EFWW
7/18 11:30

Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregessive (FAVAR) Approach

Ben S. Bernanke, Princeton University

Jean Boivin, Columbia University

Piotr Eliasz, Princeton University

First draft: April 2002

Revised: July 2003

Thanks to Christopher Sims, Mark Watson and Tao Zha for useful comments. Boivin would like to thank National Science Foundation for financial support (SES-0214104).

1. Introduction

Since Bernanke and Blinder (1992) and Sims (1992), a considerable literature has developed that employs vector autoregression (VAR) methods to attempt to identify and measure the effects of monetary policy innovations on macroeconomic variables (see Christiano, Eichenbaum, and Evans, 2000, for a survey). The key insight of this approach is that identification of the effects of monetary policy shocks requires only a plausible identification of those shocks (for example, as the unforecasted innovation of the federal funds rate in Bernanke and Blinder, 1992) and does not require identification of the remainder of the macroeconomic model. These methods generally deliver empirically plausible assessments of the dynamic responses of key macroeconomic variables to monetary policy innovations, and they have been widely used both in assessing the empirical fit of structural models (see, for example, Boivin and Giannoni, 2003; Christiano, Eichenbaum, and Evans, 2001) and in policy applications.

The VAR approach to measuring the effects of monetary policy shocks appears to deliver a great deal of useful structural information, especially for such a simple method. Naturally, the approach does not lack for criticism. For example, researchers have disagreed about the appropriate strategy for identifying policy shocks (Christiano, Eichenbaum, and Evans, 2000, survey some of the alternatives; see also Bernanke and Mihov, 1998). Alternative identifications of monetary policy innovations can, of course, lead to different inferences about the shape and timing of the responses of economic variables. Another issue is that the standard VAR approach addresses only the effects of unanticipated changes in monetary policy, not the arguably more important effects of the systematic portion of monetary policy or the choice of monetary policy rule (Sims and Zha, 1998; Cochrane, 1996; Bernanke, Gertler, and Watson, 1997).

Several criticisms of the VAR approach to monetary policy identification center around the relatively small amount of information used by low-dimensional VARs. To conserve degrees of freedom, standard VARs rarely employ more than six to eight variables.[1] This small number of variables is unlikely to span the information sets used by actual central banks, who are known to follow literally hundreds of data series, or by the financial markets and other Fed-watchers. The sparse information sets used in typical analyses lead to at least two potential sets of problems with the results. First, to the extent that central banks and the private sector have information not reflected in the VAR analysis, the measurement of policy innovations is likely to be contaminated. A standard illustration of this potential problem, which we explore in this paper, is the Sims (1992) interpretation of the so-called "price puzzle", the conventional finding in the VAR literature that a contractionary monetary policy shock is followed by a slight increase in the price level, rather than a decrease as standard economic theory would predict. Sims's explanation for the price puzzle is that it is the result of imperfectly controlling for information that the central bank may have about future inflation. If the Fed systematically tightens policy in anticipation of future inflation, and if these signals of future inflation are not adequately captured by the data series in the VAR, then what appears to the VAR to be a policy shock may in fact be a response of the central bank to new information about inflation. Since the policy response is likely only to partially offset the inflationary pressure, the finding that a policy tightening is followed by rising

[1] Leeper, Sims, and Zha (1996) increase the number of variables included by applying Bayesian priors, but their VAR systems still typically contain less than 20 variables.

prices is explained. Of course, if Sims' explanation of the price puzzle is correct, then all the measured responses of economic variables to the monetary policy innovation are incorrectly measured, not just the price response.

A second problem arising from the use of sparse information sets in VAR analyses of monetary policy is that impulse responses can be observed only for the included variables, which generally constitute only a small fraction of the variables that the researcher and policymakers care about. For example, both for policy analysis and model validation purposes, we may be interested in the effects of monetary policy shocks on variables such as total factor productivity, real wages, profits, investment, and many others. Another reason to be interested in the responses of many variables is that no single time series may correspond precisely to a particular theoretical construct. The concept of "economic activity", for example, may not be perfectly represented by industrial production or real GDP. To assess the effects of a policy change on "economic activity", therefore, one might wish to observe the responses of multiple indicators including, say, employment and sales, to the policy change.[2] Unfortunately, as we have already noted, inclusion of additional variables in standard VARs is severely limited by degrees-of-freedom problems.

