1.

Since the yields data are in nominal terms, the real business cycle factor which is reflected in the real interest rates is not modeled separately and thus

| A sketch of the mechanism of interest rate models |
|---|---|---|---|---|
| | AR | VAR | DNS | ATSM |
| Business cycle | $\frac{e^r}{1-\gamma}$ | $PC_1$ | $\beta_0$ | short rate |
| Inflation | N.A. | $PC_2, PC_3$ | $\beta_1, \beta_2$ | $a^r, b^r$ via no-arbitrage assumption |
| Shape of yield curve | $L(y_t)$ | $L(PC_1)$ | $L(\beta_t)$ | $L(Z_t)$ |
| Momentum | N.A. | N.A. | $\lambda$ | pricing kernel |

Table 6.1: Decomposition of model’s mechanisms. $L(\cdot)$ stands for the lag operator; N.A. implies absence of the factor.
always comes together with inflation factor. The shape of the yield curves mainly refers to the slope and curvature, the information of which is embedded in the cross section of yields. Pure time series models as simple as AR give no consideration to the cross section structure. VAR accounts for the shape factor by using PCs as explanatory variables which summarizes the cross sectional information. DNS is motivated by the fitting the yield curve. The no-arbitrage assumption behind ATSM gives rise to cross section restrictions, implicitly take the shape factor into account. All models have included the momentum factor by modeling the time series process of either the yields themselves (AR and VAR) or variables which are directly related to interest rates (DNS and ATSM). The risk aversion factor can only find itself in the DNS and ATSM model. The specification of the pricing kernel in ATSM models the market price of risk explicitly. To interpret $\lambda$ (see Equation (2.19)) in the DNS as risk aversion factor is less convincing. The accepted interpretation of $\lambda$ is the exponential decay towards zero of the factor loadings. However, as has been discussed in section 4.3, the correspondence of the latent states from DNS and ATSM implies that $\lambda$, which guides the DNS state factors might be subtly analogous to the state coefficient in ATSM which is subjected to no-arbitrage assumption and can be boiled down to market price of risk.

From AR to ATSM, the models get more structured and therefore captures more factors. The difference in factor accounts results in the variation in utilization of macro information. From the CSPE plots in Figure 5.5 and 5.6, as expected, macro factors improve the forecast of DNS and ATSM marginally while make a more noticeable difference for AR and VAR especially in the 3 month forecast horizon. In general, the more information contained in the generic models, the less information is extracted from the macro data, because macro factors serve as a complementary to the existing explanatory variables by construction of 3PRFRF.

A draw back of the more structured models is that their capability in dealing the short ends of the yield curve is typically poor both with respect to in-sample fit and out-of-sample forecast. Recall the $R^2$ statistics presented in Table 4.4, the lowest $R^2$ for both DNS and ATSM happen to be at the 3-month tenor. Forecasts from DNS and ATSM also prove this fact. Notice their large 3-month yield prediction error occurred at around year 2002 shown in Figure 5.1 and 5.3. Given the fact that a yield curve can be very well described using level, slope and curvature, the reason might lie in the under-identification of the level factor, which is reflected in the short rate behavior. Reading from the RMSPE table 5.1 and $R^2$ table 4.4, the structured models are specialized in the intermediate tenors, such as 2 year to 10 year yields. These tenors are mostly influenced by the slope and curvature factors. Quite the opposite, AR is good at the short end and the longer the tenor the worse its performance. Hence comes the conclusion, the more a model emphasize on time series behavior of interest rates, the better its performance at short end while the more on cross sections, the better at intermediate tenors.

6.2 Macro data incorporation methods

Two stages are involved when incorporating macro factors into the model. The first stage is to extract the macro factors from the large dataset and the second stage is to decide how to add the macro factors into the estimation and forecast.
model. All plots presented in this section are using standardized series which have mean 0 and unit variance.

### 6.2.1 A comparison of PCA, Kalman Filter and 3PRF factor extraction methods

In the first stage, three ways of extracting macro factors from the large dimension of macro dataset are discussed in this section: the principal component approach, Kalman Filter approach and the Three Pass Regression Filter approach. PCA and Kalman Filter can be performed alone regardless of the target variable whereas 3PRF requires a proxy and 3PRF factors are thus very target-dependent.

PCA is the most widely used of the three because it captures the largest variation in the dataset possible using the least number of factors. PCA is performed on the eigenvalues of the variance matrix of the data. The algorithms are simple and closed-form solutions are available. The PCs are orthogonal by construction which enable the factors to efficiently span the column space. A drawback of it is that PCs are subject to selection bias of the variables in the dataset, if the bias leads to a dramatic change in the variance structure in the data. PCA also lacks flexibility since PCs are derived explicitly and can hardly adapt to customized requests.

The Kalman Filter estimation of factors are most useful for interpolation purposes. The Kalman Filter has its advantage in dealing with missing observations and hence can be used to come up with more frequent data. The Kalman Filter factors are adaptable to estimation restrictions imposed on the estimation model. For instance, their covariance can be explicitly specified. The Kalman Filter is estimated numerically with maximum likelihood, which might be time-consuming and unstable.

