Macro Factors, Monetary Policy Analysis and Affine Term Structure Models

Lasse Bork

Ph.D. dissertation

Finance Research Group
Department of Business Studies
Aarhus School of Business, Aarhus University

Thesis advisor: Tom Engsted

January 2010

AARHUS SCHOOL OF BUSINESS
AARHUS UNIVERSITY

Contents

Preface

This thesis marks the end of my PhD studies at the Aarhus School of Business, University of Aarhus. This PhD project originated in 2005 when I was inspired to do a PhD by reading a seminal paper by Andrew Ang & Monika Piazzesi published in the Journal of Monetary Economics 2003. In this paper they show how to link macroeconomics and no-arbitrage financial theory. Back then, I worked as a fixed income analyst at Jyske Bank and I immediately found and still find the approach theoretically interesting but also practically relevant in bond market research. I therefore proposed Jyske Bank to sponsor my PhD project and I am grateful that the bank offered me the opportunity to pursue a PhD, which has proven to be very rewarding professionally and personally. I would like to thank my company supervisors Mikkel Røgild, Jan Mark Sørensen and Morten R. Fræmohs for their support and Peter F. Andersen for making this project feasible.

I would also like to express my sincere gratitude to my thesis advisor Tom Engsted for his guidance and ready assistance throughout my years as a PhD student. This thesis would have been less readable, if he had not carefully read and commented on my early drafts.

During the course of my studies, I enjoyed the friendly environment at the Department of Business Studies and I would like to thank colleagues and fellow PhD students in the Finance Research Group for making my time pleasant, fun, and rewarding.

Among the highlights of my study time were my stays at the Department of

Economics, Catholic University of Leuven in Belgium. I would like to thank my co-advisor Hans Dewachter for inviting me to be part of an outstanding, pleasant, and stimulating international research environment from which I benefitted enormously academically as well as personally. One of the chapters in this thesis is co-authored by Hans Dewachter and it has been a great privilege to have the opportunity to work with him. I would also like to thank my co-author Romain Houssa and my fellow PhD students Leonardo Iania and Pablo Rovira Kaltwasser for their hospitality and for providing such a friendly atmosphere.

Outside academia, I would like to thank my family – parents and parents in law – for their continuous support and encouragement during the years of my study.

Finally and most important of all, I would like to thank my wife Mette for her unconditional love, support and encouragement during all my ups and downs. She has been a safe haven for me during troubled times and a rock solid home base for our wonderful three daughters Camilla, Caroline and Laura. It goes without saying that I dedicate this thesis to Mette.

Skanderborg, August 2009
Lasse Bork

Updated preface

I received the evaluation report on my PhD thesis December 15, 2009.

I am grateful to the members of the assessment committee, Paul Söderlind, Ken Nyholm and Jan Bartholdy, for their careful reading of the dissertation and their many useful comments and suggestions. Many of the suggestions have been incorporated in the present version of the thesis while others remain for future work on the chapters.

Skanderborg, January 2010
Lasse Bork

Introduction to the thesis

This thesis is positioned in the research area represented by the intersection of macroeconomics, monetary policy and the financial markets. The thesis consists of three chapters, which make empirical and methodological contributions to the field of empirical monetary policy analysis and dynamic macro-finance models of the yield curve.

On a non-technical level, I propose an econometric model to evaluate empirically, how a surprise change in the US monetary policy instrument (the short-term interest rate) affects a large set of key macroeconomic variables. This enables us to assess the typical macroeconomic outcome, following an unexpected change in the monetary policy instrument.

In the above monetary policy analysis, I do not analyze potential determinants of the interest-rate setting by the central bank. This would require the monetary policy instrument to be expressed as a function of relevant macroeconomic variables; a so-called policy rule. However, questions of what should be and appears to be the economic determinants of the monetary policy rate have been discussed in a large volume of papers. In a widely cited paper, Taylor (1993) estimates a particular simple policy rule, which can be characterized as a "lean against the wind" policy. Intuitively, the central bank increases the interest rate if economic activity expands beyond its natural or potential level, or if inflation exceeds some desired rate of inflation, or both. Accordingly, measures of inflation and economic activity should be included in the set of candidate explanatory variables for the short-term interest rate.

The short-term interest rate is a crucial component in modern no-arbitrage yield curve models; aka dynamic term structure models. However, until recently macroeconomic determinants of the short-term interest rate have been largely absent in standard dynamic term structure models. As such, these models do not reflect how central banks implement their monetary policy. Consequently, I propose to describe the bond market behavior, in terms of bond pricing, by an econometric model that includes macroeconomic determinants of the bond yields. This model is then evaluated in terms of how surprises to the included macroeconomic determinants (variables) affect the yield curve.

The policy makers at the Federal Reserve Bank and the bond market participants have one thing in common: They have an abundant amount of information at their disposal, and as such the information set on which they condition the interest rate setting, and bond pricing respectively, is large. Consequently, a recurrent theme of this thesis is the approximation of the large information sets by a large panel of macroeconomic and financial time series. In particular, this thesis advances the use of dynamic factors, to approximate the conditioning information set in both monetary policy analysis and in an affine macro-finance model of the term structure. By construction, only a few of these factors are able to summarize the bulk of the information of potentially hundreds of observed time series.

In the first chapter entitled *"Estimating US Monetary Policy Shocks Using a Factor-Augmented Vector Autoregression: An EM algorithm Approach"* the economy-wide effects of shocks to the US monetary policy instrument (the federal funds rate) are estimated using an iterative maximum likelihood estimation method. The data description of the US economy is confined to a large cross-section of 120 macroeconomic and financial time series and the comovement of these time series over time is shown to be adequately described in terms of a few dynamic latent driving forces (dynamic factors) and the US federal funds rate. Technically, the 120 time series constitute the measured part in a state space system. The state transition part of this system contains the dynamics of the driving forces and is represented as a vector autoregression of the federal funds rate augmented by a few dynamic factors extracted from the large cross-section of time series. The complete state space system in turn allows for an empirical study of the response of each of the 120 observed variables following a shock to the federal funds rate. The methodological contribution of this chapter is the one-step fully parametric estimation of the Factor-Augmented VAR (FAVAR) by means of the EM algorithm as an alternative to the two-step principal component method and the one-step Bayesian method in Bernanke et al. (2005). I demonstrate empirically that the same impulse responses but better fit emerge robustly from a low order FAVAR with eight correlated factors compared to a high order FAVAR with fewer correlated factors, for instance four factors. This empirical result accords with one of the theoretical results from Bai & Ng (2007) in which it is shown that the information in complicated factor dynamics may be substituted by panel information

The dynamic factors estimated in the first chapter capture reasonably well the observed time series associated with the most dominant loading on the particular factor. A standard procedure in the literature amounts to inferring the economic interpretation of a particular dynamic factor from the dominant factor loading. However, this approach neglects the non-dominant (but possibly significant) loadings and hence does not generate unambiguous and well-defined interpretations of the factors.

The second chapter *"Identification of Macroeconomic Factors in Large Panels"* is written jointly with Hans Dewachter and Romain Houssa. In this paper we address the ambiguous economic interpretation of the exactly identified dynamic factors by using a procedure that imposes a specific and well-defined interpretation of the factors. The economic interpretation of the extracted factors is based on a set of overidentifying restrictions on the factor loadings. This model is still a Factor-Augmented Vector Autoregression, but it is now subject to linear loading restrictions. However, we show how the estimator for the loadings subject to linear restrictions can be stated in closed form within the EM algorithm. We apply this framework to the same panel of US macroeconomic series as in the first chapter of this thesis. In particular, we identify nine macroeconomic factors and discuss the economic impact of monetary policy shocks. We find that the results are theoretically plausible and in line with other findings in the literature.

In the third chapter entitled *"A multifactor Affine Term Structure Model with Macroeconomic Factors from Large Panels"* I approximate the potentially large information set of the bond market by the large panel of macroeconomic and financial time series used in the former chapters. The main motivation for the use of an expanded information set is the fact that the financial markets monitor and respond to a large set of macroeconomic variables in the process of filtering the underlying development in key macroeconomic variables. I propose to solve the bond market's filtering problem by a large panel dynamic factor analysis to derive a small set of macroeconomic state variables. In fact, these macroeconomic state variables are a subset of the well-defined macroeconomic factors derived in chapter 2. A discrete-time dynamic term structure model is then augmented with these filtered macroeconomic state variables. The focus in the chapter is primarily on bond risk premia and a forecast error variance decomposition shows that shocks to inflation and in particular unemployment are important for the risk premia on long-term bonds.

CHAPTER 1

Estimating US Monetary Policy Shocks Using a Factor-Augmented Vector Autoregression: An EM algorithm Approach

Lasse Bork[†]

January 2010

[†]Finance Research Group, Department of Business Studies, Aarhus School of Business, University of Aarhus, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. Affiliated with CREATES at the University of Aarhus, a research center funded by the Danish National Research Foundation. Email: lbo@lassebork.dk.

Abstract[*]

Economy-wide effects of shocks to the US federal funds rate are estimated in a state space model with 120 US macroeconomic and financial time series driven by the dynamics of the federal funds rate and a few dynamic factors. This state space system is denoted a factor-augmented VAR (FAVAR) by Bernanke et al. (2005). I estimate the FAVAR by the fully parametric one-step EM algorithm as an alternative to the two-step principal component method and the one-step Bayesian method in Bernanke et al. (2005). The EM algorithm which is an iterative maximum likelihood method estimates all the parameters and the dynamic factors simultaneously and allows for classical inference. I demonstrate empirically that the same impulse responses but better fit emerge robustly from a low order FAVAR with eight correlated factors compared to a high order FAVAR with fewer correlated factors, for instance four factors. This empirical result accords with one of the theoretical results from Bai & Ng (2007) in which it is shown that the information in complicated factor dynamics may be substituted by panel information.

JEL classifications: E3, E43, E51, E52, C33.
Keywords: Monetary policy, large cross-sections, factor-augmented vector autoregression, EM algorithm, state space.

[*] I thank Tom Engsted and Stig Vinther Møller for helpful comments. Any remaining errors are my own.

1.1 Introduction

This paper estimates the "economy-wide" response to shocks to the US federal funds rate using an iterative maximum likelihood estimation method. The data description of the US economy is confined to a large cross-section of 120 macroeconomic and financial time series and the comovement of these time series over time is shown to be adequately described in terms of a few dynamic latent driving forces (dynamic factors) and the US federal funds rate. Technically, the 120 time series constitute the measured part in a state space system. The state transition part of this system contains the dynamics of the driving forces and is represented as a vector autoregression of the federal funds rate augmented by a few dynamic factors extracted from the large cross-section of time series. The complete state space system in turn allows for an empirical study of the response of each of the 120 observed variables following a shock to the federal funds rate.

This setup is what Bernanke et al. (2005) denote a factor-augmented vector autoregressive (FAVAR) approach and this paper is closely related to both their approach and the data used. While Bernanke et al. (2005) estimate their FAVAR using both a two-step semi-parametric principal component method and a one-step Bayesian likelihood method, this paper contributes to the literature by estimating the FAVAR by a one-step fully parametric iterative maximum likelihood method, the Expectation Maximization (EM) algorithm. In fact, several of the future research issues that Bernanke et al. (2005) address in their conclusion are cited below and discussed in this paper:

"Future work should investigate more fully the properties of FAVARs, alternative estimation methods and alternative identification schemes. In particular, further comparison of the estimation methods based on principal components and on Gibbs sampling is likely to be worthwhile. Another interesting direction is to try to interpret the estimated factors more explicitly". Bernanke et al. (2005) page 415, §3.

Specifically, the issue of alternative estimation methods is adressed by the above-mentioned EM algorithm and the issue of alternative identification schemes is addressed by allowing for *correlated* dynamic factors in contrast to the typical application of uncorrelated dynamic factors[1]. Finally, a thorough investigation of the

[1] The issue of interpretation of the estimated factors is addressed in Bork et al. (2008) in which the EM algorithm is also applied.

properties of the FAVARs is undertaken by estimating a large number of econometric specifications of FAVARs and subsequently evaluating these in terms of statistical fit, specification tests, and implications for monetary policy analysis. Consider each of these three contributions in turn.

Similar to the one-step Bayesian method, the EM algorithm estimates all the parameters and the dynamic factors simultaneously in contrast to the two-step principal component method. The last-mentioned method extracts the factors non-parametrically from the data without imposing any dynamic properties on the factors in the first step. The second step estimates the dynamic properties of the factors through a vector autoregression treating the factors as observed[2]. One complication in the principal component method is how to separate the observed federal funds rate from the latent factors in the extraction of these factors, which in contrast is handled in a straightforward manner in the one-step method. However, the advantage of the principal component method is its computational simplicity. Finally, the fully parametric likelihood approach of the EM algorithm allows for classical inference.

The alternative identification scheme allows the factors to be correlated, which is relevant if macroeconomic interpretation is to be attached to these latent factors. For instance, if the first factor is interpreted as an industrial production factor and the second is interpreted as an unemployment factor, then we would expect these factors to be negatively correlated. The correlated factor approach in this paper allows for this feature.

Finally, the robustness of the preferred econometric model is evaluated against several model specifications in terms of the number of factors included in the FAVAR and the number of lags of these factors using various information criteria. Specifically, careful model selection leads to a preferred model characterized by eight factors with a particular parsimonious factor dynamics. This model yields an eleven percentage point better fit of the panel and reaches the same conclusions from the empirical monetary policy analysis as the benchmark model with four factors but a complicated VAR(13) factor dynamics. This finding accords with one of the theoretical results from Bai & Ng (2007) in which it is shown that complicated factor dynamics may be substituted by panel information (in terms of more factors). The

[2] The difference between the estimated factors and the true factors vanishes as the cross-section dimension and the time series dimension approach infinity, cf. Bai & Ng (2002).

eight correlated factors are found to be closely related to observed variables; for instance, the first and most important latent factor is interpreted as an industrial production factor, the second as an unemployment factor, the third as a NAPM[3] factor, and so on.

Factor models have a long tradition in applied economics, finance, and other sciences and hence only a few observations may be needed to motivate why we should continue to be interested in variants of factor models.

Firstly, factor models enable a reduction in the number of explanatory variables (factors) when the variation of a cross-section of variables can be decomposed into a low-dimensional common component reflecting the common sources of variation and a variable specific idiosyncratic component; cf. Ross (1976), Chamberlain (1983), Chamberlain & Rothschild (1983) and Geweke & Zhou (1996) for cross-section applications within finance. Macroeconomic variables tend to comove over the business cycle and therefore their common variation over time may be explained by a few dynamic factor(s); cf. Geweke (1977), Sargent & Sims (1977) and Geweke & Singleton (1981) for the first generation of the dynamic factor (index) models estimated by spectral density maximum likelihood methods. Engle & Watson (1981) propose a time domain maximum likelihood method and Watson & Engle (1983) and Quah & Sargent (1993) apply the Expectation Maximization (EM) algorithm introduced by Dempster et al. (1977).

Secondly, large cross-sections of time series are nowadays available to researchers and policy makers, including central bankers that "follow literally hundreds of data series", as expressed by Bernanke et al. (2005). The potential gains of using large information sets are more precise forecasts and a better understanding of the dynamics of the economy. In the context of the FAVAR, a much richer information set is utilized in the econometric model than in the standard vector autoregressive (VAR) model, leaving less scope for the omitted variable problem. Moreover, because macroeconomic data are prone to measurement errors[4], dynamic factor analysis of large panels may help to filter out the observed counterpart of a theoretical variable, like "inflation", which may not be well represented by a single observed time series.

Recently, a considerable amount of research has been devoted to the econometric

[3] Related to surveys by National Association of Purchasing Management.

[4] Sargent (1989) shows how the existence of measurement error leads to a dynamic factor index model.

theory and empirical analysis of large dimensional approximate[5] dynamic factor models, notably the generalized dynamic factor model by Forni et al. (2000, 2004, 2005) and the static representation of the dynamic factor model by Stock & Watson (2002a,b). Both approaches allow for a general error structure and facilitate dynamic factor analysis of large panels through a few dynamic factors that are extracted from the panel X using non-parametric dynamic and static principal component methods, respectively[6]. A vector autoregression of the factors may be considered as a second step treating the factors as observed if one is interested in structural VAR analysis; see for instance Stock & Watson (2005).

Note at this stage that in the FAVAR of Bernanke et al. (2005), the common variation of the panel dataset is not limited to being explained by a set of latent dynamic factors, as in the Stock & Watson model, but also observed variables (the federal funds rate) may enter into this set and accordingly interact dynamically with the factors.

Econometric theory of the determination of the number of factors has recently been developed, notably by Hallin & Liska (2007), Stock & Watson (2005) and Bai & Ng (2007) for the Forni, Hallin, Lippi & Reichlin class of models and by Bai & Ng (2002) for the class of dynamic factor models in the static representation. Including more factors in the factor model increases the statistical fit of the panel but at the cost of parsimony, whereas choosing too few factors means that the factor space is not sufficiently spanned by the estimated factors. The papers propose various information criteria to guide us in the selection of the number of factors but they do not provide information about the number of lags in the VAR. Consequently, the model selection problem in this paper is solved using the above-mentioned information criteria, and for a given number of factors, also the standard Akaike and Schwartz information criteria.

Since the initial work of Forni, Hallin, Lippi & Reichlin and Stock & Watson, dynamic factor models have been used in an increasing number of applications[7] starting

[5] In the first generation *exact* factor models like Ross (1976) or Geweke (1977), Sargent & Sims (1977) and Geweke & Singleton (1981), the idiosyncratic components are orthogonal. However, the *approximate* factor models allow for some "local" correlation among the idiosyncratic components.

[6] Stock & Watson (2002a) show that the space spanned by the true number of factors, F, can be consistently estimated by the non-parametric principal component method when the cross-section dimension (N) and the time dimension (T) of the panel are large and the number of principal components is at least as large as the true number of factors.

[7] A detailed account of empirical applications can be found in Reichlin (2003) and Breitung & Eickmeier (2006).

with the construction of *coincident indicator indices* as in Forni et al. (2001), *forecasting* where dynamic factors enter the forecasting equation, cf. Stock & Watson (2002*a*,*b*, 2006), and very recently *nowcasting* as in Giannone et al. (2008) where dynamic factor analysis of large panels is used to assess the current-quarter economic conditions. The use of dynamic factors in financial *asset pricing* applications includes the estimation of the conditional risk-return relation in Ludvigson & Ng (2007) and bond market applications by Mönch (2008) and Ludvigson & Ng (2008). Finally, a number of papers to which this paper is particularly related adopt the factor approach for monetary policy analyses with at least two advantages over the traditional VAR.

Firstly, the curse of dimensionality in the VAR is turned into a "blessing" of dimensionality in the factor models as expressed by Stock & Watson (2006) which is particularly useful for representing the data-rich environment in which central banks and professional forecasters actually operate.

Secondly, to assess the current and expected future state of the economy in policy decision making, the central banks are faced with a variety of data in different frequencies, with missing observations and in a preliminary or revised form. Therefore, it can be argued that empirical policy analysis researchers should look at the real-time data that the central bank had at its disposal instead of the revised data and this can be achieved by the dynamic factor model, cf. the approach by Giannone et al. (2008).

Giannone et al. (2004) perform a real-time monetary policy study and find that the US economy is driven by two stochastic shocks (real and nominal) which implies that the federal funds rate should mainly track these two shocks, they argue. Bernanke & Boivin (2003) also consider a real-time dataset in addition to a larger cross-section of revised time series. They find that the scope of the dataset (the number of variables in the cross-section, N) is more important for the forecasting performance of expected inflation and real activity in the forward-looking Taylor rule than the real-time feature. In a similar setup, Favero et al. (2005) study a revised cross-section of US and Euro area data. Common for these studies is the estimation of the factors by principal component methods which are then included in a low-order VAR in the second step to allow for impulse response analysis of monetary policy shocks and these responses are found to be more in line with the predictions from theory. However, a critical step in the empirical monetary policy analysis is a proper disentanglement of the federal funds rate from the estimated factors and the paper by Bernanke et al. (2005) is particularly clear about this identification issue.

As an alternative to the two-step principal component estimation method, one-step Bayesian estimation techniques are applied in Bernanke et al. (2005) as well as in Banbura et al. (2008). The former choose thirteen lags in their FAVAR specification while the latter also estimate this variant in addition to lag specifications determined by the BIC criterion. The fully parametric one-step EM algorithm method has recently been applied to large panels in Jungbacker & Koopman (2008) that estimate a dynamic factor model with a VAR(1) in the orthogonal factors and in Reis & Watson (2008) that estimate pure inflation with a VAR(4) in absolute-price and relative-price components.

Based on this selective literature overview there seems to be a need for exploring the consequences of model selection for not only policy evaluation but also in terms of statistical significance of parameters and statistical fit of the various components in the economy such as inflation, employment, production etc. This issue is taken up in this paper and consequently several model specifications ranging from a few correlated factors with only one lag to many correlated factors with rich factor dynamics are estimated in an EM algorithm setup. I show how identifying restrictions can easily be imposed on the parameters including restrictions on the VAR parameters, if needed. This is in contrast to the Bayesian approach where these kinds of restrictions seemingly lead to excessive computational cost, cf. Bernanke et al. (2005).

Furthermore, though the EM algorithm finds the vicinity of the maximum quickly, the convergence to the maximum is almost excruciatingly slow (linear convergence rate) and consequently hybrid methods combining the EM algorithm and the BFGS have been proposed in the literature. Therefore, I also apply the hybrid EM-BFGS as described by Jungbacker & Koopman (2008) in order to speed up the convergence.

The rest of the paper is organized as follows. The factor-augmented VAR is presented in section 1.2 while identification issues and the estimation method are presented in section 1.3. Section 1.4 details the empirical results and section 1.5 concludes. The appendices contain details on the Kalman filter and smoother as well as the EM algorithm.

1.2 Model framework: The factor-augmented VAR

Two ingredients need to be combined to set up the FAVAR. The first ingredient is the dynamic factor model and the second ingredient is the standard VAR with observed variables. Before mixing the ingredients, one thing is important to note: the federal funds rate (FFR) is both part of the observed variables in the panel (the measured part of the state space system) and also part of the state variables (the state transition equation in the state space system) which include the dynamic latent factors. Therefore, to allow for this feature the standard dynamic factor model is modified and this is described in detail below.

This section will center around the static representation of the dynamic factor model in state space form which can be seen as a special case of the large dimensional generalized dynamic factor model; see Bai & Ng (2007) for a clear exposition. Following the presentation of the dynamic factor model, the FFR is properly identified in the panel and then added to the state transition variables. This may sound like a backward description of the factor-augmented VAR but nevertheless I find this the most intuitive route towards the FAVAR.

The key implication of the dynamic factor model is that the variation of each of the N observed variables in the panel X can be decomposed into two orthogonal components, that is a component χ common to all variables and an idiosyncratic component ξ specific to each variable. The common component is driven by a few common factors and this component accounts for the covariation of the observed variables at all lags and leads. Consequently, the ith variable in the panel X[8] at time t can be written as:

$$x_{it} = \chi_{it} + \xi_{it} \tag{1.1}$$

for $i = 1, .., N$ and $t = 1, .., T$ with $E\left[\chi_{it}\xi_{js}\right] = 0 \; \forall \; i, j, t, s$ but with a potentially limited amount of correlation among the idiosyncratic components in the new generation of dynamic factor models. The following description encompasses the dynamic factor model, which is characterized by the dynamic loading on the common factors as well as the static representation of the dynamic factor model characterized by the static loadings. The distinguishing features of the models will become useful in later discussions.

[8] All variables in the panel are transformed into stationary variables with mean zero and unit variance. See section 1.4.1.

Consider as in Forni et al. (2005), the specification of the $N \times 1$ vector of the common component at time t to be dynamically explained by the q common factors f_t such that $\chi_t = \lambda^\top(L) f_t$, where $\lambda(L)$ is a $q \times N$ matrix polynomial in the lag-operator L of finite order s[9]. To facilitate an interpretation of the panel being driven entirely by q primitive iid shocks, the common component is sometimes written as $\chi_t = \beta(L)\varepsilon_t$, where $\beta(L)$ represents the impulse-response functions and accordingly for each variable records the responses in terms of sign, magnitude and lag-structure following a shock to the underlying primitive shocks, ε_t[10]. Inserting the specification of the common component in (1.1) results in a dynamic factor model driven by q dynamic factors:

$$x_{it} = \lambda_i^\top(L) f_t + \xi_{it} \tag{1.2}$$

where $\lambda_i(L) = \lambda_{i,0} + \lambda_{i,1}L + \cdots + \lambda_{i,s}L^s$. Stacking contemporaneous and s lagged values of f_t in the $q(s+1)$ dimensional vector F_t and the matching values of λ_i in $q(s+1)$ dimensional vector Λ_i results in the static representation of the dynamic factor model in (1.2), which is driven by $r = q(s+1)$ factors, F_t :

$$\begin{aligned} x_{it} &= \Lambda_i^\top F_t + \xi_{it} \qquad (1.3) \\ &= \begin{bmatrix} \lambda_{i,0} \\ \lambda_{i,1} \\ \vdots \\ \lambda_{i,s} \end{bmatrix}^\top \begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-s} \end{bmatrix} + \xi_{it} \end{aligned}$$

Notice how the dimension of $F_t, r = q(s+1)$ depends on the heterogeneity in the response of the data to the factors f_t through $\lambda(L)$ or equivalently to the primitive shocks ε_t through $\beta(L)$.

Furthermore, F_t is governed by a dynamic process which depends on how complicated the process governing f_t is relative to the response heterogeneity of the panel. Assuming that f_t is an AR(h) process, Bai & Ng (2007)[11] show that F_t can be represented as a VAR(p) process with $p = \max(1, h-s)$. Intuitively, if the dynamic process of f_t is particular simple then a VAR(1) should be sufficient. Interestingly, a sufficiently heterogeneous dynamic response of the data may substitute for some otherwise complicated dynamics of f_t, cf. the term $(h-s)$ in $\max(1, h-s)$. I will

[9] Infinite order of the lag-polynomiums is considered in the *generalized* dynamic factor model of Forni et al. (2000).

[10] Rewrite the factors in terms of the primitive shocks, $f_t = a(L)\varepsilon_t$ and as a result $\beta(L) = \lambda(L)a(L)$. See Forni et al. (2007) for a thorough discussion.

[11] They also discuss MA(h) and ARMA processes.

refer to this result later in the discussion of the empirical results.

The *static representation* of the dynamic factor model is now closed and can be written in state space form:

$$\begin{aligned} X_t &= \Lambda F_t + \xi_t \\ F_t &= \Phi(L) F_{t-1} + \Upsilon \varepsilon_t \end{aligned} \tag{1.4}$$

where $X_t = (x_{1,t}, ..., x_{N,t})^\top$, $\xi_t = \left(\xi_{1t,}, ..., \xi_{N,t}\right)^\top$ is i.i.d $N(0, R)$[12] and $\Lambda = (\Lambda_1, ..., \Lambda_N)^\top$ is a $N \times r$ loading matrix. The state transition equation is stationary so that the eigenvalues of the pth order matrix polynomial $\Phi(L)$ are less than 1 in modulus, Υ is a $r \times q$ matrix and ε_t is i.i.d $N(0, Q)$. The unknowns in this Gaussian state space model are the parameters in $\Theta = \{\Lambda, R, \Phi(L), \Upsilon, Q\}$ and the latent dynamic factors F_t.

The final step towards the FAVAR is the inclusion of the FFR in both X_t *and* F_t (FFR is added to and ordered last in F_t). Specifically, the FFR in X_t loads with unity on the last factor in F_t and zeros on the remaining latent factors, such that the corresponding row in Λ for FFR is $[0, ..., 0, 1]$. In principle, an idiosyncratic error could be attached to the FFR to capture the transition between discretionary changes in the policy rate. In line with Bernanke et al. (2005), I argue that the FFR is indeed measured without error whereas the other variables may be measured with error. Applying these minor changes to the state space form in (1.4) leads to the preferred FAVAR specification. However, some identifying restrictions need to be imposed on the econometric formulation to achieve distinct factors, which, together with the estimation procedure is the topic of the following section.

1.3 Identification and estimation by the EM algorithm

This section starts with a discussion of identification schemes and then proceeds to a brief description of the estimation procedure, that is the EM algorithm. I

[12] Note that the assumption of i.i.d idiosyncratic components in (1.4) defines an *exact* dynamic factor model. This is certainly a strong assumption, particularly in the case of large panel data where local cross-sectional correlation within a group of similar variables should be expected. As such, equation (1.4) represents a misspecified model. However, Doz et al. (2006) generate data under the assumption of an *approximate* factor model and show, for large N and T, that the exact factor model consistently estimates the factors by a Gaussian (quasi)maximum likelihood method. Specifically, they propose to use the EM algorithm.

also demonstrate how linear parameter restrictions can easily be imposed. Finally, a hybrid estimation method that combines the EM algorithm and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method with analytical derivatives is described.

The state space model in (1.4) is not econometrically identified as it is possible to form observationally equivalent models by arbitrary rotations of the latent factors, F_t, and the loadings Λ. For any non-singular matrix H we can form a model that is observationally equivalent to (1.4) by a rotation of the factors $\tilde{F}_t = HF_t$ and loadings $\tilde{\Lambda} = \Lambda H^{-1}$:

$$\begin{aligned} X_t &= \tilde{\Lambda}\tilde{F}_t + \xi_t \\ \tilde{F}_t &= \tilde{\Phi}\tilde{F}_{t-1} + \tilde{\varepsilon}_t \end{aligned}$$

where the dynamics of the factors are simplified to a VAR(1) with $\Upsilon = I$. Moreover, $\tilde{\varepsilon}_t = H\varepsilon_t$. Consequently, it is not possible to estimate a unique set of parameters $\hat{\Theta}$ with the data unless identifying restrictions are imposed on Θ. This is a well-known problem of classical factor analysis and the principal component approach to dynamic factor analysis by Stock & Watson. Typically, these models are identified by restricting the factors to be orthogonal or alternatively the loadings to be orthogonal. However, neither of these identification schemes are sufficient in the one-step estimation of the state space model because the factors are identified by both the measurement equation and state transition equation in (1.4). Therefore, more restrictions are needed to obtain an econometrically identified model and this issue is addressed by the following three requirements:

1. Ensure invariance of the model under invertible linear transformation of the factors.

2. Ensure that the number of moments in the data, $\frac{1}{2}N(N+1)$ exceeds the number of free parameters in Θ.

3. The loading matrix Λ must have full rank in order to avoid identification problems; see Geweke & Singleton (1981) and Aguilar & West (2000). Under the assumption that the number of factors is known and equal to r the rank of Λ must be r.

The second requirement is easily satisfied because the cross-sectional dimension N is very large compared to the number of factors. The third requirement is theoretically satisfied[13] from the parameter restrictions I impose to satisfy the first requirement. Consequently, I focus on the first requirement in the following.

Generally, the first identification requirement is about separating the contributions of the different latent factors to the variation in the panel X. The predominant starting point is uncorrelated factors which implies that the identification of the sources of variation in X is then a matter of imposing an identifying structure on the loading matrix; in particular a structure that embodies the separation of the contribution of the factors to the variation in X. Alternatively, the assumption about uncorrelated factors can be relaxed by allowing for *correlated* factors. However, less restricted factor dynamics would have to be paid by a more restrictive structure on the loading matrix in order to be able to separate the sources of variation. In other words, the specific identification scheme applied is a matter of choice but it does not change the underlying idea that the factors are the sources of common variation in X: either the variables in X covary because they load differently on a set of common uncorrelated factors or because they load on different factors which are themselves correlated.

In the case of uncorrelated factors, the "hierarchical" structure of the loading matrix in Geweke & Zhou (1996), Aguilar & West (2000) and chapter 8 of Harvey (1989) is a popular identification scheme that uniquely identifies the loadings and the factors by imposing a lower triangular structure on the loading matrix Λ. More specifically, Q is assumed to be an identity matrix and the upper $r \times r$ block of Λ is lower triangular with r positive diagonal elements. The term "hierarchical" stems from the lower triangular form, where the first element in X only loads on the first factor, the second variable on the first and second factor, etc[14]. Aguilar & West (2000) further restrict the diagonal of the $r \times r$ block of Λ to unity and then allow for a diagonal covariance matrix for the factors. It should be noted that this "hierarchical" approach is in fact similar to the identification scheme stated in Proposition 1 in Geweke & Singleton (1981) in their frequency domain analysis of a first generation dynamic factor model.

In the case of *correlated* factors, a more restricted and "simple" structure of the loading matrix needs to be imposed to ensure identification. Specifically, the upper

[13] The rank condition is never violated in the empirical application detailed in section 1.4.

[14] The ordering may potentially influence the statistical fit. However, I do not find such sensitivity of the empirical results in this paper to the ordering. This issue is further discussed in the empirical section. Aguilar & West (2000) present a similar discussion.

$r \times r$ block of Λ is an identity matrix. This means that *up to some measurement error* the first variable in X is assumed to be a direct measure of the first factor, the second variable a direct measure of the second variable, etc. Notice, that the unit restrictions do not guarantee that the specific factor turns out to explain the restricted variable well. On the contrary, the other of variables in X may be far better explained by a factor that is fairly different from the restricted variable and consequently the restricted variable will have a large measurement error. Interestingly, Proposition 2 in Geweke & Singleton (1981) can be used when the factors are correlated and corresponds to the identity matrix restriction on the loadings.

In this paper, the identification scheme with *correlated* factors is preferred. The reason for this preference is that if economic interpretation is to be attached to the estimated factors, for instance a "real activity factor" or an "employment factor", then it makes more sense to have correlated factors because theoretically but also empirically such economic quantities should be correlated and not orthogonal. Yet another argument for correlated factors is found in the typical view of the monetary transmission mechanism, which is investigated empirically in section 1.4. According to this view, a contractionary monetary policy shock is expected to decrease production and employment with some time lags and then even later also inflation. More precisely, the inclusion of *more* correlated factors in a *low order* VAR in the state transition equation combined with different loadings on these factors in the measurement equation is able to produce an empirically plausible monetary transmission mechanism.

The identifying restrictions in this paper can be summarized as follows. Consider the panel X with the rows reordered such that the restricted variables are found in the top r rows of X. In this case, the loading restrictions simply amount to imposing an identity matrix in the top $r \times r$ block of Λ. In the current application, the rows of X are not reordered so the loading restrictions are imposed as follows:

1. The FFR in X_t with row index ℓ_r in Λ loads only on the last dynamic factor in F_t which is a monetary policy factor (the FFR itself). Hence, for the r columnwise elements in row ℓ_r in Λ, the restricted loading is:

$$\Lambda^*_{\ell_r} = [0, ..., 0, 1] .$$

2. The remaining $(r - 1)$ latent dynamic factors ordered before the monetary policy factor in the VAR each load with unit restriction on a single "slow-moving"

variable (see below), which is assumed to respond with a lag to changes in the FFR. Let the selected slow-moving variables with restricted loadings be indexed by row $\{\ell_1, .., \ell_{r-1}\}$ of X_t, which means that the restricted rows of Λ can be written as:

$$\begin{aligned}
\Lambda^*_{\ell_1} &= \left[\begin{array}{ccc} \underset{1\times 1}{1} & \underset{1\times(r-2)}{0} & 0 \end{array} \right] \\
&\vdots \\
\Lambda^*_{\ell_j} &= \left[\begin{array}{ccc} \underset{1\times(j-1)}{0} & \underset{1\times 1}{1} & \underset{1\times(r-j)}{0} \end{array} \right] \\
&\vdots \\
\Lambda^*_{\ell_{r-1}} &= \left[\begin{array}{ccc} \underset{1\times(r-2)}{0} & \underset{1\times 1}{1} & 0 \end{array} \right]
\end{aligned} \tag{1.5}$$

whereas the remaining elements of Λ are left free.

This identification scheme allows for correlated factors and the zero restrictions on Λ ensure that the factors explain distinct parts of the variation in the panel. A separate identification issue, which is relevant for the identification of the monetary policy shocks in the VAR by a recursive identification scheme requires the factors to be associated with slow-moving variables such that $\ell_j \in \{\ell_1, .., \ell_{r-1}\}$ should be chosen from this group of variables. Therefore, Bernanke et al. (2005) propose to categorize the variables into "slow-moving" variables such as production and unemployment variables and "fast-moving" variables like financial market variables[15]; see section 1.4.1 for more details.

1.3.1 The EM algorithm

The linear Gaussian state space model in (1.4) with its latent factors F_t is well represented in a Kalman filter setting. However, the Kalman filter needs the parameters $\Theta = \{\Lambda, R, \Phi(L), \Upsilon, Q\}$ as input and therefore does not estimate these. Building on the seminal work by Dempster et al. (1977), Shumway & Stoffer (1982) introduce the Expectation Maximization (EM) algorithm to estimate the parameters in state space models as the model above. Essentially, the EM algorithm is an iterative maximum likelihood procedure applicable to models with "missing data", which in this context are the unobserved factors.

[15]Notice, that if the factors also are allowed to be fast-moving then a simultaneity problem arise in the identification of the monetary policy factor in the sense that both the monetary policy factor and the fast-moving factor(s) should be allowed to respond contemporaneously to either of these shocks. Bjørnland & Leitemo (2009) solve this by long-run restrictions.

The complete data likelihood of the Gaussian state space model in equation (1.4) is given in equation (1.19) in Appendix B.3. However, the complete data likelihood cannot be calculated due to the unobserved F_t, but it is possible to calculate the expectation of the complete data likelihood conditional on the observed data and input of parameter estimates (denoted $\Theta^{(j)}$); see Appendix B.3. Essentially, this expectation depends on smoothed moments of the unobserved variables from the Kalman smoother and hence on the data and $\Theta^{(j)}$. The Maximization step results in the following closed form estimators at iteration j

$$\begin{aligned}
\text{vec}\left(\Lambda^{(j)}\right) &= \text{vec}\left(DC^{-1}\right) &(1.6)\\
R^{(j)} &= \frac{1}{T}\left(E - DC^{-1}D^{\top}\right) &(1.7)\\
\text{vec}\left(\Phi^{(j)}\right) &= \text{vec}\left(BA^{-1}\right) &(1.8)\\
Q^{(j)} &= \frac{1}{T}\left[C - BA^{-1}B^{\top}\right] &(1.9)
\end{aligned}$$

where the following moments are available from the Kalman smoother (indicated by subscript $t|T$):

$$\begin{array}{ll}
A = \sum_{t=1}^{T}\left(\hat{F}_{t-1|T}\hat{F}_{t-1|T}^{\top} + \hat{P}_{t-1|T}\right) & B = \sum_{t=1}^{T}\left(\hat{F}_{t|T}\hat{F}_{t-1|T}^{\top} + \hat{P}_{\{t,t-1\}|T}\right)\\
C = \sum_{t=1}^{T}\left(\hat{F}_{t|T}\hat{F}_{t|T}^{\top} + \hat{P}_{t|T}\right) & D = \sum_{t=1}^{T} X_t\hat{F}_{t|T}^{\top}\\
E = \sum_{t=1}^{T} X_t X_t^{\top} &
\end{array}$$

and where F_t is approximated by $\hat{F}_{t|T} = E\left[F_t|\,\mathcal{X}_T\right]$. $\mathcal{X}_T = \{X_1, .., X_T\}$ denotes the information set, $\hat{P}_{t|T} = var\left(F_t|\,\mathcal{X}_T\right)$ is the variance and $\hat{P}_{\{t,t-1\}|T} = cov\left(F_t, F_{t-1}|\,\mathcal{X}_T\right)$ is the lag-one covariance.

These estimates can then be used in the Expectation step to compute a new set of moments from the Kalman smoother. Subsequently, the estimates are supplied to the maximization step above and the procedure continues until convergence of the likelihood.

In practical implementation, a VAR(1) usually does not pose any problem and neither should a VAR(p) because any lags of F_t can be included in an augmented state vector if the autoregressive parameters in $\Phi(L)$ are represented in a first order form (companion matrix) as in Hamilton (1994) chapter 10. The autocovariances in the B matrix needed in the Φ estimate should then follow automatically from the first order form; cf. Watson & Engle (1983). However, this paper follows a slightly different route similar to Koopman et al. (1999) but with an implementation in MATLAB[16], where the smoothed autocovariance matrix of the state variables is constructed directly and explicitly through recursions, cf. de Jong & Mackinnon (1988), de Jong (1989) and Koopman & Shephard (1992). For instance, the lag-one covariance smoother needed for the $\hat{\Phi}_1$ estimate in a VAR(1) is defined in the latter-mentioned paper as:

$$\hat{P}_{\{t,t-1\}|T} = \left[I - \hat{P}^{xx}_{t|t-1} N_{t-1} \right] L_{t-1} \hat{P}^{xx}_{t-1|t-2}$$

and the lag-two covariance smoother needed for the $\hat{\Phi}_2$ estimate in a VAR(2) is:

$$\hat{P}_{\{t,t-2\}|T} = \left[I - \hat{P}_{t|t-1} N_{t-1} \right] L_{t-1} L_{t-2} \hat{P}_{t-2|t-3}$$

where N_{t-1} and L_{t-1} in Appendix B.2 are matrices defined recursively in the Kalman smoother and Kalman filter, respectively. Furthermore, the state smoothing recursions are also stated in the appendix.

Parameter restrictions in the EM algorithm

In order to implement the identifying restrictions in (1.5), the estimators in (1.6) – (1.9) subject to linear restrictions need to be derived. Shumway & Stoffer (1982) and Wu et al. (1996) present the restricted Φ^{*}[17] and Bork et al. (2008) show how the restricted Λ^* estimator subject to a linear restriction in the form $H_\Lambda \operatorname{vec} \Lambda = \kappa_\Lambda$ can be derived:

$$\begin{aligned} \operatorname{vec}(\Lambda^*) &= \operatorname{vec}\left(DC^{-1}\right) \\ &\quad + \left(C^{-1} \otimes R\right) H_\Lambda^\top \left[H_\Lambda \left(C^{-1} \otimes R\right) H_\Lambda^\top\right]^{-1} \left\{\kappa_\Lambda - H_\Lambda \operatorname{vec}\left(DC^{-1}\right)\right\} \end{aligned} \tag{1.10}$$

where κ_Λ is a $\eta \times 1$ vector and the restriction matrix H_Λ is of dimension $\eta \times Nr$. Notice that the unrestricted estimator in (1.6) appears if $\eta = 0$ restrictions are imposed.

[16] A small dynamic factor model with $N = 12$ observed variables, $r = 2$ factors and $p = 4$ lags, was simulated and subsequently estimated with noisy initial estimates of the parameters to check the code.

[17] shown in the appendix.

1.3.2 The hybrid EM-BFGS optimization method

The EM algorithm is known to converge rather slowly due to its linear convergence rate. However, the EM algorithm robustly finds the vicinity of the maximum quickly and therefore it has been proposed by for instance Lange (1995) to combine the good properties of the EM algorithm in the early stage of the optimization process with the fast convergence properties of quasi-Newton methods in the late stage of the optimization process. This hybrid requires analytical derivatives and in an application by Jungbacker & Koopman (2008), these are derived. Moreover, whereas I often experience computing time in hours for the heavily parameterized models presented here, they report computing time in minutes. The analytical derivatives from Jungbacker & Koopman (2008) in terms of Kalman smoothed quantities are given Appendix C.

The performance of this hybrid method is here somewhat mixed. Often it is found that the EM algorithm has to get very near the optimum before it is reliably to shift to the BFGS method; otherwise the BFGS method fails to find an optimal solution. However, when the hybrid is succesful, it is indeed relatively fast and therefore continued research into this hybrid is worthwhile.

1.4 Empirical results

In this section, I present empirical evidence that a factor model with more factors but fewer lags performs equally well, if not better, in terms of statistical fit (increased R^2). Moreover, the empirical monetary policy analysis results in equally plausible impulse responses. For instance, the price puzzle is almost eliminated and comparable to Bernanke et al. (2005). Moreover, unemployment responds more negatively to contractionary monetary policy shocks but still reverts to the baseline within four years (similar to Bernanke et al. (2005)). Finally, I also show that the empirical evidence accords with the theoretical insight from section 1.2: that complicated factor dynamics (many lags) may be substituted by cross-sectional information (more factors).

Throughout this section, I compare the results that I obtain from various model specifications with the principal component FAVAR and the Bayesian FAVAR by Bernanke et al. (2005).[18] The differences in the empirical results may then be

[18]I use exactly the same dataset as these authors.

attributed to the differences in the estimation methods, i.e. the EM algorithm versus the methods of the Bernanke et al. (2005)[19] as well as the factor configuration in terms of the number of factors, r, and the number of lags, p. Accordingly, an EM algorithm equivalent to the preferred model by Bernanke et al. (2005) with four factors including the monetary policy factor and thirteen lags is calculated (abbreviated BBE-EM) and makes up a first step in the comparison. The second step in the comparison is then made with reference to the preferred model in this paper with eight factors and three lags, a model choice that is explained below. I find that the results from the BBE-EM model are comparable to the results by Bernanke et al. (2005) in the sense that a similar overall R^2 for the panel seems to be achieved as well as similar and equally plausible impulse responses. Furthermore, the preferred eight factor model with three lags improves the results significantly in the sense that a ten percentage point increase in the overall R^2 for the panel is achieved without compromising the plausibility of the impulse responses.

It should be emphasized that the empirical analysis in this paper focuses on the identification of monetary policy shocks and the economy-wide responses to these shocks while remaining agnostic about other structural shocks. Furthermore, I impose that the number of static factors equals the number of dynamic factors, i.e. $r = q$ and that $\Upsilon = I$, which generates a structural shock to each of the factors. Hence, the focus is on the determination of the number of static factors *including* the monetary policy factor, which amounts to $r = 8$ factors in this paper, rather than on the determination of the q dynamic factors driven by $q \leq r$ structural shocks.[20]

The preferred model with eight factors and three lags is the outcome of a careful model selection process where a large number[21] of estimated FAVAR models were evaluated in terms of information criteria, test statistics, and model parsimony considerations to be detailed below. The motivation for evaluating a large number of

[19] Although seemingly unreported by the authors, it seems that they employ uncorrelated factors in contrast to the correlated factors employed in this paper.

[20] For example, $f_{1,t}$ and $f_{1,t-1}$ count as $r = 2$ static factors in the static representation of the factor model whereas in the dynamic factor model, they represent the contemporanenous and lagged values of $q = 1$ dynamic factor driven by one structural shock. Accordingly, r is the rank of the covariance matrix of the common component χ whereas q is the rank of the spectral density matrix of χ. For further discussion of structural factor models, refer to Forni et al. (2007) and Stock & Watson (2005).

[21] I programmed the estimation procedure as a MATLAB function that takes the dataset, r and p as arguments and then looped over this function from $r = 3, .., 10$ and $p = 1, .., 13$. To make this excercise computationally feasible, a maximum of 10,000 iterations in the EM algorithm were allowed, which explains the few missing factor models.

models is twofold: 1) What is the sensitivity of the empirical policy analysis to the number of lags included in the VAR? The monthly frequency of the data asks for several lags, but is the thirteen lags chosen by Bernanke et al. (2005) necessary across different number of factors? Fortunately not. Nearly identical impulse responses emerge from a factor model with eight factors and three lags and from a factor model with four factors and thirteen lags[22]. I ascribe this observation to the theoretical result mentioned previously, that complicated VAR dynamics in terms of many lags can be substituted by cross-section information in terms of more factors. 2) Obviously, more factors imply a better statistical fit of the panel, but what is the optimal number of factors for this panel and which part of the panel gains from including more factors? Price indices for instance are far better explained when more than five factors are added, at least in this paper. That more factors need to be included for a proper explanation of the price indices seems to be a special feature of the correlated factor approach in this paper in contrast to the orthogonal factor approach. The reason is that although the fit is not inferior, it involves more correlated factors before the model picks up to the price dimension in the dataset.

The rest of this section now presents detailed results behind some of the conclusions stated above. Firstly, the data and the transformation of the data are described followed by an account of how the identifying restrictions are imposed. Secondly, a number of panel information criteria from Bai & Ng (2002) are calculated as well as the usual AIC/SIC information criteria and a multivariate Portmanteau test tailored to latent variables in a VAR. Moreover, the autocorrelation function for the VAR residuals and an average R-square for each factor model are plotted. All these measures guide me in the model selection choice. Thirdly, impulse responses and forecast error variance decompositions are calculated.

1.4.1 Data description and data transformation

The dataset used in this paper is exactly the same as the dataset that Bernanke et al. (2005)[23] analyze. The data consist of $N = 120$ monthly time series covering a large part of the US economy over the period 1959:1 to 2001:8; see Appendix A.1 page 162 for a description of the dataset and in particular the classification into

[22] Notice that both models involve approximately the same number of autoregressive coefficients in the VAR.

[23] I thank Jean Boivin for kindly making the data set available on his website, HEC Montréal, Canada

slow-moving variables and fast-moving variables. The time series in the panel are transformed into stationarity by taking logs and/or differencing[24]. The next step involves standardizing the transformed data so that all series have mean zero and unit variance, which is typical especially for principal component analysis. Denote by X_t the transformed and standardized data at time t consistent with equation (1.4) page 11. However, when studying impulse responses, the interest centers around the observed variables in levels (e.g. the price level) rather than the transformed variables (e.g. inflation) and therefore a reverse transformation of the responses is required, denoted by $D(L)$ such that the reverse-transformed data $\tilde{X}_t = D(L) X_t$[25].

1.4.2 The imposition of the identifying restrictions

A number of identifying restrictions need to be imposed on r rows of the loading matrix Λ as explained in equation (1.5). However, the specific set of r rows in Λ which are jointly restricted to an identity matrix needs to be determined. In other words, this amounts to choosing a set of r variables assumed to be a direct measure of the r factors.

I propose a two-step procedure to determine the specific set of r variables in the panel X which should be a direct measure of the r factors. The *first step* involves principal component analysis (PCA) where r principal components are calculated from the panel X. This choice is based on the insight that principal components consistently estimate the space spanned by the (independent) factors; cf. Bai & Ng (2002) and Forni et al. (2000, 2005). Subsequently, each of the N variables is regressed on the r principal components resulting in a $N \times r$ matrix of individual R^2. The dominant R^2's for each factors is then used to infer the characteristics of each factor. Typically, this approach reveals that the first factor can be interpreted as an industrial production factor. In the *second step*, I impose the exactly identifying restrictions on the inferred dominant factors and filter the factors with very weak priors on the initial parameter estimates. In particular, the loading matrix was filled with with zeros except for the exactly identifying unit restrictions and a complete estimation by the EM algorithm is undertaken. Finally, another evaluation

[24]The data are already transformed by Bernanke et al. (2005) to reach stationarity; see Bernanke et al. (2005) for details on the data set and on the transformation which results in a sample size of $T = 511$. The data transformation decisions are similar to Stock & Watson (2002*b*) and based on judgemental and preliminary data analysis of each series, including unit root tests.

[25]For instance, if the data in X_t are in growth rates, the diagonal elements of $D(L)$ would need to be multiplied by $\frac{1}{1-L}$ in order to have the data in levels in $\tilde{X}_t$.

of the dominant factors from the EM algorithm is undertaken as the factors are now correlated.[26]

Consequently, this pre-study reveals that the first factor is robustly associated with industrial production. The second factor is related to unemployment, the third factor is associated with NAPM indices (production or employment), the fourth factor with production hours, and the fifth factor with price indices. Based on these findings, the restrictions are imposed on the following list of variables in increasing order of the number of factors included:

$$\{\ell_1, \ell_2, .., \ell_9\} = \{11, 27, 18, 47, 112, 23, 17, 50, 16\}$$

where numbers refer to the variable number listed in Appendix A.1 page 162. Notice that the restrictions are not imposed on variables that are deemed a priori to be particularly important variables such as the unemployment rate for all workers (#26), the consumer price index all items (#108) etc. Instead, a variable that is closely related or correlated with this variable is selected such that the potentially most important variables are maximally explained and minimally restricted.

Admittedly, an alternative restriction index, $\ell_1, .., \ell_{r-1}$ may improve the overall fit although the improvement is deemed modest because of the performed two-step procedure. Finally, it should be noted that the particular characteristics of an estimated factor are not determined by the single unit restriction in a particular column in Λ but rather by how important this factor is for the fraction of variance explained. Table 1.1 supports the argument that the imposition of unit restrictions on an arbitrary set of variables does not change the underlying characteristics of the factors and the statistical fit. Hence, the statistical fit of the preferred model is robust to an alternative set of restrictions.

1.4.3 Model selection: information criteria and test statistics

An important choice in factor analysis concerns the unknown number of factors r that span the factor space. A number of papers mentioned in the introduction address this challenge and in this paper different panel information criteria developed

[26]Finally, to use a somewhat more informed starting values I use PCA of r subsets of the dataset where the principal component of each subset represents an initial estimate of one of the r factors.

by Bai & Ng (2002) are applied. Essentially, the proposed information criteria reflect the usual trade-off between model parsimony and statistical fit using a penalty function. However, this penalty function depends on both T and N so that the usual AIC/SIC cannot readily be applied and furthermore the information criteria should also take account of the fact that the factors are unobserved. However, the criteria by Bai & Ng (2002) do not address the number of lags in the VAR and therefore the AIC/SIC will have a comeback when the VAR order needs to be determined.

Principal component analysis with r factors extracted from dataset in X allows for the calculation of the sum of squared residuals $V(r) = (NT)^{-1} \sum_{t=1}^{T} \hat{\xi}_t \hat{\xi}_t^{\top}$, where $\hat{\xi}_t$ is a $N \times 1$ vector of the estimated idiosyncratic errors. Based on this quantity Bai & Ng (2002) suggest a number of information criteria of which some of the most popular are shown below:

$$\begin{aligned}
\min_r IC_{p2}(r) &= \ln(V(r)) + r\left(\frac{N+T}{NT}\right)\ln C_{NT}^2 \\
\min_r IC_{p3}(r) &= \ln(V(r)) + r\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right)
\end{aligned}$$

where the sequence of constants $C_{NT}^2 = \min\{N, T\}$ represents the convergence rate for the principal component estimator. Furthermore, the following panel information criteria are also calculated:

$$\begin{aligned}
\min_r PC_{p2}(r) &= V(r) + r\hat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln C_{NT}^2 \\
\min_r PC_{p3}(r) &= V(r) + r\hat{\sigma}^2\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right)
\end{aligned}$$

where $\hat{\sigma}^2 = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} E[\xi_t]^2$ is a penalty function scaling term and usually calculated using some maximum number of factors $r_{\max}$.

Application of the IC_{p2} and IC_{p3} however points towards a large number of factors ($r = 16$), which is similar to what Bernanke & Boivin (2003) and Forni et al. (2007) experience with this criterion. Nevertheless, instead of relying on the estimation of the sum of squared residuals from principal component analysis, I calculate $V(r), \hat{\sigma}^2$ from the actually estimated models using the EM algorithm and then calculate the above information criteria[27]. These calculations point strongly towards $r = 8$ which can be seen in figure 1.1. IC_{p2}, PC_{p2} and PC_{p3} lead to exactly

[27] I used C_{NT}^2 to represent the imperfect convergence rate for the EM algorithm estimator.

the same result and are therefore not shown.

[Insert Figure 1.1]

An alternative and less formal method consists of calculating the average explained variation of the variables in the panel relative to the total variation, the average R^2 measure, which is primarily influenced by the number of factors and less by the number of lags in the VAR. Based on the average R^2 measure adjusted for degrees of freedom, denoted $\bar{R}^2$, this alternative measure could be used to evaluate the incremental value of adding more factors. Figure 1.2 shows $\bar{R}^2$ for each estimated model and it can be seen that the incremental value of $\bar{R}^2$ diminishes as more and more factors are included in the FAVAR. A decision on when to stop adding factors is subjective, but based on these results, I maintain that $r = 8$ seems to be a good choice.

[Insert Figure 1.2]

The $\bar{R}^2$ weights each variable equally in the panel so that for instance industrial production, e.g. mining (#14), receives the same weight as the total industrial production index (#16) even though the former is probably of less interest. In other words, improved fit for some variables does not show up clearly in $\bar{R}^2$. The purpose of Figure 1.3 is to show that the fit of some variables such as unemployment and inflation, improves dramatically when more factors are added whereas others such as industrial production, e.g. mining and foreign exchange rates, are never well explained. More details about the preferred model are provided later.

[Insert Figure 1.3]

Towards a well-specified VAR

Ultimately, the preferred model is to be used for impulse response analysis of shocks to the monetary policy factor and therefore a well-specified VAR is sought for. In the previous paragraphs, I argue for eight factors but the number of lags in the VAR also needs to be determined. For this purpose, the Akaike (AIC), Schwarz (SIC) and Hannan & Quinn (HQIC) information criteria are calculated in Tables 1.2, 1.3 and 1.4 respectively. The maximum number of lags to be included does not exceed six, which is somewhat surprising. An alternative procedure would be to test if the pth autoregressive coefficient matrix is significant in terms of a likelihood ratio test. Apparently, for the preferred model with eight factors, the number of lags should be either three or six.

[Insert Tables 1.2, 1.3 and 1.4]

Given the different $\{r, p\}$ factor model specifications, the VAR residuals are also inspected to see if they are approximately white noise by tailoring the multivariate Portmanteau test to latent variables and by inspecting the VAR residuals visually. Consider the multivariate Portmanteau test which tests whether the hth order residual autocorrelation is zero. However, recall that we approximate the true factors F_t by the smoothed factors $\hat{F}_{t|T}$, i.e. $F_t = \hat{F}_{t|T} + \left(F_t - \hat{F}_{t|T}\right)$, which means that it is the residuals of the true factors that interest centers around. Accordingly, I modify the standard Portmanteau test to use smoothed quantities instead. The standard multivariate Portmanteau test statistic (see Lütkepohl (2007)) is:

$$Q(h) = T \sum_{i=1}^{h} \operatorname{tr}\left(\hat{C}_i^\top \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}\right) \sim \chi^2_{r^2(h-p)}, \quad i = 1, ..., h$$

where the (auto)covariances of the VAR residuals are:

$$\hat{C}_i = \frac{1}{T} \sum_{t=i+1}^{T} \left(\hat{\varepsilon}_t - E\left[\hat{\varepsilon}_t\right]\right)\left(\hat{\varepsilon}_{t,t-i} - E\left[\hat{\varepsilon}_{t,t-i}\right]\right)^\top, \quad i = 0, 1, ..., h$$

which are replaced by the (auto)covariances of the smoothed residuals from the Kalman smoother, cf. (1.18) page 36:

$$\begin{aligned} \hat{C}_0 &= \hat{\varepsilon}_{t|T} \hat{\varepsilon}_{t|T}^\top + P^{\varepsilon}_{t|T} \\ \hat{C}_i &= \hat{\varepsilon}_{t|T} \hat{\varepsilon}_{t-i|T}^\top + P^{\varepsilon}_{\{t,t-i\}|T}. \end{aligned}$$

The upper panel of Table 1.5 shows that all factor models reject the null hypothesis of absence of residual autocorrelation when the smoothed quantities from a VAR(1) are used. However, the lower panel of the same table shows that when a

VAR(2) is considered, the null is not rejected when a sufficient number of lags is employed $(r \geq 8)$. Table 1.6 shows that whiteness of the residuals is further improved when a VAR(3) is considered and that the null of absence of residual autocorrelation cannot be rejected for a FAVAR model with eight factors, whereas a model with four factors is rejected. However, when a VAR(4) is considered, also $r = 4$ cannot be rejected for most h. An overall conclusion from these tests, is that the number of lags needed in the VAR seems to be decreasing in the number of factors. This is particularly pronounced for $r \geq 8$ where a maximum of three lags is needed. For the benchmark FAVAR with four factors, a VAR with six or seven lags seems to do well, which is also what Bernanke & Boivin (2003) find.