Is it possible to condition VAR analyses of monetary policy on richer information sets, without giving up the statistical advantages of restricting the analysis to a small number of series? In this paper we consider one approach to this problem, which combines the standard VAR analysis with factor analysis[3]. Recent research in dynamic factor models suggests that the information from a large number of time series can be

---

[2] An alternative is to treat "economic activity" as an unobserved factor with multiple observable indicators. That is essentially the approach we take in this paper.

usefully summarized by a relatively small number of estimated indexes, or factors. For example, Stock and Watson (2002) develop an approximate dynamic factor model to summarize the information in large data sets for forecasting purposes.[4] They show that forecasts based on these factors outperform univariate autoregressions, small vector autoregressions, and leading indicator models in simulated forecasting exercises. Bernanke and Boivin (2003) show that the use of estimated factors can improve the estimation of the Fed's policy reaction function.

If a small number of estimated factors effectively summarize large amounts of information about the economy, then a natural solution to the degrees-of-freedom problem in VAR analyses is to augment standard VARs with estimated factors. In this paper we consider the estimation and properties of factor-augmented vector autoregressive models (FAVARs), then apply these models to the monetary policy issues raised above.

The rest of the paper is organized as follows. Section 2 lays out the theory and estimation of FAVARs. We consider both a two-step estimation method, in which the factors are estimated by principal components prior to the estimation of the factor-augmented VAR; and a one-step method, which makes use of Bayesian likelihood methods and Gibbs sampling to estimate the factors and the FAVAR simultaneously. Section 3 applies the FAVAR methodology and revisits the evidence on the effect of monetary policy on wide range of key macroeconomic indicators. In brief, we find that the FAVAR methodology gives broadly plausible estimates for the responses of a wide

---

[3] Giannone, Reichlin and Sala (2002) consider a related approach.
[4] In this paper we follow the Stock and Watson approach to the estimation of factors (which they call "diffusion indexes"). We also employ a likelihood-based approach not used by Stock and Watson. Sargent

variety of macroeconomic variables to monetary policy shocks. We also find that the advantages of using the computationally more burdensome Gibbs sampling procedure instead of the two-step method appear to be modest in this application. Section 4 concludes. An appendix provides more detail concerning the application of the Gibbs sampling procedure to FAVAR estimation.

2. Econometric framework and estimation

Let Y_t be an $M \times 1$ vector of observable economic variables whose dynamic properties we want to estimate, where time $t = 1, 2, \ldots, T$. For now, we do not need to specify whether our ultimate interest is in forecasting the Y_t or in uncovering structural relationships among these variables. Following the standard approach, we might proceed by estimating a VAR, a structural VAR (SVAR), or other multivariate time series model using data for the Y_t alone. However, in many applications, additional economic information, not fully captured by the Y_t, may be relevant to modeling the dynamics of these series. Let us suppose that this additional information can be summarized by an $K \times 1$ vector of unobserved factors, F_t, where K is "small". We might think of the factors as diffuse concepts such as "economic activity" or "credit conditions" that cannot easily be represented by one or two series but rather are reflected in a wide range of economic variables. Assume that the joint dynamics of (F_t, Y_t) are given by:

---

and Sims (1977) first provided a dynamic generalization of classical factor analysis. Forni and Reichlin (1996, 1998) and Forni, Hallin, Lippi, and Reichlin (2000) develop a related approach.

$$(2.1) \qquad \begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + v_t$$

where $\Phi(L)$ is a conformable lag polynomial of finite order d, which may contain a priori restrictions as in the structural VAR literature. The error term v_t is mean zero with covariance matrix Q.

Equation (2.1) is a VAR in (Y_t, F_t). This system reduces to a standard VAR in Y_t if the terms of $\Phi(L)$ that relate Y_t to F_{t-1} are all zero; otherwise, we will refer to equation (2.1) as a *factor-augmented vector autoregression*, or FAVAR. If the true system is a FAVAR, note that estimation of (2.1) as a standard VAR system in Y_t—that is, with the factors omitted—will in general lead to biased estimates of the VAR coefficients and related quantities of interest, such as impulse response coefficients.