3PRF is a new approach which specializes in discarding irrelevant information and retain only the useful information. The dependence of 3PRF factors on the proxy is both its advantage and its disadvantage. On one hand, failure to find a proxy that satisfies the assumption (i.e. the proxy should only depend on the target-relevant factors) is likely to hinge the forecast power contained in the data. In addition, the number of factors are restricted by the columns of the proxy matrix. On the other hand, 3PRF factors are easily customized by choosing a desirable proxy. An inevitable problem of the 3PRF factors is that the multiple passes of regressions are likely to build up standard errors and result in noisy factors. The summary of the features of the the aforementioned three factor extraction methods can be found in Table 6.3

I perform the PCA on the untransformed macroeconomic dataset with 40 variables which have no missing observations. I retain the first three PCs, which account for 25.8%, 19.9% and 11.0% of the total variance, respectively. I conduct the Kalman Filter estimation of macro factors on the complete macro dataset with 78 variables by specifying three latent factors. The estimation model can be found in Appendix C. I use the regression residual from the DNS transition equation (2.20) as proxy and extract 3PRF factors with the same macro data used for PCA. I compare PCs, Kalman Filter factors and 3PRF factors by regressing individual macro series separately on the factors and computing the $R^2$. Figure 6.1, 6.2 and 6.4 compare the $R^2$ in regression of each macro series on different factors from the 3 methods. The $R^2$ graphs give an
Comparison of macro factor extraction methods

<table>
<thead>
<tr>
<th></th>
<th>PCA</th>
<th>KF</th>
<th>3PRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>variance account</td>
<td>largest possible</td>
<td>arbitrary</td>
<td>arbitrary</td>
</tr>
<tr>
<td>orthogonality</td>
<td>yes</td>
<td>flexible</td>
<td>arbitrary</td>
</tr>
<tr>
<td>closed form</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>tolerance for missing observations</td>
<td>no</td>
<td>yes</td>
<td>restrictive</td>
</tr>
<tr>
<td>interpolation</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>correlation with target</td>
<td>arbitrary</td>
<td>arbitrary</td>
<td>strong</td>
</tr>
<tr>
<td>subjectiveness to selection bias</td>
<td>moderate</td>
<td>moderate</td>
<td>modest</td>
</tr>
</tbody>
</table>

Table 6.2: A comparison of macro factor extraction methods. The principal components are labeled as $PC_i$, Kalman Filter factors as $KF_i$ and 3 Pass Regression Factors as $3PRF_i$.

intuitive idea of what aspects of economy the factor captures. Figure 6.2 and Figure 6.1 show that PCs and Kalman Filter factors share their interpretations in common. The first factors from both methods load heavily on employment (see Panel A of Figure 6.1 and Figure 6.2) and the second factors capture sentiment and inflation (see Panel B). The third factors show some disagreements between PCA and Kalman Filter approaches, $PC_3$ favors output while $KF_3$ prefers housing. Figure 6.3 plots the principal components, Kalman Filter factors against some specific macro series. Indeed, the figure shows that PC1 and KF1 correlate with employment related factors like continuing jobless claims. PC2 and KF2 can be interpreted as inflation factors and PC3 and KF3 co-move with industrial production. Therefore, the three factors from both PCA and Kalman Filter can be labeled as employment, inflation and real growth factors.

Figure 6.4 pictures the $R^2$ of the 3PRF factors. The patterns are different from Figure 6.1 and 6.2 although there is overlap of information. In general, the 3PRF factors have more emphasis on the employment and sentiment aspects and less on output and inflation. Figure 6.5 plots the 3PRF factors (the red curves) against their implied macro indicators. The first 3PRF factor agrees with $PC_1$ and $KF_1$ in Figure 6.3, capturing the employment aspect of the economy. The deviations take place in the second and the third 3PRF factors, which turn out to be consumer sentiment and housing factors instead of the inflation and the real growth factors. A plausible reason is that since the interest rates have been proved to be counter cyclical, the yields panel has already contained the information about the real growth of the macro economic which excludes similar information from the model.

Whether the substantial cut on the number of variables, one might doubt, will hamper the predictive power in the macro series. As discussed by Boivin and Ng (2006), more data are not necessarily better for prediction. Still, in order to test the robustness of 3PRF factors, I delete the series with which the 3PRF factors $F$ have the largest correlation, and arrive at a new macro dataset $M'$ which has 3 less indicators than the original set $M$ and extract the 3PRF factors $F'$ following the same algorithm as before but on the reduce set $M'$. The blue curve in Figure 6.5 stands for $F'$, which almost overlaps the red curve ($F$) entirely. It implies that even when the most correlated series are omitted from the dataset, 3PRF is still able to capture the exact aspects of the economy and produce similar factors. This is partially due to the great overlapping of
Figure 6.1: $R^2$ in regression of individual macro series on each principal components. Due to the orthogonality properties of the PCs, the $R^2$ equals the square loadings on each factor.
Figure 6.2: $R^2$ in regression of individual macro series on each Kalman Filter factor.
Figure 6.3: Principal components and Kalman Filter factors plots with specific macro series. All series are standardized to have zero mean and unit variance.
Figure 6.4: $R^2$ in regression of individual macro series on each 3PRF factors.
information in the macro data and partially due to the properties of the 3PRF factors. The macro series overlap each other and the insensitivity of the factors to specific macro indicators implies that the cost of reducing the number of variables in the macro data set is bearable, if at all.

### 6.2.2 A comparison between 3PRFRF and PCAF

Turning to the second stage, the PCAF approach adopted by De Pooter, Ravazzolo, and van Dijk (2010) is to add macro factors directly into the regression model either as explanatory variables or as state variables (see Appendix B for details about PCAF). Alternatively, 3PRFRF approach adds macro factors in a selected way, as documented in section 2.3 and 2.4 for DNS-X and ATSM-X. Table 6.3 compares the Hit ratio of forecasts using the two approaches and prove that the 3PRFRF approach is able to deliver more accurate direction forecast than PCAF. The average of Hit ratio of DNS-X and ATSM-X models at both forecast horizons using 3PRFRF approach is significantly higher than the direct PCA approach. The improvement of 3PRFRF on PCAF can be attributed to 3 aspects.