[Insert Tables 1.5 and 1.6]

Finally, a visual inspection of the autocorrelation functions of the smoothed residuals is also performed and combined with the multivariate Portmanteau test, and $\bar{R}^2$ the best FAVAR specification among $r = \{3, 4, .., 10\}$ is selected. Attention to model parsimony influences the choice when competing FAVAR specifications are encountered[28]. This selection of best specifications will be used in an evaluation of the robustness and sensitivity of different factor model specifications for the empirical monetary policy analysis.

To facilitate the interpretation of the following results, I introduce some shorthand notation for the various models. The notation $r8p3$ means $r = 8$ factors including the monetary policy factors with $p = 3$ lags in the FAVAR. The notation $r8p3\,(2)$ indicates a special focus on factor number two among the total of eight factors. Likewise, $r4p13\,(4)$ indicates a special focus on the last factor among the four factors each with thirteen lags; in fact, this is the monetary policy factor as this is always the last factor. The best specifications model among $r = \{3, 4, .., 10\}$ is $\{r3p7, r4p7, r5p6, r6p4, r7p5, \mathbf{r8p3}, r9p3, r10p2\}$ with the overall preferred model in bold. Figure 1.4 shows the autocorrelation functions for best specifications versus their VAR(1) counterpart. These autocorrelation functions are calculated for the monetary policy factor residuals and it should be noted that the improvement for

[28]For instance the specification with eight factors and three lags is preferred to the specification with eight factors and six lags. Similarly, the specification with six factors and four lags is preferred to the specification with six factors and eight lags.

the other variables in the VAR is often more pronounced than for the policy factor itself.

[Insert Figure 1.4]

The list of best FAVAR specifications is shortened marginally by removing $r3p7$ because of inferior fit and because of less plausible impulse responses. Also $r10p3$ is removed because of computational complexity and because this model does not add anything in terms of fit or interpretation.

The revised list $\{r4p7, r5p6, r6p4, r7p5, \mathbf{r8p3}, r9p3\}$ is now used in the empirical monetary policy analysis against the benchmark BBE-EM model denoted $r4p13$.

Figure 1.5 illustrates the gain in terms of increased fit for each obserserved variable of using the preferred model versus the BBE-EM and the preferred model by Bernanke et al. (2005).

[Insert Figure 1.5]

For the sake of brevity, the parameter estimates are not presented in detail. However, it should be mentioned that the estimates of the loadings are generally as expected in terms of signs and magnitude. For instance, the industrial production variables all load positively on the first "industrial production" factor with a coefficient close to unity. The unemployment variables generally load positively on the second "unemployment" factor whereas the largest loadings for the employment variables are generally negativ. For the monetary policy factor, it should be noted that the bond yields are positively related to this factor with loadings for the short-duration bonds close to unity, as expected. For the autoregressive parameters in Φ it should be noted that all eigenvalues of Φ are less than 1 in modulus.implying that the system is stationary.

1.4.4 A look at the factors

Given the choice of the preferred model that involves eight factors, the following offers some description and "labeling" of these latent dynamic factors. Figures 1.6 and 1.7 show the time series properties of the factors. Figures 1.8, 1.9, 1.10 and 1.11 show the correlation coefficients with the panel.

[Insert Figures 1.6, 1.7, 1.8, 1.9, 1.10 and 1.11]

Factor one is clearly an industrial production factor with a correlation with industrial production variables often exceeding 85%. Factor two is primarily related to unemployment with a correlation often exceeding 70% and secondarily related to Moody's BAA yield spread. Factor three is labeled a NAPM factor because it is primarily related to NAPM production, PMI, NAPM employment and NAPM orders, where correlation often exceeds 80%. Factor four is an "(overtime) hours in production" factor that is negatively related to dividend yield (proxy for risk aversion) and positively related to consumer expectations. Factor five is an inflation factor with correlation with inflation variables often exceeding 80%. Factor six is an employment factor closely related to help-wanted ads. and of course negatively related to unemployment, though this factor picks up something different from the unemployment, which can be seen from the correlations in Figure 1.10. Factor seven is a capacity utilization factor[29] and factor eight is the monetary policy factor.

1.4.5 Impulse response analysis

Having estimated the FAVAR model, we would like to study the dynamic responses of the variables in the panel following a shock to the federal funds, i.e. a shock to the VAR innovation for the monetary policy factor. However, to identify this innovation as a structural monetary policy shock, identifying restrictions need to be imposed and I follow Bernanke et al. (2005) by applying a recursive identification scheme proposed by Sims (1980). The recursive identification scheme (sometimes called a Wold causal ordering) implies that the first factor in the VAR is only affected by its own shock. The second factor is affected by its own shock and the first shock and so on. The monetary policy shock is influenced by all r shocks, so that if we for a minute interpret the first factor as output, the second as employment and so on, then output and employment shocks affect the monetary policy shock contemporaneously. However, monetary policy shocks do not affect output and employment shocks contemporaneously because monetary policy affects these with a lag.

[29] This factor is quite correlated with the employment factor number six. Although the correlation coefficient is 0.83 the capacity utilization factor is still different from factor six, which is apparent in the beginning of the period. Admittedly, this may be a weakness of the correlated factor approach, that factors can become quite correlated.

This recursive structure can be achieved by specifying the VAR innovations ε_t in terms of a new set of orthogonal residuals multiplied by a lower triangular matrix, such that $\varepsilon_t = Pe_t$. This particular example corresponds to a Cholesky decomposition of the covariance of ε_t, i.e. $\hat{Q} = PP^\top$. However, shocks of size one rather than size one standard deviation are sought for, so consider instead the decomposition $\hat{Q} = W\Sigma_e W^\top$, where $\Sigma_e = DD^\top$ is diagonal and $W = PD^{-1}$ has ones along the diagonal. Accordingly, for the VAR in F the response of the jth element of F at time $t+i$ due to a change in the kth element of F at time t is:

$$\frac{\partial \hat{E}\left[F_{j,t+i}|F_{k,t}, F_{t-1}, F_{t-2}, ...\right]}{\partial F_{k,t}} = \frac{\partial \hat{E}\left[F_{j,t+i}|F_{k,t}, F_{t-1}, F_{t-2}, ...\right]}{\partial F_{k,t}} \frac{\partial \varepsilon_{k,t}}{\partial e_{k,t}} = \psi_i w_j$$

for $i = 1, .2, ...h$, where ψ_i is the VAR moving average coefficient matrix and w_j is the jth column of the matrix W. ψ_i can be calculated recursively[30] from $\Phi(L)$ in the stationary system in (1.4), and monetary policy shocks corresponding to 25 basis point are now simply a matter of multiplying w_j by this (standardized) shock size. However, interest centers around the observed variables in levels $\tilde{X}$ rather than the transformed and standardized variables in X and therefore a multiplication of the loadings Λ is required, followed by a reverse transformation of the responses, i.e. $D(L)[\Lambda\psi_i w_j]$, cf. section 1.4.1. Consequently, the figures in the following correspond to a plot of $\{D(L)[\Lambda\psi_i w_j]\}_{i=1}^{h}$ which tracks the dynamic responses of the observed variables measured in standard deviation units to a 25 basis point shock to the FFR.

Figure 1.12 shows that the FAVAR model estimated by the EM algorithm delivers robust results in terms of impulse responses. Impulse responses for each of the best specifications in $\{r4p7, r5p6, r6p4, r7p5, \mathbf{r8p3}, r9p3\}$ are plotted against the benchmark BBE-EM ($r4p13$) for key macroeconomic variables. Moreover, the responses are very much in line with the results of Bernanke et al. (2005), although including confidence intervals around the impulse responses would further sharpen the conclusions.

[Insert Figure 1.12]

Each model delivers the same shape of the impulse response functions, i.e. the industrial production decreases by 0.6-0.7 standard deviations within one year fol-

[30] $\psi_i = \sum_{j=1}^{i} \psi_{i-j}\Phi_j$ for $i = 1, 2, ...$ and $\psi_0 = I$. See Lütkepohl (2007) chapter 2.

lowing a contractionary monetary policy shock, and it can be seen that the preferred model $r8p3$ returns more quickly to the starting point than BBE-EM. However, the speed of reversion is similar to the results in Figure II in Bernanke et al. (2005). For the price index, we see that the price puzzle noted by Sims[31] is almost eliminated, as there is a pronounced decrease in the price level following a contractionary monetary policy shock. The response is similar for all models but the preferred model has a particularly small initial positive effect and a pronounced negative response after one year, which is in line with Bernanke et al. (2005). The unemployment increases more than in the aforementioned example and most in the preferred model after one year but reverts to the starting point within four years. Furthermore, the response of NAPM commodity prices, capacity utilization rate, and average hourly earnings is also more pronounced than in Bernanke et al. (2005).

To summarize the impulse response analysis, I conclude that the FAVAR models deliver robust results across different specifications. Moreover, the preferred model eliminates the price puzzle and yields plausible impulse responses as in Bernanke et al. (2005). Compared to the aforementioned result some differences in the impulse responses following a contractionary policy shock can be noted. Firstly, the NAPM variables such as commodity price index, employment, new orders, and also capacity utilization rate are comparably affected more negatively, i.e. the impulse response shapes are "deeper". Similarly, unemployment peaks at a comparably higher level. However, comparably the same magnitude of the responses is seen for industrial production, CPI and the federal funds rate.

1.4.6 Forecast error variance decomposition

An alternative way of evaluating monetary policy shocks is to consider what role these shocks play in forecast errors. Specifically, in a forecast error variance decomposition, I calculate for a given forecast horizon what fraction of the total forecast error variance for a particular variable is due to a specific shock, for instance the

[31] A typical finding in standard VAR analysis of monetary policy is an increase in the price level following a contractionary monetary policy shock - hence the notion of a price puzzle, because we would expect a decrease. This can be explained as follows. Consider a simple policy rule that is linear in current inflation, current output gap and the Fed's expectations about future inflation. If the Fed expects future inflation to rise, it will accomodate this partly by increasing the federal funds rate. Consider now a VAR in the federal funds rate, inflation and output gap. Here, the information about the Fed's expectations is for obvious reasons not included in the VAR and is left in the residuals as a positive shock which happens alongside an increase in the price level (under the assumption that the Fed predict the rise in the price level correctly.)

monetary policy shock. Hence, the forecast error variance decomposition is similar to the R^2 measure but for forecast errors at different horizons. The proportion of the forecast error variance at horizon h of variable X_j due to the kth innovation $e_{k,t}$ is given by:

$$\omega_{jk}(h) = \frac{d_{kk}^2 \sum_{i=0}^{h-1} \left(\Psi_{jk,0}^2 + \Psi_{jk,1}^2 + ... + \Psi_{jk,h-1}^2\right)}{MSE\left(\hat{X}_{j,t+h|t}\right) + R_{j,j}}$$

where the $N \times r$ matrix $\Psi_{jk,i}$ is the (j,k) element of $(\Lambda_j \psi_i W)$ as a function of horizon $i \in h$, d_{kk}^2 is the (k,k) element of the diagonal matrix $DD^\top$, $MSE\left(\hat{X}_{j,t+h|t}\right)$ is the mean square error of $\left(X_{j,t+h} - \hat{X}_{j,t+h|t}\right)$ and $R_{j,j}$ is the variance of the jth idiosyncratic term. Details about the derivation are given in Appendix A.

The percentage of the forecast error variance explained by a monetary policy shock for the group of key macroeconomic variables is shown in Figure 2.3. Generally, a monetary shock rarely explains more than 10% of the forecast error variance, except for capacity utilization rate, (un)employment and new orders where forecast error variance is roughly doubled. The results are in line with similar findings in the literature, with only minor differences to be explained below.

[Insert Figure 1.12]

As only one structural shock, the monetary policy shock, is identified in this paper, it makes little sense to comment on impulse responses and variance decompositions for the other shocks. Nevertheless, the purpose of the upper panel of Table 2.3 is to illustrate that the fraction of the total forecast error variance of all the factors accounts for 40-50% and that the idiosyncratic component accounts for a significant fraction, on average 50-60%. This is also what Stock & Watson (2005) report. The difference between employing correlated versus uncorrelated factors as in the aforementioned result also shows up in the variance decomposition in the lower panel of Table 2.3. Whereas 93% of all of the forecast error variance for industrial production is explained by the first out of their seven factors in Stock & Watson (2005), only 50% shows up in the first correlated factor in this paper and the remaining 47% is spread evenly between the remaining seven factors.

[Insert Table 2.3]

Stock & Watson (2005) also estimate a principal component variant of Bernanke et al. (2005) and despite minor differences in the dataset, some comparisons with the two aforementioned papers, the closely related paper by Ahmadi & Uhlig (2008), and this one can be made. Generally, the monetary policy shocks play a larger role in the forecast error variance in this paper than in Stock & Watson (2005), except for the FFR and the bond yields; see below. Further, the forecast error variance decompositions in this paper are generally similar to those in Ahmadi & Uhlig (2008), although in this paper we see the largest influence of monetary policy shocks on the forecast error variance of unemployment peaking around 24 months at 35% but also the NAPM related variables such as new orders and employment are highly influenced. In contrast, the numbers in Stock & Watson (2005) are almost zero for the same variables, whereas in Bernanke et al. (2005) the corresponding numbers are somewhere in between. Moreover, in this paper, we see the smallest influence of the monetary shock on the FFR itself and in particular on the bond yields, although the variance decomposition in Ahmadi & Uhlig (2008) is roughly similar. In contrast, Bernanke et al. (2005) report that the fraction of the total forecast error variance of the FFR explained by its own shock is 45% compared to 3% in this paper, around 5% in Ahmadi & Uhlig (2008) and 7% in Stock & Watson (2005) for the long horizon. Strikingly, the fraction increases to 20% and 40% for the three-month T-bill and the five-year T-bond in the last-mentioned result. Finally, it can be noted that for all four papers, the forecast error variance of consumption and money supply is generally never explained by more than roughly 5%.

1.5 Conclusion

Three important issues are addressed in this paper. Firstly, an alternative identification scheme is applied that allows for correlated factors, which is desirable if one seeks a macroeconomic interpretation of the latent factors. For instance, in the correlated factor approach here, the industrial production factor and the unemployment factor are allowed to be correlated, and they are estimated to have a correlation of 0.23.

Secondly, I investigate the EM algorithm as an alternative estimation method to the two-step principal component method and the one-step Bayesian method. In general, it is easy to impose parameter restrictions on both the measure equation and the state transition equation, which is illustrated plentifully in Bork et al. (2008) where explicit interpretation of the factors is achieved through identification.

Thirdly, the sensitivity of the statistical fit and impulse response analysis to different factor specifications is evaluated as well as a careful model selection. The combination of the panel information criteria by Bai & Ng (2002) for the number of factors and the standard Akaike, Schwarz or Hannan-Quinn information criteria for the VAR order results in a preferred FAVAR model with eight factors and only three lags. This model naturally delivers a better fit than models with fewer factors without compromising well-specified factor dynamics or the plausibility of the impulse response analysis. Interestingly, some of the key macroeconomic variables such as industrial production and employment seem to respond somewhat more in the preferred model compared to the EM algorithm equivalent of Bernanke et al. (2005) with four factors and thirteen lags. Furthermore, the NAPM indices (commodity price, new orders, and employment) as well as unemployment respond somewhat more to a monetary policy shock than in the aforementioned model(s).

Generally, it is found that the FAVAR models investigated here deliver robust results in terms of fit, impulse responses and forecast error variance decompositions across the best-specified models for the different numbers of factors included. I find that the fewer the factors used in the FAVAR the more lags are needed to achieve a well-specified model and vice versa. Hence, it seems possible to trade off a model with a few factors but necessarily many lags for a model with more factor but fewer lags; specifically, it is possible to trade off a four-factor and seven-lag model for an eight-factor and three-lag model with the benefit of a ten percentrage point increase in the overall R^2. This observation accords with the theoretical result that complicated factor dynamics may be substituted by the information in the panel dataset. One objection might be that more factors are the result of the correlated factor approach in contrast to the uncorrelated factor approach. However, besides the above-mentioned theoretical result, it should be noted that the four-factor and thirteen-lag benchmark model performs equally well in terms of fit and plausibility of the impulse responses to the uncorrelated factor approach in Bernanke et al. (2005). On this basis, there is no clear sign that the correlated factor approach needs relatively more factors to achieve the same fit.

A Forecast error variance decomposition

Consider the forecast error of the optimal $h-$step ahead forecast for the jth observed variable:

$$
\begin{aligned}
X_{j,t+h} - \hat{X}_{j,t+h|t} &= \sum_{i=1}^{h-1} \left[\ \Lambda_j \psi_i W \ \right] e_{t+h-i} + \xi_{j,t+h} \\
&= \sum_{i=1}^{h-1} \Psi_i e_{t+h-i} + \xi_{j,t+h} \\
&= \sum_{k=1}^{K} \left(\Psi_{jk,0} e_{k,t+h} + \Psi_{jk,1} e_{k,t+h-1} + ... + \Psi_{jk,h-1} e_{k,t+1} \right) + \xi_{j,t+h}
\end{aligned}
$$

where e_t is the orthogonal residual defined from the VAR residuals, $\varepsilon_t = Pe_t$ where P is the Cholesky factor from the decomposition of the covariance of ε_t into $\hat{Q} = PP^\top$. This covariance matrix is further rewritten as explained in section 1.4.5 as $\hat{Q} = W\Sigma_e W^\top$, where $\Sigma_e = DD^\top$ is diagonal and $W = PD^{-1}$ has ones along the diagonal. Moreover, $\Psi_i = \Lambda_j \psi_i W$ is a $N \times r$ matrix and $\xi_{j,t+h}$ is the jth idiosyncratic term. The mean square error of $\left(X_{j,t+h} - \hat{X}_{j,t+h|t} \right)$ is denoted $MSE\left(\hat{X}_{j,t+h|t} \right)$ and given by:

$$
MSE\left(\hat{X}_{j,t+h|t} \right) = \sum_{k=1}^{K} \left(\Psi_{jk,0}^2 d_{kk}^2 + \Psi_{jk,1}^2 d_{kk}^2 + ... + \Psi_{jk,h-1}^2 d_{kk}^2 \right) + R_{j,j}
$$

where d_{kk}^2 is the (k,k) element of the diagonal matrix $DD^\top$ and $R_{j,j}$ is the variance of the jth idiosyncratic term. The proportion of the forecast error variance at horizon h of variable X_j due to the kth innovation $e_{k,t}$ is given by:

$$
\omega_{jk}(h) = \frac{d_{kk}^2 \sum_{i=0}^{h-1} \left(\Psi_{jk,0}^2 + \Psi_{jk,1}^2 + ... + \Psi_{jk,h-1}^2 \right)}{MSE\left(\hat{X}_{j,t+h|t} \right) + R_{j,j}}
$$

B Kalman filter, Kalman smoother and the EM algorithm

B.1 The Kalman filter

The Kalman filter is an algorithm for sequentially updating a linear projection for a dynamic system. Denote the information set $\mathcal{X}_t = \{X_1, ..., X_t\}$ and by $\hat{F}_{t+1|t} = E\left[F_{t+1}|\, \mathcal{X}_t\right]$ the linear projection of F_{t+1} on $\mathcal{X}_t$. The variance is denoted $\hat{P}_{t+1|t} = var\left(F_{t+1}|\, \mathcal{X}_t\right)$. The Kalman filter recursions for $t = 1, .., T$ can then be written as:

$$
\begin{aligned}
&\hat{F}_{t+1|t} = \Phi\hat{F}_{t|t-1} + K_t\left(X_t - \Lambda\hat{F}_{t|t-1}\right) \\
&\hat{P}_{t+1|t} = \Phi\hat{P}_{t|t-1}L_t^{\top} + Q \\
&\text{where} \\
&\underset{n\times 1}{\xi_t} = X_t - \Lambda\hat{F}_{t|t-1} \\
&\underset{n\times n}{P^{\xi\xi}_{t|t-1}} = \Lambda\hat{P}_{t|t-1}\Lambda^{\top} + R \\
&\underset{k\times n}{K_t} = \Phi\hat{P}_{t|t-1}\Lambda^{\top}\left(\Lambda\hat{P}_{t|t-1}\Lambda^{\top} + R\right)^{-1} \\
&\underset{k\times k}{L_t} = \Phi - K_t\Lambda
\end{aligned}
\tag{1.11}
$$

B.2 Kalman smoothing

Kalman smoothing reconstructs the full state sequence $\{F_1, .., F_T\}$ given the observations $\{X_1, .., X_T\}$. Smoothing provides us with more accurate inference on the state variables since it uses more information than the basic filter. The Kalman smoother recursions are based on the efficient smoother by de Jong & Mackinnon (1988) and de Jong (1989) which is used in Koopman & Shephard (1992) and given by:

$$
\begin{aligned}
\hat{F}_{t|T} &= \hat{F}_{t|t-1} + \hat{P}_{t|t-1}\Lambda^{\top}\left[\hat{P}^{\xi\xi}_{t|t-1}\right]^{-1}\xi_t && (1.12)\\
&\quad + \hat{P}_{t|t-1}L_t^{\top}r_t && (1.13)\\
&= \hat{F}_{t|t-1} + \hat{P}_{t|t-1}r_{t-1} \quad \text{(alternatively)}\\
\hat{P}_{t|T} &= \hat{P}_{t|t-1} - \hat{P}_{t|t-1}N_{t-1}\hat{P}_{t|t-1} && (1.14)\\
\hat{P}_{\{t,t-1\}|T} &= \left(I - \hat{P}_{t|t-1}N_{t-1}\right)L_{t-1}\hat{P}_{t-1|t-2}, && (1.15)\\
\text{for } t &= T-1, ..., 1 && (1.16)\\
cov\left(F_t - \hat{F}_{t|T}, F_j - \hat{F}_{j|T}\right) &= \hat{P}_{t|t-1}L_t^{\top}L_{t+1}^{\top}\cdots L_{j-1}^{\top}\left[I - N_{j-1}\hat{P}_{j|j-1}\right] && (1.17)\\
\text{for } j &\geq t
\end{aligned}
$$

where:

$$
\begin{aligned}
r_{t-1} &= \Lambda^{\top}\left[\hat{P}^{\xi\xi}_{t|t-1}\right]^{-1}\xi_t + L_t^{\top}r_t, \text{ for } 1 \leq t < T \text{ and } r_T = 0\\
N_{t-1} &= \Lambda^{\top}\left[\hat{P}^{\xi\xi}_{t|t-1}\right]^{-1}\Lambda + L_t^{\top}N_tL \text{ for } 1 \leq t < T \text{ and } N_T = 0\\
L_t &= \Phi - K_t\Lambda = \Phi - \Phi\hat{P}_{t|t-1}\Lambda^{\top}\left[\hat{P}^{\xi\xi}_{t|t-1}\right]^{-1}\Lambda.
\end{aligned}
$$

The smoothed residuals given by Koopman (1993) are used in for instance the Portmanteau test:

$$
\begin{aligned}
\hat{\varepsilon}_{t|T} &= E\left[\varepsilon_t|\, \mathcal{X}_T\right] = \hat{F}_{t|T} - \Phi\hat{F}_{t-1|T}\\
&= Q\Upsilon^{\top}r_t, \quad t = 1, .., T && (1.18)
\end{aligned}
$$

and variance and covariance:

$$
\begin{aligned}
var\left(\varepsilon_t|\, \mathcal{X}_T\right) &= P^{\varepsilon}_{t|T}\\
&= Q - Q\Upsilon^{\top}N_t\Upsilon Q\\
cov\left(\varepsilon_t - \hat{\varepsilon}_{t|T}, \varepsilon_j - \hat{\varepsilon}_{j|T}\right) &= P^{\varepsilon}_{\{t,j\}|T}\\
&= -Q\Upsilon^{\top}L_{t+1}^{\top}\cdots L_{j-1}^{\top}L_j^{\top}N_j\Upsilon Q, \quad j = t+1, ..., T
\end{aligned}
$$

with the convention that $L_t^{\top}\cdots L_{T-1}^{\top} = I_r$ when $t = T$ and $L_t^{\top}\cdots L_{T-1}^{\top} = L_{T-1}^{\top}$ when $t = T-1$.

B.3 The complete data likelihood and the incomplete data likelihood

Under the Gaussian assumption including $F_0 \sim N(\mu_0, P_0)$ and ignoring the constant, the complete data likelihood of equation (1.4) page 11 assuming a VAR(1) for simplicity and ignoring Υ is written as:

$$
\begin{aligned}
-2\ln L_{\mathcal{F},\mathcal{X}}(\Theta) &= \ln|P_0| + (F_0-\mu_0)^\top P_0^{-1}(F_0-\mu_0) \\
&\quad + T\cdot\ln|Q| + \sum_{t=1}^{T}(F_t-\Phi F_{t-1})^\top Q^{-1}(F_t-\Phi F_{t-1}) \\
&\quad + T\cdot\ln|R| + \sum_{t=1}^{T}(X_t-\Lambda F_t)^\top R^{-1}(X_t-\Lambda F_t)
\end{aligned} \tag{1.19}
$$

given that we can observe the states $\mathcal{F}_T = \{F_0, .., F_T\}$ as well as the observations $\mathcal{X}_T = \{X_1, .., X_T\}$. However, given $\mathcal{X}_T$ and initial values of the parameter estimates (denoted $\Theta^{(j-1)}$), the conditional expectation of the complete data likelihood can be written as:

$$
\begin{aligned}
\mathcal{Q}\left(\Theta|\Theta^{(j-1)}\right) &= E\left[-2\ln L_{\mathcal{F},\mathcal{X}}(\Theta)\,\middle|\,\mathcal{X}_T, \Theta^{(j-1)}\right] \\
&= \ln|P_0| + \mathrm{tr}\left[P_0^{-1}\left\{\left(\hat{F}_{0|T}-\mu_0\right)\left(\hat{F}_{0|T}-\mu_0\right)^\top + P_{0|T}\right\}\right] \\
&\quad + T\cdot\ln|Q| + \mathrm{tr}\left[Q^{-1}\left\{C - B\Phi^\top - \Phi B^\top + \Phi A \Phi^\top\right\}\right] \\
&\quad + T\cdot\ln|R| \\
&\quad + \mathrm{tr}\left[R^{-1}\sum_{t=1}^{T}\left\{\left(X_t-\Lambda\hat{F}_{t|T}\right)\left(X_t-\Lambda\hat{F}_{t|T}\right)^\top + \Lambda\hat{P}_{t|T}\Lambda^\top\right\}\right]
\end{aligned} \tag{1.20}
$$

where the following moments can be calculated from the Kalman smoother listed above:

$$
\begin{array}{ll}
A = \sum_{t=1}^{T}\left(\hat{F}_{t-1|T}\hat{F}_{t-1|T}^\top + \hat{P}_{t-1|T}\right) & B = \sum_{t=1}^{T}\left(\hat{F}_{t|T}\hat{F}_{t-1|T}^\top + \hat{P}_{\{t,t-1\}|T}\right) \\
C = \sum_{t=1}^{T}\left(\hat{F}_{t|T}\hat{F}_{t|T}^\top + \hat{P}_{t|T}\right) & D = \sum_{t=1}^{T} X_t\hat{F}_{t|T}^\top \\
E = \sum_{t=1}^{T} X_t X_t^\top &
\end{array}
$$

A useful trick to arrive at (1.20) is to consider the decomposition of the true state variable $F_t = \hat{F}_{t|T} + \left(F_t - \hat{F}_{t|T}\right)$, which explains the terms in for instance C, where:

$\hat{P}_{t|T} = E\left[\left(F_t - \hat{F}_{t|T}\right)\left(F_t - \hat{F}_{t|T}\right)^{\top}\bigg|\mathcal{X}_T\right]$.

The estimator of Φ^* subject to linear restrictions is:

$$\begin{aligned}\operatorname{vec}(\Phi^*) &= \operatorname{vec}\left(BA^{-1}\right) \\ &\quad + \left(A^{-1}\otimes Q\right)H_{\Phi}^{\top}\left[H_{\Phi}\left(A^{-1}\otimes Q\right)H_{\Phi}^{\top}\right]^{-1}\left\{\kappa_{\Phi} - H_{\Phi}\operatorname{vec}\left(BA^{-1}\right)\right\}\end{aligned} \tag{1.21}$$

where κ_{Φ} is a $\varrho \times 1$ vector and the restriction matrix H_{Φ} is of dimension $\varrho \times r^2$.

C Analytical derivatives of the log likelihood function

The following is primarily from Jungbacker & Koopman (2008) and Koopman & Shephard (1992). A key result $\frac{\partial \log L_{\mathcal{Y}}(\Theta)}{\partial Q}\Big|_{Q=Q^*} = \frac{\partial \mathcal{Q}(\Theta|\Theta^*)}{\partial Q}\Big|_{Q=Q^*}$ is from Louis (1982).

Consider the following derivatives of the log likelihood function for the state space model with incomplete data:

$$\begin{aligned}\frac{\partial \log L_{\mathcal{Y}}(\Theta)}{\partial Q}\bigg|_{Q=Q^*} &= \frac{\partial \mathcal{Q}(\Theta|\Theta^*)}{\partial Q}\bigg|_{Q=Q^*} \\ &= Q^{-1}(S - T\cdot Q)Q^{-1} - \frac{1}{2}\operatorname{diag}\left(Q^{-1}(S - T\cdot Q)Q^{-1}\right)\end{aligned}$$

where:

$$S = C - B\Phi^{\top} - \Phi B^{\top} + \Phi A \Phi^{\top}$$

and where Q is the covariance matrix of the innovation error in the transition equation:

$$\frac{\partial \log L_{\mathcal{Y}}(\Theta)}{\partial \Phi}\bigg|_{\Phi=\Phi^*} = \frac{\partial \mathcal{Q}(\Theta|\Theta^*)}{\partial \Phi}\bigg|_{\Phi=\Phi^*} = Q^{-1}(B - \Phi A)$$

where Φ contains the autoregressive parameters in the transition equation. Moreover, the derivative with respect to Λ is:

$$\left.\frac{\partial \log L_{\mathcal{Y}}(\Theta)}{\partial \Lambda}\right|_{\Lambda=\Lambda^*} = \left.\frac{\partial \mathcal{Q}(\Theta|\Theta^*)}{\partial \Lambda}\right|_{\Lambda=\Lambda^*} = R^{-1}\left(\sum_{t=1}^{T} y_t \hat{F}_{t|T}^{\top} - \Lambda C\right)$$

where Λ is the loading matrix, R is the covariance of the measurement errors, y_t is the data in the panel data set at time t and $\hat{F}_{t|T}^{\top}$ is the smoothed dynamic factor. Finally, the derivative with respect to the covariance of the measurement errors is:

$$\begin{aligned} \left.\frac{\partial \log L_{\mathcal{Y}}(\Theta)}{\partial R}\right|_{R=R^*} &= \left.\frac{\partial \mathcal{Q}(\Theta|\Theta^*)}{\partial R}\right|_{R=R^*} \\ &= R^{-1}(M - T \cdot R) R^{-1} - \frac{1}{2}\operatorname{diag}\left(R^{-1}(M - T \cdot R) R^{-1}\right). \end{aligned}$$

where

$$M = E - D\Lambda^{\top} - \Lambda D^{\top} + \Lambda C \Lambda^{\top}$$

Table 1.1: The statistical fit in the preferred specification is robust to alternative restrictions and an alternative factor ordering.

	mean R^2 for groups of variables	
	Preferred restrictions	Alternative restrictions
Real output and income (21)	58%	58%
Employment and hours (27)	55%	56%
Price Indexes (16)	58%	57%
Key economic variables (20)	62%	61%

The table supports the argument that a different set of restricted variables does not significantly change the estimated factors. Moreover, the R^2's are robust to an alternative factor ordering. Consider for instance a unit restriction imposed on "Industrial Production: Mining" as an alternative to the a unit restriction on "Industrial Production: Manufacturing" as in the preferred specification. This alternative restriction does not lead to a significantly different R^2 for "Industrial Production: Mining" as R^2 stays within 8-9 percent in both specifications.

Notes: "Real output and income", "Employment and hours" and "Price Indexes" in the left column correspond to the organization of the panel into groups of similar variables; see the data appendix. The column heading "Key economic variables" corresponds to the set of variables used in the forecast error variance decomposition (the set is also used by Bernanke, Boivin and Eliasz (2005)). The numbers in parentheses refer to the number of variables in the categories. The alternative exactly identifying restrictions are imposed on variable numbers 109,24,14,46,23,106,19,77.

Table 1.2: Akaike information criterion for a given number of factors.

	number of factors							
lags	3	4	5	6	7	8	9	10
1	- 11.467	- 13.271	- 17.625	- 24.756	- 22.658	-	- 30.480	- 43.801
2	- 11.634	- 14.151	**- 23.221**	- 27.482	- 24.302	- 35.706	- 35.812	- 45.512
3	- 11.616	**- 16.646**	- 19.462	-	-	- 37.692	**- 40.605**	**- 50.218**
4	- 11.663	- 16.182	- 18.741	**- 28.607**	-	- 36.904	- 38.251	- 48.553
5	- 11.562	-	- 19.479	- 26.748	**- 30.378**	- 37.673	-	-
6	**- 11.904**	- 15.653	- 20.636	- 26.683	- 24.834	**- 38.726**	- 37.483	- 48.733
7	- 11.784	- 15.918	- 19.039	-	-	- 37.302	- 37.034	-
8	- 10.627	- 14.955	- 18.993	- 28.004	-	- 37.888	-	-
9	- 10.671	- 15.037	- 19.456	- 26.212	- 26.262	-	-	-
10	- 10.670	- 15.257	- 20.138	- 25.987	- 26.582	-	-	-
11	- 10.898	- 15.212	- 19.939	-	-	-	-	-
12	- 11.459	- 15.112	-	-	-	-	-	-
13	-	- 15.045	- 20.121	-	-	-	-	-

A bold number represents a minimum.

Table 1.3: Schwarz information criterion for a given number of factors.

	number of factors							
lags	3	4	5	6	7	8	9	10
1	- 11.392	- 13.139	- 17.418	- 24.457	- 22.252	-	- 29.809	- 42.972
2	**- 11.485**	- 13.886	**- 22.806**	- 26.885	- 23.490	- 34.645	- 34.469	- 43.854
3	- 11.393	**- 16.248**	- 18.841	-	-	**- 36.100**	**- 38.590**	**- 47.731**
4	- 11.365	- 15.651	- 17.912	**- 27.413**	-	- 34.782	- 35.565	- 45.237
5	- 11.189	-	- 18.443	- 25.255	**- 28.347**	- 35.020	-	-
6	- 11.456	- 14.857	- 19.392	- 24.893	- 22.397	- 35.543	- 33.454	- 43.759
7	- 11.262	- 14.990	- 17.588	-	-	- 33.588	- 32.333	-
8	- 10.030	- 13.894	- 17.335	- 25.616	-	- 33.643	-	-
9	- 9.999	- 13.843	- 17.590	- 23.526	- 22.606	-	-	-
10	- 9.924	- 13.930	- 18.066	- 23.002	- 22.520	-	-	-
11	- 10.078	- 13.753	- 17.659	-	-	-	-	-
12	- 10.563	- 13.520	-	-	-	-	-	-
13	-	- 13.320	- 17.426	-	-	-	-	-

Table 1.4: Hannan and Quinn information criterion for a given number of factors.

	number of factors							
lags	3	4	5	6	7	8	9	10
1	- 11.438	- 13.219	- 17.544	- 24.639	- 22.499	-	- 30.217	- 43.476
2	- 11.576	- 14.047	**- 23.058**	- 27.248	- 23.984	- 35.290	- 35.285	- 44.862
3	- 11.529	**- 16.490**	- 19.219	-	-	- 37.068	**- 39.815**	**- 49.243**
4	- 11.546	- 15.974	- 18.416	**- 28.139**	-	- 36.072	- 37.198	- 47.253
5	- 11.416	-	- 19.073	- 26.163	**- 29.582**	- 36.633	-	-
6	**- 11.729**	- 15.341	- 20.148	- 25.981	- 23.879	**- 37.478**	- 35.904	- 46.783
7	- 11.580	- 15.554	- 18.470	-	-	- 35.846	- 35.191	-
8	- 10.393	- 14.539	- 18.343	- 27.068	-	- 36.224	-	-
9	- 10.408	- 14.569	- 18.724	- 25.159	- 24.829	-	-	-
10	- 10.378	- 14.737	- 19.326	- 24.817	- 24.990	-	-	-
11	- 10.577	- 14.640	- 19.045	-	-	-	-	-
12	- 11.108	- 14.488	-	-	-	-	-	-
13	-	- 14.369	- 19.064	-	-	-	-	-

Table 1.5: Multivariate Portmanteau tests.

Test statistics based on smoothed residuals from a VAR(1)

h	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$	$r=8$	$r=9$	$r=10$
1	140.00* {0.000}	218.04* {0.000}	252.36* {0.000}	304.13* {0.000}	312.15* {0.000}	362.92* {0.000}	360.50* {0.000}	384.54* {0.000}
2	161.83* {0.000}	250.56* {0.000}	294.05* {0.000}	363.69* {0.000}	387.10* {0.000}	442.70* {0.000}	450.37* {0.000}	471.28* {0.000}
3	181.62* {0.000}	276.12* {0.000}	329.30* {0.000}	419.71* {0.000}	460.05* {0.000}	505.91* {0.000}	532.32* {0.000}	548.44* {0.000}
4	188.83* {0.000}	296.30* {0.000}	361.72* {0.000}	480.70* {0.000}	531.55* {0.000}	558.04* {0.000}	623.72* {0.000}	630.75* {0.000}
5	210.70* {0.000}	324.65* {0.000}	401.96* {0.000}	535.43* {0.000}	607.16* {0.000}	613.96* {0.000}	731.48* {0.000}	742.25* {0.000}
6	230.42* {0.000}	347.53* {0.000}	444.24* {0.000}	593.32* {0.000}	685.95* {0.000}	675.11* {0.000}	825.78* {0.000}	837.65* {0.000}
7	243.14* {0.000}	368.32* {0.000}	482.16* {0.000}	650.24* {0.000}	755.09* {0.000}	734.22* {0.000}	906.11* {0.000}	922.48* {0.000}
8	288.93* {0.000}	418.07* {0.000}	540.25* {0.000}	732.82* {0.000}	866.36* {0.000}	799.02* {0.000}	1,024.43* {0.000}	1,045.57* {0.000}
9	308.38* {0.000}	440.20* {0.000}	573.01* {0.000}	782.47* {0.000}	927.33* {0.000}	849.77* {0.000}	1,094.50* {0.000}	1,123.82* {0.000}
10	320.56* {0.000}	459.57* {0.000}	598.82* {0.000}	819.95* {0.000}	991.55* {0.000}	884.62* {0.000}	1,182.08* {0.000}	1,210.20* {0.000}
12	344.19* {0.000}	497.36* {0.000}	643.14* {0.000}	869.03* {0.000}	1,043.41* {0.000}	942.27* {0.000}	1,272.57* {0.000}	1,302.48* {0.000}
13	377.01* {0.000}	539.93* {0.000}	702.67* {0.000}	935.97* {0.000}	1,108.93* {0.000}	1,011.22* {0.000}	1,359.05* {0.000}	1,386.90* {0.000}
14	404.69* {0.000}	576.69* {0.000}	738.00* {0.000}	990.09* {0.000}	1,174.94* {0.000}	1,064.53* {0.000}	1,442.79* {0.000}	1,470.31* {0.001}
15	425.71* {0.000}	617.82* {0.000}	788.63* {0.000}	1,040.62* {0.000}	1,240.25* {0.000}	1,122.43* {0.000}	1,515.66* {0.000}	1,549.43* {0.003}

Test statistics based on smoothed residuals from a VAR(2)

h	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$	$r=8$	$r=9$	$r=10$
1	58.66* {0.000}	59.48* {0.000}	88.26* {0.000}	88.05* {0.000}	120.09* {0.000}	127.29* {0.000}	150.73* {0.000}	158.01* {0.000}
2	69.11* {0.000}	72.03* {0.000}	105.40* {0.000}	110.30* {0.003}	162.96* {0.000}	178.34* {0.002}	227.23* {0.001}	220.72 {0.150}
3	73.66* {0.000}	90.34* {0.000}	138.02* {0.000}	161.16* {0.001}	222.99* {0.000}	240.76* {0.010}	313.06* {0.002}	297.66 {0.527}
4	93.25* {0.000}	113.36* {0.000}	180.74* {0.000}	209.08* {0.000}	287.91* {0.000}	303.90* {0.021}	397.97* {0.003}	393.07 {0.588}
5	118.35* {0.000}	139.77* {0.000}	218.54* {0.000}	250.64* {0.000}	342.75* {0.000}	360.20 {0.060}	466.35* {0.019}	457.32 {0.915}
6	128.73* {0.000}	154.08* {0.000}	244.03* {0.000}	284.04* {0.001}	391.37* {0.000}	402.32 {0.250}	519.99 {0.139}	526.91 {0.986}
7	166.25* {0.000}	185.86* {0.000}	277.94* {0.000}	333.40* {0.000}	475.13* {0.000}	477.40 {0.163}	609.53 {0.105}	633.67 {0.965}
8	185.63* {0.000}	204.51* {0.000}	299.97* {0.000}	366.12* {0.001}	521.55* {0.000}	528.66 {0.296}	663.27 {0.330}	688.64 {0.998}
9	193.90* {0.000}	221.42* {0.000}	322.41* {0.000}	393.23* {0.005}	570.79* {0.000}	579.33 {0.453}	731.97 {0.462}	782.77 {0.998}
10	216.72* {0.000}	246.77* {0.000}	352.08* {0.000}	439.60* {0.003}	624.22* {0.000}	612.92 {0.773}	810.22 {0.491}	858.89 {1.000}
12	242.06* {0.000}	279.14* {0.000}	402.61* {0.000}	492.10* {0.001}	679.98* {0.000}	658.83 {0.887}	872.45 {0.665}	943.96 {1.000}
13	257.49* {0.000}	294.25* {0.001}	421.62* {0.000}	532.21* {0.001}	730.13* {0.000}	711.68 {0.927}	935.29 {0.796}	1,020.73 {1.000}
14	270.65* {0.000}	319.49* {0.000}	460.12* {0.000}	567.78* {0.001}	782.21* {0.000}	757.77 {0.969}	995.82 {0.895}	1,084.72 {1.000}

The rows represent test statistics of residual autocorrelation up to order h. H_0: Residual autocorrelation up to lag h is zero. p-values in {}. * indicates rejection on 5 pct. level.

Table 1.6: Multivariate Portmanteau tests.

Test statistics based on smoothed residuals from a VAR(3)

h	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$	$r=8$	$r=9$	$r=10$
1	26.30* {0.002}	52.15* {0.000}	49.16* {0.003}	failed max. iterations		64.70 {0.452}	100.79 {0.068}	97.81 {0.543}
2	31.19* {0.027}	71.02* {0.000}	87.50* {0.001}	-	-	102.85 {0.950}	178.52 {0.178}	160.17 {0.983}
3	46.73* {0.011}	92.27* {0.000}	118.99* {0.001}	-	-	156.50 {0.972}	247.82 {0.402}	239.47 {0.996}
4	64.98* {0.002}	106.51* {0.001}	147.55* {0.001}	-	-	194.08 {0.999}	306.45 {0.751}	299.03 {1.000}
5	74.80* {0.004}	123.08* {0.001}	174.07* {0.003}	-	-	224.64 {1.000}	344.81 {0.986}	367.33 {1.000}
6	99.72* {0.000}	155.54* {0.000}	212.95* {0.001}	-	-	284.73 {1.000}	419.23 {0.987}	446.80 {1.000}
7	114.30* {0.000}	166.46* {0.001}	231.74* {0.003}	-	-	328.48 {1.000}	468.23 {0.999}	496.79 {1.000}
8	120.55* {0.000}	185.85* {0.001}	260.84* {0.003}	-	-	370.56 {1.000}	527.45 {1.000}	574.34 {1.000}
9	134.73* {0.000}	199.94* {0.001}	279.82* {0.008}	-	-	404.49 {1.000}	593.54 {1.000}	636.62 {1.000}
10	157.52* {0.000}	212.42* {0.004}	301.64* {0.014}	-	-	446.82 {1.000}	653.77 {1.000}	709.65 {1.000}
11	170.53* {0.000}	239.61* {0.001}	345.27* {0.003}	-	-	490.68 {1.000}	708.49 {1.000}	783.20 {1.000}
12	183.29* {0.000}	253.33* {0.002}	377.71* {0.002}	-	-	527.04 {1.000}	763.42 {1.000}	841.09 {1.000}

Test statistics based on smoothed residuals from a VAR(4)

h	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$	$r=8$	$r=9$	$r=10$
1	21.00* {0.013}	28.82* {0.025}	60.72* {0.000}	24.66 {0.924}	-	64.46 {0.460}	125.89* {0.001}	105.20 {0.341}
2	37.59* {0.004}	43.90 {0.078}	92.43* {0.000}	52.71 {0.957}	-	94.92 {0.987}	171.86 {0.283}	176.64 {0.882}
3	60.32* {0.000}	63.89 {0.062}	121.33* {0.001}	75.17 {0.993}	-	141.01 {0.998}	216.79 {0.886}	230.63 {0.999}
4	67.40* {0.001}	75.83 {0.148}	139.58* {0.006}	89.52 {1.000}	-	166.00 {1.000}	256.54 {0.998}	288.91 {1.000}
5	90.86* {0.000}	104.85* {0.033}	175.20* {0.002}	126.35 {0.999}	-	219.22 {1.000}	314.99 {1.000}	351.50 {1.000}
6	103.86* {0.000}	110.88 {0.142}	191.64* {0.012}	140.07 {1.000}	-	257.89 {1.000}	372.87 {1.000}	404.90 {1.000}
7	109.69* {0.000}	126.97 {0.158}	218.54* {0.014}	162.72 {1.000}	-	292.05 {1.000}	447.16 {1.000}	478.20 {1.000}
8	121.38* {0.000}	145.09 {0.143}	239.23* {0.030}	185.96 {1.000}	-	323.48 {1.000}	512.27 {1.000}	535.13 {1.000}
9	143.21* {0.000}	165.07 {0.110}	265.68* {0.033}	217.70 {1.000}	-	363.93 {1.000}	570.38 {1.000}	599.09 {1.000}
10	155.96* {0.000}	196.74* {0.026}	314.68* {0.003}	267.49 {1.000}	-	412.06 {1.000}	635.52 {1.000}	677.01 {1.000}
11	167.10* {0.000}	212.70* {0.031}	332.92* {0.010}	284.29 {1.000}	-	447.43 {1.000}	685.14 {1.000}	728.51 {1.000}

The rows represent test statistics of residual autocorrelation up to order h. H_0: Residual autocorrelation up to lag h is zero. p-values in {}. * indicates rejection on 5 pct. level. Unreported numbers show that a VAR(5) fixes the residual autocorrelation for $r=5$, whereas for $r=3$, the problem remains.

Table 1.7: Forecast error variance decompositions for key macroeconomic variables.

	$F(8,1)$	$F(8,2)$	$F(8,3)$	$F(8,4)$	$F(8,5)$	$F(8,6)$	$F(8,7)$	**FFR**)	**F − total**	**Idio.**
6m	0.06	0.03	0.02	0.06	0.07	0.07	0.04	0.04	0.39	0.61
12m	0.06	0.03	0.04	0.07	0.07	0.06	0.06	0.05	0.43	0.57
24m	0.06	0.03	0.06	0.07	0.07	0.06	0.08	0.05	0.47	0.53
60m	0.06	0.03	0.10	0.07	0.07	0.05	0.09	0.04	0.51	0.49

60-month horizon	$F(8,1)$	$F(8,2)$	$F(8,3)$	$F(8,4)$	$F(8,5)$	$F(8,6)$	$F(8,7)$	**FFR**
77) Federal funds rate	0.03	0.06	0.35	0.23	0.03	0.08	0.19	**0.03**
16) IP: total index	0.50	0.03	0.05	0.11	0.03	0.10	0.09	**0.07**
108) CPI-U: all items	0.02	0.07	0.20	0.09	0.42	0.02	0.09	**0.02**
78) US Tbill, 3m.	0.03	0.06	0.35	0.20	0.03	0.10	0.18	**0.02**
81) Tbond const 5yr.	0.04	0.07	0.38	0.17	0.05	0.18	0.09	**0.01**
96) Monetary base	0.00	0.02	0.02	0.02	0.01	0.01	0.03	**0.00**
93) Money stock: M2	0.01	0.06	0.03	0.01	0.03	0.03	0.06	**0.01**
74) FX: Japan	0.01	0.01	0.01	0.00	0.00	0.01	0.03	**0.00**
102) NAPM commodity prices	0.02	0.06	0.17	0.05	0.10	0.07	0.11	**0.06**
17) Capacity util rate	0.03	0.01	0.17	0.12	0.09	0.04	0.21	**0.12**
49) Pers cons exp: total	0.01	0.00	0.01	0.01	0.04	0.00	0.01	**0.00**
50) Pers cons exp: tot. dur	0.01	0.00	0.01	0.00	0.01	0.00	0.01	**0.00**
51) Pers cons exp: nondur.	0.01	0.00	0.01	0.00	0.03	0.00	0.01	**0.00**
26) Unempl. rate: all wrks	0.05	0.02	0.21	0.10	0.19	0.05	0.21	**0.15**
48) NAPM Empl. Index	0.02	0.01	0.20	0.12	0.07	0.10	0.22	**0.15**
118) Avg hr earnings constr.	0.01	0.01	0.01	0.00	0.01	0.00	0.01	**0.00**
54) Housing starts: nonfarm	0.02	0.07	0.14	0.04	0.07	0.04	0.20	**0.06**
62) NAPM new orders	0.03	0.02	0.18	0.10	0.04	0.11	0.23	**0.14**
71) SP500: dividend	0.05	0.08	0.30	0.03	0.08	0.03	0.14	**0.02**
120) Consumer expec. (Mich.)	0.02	0.05	0.15	0.09	0.21	0.03	0.04	**0.05**

The upper panel illustrates the total fraction that the eight factors can explain of the forecast error variance at varying horizons. "Idio." means idiosyncratic variance. FFR means federal funds rate, which is the shock in focus. The lower table shows the 60-month ahead forecast error variance decomposition for key macroeconomic variables.

Figure 1.1: The panel information criterion IC_{p3} of Bai & Ng (2002). The criterion does not provide information about the number of lags in the VAR so the criterion as a function of the number of static factors, r, is calculated for a given number of lags. On top of each bar the number of factors is plotted. Eight factors seem to be a good choice when model parsimony is taken into account.

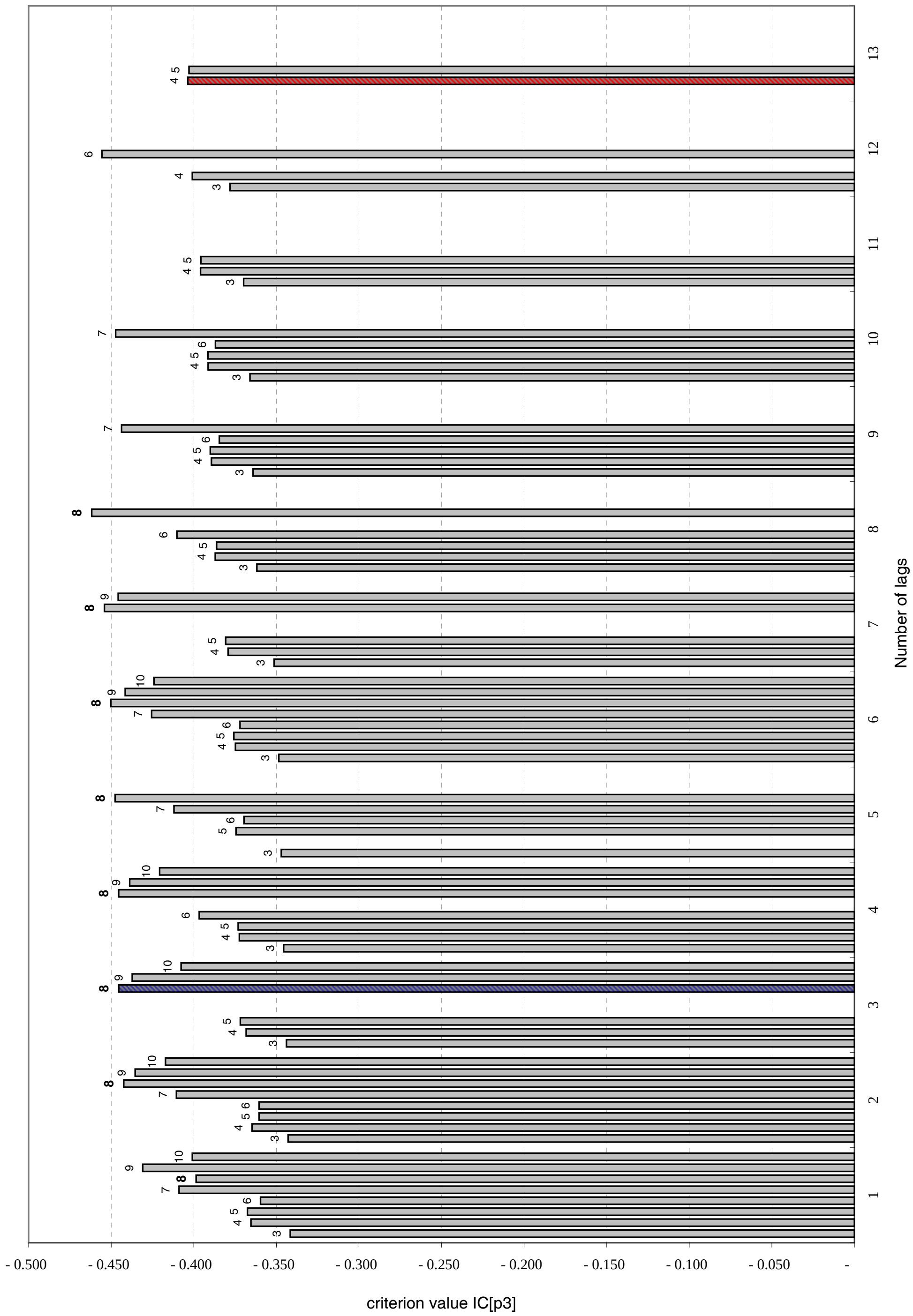


Figure 1.2: Adjusted average R^2 of all variables in the panel for all models.

For each FAVAR model with r factors and p lags in the VAR, the $\bar{R}^2$ is calculated. The number on top of each bar represents the number of lags in the VAR with r factors. Note how the incremental value of $\bar{R}^2$ diminishes as more factors are added.

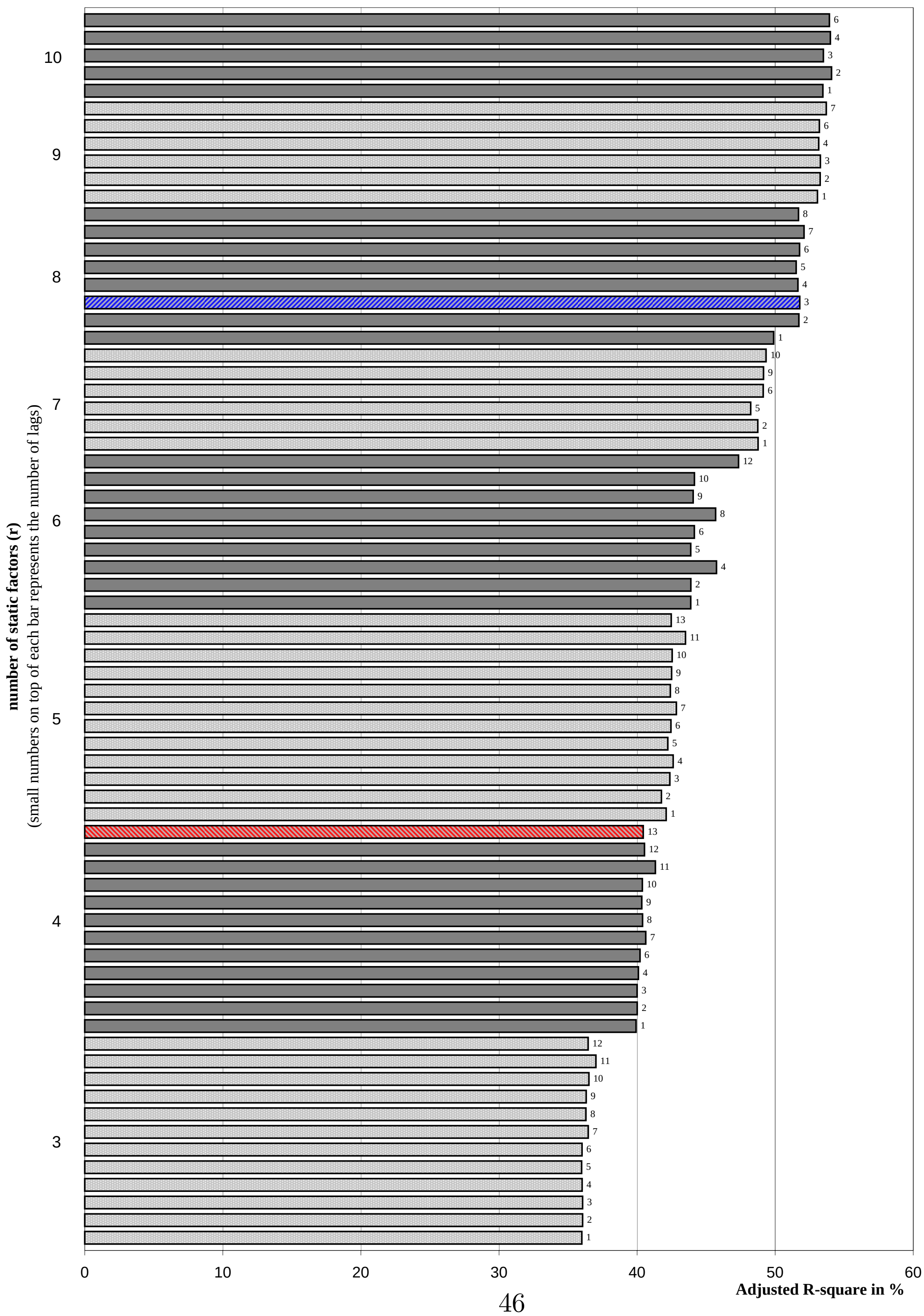


Figure 1.3: R^2 for each variable: Mean of the "15 best", "15 medium" and "15 worst" performing models.