Equation (2.1) cannot be estimated directly because the factors F_t are unobservable. However, if we interpret the factors as representing forces that potentially affect many economic variables, we may hope to infer something about the factors from observations on a variety of economic time series. For concreteness, suppose that we have available a number of background, or "informational" time series, collectively denoted by the $N \times 1$ vector X_t. The number of informational time series N is "large" (in particular, N may be greater than T, the number of time periods) and will be assumed to be much greater than the number of unobserved factors ($K << N$). We assume that the informational time series X_t are related to the factors F_t and the observable data Y_t by:

$$(2.2) \qquad X_t = \Lambda^f F_t + \Lambda^y Y_t + e_t$$

where Λ^f is an $N \times K$ matrix of factor loadings, Λ^y is $N \times M$, and the $N \times 1$ vector of error terms e_t are mean zero and will be assumed either weakly correlated or uncorrelated, depending on whether estimation is by principal components or likelihood methods (see below). Equation (2.2) captures the idea that the X_t are noisy measures of the underlying factors F_t and also permits the X_t to depend contemporaneously on the Y_t. Controlling for the Y_t in equation (2.2) has the advantage of forcing the factors F_t to span a space orthogonal to the Y_t, so that we may interpret the factors as reflecting information not otherwise accounted for by the standard VAR in Y_t only. The implication of equation (2.2) that X_t depends only on the contemporaneous factors F_t is not restrictive in practice, as F_t can be interpreted as including arbitrary lags of the fundamental factors; thus, Stock and Watson (1998) refer to equation (2.2) as a *dynamic factor model.*

In this paper we consider two approaches to estimating (2.1)-(2.2), a two-step approach based on principal components and a single-step Bayesian likelihood approach. A clear advantage of the former approach is computational simplicity. However, this approach does not exploit the structure of the transition equation in the estimation of the factors. Whether or not this is a disadvantage depends on how well specified the model is, and from a comparison of the results from the two methods we may be able to assess whether the advantages of jointly estimating the model are worth the computational costs.

The two-step procedure is analogous to that used in the forecasting exercises of Stock and Watson. First, the information set X_t is used to estimate time series of the factors, $\hat{F}_t$, $t = 1,\ldots,T$.[5] As shown in Stock and Watson (2002), when T and N are large, the factors are consistently estimated as the first K principal components of X_t.[6] In the second step, the FAVAR, equation (2.1), is estimated by standard methods, with F_t replaced by $\hat{F}_t$. As noted, this procedure has the advantages of being computationally simple and easy to implement. As discussed by Stock and Watson, it also imposes few distributional assumptions and allows for some degree of cross-correlation in the idiosyncratic error terms e_t. However, the two-step approach implies the presence of "generated regressors" in the second step. To obtain accurate confidence intervals on the impulse response functions reported below, we implement a bootstrap procedure, based on Kilian (1998), that accounts for the uncertainty in the factor estimation. The drawback of this additional step is that it reduces somewhat the computational advantage of the two-step approach.

In principle, an alternative is to estimate (2.1) and (2.2) jointly by maximum likelihood. However, for very large dimensional models of the sort considered here, the irregular nature of the likelihood function makes MLE estimation infeasible in practice. In this paper we thus consider the joint estimation by likelihood-based Gibbs sampling techniques, developed by Geman and Geman (1984), Gelman and Rubin (1992), Carter and Kohn (1994) and surveyed in Kim and Nelson (1999). Their application to large

[5] A useful feature of this framework, as implemented by an EM algorithm, is that it permits one to deal systematically with data irregularities. In their application, Bernanke and Boivin (2003) estimate factors in cases in which X includes both monthly and quarterly series, series that are introduced mid-sample or are discontinued, and series with missing values.

dynamic factor models is discussed in Eliasz (2002). Kose, Otrok and Whiteman (2003) use similar methodology to study international business cycles. The Gibbs sampling approach provides empirical approximation of the marginal posterior densities of the factors and parameters via an iterative sampling procedure. As discussed in the Appendix, we implement a multi-move version of the Gibbs sampler in which factors are sampled conditional on the most recent draws of the model parameters, and then the parameters are sampled conditional on the most recent draws of the factors. As the statistical literature has shown, this Bayesian approach, by approximating marginal likelihoods by empirical densities, helps to circumvent the high-dimensionality problem of the model. Moreover, the Gibbs-sampling algorithm is guaranteed to trace the shape of the joint likelihood, even if the likelihood is irregular and complicated.

