



SCHOOL OF ECONOMICS AND MANAGEMENT  
FACULTY OF SOCIAL SCIENCES  
AARHUS UNIVERSITY



**CREATES**  
Center for Research in Econometric  
Analysis of Time Series

**CREATES Research Paper 2009-11**

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

**Lasse Bork**

School of Economics and Management  
Aarhus University  
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C  
Denmark

# Estimating US Monetary Policy Shocks Using a Factor-Augmented Vector Autoregression: An EM Algorithm Approach\*

Lasse Bork<sup>†</sup>

Current Version: February, 2009

First Draft: July 30, 2008

## 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 for helpful comments. Any remaining errors are my own.

<sup>†</sup>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@asb.dk.

# 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 macro-economic 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 addressed 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<sup>1</sup>. Finally, a thorough investigation

---

<sup>1</sup>The issue of interpretation of the estimated factors is addressed in Bork et al. (2008) in

of the 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<sup>2</sup>. 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

---

which the EM algorithm is also applied.

<sup>2</sup>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).

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 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<sup>3</sup> 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<sup>4</sup>, dynamic factor analysis of large panels may help to filter out the observed

---

<sup>3</sup>Related to surveys by National Association of Purchasing Management.

<sup>4</sup>Sargent (1989) shows how the existence of measurement error leads to a dynamic factor

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 theory and empirical analysis of large dimensional approximate<sup>5</sup> 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 (2002*a,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<sup>6</sup>. 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

---

index model.

<sup>5</sup>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.

<sup>6</sup>Stock & Watson (2002*a*) 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.

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<sup>7</sup> starting 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

---

<sup>7</sup>A detailed account of empirical applications can be found in Reichlin (2003) and Breitung & Eickmeier (2006).

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 kind 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 2 while identification issues and the estimation method are presented in section 3. Section 4 details the empirical results and section 5 concludes. The appendices contain details on the Kalman filter and smoother as well as the EM algorithm.

## 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  $\chi$  common to all variables and an idiosyncratic component  $\xi$  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  $i$ th variable in the panel  $X$ <sup>8</sup> at

---

<sup>8</sup>All variables in the panel are transformed into stationary variables with mean zero and unit variance. See section 4.1.

time  $t$  can be written as:

$$x_{it} = \chi_{it} + \xi_{it} \quad (1)$$

for  $i = 1, \dots, N$  and  $t = 1, \dots, T$  with  $E[\chi_{it}\xi_{js}] = 0 \forall i, j, t, s$  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.

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  $\chi_t = \lambda(L) f_t$ , where  $\lambda(L)$  is a  $N \times q$  matrix polynomial in the lag-operator  $L$  of finite order  $s$ <sup>9</sup>. 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,  $\varepsilon_t$ <sup>10</sup>. Inserting the specification of the common component in (1) results in a dynamic factor model driven by  $q$  dynamic factors:

$$x_{it} = \lambda_i^\top(L) f_t + \xi_{it} \quad (2)$$

where  $\lambda_i(L) = \lambda_{i,0} + \lambda_{i,1}L + \dots + \lambda_{i,s}L^s$ . Stacking contemporaneous and  $s$  lagged values of  $f_t$  in the  $q(s+1)$  dimensional vector  $F_t$  and the matching values of  $\lambda_i$  in  $q(s+1)$  dimensional vector  $\Lambda_i$  results in the static representation of the dynamic

---

<sup>9</sup>Infinite order of the lag-polynomials is considered in the *generalized* dynamic factor model of Forni et al. (2000).

<sup>10</sup>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.

factor model in (2), which is driven by  $r = q(s + 1)$  factors,  $F_t$  :

$$\begin{aligned} x_{it} &= \Lambda_i^\top F_t + \xi_{it} \\ &= \begin{bmatrix} \lambda_{i,0} \\ \lambda_{i,1} \\ \vdots \\ \lambda_{i,s} \end{bmatrix}^\top \begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-s} \end{bmatrix} + \xi_{it} \end{aligned} \quad (3)$$

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  $\varepsilon_t$  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  $\text{AR}(h)$  process, Bai & Ng (2007)<sup>11</sup> show that  $F_t$  can be represented as a  $\text{VAR}(p)$  process with  $p = \max(1, h - s)$ . Intuitively, if the dynamic process of  $f_t$  is particularly simple then a  $\text{VAR}(1)$  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 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:

$$\begin{aligned} X_t &= \Lambda F_t + \xi_t \\ F_t &= \Phi(L) F_{t-1} + \Upsilon \varepsilon_t \end{aligned} \quad (4)$$

where  $X_t = (x_{1,t}, \dots, x_{N,t})^\top$ ,  $\xi_t = (\xi_{1,t}, \dots, \xi_{N,t})^\top$  is i.i.d  $N(0, R)$ <sup>12</sup> and  $\Lambda = [\Lambda_1^\top, \dots, \Lambda_N^\top]^\top$  is a  $N \times r$  loading matrix. The state transition equation is stationary so that the  $p$ th order matrix polynomial  $\Phi(L)$  has roots outside the unit circle,  $\Upsilon$  is a  $r \times q$  matrix and  $\varepsilon_t$  is i.i.d  $N(0, Q)$ . The unknowns in this Gaussian

<sup>11</sup>They also discuss  $\text{MA}(h)$  and  $\text{ARMA}$  processes.

<sup>12</sup>Note that the assumption of i.i.d idiosyncratic components in (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 (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.

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  $\Lambda$  for FFR is  $[0, \dots, 0, 1]$ . 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 (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.

### 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 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 FAVAR model is highly over-parameterized as it stands in (4) and we are not able to estimate a unique set of parameters,  $\hat{\Theta}$  with the data unless identifying restrictions are imposed on  $\Theta$ . This is a well-known problem of classical factor analysis including the principal component approach to dynamic factor analysis by Stock & Watson. Typically, these models are identified by restricting the covariance matrix of the factors to be an identity matrix,  $F^\top F/T = I$  but sometimes also  $\Lambda^\top \Lambda/N = I$  is used. Neither of these identification schemes is 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 (4). Briefly, for any orthogonal matrix  $P$  where  $PP^\top = I$ , it is possible to construct an observationally equivalent model with  $\tilde{\Lambda} = \Lambda P^{-1}$  and  $\tilde{F} = PF$  that still satisfies the condition  $\tilde{F}^\top \tilde{F} = I$ , and therefore more restrictions are needed.

Starting from a rank condition for the loading matrix, Geweke & Zhou (1996) propose an identification scheme that uniquely identifies the loadings and the factors by imposing additional  $r(r-1)/2$  restrictions on the loading matrix  $\Lambda$  while maintaining uncorrelated factors. Essentially, the upper  $r \times r$  block of  $\Lambda$  is lower triangular with  $r$  positive diagonal elements. Aguilar & West (2000) restrict this diagonal to unity and then allow for a diagonal covariance matrix for the factors. These approaches are often termed "hierarchical" because the first factor is only allowed to load on the first variable in the panel, the second factor on the first two variables etc. Therefore, the ordering may potentially influence the statistical fit and will be discussed in detail in the empirical section; Aguilar & West (2000) present a similar discussion.

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. Interestingly, Proposition 2 in Geweke & Singleton (1981) allows for an alternative identification scheme where the factors are *correlated*, which is a feature preferred in this paper. 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 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:

1. The FFR in  $X_t$  with row index  $\ell_r$  in  $\Lambda$  loads only on the last dynamic factor

in  $F_t$  which is a monetary policy factor (the FFR itself). Hence, for the  $r$  elements in row  $\ell_r$  in  $\Lambda$ , the restricted loading is:

$$\Lambda_{\ell_r}^* = [0, \dots, 0, 1].$$

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  $\{\ell_1, \dots, \ell_{r-1}\}$  of  $X_t$ , which means that the restricted rows of  $\Lambda$  can be written as:

$$\begin{aligned} \Lambda_{\ell_1}^* &= \begin{bmatrix} \mathbf{1}_{1 \times 1} & \mathbf{0}_{1 \times (r-2)} & 0 \end{bmatrix} \\ &\vdots \\ \Lambda_{\ell_j}^* &= \begin{bmatrix} \mathbf{0}_{1 \times (j-1)} & \mathbf{1}_{1 \times 1} & \mathbf{0}_{1 \times (r-j)} \end{bmatrix} \\ &\vdots \\ \Lambda_{\ell_{r-1}}^* &= \begin{bmatrix} \mathbf{0}_{1 \times (r-2)} & \mathbf{1}_{1 \times 1} & \mathbf{0}_{1 \times (r-j)} \end{bmatrix} \end{aligned} \tag{5}$$

where it should be noted that the last columns of  $\Lambda_{\ell_1}^*, \dots, \Lambda_{\ell_{r-1}}^*$  always contain a zero corresponding to the monetary policy factor.

This identification scheme allows for correlated factors and the zero restrictions on  $\Lambda$  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  $\ell_j \in \{\ell_1, \dots, \ell_{r-1}\}$  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<sup>13</sup>; see section 4.1 for more details.

---

<sup>13</sup>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.

### 3.1 The EM algorithm

The linear Gaussian state space model in (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.

The complete data likelihood of the Gaussian state space model in equation (4) is given in equation (18) in Appendix C.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 C.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  $j$

$$\text{vec}(\Lambda^{(j)}) = \text{vec}(DC^{-1}) \quad (6)$$

$$R^{(j)} = \frac{1}{T} (E - DC^{-1}D^\top) \quad (7)$$

$$\text{vec}(\Phi^{(j)}) = \text{vec}(BA^{-1}) \quad (8)$$

$$Q^{(j)} = \frac{1}{T} [C - BA^{-1}B^\top] \quad (9)$$

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

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

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

$cov(F_t, F_{t-1} | \mathcal{X}_T)$  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 companion matrix as in Hamilton (1994) chapter 10. The autocovariances in the  $B$  matrix needed in the  $\Phi$  estimate should then follow automatically. However, this paper follows a slightly different route similar to Koopman et al. (1999) but with an implementation in MATLAB<sup>14</sup>, where the smoothed autocovariance matrix of the state variables is constructed directly and explicitly, 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:

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

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 C.2 page 43 are matrices defined recursively in the Kalman smoother and Kalman filter, respectively. Furthermore, the state smoothing recursions are also stated in the appendix.

### 3.1.1 Parameter restrictions in the EM algorithm

In order to implement the identifying restrictions in (5), the estimators in (6)–(9) subject to linear restrictions need to be derived. Shumway & Stoffer (1982) and Wu et al. (1996) present the restricted  $\Phi^{*15}$  and Bork et al. (2008) show how the

---

<sup>14</sup>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.

<sup>15</sup>shown in the appendix.



restricted  $\Lambda^*$  estimator subject to a linear restriction in the form  $H_\Lambda \text{vec } \Lambda = \kappa_\Lambda$  can be derived:

$$\begin{aligned} \text{vec } (\Lambda^*) &= \text{vec } (DC^{-1}) \\ &\quad + (C^{-1} \otimes R) H_\Lambda^\top [H_\Lambda (C^{-1} \otimes R) H_\Lambda^\top]^{-1} \{ \kappa_\Lambda - H_\Lambda \text{vec } (DC^{-1}) \} \end{aligned} \tag{10}$$

where  $\kappa_\Lambda$  is a  $\eta \times 1$  vector and the restriction matrix  $H_\Lambda$  is of dimension  $\eta \times Nr$ . Notice that the unrestricted estimator in (6) appears if  $\eta = 0$  restrictions are imposed.

### 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 D.

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 successful, it is indeed relatively fast and therefore continued research into this hybrid is worthwhile.

## 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 (in-

creased  $R^2$ ). 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 2: that complicated factor dynamics (many lags) may be substituted by cross-sectional information (more factors).

Throughout this section natural benchmarks are the principal component FAVAR and the Bayesian FAVAR by Bernanke et al. (2005) as I use the same dataset as well as the same model as these authors. The differences in the empirical results may then be attributed to the differences in the estimation methods, i.e. the EM algorithm versus the methods of the Bernanke et al. (2005)<sup>16</sup> 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 part of this paper that involves the empirical monetary policy analysis focuses on the identification of monetary policy shocks and the economy-wide responses to these shocks while remaining agnostic about the interpretation of other structural shocks. 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

---

<sup>16</sup>Although seemingly unreported by the authors, it seems that they employ uncorrelated factors in contrast to the correlated factors employed in this paper.