1. Instead of iterated forecast in PCAF, 3PRFRF uses direct h-ahead forecast by looking for explanatory power in the h-month lagged regressors. While the direct forecast itself is by no means more advanced than iterated forecast, it requires no prediction of the macro factors and subsequently no time series models of the macro factors. By contrast, PCAF uses a VAR model to capture the dynamics of the macro factors, requires a stationary time series process of the macro factors. Subsequently, before performing PCAF, the macro dataset has to undergo necessary transformations for stationarity concern, which may lead to information loss. Nevertheless, iterated forecasts might give better performance under fast mean reversion conditions and long forecast horizon.\(^1\)

2. In case of DNS-X and ATSM-X, 3PRFRF only adds macro factors in the forecast stage but not in the estimation stage. As shown in Table 4.4, three factors are sufficient to describe the term structure of interest rates. If the information contained in the yields panel itself has explained more than 97% of the variation, there is little improvement potentials left for the macro factors. The Monte Carlo simulation results show that the identification becomes difficult when the dimension of the parameters space is expanded. After all, how much additional predictive power one benefits from adding macro factors directly into the state space model at the cost of parameter identification?

3. The 3PRFRF uses 3PRF methods and regression residual to extract factors which summarize macro information better than PCs. There are two drawbacks using principal components for forecasting interest rates. First, it is likely that information contained in PCs overlaps substantially with the information in the interest rate data, so that in a PC augmented regression, autoregressive terms absorb the influence of PCs, resulting in small and insignificant coefficients on PCs. Second, it requires prior knowledge

\(^1\)Indeed, in the 6 and 12 month forecast horizons which are not reported in the thesis, iterated forecasts perform marginally better for generic AR and VAR models.
Figure 6.5: 3PRF factors against specific macro series. “$F_i$” denotes the 3PRF factor $i$ derived from the full dataset. “$F'_i$” denotes the 3PRF factor $i$ derived from the reduced dataset. All series are standardized to have zero mean and unit variance.
of which macro PCs have the most relevant information on future yields. Stock and Watson (2002a), Stock and Watson (2005) have recognized that it is not the predominant principal components that are influencing the evolution of yields but rather the more trivial ones. Therefore instead of including the first 3 PCs, one might argue that it is actually a better practice to include, say, the first 2 and the 5th PCs. Furthermore, as the variance structure of the macro data changes with time, it might as well happen that the 5th PC which is proved be informative at time t switches to be the 6th PC at time t + 1 without any ex ante notice.

<table>
<thead>
<tr>
<th>Panel A: Hit Ratio Comparison 2000:1-2008:6, 1 month horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>DNS-X1m (3PRFRF)</td>
</tr>
<tr>
<td>DNS-X1m (PCAF)</td>
</tr>
<tr>
<td>ATSM-X1m (3PRFRF)</td>
</tr>
<tr>
<td>ATSM-X1m (PCAF)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Hit Ratio Comparison 2000:3-2008:8, 3 month horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>DNS-X3m (3PRFRF)</td>
</tr>
<tr>
<td>DNS-X3m (PCAF)</td>
</tr>
<tr>
<td>ATSM-X3m (3PRFRF)</td>
</tr>
<tr>
<td>ATSM-X3m (PCAF)</td>
</tr>
</tbody>
</table>

Table 6.3: Comparison of Hit ratio from two different approaches of incorporating macro factors. Hit ratio given by one approach that is significantly higher than the other at 90% confidence level is marked in bold.
Chapter 7

Trading Strategies

The instruments traded are limited to 2 year, 5 year and 10 year interest rate futures. Trading strategies are designed at 1 month and 3 month rebalance frequency on monthly basis. To elaborate, 1 month rebalance requires that at each month end, one clears the positions of last month and enters new positions. Whereas in the 3 month rebalance case, the positions of three months before are cleared and new positions are entered and will stay for three months until the next rebalance date. Although each position is held for three months but portfolio is still updated monthly. Trading strategies are constructed according to each model forecast, which are used to form the combined trading strategies. All strategies’ performances are evaluated with their information ratios.

7.1 Adjusting for risk exposure

Given the historical return volatility of the 2 year, 5 year and 10 year futures, $\sigma_2$, $\sigma_5$ and $\sigma_10$, and a reference amount $K_2$ on the maximum capital invested in the 2 year futures, a comparable risk exposure on the three types of futures is equivalent to:

$$\sigma_2 K_2 = \sigma_5 K_5 = \sigma_{10} K_{10}$$

(7.1)

where $K_5$ and $K_{10}$ denote maximum capital allocated to the 5 year and 10 year futures contracts.

7.2 Trading strategies based on individual model forecasts

Assume the $h$ month ahead predicted yield of tenor $\tau$ by model $m$ at time $t$ has the normal distribution

$$y_{t+h,\tau|t}^m \sim N(\hat{y}_{t+h,\tau|t}^m, \hat{\sigma}_{t+h,\tau|t}^m)$$

(7.2)

The probability that interest rate rises at time $t + h$ is given by

$$\hat{P}_{t+h,\tau|t}^m = P(y_{t+h,\tau|t}^m - y_{t,\tau} > 0) = \Phi(\frac{y_{t+h,\tau|t}^m - y_{t,\tau}}{\hat{\sigma}_{t+h,\tau|t}^m})$$

(7.3)
and the falling rate probability is
\[ \tilde{P}^m_{t+h,\tau|t} = 1 - \hat{P}^m_{t+h,\tau|t} \]  
(7.4)

The decision \( S^m_{t,\tau} \) given by model \( m \) on the positions of interest rate futures \( \tau \) at time \( t \) obeys the rule
\[ S^m_{t,\tau} = 1 - \tilde{P}^m_{t+h,\tau|t} = 1 - \hat{P}^m_{t+h,\tau|t} = 1 \]  
(7.5)

Allowing for mixed strategy turns \( S^m_{t,\tau} \) into a continuous score function \( s^m_{t,\tau} \)
\[ s^m_{t,\tau} = \tilde{P}^m_{t+h,\tau|t} + \hat{P}^m_{t+h,\tau|t} \]  
(7.6)

The value of the futures contract is the multiple of the score and the maximum capital
\[ C^m_{t,\tau} = s^m_{t,\tau} K_{\tau} \]  
(7.7)

Since the score is always between -1 and 1, the maximum capital serves as a budget constraint.