Identification

Before proceeding, we need to discuss identification of the model (2.1) – (2.2), specifically the restrictions necessary to identify uniquely the factors and the associated loadings. In two-step estimation by principal components, the factors are obtained entirely from the observation equation (2.2), and identification of the factors is standard. In this case we can choose either to restrict loadings by $\Lambda'\Lambda / N = I$ or restrict the factors by $F'F / T = I$. Either approach delivers the same common component $F\Lambda'$ and the same factor space. Here we impose the factor restriction, obtaining $\hat{F} = \sqrt{T}\hat{V}$, where the $\hat{V}$ are the eigenvectors corresponding to the K largest eigenvalues of XX', sorted in descending order. This approach identifies the factors against any rotations.

[6] Bai (2002) derives the asymptotic distribution of the estimates of the factors.

In the "one-step" (joint estimation) likelihood method, implemented by Gibbs sampling, the factors are effectively identified by both the observation equation (2.2) and the transition equation (2.1). In this case, ensuring identification also requires that we identify the factors F_t against rotations of the form $F_t^* = AF_t - BY_t$, where A is $K \times K$ and nonsingular, and B is $K \times M$. We prefer not to restrict the VAR dynamics described by equation (2.1), and so we need to impose restrictions in the observation equation, (2.2). Substituting for F_t in (2.2) we obtain

$$(2.3) \qquad X_t = \Lambda^f A^{-1} F_t^* + (\Lambda^y + \Lambda^f A^{-1} B) Y_t + e_t$$

Hence unique identification of the factors and their loadings requires $\Lambda^f A^{-1} = \Lambda^f$ and $\Lambda^y + \Lambda^f A^{-1} B = \Lambda^y$. Sufficient conditions are to set the upper $K \times K$ block of Λ^f to an identity matrix and the upper $K \times M$ block of Λ^y to zero. Since factors are only estimated up to a rotation, the choice of the block to set equal to an identity matrix does not affect the space spanned by the estimated factors. It does restrict, however, the contemporaneous impact of Y_t on those K variables and therefore such variables should be chosen for that block that do not respond contemporaneously to innovations in Y_t. The key to identification here is to make an assumption that restricts the channels by which the Y's contemporaneously affect the X's.

A separate identification issue concerns the identification of innovations in the VAR part of the model, such as identifying monetary policy innovations which is the subject of the next section. Importantly, FAVAR approach affords flexibility in identifying innovations - once factors are estimated standard procedures (e.g., structural VAR procedures as in Bernanke and Mihov, 1998) can be applied. One caveat is a time constraint when using Gibbs sampling methodology. For example, if we impose restrictions that overidentify the transition equation, we need to perform numerical optimization at each step of the Gibbs sampling procedure. This may easily become excessively time consuming. In part for computational simplicity we use a simple recursive ordering in our empirical application below.

3. Application: The dynamic effects of monetary policy

As discussed in the Introduction, an extensive literature has employed VARs to study the dynamic effects of innovations to monetary policy on a variety of economic variables. A variety of identification schemes have been employed, including simple recursive frameworks, "contemporaneous" restrictions (on the matrix relating structural shocks to VAR disturbances), "long-run" restrictions (on the shape of impulse responses at long horizons), and mixtures of contemporaneous and long-run restrictions.[7] Alternative estimation procedures have been employed as well, including Bayesian approaches (Leeper, Sims, and Zha, 1996). However, the basic idea in virtually all cases is to identify "shocks" to monetary policy with the estimated innovations to a variable or linear combination of variables in the VAR. Once this identification is made, estimating

[7] Recursive frameworks are employed, inter alia, in Bernanke and Blinder (1992), Sims (1992), Strongin (1995), and Christiano, Eichenbaum, and Evans (2000). Examples of papers with contemporaneous, non-recursive restrictions are Gordon and Leeper (1994), Leeper, Sims, and Zha (1996), and Bernanke and Mihov (1998a). Long-run restrictions are employed by Lastrapes and Selgin (1995) and Gerlach and Smets (1995). Gali (1992) and Bernanke and Mihov (1998b) use a mixture of contemporaneous and long-run restrictions. Faust and Leeper (1997) and Pagan and Robertson (1998) point out some dangers of relying too heavily on long-run restrictions for identification in VARs.

dynamic responses to monetary policy innovations (as measured by impulse response functions) is straightforward.

The fact that this simple method typically gives plausible and useful results with minimal identifying assumptions accounts for its extensive application, both by academic researchers and by practitioners in central banks. Nevertheless, a number of critiques of the approach have been made (see, for example, Rudebusch, 1998). Here we focus on two issues, both related to the fact that degrees-of-freedom problems necessarily limit the number of time series that can be included in an estimated VAR. We then evaluate the ability of FAVARs—which, potentially, can include much more information than standard VARs—to ameliorate these problems.