$q$  dynamic factors driven by  $q \leq r$  structural shocks<sup>17</sup>. Accordingly, I do not specify the matrix  $\Upsilon$  in (4) to select  $q$  shocks; rather  $r = q$  and  $\Upsilon$  is an identity matrix.

The preferred model with eight factors and three lags is the outcome of a careful model selection process where a large number<sup>18</sup> 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 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<sup>19</sup>. 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 con-

---

<sup>17</sup>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 contemporaneous 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  $\chi$  whereas  $q$  is the rank of the spectral density matrix of  $\chi$ . For further discussion of structural factor models, refer to Forni et al. (2007) and Stock & Watson (2005).

<sup>18</sup>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  $r = 3, \dots, 10$  and  $p = 1, \dots, 13$ . To make this exercise computationally feasible, a maximum of 10,000 iterations in the EM algorithm were allowed, which explains the few missing factor models.

<sup>19</sup>Notice that both models involve approximately the same number of autoregressive coefficients in the VAR.

clusions 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.

## 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)<sup>20</sup> 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 for a description of the dataset and in particular the classification into slow-moving variables and fast-moving variables. The time series in the panel are transformed into stationarity by taking logs and/or differencing<sup>21</sup>. 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 (4) page 10. 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  $\tilde{X}_t = D(L) X_t$ <sup>22</sup>.

---

<sup>20</sup>I thank Jean Boivin for kindly making the data set available on his website, HEC Montréal, Canada

<sup>21</sup>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 (2002b) and based on judgemental and preliminary data analysis of each series, including unit root tests.

<sup>22</sup>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  $\tilde{X}_t$ .

## 4.2 The imposition of the identifying restrictions

A number of identifying restrictions need to be imposed on  $r$  rows of the loading matrix  $\Lambda$  as explained in equation (5) page 5 and it should be emphasized that it is not completely unimportant which variables are selected to have restricted loading. Consider the case where the use of principal components as consistent estimators reveals that the first and most important factor can be interpreted as an industrial production factor. Then it makes sense to impose the restriction  $\Lambda_{\ell_1} = (1, 0, \dots, 0)$  on an industrial production variable and not on for instance a price level variable, which would lead to a bad fit for the particular price level variable and not really change the characteristics of the first factor. The reason is that the single "1" in the first column corresponding to the particular price level variable does not really affect the loadings on the industrial production variables<sup>23</sup>. This is important because I argue that the outcome of the estimated factors is not determined by the single unit restriction in a particular column in  $\Lambda$  but rather by how important this factor is for the fraction of variance explained.

Consequently, a pre-study is undertaken to reveal which variables each factor is primarily associated with. Specifically, this pre-study reveals that the first factor is robustly associated with industrial production. Interestingly, the second factor is often related to Moody's BAA yield spread (variable number 91 (#91) in Appendix A) which is a fast-moving variable. However, the correlation coefficient between this spread and the slow-moving help-wanted employment ads (#23) is 0.6 and similar correlation for unemployment measures. Consequently, restrictions are imposed on these slow-moving variables instead. 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:

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

where numbers refer to the variable number listed in Appendix A page 38. Notice that the restrictions are not imposed on variables that are deemed a priori to be

---

<sup>23</sup>This was actually done in the first estimations and the imposition of a single "1" for CPI-U changed absolutely nothing. See Bork et al. (2008) for an oblique transformation of the factors towards a target loading matrix such that the first factor becomes an "inflation factor".

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,  $\ell_1, \dots, \ell_{r-1}$  may improve the overall fit although the improvement is deemed to be modest.

### 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 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<sup>24</sup>. These calculations point strongly towards  $r = 8$  which can be seen in figure 1 page 52.  $IC_{p2}$ ,  $PC_{p2}$  and  $PC_{p3}$  lead to exactly the same result and are therefore not shown.

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 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.

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 3 is to show that the fit of some variables such as unemployment and inflation, improves dramatically when more factors are

---

<sup>24</sup>I used  $C_{NT}^2$  to represent the imperfect convergence rate for the EM algorithm estimator.

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.

### 4.3.1 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 and 3 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  $p$ th 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.

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  $h$ th 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.  $F_t = \hat{F}_{t|T} + (F_t - \hat{F}_{t|T})$ , 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:

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

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

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

which are replaced by the (auto)covariances of the smoothed residuals from the



Kalman smoother, cf. (17) page 44:

$$\begin{aligned} C_0 &= \hat{\varepsilon}_{t|T} \hat{\varepsilon}_{t|T}^\top + P_{t|T}^\varepsilon \\ C_i &= \hat{\varepsilon}_{t|T} \hat{\varepsilon}_{t-i|T}^\top + P_{(t,t-i)|T}^\varepsilon. \end{aligned}$$

The upper panel of Table 4 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 5 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.

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  $r = \{3, 4, \dots, 10\}$  is selected. Attention to model parsimony influences the choice when competing FAVAR specifications are encountered<sup>25</sup>. 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 pol-

---

<sup>25</sup>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.

icy factor as this is always the last factor. The best specifications model among  $r = \{3, 4, \dots, 10\}$  is  $\{r3p7, r4p7, r5p6, r6p4, r7p5, \mathbf{r8p3}, r9p3, r10p2\}$  with the overall preferred model in bold. Figure 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 the other variables in the VAR is often more pronounced than for the policy factor itself.

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

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  $\Phi$  it should be noted that all eigenvalues of  $\Phi$  are less than 1 in modulus implying that the system is stationary.

#### 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 6 and 7 show the time series properties of the factors. Figures 8, 9, 10 and 11 show the correlation coefficients with the panel.

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 10. Factor seven is a capacity utilization factor<sup>26</sup> and factor eight is the monetary policy factor.

## 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.

---

<sup>26</sup>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  $\varepsilon_t$  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  $\varepsilon_t$ , 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  $\hat{Q} = W\Sigma_eW^\top$ , 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  $j$ th element of  $F$  at time  $t + i$  due to a change in the  $k$ th element of  $F$  at time  $t$  is:

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

for  $i = 1, 2, \dots, h$ , where  $\psi_i$  is the VAR moving average coefficient matrix and  $w_j$  is the  $j$ th column of the matrix  $W$ .  $\psi_i$  can be calculated recursively<sup>27</sup> from  $\Phi(L)$  in the stationary system in (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  $\Lambda$  is required, followed by a reverse transformation of the responses, i.e.  $D(L)[\Lambda\psi_i w_j]$ , cf. section 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 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.

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 following a contractionary monetary policy shock, and it can be seen that the preferred model  $r8p3$  returns more quickly to the starting point than BBE-

---

<sup>27</sup> $\psi_i = \sum_{j=1}^i \psi_{i-j} \Phi_j$  for  $i = 1, 2, \dots$  and  $\psi_0 = I$ . See Lütkepohl (2007) chapter 2.

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<sup>28</sup> 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.

## 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

---

<sup>28</sup>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 accommodate 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.)

forecast error variance for a particular variable is due to a specific shock, for instance the 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  $k$ th innovation  $e_{k,t}$  is given by:

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

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  $DD^\top$ ,  $MSE(\hat{X}_{j,t+h|t})$  is the mean square error of  $(X_{j,t+h} - \hat{X}_{j,t+h|t})$  and  $R_{j,j}$  is the variance of the  $j$ th idiosyncratic term. Details about the derivation are given in Appendix B.

The percentage of the forecast error variance explained by a monetary policy shock for the group of key macroeconomic variables is shown in Figure 13. 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.

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 6 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 6. 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.

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%.

## 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 order 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 percentage 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.

## References

- 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.
- 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) .
- 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. & Eliasch, P. (2005), ‘Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach’, *The Quarterly Journal of Economics* pp. 387–422.
- Bjørnland, H. & Leitemo, K. (2009), ‘Identifying the interdependence between US monetary policy and the stock market’, *Journal of Monetary Economics* (forthcoming) .
- 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.
- Breitung, J. & Eickmeier, S. (2006), ‘Dynamic factor models’, *Allgemeines Statistisches Archiv* **90**(1), 27–42.
- 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.
- 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.
- 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.
- 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., 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.
- 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. (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.
- 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.
- Jungbacker, B. & Koopman, S. J. (2008), Likelihood-based analysis for dynamic factor models, Tinbergen Institute Discussion Papers 08-007/4, Tinbergen Institute.
- 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.
- Lange, K. (1995), ‘A quasi-newton acceleration of the EM algorithm’, *Statistica Sinica* **5**, 1–18.
- 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) .
- Lütkepohl, H. (2007), *New Introduction to Multiple Time Series Analysis*, Springer, Berlin.
- 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.
- 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.
- 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.
- Stock, J. H. & Watson, M. W. (2002a), ‘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. (2002b), ‘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.
- 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.
- 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.

## A Data description

Data are from Bernanke et al. (2005).

First column: A superscript indicates that an identifying restrictions has been imposed on this variable, e.g. 11<sup>[1]</sup> indicates that an identifying restriction has been imposed on this variable for the *first* factor; similarly, 27<sup>[2]</sup> indicates an identifying restriction for the *second* factor. Note that the last factor is always restricted to be the federal funds rate in 77.

The second column is a mnemonic and a \* indicates a "slow-moving" variable. The 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	$\Delta \ln$	Industrial production: products, total (1992 = 100,SA)
2	IPF*	1959:01–2001:08	$\Delta \ln$	Industrial production: final products (1992 = 100,SA)
3	IPC*	1959:01–2001:08	$\Delta \ln$	Industrial production: consumer goods (1992 = 100,SA)
4	IPCD*	1959:01–2001:08	$\Delta \ln$	Industrial production: durable cons. goods (1992 = 100,SA)
5	IPCN*	1959:01–2001:08	$\Delta \ln$	Industrial production: nondurable cons. goods (1992 = 100,SA)
6	IPE*	1959:01–2001:08	$\Delta \ln$	Industrial production: business equipment (1992 = 100,SA)
7	IPI*	1959:01–2001:08	$\Delta \ln$	Industrial production: intermediate products (1992 = 100,SA)
8	IPM*	1959:01–2001:08	$\Delta \ln$	Industrial production: materials (1992 = 100,SA)
9	IPMD*	1959:01–2001:08	$\Delta \ln$	Industrial production: durable goods materials (1992 = 100,SA)
10	IPMND*	1959:01–2001:08	$\Delta \ln$	Industrial production: nondur. goods materials (1992 = 100,SA)
11 <sup>[1]</sup>	IPMFG*	1959:01–2001:08	$\Delta \ln$	Industrial production: manufacturing (1992 = 100,SA)
12	IPD*	1959:01–2001:08	$\Delta \ln$	Industrial production: durable manufacturing (1992 = 100,SA)
13	IPN*	1959:01–2001:08	$\Delta \ln$	Industrial production: nondur. manufacturing (1992 = 100,SA)
14	IPMIN*	1959:01–2001:08	$\Delta \ln$	Industrial production: mining (1992 = 100,SA)
15	IPUT*	1959:01–2001:08	$\Delta \ln$	Industrial production: utilities (1992 = 100,SA)
16 <sup>[9]</sup>	IP*	1959:01–2001:08	$\Delta \ln$	Industrial production: total index (1992 = 100,SA)
17 <sup>[7]</sup>	IPXMCA*	1959:01–2001:08	level	Capacity util rate: manufac., total (% of capacity,SA) (frb)
18 <sup>[3]</sup>	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	$\Delta \ln$	Personal income (chained) (series #52) (bil 92\$,SAAR)
21	GMYPXQ*	1959:01–2001:08	$\Delta \ln$	Personal inc. less trans. payments (chained) (#51) (bil 92\$,SAAR)