## 7.3 Trading strategy based on the Model Confidence Set selection

Taking the one month rebalance frequency as an example, the optimal strategy is selected weighted average of the individual trading strategies. The gain/loss \( g^m_{t+h,\tau} \) resulted from score \( s^m_{t,\tau} \) will be known at time \( t + h \) for \( h \) month rebalance frequency. The vector \( \vec{g}^m_{T+h,\tau} = (g^m_{t+h,\tau}, g^m_{t+h,\tau}, \ldots, g^m_{T+h,\tau}) \) contains the historical performance of trading strategy \( m \) from \( t = 1 \) to \( T \) evaluated at time \( T + h \).

Denote the volatility of \( \vec{g} \) by \( \sigma^m_{g,\tau} \). Before applying the Model Confidence Set once again to combine the best performed trading strategies into the optimal strategy, a loss function needs to be defined with the following properties

1. \( L^m_{t,\tau} > 0 \)
2. if \( g^m_{t+h,\tau} > g^n_{t+h,\tau} \), then \( L^m_{t+h,\tau} < L^n_{t+h,\tau} \)
3. if \( g^m_{t+h,\tau} = g^n_{t,\tau} \) and \( \sigma^m_{g,\tau} > \sigma^n_{g,\tau} \), then \( L^m_{t+h,\tau} < L^n_{t+h,\tau} \)

The idea is to reward large gains but punish volatility. Similar to the second scheme of forecast combination in Chapter 5.3, the elimination results in desirable number of individual strategies contained in the Model Confidence Set \( S \).

The optimal capital allocation on future contract \( \tau \) at time \( t + h \) is given by
\[ C^*_{t,\tau} = \sum_{m \in S} w_m C^m_{t,\tau} \]  
(7.8)

where \( w_m = \frac{1/L^m_{t,\tau}}{\sum_{m \in S} 1/L^m_{t,\tau}} \).
7.4 Performance of trading strategies

I compute the individual score functions from December 1999 to May 2008. The combined trading strategies is constructed from December 2004 to May 2008. Notice that a strategy at time $t$ can only be evaluated at time $t + h$, depending on the rebalance frequency. I choose the loss function as

$$L_{t,h,\tau}^m = \exp(|-v_{t,h,\tau} - s_{t,h,\tau}^m|) - 1 \quad (7.9)$$

where $v_{t,h,\tau}$ is the direction change of actual interest rates defined in Equation (5.3) and $s_{t,h,\tau}^m$ is the score function defined in Equation (7.6). The idea of the loss function is to compare the score to the “best position” using posterior knowledge, taking exponentials to account for volatility and $-1$ subtraction to have desirable magnitude.

The confidence level is set at 20%. The optimal strategies are constructed out of the expanding window evaluation of historical loss function values. The initial evaluation period is from January 2000 to December 2003. Table 7.1 reports the survival ratio for each model in both rebalance frequencies. The application of MCS to combined trading strategies is intended differently from that in forecast combination. In the latter case, the MCS is used to eliminate the worst performed models while in case of combined trading strategies, MCS is used to detect the best ones. The selection procedure is thus more strict in this case and so explains the lots of zeros in Table 7.1. The surviving ratios show that macro forecasts dominate the optimal strategies design. It confirms the value of macro factors can be exploited in the trading strategies. I also compare the trading strategies to Momentum strategy, which duplicates the “best position” of last period. The Momentum strategy works very well in the 1 month rebalance case and even outperforms the optimal strategy on the 2 year futures. When it comes to 3 month rebalance frequency, however, the Momentum strategy ends up with all negative information ratios.

Figure 7.1 plots the cumulated profits/loss for the combined trading strategies and Figure 7.2 and 7.3 visualizes the information ratio of all trading strategies. Notice that a negative information ratio does not necessarily imply a loss at the end, because the capital allocated is different for every time period. More capital is invested if models have more confidence in their direction forecasts which is not reflected in the information ratio. Another illustration of the how information ratio ignores the variation in capital allocation is that the profits on 2 years futures in Panel B of Figure 7.1 amounts to large figures but the information ratio shown in Panel A of Figure 7.3 remains modest. The information ratios of trading strategies from models with and without macro factors are drawn next to each other for comparison in Figure 7.2 and 7.3. “E.W.” stands for Equally weighted scheme, which is the trading strategy constructed by averaging the positions of all individual trading strategies. The figures show that trading strategies from macro factors deliver a higher information ratio.
### Panel A: Model Surviving Ratio 2004:1-2008:6, 1 month rebalance

<table>
<thead>
<tr>
<th></th>
<th>2 year</th>
<th>5 year</th>
<th>10 year</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>0.00</td>
<td>0.00</td>
<td>0.28</td>
<td>0.09</td>
</tr>
<tr>
<td>VAR</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>ATSM</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>DNS</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AR-X</td>
<td>0.02</td>
<td>0.09</td>
<td>1.00</td>
<td>0.37</td>
</tr>
<tr>
<td>VAR-X</td>
<td>0.00</td>
<td>0.00</td>
<td>0.43</td>
<td>0.14</td>
</tr>
<tr>
<td>ATSM-X</td>
<td>1.00</td>
<td>1.00</td>
<td>0.28</td>
<td>0.76</td>
</tr>
<tr>
<td>DNS-X</td>
<td>0.98</td>
<td>0.91</td>
<td>0.00</td>
<td>0.63</td>
</tr>
</tbody>
</table>