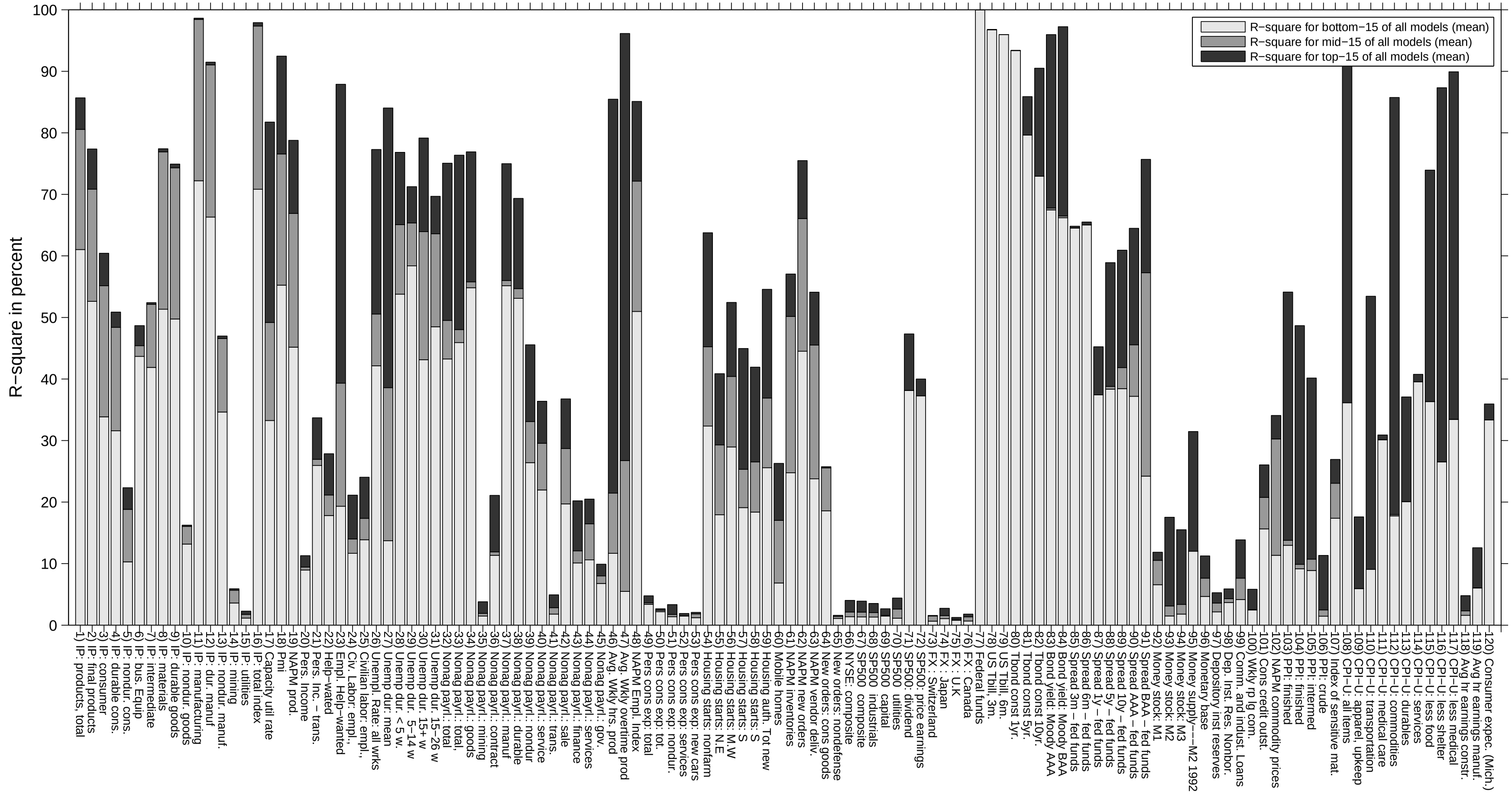
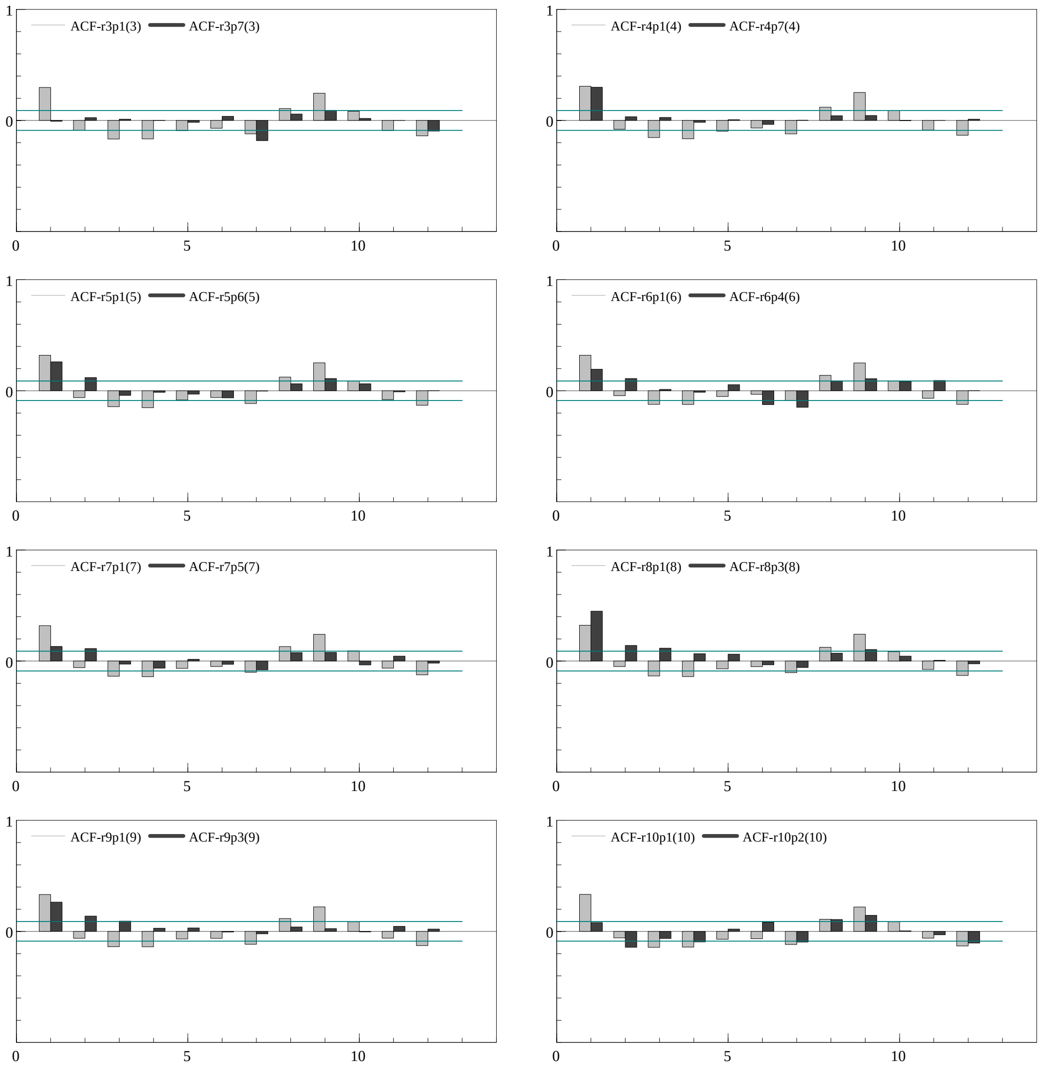


Figure 1.4: Autocorrelation functions for the preferred specification versus a VAR(1).



The smoothed residual autocorrelation functions for the preferred specifications for FAVARs with 3,4,...,10 factors versus their VAR(1) counterpart are plotted to emphasize that VAR(1) dynamics are not sufficient for whiteness in the monetary policy factor residuals. Unreported results show that there is virtually no difference in the autocorrelation function for r4p7(4) versus BBE-EM version r4p13(4).

Figure 1.5: R^2 **for the preferred model versus the BBE-EM and Bernanke, Boivin and Eliaz.**

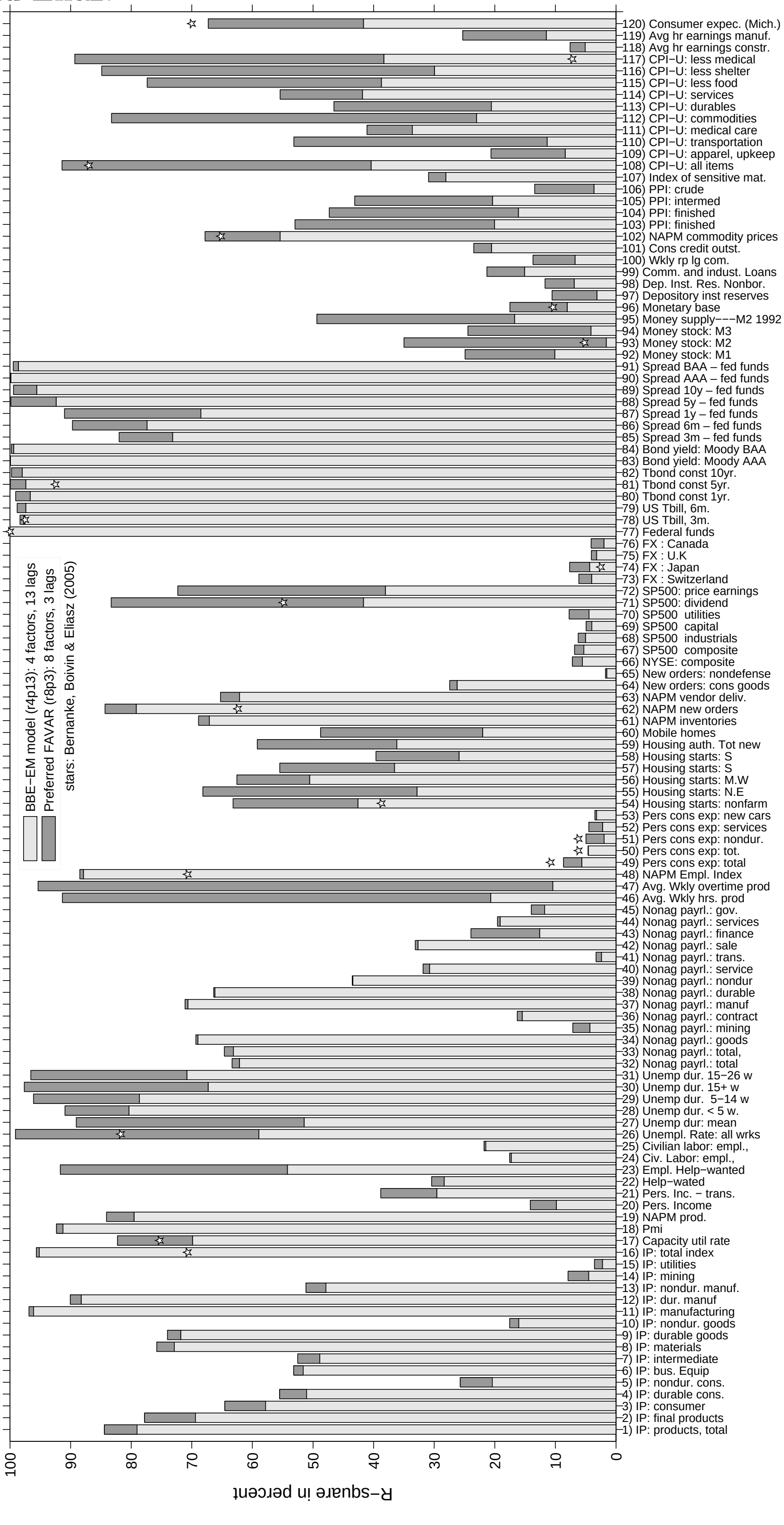


Figure 1.6: The time series of factors 1-4 from the preferred model versus related observed variables.

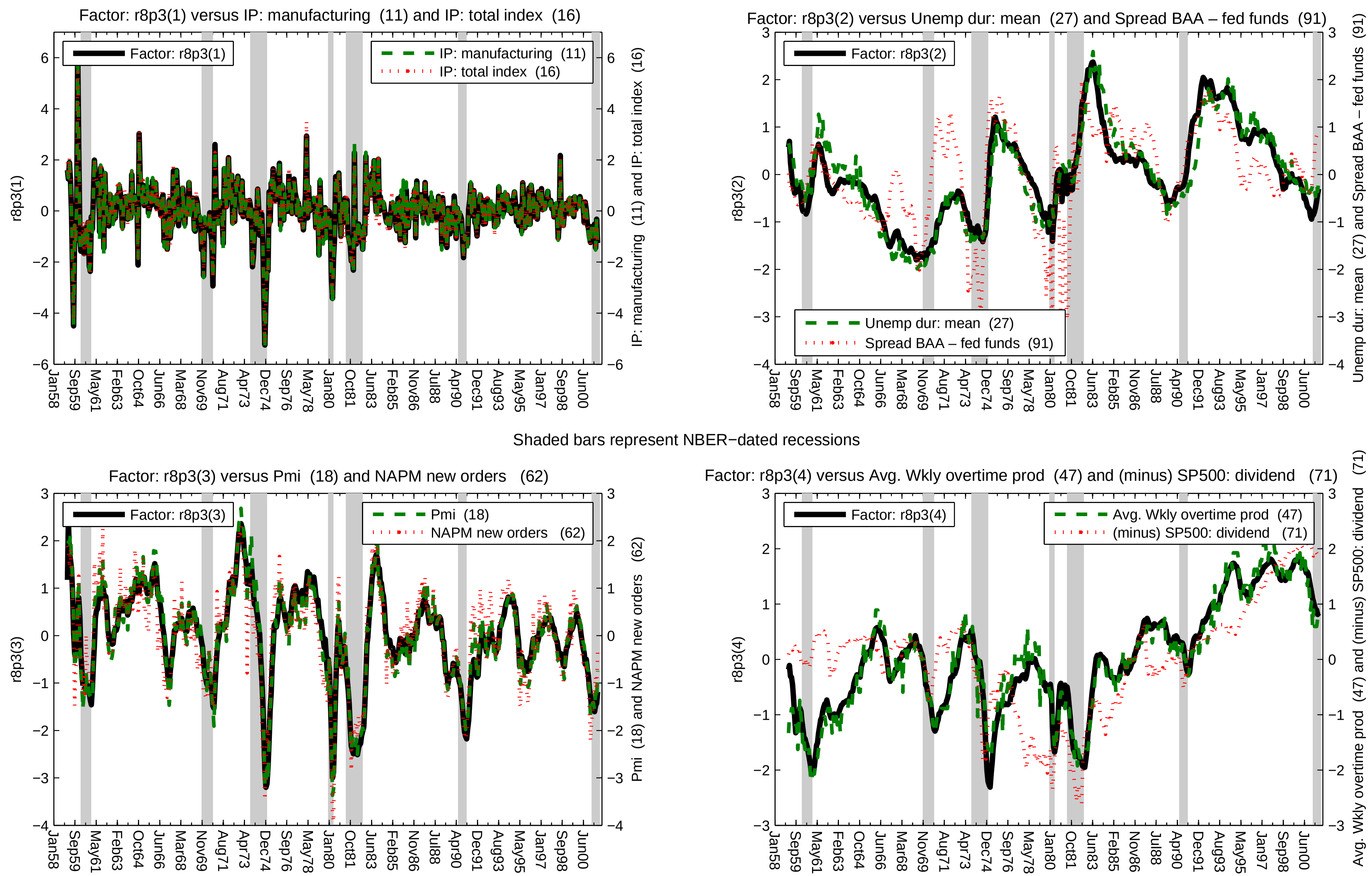


Figure 1.7: The time series of factors 5-8 from the preferred model versus related observed variables.

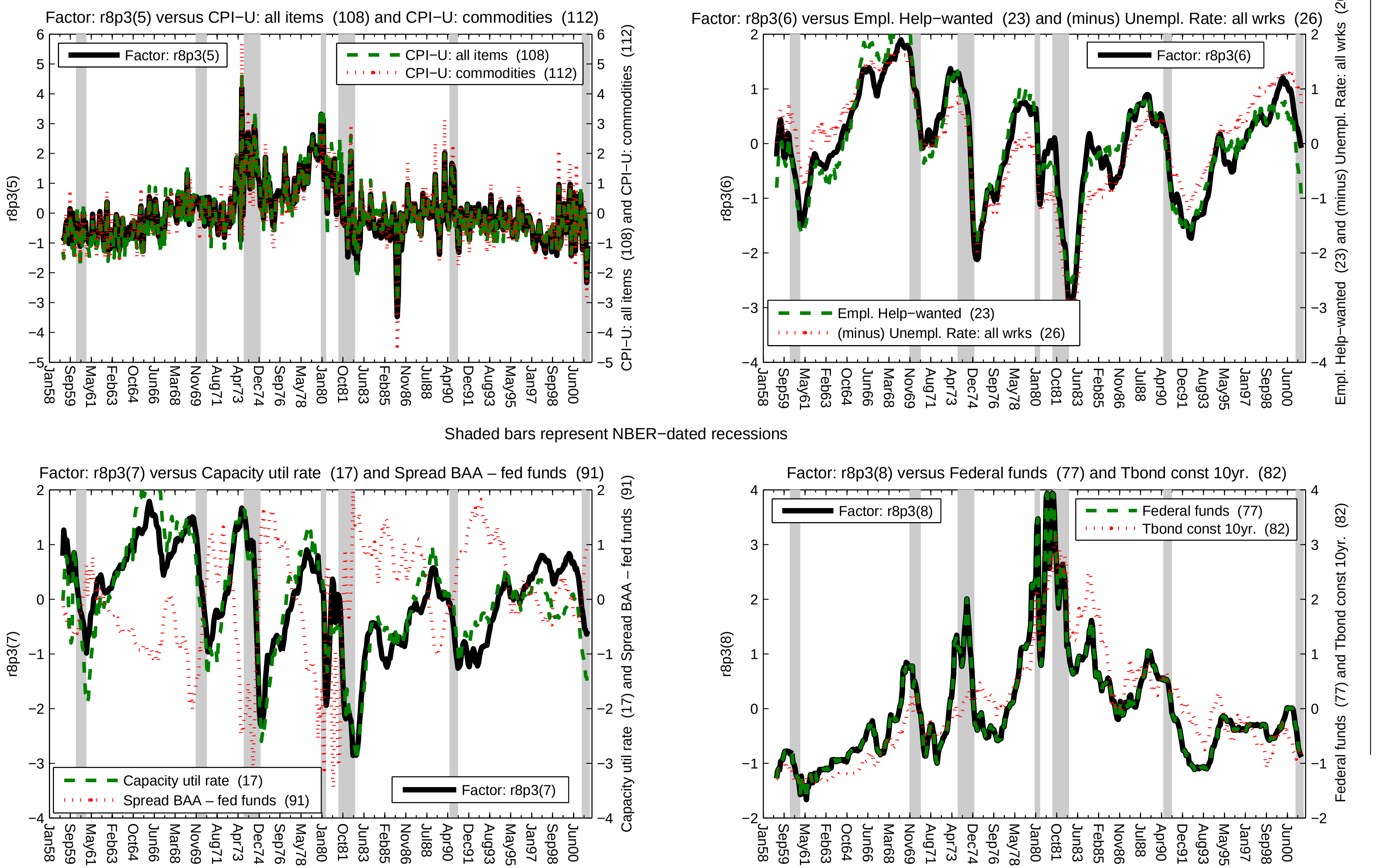
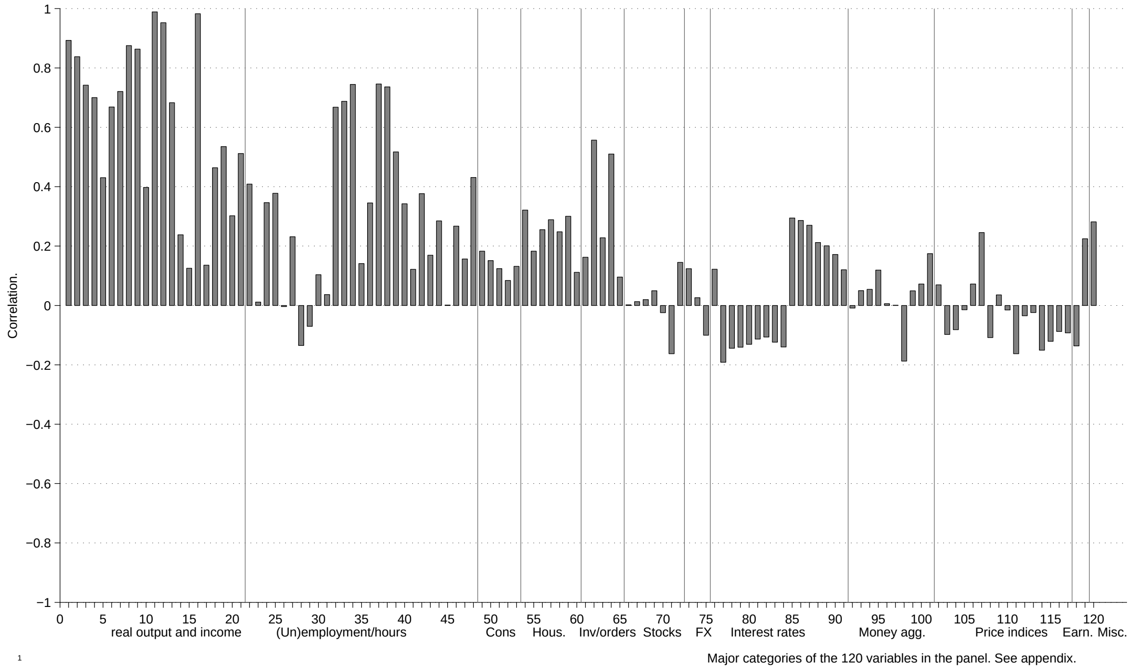
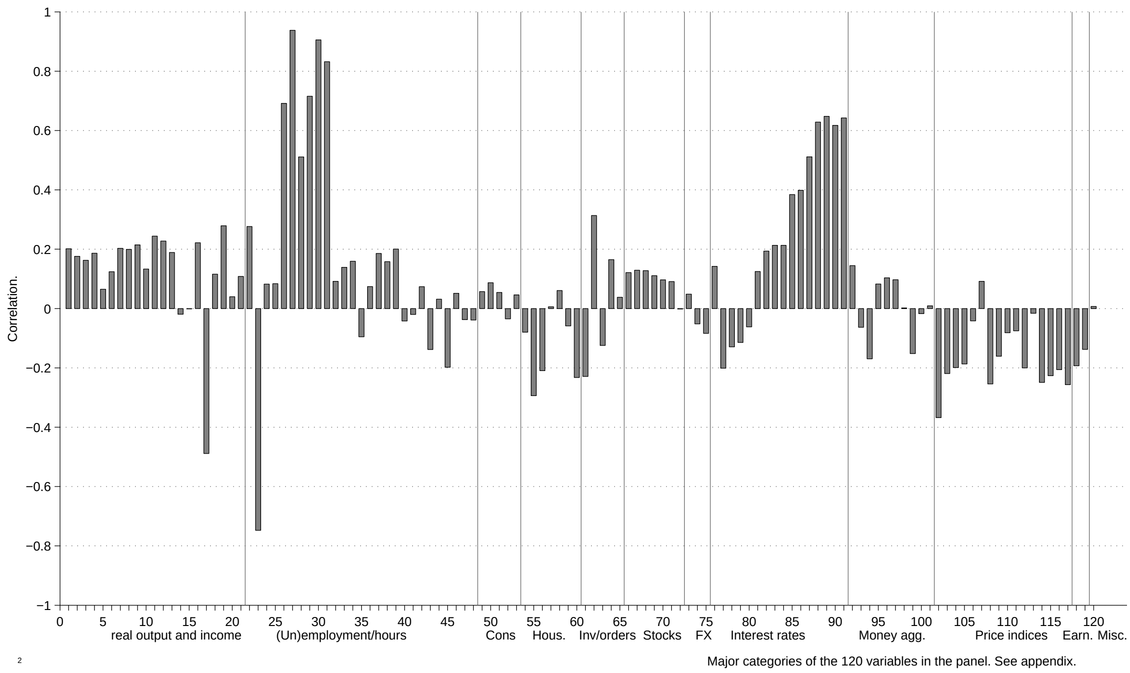


Figure 1.8: "Industrial production factor" and "unemployment factor".

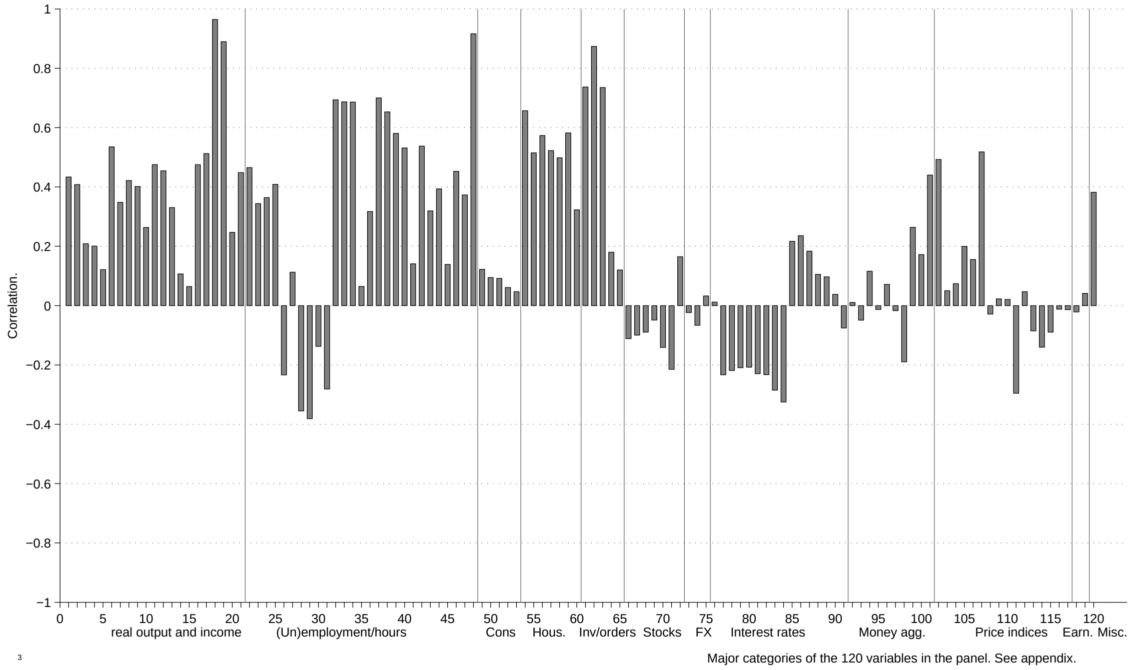


Correlation between factor 1 and each of the variables as a function of (Φ,Q,Λ).

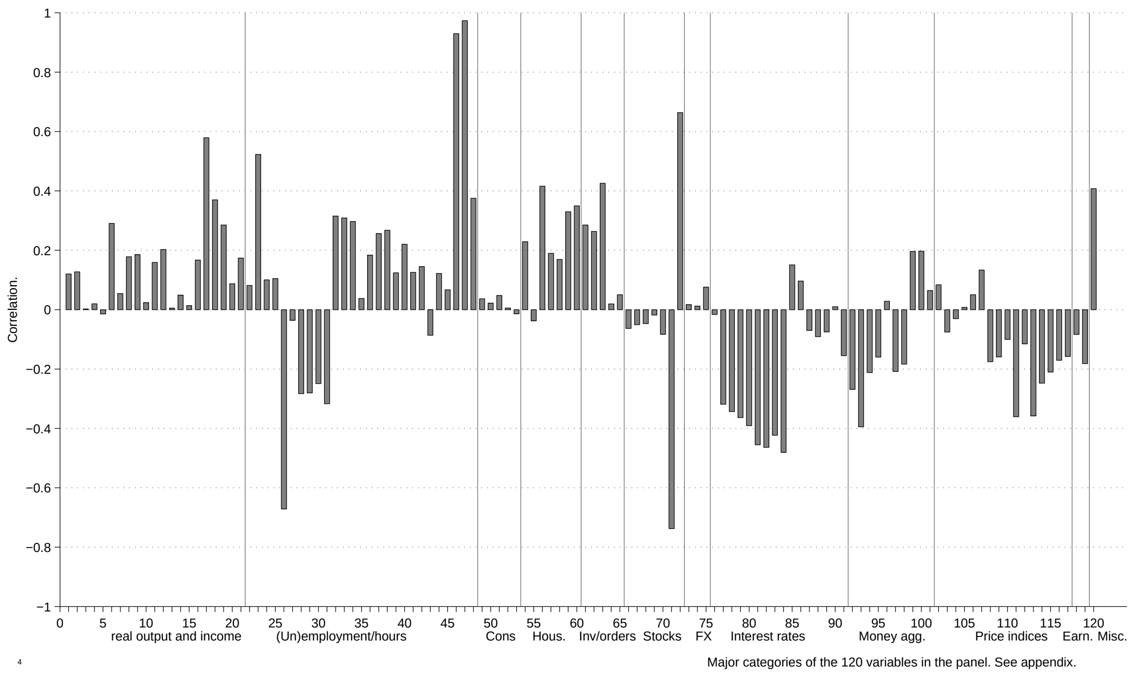


Correlation between factor 2 and each of the variables as a function of (Φ,Q,Λ).

Figure 1.9: "NAPM factor" and "(overtime)hours factor".

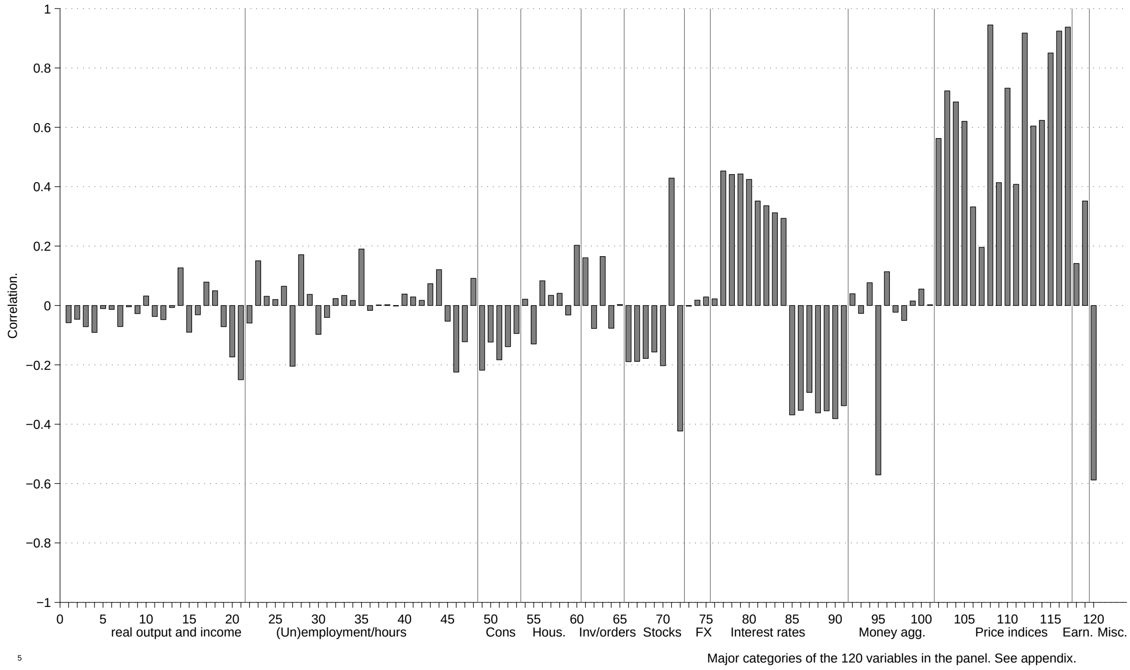


Correlation between factor 3 and each of the variables as a function of (Φ, Q, Λ).

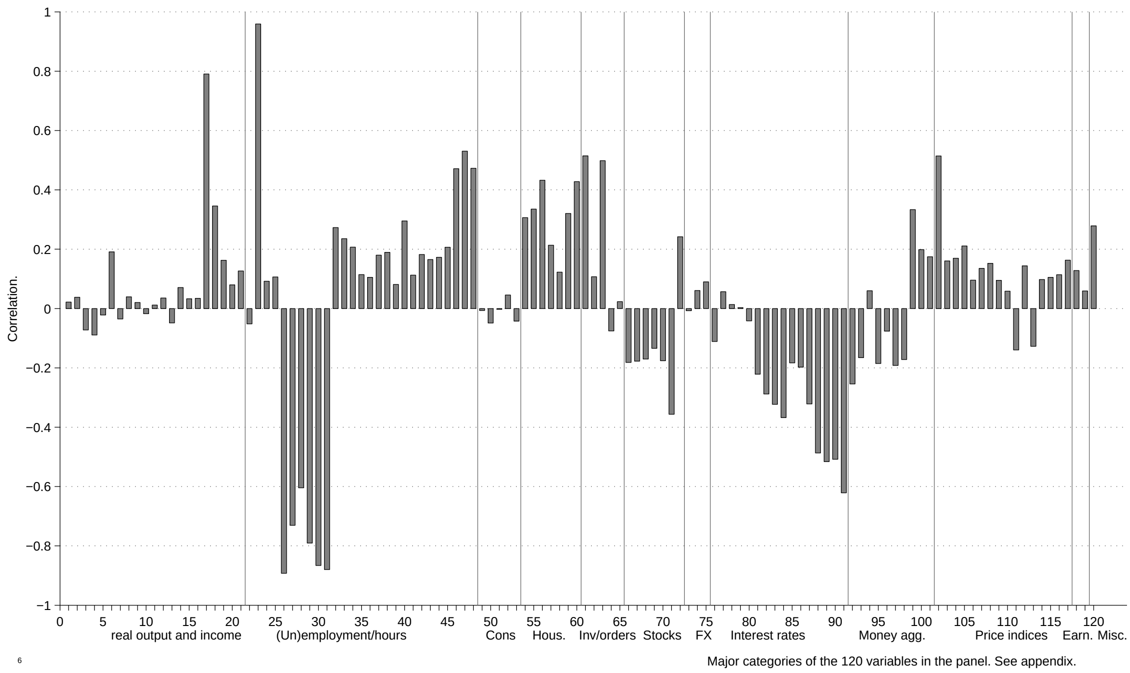


Correlation between factor 4 and each of the variables as a function of (Φ, Q, Λ).

Figure 1.10: "Inflation factor" and "employment factor".

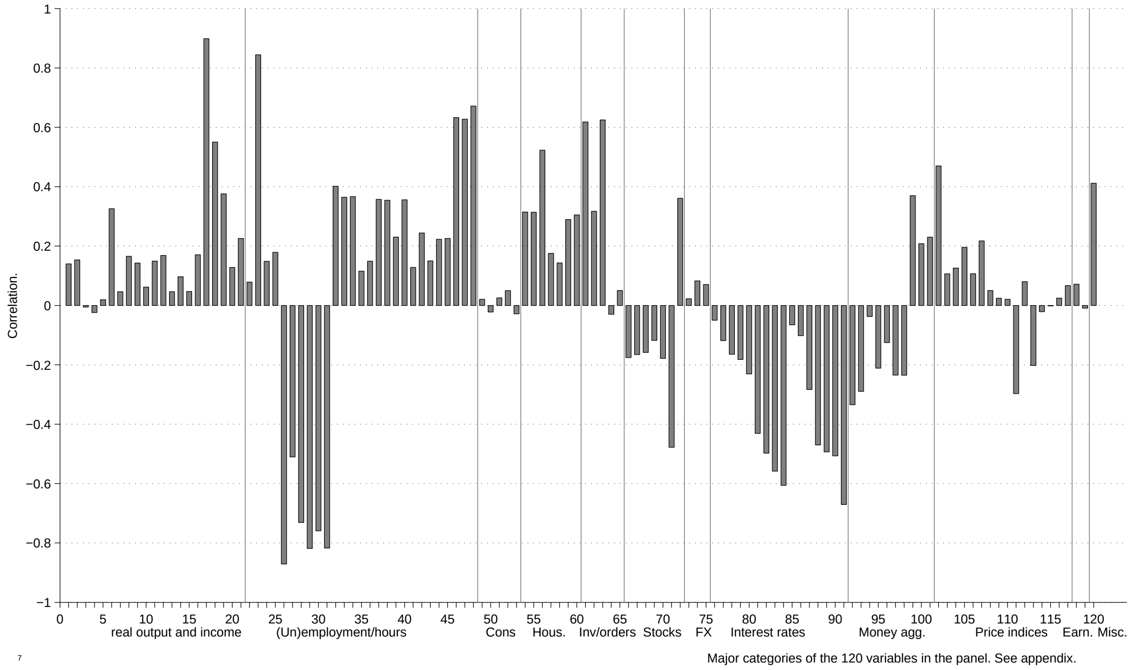


Correlation between factor 5 and each of the variables as a function of (Φ, Q, Λ).

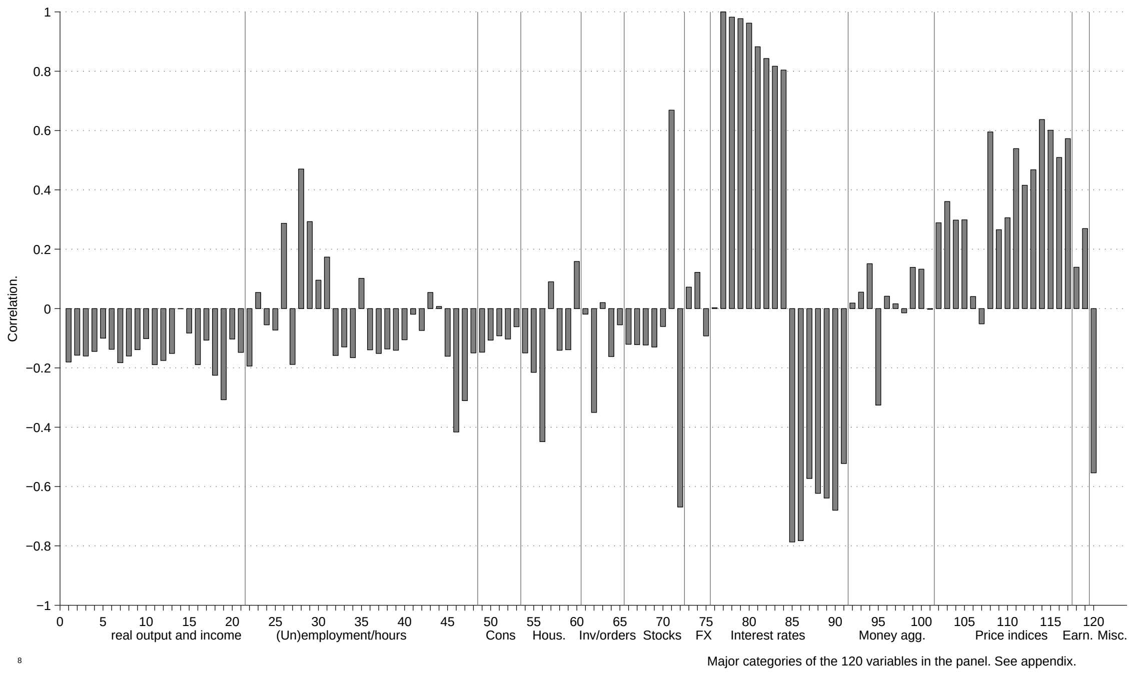


Correlation between factor 6 and each of the variables as a function of (Φ, Q, Λ).

Figure 1.11: "Capacity utilization factor" and "Monetary policy factor".

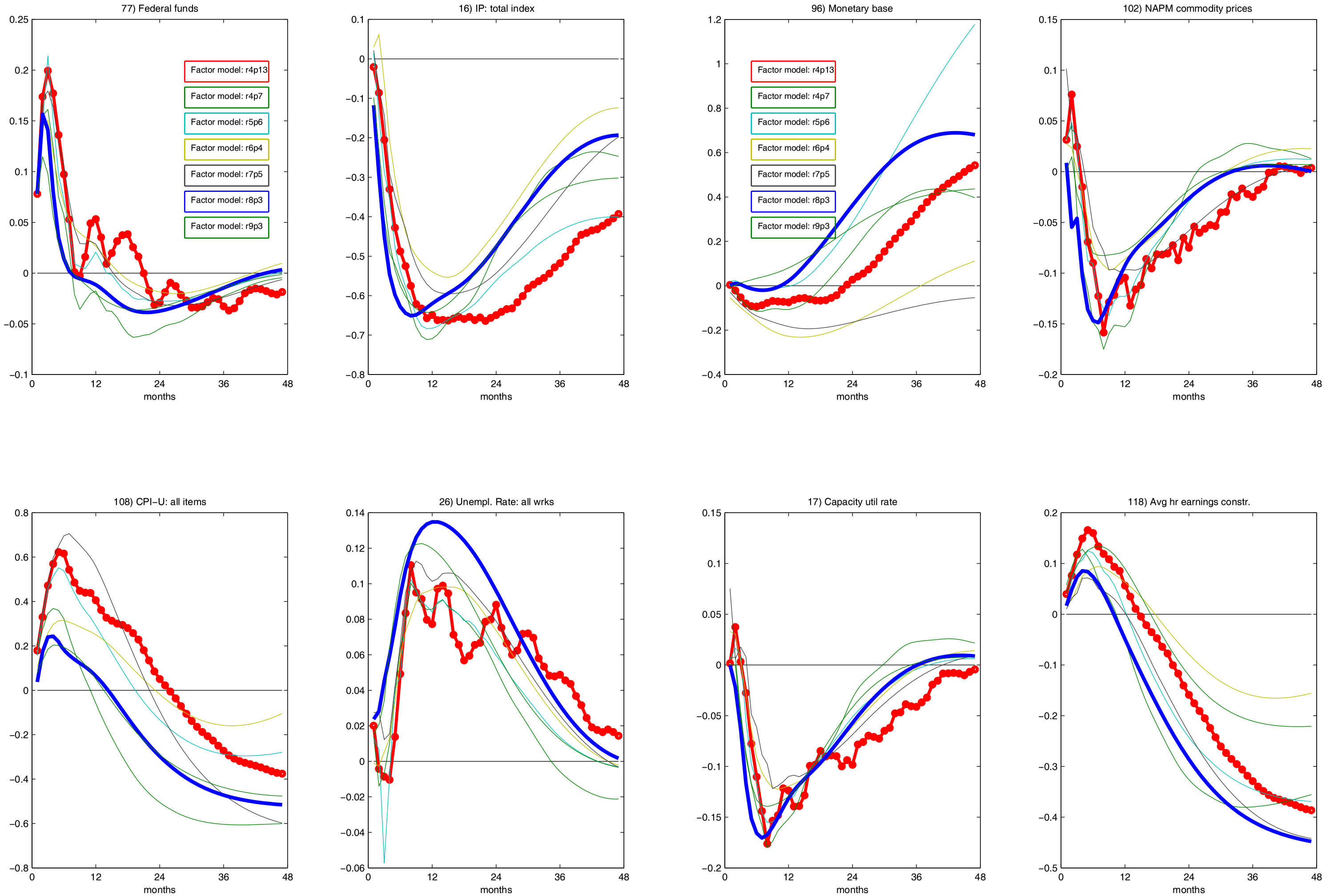


Correlation between factor 7 and each of the variables as a function of (Φ, Q, Λ).



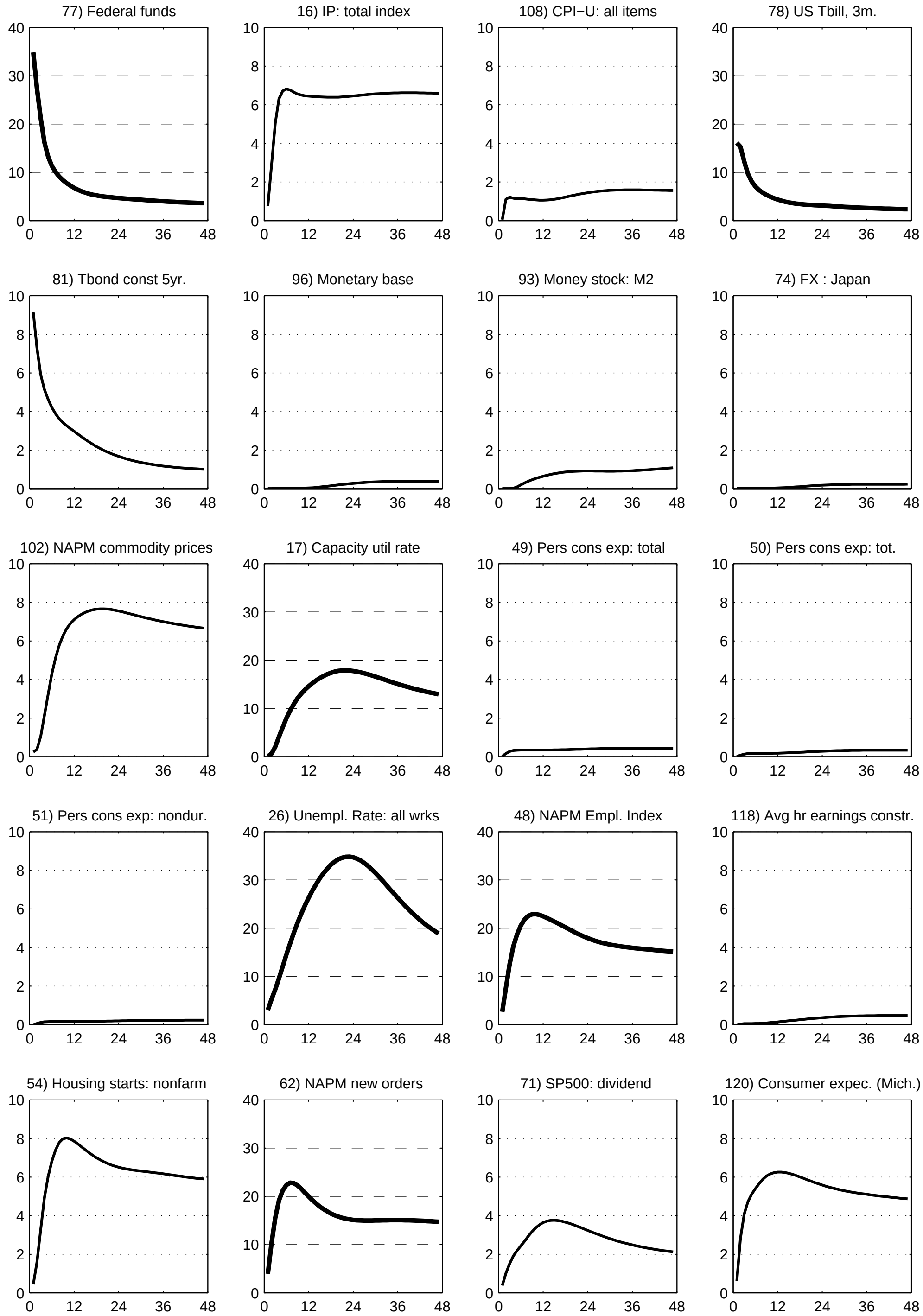
Correlation between factor 8 and each of the variables as a function of (Φ, Q, Λ).

Figure 1.12: Impulse responses in standard deviations to a 25 basis point monetary policy shock.



The figure illustrates the impulse responses of key macroeconomic variables for the best model across different numbers of factors included in the FAVAR.

Figure 1.13: Contribution of the monetary policy shock to forecast error variance decomposition.



The figure plots the forecast error variance decomposition along the forecast horizon (the horizontal axis). Dashed gridlines indicate a larger scale compared to the dotted gridlines.

CHAPTER 2

Identification of Macroeconomic Factors in Large Panels

Lasse Bork,[†] Hans Dewachter[‡] and Romain Houssa.[§]

January 2010

[†]Finance Research Group, Department of Business Studies, Aarhus School of Business, University of Aarhus, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. Affiliated with CREATES at the University of Aarhus, a research center funded by the Danish National Research Foundation. Email: lbo@lassebork.dk.

[‡]CES, University of Leuven, RSM Rotterdam and CESIFO. Hans Dewachter acknowledges financial support from FWO grant number G.0626.07.

[§]CRED and CEREFIM, University of Namur, CES, University of Leuven.

Abstract*

This paper presents a dynamic factor model in which the extracted factors and shocks are given a clear economic interpretation. The economic interpretation of the *factors* is obtained by means of a set of over-identifying loading restrictions, while the structural *shocks* are estimated following standard practices in the SVAR literature. Estimators based on the EM algorithm are developed. We apply this framework to a large panel of US monthly macroeconomic series. In particular, we identify nine macroeconomic factors and discuss the economic impact of monetary policy stocks. The results are theoretically plausible and in line with other findings in the literature.

JEL classifications: E3, E43, C51, E52, C33
Keywords: Monetary policy, Business Cycles, Factor Models, EM Algorithm.

*We thank Tom Engsted, Priscilla Toffano and participants at the Danish PhD Workshop in Finance (2008) for helpful comments and suggestions.

2.1 Introduction

In recent years, factor models have become a standard tool in applied macroeconomics and finance. They have been applied in empirical macroeconomics for predictions (Bernanke & Boivin (2003), Forni et al. (2005), and Stock & Watson (2002*a*,*b*)); for structural analysis (Forni & Reichlin (1998), Forni et al. (2008), Giannone et al. (2004, 2002), Houssa (2008*a*), Bernanke et al. (2005) and Stock & Watson (2005)); and for constructing business cycle indicators (Forni et al. (2001), Kose et al. (2003), Houssa (2008*b*), and Otrok & Whiteman (1998)). Applications of factor models in finance include the arbitrage pricing theory (Chamberlain & Rothschild (1983) and Ingersoll (1984)); the measurement of risks (Campbell et al. (1997), ch. 2); the estimation of the conditional risk-return relation in Ludvigson & Ng (2007); bond market applications (Mönch (2008), Ludvigson & Ng (2008) and Diebold et al. (2008)); and the prediction of the volatility of asset returns (Alessi et al. (2007)).

The increasing popularity of factor models can be explained by two model features. First, factor models distinguish measurement errors and other idiosyncratic (series-specific) disturbances from structural shocks. As such, dynamic factor models have a direct mapping from observed data to their theoretical counterparts[1]. Second, large data sets are becoming increasingly available and classical multivariate regression models generally perform poorly in fitting them. By contrast, Dynamic Factor Models (DFM), exploiting the dynamic and cross-sectional structure of the data set, allow a large panel to be analyzed through a (small) set of underlying extracted factors. Moreover, various estimation techniques have been developed recently to analyze factor models in large panels. For instance, Stock & Watson (2002*a*,*b*) and Forni et al. (2000) proposed a non-parametric estimation approach based on principal components. The former uses the time domain method while the latter suggests a frequency domain estimation technique. In a related literature, Otrok & Whiteman (1998) and Kim & Nelson (1999) propose a Bayesian estimation technique whereas Doz et al. (2006, 2007) and Jungbacker & Koopman (2008) use an estimation approach based on the EM algorithm.

While these studies have provided important contributions to the literature on factor models, some identification issues remain, however. In particular, it is often

[1] Typically, these theoretical counterparts are defined within a DSGE model (see for example Altug (1989), Sargent (1989) and recently Boivin & Giannoni (2006)).

the case that the (static) factors estimated in applied work do not necessarily have a well-defined and unambiguous economic interpretation[2]. A standard procedure amounts to inferring the economic interpretation of the factors from the dominant factor loadings. This approach, however, neglects the non-dominant (but possibly significant) loadings and hence does not necessarily generate unambiguous and well-defined interpretations of the factors.

In this paper we address this identification problem by using a procedure that *imposes* a specific and well-defined interpretation on the static factors. The economic interpretation of the extracted static factors is based on a set of overidentifying restrictions on factor loadings. Furthermore, a set of standard exclusion restrictions on the impact matrix is used to identify the structural shocks. We employ the iterative maximum likelihood estimation approach as in Doz et al. (2006, 2007) and Jungbacker & Koopman (2008) which is an iterative maximum likelihood method.

We illustrate our approach by revisiting the large cross-section data analyzed in Bernanke et al. (2005). We aim at identifying and extracting from the data panel nine macroeconomic factors respectively related to inflation, unemployment, economic activity, consumption, state of the business cycle, residential investments, financial markets and monetary policy. Given the identification of these factors, we assess and analyze (as in Bernanke et al. (2005)) the impact of monetary policy shocks on a number of key observable through impulse response analysis and variance decompositions.

Our paper is closely related to a number of recent studies. Boivin et al. (2009) and Reis & Watson (2008) impose loadings restrictions to identify a measure of pure inflation for the US economy. In the same way, Forni & Reichlin (2001) and Kose et al. (2003) use loading restrictions to differentiate between world, regional and country factors. Finally, Boivin & Giannoni (2006) employ loading restrictions to estimate the theoretical concepts of variables defined in DSGE model. The main difference between these studies and ours is that we employ the EM algorithm to derive closed form solutions for (linearly) restricted factor loadings. As such, we

[2] *Static* factors are related to the variance-covariance matrix of the data while *dynamic* factors capture the property of their spectral density matrix. See Forni et al. (2000) for a literature review. Recent studies provide a structural interpretation to dynamic factors (shocks), see for example Giannone et al. (2004); Houssa (2008*a*) and Forni et al. (2008). The main difference between these studies and ours is that we identify (in economic and structural terms respectively) the static and dynamic factors.

can combine various loading restrictions allowing to obtain a clear macroeconomic interpretation of the extracted factors (see sections 2 and 3).

The remainder of the paper is organized as follows. First, the methodological approach is explained in Section 2.2. We introduce a dynamic factor model and discuss the identification restrictions. In addition, closed-form solutions for the parameter estimates, consistent with the identification schemes and using results from Shumway & Stoffer (1982) and Wu et al. (1996), are presented. An empirical illustration of the impact of US monetary policy shocks on the macroeconomic factors is provided in Section 2.3. Section 2.4 concludes.

2.2 Methodology

We first introduce the DFM. More details can be found in Forni et al. (2000) and Forni & Lippi (2001). Subsequently, we employ the quasi-maximum likelihood estimation approach as in Doz et al. (2006, 2007) and Jungbacker & Koopman (2008). We take this approach one step further by imposing (over-) identifying restrictions on the loadings and on the impulse response function (IRF). This allows a clear economic interpretation of the static factors and a structural identification of the shocks.

2.2.1 Dynamic Factor Model

Consider a panel of observable economic variables $y_{i,t}$, where i denotes the cross-section unit, $i = 1, ..., N$ while t refers to the time index, $t = 1, ..., T$. The panel of observed economic variables is transformed into stationary variables with zero mean and unit variance. These transformed variables are labeled by $x_{i,t}$. Dynamic factor models assume that a variable $x_{i,t}$ can be decomposed into two components, the *common component,* χ_{it}, and the *idiosyncratic component* ξ_{it}:

$$x_{it} = \chi_{it} + \xi_{it}. \tag{2.1}$$

Furthermore, in exact dynamic factor models it is assumed that the idiosyncratic and common components are uncorrelated at all leads and lags and across all variables, $E(\xi_{i,t}\chi_{j,s}) = 0, \ \forall \ s, t, i, j$. The common component is assumed to be driven by a small number r, $r << N$, of common factors $f_t = (f_{1t}, \ f_{2t}, \cdots, f_{rt})^\top$:

$$x_{it} = \lambda_i^\top f_t + \xi_{it}, \tag{2.2}$$

where λ_i is a $r \times 1$ vector of factor loadings measuring the exposure of $x_{i,t}$ to the factors f_t. On the other hand, the idiosyncratic component is driven by variable-specific noises. Stacking equation (2.2) over all cross-section units, $x_{i,t}$, $i = 1, ..., N$, gives

$$X_t = \lambda f_t + \xi_t, \tag{2.3}$$

where $X_t = (x_{1t}, \ldots, x_{Nt})^\top$, $\xi_t = (\xi_{1t}, \ldots, \xi_{Nt})^\top$, and λ is a $N \times r$ matrix of factor loadings, $\lambda = (\lambda_1, ..., \lambda_N)^\top$. Equation (2.3) is called a *static* factor model. "Static" stands for the fact that the observed variables only load contemporaneously on the factors.

To close the model, factor dynamics have to be specified. We assume that the r-dimensional vector of common factors f_t has a VAR(p) representation

$$\phi(L) f_t = \Upsilon \eta_t, \tag{2.4}$$

where $\phi(L) = I - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$, with ϕ_j denoting a $r \times r$ matrix of autoregressive coefficients ($j = 1, \ldots, p$). Moreover, given the stationarity of the transformed panel, we impose that the roots of $\det(\phi(L))$ are outside the complex unit circle. The q-dimensional vector of dynamic factor innovations is denoted η_t and Υ represents a $r \times q$ transformation matrix. As in Doz et al. (2006) we make additional distributional assumptions: $\eta_t \sim i.i.d\ N(0, Q)$ and $\xi_t \sim i.i.d\ N(0, R)$, with Q and R denoting (semi)positive definite matrices[3].

Using equations (2.3) and (2.4), the model can be summarized in first order, with a $rp \times 1$ state vector F_t, $F_t = (f_t^\top, ..., f_{t-p+1}^\top)^\top$, by the measurement equation:

$$X_t = \Lambda F_t + \xi_t, \tag{2.5}$$

[3]Note that, by assuming $i.i.d$ idiosyncratic components, (2.3)-(2.4) define an *exact* dynamic factor model. This is certainly a strong assumption, particularly in the case of large panel data sets where cross-sectional and serial correlations are expected to be found. As such, (2.3)-(2.4) represent a misspecified model. However, Doz et al. (2006) show that, for large N and T, the exact factor model estimators are consistent quasi-maximum likelihood estimators for the *approximate* factor model.

and the transition equation:

$$F_t = \Phi F_{t-1} + VSu_t, \tag{2.6}$$

where Λ is the $N \times rp$ matrix loading, implied by λ, Φ is the $rp \times rp$ companion matrix corresponding to the VAR(p) in (2.4), $V = \left(\Upsilon^\top, 0^\top_{r(p-1)\times q}\right)^\top$, and u_t represents the structural shocks that are identified through the matrix S (see section 2.2.2 below). Inverting the VAR in (2.6) and substituting F_t in (2.5) gives

$$X_t = B(L)u_t + \xi_t, \tag{2.7}$$

where $B(L) = \Lambda(I - \Phi L)^{-1}VS$, represents the IRF to u_t.

The state-space system, defined by equations (2.5) and (2.6), is not uniquely identified. We address the econometric identification as well as the economic interpretation of the static factors in section 2.2.2 Finally, the identification of the structural shocks u_t is discussed in section 2.2.2.

2.2.2 Economic interpretation

Economic interpretation of the factors and shocks requires additional identification restrictions. We use two types of restrictions:

1. Loading restrictions allowing for a clear macroeconomic interpretation of the (static) factors.

   The section headed "Economic factors" details this approach.

2. Restrictions on the impact matrix identifying the structural shocks.

   The section headed "Structural shocks" details this approach.

Economic factors

We impose a set of restrictions on the loading matrix Λ in (2.5) and denote the restricted loading matrix by Λ^*. The linear loading restrictions take the following general form:

$$H_\Lambda vec(\Lambda^*) = \kappa_\Lambda. \tag{2.8}$$

where κ_Λ refers to a $\ell \times 1$ vector of ℓ linear combinations of restrictions of factor loadings defined by H_Λ, $H_\Lambda \in \mathbb{R}^{\ell \times Nr}$.

We use three types of loading restrictions, depending on the information content of the observables. In particular, economic identification is achieved by means of *(i)* unbiasedness restrictions *(ii)* one-to-one restrictions or *(iii)* exclusion restrictions.[4]

The *unbiasedness restriction* implies that observable x_j is an unbiased and direct information variable for factor $f_l, l = 1, 2, \ldots, r,$:

$$\Lambda^*_{j,l} = 1, \ \Lambda^*_{j,k\neq l} = 0. \tag{2.9}$$

This type of restrictions is used on observables that are assumed to be a direct measure (up to some measurement error) of the underlying factor. For instance, our empirical application assumes that the observable "*CPI-U: All items*" inflation is a direct measure for the inflation factor. As such, the unbiasedness restrictions imply a unit loading of "*CPI-U: All items*" inflation on the inflation factor and zero loadings on all other factors. Note that these unbiasedness restrictions allow for the *econometric identification* of the DFM as the static factors are now uniquely defined. Our identification approach is an application of Proposition 2 in Geweke & Singleton (1981) and further discussed in Bork (2008).

The *one-to-one restriction* implies a one-to-one link between an observable and a factor. Unlike unbiasedness restrictions, we allow other common factors to affect the observable as well, i.e. we do not impose $\Lambda^*_{j,k\neq l} = 0$. Formally, one-to one restrictions between observable x_j and factor l are ensured by imposing:

$$\Lambda^*_{j,l} = 1. \tag{2.10}$$

Finally, contemporaneous *exclusion restrictions*, i.e. the case where variable x_j is (contemporaneously) not related to the factor f_l, take the form of:

$$\Lambda^*_{j,l} = 0. \tag{2.11}$$

Note that this identification scheme formalizes and extends the standard informal identification procedures used in the literature. The standard approach identifies the factors from the principal factor loadings of the economic variables, disregarding the smaller loadings. Our identification procedure formalizes this approach by *(i)* imposing exclusion restrictions on the non-informative variables, which ensures

[4] To conform to the static factor structure of the model, all loadings on lagged factors are set to zero.

that only information of relevant variables is incorporated in the factor and *(ii)* facilitating interpretation of the factors by means of the unbiasedness or one-to-one restrictions imposing a direct mapping between the observables and the static factor.