## (Un)employment and hours

22	LHEL*	1959:01–2001:08	$\Delta$ ln	Index of help-wanted advertising in newspapers (1967 = 100;SA)
23 <sup>[6]</sup>	LHELX*	1959:01–2001:08	ln	Employment: ratio; help-wanted ads: no. unemployed clf
24	LHEM*	1959:01–2001:08	$\Delta$ ln	Civilian labor force: employed, total (thous.,SA)
25	LHNAG*	1959:01–2001:08	$\Delta$ 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 <sup>[2]</sup>	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	$\Delta$ ln	Employees on nonag. payrolls: total (thous.,SA)
33	LP*	1959:01–2001:08	$\Delta$ ln	Employees on nonag. payrolls: total, private (thous.,SA)
34	LPGD*	1959:01–2001:08	$\Delta$ ln	Employees on nonag. payrolls: goods-producing (thous.,SA)
35	LPMI*	1959:01–2001:08	$\Delta$ ln	Employees on nonag. payrolls: mining (thous.,SA)
36	LPCC*	1959:01–2001:08	$\Delta$ ln	Employees on nonag. payrolls: contract construc. (thous.,SA)
37	LPEM*	1959:01–2001:08	$\Delta$ ln	Employees on nonag. payrolls: manufacturing (thous.,SA)
38	LPED*	1959:01–2001:08	$\Delta$ ln	Employees on nonag. payrolls: durable goods (thous.,SA)
39	LPEN*	1959:01–2001:08	$\Delta$ ln	Employees on nonag. payrolls: nondurable goods (thous.,SA)
40	LPSP*	1959:01–2001:08	$\Delta$ ln	Employees on nonag. payrolls: service-producing (thous.,SA)
41	LPTU*	1959:01–2001:08	$\Delta$ ln	Employees on nonag. payrolls: trans. and public util. (thous.,SA)
42	LPT*	1959:01–2001:08	$\Delta$ ln	Employees on nonag. payrolls: wholesale and retail (thous.,SA)
43	LPFR*	1959:01–2001:08	$\Delta$ ln	Employees on nonag. payrolls: finance, ins. and real est (thous.,SA)
44	LPS*	1959:01–2001:08	$\Delta$ ln	Employees on nonag. payrolls: services (thous.,SA)
45	LPGOV*	1959:01–2001:08	$\Delta$ ln	Employees on nonag. payrolls: government (thous.,SA)
46	LPHRM*	1959:01–2001:08	level	Avg. weekly hrs. of production wkrs.: manufacturing (sa)
47 <sup>[4]</sup>	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	$\Delta$ ln	Pers cons exp (chained)—total (bil 92\$,SAAR)
50 <sup>[8]</sup>	GMCDQ*	1959:01–2001:08	$\Delta$ ln	Pers cons exp (chained)—tot. dur. (bil 96\$,SAAR)
51	GMCNQ*	1959:01–2001:08	$\Delta$ ln	Pers cons exp (chained)—nondur. (bil 92\$,SAAR)
52	GMCSQ*	1959:01–2001:08	$\Delta$ ln	Pers cons exp (chained)—services (bil 92\$,SAAR)
53	GMCANQ*	1959:01–2001:08	$\Delta$ 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, orders 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	$\Delta \ln$	New orders (net)—consumer goods and materials, 1992 \$ (bci)
65	MSONDQ	1959:01–2001:08	$\Delta \ln$	New orders, nondefense capital goods, in 1992 \$s (bci)

### Stock prices

66	FSNCOM	1959:01–2001:08	$\Delta \ln$	NYSE composite (12/31/65 = 50)
67	FSPCOM	1959:01–2001:08	$\Delta \ln$	S&P's composite (1941–1943 = 10)
68	FSPIN	1959:01–2001:08	$\Delta \ln$	S&P's industrials (1941–1943 = 10)
69	FSPCAP	1959:01–2001:08	$\Delta \ln$	S&P's capital goods (1941–1943 = 10)
70	FSPUT	1959:01–2001:08	$\Delta \ln$	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	$\Delta \ln$	Foreign exchange rate: Switzerland (Swiss franc per US\$)
74	EXRJAN	1959:01–2001:08	$\Delta \ln$	Foreign exchange rate: Japan (yen per US\$)
75	EXRUK	1959:01–2001:08	$\Delta \ln$	Foreign exchange rate: United Kingdom (cents per pound)
76	EXRCAN	1959:01–2001:08	$\Delta \ln$	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 tbill,sec mkt,3-mo. (% per ann,nsa)
79	FYGM6	1959:01–2001:08	level	Interest rate: us tbill,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 <sup>[5]</sup>	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
-----	--------	-----------------	-------	-------------------------------

## B Appendix: Forecast error variance decomposition

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

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

where  $e_t$  is the orthogonal residual defined from the VAR residuals,  $\varepsilon_t = P e_t$  where  $P$  is the Cholesky factor from the decomposition of the covariance of  $\varepsilon_t$  into  $\hat{Q} = PP^\top$ . This covariance matrix is further rewritten as explained in section 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  $j$ th idiosyncratic term. The mean square error of  $(X_{j,t+h} - \hat{X}_{j,t+h|t})$  is denoted  $MSE(\hat{X}_{j,t+h|t})$  and given by:

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

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

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

## C Appendix: Kalman filter, Kalman smoother and the EM algorithm

### C.1 The Kalman filter

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

$$\begin{aligned}\hat{F}_{t+1|t} &= \Phi \hat{F}_{t|t-1} + K_t (X_t - \Lambda \hat{F}_{t|t-1}) \\ \hat{P}_{t+1|t} &= \Phi \hat{P}_{t|t-1} L_t^\top + Q\end{aligned}\tag{11}$$

where

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

### C.2 Kalman smoothing

Kalman smoothing reconstructs the full state sequence  $\{F_1, \dots, F_T\}$  given the observations  $\{X_1, \dots, X_T\}$ . 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:

$$\hat{F}_{t|T} = \hat{F}_{t|t-1} + \hat{P}_{t|t-1} \Lambda^\top \left[ \hat{P}_{t|t-1}^{\xi\xi} \right]^{-1} \xi_t + \hat{P}_{t|t-1} L_t^\top r_t \quad (12)$$

$$= \hat{F}_{t|t-1} + \hat{P}_{t|t-1} r_{t-1} \quad (\text{alternatively})$$

$$\hat{P}_{t|T} = \hat{P}_{t|t-1} - \hat{P}_{t|t-1} N_{t-1} \hat{P}_{t|t-1} \quad (13)$$

$$\hat{P}_{\{t,t-1\}|T} = \left( I - \hat{P}_{t|t-1} N_{t-1} \right) L_{t-1} \hat{P}_{t-1|t-2}, \quad (14)$$

$$\text{for } t = T-1, \dots, 1 \quad (15)$$

$$\text{cov} \left( F_t - \hat{F}_{t|T}, F_j - \hat{F}_{j|T} \right) = \hat{P}_{t|t-1} L_t^\top L_{t+1}^\top \cdots L_{j-1}^\top \left[ I - N_{j-1} \hat{P}_{j|j-1} \right] \quad (16)$$

$$\text{for } j \geq t$$

where:

$$r_{t-1} = \Lambda^\top \left[ \hat{P}_{t|t-1}^{\xi\xi} \right]^{-1} \xi_t + L_t^\top r_t, \text{ for } 1 \leq t < T \text{ and } r_T = 0$$

$$N_{t-1} = \Lambda^\top \left[ \hat{P}_{t|t-1}^{\xi\xi} \right]^{-1} \Lambda + L_t^\top N_t L \text{ for } 1 \leq t < T \text{ and } N_T = 0$$

$$L_t = \Phi - K_t \Lambda = \Phi - \Phi \hat{P}_{t|t-1} \Lambda^\top \left[ \hat{P}_{t|t-1}^{\xi\xi} \right]^{-1} \Lambda.$$

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

$$\begin{aligned} \hat{\varepsilon}_{t|T} &= E[\varepsilon_t | \mathcal{X}_T] = \hat{F}_{t|T} - \Phi \hat{F}_{t-1|T} \\ &= Q \Upsilon^\top r_t, \quad t = 1, \dots, T \end{aligned} \quad (17)$$

and variance and covariance:

$$\text{var}(\varepsilon_t) = Q - Q \Upsilon^\top N_t \Upsilon Q$$

$$\text{cov}(\varepsilon_t - \hat{\varepsilon}_{t|T}, \varepsilon_j - \hat{\varepsilon}_{j|T}) = -Q \Upsilon^\top L_{t+1}^\top \cdots L_{j-1}^\top L_j^\top N_j \Upsilon Q, \quad j = t+1, \dots, T$$

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  $t = T-1$ .

### C.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 (4) page 10 assuming a VAR(1) for simplicity and ignoring  $\Upsilon$  is written as:

$$\begin{aligned}
-2 \ln L_{\mathcal{F}, \mathcal{X}}(\Theta) &= \ln |P_0| + (F_0 - \mu_0)^\top P_0^{-1} (F_0 - \mu_0) \\
&+ T \cdot \ln |Q| + \sum_{t=1}^T (F_t - \Phi F_{t-1})^\top Q^{-1} (F_t - \Phi F_{t-1}) \\
&+ T \cdot \ln |R| + \sum_{t=1}^T (X_t - \Lambda F_t)^\top R^{-1} (X_t - \Lambda F_t)
\end{aligned} \tag{18}$$

given that we can observe the states  $\mathcal{F}_T = \{F_0, \dots, F_T\}$  as well as the observations  $\mathcal{X}_T = \{X_1, \dots, X_T\}$ . 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:

$$\begin{aligned}
\mathcal{Q}(\Theta | \Theta^{(j-1)}) &= E[-2 \ln L_{\mathcal{F}, \mathcal{X}}(\Theta) | \mathcal{X}_T, \Theta^{(j-1)}] \\
&= \ln |P_0| + \text{tr} \left[ P_0^{-1} \left\{ \left( \hat{F}_{0|T} - \mu_0 \right) \left( \hat{F}_{0|T} - \mu_0 \right)^\top + P_{0|T} \right\} \right] \\
&+ T \cdot \ln |Q| + \text{tr} \left[ Q^{-1} \left\{ C - B\Phi^\top - \Phi B^\top + \Phi A\Phi^\top \right\} \right] \\
&+ T \cdot \ln |R| + \text{tr} \left[ R^{-1} \sum_{t=1}^T \left\{ \left( X_t - \Lambda \hat{F}_{t|T} \right) \left( X_t - \Lambda \hat{F}_{t|T} \right)^\top + \Lambda \hat{P}_{t|T} \Lambda^\top \right\} \right]
\end{aligned} \tag{19}$$

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

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

A useful trick to arrive at (19) is to consider the decomposition of the true state variable  $F_t = \hat{F}_{t|T} + (F_t - \hat{F}_{t|T})$ , which explains the terms in for instance

$C$ , where:

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

The estimator of  $\Phi^*$  subject to linear restrictions is:

$$\begin{aligned} \text{vec}(\Phi^*) &= \text{vec}(BA^{-1}) \\ &+ (A^{-1} \otimes Q) H_\Phi^\top [H_\Phi (A^{-1} \otimes Q) H_\Phi^\top]^{-1} \{ \kappa_\Phi - H_\Phi \text{vec}(BA^{-1}) \} \end{aligned} \quad (20)$$

where  $\kappa_\Phi$  is a  $\varrho \times 1$  vector and the restriction matrix  $H_\Phi$  is of dimension  $\varrho \times r^2$ .