First, as emphasized by Bernanke and Boivin (2003), central banks routinely monitor a large number of economic variables. One rationale for this practice is that many variables may contain information that is useful in predicting output, inflation, and other variables which enter into the central bank's objective function (Stock and Watson, 1999, forthcoming; Kozicki, 2001). Standard VARs of necessity include only a relatively small number of time series, implying that the information set employed by the econometrician differs from (is a subset of) that of the monetary policy-makers. To the extent that policy-makers react to variables not included in the VAR, monetary policy "shocks" and the implied dynamic responses of the economy will be mismeasured by the econometrician.[8] A possible example of the effects of shock mismeasurement is the

[8] Another source of mismeasurement arises from the fact that most VAR studies typically use revised, as opposed to "real-time" data. Croushore and Evans (1999) do not find this issue to be important for the identification of monetary policy shocks, a view consistent with evidence presented in a forecasting context by Bernanke and Boivin (2003). However, Orphanides (2001) argues that assessment of Fed policy depends sensitively on whether revised or real-time data are used.

"price puzzle" discussed in the Introduction. We will check below whether including broader information set ameliorates the price puzzle.

Even if monetary policy shocks are properly identified, standard VAR analyses have the shortcoming that the dynamic responses of only those few variables included in the estimated VAR can be observed. As discussed in the Introduction, this limitation may be a problem for at least two reasons. First, for purposes both of policy analysis and model validation, it is often useful to know the effects of monetary policy on a lengthy list of variables.[9] Second, the choice of a specific data series to represent a general economic concept (e.g., industrial production for "economic activity", the consumer price index for "the price level") is often arbitrary to some degree, and estimation results may depend on idiosyncratic features of the particular variable chosen. To assess the effects of monetary policy on a concept like "economic activity", it is of interest to observe the responses of a variety of indicators of activity, not only one or two.

The FAVAR framework is well-suited for addressing both issues. First, the estimated system (2.1)-(2.2) can be used to draw out the dynamic responses of not only the "main" variables Y_t but of any series contained in X_t. Hence the "reasonableness" of a particular identification can be checked against the behavior of many variables, not just three or four. Second, one might also consider constructing the impulse response functions of factors (or linear combinations of the factors) that can be shown to stand in for a broad concept like "economic activity."

[9] One approach to this problem is to assume no feedback from variables outside the basic VAR, that is, a block-recursive structure with the base VAR ordered first (see Bernanke and Gertler, 1995). However, the no-feedback assumption is dubious in many cases.

Empirical results

We applied both the two-step and "one-step" (joint estimation) methodologies to the estimation of monetary FAVARs. In our applications, X_t consists of a balanced panel of 120 monthly macroeconomic time series (updates of series used in Stock and Watson, 1998 and 1999). For the baseline analysis, we assume that the Fed funds rate is the only observable factor, i.e. the only variable included in Y_t. In doing so, we interpret the Fed funds rate as the monetary policy instrument, which is assumed to have pervasive effect on the economy, X_t. The latent factors are then understood as capturing real activity and general price movements. The key advantage of this specification is that we do not have to take a stand on what is the appropriate measure of the real-activity or inflation.

We order the federal funds rate last and treat its innovations as monetary policy "shocks", in the standard way. This ordering imposes the identifying assumption that output and prices do not respond to monetary policy innovations within the month. This assumption may be subject to criticism if there are "fast-moving" components of the estimated factors, not accounted for by the federal funds rate, that nevertheless respond contemporaneously to interest rate shocks.[10] To address this potential problem we have considered a specification in which we divided our dataset into "slow" and "fast" blocks, the latter including series from stock market, money and credit, interest rates and exchange rates sectors and consumer expectations. Next we extracted "slow-moving" and "fast-moving" factors from the respective blocks of data and ordered "fast-moving" factors on top of the federal funds rate in the VAR ordering. This produced "fast-moving" factors that very closely follow interest rate movements and consequently introduce collinearity in the system. We interpret results of this exercise as suggesting that there is little informational content in the "fast-moving" factors that is not already accounted for by the federal funds rate. We therefore adhere to our original formulation. The data we use span the period January 1959 through August 2001.