The economic interpretation of the factors is obtained by imposing at least one unbiasedness or a one-to-one restriction per factor. However, while exclusion and unbiasedness restrictions exclude some observables from the information set of a factor, we allow for feedback effects across factors. Specifically, through the VAR specification in (2.6) we allow for dynamic interactions among factors. As such, factors can be correlated and structural shocks are eventually transmitted across all observables.

Structural shocks

In equation (2.7), structural shocks are identified. We follow the standard identification procedure in the SVAR literature by choosing an appropriate matrix S such that the implied restricted IRF, $B(L)^*$, has an economic justification. For instance, the Blanchard & Quah (1989) long-run restrictions can be obtained by choosing S such that appropriate elements of $B(1)^*$ are equal zero. Sign restrictions, recently introduced by Uhlig (2005), can also be fulfilled by choosing S such that the time path of some elements of $B(L)^*$ have an appropriate sign. Popular sign restrictions include the fact that prices cannot increase following a negative demand shock. Finally, structural identification can be obtained by imposing the Sims (1980)'s triangular representation on the matrix S. This is the approach followed in our empirical application in section 2.3.

We first impose that the number of static factors equals the number of dynamic factors, i.e. $q = r$. This generates a structural shock to each of the static factors. Thereafter, we use the exclusion restrictions implied by the Cholesky decomposition of $Q = SS'$, with S lower triangular. The structural interpretation of the shocks is then implied by the ordering of the static factors and discussed in more details in section 2.3.

2.2.3 Estimation: the EM algorithm

Given the latent nature of the static factors, a standard EM algorithm is used to estimate the parameters and to extract the implied factors. Denote by $\Theta^* = \{\Lambda^*, R, \Phi, Q\}$ the set of parameters to be estimated with Λ^* satisfying the set of identification restrictions listed in equation (2.8). Conditional on the estimates of the factors, $\hat{F}$ (and matrices measuring uncertainty $\hat{P}$), the elements of Θ^* can be estimated by (Maximization step):

$$
\begin{aligned}
\operatorname{vec}(\Lambda^*) &= \operatorname{vec}(DC^{-1}) \\
&\quad + (C^{-1} \otimes R)\, H_\Lambda^\top \left[H_\Lambda (C^{-1} \otimes R)\, H_\Lambda^\top \right]^{-1} \\
&\quad \times \left\{ \kappa_\Lambda - H_\Lambda \operatorname{vec}(DC^{-1}) \right\}, \\
R &= \tfrac{1}{T} \left(E - DC^{-1} D^\top \right), \\
\operatorname{vec}(\Phi) &= \operatorname{vec}(BA^{-1}), \\
\Omega &= VQV^\top = \tfrac{1}{T} \left[C - BA^{-1} B^\top \right],
\end{aligned}
\tag{2.12}
$$

where the estimator for Λ^* follows from a straightforward extension of Wu et al. (1996). Appendix B.1 offers a derivation of Λ^*.

Conditional on the estimated parameters, Θ, the latent factors can be extracted by means of the Kalman smoother and the required moments can be computed (Expectation step). In particular, the following expectations are generated:

$$
\begin{aligned}
A &= \textstyle\sum_{t=1}^{T} \left(\hat{P}_{t-1|T} + \hat{F}_{t-1|T} \hat{F}_{t-1|T}^\top \right), \\
B &= \textstyle\sum_{t=1}^{T} \left(\hat{F}_{t|T} \hat{F}_{t-1|T}^\top + \hat{P}_{\{t,t-1\}|T} \right), \\
C &= \textstyle\sum_{t=1}^{T} \left(\hat{F}_{t|T} \hat{F}_{t|T}^\top + \hat{P}_{t|T} \right), \\
D &= \textstyle\sum_{t=1}^{T} X_t \hat{F}_{t|T}^\top, \\
E &= \textstyle\sum_{t=1}^{T} X_t X_t^\top
\end{aligned}
\tag{2.13}
$$

with:

$$
\begin{aligned}
\hat{F}_{t|T} &= E\left[F_t \mid \mathcal{X}_T \right], \\
\hat{P}_{t|T} &= E\left[(F_t - \hat{F}_{t|T})(F_t - \hat{F}_{t|T})^\top \mid \mathcal{X}_T \right], \\
\hat{P}_{\{t,t-1\}|T} &= E\left[(F_t - \hat{F}_{t|T})(F_{t-1} - \hat{F}_{t-1|T})^\top \mid \mathcal{X}_T \right],
\end{aligned}
\tag{2.14}
$$

where $E[\cdot \mid \mathcal{X}_T]$ denotes the expectations operator conditional on the information set $\mathcal{X}_T = \{X_1, \ldots, X_t\}$ as implied by the Kalman smoother (as a function of Θ). See for instance de Jong & Mackinnon (1988) or de Jong (1989). We iterate sequentially over the M-step in equation (2.12) and the E-step in equation (2.13) until convergence of the likelihood starting from different sets of initial values.[5]

In our empirical application discussed in section 2.3 the unrestricted model involves $1,614$ parameters to be estimated. This is computationally feasible with the EM algorithm method. Doz et al. (2006) suggest to initialize the Kalman filter by the parameters implied by principal components and then filter the factors. However, principal component analysis results in orthogonal factors and we prefer correlated factors[6]. Consequently, we suggest entertaining an oblique rotation of the orthogonal factors, which is a common tool in confirmatory factor analysis and described in Lawley & Maxwell (1971). This approach does not change the initial fit but rotates the factors towards a target loading matrix which we choose to be the exactly identifying loading restrictions. The result is a set of correlated factors from which a set of implied initial parameters[7] consistent with the identifying loading restrictions can be derived.

2.3 Empirical Application

We illustrate our procedure by revisiting the large data panel analyzed in Bernanke et al. (2005)[8]. This data set includes 120 monthly time series covering a large part of the US economy over the period 1959:1 to 2001:8[9].

[5] We define convergence using a relative tolerance of 10^{-8} for the log-likelihood.

[6] The Geweke & Singleton (1981) identification scheme allows the factors to be correlated which is relevant if any macroeconomic interpretation is going to be attached to these factors.

[7] We experimented with many different sets of starting values in order to address the sensitivity of the EM algorithm to starting values. In one of the experiments we imposed very weak priors on the initial parameter estimates $\Theta^* = \{\Lambda^*, R, \Phi, Q\}$. In particular, the loading matrix was filled with with zeros except for the exactly identifying unit restrictions as explained in equation (2.9). In another experiment, principal component analysis (PCA) on the panel X generated a set of orthogonal factors from which a set of starting values of Θ can be derived. We also experimented with PCA of r subsets of the dataset where the principal component of each subset represents an initial estimate of one of the r factors. In the end, we prefer the oblique factor rotation as this approach results in improved likelihood values and statistical fit compared to the other approaches.

[8] We thank Jean Boivin for kindly making the data set available on his website, HEC-Montréal, Canada.

[9] The data are already transformed by Bernanke et al. (2005) to reach stationarity; see Bernanke et al. (2005) for details on the data set and on the transformation which results in a sample size of $T = 511$. The data transformation decisions are similar to Stock & Watson (2002*b*) and based

The focus of our empirical analysis is to extract a number of factors with an unambiguous (macro) economic interpretation. Moreover, we analyze the economic impact of monetary policy shocks on the US economy.

We first discuss the identification of the factors in section 2.3.1. Then in section 2.3.2 the statistical and economic significance of the over-identifying restrictions are evaluated by means of standard information criteria, R-squared and a likelihood ratio test. The results are then presented in section 2.3.2 and the empirical monetary policy analysis undertaken in section 2.3.2.

2.3.1 Identification

The identification of the factors and the structural monetary policy shocks are now discussed in an empirical setting. Firstly, the number of factors is discussed in section 2.3.1. Subsequently, we follow the structure represented by the two numbered items in 2.2.2 by a discussion of 1) economic interpretation in of the identified factors in section 2.3.1 and 2) structural monetary policy shocks in section 2.3.1. Thus, the economic identification in 1) and the structural identification in 2) are discussed separately but it should be mentioned that the *order* in which the factors enters into the VAR in (2.6) is influenced by the recursive structural identification of the shocks.

Determining the number of factors

An important choice in factor analysis concerns the unknown number of static factors r that span the factor space. Bai & Ng (2007, 2002), Stock & Watson (2002*b*) and Hallin & Liska (2007) represent important contributions to the literature on the determination of the number of factors. However, applications of the proposed tests usually result in substantial variation in the number of factors. For example, Giannone et al. (2004) find that the number of shocks (dynamic factors) driving the US economy is equal to two (i.e. $\widehat{q} = 2$). Stock & Watson (2005) analyzing a similar large US data set set, but with a different method, argue that seven dynamic factors and nine static factors are required ($\widehat{q} = 7$ and $\widehat{r} = 9$). Bai & Ng (2007) and Hallin & Liska (2007) find that $\widehat{q} = 4$.

---

on judgemental and preliminary data analysis of each series, including unit root tests.

Prior to the estimation, we de-mean the series and divide them by their standard deviation such that the resulting series have zeros mean and unit variance.

Part of the difference in the number of factors can be attributed to the fact that earlier research focussed primarily on fitting the leading statistical indicators for economic activity and inflation. Stock & Watson (2005) demonstrate, however, that additional factors are required to fit the other dimensions of the data panel. Bernanke et al. (2005) do not use the popular information criteria by Bai & Ng (2002) because this test does not address the number of lags of the factors. They prefer a model with four factors of which three are latent and the last is the federal funds rate. Bork (2008) also consider the same data as in Bernanke et al. (2005) and based on various information criteria including the criteria by Bai & Ng (2002) he finds that an exactly identified factor-augmented VAR with $\hat{r} = 8$ explains the data well.

Based on the reasoning by Stock & Watson (2005) and the results of Bork (2008) we allow for nine factors and include six lags in the dynamics of the factors ($r = q = 9$ and $p = 6$)[10]. The motivation for introducing more factors is based on the observation that our approach, unlike the latent factor approach, imposes a large number of over-identifying restrictions on the loading matrix. These over-identifying restrictions most likely reduce the fit of each of the factors. This decrease in flexibility is compensated for by increasing the number of factors.

Economic interpretation of the factors

We identify the nine retained static factors using a relatively wide array of economic concepts or interpretations, relevant for empirical monetary policy analysis. The identification of seven out of the nine factors is motivated by small-scale macroeconomic theoretical models. In particular, we retain four (aggregate supply) factors: an *inflation factor* (π); an *economic activity factor* (y); an *hours in production factor* (hrs) functioning as a buffer to changes in demand and an *unemployment factor* (u_n). The standard aggregate demand equation motivates the identification of the following three factors: a *consumption factor* (c)*;* a *housing factor* (h) approximating (residential) investment; and a *monetary policy factor* (i)[11].

The remaining two factors have an interpretation either as additional information

[10] Our results are robust to including more lags and to reducing the number of lags to $p = 4$. Choosing a lower order VAR than $p = 4$ seems to leave some of the endogeneous response of the monetary policy in the VAR residuals which in turn affects the impulse response functions; see the following sections for a discussion.

[11] For more details we refer to Bernanke et al. (2005) for a nice exposition on the mapping between a small-scale macro model and a factor model.

factors or as financial factors.[12] More precisely, we identify a *stock market factor* (s) which may capture wealth effects on consumption, a Tobin's q effect on investments as well as serving as an information factor for monetary policy in the sense that deciphering the forward-looking expectations of the private sector embedded in stock prices is relevant information for policy makers. Finally, we define a *commodity price factor* ($pcom$) which is intended to indicate nascent inflation upon which the monetary policy makers may respond to. We experimented with various factor specifications and discuss these later in section 2.3.1.

[Insert Table 2.1]

Table 2.1 offers an overview of the identification restrictions. The identification of the respective factors is obtained in two steps. First, we identify the factors by imposing a set of unbiasedness restrictions. In particular, we impose unbiasedness restrictions on nine observables closest to the economic interpretation of each of the factors (see shading areas in Table 2.1).[13] This results in an exactly identified system (along the lines of *Proposition 2* in Geweke & Singleton (1981)). This exactly identified latent factor model is labelled as the "*unrestricted model*".

Second, to enhance economic interpretation of the factors we impose overidentifying restrictions in the form of exclusion restrictions (see empty boxes in Table 2.1).

Generally, the identification scheme is based on two strategies. First, exclusion restrictions are primarily imposed on slow-moving variables while fast-moving

[12]Information variables (or information factors) are assumed to be monitored by central banks because they may display relevant information that is not available in typical macroeconomic variables. See Leeper et al. (1996), Christiano et al. (1999) and very recently Bjørnland & Leitemo (2009) for a discussion. Generally, information variables are fast-moving variables that respond contemporaneously to all variables. Examples of fast moving variables include auction market commodity prices, stock prices, and options on financial instruments.

[13]The target observables of the factors are: the CPI-all items index (series 108) for the inflation factor (π); the Unemployment Rate all workers (series 26) for the unemployment factor (u_n); the Industrial Production-total index (series 16) for the economic activity factor (y); Personal Consumption Expenditure all items (series 49) for the consumption factor (c); Average weekly Hours of Production in manufacturing (series 47) for the hours in production factor (hrs); Housing Starts non-farm (series 54) for the housing factor (h); NAPM commodity price index (series 102) for the commodity price factor ($pcom$); The effective federal funds rate (series 77) by the monetary policy rate factor (i); and finally the NYSE stock price index (series 66) for the stock market factor (s). See appendix A.1 page 162 for the definition and numbers assigned to each observable in the data panel.

observables are left unrestricted (except for housing starts and stock market observations).[14] This modeling choice is motivated by the idea that fast moving variables, containing a speculative component, can be considered as general and timely information variables for macroeconomic developments. Second, we differentiate between nominal, real, information, and policy factors.

We define: one nominal factor (*inflation factor);* four real factors (*unemployment*, *economic activity*, *consumption*, and *hours in production factors*); three information factors (*housing, commodity price, and stock market factors);* and one policy factor (*monetary policy factor).*

In our identification strategy, nominal factors exclude all types of real observables as (contemporaneous) information variables. In the same way, real factors exclude nominal variables. Information factors exclude all slow-moving real and nominal observables. Finally, the policy factor loads freely on all observables (except for the necessary unbiasedness restrictions). Details on the restrictions per variable are described in more detail in Appendix A and displayed in terms of shaded and empty cells in Table 2.1.

Identification of structural monetary policy shocks

Dynamic factor models (DFMs) or related models such as FAVARs (e.g. Bernanke et al. (2005)) or large Bayesian VARs (Banbura et al. (2008)) are increasingly used to assess the economic impact of monetary policy shocks. The main advantages of these models over the commonly-used small-scale VAR models are well understood: (i) a large information set is used in the former models leaving less scope for the omitted variable problem or the fundamentalness problem (see e.g Forni et al. (2008)); (ii) the results are more robust, i.e. less dependent on the particular choice of variables than in a small-scale VAR; (iii) the formal structure of the DFMs, FAVARs or BVARs, is sufficiently strict to contain estimation problems.

The structural analysis of monetary policy shocks proceeds as in conventional small-scale VAR models in several respects. First of all, the factors in DFMs, FAVARs or BVARs play the role of the variables in the conventional VAR. Provided

[14] We use the definition of fast- and slow-moving variables of Bernanke et al. (2005) except housing starts and stock market returns, which we assume not to respond *contemporaneously* to some factors. This assumption helps empirically to distinguish a housing factor from a stock market factor.

that one of the factors is the federal funds rate, the issue of identifying the structural monetary policy shock from the reduced form VAR residuals therefore applies exactly similar to the conventional VAR. Accordingly, the identification schemes often entertained in the structural VAR (SVAR) literature apply equally well.

In the following a brief introduction to empirical monetary policy analysis using SVARs is presented; thorough expositions are given in Leeper et al. (1996) and Christiano et al. (1999). The first thing to realize is that most monetary policy actions are systematic, i.e. the actions are predominantly endogenous response to the state of the economy. As such, the systematic effects of monetary policy on the economy are difficult to assess on the basis of historical aggregate time series. However, not all of the variation in the monetary policy instrument can be characterized as a response to the state of economy. The unaccounted variation is formalized with the notion of an exogenous monetary policy shock and a non-exhaustive list of exogenous policy shocks includes changes in the mix of the board of policy makers, changes in the preferences of the board[15], and technical factors like measurement errors arising from real-time data versus revised data. An empirical identification of these shocks is interesting because only when policy makers deviate from the endogenous response it becomes possible to collect empirical evidence of the response of actual economies to such shocks. Furthermore, the empirical evidence facilitates a comparison with a similar shock in model economies; cf. Christiano et al. (1999) for a further discussion.

The VAR model introduced by Sims (1980) represents a widely used model for evaluating the dynamic response of a monetary policy shock. In particular, all the variables (or factors) in the VAR are endogenous and can be written in a moving-average (MA) form of the reduced form VAR residuals. However, these reduced form VAR residuals lack any economic content because they are linear combinations of underlying structural shocks but identification of the structural shocks of interest would lead to a structural MA form of the VAR which in turn allow for an impulse response analysis of the monetary policy shocks.

The notion of systematic responses of monetary policy and monetary policy shocks is formalized in a standard fashion by a linear reaction function f for the

[15]For instance a shift toward more weight on inflation versus unemployment.

monetary policy instrument:

$$i_t = f(\mathcal{F}_t) + \sigma_i \varepsilon_t^i$$

where $\mathcal{F}_t$ are the variables that the policy makers look at (time t information set), ε_t^i is the monetary policy shock normalized to have zero mean and unit variance and σ_i is the standard deviation of monetary policy shock. We follow Christiano et al. (1999) and more recently Bernanke et al. (2005) and apply a recursive identification approach to the identification of the structural monetary policy shock. This approach assumes that the monetary policy shock is contemporaneously orthogonal to the variables that enter the reaction function, i.e. that the variables in $\mathcal{F}_t$ do not respond contemporaneously to time t shock but instead respond with a lag. This effectively allows for a recovery of the monetary policy shock from the VAR residuals.

A standard approach in the literature is to relate the reduced form VAR residuals e_t to the structural innovations u_t by $e_t = Su_t$ where S is a lower triangular Cholesky decomposition of the covariance matrix of e_t. Moreover, this particular identification approach leaves the dynamic impact on all the variables in the VAR to a monetary policy shock invariant to the ordering of the variables. This result is due to the proposition in Christiano et al. (1999)[16]. Nonetheless, the actual ordering of the variables is based on economic theory as follows.

Inflation is ordered first in the VAR as inflation responds with long and variable lags to changes in the monetary policy through complex transmission mechanisms. Delayed response of aggregate demand to changes in monetary policy, periodic wage negotiations and staggered price-setting among firms all lead to a drawn out response of inflation. Unemployment is ordered second, industrial production ordered third and consumption fourth. Underlying this ordering is the idea that a monetary policy shock will impact consumption before industrial production which in turn affects unemployment that finally affects inflation via some rigidity in wage and price setting. The fifth factor is defined as (overtime) hours in production and this factor is assumed to respond relatively faster. The sixth factor is housing starts and the seventh factor is a commodity price factor which partly represents commodity prices determined in auction-like markets and partly represents producer prices including

[16]However, for non-policy shocks the impulse responses are sensitive to the ordering of the variables in the VAR.

crude and intermediate components of the producer price index. The eighth factor is the monetary policy instrument (federal funds rate). The ninth factor is a stock market price index factor and obviously fast-moving. We order this last serving as a informational variable for the monetary policy makers[17]. Using the previously defined notation the factors in F are therefore ordered as $\{\pi, u_n, y, c, hrs, h, pcom, i, s\}$.

Alternative factor specifications

The factor specification described above is motivated by theoretical and empirical monetary policy as well as by the characteristics of the distinct factors in the panel data. The factors representing inflation (π), output (y), and the monetary policy interest rate (i) follows readily from theoretical small-scale models within monetary theory[18] whereas another standard theoretical ingredient in the form of potential output $\left(y^N\right)$ has not been represented by a factor.

According to Giordani (2004) the omission of the potential output in empirical VARs results in a 'price puzzle' as first noted by Sims (1992).[19]. Typically, the literature on empirical monetary policy analysis includes a commodity price index in the VARs in order to mitigate the price puzzle. The reason is, that commodity prices contain useful information about potential output (to be precise the output gap given by $y - y^N$). Inspired by this we tried to approximate the output gap by the capacity utilization rate given by a single series in the data by imposing an unbiasedness restriction on this. However, in order to achieve a precise measure of the capacity utilization rate one would need to impose many exclusion restrictions on the slow-moving variables resulting in a thinly defined factor that most likely will be dominated by the fast-moving variables[20].

[17]Notice, that if structural shocks to stock market prices are considered then a simultaneity problem arise in the sense that both the federal funds rate and stock market prices should be allowed to respond contemporaneously to either of these shocks. Bjørnland & Leitemo (2009) solve this by long-run restrictions. However, we do not consider shocks to the stock market and order the stock market factors last. Ordering stock market last implies that stock prices can respond contemporaneously to all other shocks inlcuding the federal funds shock. However, the monetary policy shock respond with a lag to shocks to stock market prices; the underlying assumption is that the monetary authority wants to evaluate whether the shock is fundamental or not.

[18]See the backward-looking models of e.g. Svensson (1997) and Rudebusch & Svensson (1999).

[19]A typical finding in standard VAR analysis of monetary policy is an increase in the price level following a contractionary monetary policy shock - hence the notion of a price puzzle, because we would expect a decrease.

[20]Notice, that whenever we tried to mix the capacity utilization rate with other slow-moving variables the signal get too blurred to represent inflation expectation and/or supply shocks and consequently resulted in a price puzzle.

We suggest including more proxies for the output gap in the data set if an output gap factor should be extracted. The same problem with thinly defined factors applies to the NAPM commodity prices index, but we solve this problem by imposing one-to-one restrictions on the closely related intermediate and sensitive materials among the producer price indices.

Summing up the discussion, we choose to include a commodity price factor to proxy the output gap and to construct a hours in production factor which is quite correlated with the capacity utilization rate.

2.3.2 Empirical results

Evaluating the over-identifying restrictions

Our identification scheme involves more than 400 over-identifying restrictions and the validity of these restrictions is tested statistically. Specifically, a likelihood ratio (LR) test is used to test the over-identifying restrictions against the exactly identified "unrestricted model". As expected all the restrictions lumped together are clearly statistically level rejected at any significance level but interestingly the economic significance of the restrictions is indeed small. As discussed in details below the consequence of imposing more than 400 over-identifying restrictions is an approximately 3 percentage points decrease in overall adjusted R^2. We conclude that little is lost by imposing the over-identifying restrictions and we are willing to pay the price of a slight reduction in overall R^2 for economically interpretable factors.

Similar findings have been found by Reis & Watson (2008). In a related dynamic factor model they estimate a measure of pure inflation by imposing a unit loading on each of 187 US sectoral price indices. Their restrictions are rejected in t-tests but they find that for eighty percent of the series the decrease in terms of R^2 is less than 3%.

In order to examine whether some of our restrictions are particularly restrictive, we impose the restrictions sequentially using 23 blocks of restrictions of varying sizes. The model is re-estimated for each added set of restrictions. Table 2.1 represents the loading matrix and displays the exactly identifying restrictions, the 23 blocks of over-identifying restrictions and the free parameters of our preferred model as follows. Firstly, the exactly identifying restrictions are indicated by shaded entries with either a 1 or 0 and constitute the unrestricted model. The numbers in Table 2.1 represent estimated free parameters of our preferred model.

For the first factor (π) a total of fifteen exclusion restrictions are lumped together in block [1] and imposed on industrial production. These over-identifying restrictions are in fact accepted in a likelihood ratio test[21]. Then an additional four over-identifying restrictions in block [2] are imposed on π which are rejected, although the overall adjusted R^2 is unchanged. Subsequently, the restrictions in block [3] to block [23] are added implying that when the block [23] is reached all restrictions are imposed yielding the preferred model. For each block of restrictions we calculate the LR-test, the adjusted R^2, AIC and BIC. We also report the panel information criteria IC_{p2} from Bai & Ng (2002) targeted towards principal component dynamic factor models. However, to comply with the EM algorithm the IC_{p2} criteria is slightly modified in terms of the convergence rate of the estimated factors towards the true factors and is denoted IC^{*}_{p2}; see Doz et al. (2006) for the convergence rate of $\left(\sqrt{T}, \frac{N}{\log N}\right)$.

The rectangles represents the blocks of exclusion restrictions starting with the first block of 15 over-identifying restrictions on the first factor π in the upper left corner. Notice the small [1] in the upper right corner of the rectangle which denotes block number [1]. Within the rectangle the overall adjusted R^2 is reported along with the above mentioned criteria if space allow. We then move downwards in the π column and impose block number [2]. This particular rectangle is not completely closed because an element in the rectangle does not belong to the set of over-identifying restrictions. We intend to indicate the opening part of the rectangle in the same way as the symbol [and the closing part of the rectangle as]. One-to-one restrictions are indicated by unit integers without any decimal places. When all blocks of restrictions have been imposed and subsequently estimated we end up in the lower right rectangle of the table denoted [23]. The adjusted R^2 from this rectangle equalling 52.5% is the fit of preferred model to be compared with an adjusted R^2 of 55.9 in the unrestricted model.

[21] To save space a full table of the 23 likelihood ratio tests is not reported. A likelihood ratio test of the over-identifying restrictions in block [1] cannot be rejected at a 1%, 5% or 10% significance level. However, the restrictions in block [2] to block [23] are clearly rejected. Likelihood ratio tests were also calculated from block [i] to block [$i+1$] and in this marginal sense we could not reject the restrictions imposed on block [5] and on block [22].

As expected, the adjusted R^2 measures in Table 2.1 decrease and the AIC/SIC measures increase as more and more restrictions are imposed. Moreover, all blocks of restrictions are rejected in a likelihood ratio test, except the first block and the fifth block, where the latter is only accepted in a marginal sense, i.e. the log likelihood value for block [5] is not significantly different from the log likelihood value of block [4].

An evaluation of the economic significance of the restrictions shows that some of the restriction blocks result in a clear decrease in R^2 up to half a percentage point and a relative clear increase in the AIC/SIC. On the other hand, some of the other restriction blocks do not change the R^2 nor the AIC/SIC. Consider each of the factors in turn. Five of the six blocks of restrictions imposed on the inflation factor result in a very small decrease in R^2 whereas the last block restriction on housing starts and stock returns seems to be somewhat restrictive. We suspect that this block restriction significantly changes the nature of the first factor towards a clear inflation factor thus eliminating any potential residential investment component and stock market component in this factor which in turn may deteriorate the fit of housing starts and stock returns. For the unemployment factor, the economic activity factor, the consumption factor and the stock market factor there are no particular restrictive block restrictions. The three factors, hours in production, residential investment and commodity prices, are more narrowly defined factors and consequently one should expect reduced explanatory power of these factors compared to the fully latent factors. In fact, this illustrates the tradeoff we face between economic interpretability and statistical fit in the sense that some of the explanatory power of the fully latent factors has to be sacrificed to achieve an unambiguous economic interpretation of the factors[22]. In that respect, restriction block [16] is relatively restrictive for the residential investment factor, which could be explained by the exclusion of e.g. unemployment variables in the measurement equation[23]. For the (NAPM) commodity price factor, restriction block [18] is relative restrictive which may be explained by the exclusion of other NAPM survey measures whereas the exclusion of CPI inflation is unrestrictive.

[22] Furthermore, the defined factors should also be relevant for empirical monetary policy analysis.

[23] The exclusion of unemployment and employment variables in the housing factor may eliminate any potential *employment* component (distinct from the unemployment factor) which in turn may deteriorate the fit of the employment variables. In other words, if any employment component is embedded in the housing factor the exclusion restrictions eliminate this component such that a clearer interpretation of the factor emerges.

Implied factors

Figure 2.1 displays the factors as retrieved from the panel. Overall, these factors are well in line with the leading measures and trends in the US economy over the sample period. Specifically, the general inflation factor captures almost perfectly the overall CPI series while the economic activity factor picks up most of the peaks and troughs as identified by the NBER and indicated by shaded bars in the figure.

Insert Figure 2.1 and Table 2.2

Table 2.2 reports the factor loadings as well as the total variance explained by the common factors (R-squared) for a number of leading economic measures[24]. Overall, the statistics reported in Table 2.2 and Figure 2.1 support the economic interpretation of the latent factors. Specifically, we find that the inflation factor (π) closely tracks the *CPI-U: All items* inflation. Moreover, the R-squared is higher than the one based on the inflation factor identified by Bernanke et al. (2005) (96% instead of 87%).[25] The estimated factor loadings on other CPI and PPI inflation series are significantly positive and the common component captures a substantial part of the variation in these series.

The unemployment factor (u_n) captures almost half of the variation in all of the 23 (un)employment series while the R-squared for the four unemployment duration series is almost 70%. Moreover, this factor contributes predominantly to the fit of the capacity utilization rate measured by unreported marginal R-squared.

The economic activity factor (y) explains up to 97% of growth in industrial production and also fits reasonably well the different components of industrial production (R-squares above 50% for half of the 16 series). Moreover, loadings for industrial production components are in general positive. The economic activity factor also contributes to the variation of payroll, income and employment variables.

The consumption factor (c) is restricted to load only on the five personal expenditure series in addition to the fast-moving variables. The one-to-one restrictions help to extract a consumption factor that explains half of the variation in the personal expenditure series which is significantly higher than the 6-10% reported by

[24] To save space we have not reported the estimated covariance matrices $\hat{Q}$ and $\hat{R}$ as well as the companion matrix $\hat{\Phi}$. These results are available on request. It should be noted that all eigenvalues of $\hat{\Phi}$ are less than one in absolute values such that the system is stationary.

[25] Bernanke et al. (2005) use an exactly-identified four-factor FAVAR model.

Bernanke et al. (2005) and Bork (2008). Our identification approach therefore allows to target a group of variables in the panel which otherwise have little chance of showing up in a distinct factor. The price for this may be a reduced overall fit of the panel but as reported above this appears to be only modest in our application.

The hours in production factor (hrs) explains average weekly overtime hours for production workers in manufacturing almost perfectly. Furthermore, this factor also helps explaining capacity utilization and help-wanted ads measured by unreported marginal R-squared. The commodity factor ($pcom$) captures close to half of the variation in monthly commodity price inflation as measured by movement in the NAPM commodity price index. The housing factor (h) explains on average 66% of the variation in the seven housing start series while the stock market factor (s) explains more than 80% of the variation for four out of five stock prices.

Finally, the overall, unadjusted average R^2 of the panel is 53%, which is comparable to the R^2 reported by Bai & Ng (2007), Bork (2008) and Yu (2008) where over-identifying restrictions are not imposed[26]. Moreover, for the targeted concepts, e.g. inflation, economic activity, we obtain significantly higher values for the R-squared. Also, the reported average R^2 corresponds to the average R^2 that one would obtain from the Stock & Watson principal components approach to factor models if six factors are entertained for this particular panel data. This suggests that the over-identifying restrictions and the implied economic interpretation of the factors can be obtained without major loss in fitting the dominant sources of variation in the panel.

Measuring the impact of monetary policy

We use our model to analyze the overall impact of monetary policy shocks on the US economy. To facilitate comparison with other papers including Bernanke et al. (2005), we do not present the impulse response functions (IRFs) of the factors themselves but instead focus on the IRFs of twenty key measures covering the US economy, as implied by the factor model. More specifically, we analyze the federal funds rate, the level of industrial production, the consumer price level (CPI), monetary aggregates, the capacity utilization, the (un)employment level, the average hourly earnings, the level of consumption and consumer confidence expectations as key

[26] Stock & Watson (2002a) find 39% for $r = 6$ in a panel of 215 US monthly series.

indicators for the macroeconomy. Additionally, we cover housing starts and three financial market variables: the dividend yield on the S&P, the five year treasury yield and the USD-YEN foreign exchange rate.

Insert Figure 2.2 and Table 2.3

Figure 2.2 displays the IRFs of each of these variables to a 25 basis point monetary policy shock. The unit of the impulse response functions is the standard deviation of the respective series.

Christiano et al. (1999) survey and estimate monetary policy shocks using different VAR specifications and conclude that there is considerable agreement in the literature about the qualitative effects of a monetary policy shock. Specifically, monetary policy shocks explain only a modest fraction of the variation in output and prices as measured by forecast error variance decomposition. Furthermore, a contractionary monetary policy shock is followed by an increase in interest rates and a fall in output and employment whereas wages respond only modest and prices decrease slowly with a significant time lag. Our impulse response functions depicted in Figure 2.2, are as expected and in line with this finding[27]. Therefore, the plausibility of the impulse response functions suggests that the model is able to identify accurately the key macroeconomic transmission mechanisms and shocks. Several observations can be made in this respect.

First, in line with recent FAVAR or BVAR models and unlike standard small-scale VAR models, we do no longer observe a persistent price puzzle. The (permanent) deflationary effects of monetary policy tightening do appear with about a one-year lag. Second, we observe long-run neutrality. Monetary policy shocks only have a temporary effect on production and consumption, although it holds only marginally for the latter. Long-run neutrality also holds for both the depicted unemployment series and employment series. Third, note that the impact of temporary policy shocks is initially negative on the consumption expectations but then reverses before the impact becomes neutral in the long-run. Finally, the results show a significant impact of monetary policy shocks on financial markets. Monetary policy tightening increases the bond yields with the short-term yields responding more

[27] The impulse response functions (IRFs) for the monetary aggregates are somewhat puzzling as we would expect a decrease in the monetary aggregates following a positive monetary policy shock. However, the uncertainty surrounding these IRFs are very wide implying that any interpretation should be made with caution.

than the long-term yields, as illustrated by the IRF of the 3 month and 5 year yield. However, given the moderate persistence of the policy shocks (see the IRF of the federal funds rate), the impact on bond yields of monetary policy shocks remains relatively small and temporary. Real estate markets, as illustrated by the IRF of the housing starts, initially respond strongly to the monetary policy shock although there is no long-run effect. On the other hand, price-dividend ratio tends to adjust downwards following a monetary tightening. These IRFs match both the responses reported in Banbura et al. (2008), using a BVAR and Bernanke et al. (2005) using a FAVAR.

Table 2.3 reports the variance decomposition of the selected variables at alternative forecasting horizons. This variance decomposition allows us to assess the relative importance of monetary policy shocks in the overall variation of the series. Our results are broadly in line with those reported both in Banbura et al. (2008) and Bernanke et al. (2005). In line with these studies, we observe that monetary shocks do not have important long-run impact on the forecast error variance of a broad selection of twenty key macroeconomic and financial variables. Specifically, we find that a monetary policy shock explains less than 10% of the variation in industrial production, consumer prices, commodity prices, (un)employment, new orders for any forecast horizon and virtually zero for consumption. For the bond yields the portion of the variance explained is decreasing in the maturity of the bond and does not exceed 15% in the long run. Unlike Bernanke et al. (2005), we do not find a large significant long-run effect of monetary policy shocks on the federal funds rate. The estimates reported in Table 2.3 indicate that monetary policy shocks are only mildly persistent and only account for approximately 15% of total long-run variation in the federal funds rate. Banbura et al. (2008), reporting similarly small numbers, argue that this may be explained by the size of the model[28].

2.4 Conclusion

This paper has proposed a methodology to identify factors within the framework of Dynamic Factor Models. We impose an economic interpretation on the static factors through a set of over-identifying restrictions on the factor loadings. We

[28] The larger the model, the more shocks can be identified and the smaller the likelihood of misspecification of the monetary policy shocks. In this model we identify nine structural shocks, which is significantly higher than the number of structural shocks identified by Bernanke et al. (2005).

modify the standard estimation methodology to incorporate these over-identifying loading restrictions. In particular, following Shumway & Stoffer (1982) and Wu et al. (1996), we derive the appropriate parameter estimators and filters based on the EM algorithm.

In the application, we focus on identifying a set of nine factors with unambiguous economic interpretation. These factors represent key measures of the US economy such as inflation, unemployment, economic activity, consumption, state of the industrial production, residential investments, financial markets and monetary policy. The obtained factors are empirically plausible measures for each of the targeted key concepts, listed above. Subsequently, we use the model to assess the overall impact of monetary policy on the US economy. Our results are in line with the results those obtained using alternative methods on large panels, e.g. FAVARs or large BVARs.

The framework proposed in this paper has many other applications in economics and finance. For instance, the identification restrictions can be used to generate factor pricing models, where factors can be economically interpreted. This type of model could be used to evaluate and analyze the types and the importance of macroeconomic risks in stock and bond markets.

A Over-identifying loading restrictions

The specific set of (over-) identifying restrictions can be summarized as follows; the *inflation factor* (π) is identified by the unbiasedness restriction on "CPI-U: All items". Additionally, we allow other inflation measures to load on the inflation factor. With the inflation factor being a nominal factor, we exclude from the information set all real variables, e.g. industrial production.

For the four real factors, we impose exclusion restrictions on nominal variables (e.g. CPI inflation). Additional exclusion restrictions limit the type of real variables acting as information variables for each of the factors. In particular, the *unemployment factor* (u_n) is identified by the unbiasedness restriction on "Unemployment: All workers". Other (un)employment variables and measures of payroll statistics and capacity utilization are included as additional information variables. All other slow-moving variables are excluded from the information set. The *economic activity factor* (y), identified by the unbiasedness restriction on the "Industrial Production: Total index" uses other industrial production (IP) variables next to employment and payroll series as additional state variables. The *hours in production factor* (hrs) measures the current over (under) production and is identified (by means of an unbiasedness restriction) through the overtime hours in production and manufacturing.

As additional information variables we include variables such as capacity utilization rate, survey-based production indices (PMI, PMP) and help-wanted advertising to enter freely. We exclude (un)employment and IP growth as we consider them less informative with respect to the level of over and underproduction. The last real factor, i.e. the *consumption factor* (c) , is filtered from the observed consumption series in the panel with an unbiasedness restriction on "Personal Consumption Expenditure" series and one-to-one restrictions on two consumption observables. Moreover, due to consumption smoothing, we do not expect strong contemporaneous correlations between production employment based statistics and consumption (growth). Therefore, we impose exclusion restrictions on production related variables.

The information and the policy factors measure particular features in the economy. More precisely, the *housing factor* (h) is included as a residential investment factor. This factor is identified through an unbiasedness restriction on the total number of housing starts and uses as additional information variables other housing starts or authorization variables. We consider the housing factor to be mainly a forward-looking variable containing all relevant information. As such, exclusion restrictions are imposed on all slow-moving variables. The *commodity price factor* ($pcom$) aims at measuring cost-push factors due to price increases of raw materials or intermediate products. It is identified by means of the NAPM commodity price index. Moreover, the commodity price factor retrieves additional information from PPI data for crude and intermediate materials and from the index of sensitive materials. The *monetary policy factor* (i) is directly measured by the effective federal funds rate. Finally, the *stock market factor* (s) is related to returns on the NYSE index and uses S&P500 stock market component indices as additional state variables. We allow all other fast-moving variables to load freely on the stock market factor allowing for direct interactions across financial markets. Notice that the imposition of an unbiasedness loading restriction on the stock market factor seems to prevent this factor from responding contemporaneously to shocks to the other factors which seems inappropriate for a financial market factor. However, the combination of the stock market factor ordered last in the VAR and a lower triangular S (consistent with a standard recursive identification scheme) makes the stock market factor respond contemporaneously to all shocks.

B The EM algorithm, the Kalman filter and the Kalman smoother

The EM algorithm is an iterative maximum likelihood procedure applicable to models with "missing data", which in this context is the unobserved factors. The complete data likelihood of the Gaussian state space model in equations (2.5)-(2.6) is given in equation (2.21) below. Although the complete data likelihood cannot be calculated due to the unobserved factors, it is nevertheless possible to calculate the expectation of the complete data likelihood conditional on the observed data and inputs of parameters, denoted $\Theta^{(j)}$ at the jth iteration. Essentially, this expectation depends on smoothed moments of the unobserved variables from the Kalman smoother and hence on the data as well as parameters in $\Theta^{(j)}$. Finally, "updated" values of the parameters at iteration $j+1$ denoted $\Theta^{(j+1)}$ are available in closed form and follows from the first-order conditions of the conditional expectation of the complete data likelihood. The updated parameters $\Theta^{(j+1)}$ can then be used to filter and smooth a new set of moments to be used in the calculation of the conditional expectation of the complete data likelihood. This algorithm continues until convergence of the likelihood value.

The following offers a brief description of the Kalman filter and the Kalman smoother. Then the complete data likelihood and the incomplete data likelihood for a state space model are stated. Finally the moments used in the closed form parameters estimators in (2.12) are stated.

The Kalman filter

Denote by $\mathcal{X}_t = \{X_1, ..., X_t\}$ the information set available at time t. The conditional expectation and variance of the factor are: $\hat{F}_{t+1|t} = E\left[F_{t+1}|\,\mathcal{X}_t\right]$ and $\hat{P}_{t+1|t} = var\left(F_{t+1}|\,\mathcal{X}_t\right)$, respectively.

The Kalman filter recursions for $t = 1, .., T$ can then be written as

$$
\begin{aligned}
&\hat{F}_{t+1|t} = \Phi \hat{F}_{t|t-1} + K_t \left(X_t - \Lambda \hat{F}_{t|t-1} \right), \\
&\hat{P}_{t+1|t} = \Phi \hat{P}_{t|t-1} L_t^{\top} + Q,
\end{aligned} \tag{2.15}
$$

where

$$
\begin{aligned}
&\underset{n\times 1}{\xi_t} = X_t - \Lambda \hat{F}_{t|t-1}, \\
&\underset{n\times n}{P_t^{\xi\xi}} = \Lambda \hat{P}_{t|t-1} \Lambda^{\top} + R, \\
&\underset{k\times n}{K_t} = \Phi \hat{P}_{t|t-1} \Lambda^{\top} \left(\Lambda \hat{P}_{t|t-1} \Lambda^{\top} + R \right)^{-1}, \\
&\underset{k\times k}{L_t} = \Phi - K_t \Lambda.
\end{aligned}
$$

Kalman smoothing

Kalman smoothing is the name for the reconstruction of the full state sequence $\{F_1, .., F_T\}$ given the observations $\{X_1, .., X_T\}$. Smoothing provides us with more accurate inference on the state variables since it uses more information than the basic filter.

The Kalman smoother recursions for $t = T, ..., 1$, based on the efficient smoother by de Jong & Mackinnon (1988), de Jong (1989) and used in Koopman & Shephard (1992) are given by

$$
\begin{aligned}
\hat{F}_{t|T} &= \hat{F}_{t|t-1} + \hat{P}_{t|t-1} \Lambda^{\top} \left[\hat{P}^{\xi\xi}_{t|t-1} \right]^{-1} \xi_t + \hat{P}_{t|t-1} L_t^{\top} r_t && (2.16) \\
&= \hat{F}_{t|t-1} + \hat{P}_{t|t-1} r_{t-1} \quad \text{(alternatively)} && (2.17) \\
\hat{P}_{t|T} &= \hat{P}_{t|t-1} - \hat{P}_{t|t-1} N_{t-1} \hat{P}_{t|t-1} && (2.18) \\
\hat{P}_{\{T,T-1\}|T} &= \left[I - K_T \Lambda \right] \Phi \hat{P}_{T-1|T-1} && (2.19) \\
\hat{P}_{\{t,t-1\}|T} &= \left(I - \hat{P}_{t|t-1} N_{t-1} \right) L_{t-1} \hat{P}_{t-1|t-2}, \; t = T-1, ..., 1 && (2.20)
\end{aligned}
$$

where

$$
\begin{aligned}
r_{t-1} &= \Lambda^{\top} \left[\hat{P}^{\xi\xi}_{t|t-1} \right]^{-1} \xi_t + L_t^{\top} r_t, \text{ for } 1 \leq t < T \text{ and } r_T = 0 \\
N_{t-1} &= \Lambda^{\top} \left[\hat{P}^{\xi\xi}_{t|t-1} \right]^{-1} \Lambda + L_t^{\top} N_t L \text{ for } 1 \leq t < T \text{ and } N_T = 0 \\
L_t &= \Phi - K_t \Lambda = \Phi - \Phi \hat{P}_{t|t-1} \Lambda^{\top} \left[\hat{P}^{\xi\xi}_{t|t-1} \right]^{-1} \Lambda.
\end{aligned}
$$

The complete data likelihood and the incomplete data likelihood

Under the Gaussian assumption including $F_0 \sim N(\mu_0, P_0)$ and ignoring the constant, the complete data likelihood of Equations (2.3)-(2.4) page 64 can be written as

$$\begin{aligned} -2\ln L_{\mathcal{F},\mathcal{X}}(\Theta) &= \ln|P_0| + (F_0-\mu_0)^\top P_0^{-1}(F_0-\mu_0) \\ &\quad + T\ln|Q| + \sum_{t=1}^{T}(F_t-\Phi F_{t-1})^\top Q^{-1}(F_t-\Phi F_{t-1}) \\ &\quad + T\ln|R| + \sum_{t=1}^{T}(X_t-\Lambda F_t)^\top R^{-1}(X_t-\Lambda F_t). \end{aligned} \tag{2.21}$$

given that we can observe the states $\mathcal{F}_T = \{F_0, .., F_T\}$ as well as the observations $\mathcal{X}_T = \{X_1, .., X_T\}$. However, given $\mathcal{X}_T$ and some input of parameter estimates (denoted $\Theta^{(j-1)}$) the conditional expectation of the complete data likelihood can be written as

$$\begin{aligned} \mathcal{Q}\left(\Theta|\Theta^{(j-1)}\right) &= E\left[-2\ln L_{\mathcal{F},\mathcal{X}}(\Theta)\left|\mathcal{X}_T, \Theta^{(j-1)}\right.\right] \\ &= \ln|P_0| + \operatorname{tr}\left[P_0^{-1}\left\{\left(\hat{F}_{0|T}-\mu_0\right)\left(\hat{F}_{0|T}-\mu_0\right)^\top + P_{0|T}\right\}\right] \\ &\quad + T\cdot\ln|Q| + \operatorname{tr}\left[Q^{-1}\left\{C - B\Phi^\top - \Phi B^\top + \Phi A\Phi^\top\right\}\right] \\ &\quad + T\cdot\ln|R| \\ &\quad + \operatorname{tr}\left[R^{-1}\sum_{t=1}^{T}\left\{\left(X_t-\Lambda\hat{F}_{t|T}\right)\left(X_t-\Lambda\hat{F}_{t|T}\right)^\top + \Lambda\hat{P}_{t|T}\Lambda^\top\right\}\right] \end{aligned} \tag{2.22}$$

where the following moments can be calculated from the Kalman smoother listed above.

$$C = \sum_{t=1}^{T}\left(\hat{F}_{t|T}\hat{F}_{t|T}^\top + \hat{P}_{t|T}\right) \tag{2.23}$$

$$B = \sum_{t=1}^{T}\left(\hat{F}_{t|T}\hat{F}_{t-1|T}^\top + \hat{P}_{\{t,t-1\}|T}\right) \tag{2.24}$$

$$A = \sum_{t=1}^{T}\left(\hat{F}_{t-1|T}\hat{F}_{t-1|T}^\top + \hat{P}_{t-1|T}\right) \tag{2.25}$$

B.1 The Loading Estimator Subject to Linear Restrictions

Within the EM algorithm approach we want to derive analytically the estimate of Λ subject to linear restrictions $H_\Lambda \operatorname{vec}\Lambda = \kappa_\Lambda$. Consider therefore the part of the conditional expectation of the likelihood function in equation (2.22) which involve Λ

$$\begin{aligned} & -\frac{1}{2}\operatorname{tr}\left[R^{-1}\left\{\sum_{t=1}^{n}\left(X_t-\Lambda\hat{F}_{t|T}\right)\left(X_t-\Lambda\hat{F}_{t|T}\right)^\top+\sum_{t=1}^{n}\Lambda\hat{P}_{t|T}\Lambda^\top\right\}\right] \\ = \; & -\frac{1}{2}\operatorname{tr}\left[R^{-1}\left\{E-D\Lambda^\top-\Lambda D^\top+\Lambda C\Lambda^\top\right\}\right] \end{aligned}$$

where

$$\begin{aligned} C &= \textstyle\sum_{t=1}^{T}\left(\hat{F}_{t|T}\hat{F}_{t|T}^\top+\hat{P}_{t|T}\right) \\ D &= \textstyle\sum_{t=1}^{T}X_t\hat{F}_{t|T}^\top \\ E &= \textstyle\sum_{t=1}^{T}X_tX_t^\top \end{aligned}$$

The maximization of the likelihood problem can be rewritten as the minimization problem:

$$\begin{aligned} \text{minimize} &: \; \frac{1}{2}\operatorname{tr}\left[R^{-1}\left\{E-D\Lambda^\top-\Lambda D^\top+\Lambda C\Lambda^\top\right\}\right] \\ \text{subject to} &: \; H_\Lambda \operatorname{vec}\Lambda=\kappa_\Lambda. \end{aligned}$$

The Lagrangian is:

$$\mathcal{L}(\Lambda) := \frac{1}{2}\operatorname{tr}\left[\underset{N\times N}{R^{-1}}\left\{\underset{N\times N}{E}-\underset{N\times r}{D}\underset{r\times N}{\Lambda^\top}-\underset{N\times r}{\Lambda}\underset{r\times N}{D^\top}+\underset{N\times r}{\Lambda}\underset{r\times r}{C}\underset{r\times N}{\Lambda^\top}\right\}\right]-\lambda^\top\left[\underset{\eta\times Nr}{H_\Lambda}\underbrace{\operatorname{vec}\Lambda}_{Nr\times 1}-\underset{\eta\times 1}{\kappa_\Lambda}\right]$$

and the differential of the Lagrangian is (keeping only $\mathrm{d}\,\Lambda$), cf. Magnus & Neudecker (2007)

$$\mathrm{d}\,\mathcal{L}\left(\tilde{M}\right)=\frac{1}{2}\operatorname{tr}\left[\underset{N\times N}{R^{-1}}\left\{-\underset{N\times r}{D}\underset{r\times N}{(\mathrm{d}\,\Lambda)^\top}-\underset{N\times r}{(\mathrm{d}\,\Lambda)}\underset{r\times N}{D^\top}+\underset{N\times r}{(\mathrm{d}\,\Lambda)}\underset{r\times r}{C}\underset{r\times N}{(\mathrm{d}\,\Lambda)^\top}\right\}\right]-\lambda^\top\underset{\eta\times Nr}{H_\Lambda}\,\mathrm{d}\operatorname{vec}\Lambda$$

The first-order conditions are

$$1) \quad : \quad \left[\operatorname{vec}\left\{\left(\left(\left\{\underset{r\times r}{C}\underset{r\times N}{\Lambda^\top} - \underset{r\times N}{D^\top}\right\}\underset{N\times N}{R^{-1}}\right)^\top\right\}\right]^\top - \lambda^\top \underset{\eta\times Nr}{H_\Lambda} = 0 \tag{2.26}$$

$$2) \quad : \quad \underset{\eta\times Nr}{H_\Lambda} \operatorname{vec}\Lambda = \underset{\eta\times 1}{\kappa_\Lambda} \tag{2.27}$$

Rewrite (2.26) as

$$\left(\underset{r\times r}{C} \otimes \underset{N\times N}{R^{-1}}\right)\operatorname{vec}\left(\underset{N\times r}{\Lambda}\right) - \operatorname{vec}\left(\underset{N\times N}{R^{-1}}\underset{N\times r}{D}\right) = H_\Lambda^\top \lambda \tag{2.28}$$

pre-multiply by $H_\Lambda\left(C^{-1}\otimes R\right)$

$$\left\{\begin{array}{c} \underset{\eta\times Nr}{H_\Lambda}\left(\underset{r\times r}{C^{-1}}\otimes \underset{N\times N}{R}\right)\left(\underset{r\times r}{C}\otimes \underset{N\times N}{R^{-1}}\right)\operatorname{vec}\left(\underset{m\times p}{\tilde{M}}\right) \\ -\underset{\eta\times Nr}{H_\Lambda}\left(C^{-1}\otimes R\right)\operatorname{vec}\left(\underset{N\times N}{R^{-1}}\underset{N\times r}{D}\right)\end{array}\right\} = \underset{\eta\times Nr}{H_\Lambda}\left(\underset{r\times r}{C^{-1}}\otimes \underset{m\times m}{R}\right)H_\Lambda^\top\lambda$$

$\Leftrightarrow$

$$\underset{\eta\times Nr}{H_\Lambda}\underset{Nr\times Nr}{I}\operatorname{vec}\left(\underset{N\times r}{\Lambda}\right) - \underset{\eta\times Nr}{H_\Lambda}\left(C^{-1}\otimes R\right)\operatorname{vec}\left(\underset{N\times N}{R^{-1}}\underset{N\times r}{D}\right) = \underset{\eta\times Nr}{H_\Lambda}\left(\underset{r\times r}{C^{-1}}\otimes \underset{N\times N}{R}\right)H_\Lambda^\top\lambda$$

substitute from the constraint in (2.27)

$$\kappa_\Lambda - \underset{\eta\times Nr}{H_\Lambda}\left(\underset{r\times r}{C^{-1}}\otimes \underset{N\times N}{R}\right)\operatorname{vec}\left(\underset{N\times N}{R^{-1}}\underset{N\times r}{D}\right) = \underset{\eta\times Nr}{H_\Lambda}\left(\underset{r\times r}{C^{-1}}\otimes \underset{N\times N}{R}\right)H_\Lambda^\top\lambda$$

$\Rightarrow$

$$\lambda = \left[\underset{\eta\times Nr}{H_\Lambda}\left(\underset{r\times r}{C^{-1}}\otimes \underset{N\times N}{R}\right)H_\Lambda^\top\right]^{-1}\left\{\kappa_\Lambda - \underset{\eta\times Nr}{H_\Lambda}\left(\underset{r\times r}{C^{-1}}\otimes \underset{N\times N}{R}\right)\operatorname{vec}\left(\underset{N\times N}{R^{-1}}\underset{N\times r}{D}\right)\right\}$$

substitute λ in (2.28)

$$\begin{aligned}\left(\underset{r\times r}{C}\otimes \underset{N\times N}{R^{-1}}\right)\operatorname{vec}\left(\underset{N\times r}{\Lambda}\right) - \operatorname{vec}\left(\underset{N\times N}{R^{-1}}\underset{N\times r}{D}\right) \quad &= \quad H_\Lambda^\top\left[\underset{\eta\times Nr}{H_\Lambda}\left(\underset{r\times r}{C^{-1}}\otimes \underset{N\times N}{R}\right)H_\Lambda^\top\right]^{-1}\times \\ &\left\{\kappa_\Lambda - \underset{\eta\times Nr}{H_\Lambda}\left(\underset{r\times r}{C^{-1}}\otimes \underset{N\times N}{R}\right)\times\operatorname{vec}\left(\underset{N\times N}{R^{-1}}\underset{N\times r}{D}\right)\right\}\end{aligned} \tag{2.29}$$

Consider the term $\operatorname{vec}\left(R^{-1}D\right)$. Using Theorem 2 in ch. 2.4 of Magnus & Neudecker

(2007) which allows to rewrite $\mathrm{vec}\,(R^{-1}D) = \left(D^\top \otimes R^{-1}\right)\mathrm{vec}\,(I)$, which is going to be used in the curly brackets above, i.e. in the term $(C^{-1} \otimes R)\,\mathrm{vec}\,(R^{-1}D)$.

Consider this last mentioned term, post-multiply by $\mathrm{vec}\,(I_N)$ and use equation (4), ch. 2.4 of Magnus & Neudecker (2007)

$$\begin{aligned}
\left(\underset{r\times r}{C^{-1}} \otimes \underset{N\times N}{R}\right)\left(\underset{r\times N}{D^\top} \otimes \underset{N\times N}{R^{-1}}\right)\mathrm{vec}\left(\underset{N\times N}{I}\right) &= \left(\underset{r\times r}{C^{-1}}\underset{r\times N}{D^\top} \otimes \underset{N\times N}{R}\underset{N\times N}{R^{-1}}\right)\mathrm{vec}\left(\underset{N\times N}{I}\right) \\
&= \left(\underset{r\times r}{C^{-1}}\underset{r\times N}{D^\top} \otimes \underset{N\times N}{I}\right)\mathrm{vec}\left(\underset{N\times N}{I}\right) \\
&= \mathrm{vec}\left(\underset{N\times r}{D}\,\underset{r\times r}{C^{-1}}\right)
\end{aligned}$$

Substituting the last result into (2.29) and solving for $\mathrm{vec}\,(\Lambda)$ yields the restricted loadings estimator denoted $\mathrm{vec}\,(\Lambda^*)$ in (2.12).

$$\begin{aligned}
\mathrm{vec}\,(\Lambda^*) &= \mathrm{vec}\left(\underset{N\times r}{D}\,\underset{r\times r}{C^{-1}}\right) + \\
&\quad \left(\underset{r\times r}{C^{-1}} \otimes \underset{N\times N}{R}\right) H_\Lambda^\top \left[\underset{\eta\times Nr}{H_\Lambda}\left(\underset{r\times r}{C^{-1}} \otimes \underset{N\times N}{R}\right) H_\Lambda^\top\right]^{-1} \times \\
&\quad \left\{\kappa_\Lambda - \underset{\eta\times Nr}{H_\Lambda}\,\mathrm{vec}\left(\underset{N\times r}{D}\,\underset{r\times r}{C^{-1}}\right)\right\}
\end{aligned}$$

$\square$

Table 2.1: Estimated factor loadings and test statistics.