## D Appendix: Analytical derivatives of the log likelihood function

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

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

$$\left. \frac{\partial \log L_Y(\Theta)}{\partial Q} \right|_{Q=Q^*} = \left. \frac{\partial \mathcal{Q}(\Theta|\Theta^*)}{\partial Q} \right|_{Q=Q^*} = Q^{-1} (S - T \cdot Q) Q^{-1} - \frac{1}{2} \text{diag} (Q^{-1} (S - T \cdot Q) Q^{-1})$$

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:

$$\left. \frac{\partial \log L_Y(\Theta)}{\partial \Phi} \right|_{\Phi=\Phi^*} = \left. \frac{\partial \mathcal{Q}(\Theta|\Theta^*)}{\partial \Phi} \right|_{\Phi=\Phi^*} = Q^{-1} (B - \Phi A)$$

where  $\Phi$  contains the autoregressive parameters in the transition equation. Moreover, the derivative with respect to  $\Lambda$  is:

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

where  $\Lambda$  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:

$$\frac{\partial \log L_y(\Theta)}{\partial R} \Big|_{R=R^*} = \frac{\partial \mathcal{Q}(\Theta|\Theta^*)}{\partial R} \Big|_{R=R^*} = R^{-1} (M - T \cdot R) R^{-1} - \frac{1}{2} \text{diag} (R^{-1} (M - T \cdot R) R^{-1}).$$



**Table 1: Akaike information criterion for a given number of factors.**

lags	number of factors							
	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	- <b>23.221</b>	- 27.482	- 24.302	- 35.706	- 35.812	- 45.512
3	- 11.616	- <b>16.646</b>	- 19.462	-	-	- 37.692	- <b>40.605</b>	- <b>50.218</b>
4	- 11.663	- 16.182	- 18.741	- <b>28.607</b>	-	- 36.904	- 38.251	- 48.553
5	- 11.562	-	- 19.479	- 26.748	- <b>30.378</b>	- 37.673	-	-
6	- <b>11.904</b>	- 15.653	- 20.636	- 26.683	- 24.834	- <b>38.726</b>	- 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 2: Schwarz information criterion for a given number of factors.**

lags	number of factors							
	3	4	5	6	7	8	9	10
1	- 11.392	- 13.139	- 17.418	- 24.457	- 22.252	-	- 29.809	- 42.972
2	- <b>11.485</b>	- 13.886	- <b>22.806</b>	- 26.885	- 23.490	- 34.645	- 34.469	- 43.854
3	- 11.393	- <b>16.248</b>	- 18.841	-	-	- <b>36.100</b>	- <b>38.590</b>	- <b>47.731</b>
4	- 11.365	- 15.651	- 17.912	- <b>27.413</b>	-	- 34.782	- 35.565	- 45.237
5	- 11.189	-	- 18.443	- 25.255	- <b>28.347</b>	- 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 3: Hannan and Quinn information criterion for a given number of factors.**

lags	number of factors							
	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	- <b>23.058</b>	- 27.248	- 23.984	- 35.290	- 35.285	- 44.862
3	- 11.529	- <b>16.490</b>	- 19.219	-	-	- 37.068	- <b>39.815</b>	- <b>49.243</b>
4	- 11.546	- 15.974	- 18.416	- <b>28.139</b>	-	- 36.072	- 37.198	- 47.253
5	- 11.416	-	- 19.073	- 26.163	- <b>29.582</b>	- 36.633	-	-
6	- <b>11.729</b>	- 15.341	- 20.148	- 25.981	- 23.879	- <b>37.478</b>	- 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 4: 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* {0.000}	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 5: 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.

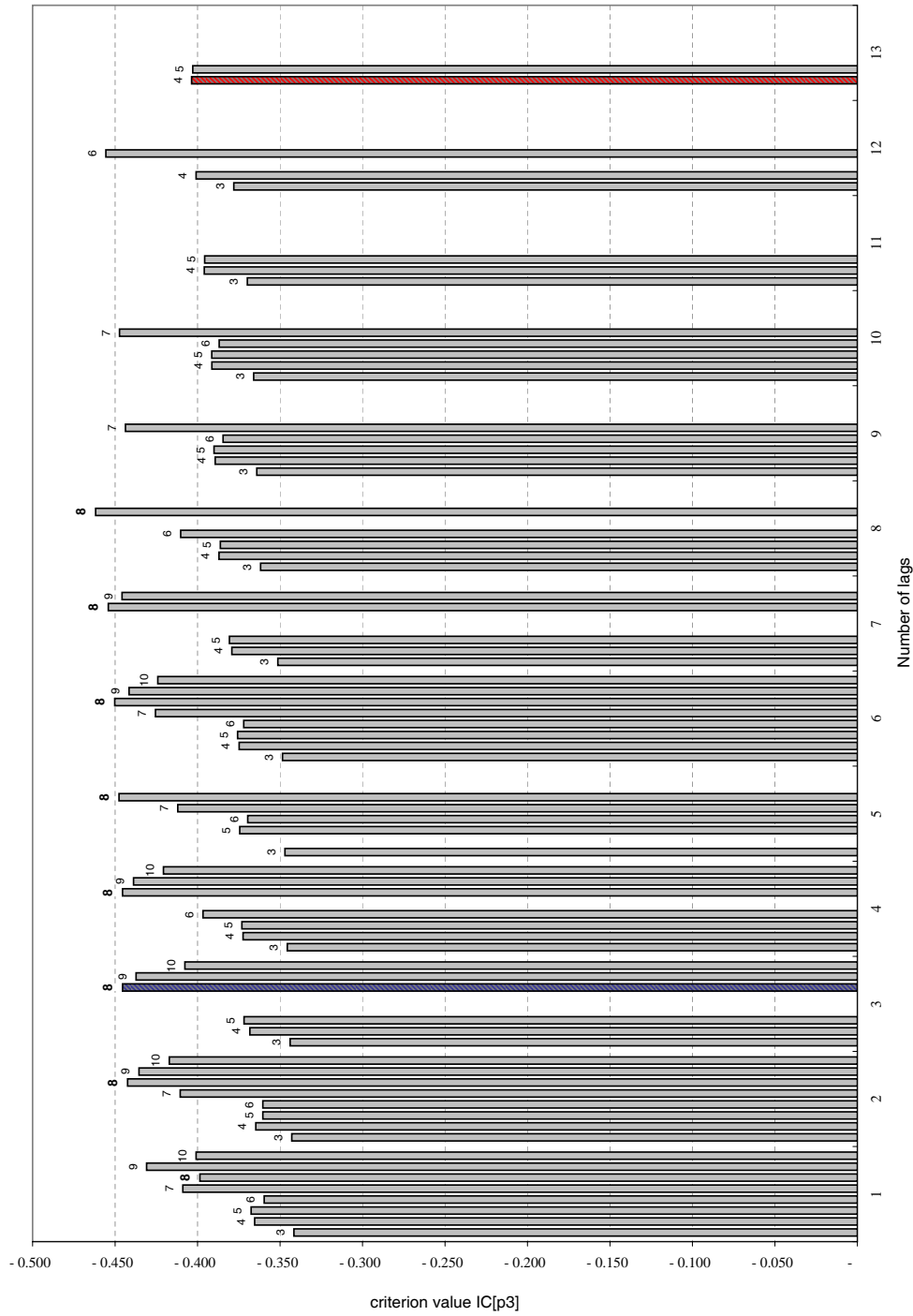
**Table 6: Forecast error variance decompositions for key macroeconomic variables.**

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

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

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: 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 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.

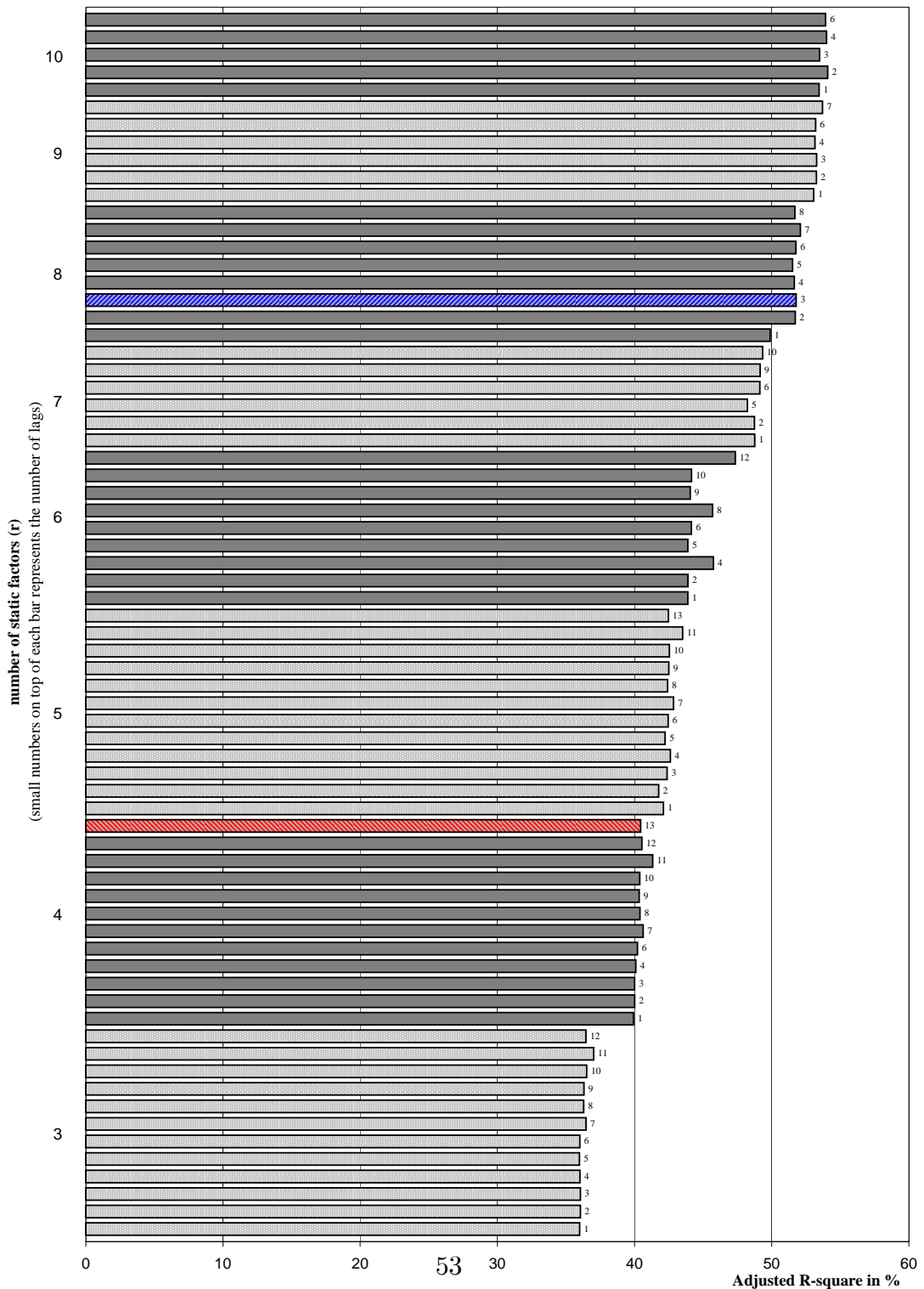
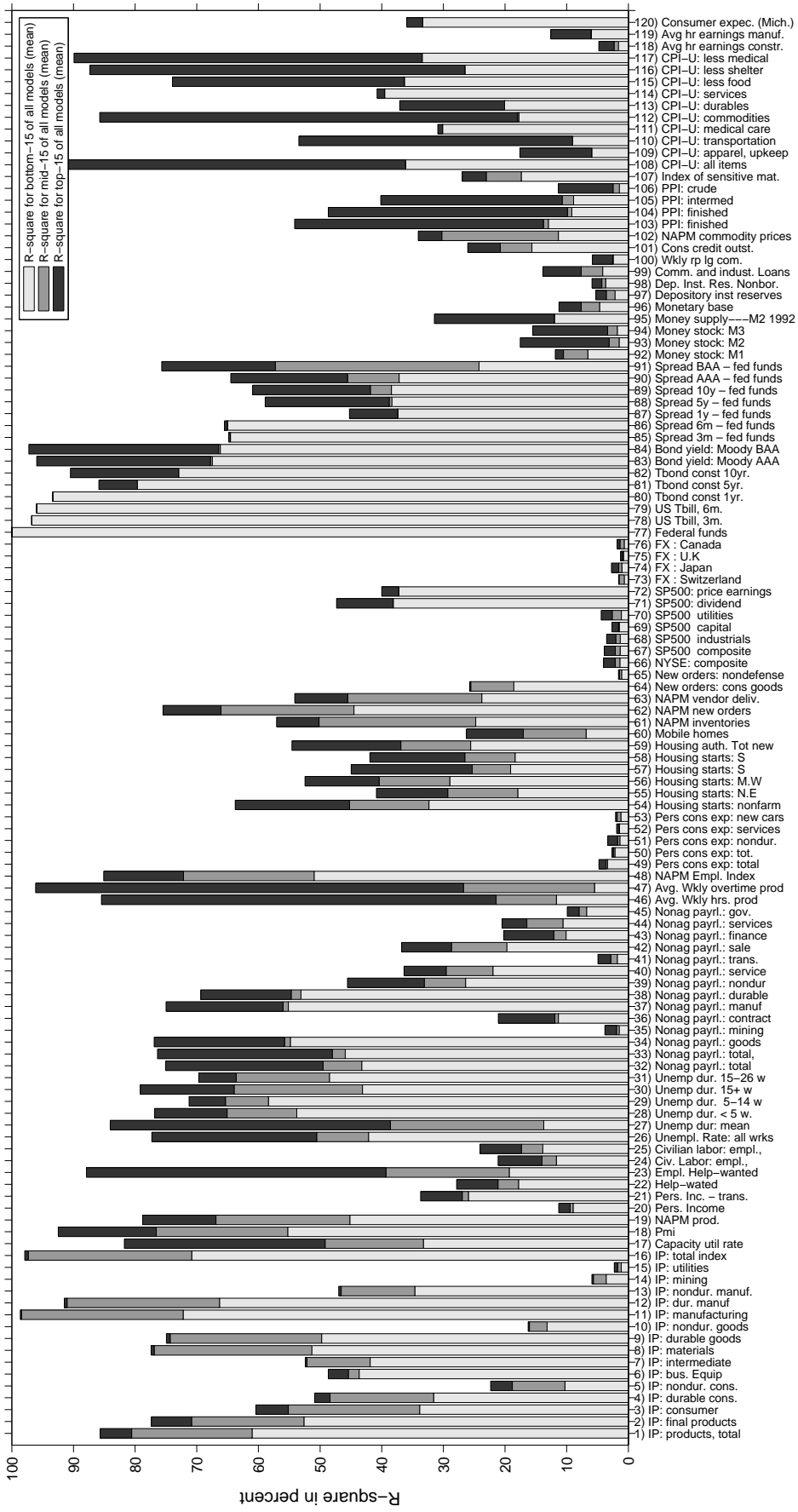
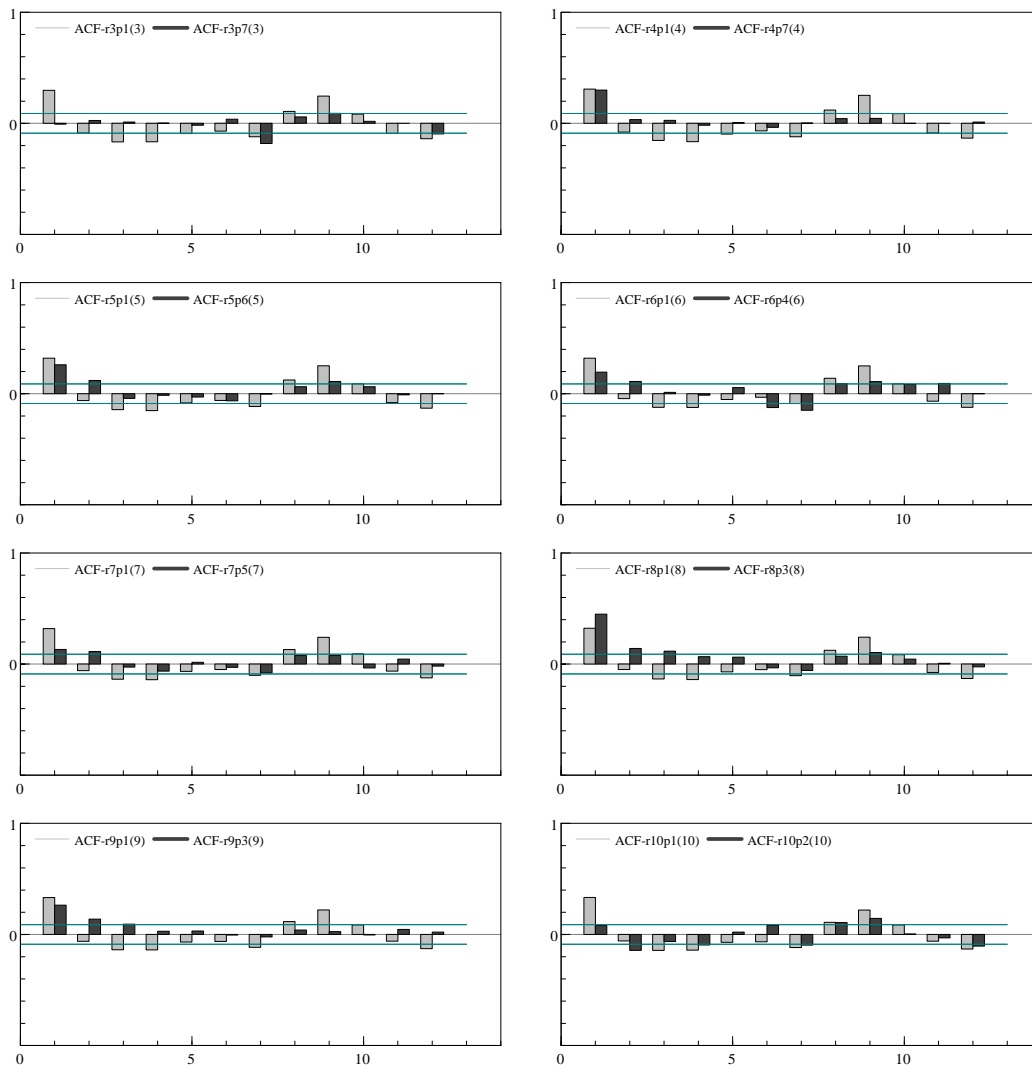