Our main results are shown in Figures 1-4 below. Each Figure shows impulse responses with 90% confidence intervals of a selection of key macroeconomic variables to a monetary policy shock. Figures 1 and 2 show the results for the FAVAR model with 3 latent factors, estimated by principal components and likelihood methods, respectively. We used 13 lags but employing 7 lags led to very similar results as found with the greater number of lags. Likelihood-based estimates employed 5,000 iterations of the Gibbs sampling procedure (of which the first 1,000 were discarded to minimize the effects of initial conditions). There seemed to be no problems achieving convergence, and alternative starting values or the use of 10,000 iterations gave essentially the same results. We standardize the monetary shock to correspond to a 25-basis-point innovation in the federal funds rate[11].

An important practical question is how many factors are needed to capture the information necessary to properly model the effect of monetary policy. Bai and Ng (2002) provide a criterion to determine the number of factors present in the data set, X_t. However, this does not address the question of how many factors should be included in the VAR and due to computational constraint cannot be readily implemented in the

[10] A slow-moving variable is one that is largely predetermined as of the current period, while a fast-moving variable—think of an asset-price factor—is highly sensitive to contemporaneous economic news or shocks.
[11] Note that the figures report the response of the standardized Fed funds rate. So the policy shock we consider is actually 0.25 divided by the standard deviation of the Fed funds rate.

likelihood-based estimation. To explore the effect of increasing the number of factors, we thus consider an alternative specification with 5 latent factors. The results are reported in Figures 3 and 4. Increasing the number of factors beyond this did not change qualitative nature of our results.

As we have discussed, an advantage of the FAVAR approach is that impulse response functions can be constructed for any variable in the informational data set, that is, for any element of X_t. This gives both more information and provides a check on the empirical plausibility of the specification. In that respect, the most successful specification, in terms of plausibility, appears to be the two-step principal component approach with 5 factors, reported in Figure 3. The responses are generally of the expected sign and magnitude: following a contractionary monetary policy shock, real activity measures decline, prices eventually go down and money aggregates decline. The response of the stock market is small and insignificant. Overall these results seem to provide a consistent and sensible measure of the effect of monetary policy. Note that we display only 20 responses of all 120 that in principle could be investigated.

The FAVAR model appears successful in capturing relevant information. First, the price-puzzle is not present in our FAVAR model estimated by two-step approach, even when only three factors are included. Given that our recursive identification of the policy shocks is consistent with existing structural VARs that display the price-puzzle, our result might suggest that a few factors are sufficient to properly capture the information that Sims argued might be missing from these VARs. Second, increasing the number of factors generally tends to produce results more consistent with conventional wisdom. This is particularly obvious when comparing the response of money aggregates for the 2-step approach in Figure 1 and 3: the apparent liquidity puzzle in Figures 1 disappears when more factors are included. For the likelihood based estimation, the liquidity puzzle remains even with 5 factors. However, in this case the price-puzzle seems mitigated. The amount of information included in the empirical analysis is thus crucial to yield a plausible picture of the effects of monetary policy, and the FAVAR approach shows some success at exploiting this information.

But, as is obvious from the likelihood-based results, information is not all the story. Responses of prices and money aggregates in this case might suggest in fact that the policy shock has not been properly identified. This is a possibility that would be worth considering in future research. It is important to stress, however, that although we considered a recursive identification of the policy shock, there is nothing in our proposed approach – other than the computational constrains mentioned above – that prevents using alternative, non-recursive, identification schemes. However, the fact that the two-step approach is relatively successful with the same identification scheme might suggest that the likelihood-based estimation suffers from the additional structure it imposes, which might not be entirely supported empirically.

While the two methods yield somewhat different responses for money aggregates and CPI, overall the point estimates of the response are quite similar. We find it remarkable that the two rather different methods give such similar results. On the other hand, the degree of uncertainty about the estimates implied by the two methods are quite different: as is apparent from the figures, the two-step approach yields much wider confidence intervals. Most of the difference comes in fact from the factor estimation step, and our implementation of the bootstrap might overstate the true degree of uncertainty. In any case, one would expect the joint likelihood-based estimation approach to produce tighter estimates.