	π	u	y	c	hrs	h	$pcom$	i	s
1) IP: products, total	[1]	[7]	0.92	[12]	[15]	[16]	[18]	-0.01	[22]
2) IP: final products	$\bar{R}^2_a$ =	$\bar{R}^2_a$ =	0.87	$\bar{R}^2_a$ =	$\bar{R}^2_a$ =	$\bar{R}^2_a$ =	$\bar{R}^2_a$ =	0.01	$\bar{R}^2_a$ =
3) IP: consumer	55.9%	55.4%	0.75	55.1%	54.4%	53.9%	53.2%	-0.02	52.5%
4) IP: durable cons.			0.72					-0.00	
5) IP: nondurable cons.	AIC: 1.449	AIC: 1.457	0.43	AIC: 1.462	AIC: 1.480	AIC: 1.490	AIC: 1.502	-0.04	AIC: 1.518
6) IP: bus. Equip	SIC: 1.684	SIC: 1.683	0.71	SIC: 1.672	SIC: 1.676	SIC: 1.680	SIC: 1.686	0.01	SIC: 1.688
7) IP: intermediate	IC(2) : -0.410	IC(2) : -0.398	0.73	IC(2) : -0.386	IC(2) : -0.367	IC(2) : -0.355	IC(2) : -0.340	-0.05	IC(2) : -0.321
8) IP: materials	IC(2)*: -0.565	IC(2)*: -0.553	0.87	IC(2)*: -0.540	IC(2)*: -0.522	IC(2)*: -0.510	IC(2)*: -0.495	0.01	IC(2)*: -0.476
9) IP: durable goods			0.87					0.04	
10) IP: nondur. Goods			0.40					-0.04	
11) IP: manufacturing			1.01					0.01	
12) IP: dur. Manuf			0.97					0.02	
13) IP: nondur. Manuf.			0.70					-0.04	
14) IP: mining			0.23					0.03	
15) IP: utilities			0.12					-0.07	
16) IP: total index	0	0	**1**	0	0	0	0	0	0
17) Capacity util rate	[2]	-0.74	0.16	[12]	0.25	[16]	[18]	0.19	[22]
18) Pmi	$\bar{R}^2_a$ =	[7]	0.50		0.25			-0.11	
19) NAPM prod.	55.8%		0.54		0.14			-0.20	
20) Pers. Income	AIC: 1.450		0.31		[15]			-0.05	
21) Pers. Inc. - trans.	SIC: 1.685		0.54					-0.05	
22) Help-wanted	[3]	0.03	0.44		-0.01		[19]	-0.13	
23) Empl. Help-wanted		-0.71	0.01		0.31			0.33	
24) Civ. Labor: empl.,		-0.06	0.39		[15]			0.04	
25) Civilian labor: empl.,		-0.10	0.43					0.03	
26) Unempl. Rate: all wrks	0	**1**	0	0	0	0	0	0	0
27) Unemp dur: mean	[3]	0.62	0.22	[12]	[15]	[16]	[19]	-0.29	[22]
28) Unemp by dur. < 5 wks		0.71	-0.04					0.28	
29) Unemp by dur. 5-14 w		0.79	-0.02					0.11	
30) Unemp by dur. 15+ w		0.80	0.13					-0.06	
31) Unemp by dur. 15-26 w	$\bar{R}^2_a$ =	0.82	0.06				$\bar{R}^2_a$ =	-0.01	
32) Nonag payrl.: total	55.8%	-0.21	0.73				52.7%	0.05	
33) Nonag payrl.: total,		-0.16	0.76					0.08	
34) Nonag payrl.: goods	AIC: 1.451	-0.18	0.81				AIC: 1.516	0.05	
35) Nonag payrl.: mining	SIC: 1.682	-0.11	0.18				SIC: 1.696	0.17	
36) Nonag payrl.: contract	IC(2) : -0.407	-0.04	0.36				IC(2) : -0.327	-0.04	
37) Nonag payrl.: manuf	IC(2)*: -0.561	-0.18	0.81				IC(2)*: -0.481	0.05	
38) Nonag payrl.: durable		-0.18	0.80					0.07	
39) Nonag payrl.: nondur		-0.11	0.56					-0.03	
40) Nonag payrl.: service		-0.24	0.38					0.03	
41) Nonag payrl.: trans.		-0.07	0.14					0.04	
42) Nonag payrl.: sale		-0.12	0.42					0.04	
43) Nonag payrl.: finance		-0.19	0.21					0.11	
44) Nonag payrl.: services		-0.15	0.33					0.10	
45) Nonag payrl.: gov.		-0.25	-0.02					-0.13	
46) Avg. Wkly hrs. prod	$\bar{R}^2_a$ = [4]	[7]	[11]	[12]	0.97			-0.15	
47) Avg. Wkly overtime prod	0	0	0	0	**1**	0	0	0	0
48) NAPM Empl. Index	55.7%	-0.53	0.48	[12]	[15]	[16]	[19]	0.07	[22]
49) Pers cons Exp: total	0	0	0	**1**	0	0	0	0	0
50) Pers cons Exp: tot.	$\bar{R}^2_a$ = [5]	[8]	[11]	1	[14]	[16]	[20]	0.05	[23]
51) Pers cons Exp: nondur.	55.7%			1				0.07	
52) Pers cons Exp: services	AIC: 1.452			0.16				-0.09	
53) Pers cons Exp: new cars	SIC: 1.682			1.04				0.10	
54) Housing starts: nonfarm	0	0	0	0	0	**1**	0	0	0
55) Housing starts: N.E	$\bar{R}^2_a$ = [6]	$\bar{R}^2_a$ = [8]	$\bar{R}^2_a$ = [11]	$\bar{R}^2_a$ = [13]	$\bar{R}^2_a$ = [14]	0.48	$\bar{R}^2_a$ = [20]	-0.21	$\bar{R}^2_a$ = [23]
56) Housing starts: M.W	55.4%	55.3%	55.2%	55.1%	54.9%	0.57	52.5%	-0.38	52.5%
57) Housing starts: S	AIC: 1.457	AIC: 1.461	AIC: 1.460	AIC: 1.462	AIC: 1.467	0.94	AIC: 1.519	0.27	AIC: 1.518
58) Housing starts: S	SIC: 1.686	SIC: 1.684	SIC: 1.677	SIC: 1.669	SIC: 1.670	0.85	SIC: 1.697	0.00	SIC: 1.685
59) Housing auth. Tot new	IC(2) : -0.399	IC(2) : -0.395	IC(2) : -0.393	IC(2) : -0.385	IC(2) : -0.381	1.00	IC(2) : -0.323	0.06	IC(2) : -0.321
60) Mobile homes	IC(2)*: -0.553	IC(2)*: -0.549	IC(2)*: -0.548	IC(2)*: -0.539	IC(2)*: -0.535	0.61	IC(2)*: -0.478	0.32	IC(2)*: -0.475

The consequence of imposing more and more over-identifying loadings restrictions is shown in this table and explained in details in the text. 23 blocks of restrictions of varying sizes are imposed sequentially. Each block is denoted by a number in a bracket in the upper right corner of the rectangles. For each added set of restrictions we estimate the model and calculate the mean adjusted R-squared $\left(\bar{R}^2_a\right)$, AIC and SIC as well as the IC_{p2} panel information criteria from Bai & Ng (2002) and IC^*_{p2} which takes into account the convergence rate of Doz et al. (2006).

Table 2.1 continued

	π	u	y	c	hrs	h	$pcom$	i	s
61) NAPM inventories	0.06	-0.39	0.08	0.00	0.01	0.27	0.16	0.04	-0.05
62) NAPM new orders	-0.09	0.08	0.39	-0.01	0.07	0.26	0.30	-0.29	0.00
63) NAPM vendor deliv.	0.04	-0.29	0.15	-0.01	0.26	0.21	0.15	0.14	-0.09
64) New orders: cons goods	0.00	0.08	0.50	0.13	-0.06	0.00	-0.01	-0.07	0.07
65) New orders: nondefense	0.04	0.01	0.06	0.12	0.03	0.03	0.03	-0.05	0.00
66) NYSE: composite	0	0	0	0	0	0	0	0	**1**
67) S&P composite	[6]	[8]	[11]	[13]	[14]	[17]	[20]	0.00	1.01
68) S&P industrials								-0.00	1.00
69) S&P capital								-0.02	0.92
70) S&P utilities								0.01	0.61
71) S&P: dividend	0.08	0.30	-0.02	-0.00	-0.49	-0.09	0.32	0.31	-0.03
72) S&P: price earnings	-0.06	-0.17	0.01	-0.01	0.48	0.05	-0.30	-0.37	-0.00
73) FX : switzerland	-0.08	-0.02	0.17	0.08	0.06	-0.19	0.08	0.14	0.07
74) FX : japan	-0.12	-0.13	0.09	-0.00	0.01	-0.18	0.06	0.19	-0.04
75) FX : united	0.10	-0.03	-0.16	-0.05	0.03	0.13	-0.01	-0.15	0.01
76) FX : canada	-0.01	0.07	0.13	0.04	-0.01	-0.04	0.01	-0.02	-0.24
77) Federal funds	0	0	0	0	0	0	0	**1**	0
78) US Tbill, 3m.	-0.06	0.12	0.02	-0.01	0.03	-0.03	0.12	0.96	0.01
79) US Tbill, 6m.	-0.08	0.16	0.01	-0.01	0.03	-0.03	0.17	0.94	0.00
80) Tbond const 1yr.	-0.12	0.24	0.01	-0.02	0.04	-0.04	0.23	0.91	0.00
81) Tbond const 5yr.	-0.17	0.54	-0.02	-0.02	0.10	-0.02	0.28	0.76	-0.01
82) Tbond const 10yr.	-0.14	0.62	-0.02	-0.02	0.13	-0.00	0.26	0.69	-0.01
83) Bond yield: Moody AAA	-0.06	0.65	-0.05	-0.02	0.19	0.07	0.12	0.65	-0.02
84) Bond yield: Moody BAA	-0.06	0.65	-0.05	-0.01	0.11	0.08	0.10	0.64	-0.00
85) Spread 3m – fed funds	-0.21	0.38	0.07	-0.03	0.10	-0.09	0.39	-0.88	0.04
86) Spread 6m – fed funds	-0.24	0.47	0.04	-0.03	0.08	-0.09	0.50	-0.92	0.01
87) Spread 1y – fed funds	-0.38	0.75	0.02	-0.05	0.11	-0.12	0.70	-0.79	0.01
88) Spread 5y – fed funds	-0.29	0.93	-0.03	-0.04	0.18	-0.04	0.49	-0.85	-0.01
89) Spread 10y – fed funds	-0.21	0.93	-0.03	-0.03	0.19	-0.00	0.40	-0.87	-0.01
90) Spread AAA – fed funds	-0.09	0.91	-0.07	-0.02	0.26	0.10	0.17	-0.85	-0.02
91) Spread BAA – fed funds	-0.09	1.00	-0.08	-0.02	0.17	0.12	0.15	-0.72	-0.01
92) Money stock: M1	0.17	0.31	-0.05	0.08	-0.21	0.29	-0.08	-0.13	0.05
93) Money stock: M2	0.02	0.03	0.00	0.03	-0.59	0.51	-0.14	0.02	0.04
94) Money stock: M3	0.03	-0.12	-0.04	0.06	-0.44	0.59	-0.07	0.18	0.06
95) Money supply---M2 1992	-0.53	-0.01	0.02	0.03	-0.44	0.40	-0.13	-0.00	0.04
96) Monetary base	0.25	0.25	-0.04	0.01	0.14	0.23	-0.13	-0.05	0.02
97) Depository inst reserves	0.04	0.16	0.02	-0.06	-0.21	0.17	-0.09	-0.05	-0.00
98) Dep. Inst. Res. Nonbor.	0.10	0.07	-0.15	-0.01	-0.16	0.07	-0.18	-0.09	0.06
99) Comm. & indust. Loans	-0.24	-0.22	0.03	0.03	0.19	-0.08	0.23	0.31	0.02
100) Wkly rp lg com.	-0.13	0.02	0.03	-0.02	0.34	-0.06	0.22	0.23	0.10
101) Cons credit outst.	-0.21	-0.06	0.02	0.05	-0.09	0.36	0.29	0.08	-0.03
102) NAPM cmodity prices	0	0	0	0	0	0	**1**	0	0
103) PPI: finished	0.79	[9]	[10]	[13]	[14]	[17]	0.03	-0.12	[23]
104) PPI: finished	0.76						0.05	-0.17	
105) PPI: intermed	[6]						0.28	0.23	
106) PPI: crude							0.20	-0.01	
107) Index of sensitive mat.							0.33	-0.14	
108) CPI-U: all items	**1**	0	0	0	0	0	0	0	0
109) CPI-U: apparel & upkeep	0.44	[9]	[10]	[13]	[14]	[17]	[21]	-0.02	[23]
110) CPI-U: transportation	0.85	$\bar{R}_a^2$ =	$\bar{R}_a^2$ =			$\bar{R}_a^2$ =	$\bar{R}_a^2$ =	-0.20	
111) CPI-U: medical care	0.23	55.2%	55.2%			53.8%	52.5%	0.41	
112) CPI-U: commodities	1.02						0.06	-0.21	
113) CPI-U: durables	0.58						[21]	0.11	
114) CPI-U: services	0.51	AIC: 1.461	AIC: 1.461			AIC: 1.491	AIC: 1.519	0.33	
115) CPI-U: less food	0.85	SIC: 1.683	SIC: 1.680			SIC: 1.678	SIC: 1.696	0.10	
116) CPI-U: less shelter	1.01	IC(2) : -0.393	IC(2) : -0.392			IC(2) : -0.353	IC(2)*: -0.322	-0.09	
117) CPI-U: less medical	1.00	IC(2)*: -0.548	IC(2)*: -0.547			IC(2)*: -0.508	IC(2)*: -0.477	-0.02	
118) Avg hr earnings constr.	0.10	-0.15	-0.13	0.07	-0.12	0.04		0.07	-0.03
119) Avg hr earnings manuf.	0.30	-0.04	0.31	-0.03	-0.21	0.03		0.11	-0.00
120) U. Of mich. Index	-0.67	-0.23	0.12	0.00	0.11	0.03	0.23	-0.12	0.02

The consequence of imposing more and more over-identifying loadings restrictions is shown in this table and explained in details in the text. 23 blocks of restrictions of varying sizes are imposed sequentially. Each block is denoted by a number in a bracket in the upper right corner of the rectangles. For each added set of restrictions we estimate the model and calculate the mean adjusted R-squared $\left(\bar{R}_a^2\right)$, AIC and SIC as well as the IC_{p2} panel information criteria from Bai & Ng (2002) and IC_{p2}^* which takes into account the convergence rate of Doz et al. (2006).

Table 2.2: Estimated factor loadings.

Variable Names	π	u	y	c	hrs	h	$pcom$	i	s	R^2
1) IP: products, total			**0.92**					−0.01		79.7
2) IP: final products			**0.87**					0.01		70.5
3) IP: consumer			**0.75**					−0.02		53.0
4) IP: durable cons.			**0.72**					0.00		47.0
5) IP: nondur. cons.			**0.43**					−0.04		16.4
6) IP: bus. Equip			**0.71**					0.01		46.5
7) IP: intermediate			**0.73**					−0.05		50.8
8) IP: materials			**0.87**					0.01		76.5
9) IP: durable goods			**0.87**					0.04		74.8
10) IP: nondur. goods			**0.40**					−0.04		14.7
11) IP: manufacturing			**1.01**					0.01		97.4
12) IP: dur. manuf			**0.97**					0.02		90.9
13) IP: nondur. manuf.			**0.70**					−0.04		46.5
14) IP: mining			**0.23**					0.03		3.5
15) IP: utilities			**0.12**					**−0.07**		0.5
16) IP: total index			1							96.4
17) Capacity util rate		**−0.74**	**0.16**		**0.25**			**0.19**		72.7
18) Pmi			**0.50**		**0.24**			**−0.11**		36.6
19) NAPM prod.			**0.54**		**0.14**			**−0.20**		41.1
20) Pers. Income			**0.31**					−0.05		8.4
21) Pers. Inc. - trans.			**0.54**					−0.05		27.4
22) Help-wated		0.03	**0.44**		−0.01			**−0.13**		20.3
23) Empl. Help-wanted		**−0.71**	0.01		**0.31**			**0.33**		67.7
24) Civ. Labor: empl.,		−0.06	**0.39**					0.04		13.2
25) Civilian labor: empl.,		**−0.10**	**0.43**					0.03		16.6
26) Unemp rate: all		1								72.7
27) Unemp dur: mean		**0.62**	**0.22**					**−0.29**		42.4
28) Unemp dur. < 5 w.		**0.71**	−0.04					**0.28**		75.4
29) Unemp dur. 5-14 w		**0.79**	−0.02					**0.11**		73.3
30) Unemp dur. 15+ w		**0.80**	**0.13**					**−0.06**		66.1
31) Unemp dur. 15-26 w		**0.82**	0.06					−0.01		70.4
32) Nonag payrl.: total		**−0.21**	**0.73**					**0.05**		54.6
33) Nonag payrl.: total,		**−0.16**	**0.76**					**0.08**		55.6
34) Nonag payrl.: goods		**−0.18**	**0.81**					**0.05**		64.8
35) Nonag payrl.: mining		**−0.11**	**0.18**					**0.17**		3.5
36) Nonag payrl.: contrct		−0.04	**0.36**					−0.04		12.0
37) Nonag payrl.: manuf		**−0.18**	**0.81**					0.05		65.3
38) Nonag payrl.: durable		**−0.18**	**0.80**					**0.07**		63.0
39) Nonag payrl.: nondur		**−0.11**	**0.56**					−0.03		31.7
40) Nonag payrl.: service		**−0.24**	**0.38**					0.03		19.1
41) Nonag payrl.: trans.		**−0.07**	**0.14**					0.04		0.6
42) Nonag payrl.: sale		**−0.12**	**0.42**					0.04		17.2
43) Nonag payrl.: finance		**−0.19**	**0.21**					**0.11**		6.1
44) Nonag payrl.: services		**−0.15**	**0.33**					**0.10**		10.7
45) Nonag payrl.: gov.		**−0.25**	−0.02					**−0.13**		8.6
46) Avg. wkly hrs. prod					**0.97**			**−0.15**		87.6
47) Avg. wkly overtime					1					93.0
48) NAPM Empl. Index		**−0.53**	**0.48**					0.07		52.1
49) Pers cons exp: total				1						66.6
50) Pers cons exp: tot.				1				0.05		94.4
51) Pers cons exp: nondur.				1				**0.07**		9.9
52) Pers cons exp: services				**0.16**				**−0.09**		1.7
53) Pers cons exp: new cars				**1.04**				**0.10**		84.9
54) Housing starts: n'farm						1				92.5
55) Housing starts: N.E						**0.48**		**−0.21**		28.6
56) Housing starts: M.W						**0.57**		**−0.38**		50.3
57) Housing starts: S						**0.94**		**0.27**		85.5
58) Housing starts: S						**0.85**		0.00		69.3
59) Housing auth. Tot new						**1.00**		0.06		95.1
60) Mobile homes						**0.61**		**0.32**		41.1

The factors are denoted by the symbols $\{\pi, u, y, c, hrs, h, pcom, i, s\}$ and describe general inflation, unemployment, economic activity (growth), consumption growth, hours in production, residential investments, commodity pric inflation, federal funds rate and stock markets returns respectively. R^2 denotes R-squared. Coefficients in bold are statistically significant at the 5% level (the standard errors are two-sided finite difference approximations of the gradient of the likelihood function.

Table 2.2 continued

Variable Names	π	u	y	c	hrs	h	$pcom$	i	s	R^2
61) NAPM inventories	0.06	**−0.39**	**0.08**	0.00	0.01	**0.27**	**0.16**	0.04	−0.05	44.7
62) NAPM new orders	**−0.09**	**0.08**	**0.39**	−0.01	0.07	**0.26**	**0.30**	**−0.29**	0.00	60.1
63) NAPM vendor deliv.	0.04	**−0.29**	**0.15**	−0.01	**0.26**	**0.21**	**0.15**	**0.14**	**−0.09**	45.3
64) New orders: cons goods	0.00	0.08	**0.50**	**0.13**	−0.06	0.00	−0.01	**−0.07**	0.07	28.1
65) New orders: nondefense	0.04	0.01	0.06	**0.12**	0.03	0.03	0.03	−0.05	0.00	1.1
66) NYSE: composite									1	97.5
67) SP500 composite								0.00	**1.01**	100.0
68) SP500 industrials								0.00	**1.00**	98.7
69) SP500 capital								−0.02	**0.92**	82.7
70) SP500 utilities								0.01	**0.61**	35.4
71) SP500: dividend	**0.08**	**0.30**	−0.02	0.00	**−0.49**	**−0.09**	**0.32**	**0.31**	−0.03	80.4
72) SP500: price earnings	−0.06	**−0.17**	0.01	−0.01	**0.48**	0.05	**−0.30**	**−0.37**	0.00	69.3
73) FX : Switzerland	**−0.08**	−0.02	**0.17**	**0.08**	0.06	**−0.19**	**0.08**	**0.14**	**0.07**	4.4
74) FX : Japan	**−0.12**	**−0.13**	**0.09**	0.00	0.01	**−0.18**	0.06	**0.19**	−0.04	4.5
75) FX : U.K	**0.10**	−0.03	**−0.16**	−0.05	0.03	**0.13**	−0.01	**−0.15**	0.01	3.1
76) FX : Canada	−0.01	0.07	**0.13**	0.04	−0.01	−0.04	0.01	−0.02	**−0.24**	5.3
77) Federal funds								1		100.0
78) US Tbill, 3m.	**−0.06**	**0.12**	0.02	−0.01	0.03	−0.03	**0.12**	**0.96**	0.01	98.2
79) US Tbill, 6m.	**−0.08**	**0.16**	0.01	−0.01	0.03	−0.03	**0.17**	**0.94**	0.00	98.7
80) Tbond const 1yr.	**−0.12**	**0.24**	0.01	−0.02	0.04	−0.04	**0.23**	**0.91**	0.00	99.0
81) Tbond const 5yr.	**−0.17**	**0.54**	−0.02	−0.02	**0.10**	−0.02	**0.28**	**0.76**	−0.01	100.0
82) Tbond const 10yr.	**−0.14**	**0.62**	−0.02	−0.02	**0.13**	0.00	**0.26**	**0.69**	−0.01	99.7
83) Bond yield: AAA	**−0.06**	**0.65**	**−0.05**	−0.02	**0.19**	**0.07**	**0.12**	**0.65**	−0.02	100.0
84) Bond yield: BAA	**−0.06**	**0.65**	**−0.05**	−0.01	**0.11**	**0.08**	**0.10**	**0.64**	0.00	99.7
85) Spread 3m − FF	**−0.21**	**0.38**	**0.07**	**−0.03**	**0.10**	**−0.09**	**0.39**	**−0.88**	**0.04**	80.2
86) Spread 6m − FF	**−0.24**	**0.47**	**0.04**	**−0.03**	**0.08**	**−0.09**	**0.50**	**−0.92**	0.01	88.5
87) Spread 1y − FF	**−0.38**	**0.75**	**0.02**	**−0.05**	**0.11**	**−0.12**	**0.70**	**−0.79**	0.01	90.3
88) Spread 5y − FF	**−0.29**	**0.93**	−0.03	−0.04	**0.18**	−0.04	**0.49**	**−0.85**	−0.01	100.0
89) Spread 10y − FF	**−0.21**	**0.93**	**−0.03**	**−0.03**	**0.19**	0.00	**0.40**	**−0.87**	−0.01	99.3
90) Spread AAA − FF	**−0.09**	**0.91**	**−0.07**	−0.02	**0.26**	**0.10**	**0.17**	**−0.85**	−0.02	100.0
91) Spread BAA − FF	**−0.09**	**1.00**	**−0.08**	**−0.02**	**0.17**	**0.12**	**0.15**	**−0.72**	−0.01	99.2
92) Money stock: M1	**0.17**	**0.31**	−0.05	**0.08**	**−0.21**	**0.29**	**−0.08**	**−0.13**	0.05	21.7
93) Money stock: M2	0.02	0.03	0.00	0.03	**−0.59**	**0.51**	**−0.14**	0.02	0.04	38.7
94) Money stock: M3	0.03	**−0.12**	−0.04	0.06	**−0.44**	**0.59**	−0.07	**0.18**	0.06	35.5
95) Money supply—M2(92)	**−0.53**	−0.01	0.02	0.03	**−0.44**	**0.40**	**−0.13**	0.00	0.04	52.9
96) Monetary base	**0.25**	**0.25**	−0.04	0.01	**0.14**	**0.23**	**−0.13**	−0.05	0.02	13.9
97) Depository inst res	0.04	**0.16**	0.02	−0.06	**−0.21**	**0.17**	**−0.09**	−0.05	0.00	8.2
98) Dep. Inst. Res..	**0.10**	**0.07**	**−0.15**	−0.01	**−0.16**	**0.07**	**−0.18**	**−0.09**	0.06	10.0
99) Comm. and indust. L	**−0.24**	**−0.22**	0.03	0.03	**0.19**	**−0.08**	**0.23**	**0.31**	0.02	19.2
100) Wkly rp lg com.	**−0.13**	0.02	0.03	−0.02	**0.34**	**−0.06**	**0.22**	**0.23**	**0.10**	14.5
101) Cons credit outst.	**−0.21**	−0.06	0.02	0.05	**−0.09**	**0.36**	**0.29**	**0.08**	−0.03	29.0
102) NAPM comm. prices							1			38.0
103) PPI: finished	**0.79**						0.03	**−0.12**		52.0
104) PPI: finished	**0.76**						0.05	**−0.17**		46.5
105) PPI: intermed							**0.28**	**0.23**		17.2
106) PPI: crude							**0.20**	−0.01		3.6
107) Index of sensitive mat.							**0.33**	**−0.14**		12.4
108) CPI-U: all items	1									95.6
109) CPI-U: apparel,	**0.44**							−0.02		16.3
110) CPI-U: transport.	**0.85**							**−0.20**		52.5
111) CPI-U: medical	**0.23**							**0.41**		32.0
112) CPI-U: comm.	**1.02**						0.06	**−0.21**		85.8
113) CPI-U: durables	**0.58**							**0.11**		39.5
114) CPI-U: services	**0.51**							**0.33**		55.1
115) CPI-U: less food	**0.85**							**0.10**		79.5
116) CPI-U: less shelter	**1.01**							**−0.09**		88.1
117) CPI-U: less medical	**1.00**							−0.02		93.6
118) Avg hr earn. constr.	**0.10**	**−0.15**	**−0.13**	0.07	**−0.12**	0.04		0.07	−0.03	5.2
119) Avg hr earn. manuf.	**0.30**	−0.04	**0.31**	−0.03	**−0.21**	0.03		**0.11**	0.00	22.3
120) Consumer expec.	**−0.67**	**−0.23**	**0.12**	0.00	**0.11**	0.03	**0.23**	**−0.12**	0.02	67.2

The factors are denoted by the symbols $\{\pi, u, y, c, hrs, h, pcom, i, s\}$ and describe general inflation, unemployment, economic activity (growth), consumption growth, hours in production, residential investments, commodity pric inflation, federal funds rate and stock markets returns respectively. R^2 denotes R-squared. Coefficients in bold are statistically significant at the 5% level (the standard errors are two-sided finite difference approximations of the gradient of the likelihood function.

Table 2.3: Forecast error variance due to monetary policy shocks.

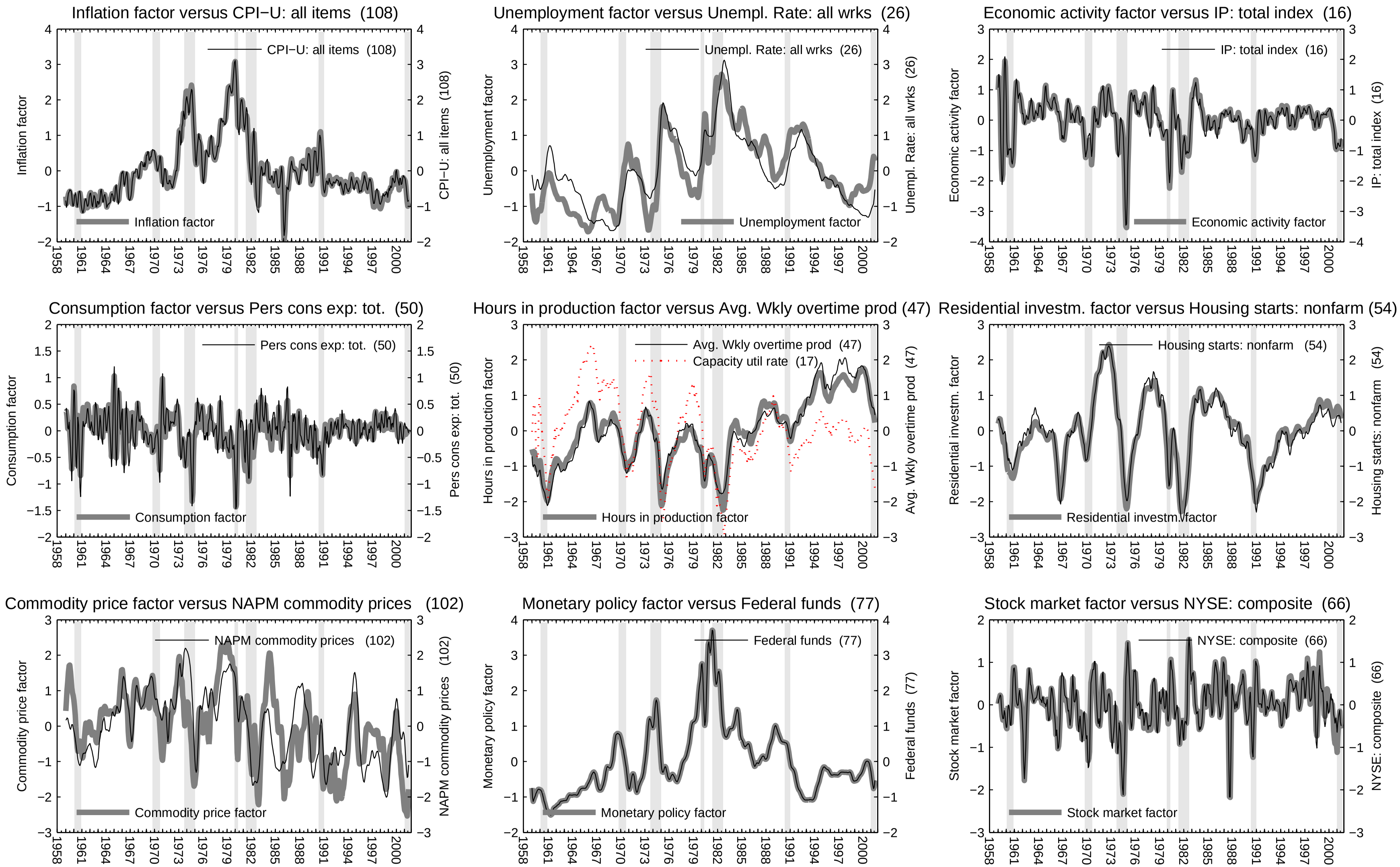
Average (all variables)	π	u	y	c	hrs	h	$pcom$	i	s	*total*	*Idio.*
6 month	0.03	0.05	0.06	0.03	0.07	0.04	0.04	**0.05**	0.04	0.41	0.59
12 month	0.04	0.05	0.06	0.03	0.09	0.06	0.04	**0.05**	0.05	0.46	0.54
24 month	0.04	0.05	0.06	0.03	0.10	0.08	0.03	**0.05**	0.06	0.50	0.50
60 month	0.06	0.05	0.06	0.03	0.10	0.11	0.03	**0.04**	0.06	0.53	0.47

12 month horizon	π	u	y	c	hrs	h	$pcom$	i	s	*total*	*Idio.*
77) Federal funds rate	0.02	0.06	0.06	0.03	0.41	0.15	0.04	**0.16**	0.08	1.00	0.00
16) IP: totalindex	0.05	0.21	0.32	0.01	0.12	0.09	0.02	**0.09**	0.05	0.95	0.05
108) CPI-U: all items	0.37	0.03	0.03	0.03	0.24	0.11	0.04	**0.03**	0.01	0.91	0.09
78) US Tbill, 3m.	0.03	0.03	0.04	0.02	0.38	0.15	0.09	**0.12**	0.08	0.94	0.06
81) Tbond const 5yr.	0.06	0.02	0.00	0.01	0.31	0.12	0.34	**0.08**	0.04	1.00	0.00
96) Monetary base	0.02	0.01	0.01	0.00	0.00	0.00	0.01	**0.00**	0.00	0.06	0.94
93) Money stock: M2	0.02	0.01	0.01	0.00	0.12	0.03	0.04	**0.01**	0.03	0.25	0.75
74) FX:Japan	0.01	0.00	0.01	0.00	0.01	0.00	0.01	**0.00**	0.01	0.05	0.95
102) NAPM comm prices	0.02	0.01	0.01	0.01	0.04	0.10	0.28	**0.05**	0.02	0.54	0.46
17) Capacity util rate	0.02	0.07	0.04	0.01	0.10	0.09	0.00	**0.10**	0.10	0.52	0.48
49) Pers cons : total	0.02	0.01	0.02	0.57	0.01	0.02	0.01	**0.02**	0.00	0.69	0.31
50) Pers cons : tot. dur	0.02	0.02	0.02	0.73	0.01	0.02	0.02	**0.02**	0.01	0.87	0.13
51) Pers cons : nondur.	0.01	0.01	0.01	0.35	0.01	0.01	0.01	**0.01**	0.00	0.41	0.59
26) Unempl.Rate: all	0.01	0.15	0.04	0.01	0.04	0.06	0.00	**0.09**	0.06	0.45	0.55
48) NAPM Empl. Index	0.02	0.06	0.09	0.00	0.06	0.06	0.00	**0.05**	0.05	0.40	0.60
118) Avg hr earn. constr.	0.00	0.01	0.01	0.00	0.00	0.00	0.00	**0.00**	0.00	0.03	0.97
54) Housing starts: n'farm	0.01	0.01	0.07	0.01	0.24	0.39	0.03	**0.10**	0.01	0.86	0.14
62) NAPM new orders	0.02	0.06	0.09	0.00	0.08	0.08	0.04	**0.11**	0.02	0.50	0.50
71) SP500: div. yield	0.02	0.00	0.02	0.02	0.06	0.02	0.21	**0.04**	0.01	0.40	0.60
120) Consumer expec.	0.20	0.01	0.03	0.01	0.07	0.03	0.02	**0.03**	0.03	0.43	0.57

60 month horizon	π	u	y	c	hrs	h	$pcom$	i	s	*total*	*Idio.*
77) Federal funds rate	0.10	0.04	0.04	0.05	0.19	0.38	0.06	**0.07**	0.07	1.00	0.00
16) IP: totalindex	0.05	0.19	0.29	0.01	0.14	0.12	0.01	**0.09**	0.07	0.96	0.04
108) CPI-U: all items	0.35	0.04	0.04	0.04	0.19	0.19	0.03	**0.05**	0.03	0.94	0.06
78) US Tbill, 3m.	0.11	0.03	0.03	0.05	0.19	0.37	0.08	**0.05**	0.06	0.98	0.02
81) Tbond const 5yr.	0.16	0.02	0.02	0.05	0.19	0.37	0.14	**0.03**	0.03	1.00	0.00
96) Monetary base	0.02	0.02	0.01	0.00	0.01	0.01	0.02	**0.00**	0.01	0.10	0.90
93) Money stock: M2	0.02	0.01	0.01	0.01	0.13	0.06	0.05	**0.02**	0.05	0.35	0.65
74) FX:Japan	0.01	0.00	0.01	0.00	0.02	0.00	0.01	**0.01**	0.01	0.07	0.93
102) NAPM comm prices	0.03	0.01	0.01	0.02	0.06	0.12	0.26	**0.05**	0.04	0.61	0.39
17) Capacity util rate	0.05	0.05	0.04	0.02	0.15	0.19	0.02	**0.11**	0.08	0.71	0.29
49) Pers cons : total	0.02	0.02	0.02	0.56	0.02	0.02	0.01	**0.02**	0.01	0.69	0.31
50) Pers cons : tot. dur	0.02	0.02	0.02	0.72	0.02	0.03	0.02	**0.03**	0.01	0.87	0.13
51) Pers cons : nondur.	0.01	0.01	0.01	0.35	0.01	0.01	0.01	**0.01**	0.00	0.42	0.58
26) Unempl.Rate: all	0.05	0.09	0.03	0.02	0.17	0.19	0.01	**0.10**	0.06	0.72	0.28
48) NAPM Empl. Index	0.03	0.05	0.08	0.01	0.11	0.10	0.01	**0.05**	0.05	0.50	0.50
118) Avg hr earn. constr.	0.01	0.01	0.01	0.01	0.01	0.02	0.00	**0.01**	0.00	0.07	0.93
54) Housing starts: n'farm	0.02	0.05	0.07	0.01	0.37	0.25	0.02	**0.08**	0.04	0.93	0.07
62) NAPM new orders	0.03	0.06	0.08	0.01	0.16	0.11	0.04	**0.10**	0.05	0.63	0.37
71) SP500: div. yield	0.16	0.01	0.02	0.06	0.08	0.22	0.19	**0.03**	0.01	0.78	0.22
120) Consumer expec.	0.22	0.02	0.03	0.03	0.11	0.17	0.02	**0.03**	0.02	0.66	0.34

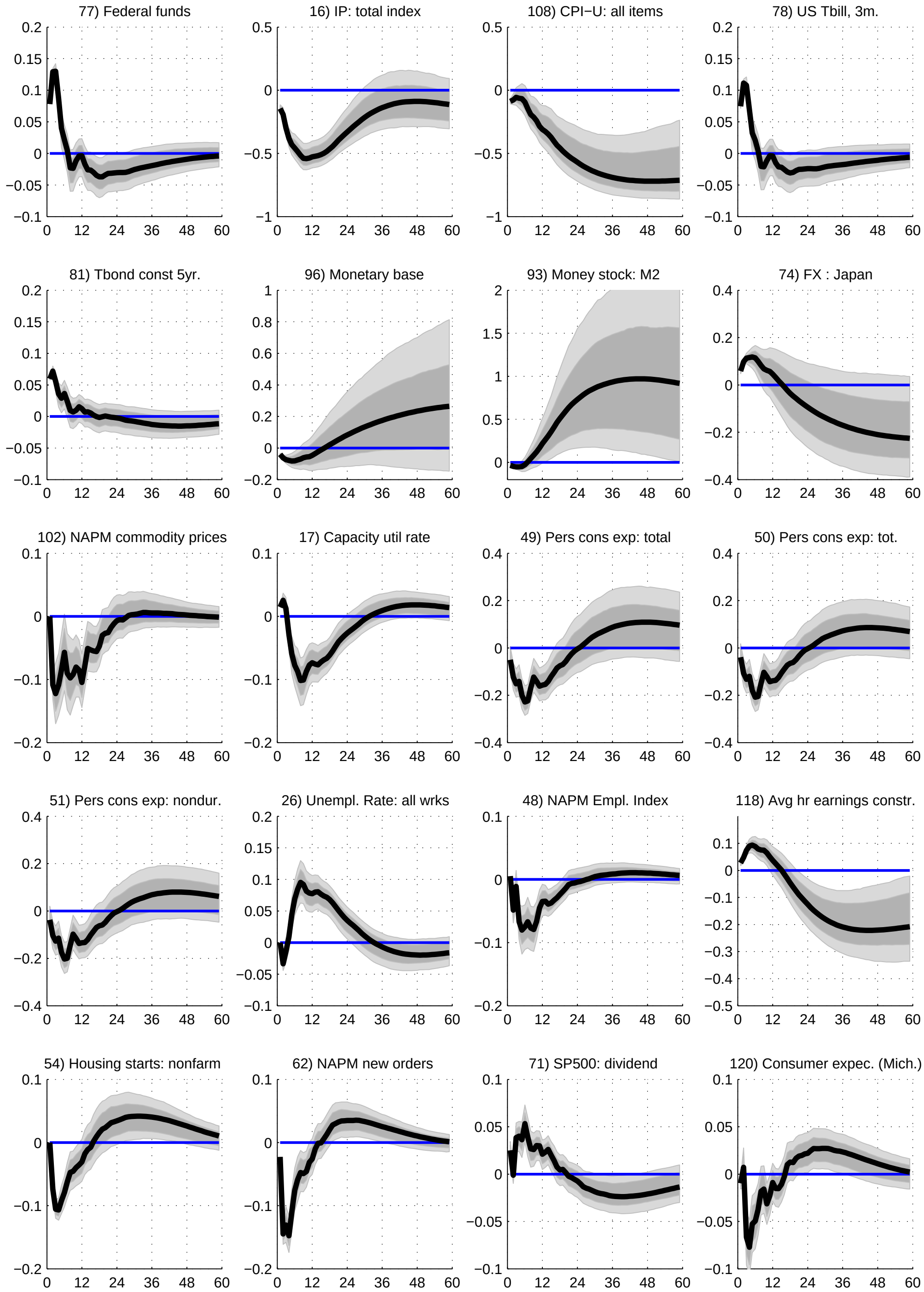
The upper panel illustrates the total fractions that the eight factors can explain of the forecast error variance on average for the panel at varying horizon. "Idio." means idiosyncratic variance. The factors are denoted by the symbols $\{\pi, u, y, c, hrs, h, pcom, i, s\}$ and describes general inflation, unemployment, economic activity (growth), consumption growth, hours in production, residential investments, commodity price inflation, federal funds rate and stock markets returns respectively. The middle and lower panel shows the 12 month ahead and 60 month ahead forecast error variance decomposition for key macroeconomic variables.

Figure 2.1: The time series of the factors versus related observed variables.



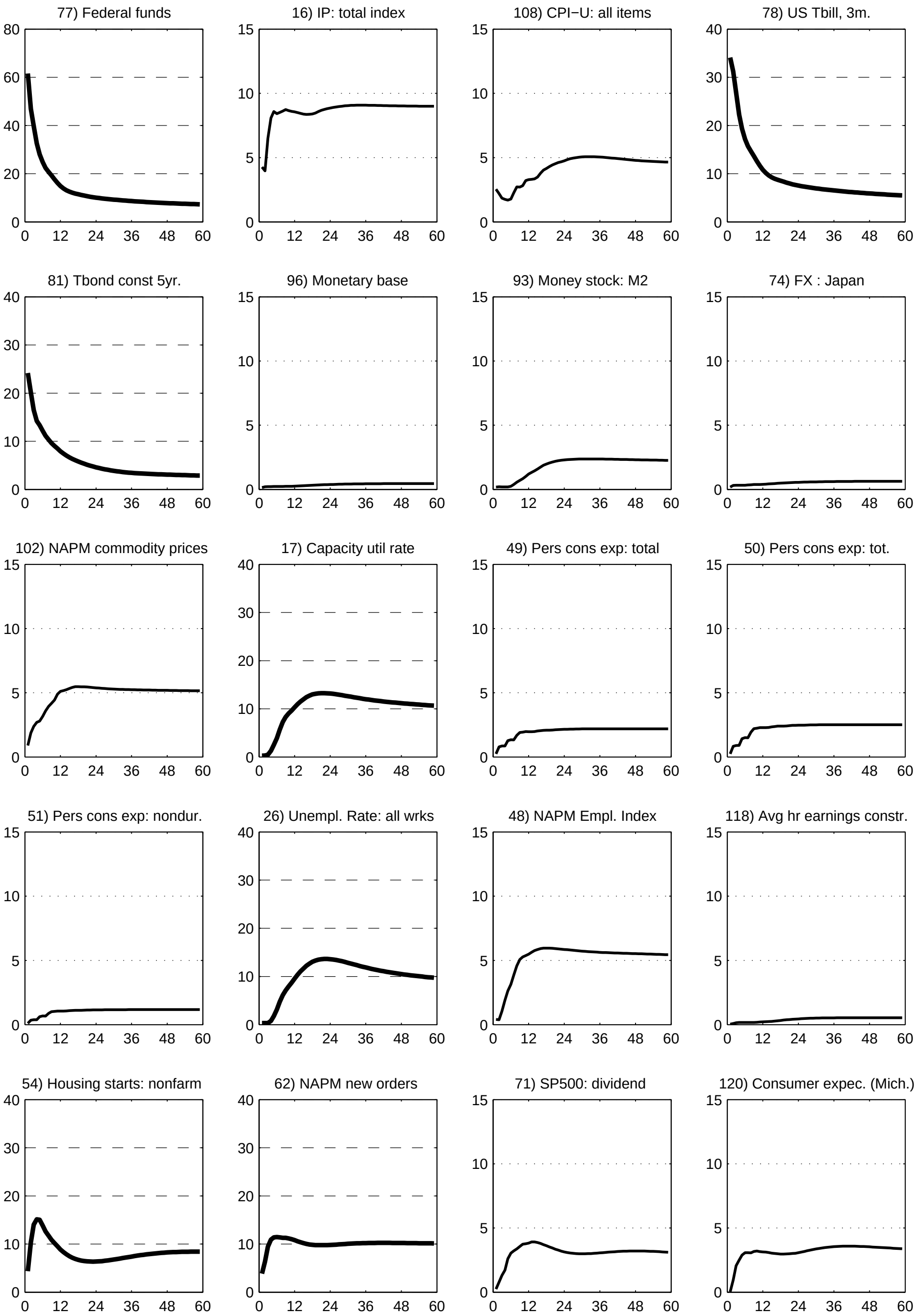
The shaded bars represent NBER-dated recessions. Numbers in parenthesis refer to the variable number in the panel; see the data appendix. To smooth series we have taken two-sided moving-averages of the original series.

Figure 2.2: Impulse responses to a 25 basis point monetary policy shock.



The figure illustrates the impulse responses in standard deviations of key macroeconomic variables following a 25 basis point monetary policy shock. The horizontal axis denotes the forecast horizon in months. Confidence intervals are represented by dark bands (68 percent) and light bands (95 percent).

Figure 2.3: Forecast error variance due to monetary policy shocks.



The figure plots the contribution of the monetary policy shock to the forecast error variance decomposition of key macroeconomic variables along the foreast horizon (the horizontal axis). Dashed gridlines indicate a larger scale compared to the dotted grid lines. Numbers in parenthesis refer to the variable number in the panel, see the data appendix.

CHAPTER 3

A Multifactor Affine Term Structure Model with Macroeconomic Factors from Large Panels

Lasse Bork[†]

January 2010

[†]Finance Research Group, Department of Business Studies, Aarhus School of Business, University of Aarhus, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. Affiliated with CREATES at the University of Aarhus, a research center funded by the Danish National Research Foundation. Email: lbo@lassebork.dk.

Abstract[*]

The bond market is filtering an abundant amount of information in the process of assessing the current state of the economy and its implications for bond pricing and bond risk premia. I propose to solve the filtering problem by a dynamic factor analysis of a large panel of US macroeconomic and financial time series to derive a small set of macroeconomic state variables. A discrete-time dynamic term structure model is then augmented with these filtered macroeconomic state variables. A forecast error variance decomposition shows that shocks to inflation and in particular unemployment are important for the risk premia on long-term bonds.

JEL classifications: C13, C32, C33, E43, E44, E52
Keywords: Monetary policy, Discrete-time Affine Term Structure Models, Financial markets and the macroeconomy, macroeconomic factors, Kalman filter.

[*] I thank Tom Engsted, Hans Dewachter and Claus Bajlum for helpful comments. Any remaining errors are my own.

3.1 Introduction

The affine class of dynamic term structure models proposed by Duffie & Kan (1996) and generalized by Dai & Singleton (2000) has been successful in modeling the evolution of bond yields linearly in typically two or three latent state variables that evolve over time according to some specified law of motion. However, given the purely latent nature of the state variables, these models offer little economic insight into the underlying driving forces of the yield curve.

However, from an economic point of view a macroeconomic underpinning of the state variables is preferred. In particular, dynamic term structure models should reflect how central banks implement their monetary policy through the adjustment of the short term interest rate controlled by the central bank. Being an important regulator of the economy, the economic determinants of the monetary policy rate are of central interest in macroeconomics and in particular within the field of monetary economics. Questions of what should be and appears to be the economic determinants of the monetary policy rate have been discussed in a large volume of papers and in book lengths[1]. In a widely cited paper Taylor (1993) estimates a remarkably simple empirical monetary policy rule as a linear function of the deviation of current inflation from an inflation target and the deviation from current GDP from the potential GDP (output gap). Intuitively, the central bank "leans against the wind" in the sense that the monetary policy rate is raised if economic activity expands beyond its natural or potential level or if inflation exceeds a desired rate of inflation or both.

However, macroeconomic influence is not limited to the short end of the yield curve. Some macroeconomic underpinning of the risk premia demanded for holding bonds of different time to maturity is also preferred as we would expect risk premia to be high at the trough of the business cycle and low at the peak of the business cycle.

To bridge no-arbitrage financial theory and macroeconomic theory, the recent and rapidly growing *"macro-finance"* literature integrates more or less structural macroeconomic models into no-arbitrage dynamic term structure models which in turn allow for a macroeconomic explanation of the dynamics of the yield curve including the monetary policy rate and the time-varying bond risk premia.

[1] A general treatment is found in Woodford (2003) and Walsh (2003) to mention only a few. Bernanke et al. (1999) discuss inflation targeting as the monetary policy strategy.

This paper contributes to the macro-finance literature by significantly *expanding the macroeconomic information set* used in the *affine* class of dynamic term structure model.

The main motivation for the use of an expanded information set is the fact that the financial markets monitor and respond to a large set of macroeconomic variables in the assessment of the current state of the economy. Therefore, including e.g. a single specific consumer price index and a single specific series for production (or unemployment) in a macro-finance term structure model may not carry enough information compared with the potential macroeconomic information embedded in bond prices. Furthermore, most macroeconomic series are prone to measurement errors implying that the financial markets filter key underlying economic concepts (like inflation) from many different sources (e.g. from a number of different price indices).[2]

To imitate the potential information set and solve the bond markets filtering problem, I propose a large panel dynamic factor analysis of a panel of 120 US macroeconomic and financial time series, from which key macroeconomic factors like inflation, production, and unemployment are filtered. Effectively controlling for the short-term interest rate in the dynamic factor analysis, I use these macroeconomic factors as observed state variables in an affine multi-factor Gaussian term structure model. This setup allows for an empirical analysis of the dynamic responses of the bond yields and bond risk premia (excess returns) to macroeconomic shocks.

The focus in this paper is on potential macroeconomic sources of variation in expected excess returns on bonds. An impulse response analysis of the model-implied expected excess return reveals that an inflation factor and an unemployment factor are the most important among five candidate macroeconomic factors. A one standard deviation shock to unemployment initially raises the expected excess return by 17 basis points on an annually basis for a five-year bond held for one year. The intuition is clear: risk premia are time-varying and counter-cyclical. Hence, in business cycle troughs we see rising unemployment and investors are demanding a higher risk premium to buy risky assets. Continuing with the same example, I find that a one standard deviation shock to inflation lowers the expected excess return by 9 basis points. Higher inflation increases the possibility that the Federal Reserve

[2]Notice that the same applies to central banks in the sense that central banks "monitor literally hundreds of economic variables in the process of policy formulation" as expressed in Bernanke et al. (2005).

Board leans against the wind and raises the interest rate. Open positions in long bonds would then probably lose money.

The findings are related to some of the existing literature as follows. Joslin et al. (2009) also consider an impulse response analysis of the excess bond returns in an affine term structure model. Their macroeconomic state variables are an individual industrial production series and an individual inflation series. The impulse responses in Joslin et al. (2009) and this paper are strikingly similar in terms of magnitude and form[3]. However, I find that excess return responds more to unemployment than industrial production. In contrast to Joslin et al. (2009) I examine longer holding-periods and find that the longer the holding-period the larger the response of expected excess return. Moreover, the longer the bond the larger the response. This insight conforms to the findings in Cochrane & Piazzesi (2005) where one-year horizon excess return regressions are the key to uncovering a single return-forecasting factor.

The unique feature of this paper is its focus on the response of excess returns, as implied by an affine term structure model, to shocks to large-panel dynamic macroeconomic factors.

Several other papers also analyze bond excess return but these papers do not entertain all three ingredients (excess returns, affine term structure model, large-panel dynamic macroeconomic factors). For the well informed in this literature, I use the five-factor Gaussian affine term structure model from Ang & Piazzesi (2003), replace their macroeconomic factor by large-panel dynamic macroeconomic factors as in Mönch (2008) and focus on model-implied bond excess returns as in Joslin et al. (2009). The following offers a brief introduction to closely related papers in the literature on bond excess returns.

Duffee (2007), Joslin et al. (2009) and Ludvigson & Ng (2008) also focus on bond risk premia. However, Duffee (2007) and Joslin et al. (2009) do not use large panel macroeconomic state variables in their affine multi-factor Gaussian term structure model whereas the large panel dynamic factors in Ludvigson & Ng (2008) are used in excess return regressions only and not in a dynamic term structure model.

Dynamic term structure models are used in Mönch (2008) and Ang & Piazzesi (2003) but they do not focus on bond risk premia. Mönch (2008) includes large

[3]That is, if I perform the impulse response analysis with similar state variables compared to Joslin et al. (2009) I get strikingly similar results.

panel dynamic factors but does not use latent term structure factors in his affine term structure model. Furthermore, Mönch (2008) relies on a two-step principal component method to extract the dynamic factors whereas a fully parametric one-step iterative maximum likelihood method is used in this paper to estimate the factors.

Finally, a distinguishing econometric feature of this paper is the recurring use of the Kalman filter to estimate the large panel dynamic factors and the affine term structure model.[4] Having stated how this paper differs from the most closely related papers the following now contains a brief summary of these papers as well as other papers in the macro-finance literature.

In the seminal paper by Ang & Piazzesi (2003) a standard three-factor affine term structure model is augmented with two macroeconomic state variables and they find that bond yields respond significantly to shocks to these state variables. However, the three latent factors continue to play an important role in the variation of the long bond yields. Mönch (2008) examines the forecasting power of multifactor dynamic term structure models where the state variables include the monetary policy rate as well as factors derived from large panel principal component methods. These factors are shown to have good forecasting properties but the factors lack a well-defined economic interpretation as opposed to this paper where the economic interpretation of the factors is obtained by means of a set of overidentifying restrictions. Ludvigson & Ng (2008) use large panel dynamic factor analysis to obtain dynamic factors which are subsequently used as explanatory variables in excess return regressions. They find that dynamic factors which are correlated with measures of inflation and with measures of real output and employment are the key to explain cyclical variation in bond risk premia. However, the previous critique with respect to economic interpretability also applies here. Still in the context of bond excess returns, Cochrane & Piazzesi (2005) find impressive forecasting properties of a tent-shaped combination of forward rates. Cochrane & Piazzesi (2009) analyze in an affine term structure model how much of a given yield curve that corresponds to expectations of future interest rates, and how much that corresponds to bond risk premia.

On a more general level, this paper resides in the above-mentioned macro-finance research area that was pioneered by the work of Ang & Piazzesi (2003), Dewachter

[4]Duffee (2007) also use the Kalman filter to estimate his dynamic term structure model.

et al. (2006) and Dewachter & Lyrio (2006*b*). These papers are characterized by the inclusion of macroeconomic variables among the state variables in dynamic no-arbitrage *affine* term structure models[5]. However, real structural macroeconomic theory in these models commenced with the papers by Hördahl et al. (2006), Bekaert et al. (2005) and Wu (2006) in which New Keynesian macroeconomic models are integrated with affine term structure models. Recently, also learning theory has been introduced into macro-finance term structure models by Dewachter & Lyrio (2006*a*), Laubach et al. (2006) and Dewachter (2008) where agents learn about the state of the economy for instance in terms of the inflation target, the long-run inflation or the real interest rate. This approach seems promising in generating persistent state variables, which is a decisive for empirical term structure models; in particular for the long end of the yield curve.

This paper is organized as follows. Section 3.2 presents the discrete-time Gaussian affine term structure model and a variant of a dynamic factor model that allows me to derive a set of macroeconomic state variables while controlling for the interest rates. Section 3.3 addresses identification and estimation issues in both models which in turn allows for an empirical application with respect to yield curve modeling using macroeconomic state variables filtered from a large panel of US data. Section 3.5 concludes by summarizing the main findings of this paper.

3.2 The modeling framework

Two dynamic models for panel data are used in this paper and now presented in turn. As a *first step,* I extract a few dynamic macroeconomic factors from a large panel of macroeconomic and financial time series, which represent the large information set of the bond traders. To do this, the factor-augmented VAR model of Bernanke et al. (2005) is used, which is a variant of a dynamic factor model that effectively controls for the short-term interest rate.

The identification and dynamic interaction of the short-term interest rate with the macroeconomic factors imply some advantages compared to the existing approaches in the dynamic factor analysis literature. Firstly, I do not want an interest

[5]Yet another interesting branch in the macro-finance literature has evolved around dynamic extensions of the parametric Nelson-Siegel yield curve model. See Diebold et al. (2005), Diebold et al. (2006), Coroneo et al. (2008) and Christensen et al. (2009).

rate factor in disguise to explain the yield curve, so the macroeconomic dynamic factors are carefully estimated under an identification procedure similar to Bernanke et al. (2005) that ensures that each factor is not an interest rate factor. Accordingly, the short-term interest rate is explicitly modeled and identified as an observed factor that interacts dynamically with the macroeconomic factors. Secondly, yield to redemption on *coupon* bonds are not excluded from the panel as in Mönch (2008); on the contrary all redemption yields are included as a subset of financial market variables based on the idea that financial variables serve as timely information variables for macroeconomic variables. To accommodate potential concern that one of the factors may be a "redemption yield factor" in disguise it should be noted that the redemption yields are highly correlated with the identified short-term interest rate such that there is little need for a separate yield factor. Furthermore, the estimated macroeconomic factors are shown empirically to be well in line with leading macroeconomic measures of US economy. Details about the factor-augmented VAR (FAVAR) are presented in section 3.2.2.

In the *second step,* the dynamic macroeconomic factors are used as state variables in a dynamic no-arbitrage Gaussian multifactor term structure model to explain a panel of US bond yields. The standard Gaussian affine term structure model is presented first in section 3.2.1 in terms of a general state vector, which is responsible for the dynamic evolution of the yield curve. The details of the partitioning of the state vector into latent state variables and macroeconomic state variables is postponed until the econometric formulation of the model in section 3.3, as I find it more natural to begin to partition things there. Until then, it is sufficient to think about the no-arbitrage term structure model as driven by both latent and macroeconomic state variables. Finally, this section ends with a brief introduction to the factor-augmented VAR model, which delivers the macroeconomic state variables.

3.2.1 Gaussian multifactor affine term structure model

In this section I present a discrete time multifactor affine term structure model (ATSM)[6] where the dynamics of the yield curve are explained in terms of a small set of latent variables.

[6] Backus et al. (1996), Backus et al. (1998) and Piazzesi (2009) present multifactor ATSMs in discrete time whereas Dai & Singleton (2000) and Singleton (2006) among many other papers present ATSMs in continuous time. Lund (1997) presents the algebra in getting from the continuous time form to the discrete time representation.

In the end the no-arbitrage ATSM is written in state-space form so the following presentation will take an unrestricted state space form as a simple starting point and later impose the no-arbitrage restrictions. Denote by y_t^n the yield of a default-free zero-coupon bond, which always has n periods to maturity at any time t. Stack $\mathcal{N}$ of these yields varying in terms of n in the $\mathcal{N} \times 1$ vector Y_t. The dynamics of the $\mathcal{N}$ yields are described linearly in terms of a small set of $K < \mathcal{N}$ dynamic latent state variables X in accordance with the empirical findings of Litterman & Scheinkman (1991), where $\mathcal{N} = 3$ latent variables (factors) can explain the vast majority of the variation in the yields. The law of motion of the unobserved X is assumed to be described in terms of a first-order autoregressive system and represents the state transition equation of the state-space system whereas the observed yields represent the observation equation in (3.1) below:

$$Y_t = A + BX_t + v_t \tag{3.1}$$

$$X_t = \mu + \Phi X_{t-1} + \Sigma \varepsilon_t \tag{3.2}$$

where A is an $\mathcal{N} \times 1$ vector of constants, B is an $\mathcal{N} \times K$ matrix that allows the $\mathcal{N}$ different yields in the observation equation to load with different weights on the K state variables in X, μ is a $K \times 1$ vector of constants, Φ is a $K \times K$ matrix that contains the autoregressive parameters where stationarity of the system implies that the eigenvalues of Φ are less than one in modulus, ε_t is a $K \times 1$ vector with zero mean and unit variance, Σ is a lower triangular $K \times K$ matrix that is the result of a Cholesky decomposition of $\Sigma \varepsilon_t \sim N(0, \Omega)$, where $\Omega = \Sigma\Sigma^\top$. Although more will be said about this state-space system, the two equations in (3.1)-(3.2) illustrate in a simple way how the time evolution of the yield curve is analyzed in terms of a few driving forces, X, and how the dynamic response of a particular yield to shocks (ε) to the driving forces can be analyzed within the same framework. The measurement errors v_t are assumed to be cross-sectionally independent Gaussian white noises, i.e. $v_t \sim N(0, R)$, with R being an $\mathcal{N} \times \mathcal{N}$ diagonal matrix. Variations about the distributional assumptions in the literature for R are discussed in the empirical section.

The state space system in (3.1)-(3.2) is too general to be econometrically identified, and it does not rule out arbitrage opportunities among the yields included in Y. No-arbitrage is the fundamental building block in every standard asset pricing model within finance and dates back to path-breaking contributions by Black &

Scholes (1973), Merton (1973) and Harrison & Kreps (1979), to mention just a few important papers. The idea is that, under the assumption of no-arbitrage, there exists a risk-neutral measure $\mathbb{Q}$ under which we can calculate the price P of an asset as the discounted expected value of the payoff of the particular asset using the risk-free rate i as the discount rate. Duffie & Kan (1996) apply the notion of no-arbitrage to multifactor ATSMs driven by latent yield curve factors and characterize the class of ATSMs formally. Specifically, they demonstrate that if the bond price is exponential affine in the state variables X, then the drift and volatility of the state variables are also affine[7] and A and B in (3.1) must obey a set of recursive restrictions. Before the exact no-arbitrage cross-section restrictions on A and B can be stated, the following definition presents briefly the necessary assumptions needed to derive these restrictions and set up the ATSM:

Definition 1 (Gaussian ATSM) *The Gaussian ATSM in discrete time is constructed by the following three ingredients:*

1. *The one-period interest rate, i_t, is affine in the K-dimensional vector of state variables X_t*
$$i_t = \delta_0 + \delta_1^\top X_t \tag{3.3}$$
where δ_0 is a scalar and δ_1 is a $K \times 1$ vector.