Figure 3:  $R^2$  for each variable: Mean of the "15 best", "15 medium" and "15 worst" performing models.



**Figure 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 5:  $R^2$  for the preferred model versus the BBE-EM and Bernanke, Boivin and Eliaz.

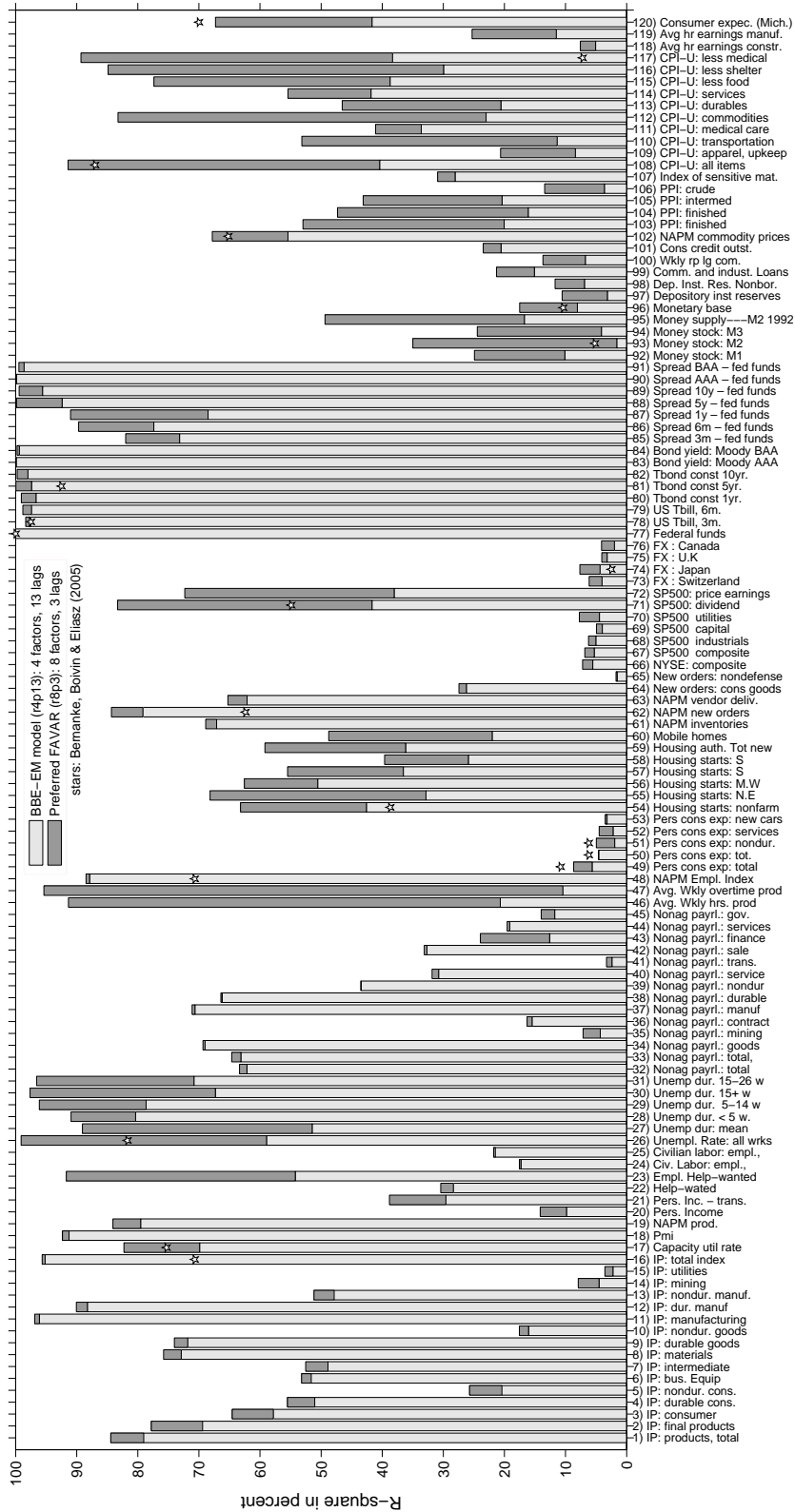


Figure 6: The time series of factors 1-4 from the preferred model versus related observed variables.

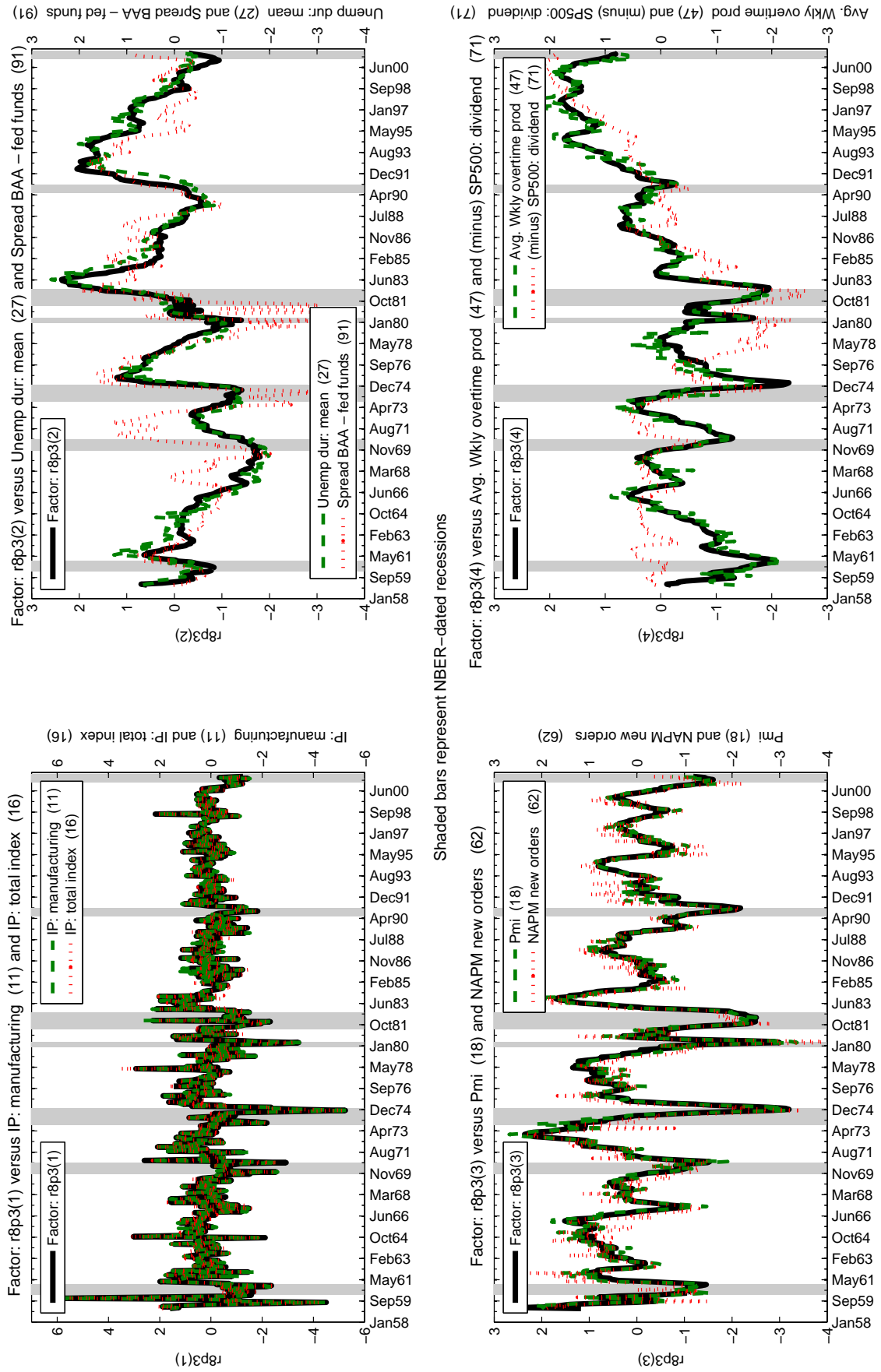


Figure 7: The time series of factors 5-8 from the preferred model versus related observed variables.