### Panel B: Model Surviving Ratio 2004:3-2008:8, 3 month rebalance

<table>
<thead>
<tr>
<th></th>
<th>2 year</th>
<th>5 year</th>
<th>10 year</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>VAR</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>ATSM</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>DNS</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AR-X</td>
<td>0.83</td>
<td>1.00</td>
<td>1.00</td>
<td>0.94</td>
</tr>
<tr>
<td>VAR-X</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>ATSM-X</td>
<td>1.00</td>
<td>0.00</td>
<td>0.47</td>
<td>0.49</td>
</tr>
<tr>
<td>DNS-X</td>
<td>0.49</td>
<td>1.00</td>
<td>0.53</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Table 7.1: Model surviving ratio in combined trading strategies construction. The confidence level is set at 20% while maintaining at least two models in the MCS.
Figure 7.1: Cumulated Profits/loss of combined trading strategies
Figure 7.2: Information ratio, 1 month rebalance frequency
Figure 7.3: Information ratio, 3 month rebalance frequency
Chapter 8

Analysis of interest rate models in the period of financial crisis

Now it’s time to extend the forecast and trading strategies to the abnormal period of financial crisis after May 2008. Since the previous sections have proved that the 3PRF Residual forecast does a better job in interest rate forecast as well as the trading strategies design, I apply the same approach to the last sample from July 2008 to May 2011. Figure 8.1 plots forecasts of the short rate against the actual realized value. The figure gives two remarks about this particular period. First, DNS and ATSM give outrageous forecasts of short rate compared to AR and VAR. The failure of DNS and ATSM models can be attributed to the violation of the cross section assumptions, which implies existence of arbitrage opportunities during crisis. Second, comparing forecast performances of models with macro factors to those without, it seems that macro factors are not as illuminating as they used to be, if not misleading. The result does not come as a surprise since similar patterns are also observed during the last crisis around year 2002, see short rate forecast in Figure 5.1. Figure 8.2, 8.3 and 8.4 are 2 year, 5 year and 10 year plots. For intermediate tenors, structured models’ forecasts are not as outstanding as they are in the short rate forecast but still the forecast valleys created by macro data in the middle of 2008 are much lower than actual values. The story behind the observation is that both momentum and macro fundamentals insist that the interest rates would continue to fall since there is no restriction which guarantees that yields can’t go negative. In reality, the Fed would rather turn to quantitative easing or fiscal policies to stimulate the economy than to stabilize interest rate marginally above 0.

Poor forecast performance results in poor trading strategies. Figure 8.5 and 8.6 draw the information ratio of the trading strategies of the models. While the information ratios on 2 years futures (see Panel A of Figure 8.5 and 8.6) are still satisfactory given the positive returns generated by macro models, the trading strategies of 5 and 10 year futures are rather disappointing. Macro factors do not consistently add value to the design of trading strategies any more and models like DNS-X and ATSM-X which used to dominate the optimal trading strategies’ composition, turn out to be the worse ones, if not the worst. By
Figure 8.1: 3 month yield forecast during financial crisis

Figure 8.2: 2 year yield forecast during financial crisis
Figure 8.3: 5 year yield forecast during financial crisis

Figure 8.4: 10 year yield forecast during financial crisis
contrast, the momentum strategy during the crisis performs surprisingly well, almost beating all other trading strategies.

To sum up, because the assumption of no-arbitrage opportunities does not hold in the financial crisis and macro fundamentals fail to point out the right direction of interest rate change, the conclusions drawn from normal period no longer apply. It is sometimes the simple models that give better forecast and trading strategies rather than structured models. To adapt the structured models to abnormal period might involve regime switching assumptions which I leave for future research.
Figure 8.5: Information ratio during crisis, 1 month rebalance frequency
Figure 8.6: Information ratio during crisis, 3 month rebalance frequency
Chapter 9

Conclusion

The thesis conducts estimation and forecast using interest rates models of AR, VAR, Dynamic Nelson Siegel and Affine Term Structure model. I demonstrate the application of 3PRFRF approach to perform forecasts using macro factors. I combine the individual forecasts using two schemes and evaluate the forecasts in various dimensions. I comment on the mechanisms and pros and cons of the interest rate models as well as the macro incorporation methods with the estimation and forecast results from the previous chapters. The forecasts are further carried on to design trading strategies on the 2-year, 5-year and 10-year futures contracts. I show how to construct an combined trading strategies by picking and weighted averaging the historically best performed individual trading strategy at each time period using the Model Confidence Set approach. I evaluate both individual and optimal trading strategies by their information ratios. Lastly, I extend the forecast and trading strategy construction framework to recent period of financial crisis.

The main findings of the paper fall into three domains. First, the comparison of the estimation and forecast results give insights of the interest rate models. From sheer time series model, AR, to highly structured model ATSM, more market factors are accounted for. Simple models perform better at short forecast horizon when interest rates are mostly driven by momentum. As the forecast horizon extends, structured models become more effective. Variation of forecast performance also differs across tenors within a single model. Time series models specialize in forecasting the short end of the yield curve while structured models do better at intermediate tenors. The fact that the latent states dynamics from DNS and ATSM estimations exhibit significant similarities implies, first, no-arbitrage assumption might be embedded in the Nelson Siegel model; second, there could over be identification in the affine model since the Dynamic Nelson Siegle model demonstrates that one parameter already suffices to ensure no-arbitrage. Second, the forecast results and the performance of trading strategies prove that macro data has forecast power for interest rates and can be exploited if incorporated properly. A comparison of the Hit ratio of the forecasts from with and without macro factors shows that macro factors are particularly useful in direction prediction. The information ratios of trading strategies further confirm the fact that the prediction power contained in macro data can be exploited to produce a positive return. On the other hand, the way to incorporate macro data is crucial. I show that the PCAF approach of incorporating macro data gives
less satisfactory results than 3PRFRF approach. Third, in abnormal period like financial crisis, cross section models typically perform worse than simple time series models, which might be regarded as a signal of existence of arbitrage opportunities. In addition, macro fundamentals no longer work as effectively as they do in the normal period. The financial crisis study warns practitioners’ to be more cautious in applying fundamental studies to abnormal period.