2. *The dynamics of the state variables is given by a VAR(1):*[8]
$$X_t = \mu + \Phi X_{t-1} + u_t; \quad u_t \sim N(0, \Omega) \tag{3.4}$$
where the conditional mean is $\mathbb{M}_{t-1} = \mu + \Phi X_{t-1}$. The covariance matrix Ω is Cholesky decomposed into $\Omega = \Sigma\Sigma^\top$ such that $u_t = \Sigma\varepsilon_t$ where $\varepsilon_t \sim N(0, I)$ follows from the ATSM.[9]

3. *The assumption of no-arbitrage guarantees the existence of a pricing kernel. Specifically, the (nominal) bond pricing kernel $\frac{M_{t+1}}{M_t}$ is given by*
$$\frac{M_{t+1}}{M_t} = \exp\{-i_t\} \left(\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}}\right)^{\mathcal{D}}_{t,t+1}$$

[7] In fact, the converse also holds, i.e. if the drift, the volatility and the short rate are affine in X then the price is exponential affine in X; cf. the proposition in Duffie & Kan (1996).

[8] A VAR with p lags can be encompassed in a VAR(1) by a square companion matrix.

[9] Specifically, the discrete time representation of the continuous time diffusion process for X involves an integration of Brownian motions which are normally distributed; cf. Lund (1997).

where $\left(\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}}\right)^{\mathcal{D}}_{t,t+1}$ *denotes the Radon-Nikodym derivative which links the conditional distributions of* X_{t+1} *under the risk-neutral measure* $\mathbb{Q}$ *and the data generating measure* $\mathbb{P}$[10] *and is characterized by*

$$\left(\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}}\right)^{\mathcal{D}}_{t,t+1} = \exp\left\{-\frac{1}{2}\lambda_t^{\top}\lambda_t - \lambda_t^{\top}\Sigma^{-1}\left[X_{t+1} - \mathbb{M}_t\right]\right\} = \exp\left\{-\frac{1}{2}\lambda_t^{\top}\lambda_t - \lambda_t^{\top}\varepsilon_{t+1}\right\}$$

where λ_t *is a* K*-dimensional vector of possible time-varying market prices of risk associated with shocks to the state variables; cf. Duffee (2002):*

$$\lambda_t = \lambda_0 + \lambda_1 X_t \tag{3.5}$$

Thus, the entire yield curve and its dynamics are characterized by (1) the functional relation between the short rate and the state variables, (2) the dynamics of the state variables, and (3) the risk premia specification. Essentially, these three ingredients specify a time-series process for the pricing kernel.

□

The discrete-time model setup in Definition 1 is quite standard and has been used in various forms in e.g. Ang & Piazzesi (2003), Duffee (2007, 2008), Ang et al. (2005) and Pericoli & Taboga (2008). Le et al. (2009) and Singleton (2006) characterize discrete-time ATSMs in terms of conditional characteristic (or moment generating) functions which are utilized below in characterizing the formal relation between moments under the risk-neutral measure and the data generating measure. In the literature, the market price of risk as specified in (3.5) is sometimes scaled with the inverse of Σ, which in turn affects the relation between μ and Φ under the two measures. However, if λ_t is expressed in terms of the μ's, Φ's and Σ, the same equation emerges irrespective of whether (3.5) is scaled by Σ^{-1} or whether μ and Φ is scaled:

$$\lambda_t = \Sigma^{-1}\left(\mu - \mu^{\mathbb{Q}}\right) + \Sigma^{-1}\left(\Phi - \Phi^{\mathbb{Q}}\right)X_t.$$

which follows from the formal relation between the conditional means $\mathbb{M}_t$ and $\mathbb{M}_t^{\mathbb{Q}}$ that can be calculated using the conditional moment-generating function and de-

[10]Whenever a superscript $\mathbb{Q}$ is used, this refers to a moment or parameter (like $\mu^{\mathbb{Q}}$) belonging to the risk-netural measure. To simplify the notation I do not use superscript $\mathbb{P}$ to denote moments or parameters under the data generating measure, for instance I do not use $\mu^{\mathbb{P}}$.

tailed in the Appendix A.1. In particular, the definition in (3.5) implies that

$$\begin{aligned}\mu &= \mu^{\mathbb{Q}} + \Sigma\lambda_0 \\ \Phi &= \Phi^{\mathbb{Q}} + \Sigma\lambda_1\end{aligned}$$

After these formal definitions, a natural starting point towards the no-arbitrage restrictions on A and B is the fundamental asset pricing equation $1 = E_t\left[\frac{M_{t+1}}{M_t}R_{t+1}\right]$, where R_{t+1} is the one period gross return for the particular asset. In the case of zero-coupon bonds this can be rewritten as:

$$P_{n+1,t} = E_t\left[\frac{M_{t+1}}{M_t}P_{n,t+1}\right] \tag{3.6}$$

where $P_{n+1,t}$ is the price at time t of a zero-coupon bond maturing in $n+1$ periods. However, this equation is still too general to be of any practical interest in pricing zero-coupon bonds, but a closed form equation can be derived as follows. With the insight from Duffie & Kan (1996) the bond price equation is proposed to be exponential affine in the state variables

$$P_{n,t} = \exp\left\{\mathcal{A}_n + \mathcal{B}_n^\top X_t\right\} \tag{3.7}$$

where $\mathcal{A}_n$ and $\mathcal{B}_n$ each depends on the maturity of the zero-coupon bond and each needs to satisfy recursive restrictions. Appendix A.2 contains a proof that the proposed bond price equation is compatible with the fundamental asset pricing equation in (3.6) and Definition 1. Moreover, the cross-sectional restrictions on $\mathcal{A}_n$ and $\mathcal{B}_n$ consistent with no-arbitrage are also specified. However, the measurement equation in (3.1) maps the *yields* to the state variables and not the prices as in (3.7) but the yield mapping in terms of A_n and B_n is easily found from the relation between the n-period zero-coupon bond yield $y_t^{(n)}$ and the price

$$\begin{aligned}y_t^{(n)} &= -\frac{\log P_{n,t}}{n} = -\frac{p_{n,t}}{n} \\ &= A_n + B_n^\top X_t \qquad (3.8)\end{aligned}$$

where the scalar $A_n = -\frac{\mathcal{A}_n}{n}$ and the $K \times 1$ vector $B_n = -\frac{\mathcal{B}_n}{n}$ is a straightforward application of the definition of $\mathcal{A}_n$ and $\mathcal{B}_n$ in Appendix A.2:

$$B_n = \frac{1}{n}\delta_1^\top \left(I - \Phi^{\mathbb{Q}}\right)^{-1} \left(I - \left[\Phi^{\mathbb{Q}}\right]^n\right) \tag{3.9}$$

$$\begin{aligned} A_n &= \delta_0 + \frac{\delta_1^\top}{n}\left[n \cdot I - \left(I - \Phi^{\mathbb{Q}}\right)^{-1}\left(I - \left[\Phi^{\mathbb{Q}}\right]^n\right)\right]\left(I - \Phi^{\mathbb{Q}}\right)^{-1}\mu^{\mathbb{Q}} \\ &\quad -\frac{1}{2n}\sum_{i=0}^{n-1} i^2 B_i^\top \Sigma\Sigma^\top B_i \end{aligned} \tag{3.10}$$

Referring to the initial state space system in (3.1) and (3.2) it is now possible to characterize this system as a no-arbitrage state space system if the cross-sectional restrictions in (3.9) and (3.10) are imposed. Specifically, A in (3.1) is replaced by a new $\mathcal{N} \times 1$ vector $A = [A_{n_1}, ..., A_{n_\mathcal{N}}]^\top$ in (3.10) and B is replaced by a new $\mathcal{N} \times K$ matrix $B = \left[B_{n_1}^\top, ..., B_{nN}^\top\right]^\top$ in (3.9). Notice that the components of the replaced A and B are highly nonlinear in the parameters of interest $\theta = \{\delta_0, \delta_1, \mu, \Phi, \Sigma, \lambda_0, \lambda_1\}$ and depend on the maturity n which is emphasized by writing $A = A(n, \theta)$ and $B(n, \theta)$.

The linear mapping between the yields and the dynamics of the state variables allows for an analysis of the dynamic response of the yield of any maturity to a shock ε to the state variables. For instance, given that one of the state variables is a time series of inflation, the model allows us to trace through time how e.g. the five-year yield responds to an inflation shock. The same type of analysis can be extended to bond returns. Specifically, I analyze in Section 3.4.3 how (expected) bond returns are affected by shocks to the state variables. For this reason, I derive the affine relation between the expected excess holding period bond returns and the state variables below.

Consider at time t, a buy-and-hold of an n-period zero-coupon bond for m periods. Sell this bond at $t + m$ which is now an $(n - m)$ period bond. Denote by $rx_{t,t+m}^{(n)}$ the resulting log return in excess of holding a bond for m periods[11] which is detailed in the Appendix A.3 to be

$$rx_{t,t+m}^{(n)} = p_{t+m}^{(n-m)} - p_t^{(n)} + p_t^{(m)}$$

where $p_{t+m}^{(n-m)}$ is the log price of an $n - m$ period bond at time $t + m$, etc. Inserting

[11] Ideally, for excess return calculations it may be preferred to roll over a one-period T-bill for m periods as the alternative to invest-and-hold for m periods a longer bond. However, this is approximated here by the m-period bond.

the appropriate $\mathcal{A}_n$ and $\mathcal{B}_n$ from (3.25) and (3.24) in Appendix A.2 yields:

$$\begin{aligned} rx_{t,t+m}^{(n)} &= \mathcal{A}_{n-m} - \mathcal{A}_n + \mathcal{A}_m + \mathcal{B}_{n-m}^\top X_{t+m} - \mathcal{B}_n^\top X_t + \mathcal{B}_m^\top X_t \\ &= a_{n,m} + b_{n,m}^\top X_t + \epsilon_{t+1,t+m}^{(n)} \end{aligned} \tag{3.11}$$

where $a_{n,m}$, $b_{n,m}$ and $\epsilon_{t+1,t+m}^{(n)}$ are specified in the appendix to be:

$$a_{n,m} = \mathcal{B}_{n-m}^\top \sum_{i=0}^{m-1} \Phi^i \mu + \mathcal{A}_{n-m} - \mathcal{A}_n + \mathcal{A}_m \tag{3.12}$$

$$b_{n,m}^\top = -\delta_1^\top \left(I - \Phi^{\mathbb{Q}}\right)^{-1} \left\{ \left(I - \left[\Phi^{\mathbb{Q}}\right]^{n-m}\right) \Phi^m + \left[\Phi^{\mathbb{Q}}\right]^n - \left[\Phi^{\mathbb{Q}}\right]^m \right\} \tag{3.13}$$

$$\epsilon_{t+1,t+m}^{(n)} = \mathcal{B}_{n-m}^\top \sum_{i=0}^{m-1} \Phi^i u_{t+m-i}$$

Hence, the expected excess return is also affine in the state variable X_t:

$$E_t\left[rx_{t,t+m}^{(n)}\right] = a_{n,m} + b_{n,m}^\top X_t \tag{3.14}$$

Notice that risk premia need to depend on the state variables if the excess return in ATSMs should be forecastable. That $\lambda_1 \neq 0$ is needed becomes particularly clear if a holding period of $m = 1$ is considered

$$\begin{aligned} b_{n,1}^\top &= -\delta_1^\top \left(I - \Phi^{\mathbb{Q}}\right)^{-1} \left\{ \left(I - \left[\Phi^{\mathbb{Q}}\right]^{n-1}\right) \Phi + \left[\Phi^{\mathbb{Q}}\right]^n - \Phi^{\mathbb{Q}} \right\} \\ &= -\delta_1^\top \left(I - \Phi^{\mathbb{Q}}\right)^{-1} \left\{ \left(I - \left[\Phi^{\mathbb{Q}}\right]^{n-1}\right) \left[\Phi^{\mathbb{Q}} + \Sigma\lambda_1\right] + \left[\Phi^{\mathbb{Q}}\right]^n - \Phi^{\mathbb{Q}} \right\} \\ &= \mathcal{B}_{n-1}^\top \Sigma \lambda_1 \end{aligned}$$

So far the state variables in X have been treated rather generically. A significant part of the empirical literature on ATSMs treats X as unobserved latent variables[12], i.e. X is an implicit function of the parameter vector that we choose such that the joint likelihood of the observed yields and state variables is maximized. Alternatively, some of the Xs may be observed as proposed by Ang & Piazzesi (2003) but this does not require a change of the theoretical model outlined above; only the econometric model outlined in section 3.3 is affected. The observed variables used in this paper are derived from a large panel of macroeconomic and financial

[12] Examples in the literature are Chen & Scott (1993), Duffie & Singleton (1997), Dai & Singleton (2000, 2002) and Duffee (2002)

time series and the following section presents the theoretical model for extracting dynamic macroeconomic factors from large panels.

3.2.2 Large panel factor analysis: A factor-augmented VAR

Recent advances in the econometric theory put forward by notably Forni et al. (2000) and Stock & Watson (2002*a*) allow us to analyze large panels of potentially hundreds of time series in terms of a few (< 10) dynamic factors[13]. As mentioned previously, the idea pursued here is to imitate the large information set of the bond traders by the large panel and then extract a few common dynamic macroeconomic factors that can explain the majority of the variation in the data panel. Subsequently these factors serve as the state variables X in the ATSM.

The model approach in this paper is similar to the factor-augmented VAR (FAVAR) of Bernanke et al. (2005). The FAVAR is particularly interesting in the way the monetary policy rate enters both as an observed variable (in the measurement equation) and through the augmentation of the state variable with the policy rate such that the policy rate interacts dynamically with the factors in the VAR dynamics - hence the term factor-augmented VAR. This means that it is possible to control for the short-term interest rate when macroeconomic factors are estimated. Identification in general is addressed in section 3.3.

As in the previous section the starting point is once again a state space model. Consider a panel of observable economic and financial variables $\bar{x}_{i,t}$, where i denotes the cross-section unit, $i = 1, ..., N$, while t refers to the time index, $t = 1, ..., T$. The panel of observed economic variables is transformed into stationary variables with zero mean and unit variance. These transformed variables are labeled $x_{i,t}$. Dynamic factor models assume that a variable x_{it} can be decomposed into two components, the *common component,* χ_{it}, and the *idiosyncratic component* ξ_{it}:

$$x_{it} = \chi_{it} + \xi_{it}.$$

Furthermore, in exact dynamic factor models it is assumed that the idiosyncratic and common components are uncorrelated at all leads and lags and across all variables, $E(\xi_{i,t}\chi_{j,s}) = 0, \ \forall \ s, t, i, j$. The common component is assumed to be driven by a small number r, $r << N$, of *common factors* $f_t = (f_{1t}, \ f_{2t}, \cdots, f_{rt})^\top$:

[13] Reichlin (2003) presents an empirical review of dynamic factor models.

$$x_{it} = \lambda_i^\top f_t + \xi_{it} \tag{3.15}$$

where λ_i is an $r \times 1$ vector of factor loadings measuring the exposure of x_{it} to the factors f_t. On the other hand, the idiosyncratic component is driven by variable-specific noises. Stacking equation (3.15) over all cross-section units, x_{it}, $i = 1, ..., N$, gives

$$x_t = \lambda f_t + \xi_t, \tag{3.16}$$

where $x_t = (x_{1t}, \ldots, x_{Nt})^\top$, $\xi_t = (\xi_{1t}, \ldots, \xi_{Nt})^\top$, and λ is an $N \times r$ matrix of factor loadings, $\lambda = (\lambda_1, ..., \lambda_N)^\top$. Equation (3.16) is called a *static* factor model.[14]

To close the model, factor dynamics have to be specified. We assume that the r-dimensional vector of common factors f_t has a VAR(p) representation

$$\varphi(L) f_t = \eta_t, \tag{3.17}$$

where $\varphi(L) = I - \varphi_1 L - \varphi_2 L^2 - \ldots - \varphi_p L^p$, with φ_j denoting an $r \times r$ matrix of autoregressive coefficients ($j = 1, \ldots, p$). Moreover, given the stationarity of the transformed panel, we impose that the roots of det $(\phi(L))$ are outside the complex unit circle. The r-dimensional vector of dynamic factor innovations is denoted η_t. As in Doz et al. (2006), I make the distributional assumptions that $\eta_t \sim i.i.d\ N(0, Q)$ and $\xi_t \sim i.i.d\ N(0, R)$, with Q and R denoting (semi)positive definite matrices[15].

Using equations (3.16) and (3.17), the model can be summarized in first order form, with state vector F_t, $F_t = (f_t^\top, ..., f_{t-p+1}^\top)^\top$ by the measurement equation:

$$x_t = \Lambda F_t + \xi_t, \tag{3.18}$$

and the transition equation

$$F_t = \phi F_{t-1} + U_t, \tag{3.19}$$

[14] "Static" stands for the fact that the observed variables only load contemporaneously on the factors.

[15] Note that, by assuming *i.i.d* idiosyncratic components, (3.16)-(3.17) define an *exact* dynamic factor model. This is certainly a strong assumption, particularly in the case of large panel data sets where some local cross-sectional and serial correlations are expected to be found. As such, (3.16)-(3.17) represent a misspecified model. However, Doz et al. (2006) show that, for large N and T, the exact factor model estimators are consistent quasi-maximum likelihood estimators for the *approximate* factor model.

where ϕ is the $rp \times rp$ companion matrix corresponding to $\varphi(L)$ and $U = \left(I_r, 0^{\top}_{r(p-1)\times r}\right)^{\top}$

Two state space models for panel data have been presented in this section and they are summarized briefly below as:

1. A no-arbitrage Gaussian ATSM in state space form

$$
\begin{aligned}
Y_t &= A(n,\theta) + B(n,\theta) X_t + v_t \\
X_t &= \mu + \Phi X_{t-1} + \Sigma \varepsilon_t
\end{aligned}
$$

   where $\theta = \{\delta_0, \delta_1, \mu, \Phi, \Sigma, \lambda_0, \lambda_1\}$ contains all the underlying parameters and where some of the observed state variables in X_t are macroeconomic factors (F_t) from:

2. A factor-augmented VAR model for large panel dynamic factor analysis

$$
\begin{aligned}
x_t &= \Lambda F_t + \xi_t \\
F_t &= \phi F_{t-1} + U_t
\end{aligned}
$$

However, in their current form neither of the two state space models are econometrically identified, as it is possible to form observationally equivalent models with different parameters and state variables[16]. The following section will thus address identification and also estimation methods.

3.3 Estimation and identification

3.3.1 Identification issues in Gaussian ATSMs with observed and unobserved state variables

The presence of latent state variables in standard multifactor ATSMs implies that not all model parameters are econometrically identified. The identification approach taken in this paper is quite standard but is nevertheless briefly discussed.

[16]Consider a rotation of the FAVAR with the invertible matrix H such that $x_t = \tilde{\Lambda}\tilde{F}_t + \xi_t$ and $\tilde{F}_t = \tilde{\Phi}\tilde{F}_{t-1} + \tilde{U}_t$ with $\tilde{\Lambda} = \Lambda H^{-1}$, $\tilde{F}_t = HF_t$, $\tilde{\Phi} = H\Phi H^{-1}$ and $\tilde{U} = MU_t$. This model is clearly observationally equivalent to the model above and the parameters are therefore not identified. The same applies to ATSM.

Identification of ATSMs with latent state variables is thoroughly discussed in Dai & Singleton (2000) where the notion of a canonical model defines a model that is admissible, econometrically identified and still maximally flexible within a family of models. Implicit in the canonical model specification is a set of normalizations required for identification, which makes it impossible to rotate the state vector without changing the short rate and thus the bond price. However, because the state vector is latent, it is still possible to make "invariant" transformations (rotations and translations of the state vector) that preserve admissibility and identification without changing the short rate; cf. Dai & Singleton (2000). When observed variables are included in the state vector, the usual rotations or translations would also change the observed part of the state vector. Consequently Pericoli & Taboga (2008) redefine the canonical ATSM when observed variables are included among the state variables.

Ang & Piazzesi (2003) assume that the macro variables are exogenous to the yield curve. Implicitly, this implies a set of overidentifying overidentifying parameter restrictions compared with the canonical model in Pericoli & Taboga (2008). However, these overidentifying restrictions are by all appearances mainly imposed to keep the estimation of these highly parameterized models manageable.

In order to clarify how and where the identifying restrictions are imposed a no-arbitrage state space model is presented below which distinguishes between observed state variables and unobserved (latent) state variables. Consider first the *measurement* equation that consists of the $\mathcal{N}$ observed yields in Y_t and now also p lags of K_1 observed macroeconomic variables stacked in a $K_1 \cdot p$ dimensional vector X_t^o. These observed $(\mathcal{N} + K_1 \cdot p)$ variables are affine in the state vector X_t which is partitioned into $K_1 \cdot p$ observed macro variables and K_2 latent variables in X_t^u:

$$\begin{bmatrix} X_t^o \\ Y_t \end{bmatrix} = \begin{bmatrix} 0 \\ A \end{bmatrix} + \begin{bmatrix} I & 0 \\ B^o & B^u \end{bmatrix} \begin{bmatrix} X_t^o \\ X_t^u \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & B^m \end{bmatrix} \begin{bmatrix} w_t \\ v_t \end{bmatrix} \tag{3.20}$$

Accordingly, X_t is a $K = K_1 \cdot p + K_2$ dimensional vector. Furthermore, the $\mathcal{N} \times 1$ vector A and the $\mathcal{N} \times K$ loadings matrix $\begin{bmatrix} B^o & B^u \end{bmatrix}$ follow from the no-arbitrage cross-section restrictions in (3.9)-(3.10). Thus, it can be seen that the yields now load on both macroeconomic variables through B^o which is an $\mathcal{N} \times (K - K_2)$ matrix and on the latent variables through B^u which is $\mathcal{N} \times K_2$. The macroeconomic variables are observed and therefore assumed to be measured without error, i.e. $w_t = 0$[17].

[17] To be precise, the dynamic factor analysis already allows for series specific idiosyncratic dis-

Furthermore $v_t \sim iid\ N(0, I)$ implying that the measurement errors are $B^m v_t \sim N\left(0, B^m B^{m\top}\right)$, where B^m is an $\mathcal{N} \times K_2$ matrix.

The *transition* equation for the state variables X_t is the same as (3.4) but written out slightly to emphasize the macroeconomic state variables:

$$\begin{bmatrix} X_t^o \\ X_t^u \end{bmatrix} = \begin{bmatrix} \mu^o \\ \mu^u \end{bmatrix} + \begin{bmatrix} \Phi^{oo} & \Phi^{ou} \\ \Phi^{uo} & \Phi^{uu} \end{bmatrix} \begin{bmatrix} X_{t-1}^o \\ X_{t-1}^u \end{bmatrix} + \begin{bmatrix} \Sigma^{oo} & \Sigma^{ou} \\ \Sigma^{uo} & \Sigma^{uu} \end{bmatrix} \begin{bmatrix} \varepsilon_t^o \\ \varepsilon_t^u \end{bmatrix} \tag{3.21}$$

where an exact listing of the dimensions of the vectors and matrices are deferred to Appendix A.4.

The exactly identifying restrictions are probably most easily stated if the starting point is the model in (3.20)-(3.21) with no macroeconomic state variables, i.e. $K_1 = 0$. In this case and following Dai & Singleton (2000) Σ^{uu} is normalized to an identity matrix which allows us to estimate δ_1^u freely and which allows the latent state variables to be correlated through a lower triangular Φ^{uu}. Furthermore, a zero restriction on μ^u allows for a free estimate of δ_0 in the short rate equation. The upper left $K_2 \times K_2$ block of λ_1^u is estimated freely, but this in turn requires one element in λ_0^u to be restricted to zero; cf. de Jong (2000). For a three-factor model, this implies twenty-one parameters to be estimated plus the covariance matrix of the measurement errors.

Ang & Piazzesi (2003) apply this identification scheme and additionally impose that $\Phi^{ou} = \Phi^{uo} = \Sigma^{ou} = \Sigma^{uo} = 0$, that μ^o is zero[18] and that Σ^{oo} is lower triangular as a result of a Cholesky decomposition. These restrictions are a consequence of the exogenous treatment of the macroeconomic state variables in the term structure where Φ^{oo}, δ_1^o and Σ^{oo} are estimated consistently in a first step prior to the term structure estimation and subsequently kept fixed in the second step estimation of the ATSM. However, the upper left $K_1 \times K_1$ block of λ_1^o is estimated freely such that the market prices of risk also depend on the state of the macro economy. The restrictions $\Phi^{ou} = \Phi^{uo} = \Sigma^{ou} = \Sigma^{uo} = 0$ are overidentifying restrictions according to Pericoli & Taboga (2008), i.e. it is in fact possible to achieve a more flexible yet identified model without imposing these restrictions. However, treating $\{\Phi^{ou}, \Phi^{uo}, \Sigma^{ou}, \Sigma^{uo}\}$ as free greatly increases the computational burden for the model in Ang & Piazzesi

turbances in the extraction of the common dynamic factors F_t. Hence, yet another source of disturbances through w_t is not preferred here.

[18] The macro state variables are de-meaned prior to estimation.

(2003) as additionally ninety parameters need to be estimated when $K_1 = 2$ and $p = 12$. I experienced the same type of computational challenges and therefore opt for the same identification scheme as in Ang & Piazzesi (2003) including the above-mentioned overidentifying restrictions.

3.3.2 Estimation of Gaussian ATSMs

Two methods for the estimation of dynamic term structure models with latent state variables are often used in the literature.

The *first method* is the maximum likelihood approach by Chen & Scott (1993), where time series of K_2 latent state variables are inverted from a more or less arbitrarily chosen set of K_2 perfectly measured zero-coupon bond yields.[19] The conditional density of the observed yields then follows from the then known density of the state vector and the Jacobian. Appendix A.4 presents the inversion of the latent state variables and the likelihood function. This method has been used in a number of papers including Duffee (2002), Dai & Singleton (2002) and the Gaussian macro-finance term structure models by Ang & Piazzesi (2003), Hördahl et al. (2006) and Pericoli & Taboga (2008).

The *second method* is the Kalman filter which recursively filters the latent state variables conditional on a parameter vector and conditional on observing a history of yields of different maturities and where *all* yields may be measured with errors. The filter is recursive in the sense that each time a new observation arrives, a forecast error can be calculated which in turn enables an update of the conditional moments for the state vector X. Based on the updated conditional moment of X a new one-period ahead forecast of the observed variable can be computed. For Gaussian models, the linear Kalman filter is the optimal linear estimator within the class of linear estimators, and the exact likelihood follows from the prediction error decomposition; cf. equation (3.26) in Appendix A.4, which contains more details[20].

It can be noted that the estimation of Gaussian term structure by the method of Chen & Scott (1993) can be seen as a special case of the Kalman filter if the same measurement errors are restricted to zero as in former method. The Kalman

[19] I find that estimation results are sensitive to which yields that are measured perfectly.

[20] For non-Gaussian models like the Cox et al. (1985) model, the exact likelihood is unknown but quasi maximum likelihood methods based on the conditional first and second moments of the state variables have been used in the literature; cf. Lund (1997) for a discussion.

filter method has been applied to Gaussian macro-finance term structure models by Dewachter et al. (2006), Dewachter & Lyrio (2006*b*) and Duffee (2007).

In this paper, I opt for the Kalman filter, primarily to avoid measuring some of the yields without error, but also because of the generality and flexibility of this method. For instance, the Kalman filter method allows measuring some of the observed variables perfectly if needed or even handling missing data. However, the computational cost of using the Kalman filter is larger than using the inversion method of Chen & Scott (1993). For each candidate parameter vector in the optimization routine, the Kalman filter loops recursively through the T observations, each involving a matrix inversion and a set of matrix multiplications in order to filter the latent state vector X_t. The inversion method is much faster in this respect, as only a single matrix inversion is needed for each candidate parameter vector.

3.3.3 Identification in the factor-augmented VAR and estimation by the EM algorithm

The FAVAR is not econometrically identified as it stands in the state space model of (3.18) and (3.19). The identification scheme and the estimation method are different from Bernanke et al. (2005) as I allow for *correlated* factors estimated by the EM algorithm as opposed to orthogonal factors estimated by Bayesian methods in the latter.

As discussed in details in Bork (2008), identification in factor models is about separating the contributions of the different latent factors to the variation in the panel x. The predominant starting point is uncorrelated factors which, implies that the identification of the sources of variation in x is then a matter of imposing an identifying structure on the loading matrix; in particular a lower triangular block structure of the loading matrix. Alternatively, the assumption about uncorrelated factors can be relaxed by allowing for *correlated* factors. However, less restricted factor dynamics would have to be paid by a more restrictive structure on the loading matrix; in particular the lower triangular block mentioned above is replaced by an identity matrix of the same size. In other words: Either the variables in x covary because they load differently on a set of uncorrelated factors or because they load on different factors which are themselves correlated.

Thus, a sufficient condition for an exactly identified FAVAR model is to impose an identity matrix restriction on r of the N observed variables.l[21]. Therefore, for an exactly identified model I specify the restricted loading matrix Λ^* by imposing:

$$\Lambda^*_{j,l} = 1, \ \Lambda^*_{j,k\neq l} = 0 \ \text{ for } j = 1, .., r \tag{3.22}$$

but as shown in Bork et al. (2008) additional restrictions may be imposed by the general form of loading restrictions

$$H_\Lambda \operatorname{vec}(\Lambda^*) = \kappa_\Lambda$$

where κ_Λ is a vector and H_Λ is the restriction matrix H_Λ. In fact, to identify the monetary policy rate that enters both the measurement and state transition equation I make use of this general restriction. Specifically, the policy rate in x_t loads with unity on the last factor in F_t and zeros on the remaining latent factors, such that the corresponding row in Λ^* for the policy rate is $[0, ..., 0, 1]$. In line with Bernanke et al. (2005), I argue that the federal funds rate is measured without error whereas the other variables may be measured with error

These restrictions are easily imposed when the EM algorithm is used as the estimation method. The EM algorithm is an iterative maximum likelihood procedure which is useful for models with "missing data", which in this context are the unobserved dynamic factors; cf. Dempster et al. (1977), Shumway & Stoffer (1982) and Watson & Engle (1983) for important contributions[22]. In fact, for the linear state space model in (3.18)-(3.19) it is possible to obtain closed form solutions for the parameters of interest $\{\Lambda^*, \phi, Q, R\}$ as specified in Appendix A.4 which offers a self-contained introduction to the EM algorithm including the application of the Kalman filter and the Kalman smoother.

[21]Furthermore, in this identification scheme, the restrictions are imposed on the so-called slow-moving variables like production, employment as opposed to fast-moving variables like exchange rates, stock prices, interest rates and consumer survey measures. This division into slow and fast-moving variables follows Bernanke et al. (2005). See Appendix A.1 page 162 for more details on this division.

[22]It seems that it is not possible to apply the EM algorithm in a standard way to ATSMs as the objective function is highly non-linear in the parameters.

3.4 Empirical application

In this section, a multifactor ATSM with filtered macroeconomic state variables conditioned on a large information set is taken to the data. I will show that filtered macroeconomic state variables from a large panel data set matter for model performance in several respects and that shocks to fundamental macroeconomic state variables play an important role in the response of bond yields and bond risk premia. In this respect, inflation and unemployment are particularly important, but the relations between bond risk premia and shocks to these fundamental macroeconomic state variables are only uncovered if the dynamics of the state variables are sufficiently rich.

In the following, the data is presented first, then some comments on the econometric model specification and estimation issues are given, and finally the empirical results are presented.

3.4.1 Data

To illustrate how the bond market might filter the state of the macroeconomy from many different data sources I revisit the large data panel of macroeconomic and financial time series analyzed in Bernanke et al. (2005)[23]. This dataset consists of 120 monthly time series and therefore captures the dynamics of a wide range of economic as well as financial developments in the US economy over the period 1959:1 to 2001:8[24]. Specifically, the dataset contains several measures of industrial production, income, (un)employment, consumption, housing starts, inventories, price indices and other economic measures. Furthermore, financial market variables such as stock prices, foreign exchange rates and coupon bond yields are also included.

I follow Ang & Piazzesi (2003) and use the continuously compounded zero-coupon bond yields of maturities 1, 3, 12, 24 and 60 months from the CRSP covering

[23] I thank Jean Boivin for kindly making the data set available on his website, HEC-Montréal, Canada.

[24] The data are already transformed by Bernanke et al. (2005) to reach stationarity; see Appendix A.1 page 162 in this paper and Bernanke et al. (2005) for details on the data set and on the transformation which results in a sample size of $T = 511$. The data transformation decisions are similar to Stock & Watson (2002*b*) and based on judgemental and preliminary data analysis of each series, including unit root tests.

Prior to the estimation, we de-mean the series and divide them by their standard deviation such that the resulting series have zero mean and unit variance.

the same period as the data above (the macro data).

To get a sense of the data, an example of a common inflation factor estimated from the FAVAR model is plotted against the log difference of two relevant specific price indices in Figure 3.1. It can be noted that although monthly growth rates in price indices are quite volatile, the common inflation factor does not match every movement in the price indices; some of movements are thus series specific idiosyncratic movements.

[Insert Figure 3.1 here]

3.4.2 Model specification

With hundreds of time series in the macro panel and potentially highly parameterized macro-finance ATSMs there is indisputably a large set of candidate models to choose between. Furthermore, both the FAVAR model and the Kalman filtration of the ATSM can be time consuming, which means that there is a timewise limit as to the depth of model exploration.

I choose to focus on three dimensions in the model specification. *Firstly*, five macroeconomic dynamic factors are considered as candidate state variables motivated by standard macroeconomic theory and derived from the FAVAR. The interest centers around what role these fundamental macroeconomic key variables play in the response of the yield curve and the bond risk premia to macroeconomic shocks. Typically, two of these enter the ATSM as well as typically three latent variables similarly to Ang & Piazzesi (2003).

Secondly, I examine whether yield and risk premia responses to fundamental shocks depend on the assumed dynamic complexity of the state variables in the VAR. Typically, VARs with monthly macroeconomic data require a significant lag length; Ang & Piazzesi (2003) find that $p = 12$. The survival of such highly parameterized macro-finance ATSMs in AIC or BIC criteria is low. Pericoli & Taboga (2008) apply the AIC and BIC criteria to a macro-finance model with quarterly data and find that p should be in the range 0-2 depending on the specification. *Thirdly*, I choose to focus on the estimation of the most parsimonious version of a given ATSM by carefully eliminating statistically insignificant parameters in order to reduce overparameterization. This approach is combined with the use of a global optimization routine (simulated annealing) in an attempt to avoid inferior local maximums.

The five macroeconomic dynamic factors are a subset of the nine factors analyzed in Bork et al. (2008) and each factor is given a clear economic interpretation by imposing overidentifying loading restrictions on Λ^* in (3.22) as proposed and described in detail by Bork et al. (2008).[25] Specifically, they define the following nine factors. Four of these factors are related to aggregate supply: an *inflation factor*; an *economic activity* factor; an *unemployment* factor and a *hours in production* factor (functioning as a buffer to changes in demand). Furthermore, they define three factors related to aggregate demand: a *consumption* factor; a *housing* factor approximating (residential) investment; and a *monetary policy* factor. The final two factors have the interpretation of an information factor (*commodity price* factor) and as a financial factor (*stock market* factor). The five macroeconomic factors retained in this paper is the aggregate supply factors and aggregate demand factors except for the housing factor and the monetary policy factor. The housing factor is a priori excluded as I expect this factor to lag interest rates whereas the monetary policy factor is excluded as the short-term interest rate is modelled within the ATSM as a Taylor-style policy rule.[26]

Obviously a general inflation factor should be a candidate state variable because inflation is an important target variable in monetary policy (the Taylor rule) and because inflation affects the purchasing power of the payoffs to bond investments. Other highly relevant fundamental macroeconomic state variables are economic activity and unemployment, which affect both the interest setting by the Federal Reserve Board and income and consumption possibilities of the consumers. The consumption growth factor is related to investor's marginal utility of consumption and therefore to the pricing kernel as in Breeden (1979) or more recently as in Campbell & Cochrane (1999). Finally, a relative responsive state variable is defined in terms of overtime hours in production.

[25] Accordingly, exactly the same data and exactly the same overidentifying restrictions are used in this paper as in Bork et al. (2008). Furthermore, they report that their results are robust to including more lags and to reducing the number of lags in the FAVAR to $p = 4$ and I expect these results to carry over to the FAVARs estimated in this paper using exactly the same estimation routine.

[26] The stock market factor is not included because I opt for a fundamental macroeconomic explanation of the term structure dynamics. The commodity price factor was included as a state variable in the ATSM but it did not show up as an important state variable.

A comment on the estimation of the ATSM will now be given. Feeding the optimization routine with good initial values for the parameters is the key to find the (hopefully) global maximum, and the simulated annealing method[27] is one mean towards this. I proceed towards the most parsimonious model as follows. In a first sequence of estimations I estimate (using simulated annealing) the expectation hypothesis version of the ATSM with $\lambda_0 = \lambda_1 = 0$ while $\{\Phi, \delta_1\}$ is estimated freely[28]. Then Φ and λ_0 are kept fixed while λ_1 is estimated. After this λ_0 is estimated while keeping $\{\Phi, \lambda_1\}$ fixed, and finally all parameters are free in a joint estimation. Given the resulting new set of initial parameter values, this sequence of estimations with some parameters kept fixed and other parameters free is repeated again (with the faster Nelder-Mead simplex method). However, this time insignificant parameters at a 10% significance level are set to zero. Notice that the purpose of this is only to build up good starting values. In the final sequence, insignificant parameters at a 10% significance level are removed one at a time from the parameter vector, which is reestimated after each removed parameter. Admittedly, this method involves many estimations, but it delivers a parsimonious model and is capable of finding the true parameters of a simulated three-factor ATSM[29]. Moreover, this approach also proves successful in an earlier replication of the results in Ang & Piazzesi (2003).

One of the attractive properties of the Kalman filter is the recursive calculation of the *filtered* state vector $X_{t|t}$ which is the time t expectation of the state vector conditional on observations up to time t. Assuming that the parameter vector of the FAVAR is known by the bond market $X_{t|t}$ may be interpreted as the time t beliefs of the bond market about the state of the economy[30]. Therefore, in terms of model specification the filtered state vector is preferred in the estimations. Finally, the Kalman filter allows for heteroskedastic measurement errors which are theoretically attractive but nevertheless increase the computation time significantly[31]. Therefore,

[27] Simulated annealing is a global stochastic optimization technique which is often explained with reference to the cooling *process* of molten metal (thermodynamics) where a slow cooling (annealing) leads to a low energy state (the minimum) whereas a quick cooling might lead to a *local* minimum only. See Goffe et al. (1994) for more details.

[28] δ_0 is always fixed at the unconditional mean of the short rate.

[29] This approach grew out of unfortunate experience with local maximums using initially the simplex method and out of a brief comment from Ang & Piazzesi (2003) page 763 on estimation issues.

[30] This argument ignores the actual vintage of the real-time data that the bond market actually had at that time.

[31] If there are good reasons to expect that the variance of the measurement errors in the bond market should be heteroskedastic, for instance if there are thinly traded bonds or "on-the run" versus "off-the-run" effects, then it may be a good idea to allow for heteroskedasticity. On the

I assume homoskedastic measurement errors as in Duffee (2007, 2008).

3.4.3 Empirical results

The results of the exploration into potential macroeconomic sources of the variation in bond excess return are now presented. The exploration is limited to an assessment of which role the five macroeconomic dynamic factors plays in the variation of excess returns. This assessment is performed in terms of impulse response analysis and forecast error variance decomposition (FEVD) of excess bond returns. In Appendix A.5 I derive the FEVD for the excess bond returns with a general holding period. The preferred ATSM model includes the filtered inflation factor and the filtered unemployment factor among the $K_1 = 2$ macroeconomic state variables and the $K_2 = 3$ latent variables. The background for this preference, is mainly that unemployment is the most important source of variation in bond excess return, among the analyzed macroeconomic factors. This specification is then evaluated against other candidate models; for instance by comparing model fit and varying the number of lags in both the FAVAR and the ATSM[32].

Some of the initial estimations are not included in this paper. Specifically, I quickly realized that the use of $K_1 = 3$ macroeconomic factors and $K_2 = 2$ latent term structure factors resulted in a significant inferior fit compared to models with $K_1 = 2$ and $K_2 = 3$. Therefore, only models with the latter model specification are reported. Furthermore, the inflation factor turns out to be an important state variable as in practically all other macro-finance models in the literature. Consequently, inflation is always one of the macroeconomic state variables in the ATSM while the other are either unemployment, economic activity, consumption or hours-in-production. Finally, I also allow the number of lags in both the FAVAR and the ATSM to be either $p = \{4, 8, 12\}$. It turns out that four lags is not sufficient in the sense that counterintuitive impulse responses emerge, which may be the result of a omitted variable problem, and muted endogeneous responses in the VAR residuals. Finally, the results with twelve lags are not very different from the results with eight lags and therefore dismissed on the basis of the AIC or BIC criteria.

other hand, adding more parameters may worsen the overparameterization problem.

[32] In a previous version of this paper a comparison with a model that do not use filtered dynamic factors was performed. For the last mentioned model, the fit was somewhat inferior and the impulse responses were different.

The empirical results for the models with eight or twelve lags are now summarized and then subsequently discussed in details. Firstly, the model-implied expected excess return (EER) responds as expected following a shock to the five macroeconomic state variables. EER responds negatively to shocks to inflation, economic activity, consumption and hours-in-production. Intuitively, shocks to the last three mentioned variables would correspond to an improved state of the economy and therefore the demanded risk-premium decreases. Secondly, the FEVD of the EER is broadly similar for all five macroeconomic factors and the factors account for no more than 30% of the total forecast error variation at any forecast horizon. Thirdly, all four models are able to fit the yield curve well. The mean of the absolute values of the deviation between observed yields and model implied yields, is in the range of about 5-10 basis points. Finally, the impulse response analyses of the yield curve following macroeconomic shocks also display the expected responses of the yields of different maturites. For instance, shocks to industrial production and/or inflation raise the short-term interest rate but also the longer yields.

I therefore conclude that the estimated models with either eight or twelve lags seem to be well-specified. These models deliver empirical impulse responses of the bond yields or expected excess returns that are consistent with theory and common sense[33].

The retained five common dynamic factors - inflation, unemployment, economic activity, and consumption - are depicted in Figure 3.2 as well as the other four factors in Bork et al. (2008). Focusing on the five factor it can be seen that these candidate state variables in the macro-finance ATSM are well in line with the leading measures and trends in the US economy over the sample period. Specifically, the general inflation factor captures very well the overall CPI series while the unemployment factor, the economic activity factor and the hours in production factor also capture the development quite good. Moreover, these factors also capture the peaks and troughs of the business cycle well.

[Insert Figure 3.2 here]

Figures 3.3, 3.4, 3.5 and 3.6 display the impulse response functions of expected excess return of bonds following a shock to 1) the inflation factor plus one of the factors from {unemployment, hours-in-production, economic activity or consumption growth}, respectively. It can be seen that the shapes of the impulse response

[33] As already noted, the models with four lags result in counterintuitive responses.

functions for inflation shocks are broadly similar. Moreover, the unemployment factor in Figure 3.3 is seen to be an important source of variation in expected excess return.

Generally, it can be seen that the longer the holding-period the larger the response of expected excess return. Moreover, the longer the bond the larger the response. A one standard deviation shock to unemployment initially raises the expected excess return by 17 basis points on an annually basis for a five-year bond held for one year. This empirical evidence conforms with the conventional view that risk premia are countercyclical. Notice also that a one standard deviation shock to inflation lowers the expected excess return by almost 20 basis points. The responses are weaker for the remaining shocks.

[Insert Figures 3.3, 3.4, 3.5 and 3.6 here]

Another related technique to analyze how the variation in bond yields and bond risk premia is affected by macroeconomic state variables is the forecast error variance decomposition. Specifically, in a forecast error variance decomposition (FEVD), I calculate for a given forecast horizon what fraction of the model-implied total forecast error variance for a particular variable that is due to a specific shock. I use the FEVD to analyze how macroeconomic variables affect the expected excess return on bonds. In Appendix A.5 I derive the FEVD for excess return on bonds with a general holding period. A step towards the FEVD is the mean squared forecast error, $MSFE_t$, given by:

$$MSFE_t\left(rx_{t+s,t+m+s}^{(n)}\right) = \sum_{j=0}^{s-1} b_{n,m}^{\top}\Phi^j\Omega\left[\Phi^j\right]^{\top} b_{n,m} + \sum_{i=0}^{m-1} \mathcal{B}_{n-m}^{\top}\Phi^i\Omega\left[\Phi^i\right]\mathcal{B}_{n-m}$$

where $b_{n,m}$ is given in (3.13) and represents the loading for the involved bonds until the forecast starts. The matrices Φ, Ω are seen in Definition 1. The price loading $\mathcal{B}_{n-m}^{\top}$ is seen in (3.7) and relates to the uncertainty of the selling price of the $n-m$ period bond at the end of the forecast period s. The holding period is represented by m. The second term dominates in this expression and therefore the FEVD is generally stable and almost constant throughout the forecast period. Consequently, the FEVD for the five macroeconomic state variables is quite similar and therefore only the FEVD for the preferred model is shown in Figure 3.8. It can be seen that the inflation factor (denoted obs. state variable 1) is the far most important macroeconomic variable in *forecasting* excess returns of bonds in this model.

[Insert Figure 3.8 here]

The preferred model is now related to other candidate models to evaluate the empirical fit. Consider Table 3.1 where the statistical fit of ATSMs with four sets of macroeconomic is evaluated. Specifically, the preferred model is evaluated against the following sets: {inflation, hours-in-production} , {inflation, economic activity}, {inflation, hours-in-production} and {inflation, consumption} . Generally, all four models are able to fit the yield curve well with a mean of the absolute values of the deviation between observed yields and model implied yields of about 7-8 basis points. However, it is difficult to discriminate in sample between different models, although the second model in Panel B seems to fit the yield curve less well. On this background I conclude that an interesting account of some of the variability of excess returns is obtained via the preferred model, without sacrificing the statistical fit.

A prominent benchmark model is the Ang & Piazzesi (2003) affine term structure model with $K_1 = 2$ macroeconomic variables and $K_2 = 3$ latent term structure factors. Interestingly, Ang & Piazzesi (2003) perform a "mini" dynamic factor analysis where inflation and real activity measures are constructed separately by principal components analysis of a few observed indices. Their original data is extended to match the sample period in this paper. Table 3.2 reports parameter estimates and statistical fit of a Ang & Piazzesi like model. The model is estimated by the Chen & Scott (1993) method to comply with the approach of Ang & Piazzesi. Figure 3.7 displays the impulse responses of expected excess returns of bonds. A comparison with the {inflation, economic activity} model of this paper reveals that excess returns in the latter models responds a little more; especially for the longer holding periods. The same holds for the preferred model .

The final robustness analysis concerns impulse response analysis of the bond yields. From the analysis of how a shock to a state variable propagates through the state space system and affects bond yields, we can learn how different segments of the maturity spectrum of the yield curve responds over time to such shocks. For instance, it might be interesting in scenario analysis to know the effect on different yields following a shock to inflation. Appendix A.5 shows that the n-period yield can be written in a moving-average form which essentially shows how an uncorrelated shock realized generically in the past affects today's yield. Illustrative examples are given below.

Figure 3.9 displays how the 1-month, 12-month and 60-month yield responds to a one standard deviation inflation shock and unemployment shock. It can be seen that 100 bp. deviation from baseline (standardized) inflation in this particular model would make the short yields respond immediately by 30 bp, and about 12 bp. for the long bond yield but then the effect dies out slowly. This is as expected in the sense that the short yields are affected most if the Federal Reserve Board accommodates a rise in inflation by an increase in the Federal funds rate. Similarly, for an unemployment shock, short yields show little response as expected with a more pronounced effect for the long term yield.

3.5 Conclusion

This paper can be seen as a new empirical approach to no-arbitrage bond pricing that takes into account the abundant amount of macroeconomic and financial information in the bond market. Part of this information needs to be processed in order to asses the current state of the economy, which is relevant for an assessment of the state of the monetary policy rate, i.e. for the short end of the yield curve. To obtain a broad based assessment, it is preferable to look at several relevant economic as well as financial variables in order to distinguish series specific idiosyncratic measurement errors from the underlying common component of relevant key macroeconomic variables. I propose to use a factor-augmented VAR to filter relevant key macroeconomic variables and explain empirically the short rate by a few of these macroeconomic variables; specifically I argue for a general inflation factor and a general unemployment factor. Furthermore, the exactly identifying restrictions imposed on the loadings of the factor-augmented VAR ensure that the factors and the monetary policy rate are identified and thus allow me to use financial variables as timely information variables for the macroeconomic development. The overidentifying restrictions ensure a clear macroeconomic interpretation of the factors.

However, the intersection between the macroeconomy and the bond market is not limited to the short end of the yield curve. Risk premiums depend on the state of the economy as well. In bad times investors would require a higher premium to hold a long-term bond. A multi-factor Gaussian affine term structure model is therefore augmented with (Kalman) filtered state variables derived from the dynamic factor analysis. I find that among five different macroeconomic state variables inflation and unemployment perform particularly well. Firstly, in terms of yield curve fit they perform at least as well as the updated factors from Ang & Piazzesi. Secondly, in a forecast error variance decomposition the filtered state macroeconomic variables show a significant ability to account for the variation in excess returns of bonds. However, the key to an important role for macroeconomic variables in excess returns of bonds is a sufficiently rich dynamic complexity in the term structure model.

A Appendix

A.1 Conditional expectation of state variables under $\mathbb{Q}$ and $\mathbb{P}$.

The conditional expectation of X_{t+1} under the $\mathbb{P}$ measure is denoted $\mathbb{M}_t$ and the conditional expectation of X_t under the $\mathbb{Q}$ measure is denoted $\mathbb{M}_t^{\mathbb{Q}}$. The two moments are related by the Radon-Nikodym derivative. Dai et al. (2007) show how the conditional distribution of X_{t+1} under $\mathbb{P}$ is fully characterized by conditional moment generating function $\psi_t^{\mathbb{P}}$:

$$\begin{aligned} \psi_t^{\mathbb{P}}(\upsilon) &= E_t\left[\exp\left\{\upsilon^\top X_{t+1}\right\}\right] = E_t^{\mathbb{Q}}\left[\exp\left\{\upsilon^\top X_{t+1}\right\}\left(\frac{\mathrm{d}\,\mathbb{P}}{\mathrm{d}\,\mathbb{Q}}\right)^{\mathcal{D}}_{t,t+1}\right] \\ &= \frac{\psi_t^{\mathbb{Q}}\left(\Sigma^{-1}\lambda_t + \upsilon\right)}{\psi_t^{\mathbb{Q}}\left(\Sigma^{-1}\lambda_t\right)} = \exp\left\{\upsilon^\top\left(\mathbb{M}_t^{\mathbb{Q}} + \Sigma\lambda_t\right) + \frac{\upsilon^\top\Sigma\Sigma^\top\upsilon}{2}\right\} \end{aligned} \qquad (3.23)$$

From this expression it is simple to calculate the first moment by differentiating (3.23) with respect to υ and evaluating at $\upsilon = 0$ which implies

$$\mathbb{M}_t = \mathbb{M}_t^{\mathbb{Q}} + \Sigma\lambda_t$$

and inserting the definitions of $\mathbb{M}_t$ and $\mathbb{M}_t^{\mathbb{Q}}$ from Definition 1 yields

$$\begin{aligned} \mu + \Phi X_t &= \mu^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_t + \Sigma\lambda_t \\ &= \mu^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_t + \Sigma\left(\lambda_0 + \lambda_1 X_t\right) \end{aligned}$$

and matching coefficients implies

$$\begin{aligned} \mu &= \mu^{\mathbb{Q}} + \Sigma\lambda_0 \\ \Phi &= \Phi^{\mathbb{Q}} + \Sigma\lambda_1 \end{aligned}$$

Alternatively, we could express λ_t as in the main text as

$$\lambda_t = \Sigma^{-1}\left(\mu - \mu^{\mathbb{Q}}\right) + \Sigma^{-1}\left(\Phi - \Phi^{\mathbb{Q}}\right) X_t$$

A.2 $\mathcal{A}$ and $\mathcal{B}$ in the bond price equation

The following proposition and proof are broadly similar to Campbell et al. (1997) chapter 11 and Ang & Piazzesi (2003). To prove that the proposed bond price equation is compatible with the fundamental asset pricing equation in (3.6) and Definition 1 the proposed bond price equation is substituted into $\exp\left\{\mathcal{A}_n + \mathcal{B}_n^\top X_{t+1}\right\}$ in (3.6) and then I show that $\mathcal{A}_n$ and $\mathcal{B}_n$ each must satisfy a cross-sectional restriction.

Proposition 2 *The bond price equation in a discrete-time affine term structure models is exponentially affine in the state variables X*

$$P_{n,t} = \exp\left\{\mathcal{A}_n + \mathcal{B}_n^\top X_t\right\}$$

where $\mathcal{A}_n$ and $\mathcal{B}_n$ must satisfy the following to be compatible with no-arbitrage as specified in Definition 1 :

$$\mathcal{B}_n = -\delta_1^\top \left(I - \Phi^{\mathbb{Q}}\right)^{-1}\left(I - \left[\Phi^{\mathbb{Q}}\right]^n\right) \tag{3.24}$$

$$\begin{aligned}\mathcal{A}_n &= -n\delta_0 - \delta_1^\top\left[n \cdot I - \left(I - \Phi^{\mathbb{Q}}\right)^{-1}\left(I - \left[\Phi^{\mathbb{Q}}\right]^n\right)\right]\left(I - \Phi^{\mathbb{Q}}\right)^{-1}\mu^{\mathbb{Q}} \\ &\quad + \frac{1}{2}\sum_{i=0}^{n-1}\mathcal{B}_i^\top \Sigma\Sigma^\top \mathcal{B}_i\end{aligned} \tag{3.25}$$

Proof. Substitute the zero-coupon bond price at time $t+1$ which now has $n-1$ periods to maturity into the fundamental asset pricing equation in (3.6), i.e. substitute $P_{n-1,t+1} = \exp\left\{\mathcal{A}_{n-1} + \mathcal{B}_{n-1}^\top X_{t+1}\right\}$ into $P_{n,t} = E_t\left[\frac{M_{t+1}}{M_t}P_{n-1,t+1}\right]$:

$$\begin{aligned}P_{n,t} &= E_t\left[\frac{M_{t+1}}{M_t}\exp\left\{-i_t - \frac{1}{2}\lambda_t^\top\lambda_t - \lambda_t^\top\varepsilon_{t+1}\right\}\exp\left\{\mathcal{A}_{n-1} + \mathcal{B}_{n-1}^\top X_{t+1}\right\}\right] \\ &= E_t\left[\exp\left\{-\delta_0 - \delta_1^\top X_t - \frac{1}{2}\lambda_t^\top\lambda_t - \lambda_t^\top\varepsilon_{t+1}\right\} \times \exp\left\{\mathcal{A}_{n-1} + \mathcal{B}_{n-1}^\top\left(\mu + \Phi X_t + \Sigma\varepsilon_{t+1}\right)\right\}\right] \\ &= \exp\left\{-\delta_0 - \delta_1^\top X_t - \frac{1}{2}\lambda_t^\top\lambda_t + \mathcal{A}_{n-1} + \mathcal{B}_{n-1}^\top\left(\mu + \Phi X_t\right)\right\} \times \\ &\quad E_t\left[\exp\left\{\left(-\lambda_t^\top + \mathcal{B}_{n-1}^\top\Sigma\right)\varepsilon_{t+1}\right\}\right]\end{aligned}$$

and from the definition of the mean of a lognormal variable this is equal to

$$\begin{aligned}
P_{n,t} &= \exp\left\{-\delta_0-\delta_1^\top X_t-\frac{1}{2}\lambda_t^\top\lambda_t+\mathcal{A}_{n-1}+\mathcal{B}_{n-1}^\top\left(\mu+\Phi X_t\right)\right\}\times \\
&\quad \exp\left\{\frac{1}{2}\lambda_t^\top\lambda_t+\frac{1}{2}\mathcal{B}_{n-1}^\top\Sigma\Sigma^\top\mathcal{B}_{n-1}-\mathcal{B}_{n-1}^\top\Sigma\lambda_t\right\} \\
&= \exp\left\{-\delta_0-\delta_1^\top X_t+\mathcal{A}_{n-1}+\frac{1}{2}\mathcal{B}_{n-1}^\top\Sigma\Sigma^\top\mathcal{B}_{n-1}+\mathcal{B}_{n-1}^\top\left(\mu+\Phi X_t\right)\right\}\times \\
&\quad \exp\left\{-\mathcal{B}_{n-1}^\top\left[\left(\mu-\mu^{\mathbb{Q}}\right)+\left(\Phi-\Phi^{\mathbb{Q}}\right)X_t\right]\right\} \\
&= \exp\left\{-\delta_0+\mathcal{A}_{n-1}+\frac{1}{2}\mathcal{B}_{n-1}^\top\Sigma\Sigma^\top\mathcal{B}_{n-1}+\mathcal{B}_{n-1}^\top\mu^{\mathbb{Q}}\right\}\times \\
&\quad \exp\left\{\left[-\delta_1^\top+\mathcal{B}_{n-1}^\top\Phi^{\mathbb{Q}}\right]X_t\right\}
\end{aligned}$$

Matching coefficients on the right hand side $P_{n,t}=\exp\left\{\mathcal{A}_n+\mathcal{B}_n^\top X_t\right\}$ with the left hand side yields

$$\begin{aligned}
\mathcal{B}_n &= -\delta_1^\top+\mathcal{B}_{n-1}^\top\Phi^{\mathbb{Q}} \\
\mathcal{A}_n &= -\delta_0+\mathcal{A}_{n-1}+\frac{1}{2}\mathcal{B}_{n-1}^\top\Sigma\Sigma^\top\mathcal{B}_{n-1}+\mathcal{B}_{n-1}^\top\mu^{\mathbb{Q}}
\end{aligned}$$

These recursions can in turn be written more compactly as in (3.24) and (3.25). The compact version of $\mathcal{B}_n$ is derived by substituting recursively from $\mathcal{B}_{n-1}$ to $\mathcal{B}_0$ which result in $\mathcal{B}_n=-\delta_1^\top\sum_{i=0}^{n-1}\left(\Phi^{\mathbb{Q}}\right)^i$ which in turn is (3.24). Similarly for $\mathcal{A}_n$. ■

A.3 Excess holding period bond returns

This section derives the conditional expected excess holding period bond return for a general holding period of length m periods.