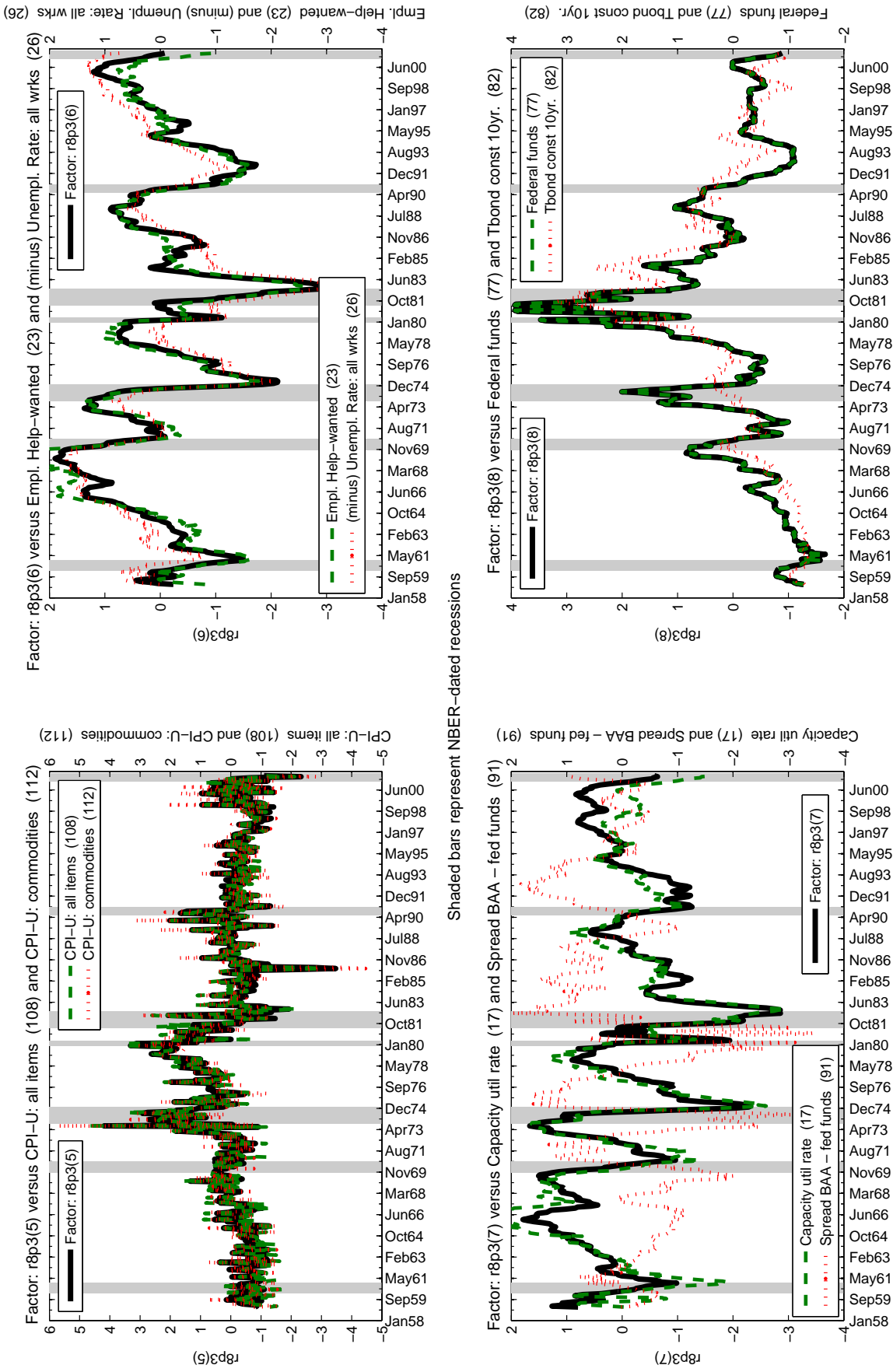
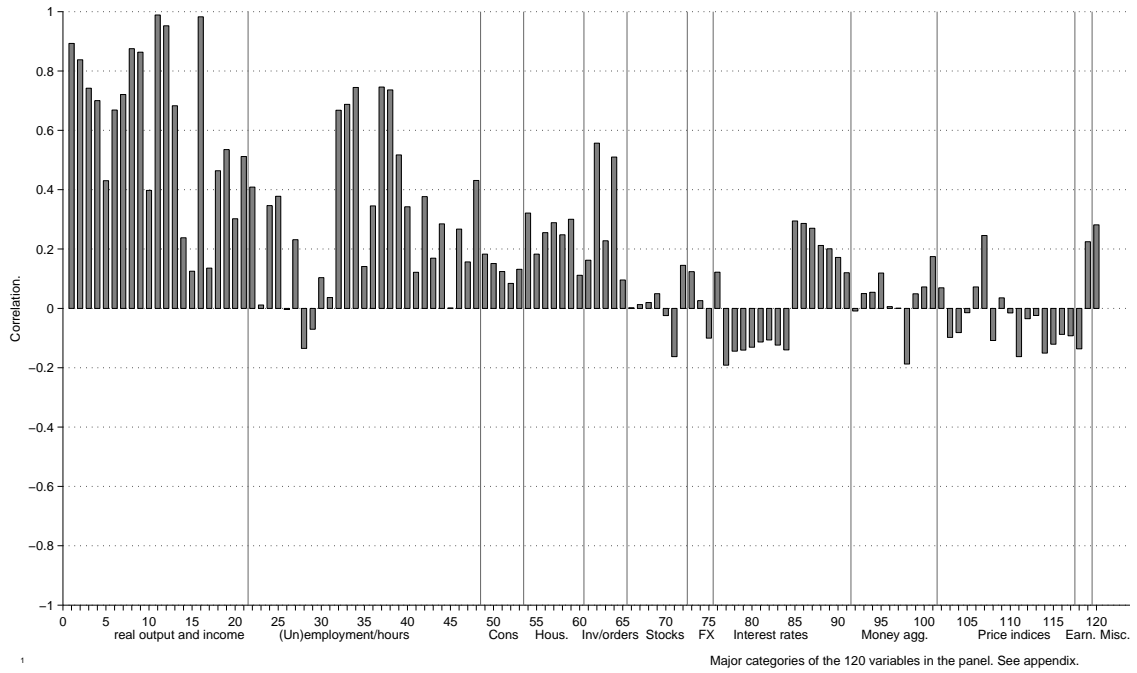
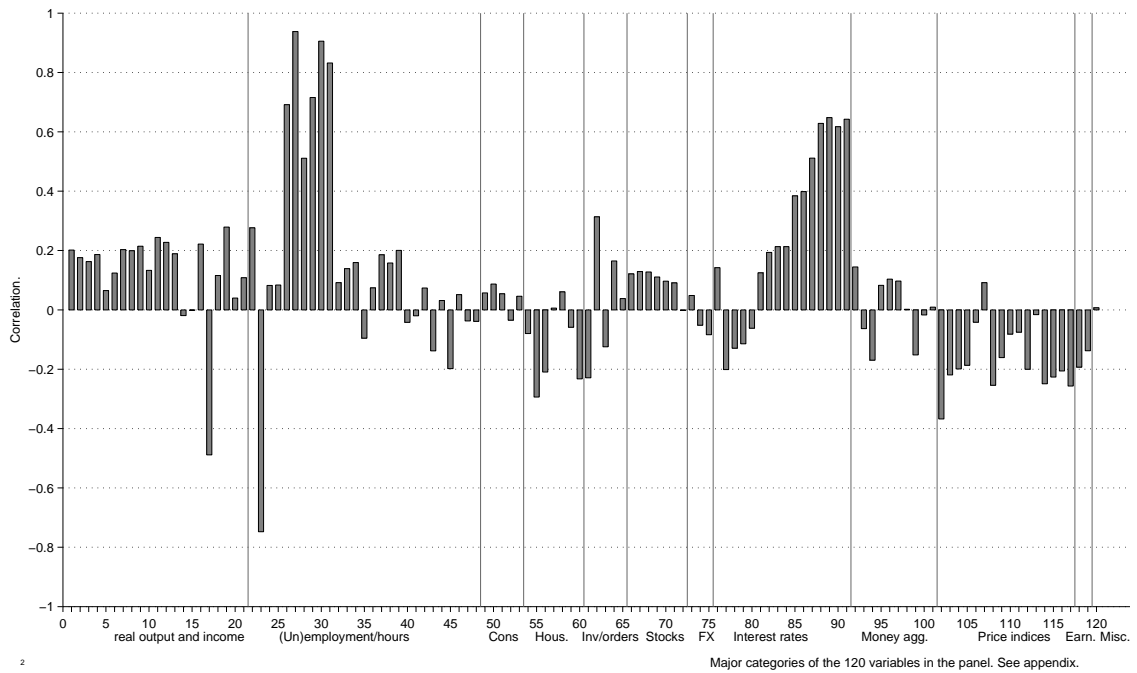


Figure 8: "Industrial production factor" and "unemployment factor".

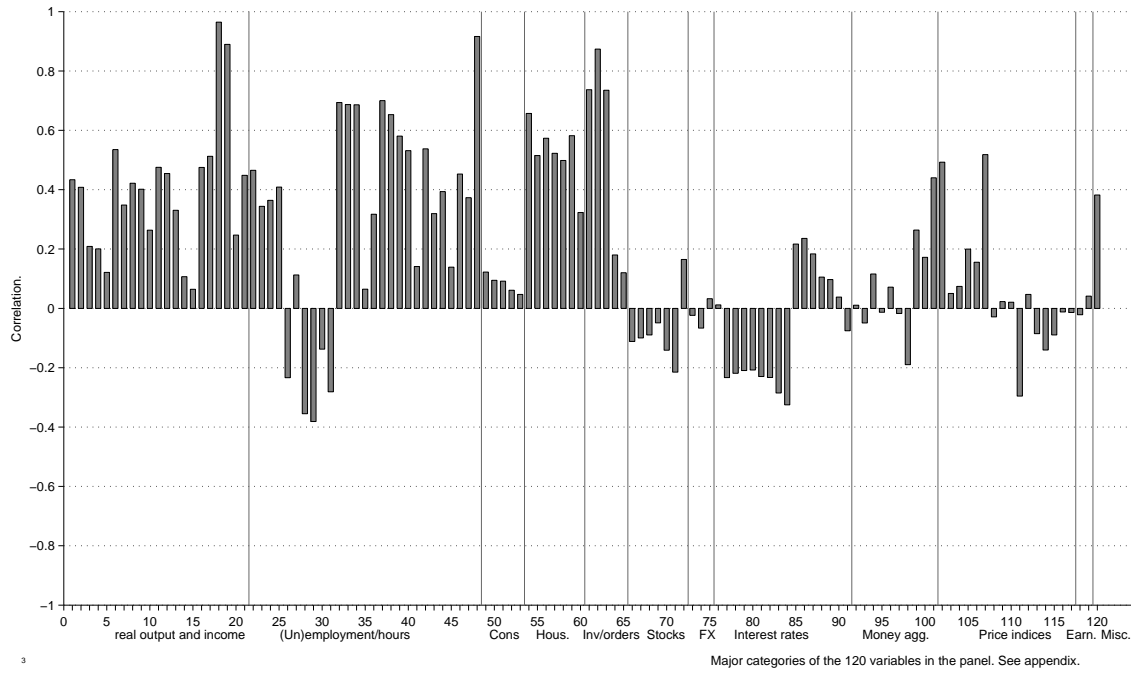


Correlation between factor 1 and each of the variables as a function of  $(\Phi, Q, \Lambda)$ .