The paper motivates future research in a number of areas. First, the analogy of Dynamic Nelson Siegel to Affine Term Structure Models enables no-arbitrage interpretation of the parameter $\lambda$. It raises the question of decomposing $\lambda$ to see the how the no-arbitrage assumption is built in. It might be possible to come up with a regression model to estimate $\lambda$ on the basis of the first fundamental theory of asset pricing. Second, the byproducts of the 3PRFRF approach, the 3PRF factors as well as their loadings, contain substantial information about the dynamic macro economy driving force of interest rate changes, which may be valuable for fundamental studies. Third, the poor performance of structured interest rate models during financial crisis possibly indicates a switching of regime in the crisis. If so, a regime switching model may produce better forecasts in this regard.
Appendix A

Monte Carlo simulations

Table A.1 reports the 1000 Monte Carlo Simulation results for DNS and ATSM. Parameters of variance of regression residuals are not reported, since variance parameters are typically very well identified. The Table shows that DNS identifies true parameters effectively whereas there are identification problems with regard to $\delta_1$ and $\Sigma \lambda_1$. 
### Panel A: Monte Carlo Simulation results for DNS

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True par</th>
<th>first quartile</th>
<th>Median</th>
<th>third quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>0.719</td>
<td>0.681</td>
<td>0.717</td>
<td>0.759</td>
</tr>
<tr>
<td></td>
<td>-0.144</td>
<td>-0.187</td>
<td>-0.147</td>
<td>-0.105</td>
</tr>
<tr>
<td></td>
<td>-0.0482</td>
<td>-0.0665</td>
<td>-0.0484</td>
<td>-0.0304</td>
</tr>
<tr>
<td>$\text{diag}(\Gamma)$</td>
<td>0.7</td>
<td>0.656</td>
<td>0.689</td>
<td>0.721</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.756</td>
<td>0.788</td>
<td>0.816</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.548</td>
<td>0.589</td>
<td>0.623</td>
</tr>
</tbody>
</table>

### Panel B: Monte Carlo Simulation results for ATSM

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True par</th>
<th>first quartile</th>
<th>Median</th>
<th>third quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{diag}(\Psi)$</td>
<td>0.905</td>
<td>0.899</td>
<td>0.949</td>
<td>0.979</td>
</tr>
<tr>
<td></td>
<td>0.994</td>
<td>0.888</td>
<td>0.925</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td>0.993</td>
<td>0.957</td>
<td>0.973</td>
<td>0.985</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.0556</td>
<td>-0.158</td>
<td>0.0303</td>
<td>0.459</td>
</tr>
<tr>
<td></td>
<td>0.485</td>
<td>-0.0129</td>
<td>0.0139</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>-0.131</td>
<td>0.136</td>
<td>0.347</td>
<td>0.628</td>
</tr>
<tr>
<td>$\text{vech}(\Sigma_{\lambda_1})$</td>
<td>0.0493</td>
<td>-0.0263</td>
<td>-0.00947</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>-0.00663</td>
<td>-0.0772</td>
<td>-0.000403</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>2.06</td>
<td>-0.498</td>
<td>-0.0764</td>
<td>0.459</td>
</tr>
<tr>
<td></td>
<td>-0.00336</td>
<td>-0.0141</td>
<td>0.0257</td>
<td>0.0798</td>
</tr>
<tr>
<td></td>
<td>0.382</td>
<td>-0.455</td>
<td>-0.0984</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>0.13</td>
<td>0.0586</td>
<td>0.102</td>
<td>0.132</td>
</tr>
</tbody>
</table>

Table A.1: Monte Carlo Simulation results. The first column specifies the parameters in DNS regression Equation (2.20) and ATSM equations (2.28), (2.29) and (2.30). “True Par” refers to the data generating parameters’ value; “first quartile” refers to 25th percentile of the estimates; “Median” refers to the 50th percentile and “third quartile” refers to 75th percentile.
Appendix B

PC Augmented Forecasting approach of incorporating macro factors

The PC Augmented Forecasting (PCAF) approach of incorporating macro factors are adopted by De Pooter, Ravazzolo, and van Dijk (2010). Under PCAF, DNS-X is constructed as

\[
\begin{pmatrix}
\beta_{t+1} \\
\alpha_{t+1} \\
PC_t \\
Y_t
\end{pmatrix} =
\begin{pmatrix}
a \\
0 \\
0 \\
0
\end{pmatrix} +
\begin{pmatrix}
\Gamma_1 & \Gamma_2 \\
0 & \Lambda \\
0 & 0 \\
X & 0
\end{pmatrix}
\begin{pmatrix}
\beta_t \\
\alpha_t \\
PC_t \\
Y_t
\end{pmatrix} +
\begin{pmatrix}
\varepsilon_t \\
\eta_t \\
0 \\
u_t
\end{pmatrix}
\]

(B.1)

where \( PC_t \) is a \( 3 \times 1 \) vector of macro factors. PC Augmented approach uses iterated forecast rather than direct forecast since the dynamics of macro factors are modeled explicitly. The \( h \) ahead forecast of \( \hat{y}_{t+h} \) is given by

\[
\hat{y}_{t+h} = X\hat{\beta}_{t+h}
\]

(B.2)

where \( \hat{beta}_{t+h} \) is computed iteratively through

\[
\hat{\beta}_{t+h} = \hat{\Gamma}_1\hat{\beta}_{t+h-1} + \hat{\Gamma}_2\hat{\alpha}_{t+h-1}
\]

(B.3)

\[
\hat{\alpha}_{t+h-1} = PC_{t+h-1} = \Lambda PC_{t+h-2}
\]

(B.4)

The application of PCAF to ATSM-X follows

\[
\begin{pmatrix}
Z_{t+1} \\
\alpha_{t+1} \\
PC_t \\
Y_t
\end{pmatrix} =
\begin{pmatrix}
\mu \\
0 \\
0 \\
a
\end{pmatrix} +
\begin{pmatrix}
\Psi & 0 \\
0 & \Lambda \\
0 & 0 \\
I & b
\end{pmatrix}
\begin{pmatrix}
Z_t \\
\alpha_t \\
PC_t \\
Y_t
\end{pmatrix} +
\begin{pmatrix}
\varepsilon_t \\
\eta_t \\
0 \\
u_t
\end{pmatrix}
\]