4. Conclusion

This paper has introduced a method for incorporating a broad range of conditioning information, summarized by a small number of factors, in otherwise standard VAR analyses. We have shown how to identify and estimate a factor-augmented vector autoregression, or FAVAR, by both a two-step method based on estimation of principal components and a more computationally demanding, Bayesian method based on Gibbs sampling. Another key advantage of this method, is to obtain the responses of a large set of variables to monetary policy, which allows both broader picture of its effect and a more complete check on the empirical plausibility of the specification.

In our monetary application of FAVAR methods, we find that overall the two methods produce qualitatively similar results, although the two-step approach tends to produce more plausible responses. Moreover, the results provide some support for the view that the "price puzzle" results from the exclusion of conditioning information. The conditioning information is also instrumental to produce reasonable responses of money aggregates. These results thus suggest that there is a scope to exploit more information in empirical macroeconomic modeling.

Future work should investigate more fully the properties of FAVARs, alternative estimation methods and alternative identification scheme. 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. For example, according to the original Sims (1992) hypothesis, if addition of factors mitigates the price puzzle, then the factors should contain information about future inflation not otherwise captured in the VAR. The marginal contribution of the estimated factors for forecasting inflation can be checked directly.[12]

[12] Stock and Watson (1999) and Bernanke and Boivin (2003) have shown that, generally, factor methods are useful for forecasting inflation.

Appendix: Estimation by Likelihood-Based Gibbs Sampling

This appendix discusses the estimation of FAVARs by likelihood-based Gibbs sampling. For further details see Eliasz (2002).

To estimate equations (2.1) and (2.2) jointly via likelihood methods, we transform the model into the following state-space form:

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \Lambda^f & \Lambda^y \\ 0 & I \end{bmatrix} \begin{bmatrix} F_t \\ Y_t \end{bmatrix} + \begin{bmatrix} e_t \\ 0 \end{bmatrix} \tag{A.1}$$

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + v_t \tag{A.2}$$

where Y_t is an $M \times 1$ vector of observable economic variables in whose dynamic properties we are interested, F_t is an $K \times 1$ vector of unobserved factors, and X_t is an $N \times 1$ vector of time series that incorporate information about the unobserved factors, all as described in the text. Time is indexed $t = 1,2,\dots,T$. The coefficient matrices Λ^f and Λ^y are $N \times K$ and $N \times M$, respectively, and $\Phi(L)$ is a conformable lag polynomial of finite order d. The loadings Λ^f and Λ^y are restricted as discussed in the text. The error vectors e_t and v_t are $N \times 1$ and $(K+M) \times 1$, respectively, and are assumed to be distributed according to $e_t \sim N(0,R)$ and $v_t \sim N(0,Q)$, with e_t and v_t independent and R diagonal.

(A.1) is the measurement or observation equation, and (A.2), which is identical to (2.1), is the transition equation. Inclusion of Y_t in the measurement equation (A.1) as well as in the transition equation (A.2) does not change the model but allows for both notational and computational simplification.

We take a Bayesian perspective, treating the model's parameters $\theta = (\Lambda^f, \Lambda^y, R, vec(\Phi), Q)$ as random variables; and where we define $vec(\Phi)$ as a column vector of the elements of the stacked matrix Φ of the parameters of the lag operator $\Phi(L)$. Likelihood estimation by multi-move Gibbs sampling (Carter and Kohn, 1994), proceeds by alternately sampling the parameters θ and the unobserved factors F_t. To be more specific, define $\mathbf{X}_t' = (X_t', Y_t')'$, $\mathbf{e}_t' = (e_t', 0)'$, and $\mathbf{F}_t' = (F_t', Y_t')'$ and rewrite the measurement and transition equations (A.1) and (A.2) as

(A.3) $$\mathbf{X}_t = \Lambda \mathbf{F}_t + \mathbf{e}_t$$

(A.4) $$\mathbf{F}_t = \Phi(L)\mathbf{F}_{t-1} + v_t$$

where Λ is the loading matrix from (A.1) and $\mathbf{R} = \text{cov}(\mathbf{e}_t \mathbf{e}_t')$ is the covariance matrix R augmented by zeros in the obvious way. For this exposition we assume that the order d of $\Phi(L)$ equals one, otherwise we would rewrite (A.4) in a standard way to express it as a first-order Markov process (see Eliasz, 2002). Further, let $\widetilde{\mathbf{X}}_T = (\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_T)$ be the history of $\mathbf{X}$ from period 1 through period *T*, and likewise define $\widetilde{\mathbf{F}}_T = (\mathbf{F}_1, \mathbf{F}_2, ..., \mathbf{F}_T)$.