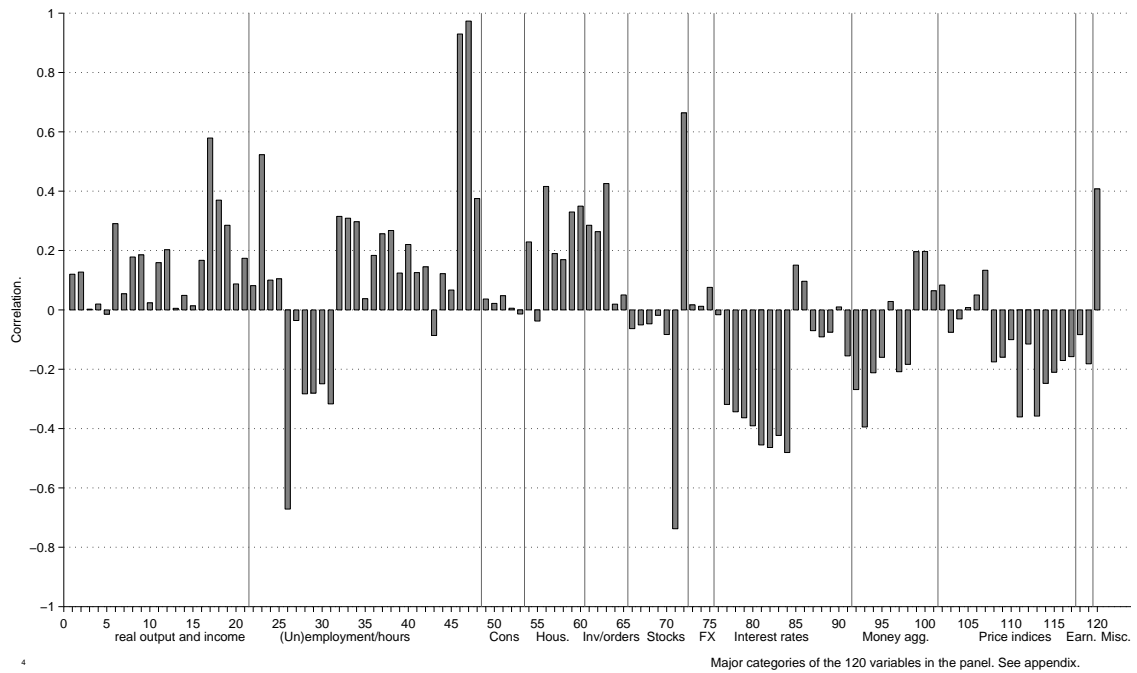


Correlation between factor 2 and each of the variables as a function of  $(\Phi, Q, \Lambda)$ .

Figure 9: "NAPM factor" and "(overtime)hours factor".

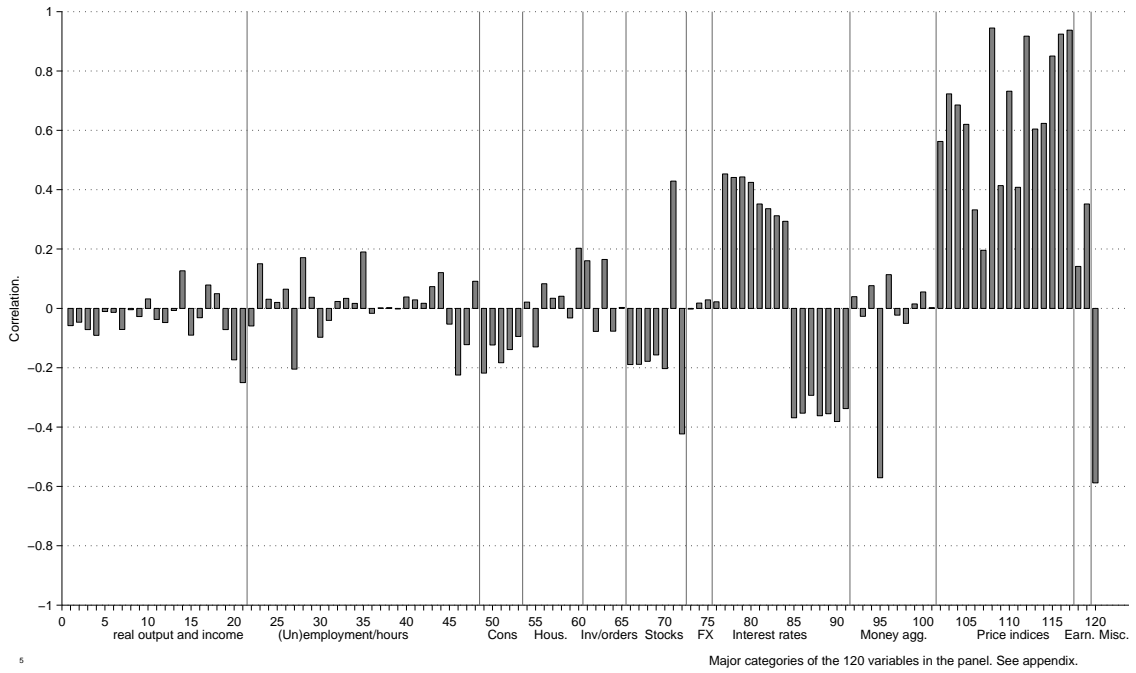


Correlation between factor 3 and each of the variables as a function of  $(\Phi, Q, \Lambda)$ .

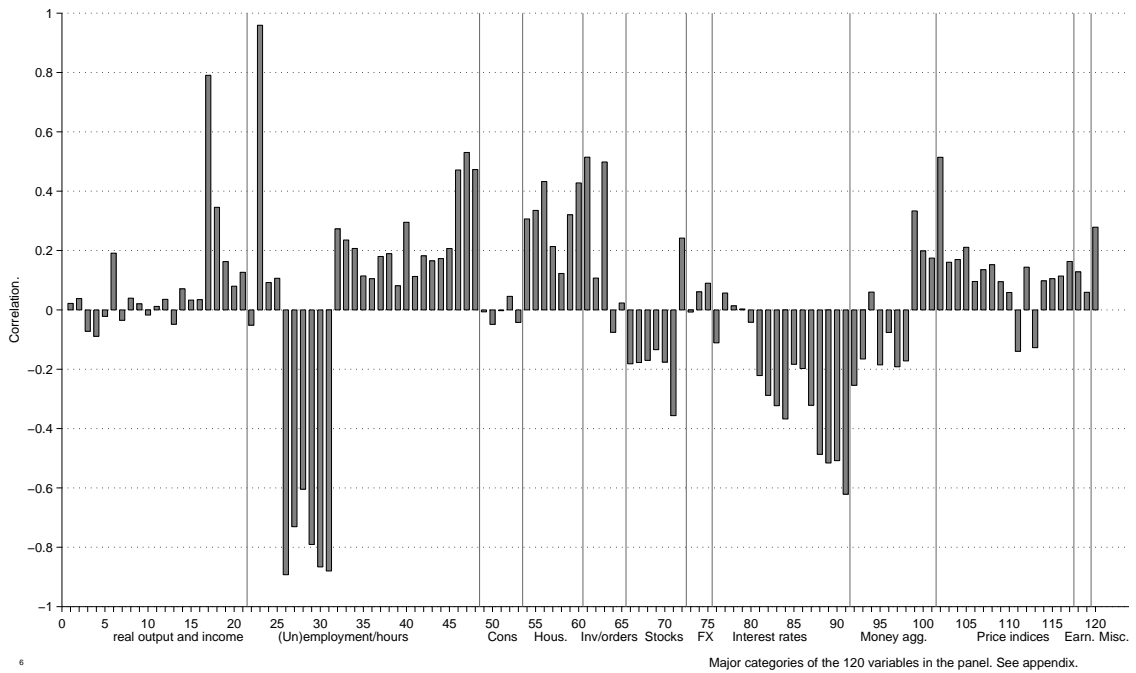


Correlation between factor 4 and each of the variables as a function of  $(\Phi, Q, \Lambda)$ .

Figure 10: "Inflation factor" and "employment factor".

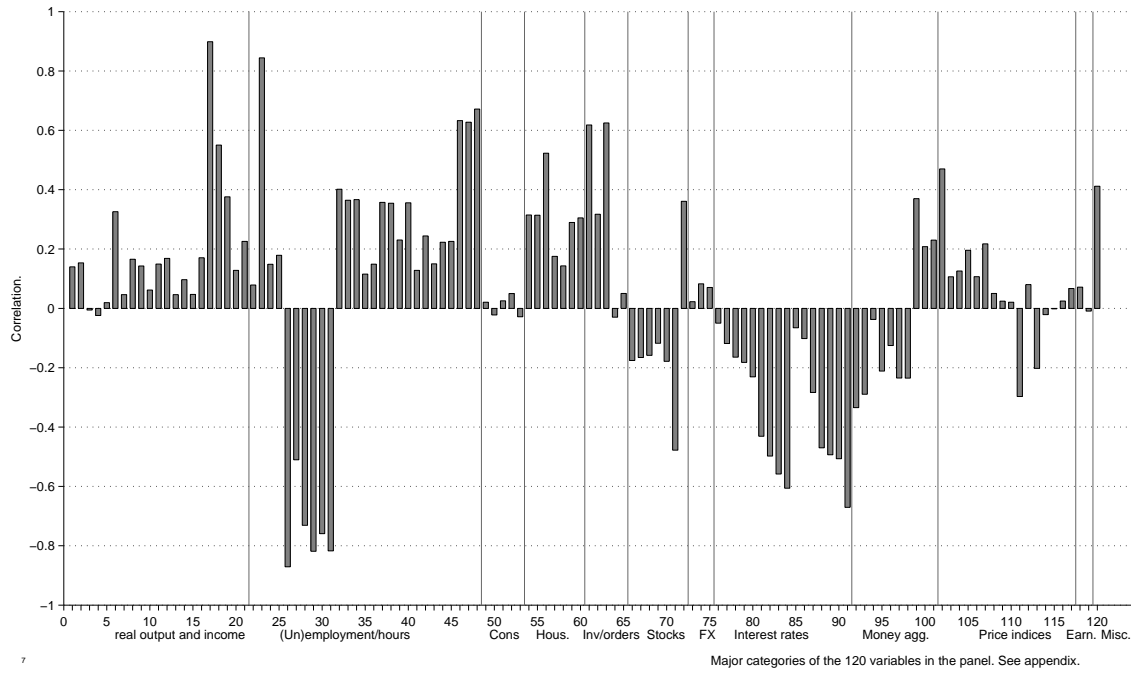


Correlation between factor 5 and each of the variables as a function of  $(\Phi, Q, \Lambda)$ .