(B.5)

Notice that unlike DNS-X where \( PC_t \) does not enter the measurement equation directly, it has direct influence on \( y_t \) in the ATSM-X case. The forecast is given by

\[
\hat{y}_{t+h} = a + b \begin{pmatrix}
\hat{Z}_{t+h} \\
\hat{\alpha}_{t+h}
\end{pmatrix}
\]

(B.6)
where \( \left( \hat{Z}_{t+h} \right) \) are computed iteratively through

\[
\begin{pmatrix}
\hat{Z}_{t+h} \\
\hat{\alpha}_{t+h}
\end{pmatrix} =
\begin{pmatrix}
\hat{\mu} \\
0
\end{pmatrix} +
\begin{pmatrix}
\hat{P}_{si} & 0 \\
0 & \hat{\Lambda}
\end{pmatrix}
\begin{pmatrix}
\hat{\beta}_{t+h-1} \\
PC_{t+h-1}
\end{pmatrix}
\]

(B.7)

Since the PCAF approach involves VAR model of macro factors dynamics, the macro data has to undergo necessary transformation to ensure stationarity.

Stacking macro factors into the state vectors but not necessarily PCs are also adopted by Diebold and Li (2006), Ang and Piazzesi (2003), Moench (2008), etc by simply substitute \( PC_t \) by other macro factors.
Appendix C

Kalman Filter estimation of macro factors

Kalman Filter estimation of macro factors are practiced by Joslin, Le, and Singleton (2011). Denote the extracted macro factors as $K$, the estimation model is given by

$$K_t = \Phi K_{t-1} + \eta_t \quad \text{(C.1)}$$

$$M_t = BK_t + \Lambda M_{t-1} + \varepsilon_t \quad \text{(C.2)}$$

where $M_t$ is the full macro dataset and $\Lambda$ is restricted to be a diagonal matrix.
Appendix D

List of macro series in the macroeconomic dataset

Table D.1 lists the number, category, description, frequency, timeliness and revision condition of macro series in the macroeconomic dataset used for 3PRF factors extraction. Table reftbmacroseries2 lists additional indicators in the Kalman Filter estimation. “MoM” stands for “Month over Month change”; “SA” means “Seasonal Adjusted”; “SAAR” denotes “Seasonal Adjusted Annualized Rate”. Indicators of monthly released frequency is denoted by “M”, weekly frequency by “W” and “D” for intraday frequency. “Timeliness” refers to the time lags between accrued date and indicator’s released date. Indicators of 1 month’s lag is labeled “LM”, “L^2M” for the 2 months’ lags and “U” for current month value available. The revisions are measured using the signal-to-noise ratio. I group indicators with a signal-to-noise ratio smaller than .6 as series with large revisions and are labeled “l”, those whose signal-to-noise-ratios are between .6 and .9 are grouped to series with medium revisions and are labeled “m” and the rest belong to indicators with small revisions and are labeled “s”.
<table>
<thead>
<tr>
<th>No.</th>
<th>Category</th>
<th>Indicators</th>
<th>Frequency</th>
<th>Timeliness</th>
<th>Revision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Output</td>
<td>personal income</td>
<td>M</td>
<td>LM</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>ISM manufacturing</td>
<td>M</td>
<td>LM</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Chicago purchasing managers</td>
<td>M</td>
<td>U</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Philadelphia Fed business outlook survey</td>
<td>M</td>
<td>U</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>industrial production (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>capacity utilization</td>
<td>M</td>
<td>LM</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>composite index of leading economic indicators</td>
<td>M</td>
<td>LM</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>manufacturing and trade inventories and sales</td>
<td>M</td>
<td>$^2M$</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>durables ex transportation (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>manufacturers’ new orders (MoM, SA)</td>
<td>M</td>
<td>$^2M$</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Employment</td>
<td>unemployment rate (SA)</td>
<td>M</td>
<td>LM</td>
<td>s</td>
</tr>
<tr>
<td>12</td>
<td>change in nonfarm payrolls (SA)</td>
<td>M</td>
<td>LM</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>change in nonfarm payrolls (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>continuing jobless claims (SA)</td>
<td>W</td>
<td>LM</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>initial jobless claims (SA)</td>
<td>W</td>
<td>LM</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Consumer Sentiment</td>
<td>auto sales (SAAR)</td>
<td>M</td>
<td>LM</td>
<td>m</td>
</tr>
<tr>
<td>17</td>
<td>personal consumption expenditures (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>consumer confidence</td>
<td>M</td>
<td>U</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>consumer sentiment</td>
<td>M</td>
<td>U</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>dollar Swiss Franc FX</td>
<td>D</td>
<td>U</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>dollar Japanese Yen FX</td>
<td>D</td>
<td>U</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>S&amp;P 500 index</td>
<td>D</td>
<td>U</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Housing</td>
<td>housing starts (SAAR)</td>
<td>M</td>
<td>LM</td>
<td>s</td>
</tr>
<tr>
<td>24</td>
<td>building permits (SAAR)</td>
<td>M</td>
<td>LM</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>new single-family home sales (SAAR)</td>
<td>M</td>
<td>LM</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Inflation</td>
<td>CPI urban change (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>m</td>
</tr>
<tr>
<td>27</td>
<td>CPI urban change less food &amp; energy (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>PCE core price index (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>PPI for finished goods (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>PPI ex food &amp; energy (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>M1 Index</td>
<td>W</td>
<td>U</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>M2 Index</td>
<td>W</td>
<td>U</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>monetary base (SA)</td>
<td>M</td>
<td>LM</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Miscellaneous</td>
<td>consumer credit (SA)</td>
<td>M</td>
<td>$^2M$</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>Federal budget balance</td>
<td>M</td>
<td>LM</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Aggregate Reserves Depository Institutions (SA)</td>
<td>W</td>
<td>U</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Aggregate Reserves Depository Institutions Non Borrowed (SA)</td>
<td>W</td>
<td>U</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>treasury international capital system</td>
<td>M</td>
<td>$^2M$</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>foreign net transactions</td>
<td>M</td>
<td>$^2M$</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>Federal Funds target rate</td>
<td>D</td>
<td>U</td>
<td>s</td>
<td></td>
</tr>
</tbody>
</table>