The return from investing in a n-period bond for m periods is:

$$R_{t,t+m}^{(n)}=\frac{P_{t+m}^{(n-m)}-P_t^{(n)}}{P_t^{(n)}}$$

and the log return is then

$$\log\left(1+R_{t,t+m}^{(n)}\right)\equiv r_{t,t+m}^{(n)}=p_{t+m}^{(n-m)}-p_t^{(n)}$$

where the log price at time $t+m$ of a $n-m$ period zero-coupon bond is denoted $p_{t+m}^{(n-m)}$, etc. This return in excess of holding a m-period zero-coupon bond to maturity is

$$rx_{t,t+m}^{(n)} = p_{t+m}^{(n-m)} - p_t^{(n)} + p_t^{(m)}.$$

Inserting from (3.24) and (3.25) yields

$$rx_{t,t+m}^{(n)} = \mathcal{A}_{n-m} - \mathcal{A}_n + \mathcal{A}_m + \mathcal{B}_{n-m}^\top X_{t+m} - \mathcal{B}_n^\top X_t + \mathcal{B}_m^\top X_t$$

where $X_{t+m} = \sum_{i=0}^{m-1} \Phi^i \mu + \Phi^m X_t + \sum_{i=0}^{m-1} \Phi^i u_{t+m-i}$ from (3.4) . Thus, $rx_{t,t+m}^{(n)}$ can be rewritten as

$$\begin{aligned} rx_{t,t+m}^{(n)} &= \mathcal{B}_{n-m}^\top \sum_{i=0}^{m-1} \Phi^i \mu + \mathcal{A}_{n-m} - \mathcal{A}_n + \mathcal{A}_m \\ &\quad + \left(\mathcal{B}_{n-m}^\top \Phi^m - \mathcal{B}_n^\top + \mathcal{B}_m^\top\right) X_t \\ &\quad + \mathcal{B}_{n-m}^\top \sum_{i=0}^{m-1} \Phi^i u_{t+m-i} \\ &= a_{n,m} + b_{n,m}^\top X_t + \epsilon_{t+1,t+m} \end{aligned}$$

where

$$\begin{aligned} a_{n,m} &= \mathcal{B}_{n-m}^\top \sum_{i=0}^{m-1} \Phi^i \mu + \mathcal{A}_{n-m} - \mathcal{A}_n + \mathcal{A}_m \\ b_{n,m}^\top &= \mathcal{B}_{n-m}^\top \Phi^m - \mathcal{B}_n^\top + \mathcal{B}_m^\top \\ &= -\delta_1^\top \left(I - \Phi^{\mathbb{Q}}\right)^{-1} \left\{ \left(I - \left[\Phi^{\mathbb{Q}}\right]^{n-m}\right) \Phi^m + \left[\Phi^{\mathbb{Q}}\right]^n - \left[\Phi^{\mathbb{Q}}\right]^m \right\} \\ \epsilon_{t+1,t+m}^{(n)} &= \mathcal{B}_{n-m}^\top \sum_{i=0}^{m-1} \Phi^i u_{t+m-i} \end{aligned}$$

A.4 Estimation methods

The EM algorithm and Kalman smoothing recursions

The Expectation Maximization (EM) algorithm is an iterative maximum likelihood procedure applicable to models with "missing data", which in this context is the unobserved factors. The complete data likelihood of the Gaussian state space model in equations (3.18)-(3.19) is given in equation (3.32) below. Although the complete data likelihood cannot be calculated due to the unobserved factors, it is nevertheless possible to calculate the expectation of the complete data likelihood conditional on the observed data and inputs of parameters, denoted $\Theta^{(j)}$ at the jth iteration. Essentially, this expectation depends on smoothed moments of the unobserved variables from the Kalman smoother and hence on the data as well as parameters in $\Theta^{(j)}$. Finally, "updated" values of the parameters at iteration $j+1$ denoted $\Theta^{(j+1)}$ are available in closed form and follows from the first-order conditions of the conditional expectation of the complete data likelihood. The updated parameters $\Theta^{(j+1)}$ can then be used to filter and smooth a new set of moments to be used in the calculation of the conditional expectation of the complete data likelihood. This algorithm continues until convergence of the likelihood value.

The following offers a brief description of the Kalman filter and the Kalman smoother. Then the complete data likelihood and the incomplete data likelihood for a state space model are stated. Finally the moments used in the closed form parameters estimators in (3.34)-(3.37) are stated.

The Kalman filter Denote by $\mathcal{X}_t = \{X_1, ..., X_t\}$ the information set available at time t. The conditional expectation and variance of the factor are: $\hat{F}_{t+1|t} = E\left[F_{t+1}|\, \mathcal{X}_t\right]$ and $\hat{P}_{t+1|t} = var\left(F_{t+1}|\, \mathcal{X}_t\right)$, respectively.

The Kalman filter recursions for $t = 1, .., T$ can then be written as

$$\begin{aligned} \hat{F}_{t+1|t} &= \phi \hat{F}_{t|t-1} + K_t \left(X_t - \Lambda \hat{F}_{t|t-1} \right), \\ \hat{P}_{t+1|t} &= \phi \hat{P}_{t|t-1} L_t^{\top} + Q, \end{aligned}$$

where

$$\begin{aligned} \xi_t &= X_t - \Lambda \hat{F}_{t|t-1}, \\ P_t^{\xi\xi} &= \Lambda \hat{P}_{t|t-1} \Lambda^{\top} + R, \end{aligned}$$

such that the Kalman gain matrix K used in $\hat{F}_{t+1|t}$ and $\hat{P}_{t+1|t}$ is calculated as

$$K_t = \phi \hat{P}_{t|t-1} \Lambda^{\top} \left(\Lambda \hat{P}_{t|t-1} \Lambda^{\top} + R \right)^{-1}.$$

If the initial state vector and the error terms have proper normal distributions a useful output of the Kalman filter is the *prediction error decomposition* form of the likelihood:

$$\log L = -\frac{NT}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log\left|\hat{P}^{\xi\xi}_{t|t-1}\right| - \frac{1}{2}\sum_{t=1}^{T}\xi_t^{\top}\phi^{-1}\xi_t \tag{3.26}$$

cf. Harvey (1989) chapter 3.

Kalman smoothing Kalman smoothing is the name for the reconstruction of the full state sequence $\{F_1, .., F_T\}$ given the observations $\{X_1, .., X_T\}$. Smoothing provides us with more accurate inference on the state variables since it uses more information than the basic filter.

The Kalman smoother recursions for $t = T, ..., 1$, based on the efficient smoother by de Jong & Mackinnon (1988), de Jong (1989) and used in Koopman & Shephard (1992) are given by

$$\begin{aligned}
\hat{F}_{t|T} &= \hat{F}_{t|t-1} + \hat{P}_{t|t-1}\Lambda^{\top}\left[\hat{P}^{\xi\xi}_{t|t-1}\right]^{-1}\xi_t + \hat{P}_{t|t-1}L_t^{\top}r_t && (3.27)\\
&= \hat{F}_{t|t-1} + \hat{P}_{t|t-1}r_{t-1} \quad \text{(alternatively)} && (3.28)\\
\hat{P}_{t|T} &= \hat{P}_{t|t-1} - \hat{P}_{t|t-1}N_{t-1}\hat{P}_{t|t-1} && (3.29)\\
\hat{P}_{\{T,T-1\}|T} &= \left[I - K_T\Lambda\right]\phi\hat{P}_{T-1|T-1} && (3.30)\\
\hat{P}_{\{t,t-1\}|T} &= \left(I - \hat{P}_{t|t-1}N_{t-1}\right)L_{t-1}\hat{P}_{t-1|t-2},\ t = T-1, ..., 1 && (3.31)
\end{aligned}$$

where

$$\begin{aligned}
r_{t-1} &= \Lambda^{\top}\left[\hat{P}^{\xi\xi}_{t|t-1}\right]^{-1}\xi_t + L_t^{\top}r_t,\ \text{for } 1 \le t < T \text{ and } r_T = 0\\
N_{t-1} &= \Lambda^{\top}\left[\hat{P}^{\xi\xi}_{t|t-1}\right]^{-1}\Lambda + L_t^{\top}N_tL \text{ for } 1 \le t < T \text{ and } N_T = 0\\
L_t &= \phi - K_t\Lambda = \phi - \phi\hat{P}_{t|t-1}\Lambda^{\top}\left[\hat{P}^{\xi\xi}_{t|t-1}\right]^{-1}\Lambda.
\end{aligned}$$

The complete data likelihood and the incomplete data likelihood Under the Gaussian assumption including $F_0 \sim N(\mu_0, P_0)$ and ignoring the constant, the complete data likelihood of equations (3.18)-(3.19) page 3.19 can be written as

$$
\begin{aligned}
-2\ln L_{\mathcal{F},\mathcal{X}}(\Theta) &= \ln|P_0| + (F_0-\mu_0)^\top P_0^{-1}(F_0-\mu_0) \\
&\quad + T\ln|Q| + \sum_{t=1}^{T}(F_t-\phi F_{t-1})^\top Q^{-1}(F_t-\phi F_{t-1}) \\
&\quad + T\ln|R| + \sum_{t=1}^{T}(X_t-\Lambda F_t)^\top R^{-1}(X_t-\Lambda F_t)\,. \qquad (3.32)
\end{aligned}
$$

given that we can observe the states $\mathcal{F}_T = \{F_0, .., F_T\}$ as well as the observations $\mathcal{X}_T = \{X_1, .., X_T\}$. However, given $\mathcal{X}_T$ and some input of parameter estimates (denoted $\Theta^{(j-1)}$) the conditional expectation of the complete data likelihood can be written as

$$
\begin{aligned}
\mathcal{Q}\left(\Theta|\Theta^{(j-1)}\right) &= E\left[-2\ln L_{\mathcal{F},\mathcal{X}}(\Theta)\,\middle|\mathcal{X}_T, \Theta^{(j-1)}\right] \\
&= \ln|P_0| + \operatorname{tr}\left[P_0^{-1}\left\{\left(\hat{F}_{0|T}-\mu_0\right)\left(\hat{F}_{0|T}-\mu_0\right)^\top + P_{0|T}\right\}\right] \\
&\quad + T\cdot\ln|Q| + \operatorname{tr}\left[Q^{-1}\left\{C - B\phi^\top - \phi B^\top + \phi A\phi^\top\right\}\right] \\
&\quad + T\cdot\ln|R| \\
&\quad + \operatorname{tr}\left[R^{-1}\sum_{t=1}^{T}\left\{\left(X_t-\Lambda\hat{F}_{t|T}\right)\left(X_t-\Lambda\hat{F}_{t|T}\right)^\top + \Lambda\hat{P}_{t|T}\Lambda^\top\right\}\right]
\end{aligned}
\qquad (3.33)
$$

where the following moments can be calculated from the Kalman smoother above:

$$
\begin{array}{ll}
A = \sum_{t=1}^{T}\left(\hat{F}_{t-1|T}\hat{F}_{t-1|T}^\top + \hat{P}_{t-1|T}\right) & B = \sum_{t=1}^{T}\left(\hat{F}_{t|T}\hat{F}_{t-1|T}^\top + \hat{P}_{\{t,t-1\}|T}\right) \\
C = \sum_{t=1}^{T}\left(\hat{F}_{t|T}\hat{F}_{t|T}^\top + \hat{P}_{t|T}\right) & D = \sum_{t=1}^{T} X_t\hat{F}_{t|T}^\top \\
E = \sum_{t=1}^{T} X_t X_t^\top &
\end{array}
$$

Given these smoothed moments the Maximization step results in the following closed form estimators at iteration j

$$\begin{aligned}
\text{vec}\left(\Lambda^{(j)}\right) &= \text{vec}\left(DC^{-1}\right) && (3.34)\\
R^{(j)} &= \frac{1}{T}\left(E - DC^{-1}D^{\top}\right) && (3.35)\\
\text{vec}\left(\phi^{(j)}\right) &= \text{vec}\left(BA^{-1}\right) && (3.36)\\
Q^{(j)} &= \frac{1}{T}\left[C - BA^{-1}B^{\top}\right] && (3.37)
\end{aligned}$$

where F_t is approximated by $\hat{F}_{t|T} = E\left[F_t | \mathcal{X}_T\right]$. $\mathcal{X}_T = \{X_1, .., X_T\}$ denotes the full information set, $\hat{P}_{t|T} = var\left(F_t | \mathcal{X}_T\right)$ is the variance and $\hat{P}_{\{t,t-1\}|T} = cov\left(F_t, F_{t-1} | \mathcal{X}_T\right)$ is the lag-one covariance.

These estimates can then be used in the Expectation step to compute a new set of moments from the Kalman smoother. Subsequently, these estimates are supplied to the maximization step above and the procedure continues until convergence of the likelihood.

Likelihood function

The maximum likelihood method of Chen & Scott (1993) for a Gaussian ATSM with both macroeconomic variables and latent state variables is now presented and follows closely Ang & Piazzesi (2003) although with somewhat more details here. The measurement and state transition equation in (3.20)-(3.21) page 118 is repeated below for convenience

$$\begin{bmatrix} X_t^o \\ Y_t \end{bmatrix} = \begin{bmatrix} 0 \\ A \end{bmatrix} + \begin{bmatrix} I & 0 \\ B^o & B^u \end{bmatrix} \begin{bmatrix} X_t^o \\ X_t^u \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & B^m \end{bmatrix} \begin{bmatrix} w_t \\ v_t \end{bmatrix} \tag{3.20}$$

and

$$\begin{bmatrix} X_t^o \\ X_t^u \end{bmatrix} = \begin{bmatrix} \mu^o \\ \mu^u \end{bmatrix} + \begin{bmatrix} \Phi^{oo} & \Phi^{ou} \\ \Phi^{uo} & \Phi^{uu} \end{bmatrix} \begin{bmatrix} X_{t-1}^o \\ X_{t-1}^u \end{bmatrix} + \begin{bmatrix} \Sigma^{oo} & \Sigma^{ou} \\ \Sigma^{uo} & \Sigma^{uu} \end{bmatrix} \begin{bmatrix} \varepsilon_t^o \\ \varepsilon_t^u \end{bmatrix} \tag{3.21}$$

where $\{\mu^o, \varepsilon_t^o\}$ are $K_1 \cdot p$ vectors with zeros except for the upper K_1 elements, Φ^{oo} is a $K_1 \cdot p \times K_1 \cdot p$ companion matrix representing the pth order lag polynomial, Σ^{oo} is a $K_1 \cdot p \times K_1 \cdot p$ matrix padded with zeros except for the upper left $K_1 \times K_1$ block and $\{\Phi^{ou}, \Sigma^{ou}\}$ are $K_1 \cdot p \times K_2$ matrices. For the latent variables $\{\mu^u, \varepsilon_t^u\}$ are K_2 vectors and Σ^{uu} is a $K_2 \times K_2$ matrix. Finally, also partition the $K \times 1$ vector

δ_1 in Definition 1 into δ_1^o of dimension $K_1 \cdot p$ and δ_1^u of dimension K_2. The market prices of risk load only on contemporaneous values of X_t such that λ_0 has the same structure as μ and λ_1 has the same structure as Σ.

The parameters of interest implicit in the state space system above are $\theta = \{\delta_0, \delta_1, \mu, \Phi, \Sigma, \lambda_0, \lambda_1\}$ and need to be estimated using a maximum likelihood method that involves the joint conditional density of (Y_t, X_t^o). However, this density is not known but given the distributional assumptions about the measurement errors and the multivariate normal distribution of the one-period ahead latent state variables $X_t|X_{-1}$ it is possible to apply a change-of-variable technique to relate the density of (Y_t, X_t^o) to the density of (X_t^o, X_t^u, v_t) if X_t^u is known. Following the approach by Chen & Scott (1993) the unobserved X_t^u are inverted from a subset of Y_t by assuming that K_2 yields are measured without error leaving $\mathcal{N} - K_2$ of the yields to be measured with errors such that the measurement matrix B^m is padded with zeros except for $\mathcal{N} - K_2$ non-zero elements. The part of the measurement equation in (3.1) that represents the yields in Y_t is now partitioned into a part with K_2 rows that is measured without error (denoted by a bar) and a part with $(\mathcal{N} - K_2)$ rows that is measured with errors (denoted by a tilde)[34] such that (3.20) can be written in a reordered form as:

$$\begin{bmatrix} X_t^o \\ \bar{Y}_t \\ \tilde{Y}_t \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{A} \\ \tilde{A} \end{bmatrix} + \begin{bmatrix} I & 0 & 0 \\ \bar{B}^o & \bar{B}^u & \bar{B}^m \\ \tilde{B}^o & \tilde{B}^u & \tilde{B}^m \end{bmatrix} \begin{bmatrix} X_t^o \\ X_t^u \\ v_t \end{bmatrix} \tag{3.38}$$

where the zero measurement errors of $\bar{Y}_t$ implies that $\bar{B}^m$ contains only zeros and $\tilde{B}^m v_t \equiv \eta_t \sim N(0, H)$ where v_t is $K_2 \times 1$ and H is $K_2 \times K_2$. The latent state variables can now be recovered as a function of the perfectly observed yields

$$X_t^u = \left[\bar{B}^u\right]^{-1} \left(\bar{Y}_t - \bar{A} - \bar{B}^o \bar{X}_t^o\right). \tag{3.39}$$

Given the Jacobian J of the transformation from Y_t to X_t^u :

$$J = \begin{bmatrix} I & 0 & 0 \\ \bar{B}^o & \bar{B}^u & 0 \\ \tilde{B}^o & \tilde{B}^u & \tilde{B}^m \end{bmatrix}$$

[34] $\bar{Y}_t$ is $K_2 \times 1$, $\bar{A}$ is $K_2 \times 1$, $\tilde{B}^o$ is $K \times (K - K_2)$, $\bar{B}^u$ is $K_2 \times K_2$, $\bar{B}^m$ is $K_2 \times (N - K_2)$, $\tilde{Y}_t$ is $(N - K_2) \times 1$, $\tilde{A}$ is $(N - K_2) \times 1$, $\tilde{B}^o$ is $(N - K_2) \times (K - K_2)$, and $\tilde{B}^m$ is $(N - K_2) \times K_2$

the joint conditional density of (Y_t, X_t^o) can be written as

$$\begin{aligned} f\left(Y_t, X_t^o \middle| X_{t-1}^o, X_{t-1}^o, \mathcal{I}_t; \Psi\right) &= \frac{1}{|\det(J)|} f\left(X_t^o, X_t^u, v_t \middle| X_{t-1}^o, X_{t-1}^u, \mathcal{I}_t; \Psi\right) \\ &= \frac{1}{|\det(J)|} f\left(X_t^o, X_t^u, \eta_t \middle| X_{t-1}^o, X_{t-1}^u; \Psi\right) f\left(\eta_t \middle| \eta_{t-1}; \Psi\right) \end{aligned}$$

where $\mathcal{I}_t$ contains lagged values of the conditioning variables. The second line follows from the Markovian structure of the state variables, the definition of η_t and the assumption that the measurement errors η_t are uninformative about the states

The joint likelihood $\mathcal{L}(\Psi)$ is then given by

$$\mathcal{L}(\Psi) = \prod_{t=2}^{T} \frac{1}{|\det(J)|} f\left(X_t^o, X_t^u, \eta_t \middle| X_{t-1}^o, X_{t-1}^u; \Psi\right) f\left(\eta_t \middle| \eta_{t-1}; \Psi\right)$$

and the log likelihood is then

$$\begin{aligned} \log \mathcal{L}(\Psi) &= -(T-1)\log(|\det(J)|) + \sum_{t=2}^{T} \log f\left(X_t^o, X_t^u, \eta_t \middle| X_{t-1}^o, X_{t-1}^u; \Psi\right) + \\ & \sum_{t=2}^{T} \log f\left(\eta_t \middle| \eta_{t-1}; \Psi\right) \\ &= -(T-1)\log(|\det(J)|) - \frac{(T-1)K}{2}\log(2\pi) - \frac{(T-1)}{2}\log(\det(\Omega)) \\ & -\frac{1}{2}\sum_{t=2}^{T} (X_t - \mu - \Phi X_{t-1})^\top \Omega^{-1} (X_t - \mu - \Phi X_{t-1}) \\ & -\frac{(T-1)(\mathcal{N}-K)}{2}\log(2\pi) - \frac{(T-1)}{2}\log(\det(H)) - \frac{1}{2}\sum_{t=2}^{T} \eta_t^\top H^{-1} \eta_t \end{aligned}$$

A.5 Impulse responses and forecast error variance decompositions

Impulse response functions for the yields

The following self-contained presentation is closely related to Ang & Piazzesi (2003) as they offer a quite pedagogical presentation of how the loadings on lagged state variables load similarly on the moving-average coefficients of the state variables in the moving-average representation of the VAR in (3.1)[35].

The VAR(1) companion form[36] of the VAR with p lags is rearranged into a VAR(p) is rearranged as follows. The $(K_1 \cdot p + K_2) \times 1$ state vector X_t of current and lagged state variables $X_t = \begin{bmatrix} X_t^{o\top} & X_t^{u\top} \end{bmatrix}^\top$ with $X_t^o = \begin{bmatrix} f_t^{o\top} & f_{t-1}^{o\top} & \cdots & f_{t-p-1}^{o\top} \end{bmatrix}^\top$ and $X_t^u = \begin{bmatrix} f_t^{u\top} \end{bmatrix}$ is now split up in the yield equation in order detail how the IRF are multiplied by the loading matrix. Consider the n-period yield

$$\begin{aligned} y_t^{(n)} &= A_n + B_n^\top X_t \\ &= A_n + B_{n,0}^{o\top} f_t^o + ... + B_{n,p-1}^{o\top} f_{t-p-1}^o + B_{n0}^{u\top} f_t^u \\ &= A_n + B_{n,o}^{ou\top} F_t^{ou} + ... + B_{n,p-1}^{ou\top} F_{t-p-1}^{ou} \end{aligned}$$

where

$$F_t^{ou} = \begin{bmatrix} f_t^o \\ f_t^u \end{bmatrix}, B_{n,0}^{ou} = \begin{bmatrix} B_{n,0}^o \\ B_{n,0}^u \end{bmatrix}, \left\{F_{t-i}^{ou}\right\}_{i=1}^{p-1} = \begin{bmatrix} f_{t-i}^o \\ 0 \end{bmatrix}, \left\{B_{n,i}^{ou}\right\}_{i=1}^{p-1} = \begin{bmatrix} B_{n,i}^o \\ 0 \end{bmatrix}$$

The VAR is now with p lags:

$$F_t^{ou} = \Phi_0 + \Phi_1^{ou} F_{t-1}^{ou} + ... + \Phi_p^{ou} F_{t-p}^{ou} + U_t$$

where

$$\Phi_1^{ou} = \begin{bmatrix} \Phi_1^o & 0 \\ 0 & \Phi_1^u \end{bmatrix}, \{\Phi_i^{ou}\}_{i=2}^p = \begin{bmatrix} \Phi_i^o & 0 \\ 0 & 0 \end{bmatrix}, U_t = \begin{bmatrix} u_t^o \\ u_t^u \end{bmatrix}$$

and where $\Phi_0 = 0$, Φ_i^{ou} is a $K_1 \times K_1$ matrix corresponding to the observed state variables and Φ_1^u is a $K_2 \times K_2$ matrix corresponding to the unobserved state variables. Finally, $var\,(U) = \Omega$ which in turn is Cholesky decomposed into $\Omega = PP^\top$ such that $U_t = Pe_t$, with e_t being iid.

The moving average representation MA(∞) of F_t^{ou} is then

$$F_t = \sum_{i=0}^{\infty} \Psi_i U_{t-i} = \sum_{i=0}^{\infty} \Theta_i e_{t-i}$$

[35] However, the IRFs can be calculated equivalently as $Y_t = B_n^\top \sum_{i=1}^\infty \Phi^i u_{t-i}$ where the practical implementation involves a choice of the response period to analyze, i.e. letting i run from 1 to say s and then recursively calculate the expression above; cf. Canova (2007) chapter 4. Both approaches were applied in this paper and it can be confirmed that both approaches give similar results.

[36] Recall that Φ denotes the $(K_1 \cdot p + K_2) \times (K_1 \cdot p + K_2)$ companion matrix whereas Φ's with the subscripts $\{\Phi_i\}_{i=1}^p$ refer to the VAR with p lags.

where $e_t = P^{-1}U_t$ and $\Theta_i = \Psi_i P$ is found by the following recursion; cf. Lütkepohl (2007):

$$\Theta_s = \sum_{j=1}^{s} \Theta_{s-j}\Phi_j, \quad s = 1, 2, ..., \quad \Theta_0 = P$$

The MA(∞) form of the n-period yield is then given by:

$$y_t^{(n)} = A_n + \sum_{i=0}^{\infty} \psi_i^{(n)} e_{t-i} \tag{3.40}$$

where

$$\begin{aligned}
\psi_0^{(n)} &= B_{n,0}^{ou\top}\Theta_0 \\
\psi_1^{(n)} &= B_{n,0}^{ou\top}\Theta_1 + B_{n,1}^{ou\top}\Theta_0 \\
\psi_2^{(n)} &= B_{n,0}^{ou\top}\Theta_2 + B_{n,1}^{ou\top}\Theta_1 + B_{n,2}^{ou\top}\Theta_0 \\
&\vdots \\
\psi_{p-1}^{(n)} &= B_{n,0}^{ou\top}\Theta_{p-1} + B_{n,1}^{ou\top}\Theta_{p-2} + ... + B_{n,p-1}^{ou\top}\Theta_0 \\
&\vdots \\
\psi_i^{(n)} &= B_{n,0}^{ou\top}\Theta_i + B_{n,1}^{ou\top}\Theta_{i-2} + ... + B_{n,p-1}^{ou\top}\Theta_{i-(p-1)} \quad \text{for } i \geq (p-1)
\end{aligned}$$

Forecast error variance decompositions for the yields

Consider the jth component of the vector of yields Y_t when there are K state variables (K shocks). Furthermore, calculate the forecast error for a forecast horizon of s periods using (3.40)

$$\begin{aligned}
y_{t+s}^{(n)} - E_t\left[y_{t+s}^{(n)}\right] &= \psi_0^{(n)}e_{t+s} + \psi_1^{(n)}e_{t+s-1} + ... + \psi_{s-1}^{(n)}e_{t+1} \\
&\sum_{k=1}^{K} \psi_{k,0}^{(n)}e_{k,t+s} + \psi_{k,1}^{(n)}e_{k,t+s-1} + ... + \psi_{k,s-1}^{(n)}e_{k,t+1}
\end{aligned}$$

The corresponding mean squared error is then

$$
\begin{aligned}
MSE\left(y_{t+s|t}^{(n)}\right) &= E\left[\left(y_{t+s}^{(n)} - E_t\left[y_{t+s}^{(n)}\right]\right)\left(y_{t+s}^{(n)} - E_t\left[y_{t+s}^{(n)}\right]\right)^{\top}\right] \\
&= \sum_{k=1}^{K}\left[\left(\psi_{k,0}^{(n)}\right)^2 + \left(\psi_{k,1}^{(n)}\right)^2 + \ldots + \left(\psi_{k,s-1}^{(n)}\right)^2\right]
\end{aligned}
$$

The contribution $\omega_{k,s}^{(n)}$ of the kth factor to the MSE of the s-step ahead forecast of the n-period yield is

$$
\omega_{k,s}^{(n)} = \frac{\sum_{i=0}^{s-1}\left[\psi_{k,i}^{(n)}\right]^2}{MSE\left(y_{t+s|t}^{(n)}\right)}
$$

Notice, it would also be possible to treat the idiosyncratic noise (measurement error) in the measurement equation[37]. However, the fraction of the measurement error in (FEVD) is negligible because of the very small measurement errors.

Forecast error variance decomposition for the excess bond returns

In the previous section I showed how to compute a forecast as of time t of a future n-period yield at time $t+s$ and then evaluate the time t mean squared error of this forecast. In this section the future yield y_{t+s} is replaced by a future excess m-period holding period return $rx_{t+s,t+m+s}^{(n)}$. Then the time t mean squared error of this forecast is constructed. The whole purpose of this exercise is to show how forecasts of excess returns are influenced by shocks to the state variables at difference forecast horizons.

The starting point is the m-period holding period return as of time t in equation (3.11) and (3.14) repeated below for convenience

$$
\begin{aligned}
rx_{t,t+m}^{(n)} &= \mathcal{A}_{n-m} - \mathcal{A}_n + \mathcal{A}_m + \mathcal{B}_{n-m}^{\top}X_{t+m} - \mathcal{B}_n^{\top}X_t + \mathcal{B}_m^{\top}X_t \\
&= a_{n,m} + b_{n,m}^{\top}X_t + \mathcal{B}_{n-m}^{\top}\sum_{i=0}^{m-1}\Phi^i u_{t+m-i}
\end{aligned}
$$

[37] See Bork (2008) for this calculation.

where

$$\begin{aligned} a_{n,m} &= \mathcal{B}_{n-m}^{\top}\sum_{i=0}^{m-1}\Phi^{i}\mu+\mathcal{A}_{n-m}-\mathcal{A}_{n}+\mathcal{A}_{m} \\ b_{n,m}^{\top} &= \mathcal{B}_{n-m}^{\top}\Phi^{m}-\mathcal{B}_{n}^{\top}+\mathcal{B}_{m}^{\top} \end{aligned}$$

Now, consider the same return but one period later, i.e. the excess holding period return from $t+1$ to $t+m+1$:

$$\begin{aligned} rx_{t+1,t+m+1}^{(n)} &= \mathcal{A}_{n-m}-\mathcal{A}_{n}+\mathcal{A}_{m}+\mathcal{B}_{n-m}^{\top}X_{t+m+1}-\mathcal{B}_{n}^{\top}X_{t+1}+\mathcal{B}_{m}^{\top}X_{t+1} \\ &= a_{n,m}+b_{n,m}^{\top}\left(\mu+\Phi X_{t}+u_{t+1}\right)+\mathcal{B}_{n-m}^{\top}\sum_{i=0}^{m-1}\Phi^{i}u_{t+m+1-i} \end{aligned}$$

implying that the forecast error can be calculated as

$$rx_{t+1,t+m+1}^{(n)}-E_{t}\left[rx_{t+1,t+m+1}^{(n)}\right]=b_{n,m}^{\top}u_{t+1}+\mathcal{B}_{n-m}^{\top}\sum_{i=0}^{m-1}\Phi^{i}u_{t+m+1-i}$$

whereas the forecast error for the period $t+2$ to $t+m+2$ is:

$$rx_{t+2,t+m+2}^{(n)}-E_{t}\left[rx_{t+2,t+m+2}^{(n)}\right]=b_{n,m}^{\top}\left(\Phi u_{t+1}+u_{t+2}\right)+\mathcal{B}_{n-m}^{\top}\sum_{i=0}^{m-1}\Phi^{i}u_{t+m+2-i}$$

and generally for time $t+s$ to $t+m+s$

$$rx_{t+s,t+m+s}^{(n)}-E_{t}\left[rx_{t+s,t+m+s}^{(n)}\right]=b_{n,m}^{\top}\sum_{j=0}^{s-1}\Phi^{j}u_{t+s-j}+\mathcal{B}_{n-m}^{\top}\sum_{i=0}^{m-1}\Phi^{i}u_{t+m+s-i}$$

Notice, that the first term relates to the uncertainty of all prices until the forecast calculations starts whereas the second term relates to the uncertainty of the selling price of the $n-m$ period bond. The mean squared forecast error (MSFE) as of time t of these forecasts up to horizon s is then simply

$$MSFE_{t}\left(rx_{t+s,t+m+s}^{(n)}\right)=\sum_{j=0}^{s-1}b_{n,m}^{\top}\Phi^{j}\Omega\left[\Phi^{j}\right]^{\top}b_{n,m}+\sum_{i=0}^{m-1}\mathcal{B}_{n-m}^{\top}\Phi^{i}\Omega\left[\Phi^{i}\right]\mathcal{B}_{n-m}$$

Apply a Cholesky decomposition to Ω such that $\Omega=\Sigma\Sigma^{\top}$ and denote by the $1\times K$

vectors

$$\begin{aligned}\varsigma_{n,m}^{(j)} &= b_{n,m}^{\top}\Phi^{j}\Sigma \\ \vartheta_{n-m}^{(i)} &= \mathcal{B}_{n-m}^{\top}\Phi^{i}\Sigma\end{aligned}$$

which then allows us to rewrite the MSFE as

$$MSFE_t\left(rx_{t+s,t+m+s}^{(n)}\right) = \sum_{j=0}^{s-1} \varsigma_{n,m}^{(j)} \left[\varsigma_{n,m}^{(j)}\right]^{\top} + \sum_{i=0}^{m-1} \vartheta_{n-m}^{(i)} \left[\vartheta_{n-m}^{(i)}\right]^{\top}$$

In particular the contribution from the kth shock

$$MSFE_t\left(rx_{t+s,t+m+s}^{(n)}(k)\right) = \sum_{j=0}^{s-1} \left[\varsigma_{n,m,k}^{(j)}\right]^{2} + \sum_{i=0}^{m-1} \left[\vartheta_{n-m,k}^{(i)}\right]^{2}$$

The contribution $w_{k,s}^{(n-m)}$ of the kth factor to the MSFE of the s-step ahead forecast of the m-period holding period excess return for a n-period bond is then:

$$w_{k,s}^{(n-m)} = \frac{\sum_{j=0}^{s-1}\left[\varsigma_{n,m,k}^{(j)}\right]^{2} + \sum_{i=0}^{m-1}\left[\vartheta_{n-m,k}^{(i)}\right]^{2}}{MSFE_t\left(rx_{t+s,t+m+s}^{(n)}\right)}$$

Table 3.1: Different sets of macroeconomic state variables and yield fit

Panel A: Filtered state variables $\left(X_{t|t}\right)$ from large panel dynamic factor analysis: Inflation and unemployment

Maturity	$\overline{y_t - \hat{y}_t}$	$\overline{\lvert y_t - \hat{y}_t \rvert}$	$\max\left(\overline{\lvert y_t - \hat{y}_t \rvert}\right)$
1m	-0.0129	0.0743	0.7067
3m	0.0139	0.1256	1.0529
12m	-0.0090	0.0825	0.5541
36m	0.0054	0.0615	0.3602
60m	0.0008	0.0558	0.3481

Panel B: Filtered state variables $\left(X_{t|t}\right)$ from large panel dynamic factor analysis: Inflation and hours-in-production

Maturity	$\overline{y_t - \hat{y}_t}$	$\overline{\lvert y_t - \hat{y}_t \rvert}$	$\max\left(\overline{\lvert y_t - \hat{y}_t \rvert}\right)$
1m	-0.0167	0.0734	0.6560
3m	0.0153	0.1309	1.1571
12m	-0.0077	0.0907	0.5817
36m	0.0053	0.0624	0.3922
60m	0.0019	0.0618	0.3877

Panel C: Filtered state variables $\left(X_{t|t}\right)$ from large panel dynamic factor analysis: Inflation and economic activity

Maturity	$\overline{y_t - \hat{y}_t}$	$\overline{\lvert y_t - \hat{y}_t \rvert}$	$\max\left(\overline{\lvert y_t - \hat{y}_t \rvert}\right)$
1m	-0.0168	0.0727	0.6603
3m	0.0159	0.1281	1.2331
12m	-0.0103	0.0828	0.6158
36m	0.0069	0.0631	0.4278
60m	0.0032	0.0579	0.4063

Panel D: Filtered state variables $\left(X_{t|t}\right)$ from large panel dynamic factor analysis: Inflation and consumption

Maturity	$\overline{y_t - \hat{y}_t}$	$\overline{\lvert y_t - \hat{y}_t \rvert}$	$\max\left(\overline{\lvert y_t - \hat{y}_t \rvert}\right)$
1m	-0.0138	0.0680	0.6790
3m	0.0114	0.1220	1.2992
12m	-0.0069	0.0864	0.7109
36m	0.0058	0.0627	0.4267
60m	0.0027	0.0578	0.3865

The model specification follows the discussion of identification in 3.3.1 page 3.3.1 with $K_1 = 2$, $p = 8$ and $K_2 = 3$.The models are estimated by the Kalman filter. The column heading $\overline{y_t - \hat{y}_t}$ means the average of fitting errors of the yields measured in percentage points, $\overline{\lvert y_t - \hat{y}_t \rvert}$ means average of the absolute value of the errors and $\max\left(\overline{\lvert y_t - \hat{y}_t \rvert}\right)$ means the maximum of the absolute value of the fitting errors.

Table 3.2: Ang and Piazzesi like model

$$\delta_1^u = \begin{bmatrix} \underset{(0.000024)}{0.000110^{**}} \\ \underset{(0.000010)}{-0.000435^{**}} \\ \underset{(0.000023)}{0.000312^{**}} \end{bmatrix} \qquad \lambda_0^u = \begin{bmatrix} \underset{(0.0041)}{-0.0437^{**}} \\ \\ \end{bmatrix}$$

$$\Phi^{uu} = \begin{bmatrix} \underset{(0.0047)}{0.9923^{**}} & & \\ & \underset{(0.0077)}{0.9712^{**}} & \\ & \underset{(0.0153)}{-0.0452^{**}} & \underset{(0.0216)}{0.8267^{**}} \end{bmatrix} \qquad \lambda_1^{uu} = \begin{bmatrix} \underset{(0.0047)}{-0.0044} & & \\ \underset{(0.0074)}{0.0543^{**}} & . & \underset{(0.0234)}{-0.2505^{**}} \\ \underset{(0.0121)}{-0.0507^{**}} & & \underset{(0.0296)}{0.2421^{**}} \end{bmatrix}$$

$$\sigma_H = \begin{bmatrix} \underset{(0.000005)}{0.000247^{**}} & \\ & \underset{(0.000002)}{0.000096^{**}} \end{bmatrix} \qquad \lambda_1^{oo} = \begin{bmatrix} \underset{(0.2421)}{1.5244^{**}} & \underset{(0.0930)}{-0.0096} \\ \underset{(0.2428)}{2.0839^{**}} & \underset{(0.1070)}{0.2000^{\diamond}} \end{bmatrix}$$

Note:The remaining macroeconomic parameters are kept fixed.

Maturity	$\overline{y_t - \hat{y}_t}$	$\overline{\lvert y_t - \hat{y}_t \rvert}$	$\max\left(\overline{\lvert y_t - \hat{y}_t \rvert}\right)$
1m	-0.0000	0.0000	0.0000
3m	0.0626	0.2132	1.7863
12m	-0.0000	0.0000	0.0000
36m	0.0029	0.0883	0.6883
60m	-0.0000	0.0000	0.0000

This model is based on the same specification and scaling as the model in Table 7 in Ang & Piazzesi (2003) and the same data adjusted to the sample period in this paper. The model specification follows the discussion of identification in Section 3.3.1 with $K_1 = 2, p = 12, K_2 = 3$ and is estimated by the Chen & Scott (1993) method described in Appendix A.4. The column heading $\overline{y_t - \hat{y}_t}$ means the average of the fitting errors of the yields measured in basis points., $\overline{\lvert y_t - \hat{y}_t \rvert}$ means the average of the absolute value of the errors and $\max\left(\overline{\lvert y_t - \hat{y}_t \rvert}\right)$ means the maximum of the absolute value of the fitting errors. Standard errors in parenthesis are calculated by the outer product of the gradient, where the gradient is calculated numerically using two-sided finite difference calculations.

Table 3.3: The preferred ATSM.

$$
\delta_1^u = \begin{bmatrix} \underset{(0.0319)}{0.1429^{**}} \\ \underset{(0.0268)}{-0.2841^{**}} \\ \underset{(0.0173)}{0.5211^{**}} \end{bmatrix}
\qquad
\lambda_0^u = \begin{bmatrix} \underset{(0.0103)}{-0.0477^{**}} \\ \\ \end{bmatrix}
$$

$$
\Phi^{uu} = \begin{bmatrix} \underset{(0.0017)}{0.9949^{**}} & & \\ & \underset{(0.0085)}{0.9256^{**}} & \\ & \underset{(0.0199)}{-0.0631^{**}} & \underset{(0.0246)}{0.8432^{**}} \end{bmatrix}
\qquad
\lambda_1^{uu} = \begin{bmatrix} & & \\ & & \underset{(0.0114)}{0.0456^{**}} \\ \underset{(0.0157)}{-0.1167^{**}} & \underset{(0.0290)}{0.2082^{**}} & \underset{(0.0240)}{0.3138^{**}} \end{bmatrix}
$$

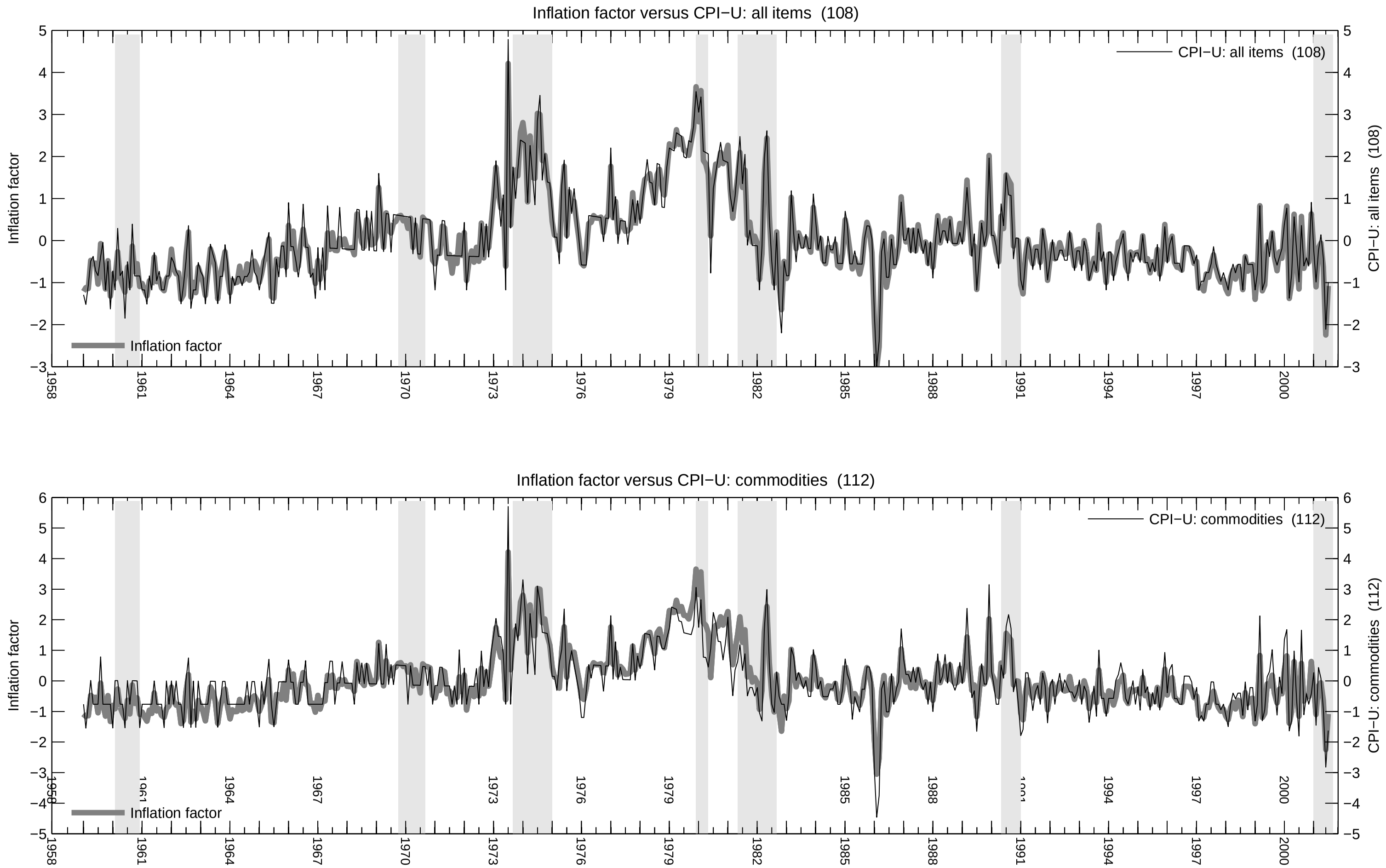
$$
\sigma_H = \begin{bmatrix} \underset{(0.0018)}{0.1578^{**}} & & \\ & \ddots & \\ & & 0.1578 \end{bmatrix}
\qquad
\lambda_1^{oo} = \begin{bmatrix} \underset{(0.0582)}{0.2406^{**}} & \underset{(0.0139)}{-0.0695^{**}} \\ \underset{(0.0916)}{-0.3346^{**}} & \underset{(0.0446)}{-0.2332^{**}} \end{bmatrix}
$$

Note:The remaining macroeconomic parameters are kept fixed.

Maturity	$\overline{y_t - \hat{y}_t}$	$\overline{\lvert y_t - \hat{y}_t \rvert}$	$\max\left(\overline{\lvert y_t - \hat{y}_t \rvert}\right)$
1m	-0.0129	0.0743	0.7067
3m	0.0139	0.1256	1.0529
12m	-0.0090	0.0825	0.5541
36m	0.0054	0.0615	0.3602
60m	0.0008	0.0558	0.3481

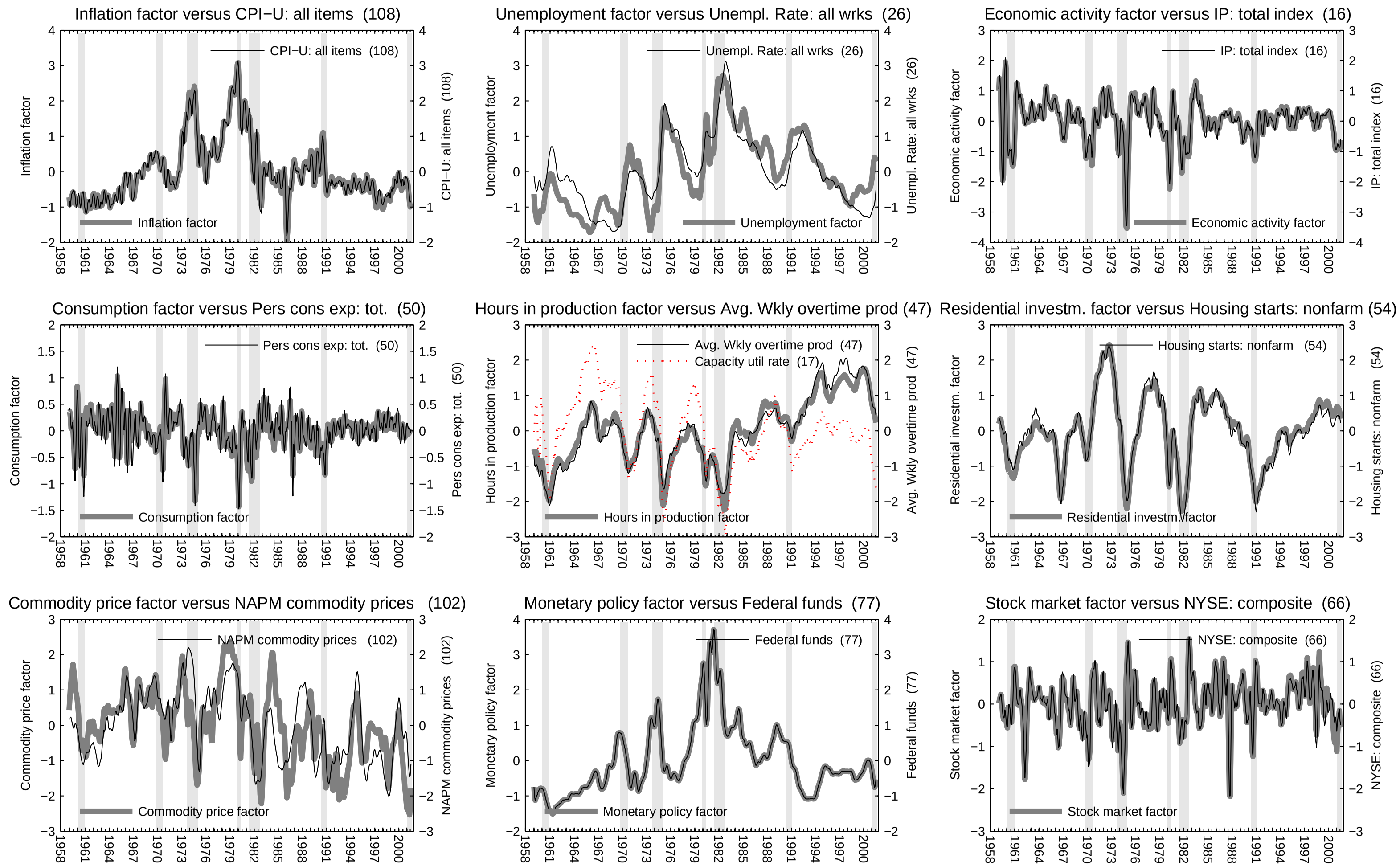
The model specification follows the discussion of identification in 3.3.1 page 3.3.1 with $K_1 = 2, p = 8, K_2 = 3$ and is estimated by the Kalman filter method. The column heading $\overline{y_t - \hat{y}_t}$ means the average of fitting errors of the yields measured in percentage points, $\overline{\lvert y_t - \hat{y}_t \rvert}$ means the average of the absolute value of the errors and $\max\left(\overline{\lvert y_t - \hat{y}_t \rvert}\right)$ means the maximum of the absolute value of the fitting errors. Standard errors in parenthesis are calculated by the outer product of the gradient, where the gradient is calculated numerically using two-sided finite difference calculations.

Figure 3.1: The time series of the the general inflation state variable versus two related observed inflation variables.



The shaded bars represent NBER-dated recessions. Numbers in parenthesis refer to the variable number in the panel; see the data appendix. To smooth series we have taken two-sided moving-averages of the original series.

Figure 3.2: The time series of the nine factors derived in Bork, Dewachter and Houssa - plotted against related observed variables.



The shaded bars represent NBER-dated recessions. Numbers in parenthesis refer to the variable number in the panel; see the data appendix. To smooth series we have taken two-sided moving-averages of the original series.

Figure 3.3: IRF for expected excess returns on bonds: The case of $\hat{X}_{t|t} = \{$inflation, unemployment$\}$ with $p = 8$

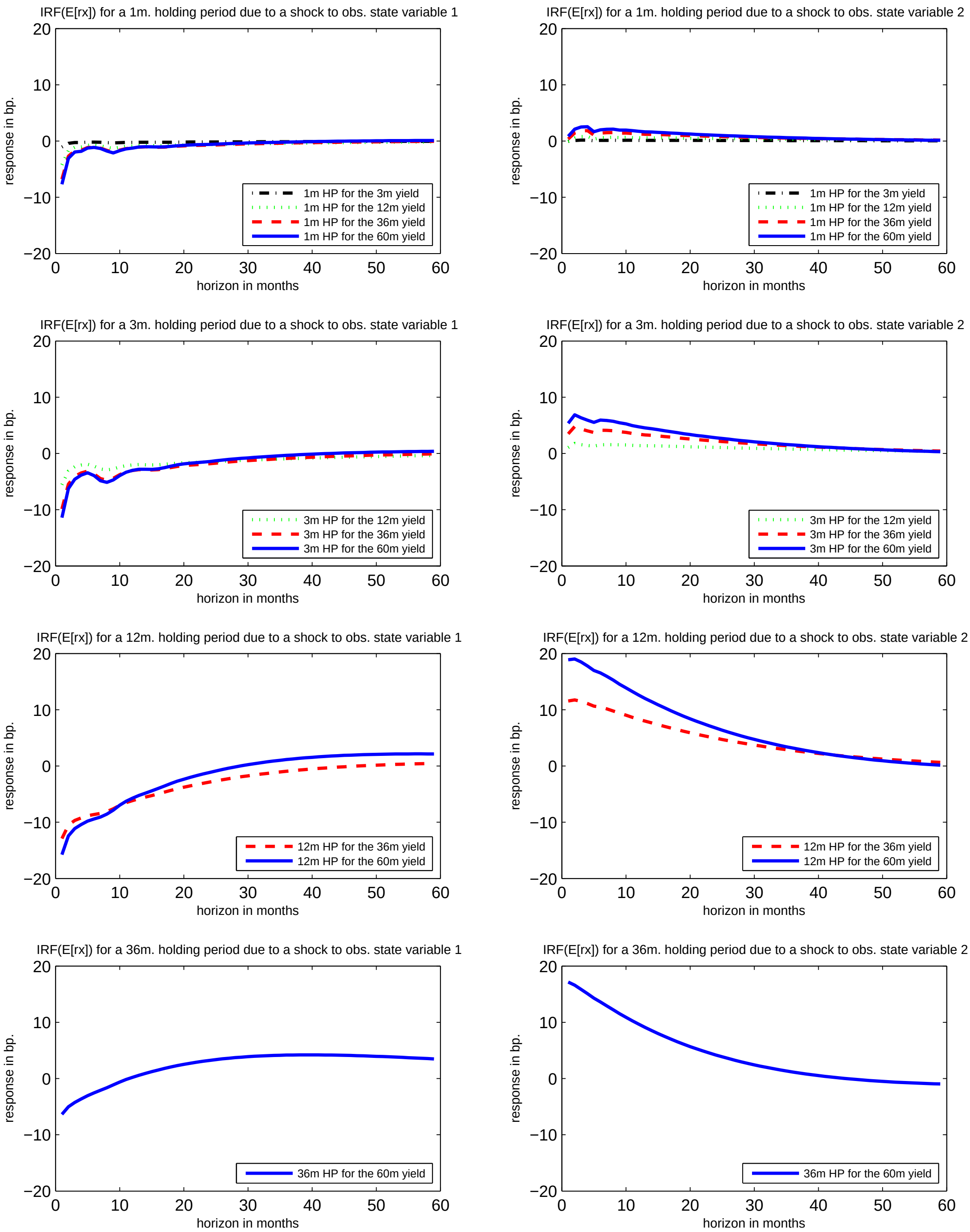


Figure 3.4: IRF for expected excess returns on bonds: The case of $\hat{X}_{t|t} = \{\text{inflation, hours-in-production}\}$ with $p = 8$

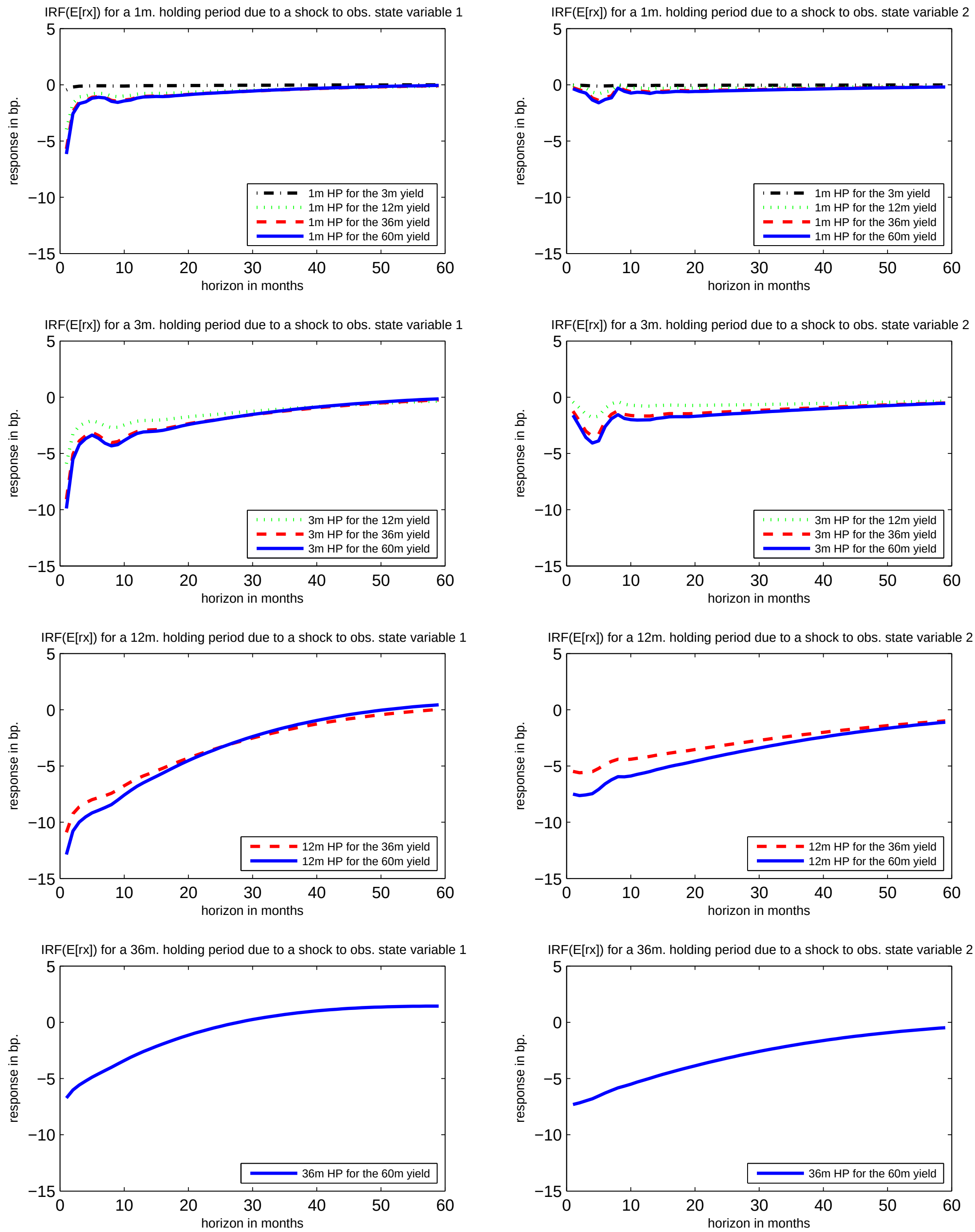


Figure 3.5: IRF for expected excess returns on bonds: The case of $\hat{X}_{t|t} =$ {inflation, economic activity} with $p = 8$

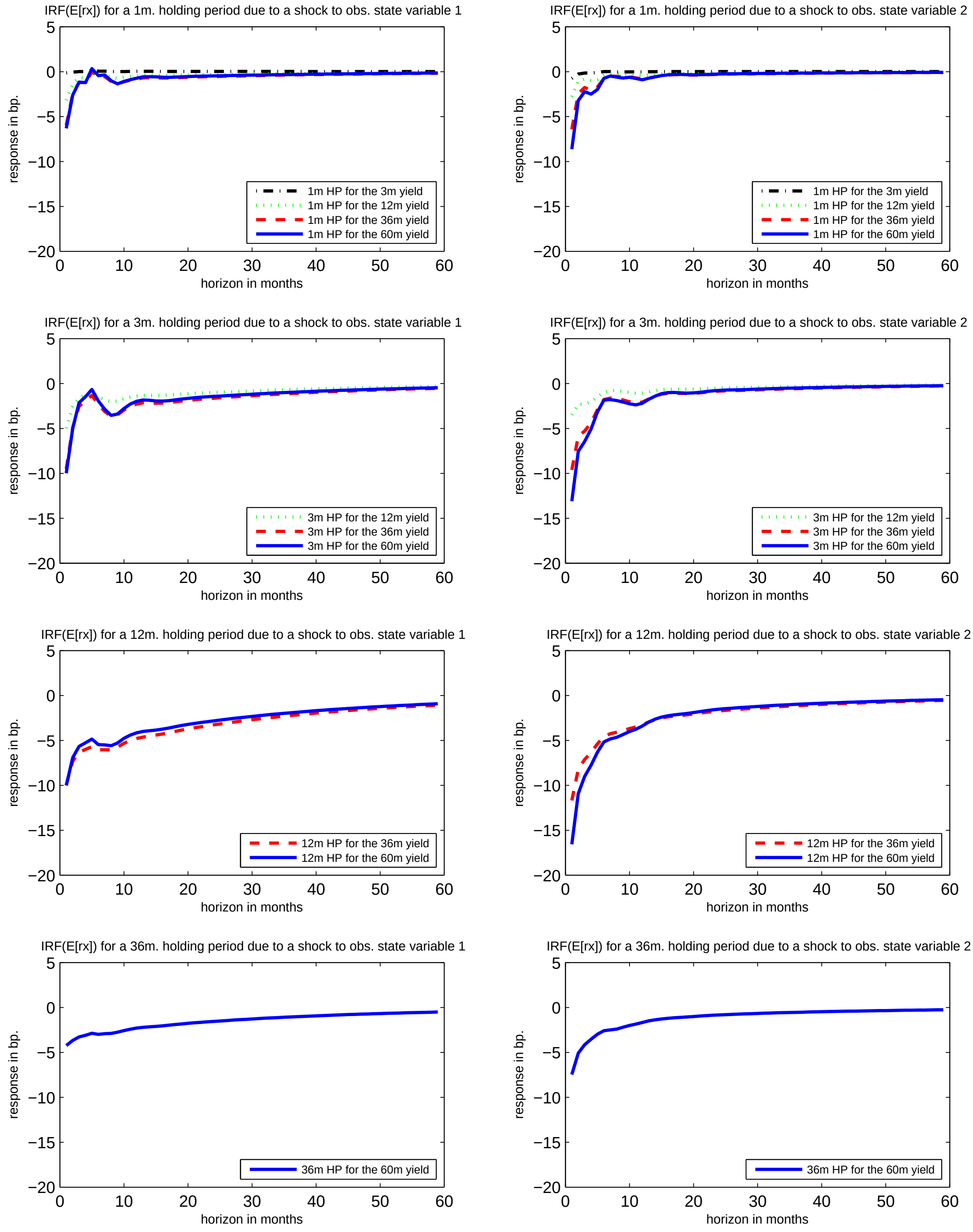


Figure 3.6: IRF for expected excess returns on bonds: The case of $\hat{X}_{t|t} = \{\text{inflation, consumption}\}$ with $p = 8$