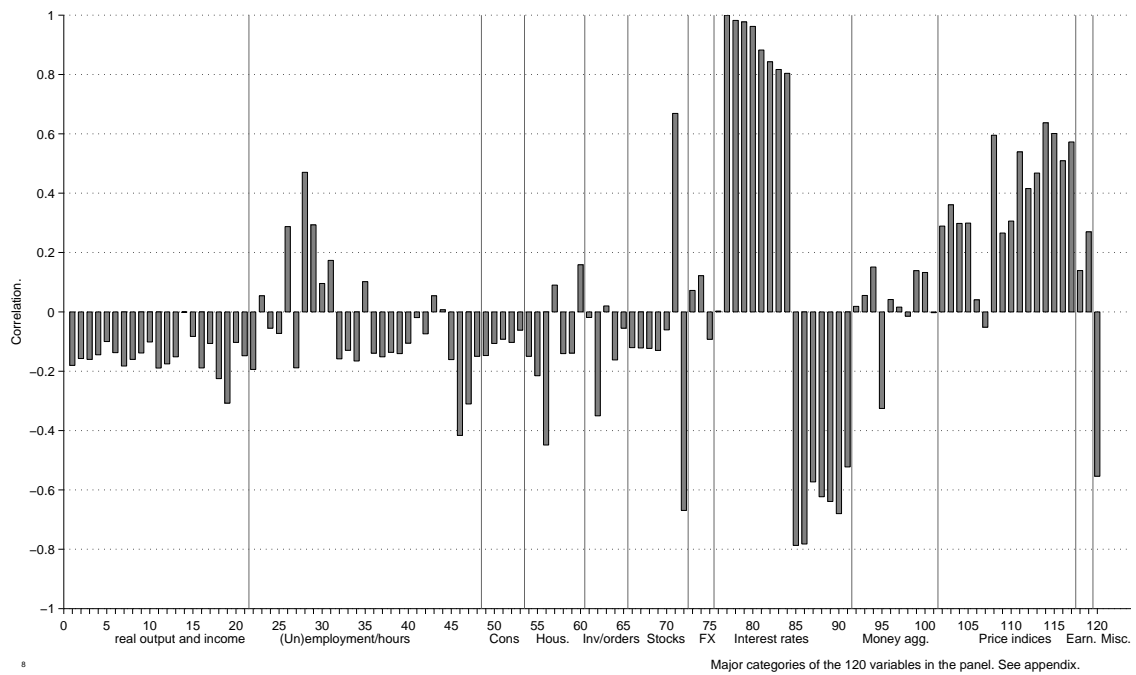


Correlation between factor 6 and each of the variables as a function of  $(\Phi, Q, \Lambda)$ .

Figure 11: "Capacity utilization factor" and "Monetary policy factor".

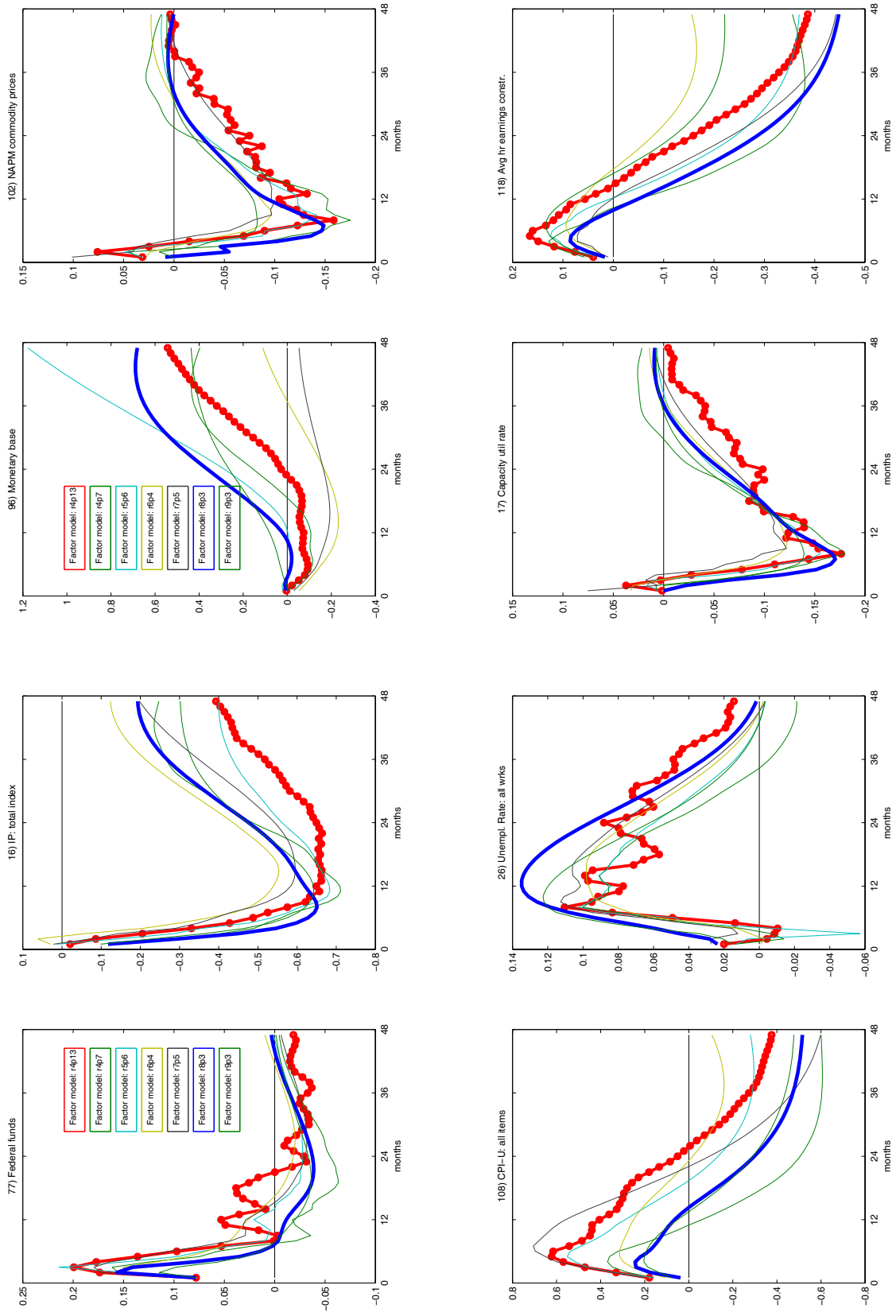


Correlation between factor 7 and each of the variables as a function of  $(\Phi, Q, \Lambda)$ .



Correlation between factor 8 and each of the variables as a function of  $(\Phi, Q, \Lambda)$ .

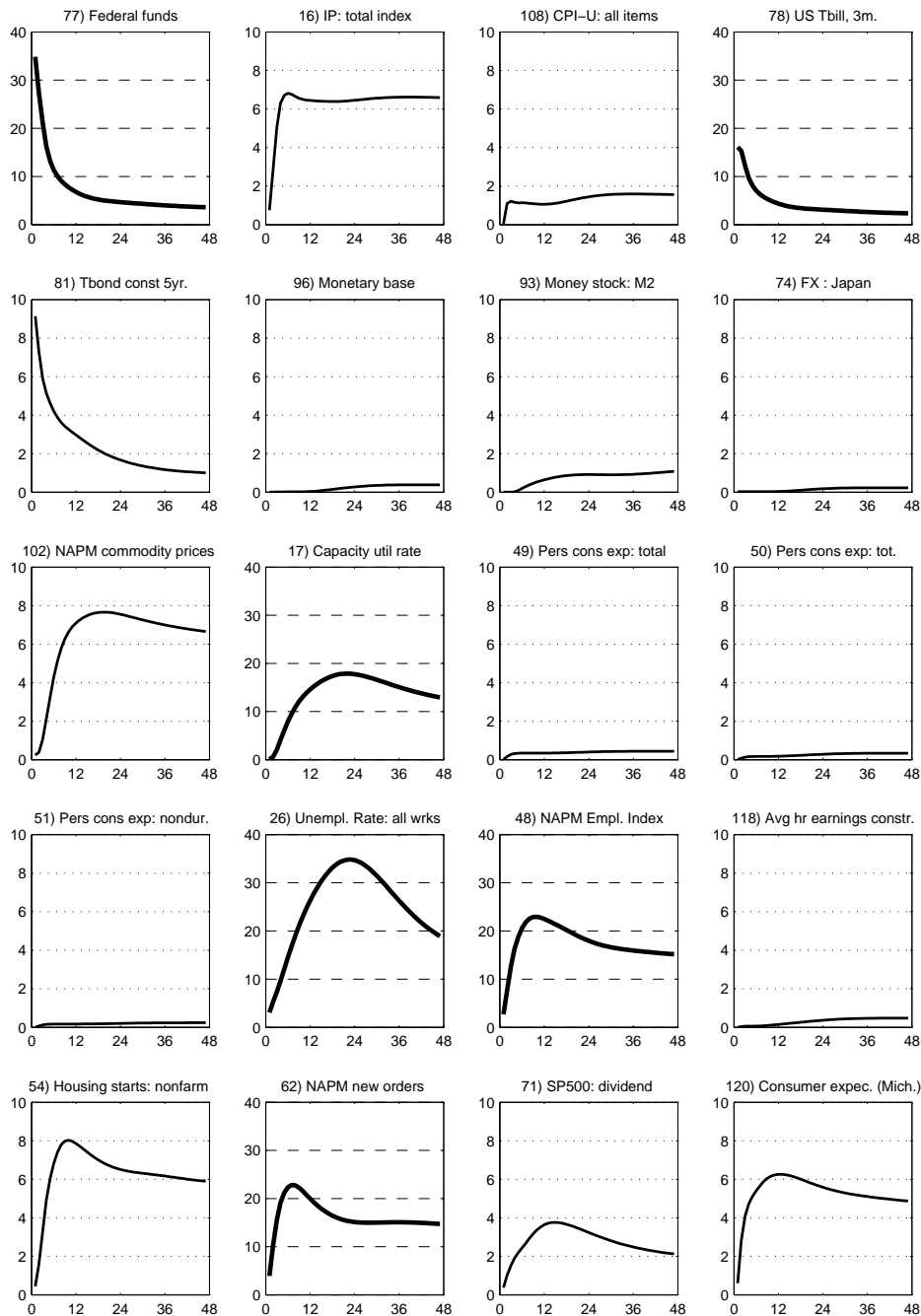
Figure 12: Impulse responses in standard deviations to a 25 basis point monetary policy shock.



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



**Figure 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.

# Research Papers 2009



- 2008-60: Thomas Q. Pedersen: Intertemporal Asset Allocation with Habit Formation in Preferences: An Approximate Analytical Solution
- 2008-61: Jean Jacod, Mark Podolskij and Mathias Vetter: Limit theorems for moving averages of discretized processes plus noise
- 2008-62: Giuseppe Cavaliere, David I. Harvey, Stephen J. Leybourne and A.M. Robert Taylor: Testing for Unit Roots in the Presence of a Possible Break in Trend and Non-Stationary Volatility
- 2008-63: Ole E. Barndorff-Nielsen, Peter Reinhard Hansen, Asger Lunde and Neil Shephard: Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading
- 2009-01: Roman Frydman, Michael D. Goldberg, Søren Johansen and Katarina Juselius: A Resolution of the Purchasing Power Parity Puzzle: Imperfect Knowledge and Long Swings
- 2009-02: Morten Ørregaard Nielsen: Nonparametric Cointegration Analysis of Fractional Systems With Unknown Integration Orders
- 2009-03: Andrés González, Kirstin Hubrich and Timo Teräsvirta: Forecasting inflation with gradual regime shifts and exogenous information
- 2009-4: Theis Lange: First and second order non-linear cointegration models
- 2009-5: Tim Bollerslev, Natalia Sizova and George Tauchen: Volatility in Equilibrium: Asymmetries and Dynamic Dependencies
- 2009-6: Anders Tolver Jensen and Theis Lange: On IGARCH and convergence of the QMLE for misspecified GARCH models
- 2009-7: Jeroen V.K. Rombouts and Lars Stentoft: Bayesian Option Pricing Using Mixed Normal Heteroskedasticity Models
- 2009-8: Torben B. Rasmussen: Jump Testing and the Speed of Market Adjustment
- 2009-9: Dennis Kristensen and Andrew Ang: Testing Conditional Factor Models
- 2009-10: José Fajardo and Ernesto Mordecki: Skewness Premium with Lévy Processes
- 2009-11: Lasse Bork: Estimating US Monetary Policy Shocks Using a Factor-Augmented Vector Autoregression: An EM Algorithm Approach