Table D.1: Macro series universe for 3PRF and PCA extraction. All indicators are full monthly series without missing values.
### Additional series in Kalman Filter estimation

<table>
<thead>
<tr>
<th>No.</th>
<th>Category</th>
<th>Indicators</th>
<th>Frequency</th>
<th>Timeliness</th>
<th>Revision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Output</td>
<td>GDP advanced</td>
<td>Q</td>
<td>L³M</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>ISM nonmanufacturing</td>
<td>M</td>
<td>LM</td>
<td>s</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Dallas Fed manufacturing outlook survey</td>
<td>M</td>
<td>U</td>
<td>s</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Empire state manufacturing survey</td>
<td>M</td>
<td>U</td>
<td>s</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Total Milwaukee Index</td>
<td>M</td>
<td>U</td>
<td>s</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>wholesale inventories</td>
<td>M</td>
<td>L²M</td>
<td>m</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>durable goods orders (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>l</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>manufacturers’ new orders (MoM, SA)</td>
<td>M</td>
<td>L²M</td>
<td>s</td>
</tr>
<tr>
<td>9</td>
<td>Employment</td>
<td>ADP National Employment Report (SA)</td>
<td>M</td>
<td>LM</td>
<td>m</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>help-wanted advertising index</td>
<td>M</td>
<td>LM</td>
<td>s</td>
</tr>
<tr>
<td>11</td>
<td>Consumer Sentiment</td>
<td>ABC news consumer comfort index</td>
<td>W</td>
<td>U</td>
<td>s</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>retail sales (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>m</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>retail sales less auto (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>m</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>chain store sales (SA)</td>
<td>W</td>
<td>U</td>
<td>l</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>personal consumption change (SAAR)**</td>
<td>Q</td>
<td>L³M</td>
<td>m</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>real effective exchange rate</td>
<td>Q</td>
<td>U</td>
<td>s</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>S&amp;P 500 index</td>
<td>Q</td>
<td>U</td>
<td>s</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>S&amp;P 500 industrials index</td>
<td>Q</td>
<td>U</td>
<td>s</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>S&amp;P 500 dividend yields</td>
<td>Q</td>
<td>U</td>
<td>s</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>Volatility index</td>
<td>Q</td>
<td>U</td>
<td>s</td>
</tr>
<tr>
<td>21</td>
<td>Housing</td>
<td>existing home sales (SAAR)</td>
<td>M</td>
<td>LM</td>
<td>s</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>existing home sales (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>s</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>pending home sales (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>m</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>construction spending (MoM, SA)*</td>
<td>M</td>
<td>L²M</td>
<td>l</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>housing market index (SA)</td>
<td>M</td>
<td>LM</td>
<td>s</td>
</tr>
<tr>
<td>26</td>
<td></td>
<td>house price index (MoM)</td>
<td>M</td>
<td>L²M</td>
<td>m</td>
</tr>
<tr>
<td>27</td>
<td></td>
<td>mortgage application index (weekly change, SA)</td>
<td>W</td>
<td>LW</td>
<td>s</td>
</tr>
<tr>
<td>28</td>
<td></td>
<td>S&amp;P/Case-Shiller home price index</td>
<td>M</td>
<td>L²M</td>
<td>s</td>
</tr>
<tr>
<td>29</td>
<td></td>
<td>S&amp;P/Case-Shiller home price index (YoY)</td>
<td>M</td>
<td>L²M</td>
<td>s</td>
</tr>
<tr>
<td>30</td>
<td>Inflation</td>
<td>GDP-based price index (QoQ SAAR)</td>
<td>Q</td>
<td>U</td>
<td>l</td>
</tr>
<tr>
<td>31</td>
<td></td>
<td>employment cost index (QoQ, SA)</td>
<td>Q</td>
<td>L³M</td>
<td>s</td>
</tr>
<tr>
<td>32</td>
<td></td>
<td>productivity and cost (QoQ, SA)</td>
<td>Q</td>
<td>L³M</td>
<td>m</td>
</tr>
<tr>
<td>33</td>
<td></td>
<td>unit labor cost (QoQ, SAAR)</td>
<td>Q</td>
<td>L³M</td>
<td>l</td>
</tr>
<tr>
<td>34</td>
<td></td>
<td>import price (MoM, SA)</td>
<td>M</td>
<td>LM</td>
<td>s</td>
</tr>
<tr>
<td>35</td>
<td>Miscellaneous</td>
<td>commercial and industrial loans</td>
<td>Q</td>
<td>U</td>
<td>s</td>
</tr>
<tr>
<td>36</td>
<td></td>
<td>Government debt</td>
<td>Q</td>
<td>L³M</td>
<td>s</td>
</tr>
<tr>
<td>37</td>
<td></td>
<td>government debt growth</td>
<td>Q</td>
<td>L³M</td>
<td>s</td>
</tr>
<tr>
<td>38</td>
<td></td>
<td>current account balance</td>
<td>Q</td>
<td>L³M</td>
<td>s</td>
</tr>
<tr>
<td>39</td>
<td></td>
<td>Citigroup surprise index</td>
<td>M</td>
<td>LM</td>
<td>s</td>
</tr>
</tbody>
</table>

Table D.2: Macro series universe which are included in the Kalman Filter estimation but not in the 3PRF or PCA extraction. These are the indicators which are released of quarterly frequency or with missing values.