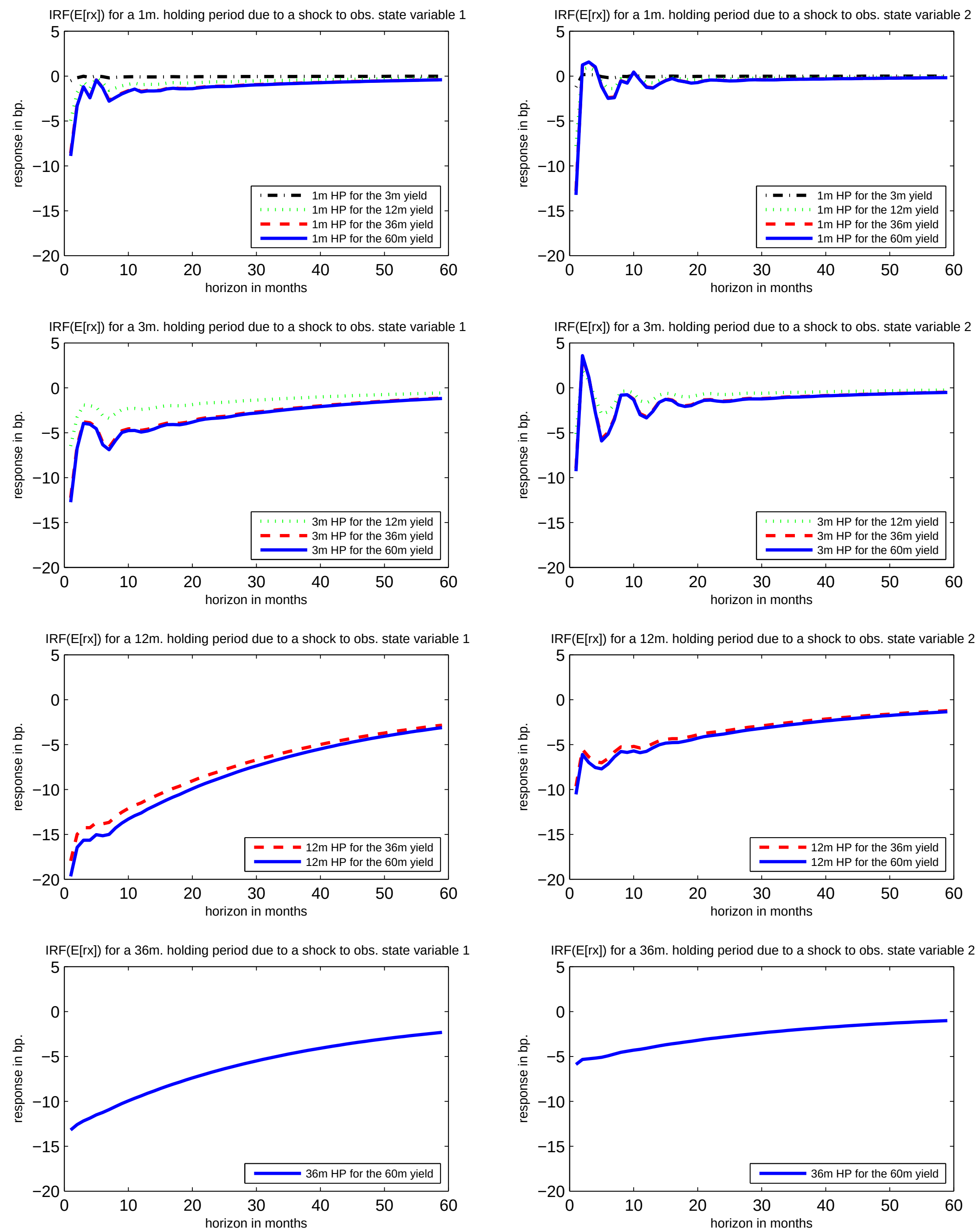


Figure 3.7: IRF for expected excess returns on bonds: The Ang and Piazzesi like model with $p = 8$

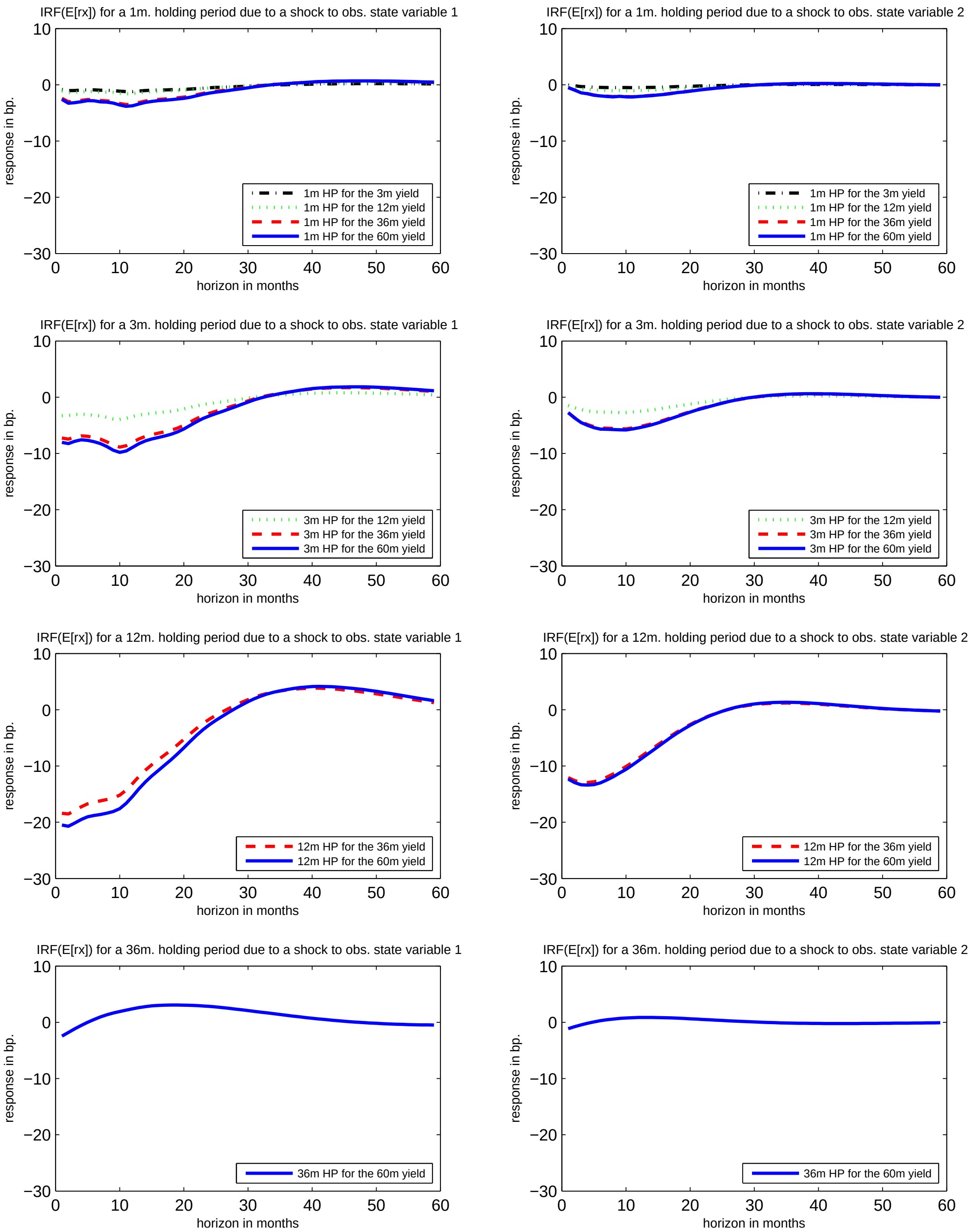
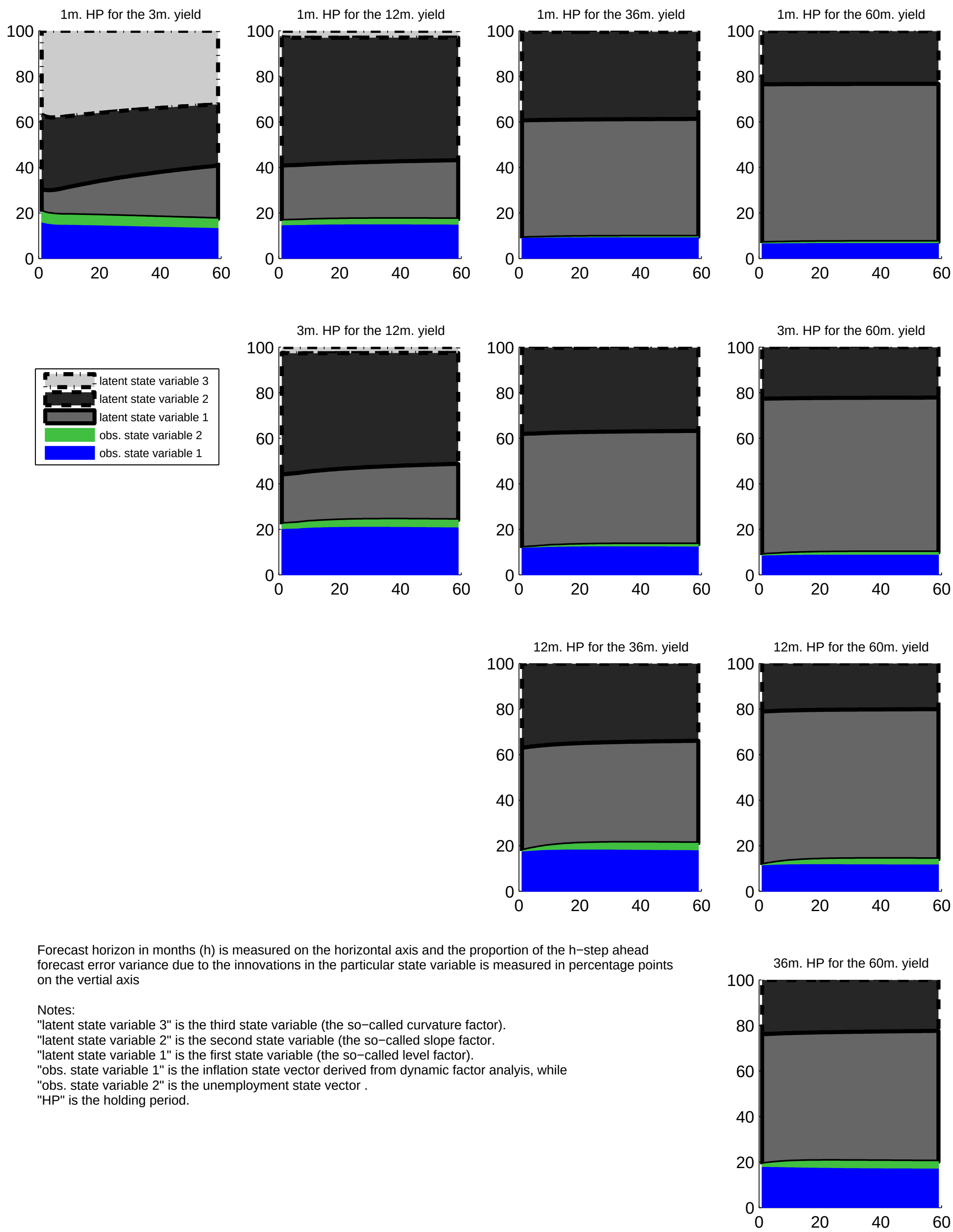


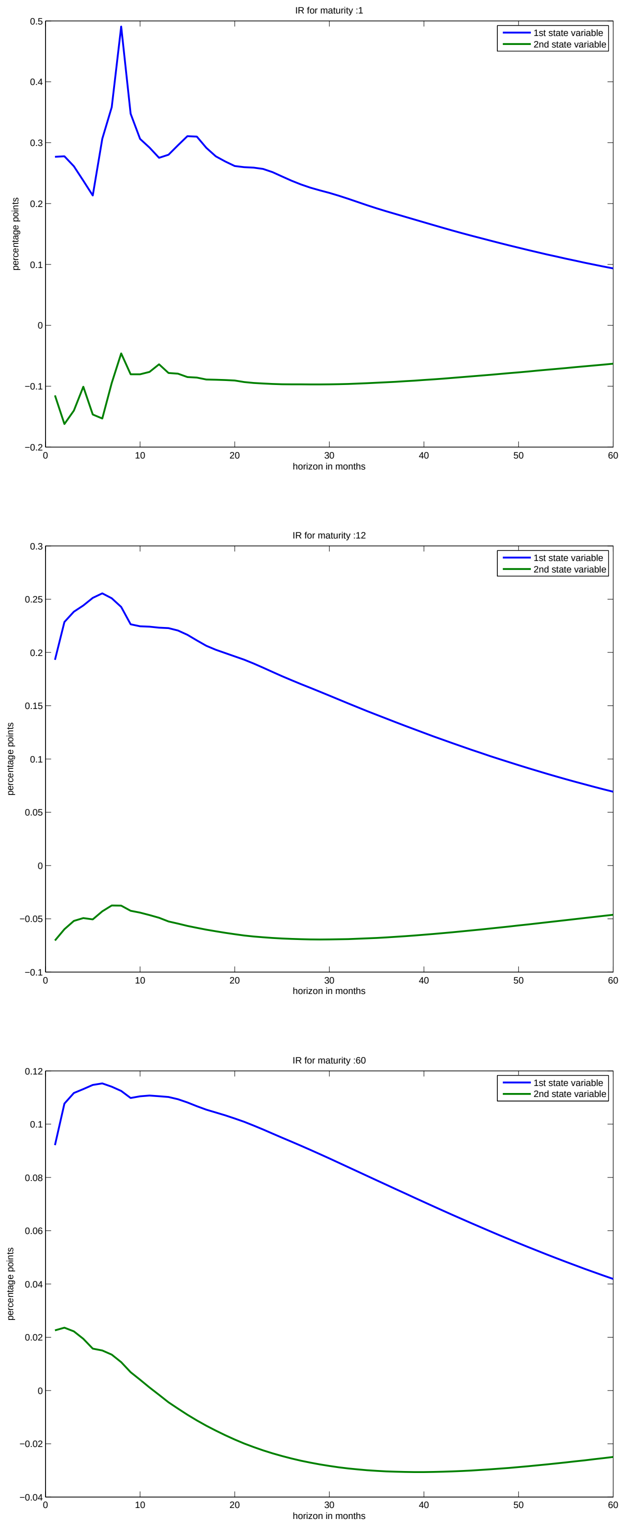
Figure 3.8: FEVD for the bond excess return: The case of $\hat{X}_{t|t} = \{\text{inflation, unemployment}\}$ with $p = 8$



Forecast horizon in months (h) is measured on the horizontal axis and the proportion of the h−step ahead forecast error variance due to the innovations in the particular state variable is measured in percentage points on the vertial axis

Notes:
"latent state variable 3" is the third state variable (the so−called curvature factor).
"latent state variable 2" is the second state variable (the so−called slope factor.
"latent state variable 1" is the first state variable (the so−called level factor).
"obs. state variable 1" is the inflation state vector derived from dynamic factor analyis, while
"obs. state variable 2" is the unemployment state vector .
"HP" is the holding period.

Figure 3.9: "Impulse response functions for 12, 36 and 60-month yields following a one standard deviation shock to inflation and unemployment".



APPENDIX A

Appendices

A.1 Data description

Data are from Bernanke et al. (2005).

The second column is a mnemonic and a * indicates a "slow-moving" variable. Fourth column contains transformation codes. "level" indicates an untransformed variable, say x_t. "ln" means $\ln x_t$ and "$\Delta \ln$" means $\ln x_t - \ln x_{t-1}$.

Real output and income

1	IPP*	1959:01–2001:08	Δ ln	Industrial production: products, total (1992=100,SA)
2	IPF*	1959:01–2001:08	Δ ln	Industrial production: final products (1992=100,SA)
3	IPC*	1959:01–2001:08	Δ ln	Industrial production: consumer goods (1992=100,SA)
4	IPCD*	1959:01–2001:08	Δ ln	Industrial production: durable cons. goods (1992=100,SA)
5	IPCN*	1959:01–2001:08	Δ ln	Industrial production: nondurable cons. goods (1992=100,SA)
6	IPE*	1959:01–2001:08	Δ ln	Industrial production: business equipment (1992=100,SA)
7	IPI*	1959:01–2001:08	Δ ln	Industrial production: intermediate products (1992=100,SA)
8	IPM*	1959:01–2001:08	Δ ln	Industrial production: materials (1992=100,SA)
9	IPMD*	1959:01–2001:08	Δ ln	Industrial production: durable goods materials (1992=100,SA)
10	IPMND*	1959:01–2001:08	Δ ln	Industrial production: nondur. goods materials (1992=100,SA)
11	IPMFG*	1959:01–2001:08	Δ ln	Industrial production: manufacturing (1992=100,SA)
12	IPD*	1959:01–2001:08	Δ ln	Industrial production: durable manufacturing (1992=100,SA)
13	IPN*	1959:01–2001:08	Δ ln	Industrial production: nondur. manufacturing (1992=100,SA)
14	IPMIN*	1959:01–2001:08	Δ ln	Industrial production: mining (1992=100,SA)
15	IPUT*	1959:01–2001:08	Δ ln	Industrial production: utilities (1992=100,SA)
16	IP*	1959:01–2001:08	Δ ln	Industrial production: total index (1992=100,SA)
17	IPXMCA*	1959:01–2001:08	level	Capacity util rate: manufac., total (% of capacity,SA) (frb)
18	PMI*	1959:01–2001:08	level	Purchasing managers' index (SA)
19	PMP*	1959:01–2001:08	level	NAPM production index (percent)
20	GMPYQ*	1959:01–2001:08	Δ ln	Personal income (chained) (series #52) (bil 92$,SAAR)
21	GMYXPQ*	1959:01–2001:08	Δ ln	Personal inc. less trans. payments (chained) (#51) (bil 92$,SAAR)

(Un)employment and hours

22	LHEL*	1959:01–2001:08	Δ ln	Index of help-wanted advertising in newspapers (1967=100;SA)
23	LHELX*	1959:01–2001:08	ln	Employment: ratio; help-wanted ads: no. unemployed clf
24	LHEM*	1959:01–2001:08	Δ ln	Civilian labor force: employed, total (thous.,SA)
25	LHNAG*	1959:01–2001:08	Δ ln	Civilian labor force: employed, nonag. industries (thous.,SA)
26	LHUR*	1959:01–2001:08	level	Unemployment rate: all workers, 16 years and over (%,SA)
27	LHU680*	1959:01–2001:08	level	Unemploy. by duration: average (mean) duration in weeks (SA)
28	LHU5*	1959:01–2001:08	level	Unemploy. by duration: pers unempl. less than 5 wks (thous.,SA)
29	LHU14*	1959:01–2001:08	level	Unemploy. by duration: pers unempl. 5 to 14 wks (thous.,SA)
30	LHU15*	1959:01–2001:08	level	Unemploy. by duration: pers unempl. 15 wks=(thous.,SA)
31	LHU26*	1959:01–2001:08	level	Unemploy. by duration: pers unempl. 15 to 26 wks (thous.,SA)
32	LPNAG*	1959:01–2001:08	Δ ln	Employees on nonag. payrolls: total (thous.,SA)
33	LP*	1959:01–2001:08	Δ ln	Employees on nonag. payrolls: total, private (thous.,SA)
34	LPGD*	1959:01–2001:08	Δ ln	Employees on nonag. payrolls: goods-producing (thous.,SA)
35	LPMI*	1959:01–2001:08	Δ ln	Employees on nonag. payrolls: mining (thous.,SA)
36	LPCC*	1959:01–2001:08	Δ ln	Employees on nonag. payrolls: contract construc. (thous.,SA)
37	LPEM*	1959:01–2001:08	Δ ln	Employees on nonag. payrolls: manufacturing (thous.,SA)
38	LPED*	1959:01–2001:08	Δ ln	Employees on nonag. payrolls: durable goods (thous.,SA)
39	LPEN*	1959:01–2001:08	Δ ln	Employees on nonag. payrolls: nondurable goods (thous.,SA)
40	LPSP*	1959:01–2001:08	Δ ln	Employees on nonag. payrolls: service-producing (thous.,SA)
41	LPTU*	1959:01–2001:08	Δ ln	Employees on nonag. payrolls: trans. and public util. (thous.,SA)
42	LPT*	1959:01–2001:08	Δ ln	Employees on nonag. payrolls: wholesale and retail (thous.,SA)
43	LPFR*	1959:01–2001:08	Δ ln	Employees on nonag. payrolls: finance, ins. and real est (thous.,SA)
44	LPS*	1959:01–2001:08	Δ ln	Employees on nonag. payrolls: services (thous.,SA)
45	LPGOV*	1959:01–2001:08	Δ ln	Employees on nonag. payrolls: government (thous.,SA)
46	LPHRM*	1959:01–2001:08	level	Avg. weekly hrs. of production wkrs.: manufacturing (sa)
47	LPMOSA*	1959:01–2001:08	level	Avg. weekly hrs. of prod. wkrs.: mfg., overtime hrs. (sa)
48	PMEMP*	1959:01–2001:08	level	NAPM employment index (percent)

Consumption

49	GMCQ*	1959:01–2001:08	Δ ln	Pers cons exp (chained)—total (bil 92$,SAAR)
50	GMCDQ*	1959:01–2001:08	Δ ln	Pers cons exp (chained)—tot. dur. (bil 96$,SAAR)
51	GMCNQ*	1959:01–2001:08	Δ ln	Pers cons exp (chained)—nondur. (bil 92$,SAAR)
52	GMCSQ*	1959:01–2001:08	Δ ln	Pers cons exp (chained)—services (bil 92$,SAAR)
53	GMCANQ*	1959:01–2001:08	Δ ln	Personal cons expend (chained)—new cars (bil 96$,SAAR)

Housing starts and sales

54	HSFR	1959:01–2001:08	ln	Housing starts: nonfarm (1947–1958); tot. (
55	HSNE	1959:01–2001:08	ln	Housing starts: northeast (thous.u.)s.a.
56	HSMW	1959:01–2001:08	ln	Housing starts: midwest (thous.u.)s.a.
57	HSSOU	1959:01–2001:08	ln	Housing starts: south (thous.u.)s.a.
58	HSWST	1959:01–2001:08	ln	Housing starts: west (thous.u.)s.a.
59	HSBR	1959:01–2001:08	ln	Housing authorized: total new priv housing (thous.,SAAR)
60	HMOB	1959:01–2001:08	ln	Mobile homes: manufacturers' shipments (thous. of units,SAAR)

Real inventories, ordes and unfilled orders

61	MNV	1959:01–2001:08	level	NAPM inventories index (percent)
62	PMNO	1959:01–2001:08	level	NAPM new orders index (percent)
63	PMDEL	1959:01–2001:08	level	NAPM vendor deliveries index (percent)
64	MOCMQ	1959:01–2001:08	Δ ln	New orders (net)—consumer goods and materials, 1992 $ (bci)
65	MSONDQ	1959:01–2001:08	Δ ln	New orders, nondefense capital goods, in 1992 $s (bci)

Stock prices

66	FSNCOM	1959:01–2001:08	Δ ln	NYSE composite (12/31/65=50)
67	FSPCOM	1959:01–2001:08	Δ ln	S&P's composite (1941–1943=10)
68	FSPIN	1959:01–2001:08	Δ ln	S&P's industrials (1941–1943=10)
69	FSPCAP	1959:01–2001:08	Δ ln	S&P's capital goods (1941–1943=10)
70	FSPUT	1959:01–2001:08	Δ ln	S&P's utilities (1941–1943=10)
71	FSDXP	1959:01–2001:08	level	S&P's composite common stock: dividend yield (% per annum)
72	FSPXE	1959:01–2001:08	level	S&P's composite common stock: price-earnings ratio (%,NSA)

Foreign exchange rates

73	EXRSW	1959:01–2001:08	Δ ln	Foreign exchange rate: Switzerland (swiss franc per US$)
74	EXRJAN	1959:01–2001:08	Δ ln	Foreign exchange rate: Japan (yen per US$)
75	EXRUK	1959:01–2001:08	Δ ln	Foreign exchange rate: United Kingdom (cents per pound)
76	EXRCAN	1959:01–2001:08	Δ ln	Foreign exchange rate: Canada (canadian $ per US$)

Interest rates and spreads

77	FYFF	1959:01–2001:08	level	Interest rate: federal funds (effective) (% per annum,nsa)
78	FYGM3	1959:01–2001:08	level	Interest rate: US T-bill,sec mkt,3-mo. (% per ann,nsa)
79	FYGM6	1959:01–2001:08	level	Interest rate: US T-bill,sec mkt,6-mo. (% per ann,nsa)
80	FYGT1	1959:01–2001:08	level	Interest rate: UST const matur., 1-yr. (% per ann,nsa)
81	FYGT5	1959:01–2001:08	level	Interest rate: UST const matur., 5-yr. (% per ann,nsa)
82	FYGT10	1959:01–2001:08	level	Interest rate: UST const matur., 10-yr. (% per ann,nsa)
83	FYAAAC	1959:01–2001:08	level	Bond yield: Moody's AAA corporate (% per annum)
84	FYBAAC	1959:01–2001:08	level	Bond yield: Moody's Baa corporate (% per annum)
85	SFYGM3	1959:01–2001:08	level	Spread fygM3—fyff
86	SFYGM6	1959:01–2001:08	level	Spread fygm6—fyff
87	SFYGT1	1959:01–2001:08	level	Spread fygt1—fyff
88	SFYGT5	1959:01–2001:08	level	Spread fygt5—fyff
89	SFYGT10	1959:01–2001:08	level	Spread fygt10—fyff
90	SFYAAAC	1959:01–2001:08	level	Spread fyaaac—fyff
91	SFYBAAC	1959:01–2001:08	level	Spread fybaac—fyff

Money and credit quantity aggregates

92	FM1	1959:01–2001:08	Δ ln	Money stock: M1 (bil$,SA)
93	FM2	1959:01–2001:08	Δ ln	Money stock: M2 (bil$,SA)
94	FM3	1959:01–2001:08	Δ ln	Money stock: M3 (bil$,SA)
95	FM2DQ	1959:01–2001:08	Δ ln	Money supply—M2 in 1992 $s (bci)
96	FMFBA	1959:01–2001:08	Δ ln	Monetary base, adj for reserve requirement changes (mil$,SA)
97	FMRRA	1959:01–2001:08	Δ ln	Depository inst reserves: total, adj for res. req chgs (mil$,SA)
98	FMRNBA	1959:01–2001:08	Δ ln	Depository inst reserves: nonbor., adj res req chgs (mil$,SA)
99	FCLNQ	1959:01–2001:08	Δ ln	Commercial and indust. loans outstanding in 1992 $s (bci)
100	FCLBMC	1959:01–2001:08	level	Wkly rp lg com. banks: net change com and ind. loans (bil$,SAAR)
101	CCINRV	1959:01–2001:08	Δ ln	Consumer credit outstanding nonrevolving g19

Price indexes

102	PMCP	1959:01–2001:08	level	NAPM commodity prices index (%)
103	PWFSA*	1959:01–2001:08	Δ ln	PPI: finished goods (82=100,SA)
104	PWFCSA*	1959:01–2001:08	Δ ln	PPI: finished consumer goods (82=100,SA)
105	PWIMSA*	1959:01–2001:08	Δ ln	PPI: intermed mat. sup and components (82=100,SA)
106	PWCMSA*	1959:01–2001:08	Δ ln	PPI: crude materials (82=100,SA)
107	PSM99Q*	1959:01–2001:08	Δ ln	Index of sensitive materials prices (1990=100) (bci-99a)
108	PUNEW*	1959:01–2001:08	Δ ln	CPI-U: all items (82–84=100,SA)
109	PU83*	1959:01–2001:08	Δ ln	CPI-U: apparel and upkeep (82–84=100,SA)
110	PU84*	1959:01–2001:08	Δ ln	CPI-U: transportation (82–84=100,SA)
111	PU85*	1959:01–2001:08	Δ ln	CPI-U: medical care (82–84=100,SA)
112	PUC*	1959:01–2001:08	Δ ln	CPI-U: commodities (82–84=100,SA)
113	PUCD*	1959:01–2001:08	Δ ln	CPI-U: durables (82–84=100,SA)
114	PUS*	1959:01–2001:08	Δ ln	CPI-U: services (82–84=100,SA)
115	PUXF*	1959:01–2001:08	Δ ln	CPI-U: all items less food (82–84=100,SA)
116	PUXHS*	1959:01–2001:08	Δ ln	CPI-U: all items less shelter (82–84=100,SA)
117	PUXM*	1959:01–2001:08	Δ ln	CPI-U: all items less medical care (82– 84=100,SA)

Average hourly earnings

118	LEHCC*	1959:01–2001:08	Δ ln	Avg hr earnings of constr wkrs: construction ($,SA)
119	LEHM*	1959:01–2001:08	Δ ln	Avg hr earnings of prod wkrs: manufacturing ($,SA)

Miscellaneous

120	HHSNTN	1959:01–2001:08	level	U. of Mich. index of consumer

Summary

English summary

The policy makers at the Federal Reserve Bank and the bond market participants have one thing in common: They have an abundant amount of information at their disposal, and as such the information set on which they condition the interest rate setting, and bond pricing respectively, is large. Consequently, a recurrent theme of this thesis is the approximation of the large information sets by a large panel of macroeconomic and financial time series. In particular, this thesis advances the use of dynamic factors, to approximate the conditioning information set in both monetary policy analysis and in bond pricing. By construction, only a few of these factors are able to summarize the bulk of the information of potentially hundreds of observed time series.

Chapter 1:

Central banks monitor literally hundreds of economic variables in the process of policy formulation as expressed by Federal Reserve Board chairman Ben Bernanke and his co-authors in Bernanke et al. (2005). Classical multivariate regression models generally perform poorly in fitting such large cross-sections of time series (aka. large panels). However, in recent years econometric estimation techniques have been developed which allow these large panels to be analyzed through a small set of underlying extracted dynamic factors. These Dynamic Factor Models (DFM) and the related Factor-Augmented Vector Autoregressive models (FAVAR) are typically estimated by principal component methods or by Bayesian methods as in Bernanke et al.

(2005). The methodological contribution of this chapter is a one-step fully parametric estimation of the FAVAR by means of the EM algorithm as an alternative to the two-step principal component method and the one-step Bayesian methods. In the empirical section of this chapter I analyze a cross-section of 120 US macroeconomic and financial time series and find that the comovement of these time series over time is shown to be adequately described in terms of eight dynamic latent driving forces (dynamic factors) and the US federal funds rate. Subsequently, I study the dynamic response of a set of key economic variables following a shock to the federal funds rate. Finally, I demonstrate empirically that the same dynamic responses but better statistical fit emerge robustly from a low order FAVAR with eight correlated factors compared to a high order FAVAR with fewer correlated factors, for instance four factors. In other words, I find empirical evidence that the information in complicated factor dynamics (high order FAVAR) as in Bernanke et al. (2005) may be substituted by panel information and a low order FAVAR.

Chapter 2:

The second chapter is written jointly with Hans Dewachter and Romain Houssa.

The starting point is the Factor-Augmented Vector Autoregressive model (FAVAR) entertained in chapter one. We focus on the economic interpretation of the latent (unobserved) factors that typically emerge from both DFMs and FAVARs. A standard procedure in the literature amounts to inferring the economic meaning of the factors from the dominant factor loadings, i.e. from the observed variables in the panel that are mostly related to the particular factor. However, this approach does not necessarily generate unambiguous and well-defined interpretations of the factors. In this paper we address the ambiguous economic interpretation of the exactly identified dynamic factors by using a procedure that imposes a specific and well-defined interpretation of the factors. The economic interpretation of the extracted factors is based on a set of overidentifying restrictions on the factor loadings. This model is still a Factor-Augmented Vector Autoregression, but it is now subject to linear loading restrictions. We apply this framework to the same panel of US macroeconomic series as in the first chapter of this thesis. In particular, we identify nine macroeconomic factors and discuss the economic impact of monetary policy shocks. We find that the results are theoretically plausible and in line with other findings in the literature.

Chapter 3:

The bond market is monitoring an abundant amount of information in its assessment of the state of the economy and its implications for bond pricing and bond risk premia. In this chapter, I propose to imitate the potential large information set and solve the bond markets filtering problem by a dynamic factor analysis of the large panel of macroeconomic and financial time series used in the former chapters. The identification approach proposed in chapter two allow me to estimate a few dynamic factors with a clear macroeconomic interpretation. Subsequently, these dynamic factors represent the macroeconomic state variables in a discrete-time dynamic term structure model that allows me to calculate model-implied bond prices, bond yields and even bond risk premia (excess returns). The focus is on potential macroeconomic sources of variation in expected excess returns on bonds. The dynamic responses of the model-implied expected excess return reveal that an inflation factor and an unemployment factor are the most important among five candidate macroeconomic factors. A one standard deviation shock to unemployment initially raises the expected excess return by 17 basis points on an annually basis for a five-year bond held for one year. The intuition is clear: risk premia are time-varying and counter-cyclical. Hence, in business cycle troughs we see rising unemployment and investors are demanding a higher risk premium to buy risky assets.

Dansk resumé

Den amerikanske nationalbank (centralbank) og det amerikanske obligationsmarked har én ting til fælles: De overvåger og har adgang til en meget stor mængde information. Baseret på denne rigelige information fastsætter begge parter henholdsvis den officielle rentesats og forskellige obligationspriser. Disse kendsgerninger har motiveret det gennemgående tema i afhandlingen, hvor denne rigelige mængde af information approksimeres med et stort panel af makroøkonomiske og finansielle data. Specifikt videreudvikles nogle af de økonometriske teknikker, som i de senere år er udviklet til at kunne håndtere store panel datasæt, således at rigelig information kan indarbejdes i empirisk pengepolitisk analyse samt i empirisk obligationsprisfastsættelse.

Kapitel 1:

Centralbanker overvåger bogstaveligt talt hundredvis af økonomiske tidsserier med henblik på at basere den pengepolitiske rentefastsættelse på så megen kvalificeret information som muligt. Sådan har den nuværende amerikanske centralbankchef Ben Bernanke og hans medforfattere udtalt i en indflydelsesrig videnskabelig artikel; jfr. Bernanke et al. (2005). Den klassiske multivariate regressionsanalyse er ikke specielt velegnet til at beskrive store panel datasæt (store i tidsseriedimensionen og i krydssektionen). I de senere år, er der dog sket store fremskridt i udviklingen af økonometriske teknikker, den såkaldte dynamiske faktoranalyse eller den beslægtede faktor-udvidede vektor autoregressive model (FAVAR). Med den dynamiske faktoranalyse er det muligt at beskrive disse store panel datasæt med nogle få underliggende latente dynamiske faktorer. Hidtil er disse modeller typisk estimeret ved principal komponentmetoden eller ved hjælp af Bayesianske teknikker, jfr. Bernanke et al. (2005).

Dette kapitel bidrager til litteraturen ved at beskrive og anvende en alternativ fuldt parametrisk, iterativ maksimum likelihood metode, til at estimere den faktorudvidede vektor autoregressive model ved hjælp af EM algoritmen. I den empiriske analyse i kapitlet, analyserer jeg et stort panel datasæt bestående af 120 amerikanske tidsserier af makroøkonomisk og finansiel karakter. Jeg finder, at samvariationen over tid for disse mange tidsserier kan beskrives tilfredsstillende, ved hjælp af otte dynamiske latente faktorer (drivkræfter). Derefter analyserer jeg den dynamiske respons af en række økonomiske nøglevariable, som følge af et pengepolitisk stød dvs. som følge af en pengepolitisk overraskende renteforhøjelse. Slutteligt demonstrerer jeg, at den samme dynamiske respons, men et markant bedre statistisk fit, kan opnås ved at bruge flere dynamiske faktorer med sparsom dynamisk kompleksitet sammenlignet med færre faktorer med væsentlig mere kompliceret dynamisk kompleksitet. FAVAR modellen i Bernanke et al. (2005) er netop karakteriseret ved en væsentlige mere kompliceret faktordynamik.

Kapitel 2: .

Dette kapitel er skrevet i samarbejde med Hans Dewachter og Romain Houssa.

Udgangspunktet er den faktor-udvidede vektor autoregressive model (FAVAR) fra kapitel 1. I dette kapitel fokuserer vi på den økonomiske fortolkning af de latente (uobserverede) dynamiske faktorer, som følger af den dynamiske faktoranalyse eller FAVAR modellerne. Typisk for litteraturen udleder man den økonomiske fortolkning af de uobserverede faktorer ved at betragte de mest betydningsfulde faktorvægte (factor loadings på engelsk). Denne tilgang udelukker dog mindre dominerende faktorvægte og under alle omstændigheder, er det ikke muligt at opnå utvetydige og veldefinerede fortolkninger af faktorerne ved denne tilgang.

I dette kapitel adresserer vi dette problem, ved at pålægge overidentificerende restriktioner på faktorvægtene, således en utvetydig og veldefineret fortolkning af faktorerne bliver mulig. I den empiriske analyse betragter vi igen det amerikanske datasæt anvendt i kapitel 1. Vi estimerer ni økonomisk fortolkbare faktorer fra et datarigeligt miljø og anvender disse i en pengepolitisk analyse. Resultaterne er teoretisk plausible og i overensstemmelse med andre resultater i litteraturen.

Kapitel 3:

Obligationsmarkedet overvåger og reagerer på en stor mængde tilgængelig information i dets bestræbelser på at vurdere (filtrere) den økonomiske tilstand i økonomien, hvilket har indflydelse på obligationskurserne og risikopræmierne. I dette kapitel foreslår jeg, at imitere den store informationsmængde og løse obligationsmarkedets filtreringsproblem ved hjælp af dynamisk faktoranalyse af et stort amerikansk datasæt; jfr. ovenfor. Faktoranalysen bidrager med nogle makroøkonomiske tilstandsvariable, udledt fra et datarigeligt miljø, og disse tilstandsvariable (faktorer) repræsenterer den underliggende udvikling i økonomiske nøglevariable, eksempelvis den underliggende inflation. Disse tilstandsvariable anvendes efterfølgende i en dynamisk rentestrukturmodel i diskret tid, som muliggør beregning af teoretiske nulkuponobligationskurser, nulkuponrenter, og endog obligationsrisikopræmier.

I den empiriske analyse fokuserer jeg på potentielle makroøkonomiske årsager til variation i obligationsrisikopræmier. Jeg finder, at den dynamiske respons af obligationsrisikopræmier som følger af stød til eksempelvis den underliggende inflation eller arbejdsløsheden, udgør de væsentligste kilder bag variationen. Et positivt stød på 1 standardafvigelse til arbejdsløsheden indebærer en modelimpliceret stigning i obligationsrisikopræmien på 17 basispunkter p.a. for en femårig nulkuponobligation med en investeringshorisont på 1 år. Intuitionen er, at risikopræmier generelt er tidsvarierende og procykliske i konjunkturmæssig forstand. Derfor ser vi, at investorerne i konjunkturmæssige lavpunkter, repræsenteret ved eksempelvis høj arbejdsløshed, kræver et store risikopræmier for at købe risikobetonede finansielle aktiver.

Bibliography

Aguilar, O. & West, M. (2000), 'Bayesian dynamic factor models and portfolio allocation', *Journal of Business and Economic Statistics* **3**(18), 338–357.

Ahmadi, P. A. & Uhlig, H. (2008), Identifying monetary policy shocks in a data-rich environment: A sign restriction approach in a bayesian factor-augmented var. Conference presentation, European Economic Association & Econometric Society, August 2008.

Alessi, L., Barigozzi, M. & Capasso, M. (2007), 'Dynamic factor GARCH: Multivariate volatility forecast for a large number of series', *Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies* .

Altug, S. (1989), 'Time to build and aggregate fluctuations: some new evidence', *International Economic Review* **30**(4), 889–920.

Ang, A., Dong, S. & Piazzesi, M. (2005), No-arbitrage taylor rules. NBER working paper 13448.

Ang, A. & Piazzesi, M. (2003), 'A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables', *Journal of Monetary Economics* **50**, 745–787.

Backus, D., Foresi, S. & Telmer, C. (1996), Affine models of currency pricing. NBER WP 5623.

Backus, D., Foresi, S. & Telmer, C. (1998), Affine models of currency pricing: Accounting for the forward premium anomaly. New York Univerisity Working Paper (revision of NBER WP 5623).

Bai, J. & Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.

Bai, J. & Ng, S. (2007), 'Determining the number of primitive shocks in factor models', *Journal of Business & Economic Statistics* **25**, 52–60.

Banbura, M., Giannone, D. & Reichlin, L. (2008), 'Large bayesian VARs', *Journal of Applied Econometrics (forthcoming)* .

Bekaert, G., Cho, S. & Moreno, A. (2005), New-keynesian macroeconomics and the term structure. Working paper, Graduate School of Business, Columbia University.

Bernanke, B. S. & Boivin, J. (2003), 'Monetary policy in a data-rich environment', *Journal of Monetary Economics* **50**(3), 525–546.

Bernanke, B. S., Boivin, J. & Eliasz, P. (2005), 'Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach', *The Quarterly Journal of Economics* pp. 387–422.

Bernanke, B. S., Laubach, T., Mishkin, F. S. & Posen, A. S. (1999), *Inflation Targeting: Lessons from the International Experience*, Princeton University Press.

Bjørnland, H. & Leitemo, K. (2009), 'Identifying the interdependence between US monetary policy and the stock market', *Journal of Monetary Economics (forthcoming)* .

Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**, 637–659.

Blanchard, O. J. & Quah, D. (1989), 'The dynamic effects of aggregate demand and supply disturbances', *American Economic Review* **79**(4), 655–73.

Boivin, J. & Giannoni, M. (2006), Dsge models in a data-rich environment, Working Paper 12772, National Bureau of Economic Research.

Boivin, J., Giannoni, M. P. & Mihov, I. (2009), 'Sticky prices and monetary policy: Evidence from disaggregated us data', *American Economic Review* **99**(1), 350–84.

Bork, L. (2008), Estimating US monetary policy shocks using a factor-augmented vector autoregression: An EM algorithm approach. CREATES Research Paper 2009-11.

Bork, L., Dewachter, H. & Houssa, R. (2008), Identification of macroeconomic factors in large panels. Manuscript, Catholic University of Leuven and Aarhus School of Business, University of Aarhus.

Breeden, D. T. (1979), 'An intertemporal asset pricing model with stochastic consumption and investment opportunities', *Journal of Financial Economics* **7**, 265–96.

Breitung, J. & Eickmeier, S. (2006), 'Dynamic factor models', *Allgemeines Statistisches Archive* **90**(1), 27–42.

Campbell, J. Y. & Cochrane, J. H. (1999), 'By force of habit: A consumption based explanation of aggregate stock market behavior', *Journal of Political Economy* **107**(2), 205–251.

Campbell, J. Y., Lo, A. & MacKinlay, C. (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton.

Canova, F. (2007), *Methods for Applied Macroeconomic Research*, Princeton University Press.

Chamberlain, G. (1983), 'Funds, factors, and diversification in arbitrage pricing models', *Econometrica* **51**(5), 1305–1323.

Chamberlain, G. & Rothschild, M. (1983), 'Arbitrage, factor structure, and mean-variance analysis on large asset markets', *Econometrica* **51**(5), 1281–1304.

Chen, R. & Scott, L. (1993), 'Maximum likelihood estimation for a multifactor general equilibrium model of the term structure of interest rates', *Journal of Fixed Income* **3**, 14—31.

Christensen, J. H. E., Diebold, F. X. & Rudebusch, G. D. (2009), 'An arbitrage-free generalized Nelson-Siegel term structure model', *The Econometrics Journal* **146**, 351–363.

Christiano, L. J., Eichenbaum, M. & Evans, C. L. (1999), Monetary policy shocks: What have we learned and to what end?, *in* J. B. Taylor & M. Woodford, eds, 'Handbook of Macroeconomics', Vol. 1A, Elsevier Science, North Holland, New York, pp. 65–148.

Cochrane, J. H. & Piazzesi, M. (2005), 'Bond risk premia', *American Economic Review* **95**, 138–160.

Cochrane, J. H. & Piazzesi, M. (2009), Decomposing the yield curve. Working paper, University of Chicago Booth School of Business.

Coroneo, L., Nyholm, K. & Vidova-Koleva, R. (2008), How arbitrage-free is the Nelson-Siegel model?, Technical report, ECB Working Paper no. 874.

Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**, 385–407.

Dai, Q. & Singleton, K. J. (2000), 'Specification analysis of affine term structure models', *The Journal of Finance* **55**(5), 1943–1978.

Dai, Q. & Singleton, K. J. (2002), 'Expectation puzzles, time-varying risk premia, and affine models of the term structure', *Journal of Financial Economics* **63**(3), 415–441.

Dai, Q., Singleton, K. J. & Yang, W. (2007), 'Regime shifts in a dynamic term structure model of U.S. treasury bond yields', *Review of Financial Studies* **20**, 1669–1706.

de Jong, F. (2000), 'Time series and cross-section information in affine term-structure models', *Journal of Business & Economic Statistics* **18**(3), 300–314.

de Jong, P. (1989), 'Smoothing and interpolation with the state-space model', *Journal of the American Statistical Association* **84**, 1085–1088.

de Jong, P. & Mackinnon, M. J. (1988), 'Covariances for smoothed estimates in state space models', *Biometrika* (75), 601–602.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society* **39**(1), 1–38.

Dewachter, H. (2008), Imperfect information, macroeconomic dynamics and the yield curve: An encompassing macro-finance model, Research series 200810-19, National Bank of Belgium.

Dewachter, H. & Lyrio, M. (2006*a*), Learning, macroeconomic dynamics and the term structure of interest rates, *in* J. Y. Campbell, ed., 'Asset Prices and Monetary Policy', National Bureau of Economic Research, Forthcoming from The University of Chicago Press.

Dewachter, H. & Lyrio, M. (2006*b*), 'Macro factors and the term structure of interest rates', *Journal of Money, Credit, and Banking* **38**(1), 119–140.

Dewachter, H., Lyrio, M. & Maes, K. (2006), 'A joint model for the term structure of interest rates and the macroeconomy', *Journal of Applied Econometrics* **21**(4), 439–462.

Diebold, F. X., Li, C. & Yue, V. Z. (2008), 'Global yield curve dynamics and interactions: A dynamic Nelson-Siegel approach', *Journal of Econometrics* **146**(2), 351–363.

Diebold, F. X., Piazzesi, M. & Rudebusch, G. D. (2005), 'Modeling bond yields in finance and macroeconomics', *American Economic Review, Papers and Proceedings* **95**, 415–420.

Diebold, F. X., Rudebusch, G. D. & Aruoba, S. B. (2006), 'The macroeconomy and the yield curve: a dynamic latent factor approach', *Journal of Econometrics* **131**(March-April), 309–338.

Doz, C., Giannone, D. & Reichlin, L. (2006), A quasi maximum likelihood approach for large approximate dynamic factor models. European Central Bank Working Paper no. 674.

Doz, C., Giannone, D. & Reichlin, L. (2007), A two-step estimator for large approximate dynamic factor models based on Kalman filtering, CEPR Discussion Paper 6043, Centre for Economic Policy Research.

Duffee, G. (2002), 'Term premia and interest rate forecasts in affine models', *Journal of Finance* **57**, 405–443.

Duffee, G. R. (2007), Are variations in term premia related to the macroeconomy. Working paper, Johns Hopkins University.

Duffee, G. R. (2008), Information in (and not in) the term structure. Working paper, Johns Hopkins University.

Duffie, D. & Kan, R. (1996), 'A yield-factor model of interest rates', *Mathematical Finance* **6**, 379–406.

Duffie, D. & Singleton, K. J. (1997), 'An econometric model of the term structure of interest-rate swap yields', *Journal of Finance* **52**(4), 1287–1321.

Engle, R. & Watson, M. (1981), 'A one-factor multivariate time series model of metropolitan wage rates', *Journal of the American Statistical Association* **76**, 774–781.

Favero, C. A., Marcellino, M. & Neglia, F. (2005), 'Principal components at work: the empirical analysis of monetary policy with large data sets', *Journal of Applied Econometrics* **20**(5), 603–620.

Forni, M., Giannone, D., Lippi, M. & Reichlin, L. (2007), 'Opening the black box: Structural factor models with large cross-sections', *Econometric Theory (forthcoming)* .

Forni, M., Giannone, D., Lippi, M. & Reichlin, L. (2008), 'Opening the black box: Structural factor models with large cross-sections', *Econometric Theory (forthcoming)* .

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2000), 'The generalized dynamic-factor model: Identification and estimation', *The Review of Economics and Statistics* **82**(4), 540–554.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2001), 'Coincident and leading indicators for the euro area', *Economic Journal* **111**(471), 62–85.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2004), 'The generalised dynamic factor model: consistency and rates', *Journal of Econometrics* **119**(2), 231–255.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2005), 'The generalized dynamic factor model: one sided estimation and forecasting', *Journal of the American Statistical Association* **100**(471), 830–840.

Forni, M. & Lippi, M. (2001), 'The generalized dynamic factor model: Representation theory', *Econometric Theory* **17**(06), 1113–1141.

Forni, M. & Reichlin, L. (1998), 'Let's get real: A factor analytical approach to disaggregated business cycle dynamics', *Review of Economic Studies* **65**(3), 453–73.

Forni, M. & Reichlin, L. (2001), 'Federal policies and local economies: Europe and the us', *European Economic Review* **45**(1), 109–134.

Geweke, J. (1977), The dynamic factor analysis of economic time series, *in* D. J. Aigner & A. S. Golderger, eds, 'Latent Variables in Socioeconomic Models', North Holland.

Geweke, J. F. & Singleton, K. J. (1981), 'Maximum likelihood "confirmatory" factor analysis of economic time series', *International Economic Review* **22**(1), 37–54.

Geweke, J. & Zhou, G. (1996), 'Measuring the pricing error of the arbitrage pricing theory', *Review of Financial Studies* **9**(2), 557–587.

Giannone, D., Reichlin, L. & Sala, L. (2002), Tracking Greenspan: Systematic and unsystematic monetary policy revisited. CEPR Discussion Papers no. 3550.

Giannone, D., Reichlin, L. & Sala, L. (2004), Monetary policy in real time, *in* M. Gertler & K. Rogoff, eds, 'NBER Macroeconomic Annual 2004', Vol. 19, MIT Press.

Giannone, D., Reichlin, L. & Small, D. (2008), 'Nowcasting: The real time informational content of macroeconomic data releases', *Journal of Monetary Economics* **55**, 665–676.

Giordani, P. (2004), 'An alternative explanation of the price puzzle', *Journal of Monetary Economics* **51**(6), 1271–1296.

Goffe, W. L., Ferrier, G. D. & Rogers, J. (1994), 'Global optimization of statistical functions with simulated annealing', *Journal of Econometrics* **60**, 65–100.

Hallin, M. & Liska, R. (2007), 'Determining the number of factors in the generalized factor model', *Journal of the American Statistical Association* **102**, 603–617.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton.

Harrison, J. M. & Kreps, D. M. (1979), 'Martingales and arbitrage in multiperiod securities markets', *Journal of Economic Theory* **20**, 381–408.

Harvey, A. C. (1989), *Forecasting structural time series models and the Kalman filter*, Cambridge University Press.

Hördahl, P., Tristani, O. & Vestin, D. (2006), 'A joint econometric model of macroeconomic and term-structure dynamics', *Journal of Econometrics* **131**(1-2), 405–444.

Houssa, R. (2008*a*), 'Monetary union in west africa and asymmetric shocks: A dynamic structural factor model approach', *Journal of Development Economics* **Vol. 85**, 319–347.

Houssa, R. (2008*b*), Sources of Fluctuations: World, Regional, and National Factors, In Macroeconomic Fluctuations in Developing Countries, PhD thesis, KULeuven.

Ingersoll, Jonathan E, J. (1984), 'Some results in the theory of arbitrage pricing', *Journal of Finance* **39**(4), 1021–39.

Joslin, S., Priebsch, M. & Singleton, K. J. (2009), Risk premium accounting in macro-dynamic term structure models. Working Paper, Stanford University.

Jungbacker, B. & Koopman, S. J. (2008), Likelihood-based analysis for dynamic factor models, Tinbergen Institute Discussion Papers 08-007/4, Tinbergen Institute.

Kim, C.-J. & Nelson, C. R. (1999), *State-Space Models with Regime Switching*, The MIT Press.

Koopman, S. J. (1993), 'Disturbance smoother for state space models', *Biometrika* **80**, 117–126.

Koopman, S. J. & Shephard, N. (1992), 'Exact score for time series models in state space form', *Biometrika* **79**(4), 823–826.

Koopman, S. J., Shephard, N. & Doornik, J. A. (1999), 'Statistical algorithms for models in state space using ssfpack 2.2', *Econometrics Journal* **2**(1), 107–160.

Kose, A. M., Otrok, C. & Whiteman, C. (2003), 'International business cycles: World, region, and country-specific factors', *American Economic Review* **93**(4), 1216–1239.

Lange, K. (1995), 'A quasi-newton acceleration of the EM algorithm', *Statistica Sinica* **5**, 1–18.

Laubach, T., Tetlow, R. J. & Williams, J. C. (2006), 'Macroeconomic factors in the term structure of interest rates when agents learn', *Society for Computational Economics* (83).

Lawley, D. N. & Maxwell, A. E. (1971), *Factor Analysis as a Statistical Method (Butterworths Mathematical Texts)*, 2nd. edn, Butterworths, London.

Le, A., Singleton, K. J. & Dai, Q. (2009), 'Discrete-time affine term structure models with generalized market price of risk', *Review of Financial Studies (forthcoming)* .

Leeper, E. M., Sims, C. A. & Zha, T. (1996), 'What does monetary policy do?', *Brookings Papers on Economic Activity* **27**(1996-2), 1–78.

Litterman, R. & Scheinkman, J. (1991), 'Common factors affecting bond returns', *Journal of Fixed Income* **1**, 54–61.

Louis, T. A. (1982), 'Finding the observed information matrix when using the EM algorithm', *Journal of the Royal Statistical Society. Series B* **44**(2), 226–233.

Ludvigson, S. C. & Ng, S. (2007), 'The empirical risk-return relation: A factor analysis approach', *Journal of Financial Economics* **83**(1), 171–222.

Ludvigson, S. C. & Ng, S. (2008), 'Macro factors in bond risk premia', *The Review of Financial Studies (forthcoming)* .

Lund, J. (1997), Econometric analysis of continuous-time arbitrage-free models of the term structure of interest rates. Working Paper, Aarhus School of Business.

Lütkepohl, H. (2007), *New Introduction to Multiple Time Series Analysis*, Springer, Berlin.

Magnus, J. R. & Neudecker, H. (2007), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 2nd edn, John Wiley & Sons.

Merton, R. C. (1973), 'Theory of rational option pricing', *Bell Journal of Economics and Management Science* **4**, 141–183.

Mönch, E. (2008), 'Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach', *Journal of Econometrics* **146**(1), 26–43.

Otrok, C. & Whiteman, C. H. (1998), 'Bayesian leading indicators: Measuring and predicting economic conditions in Iowa', *International Economic Review* **39**(4), 997–1014.

Pericoli, M. & Taboga, M. (2008), 'Canonical term-structure models with observable factors and the dynamics of bond risk premiums', *Journal of Money, Credit and Banking* **40**, 1471–1488.

Piazzesi, M. (2009), Affine term structure models, *in* Y. Ait-Sahalia & L. Hansen, eds, 'Handbook of Financial Econometrics', Elsevier.

Quah, D. & Sargent, T. J. (1993), A dynamic index model for large cross sections, *in* J. Stock & M. Watson, eds, 'Business Cycles, Indicators and Forecasting', University of Chicago Press for NBER, pp. 285–310.

Reichlin, L. (2003), Factor models in large cross sections of time series, *in* M. Dewatripont, P. Hansen & S. Turnowsky, eds, 'Advances in economics and econometrics: theory and applications', Vol. 111 of *8th world congress of the econometric society*, Cambridge University Press.

Reis, R. & Watson, M. W. (2008), Relative goods' prices, pure inflation, and the Philips correlation. Princeton University and Columbia University.

Ross, S. A. (1976), 'The arbitrage theory of capital asset pricing', *Journal of Economic Theory* **13**, 341–360.

Rudebusch, G. D. & Svensson, L. E. O. (1999), Policy rules for inflation targeting, *in* J. B. Taylor, ed., 'Monetary Policy Rules', University of Chicago Press, Chicago.

Sargent, T. J. (1989), 'Two models of measurements and the investment accelerator', *Journal of Political Economy* **97**(2), 251–287.

Sargent, T. J. & Sims, C. A. (1977), Business cycle modeling without pretending to have too much a priori economic theory, *in* C. A. Sims, ed., 'New Methods in Business Cycle Research', Federal Reserve Bank on Minneapolis, pp. 45–110.

Shumway, R. H. & Stoffer, D. S. (1982), 'An approach to time series smoothing and forecasting using the EM algorithm', *Journal of Time Series Analysis* **3**, 253–226.

Sims, C. (1980), 'Macroeconomics and reality', *Econometrica* **48**, 1–48.

Sims, C. A. (1992), 'Interpreting the macroeconomic time series facts : The effects of monetary policy', *European Economic Review* **36**(5), 975–1000.

Singleton, K. J. (2006), *Empirical Dynamic Asset Pricing*, Princeton University Press.

Stock, J. H. & Watson, M. W. (2002*a*), 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association* **97**, 1167–1179.

Stock, J. H. & Watson, M. W. (2002*b*), 'Macroeconomic forecasting using diffusion indexes', *Journal of Business Economics and Statistics* **XX:II**, 147–162.

Stock, J. H. & Watson, M. W. (2005), Implications of dynamic factor models for VAR analysis, Working Paper 11467, National Bureau of Economic Research.

Stock, J. H. & Watson, M. W. (2006), Forecasting with many predictors, *in* G. Elliott, C. W. Granger & A. Timmermann, eds, 'Handbook of Economic Forecasting (forthcoming)', North Holland, chapter 10.

Svensson, L. E. O. (1997), 'Inflation forecast targeting: Implementing and monitoring inflation targets', *European Economic Review* **41**(6), 1111–1146.

Taylor, J. (1993), 'Discretion versus policy rules in practice', *Carnegie-Rochester Conference Series on Public Policy* **39**, 195–214.

Uhlig, H. (2005), 'What are the effects of monetary policy on output? Results from an agnostic identification procedure', *Journal of Monetary Economics* **52**(2), 381–419.

Walsh, C. E. (2003), *Monetary Theory and Policy*, 2nd. edn, MIT Press.

Watson, M. W. & Engle, R. F. (1983), 'Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models', *Journal of Econometrics* **23**(3), 385–400.

Woodford, M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press.

Wu, L. S.-Y., Pai, J. S. & Hosking, J. (1996), 'An algorithm for estimating parameters of state-space models', *Statistics and Probability Letters* **28**, 99–106.

Wu, T. (2006), 'Macro factors and the affine term structure of interest rates', *Journal of Money, Credit, and Banking* **38**(7), 1847–1875.

Yu, W.-C. (2008), 'Macroeconomic and financial market volatilities: an empirical evidence of factor model', *Economics Bulletin* **3**(33), 1–18.