Our problem is to characterize the marginal posterior densities of $\widetilde{\mathbf{F}}_T$ and θ, respectively $p(\widetilde{\mathbf{F}}_T) = \int p(\widetilde{\mathbf{F}}_T, \theta) d\theta$ and $p(\theta) = \int p(\widetilde{\mathbf{F}}_T, \theta) d\widetilde{\mathbf{F}}_T$, where $p(\widetilde{\mathbf{F}}_T, \theta)$ is the joint posterior density and the integrals are taken with respect to the supports of θ and

$\widetilde{\mathbf{F}}_T$, respectively. Given these marginal posterior densities, estimates of $\widetilde{\mathbf{F}}_T$ and θ can be obtained as the medians or means of these densities.

To obtain empirical approximations to these densities, we follow Kim and Nelson (1999, chapter 8) and apply multi-move Gibbs sampling to the state-space model (A.3)-(A.4). The Gibbs sampling methodology proceeds as follows: First, choose a set of starting values for the parameters θ, say θ^0. Second, conditional on θ^0 and the data $\widetilde{\mathbf{X}}_T$, draw a set of values for $\widetilde{\mathbf{F}}_T$, say $\widetilde{\mathbf{F}}_T^1$ from the conditional density $p(\widetilde{\mathbf{F}}_T|\widetilde{\mathbf{X}}_T,\theta^0)$. Third, conditional on the sampled values of $\widetilde{\mathbf{F}}_T$ and the data, draw a set of values of the parameters θ, say θ^1, from the conditional distribution $p(\theta|\widetilde{\mathbf{X}}_T,\widetilde{\mathbf{F}}_T^1)$. The final two steps constitute one iteration, and are repeated until the empirical distributions of $\widetilde{\mathbf{F}}_T^s$ and θ^s converge, where s indexes the iteration. It has been shown (Geman and Geman, 1994), that as the number of iterations $s \to \infty$, the marginal and joint distributions of the sampled values of $\widetilde{\mathbf{F}}_T^s$ and θ^s converge to the true corresponding distributions at an exponential rate. In practice, though, convergence can be slow and should be carefully checked, for example by using alternative starting values. More details on each step are given below.

1. Choice of θ^0

In general, it is good practice to try a variety of starting parameter values to see if they generate similar empirical distributions. As Gelman and Rubin (1992) argue, a single sequence from the Gibbs sampler, even if it has apparently converged, may give a “false sense of security”. At the same time, in a problem as large as the one at hand, for which computational capacity constrains the number of feasible runs, a meaningful choice of θ^0 may be advisable. An obvious choice was to use parameter estimates obtained from principal components estimation of (A.1) and the vector autoregression (A.2). We constrained these parameter estimates to satisfy the normalization, discussed in the text, that the upper $K \times (K+M)$ block of loadings Λ is restricted to be $[\mathbf{I}_K, \mathbf{0}_{K\times M}]$. We used these parameter estimates as starting values for θ in most runs, but we have confirmed the robustness of the key results for alternative starting values. For example, we also tried starting values such that (1) $vec(\Phi) = \mathbf{0}$, (2) $Q = \mathbf{I}$, (3) $\Lambda^f = \mathbf{0}$, (4) Λ^y = OLS estimates from the regression of X on Y, and (5) R = residual covariance matrix from the regression of X on Y, and obtained similar results to those reported in the text.

2. Drawing from the conditional distribution $p(\widetilde{\mathbf{F}}_T | \widetilde{\mathbf{X}}_T, \theta)$

As in Nelson and Kim (p. 191), the conditional distribution of the whole history of factors $p(\widetilde{\mathbf{F}}_T | \widetilde{\mathbf{X}}_T, \theta)$ can be expressed as the product of conditional distributions of factors at each date t as follows:

$$\text{(A.5)} \qquad p(\widetilde{\mathbf{F}}_T | \widetilde{\mathbf{X}}_T, \theta) = p(\mathbf{F}_T | \widetilde{\mathbf{X}}_T, \theta) \prod_{t=1}^{T-1} p(\mathbf{F}_t | \mathbf{F}_{t+1}, \widetilde{\mathbf{X}}_t, \theta)$$

where $\widetilde{\mathbf{X}}_t = (\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_t)$. (A.5) relies on the Markov property of $\mathbf{F}_t$, which implies that $p(\mathbf{F}_t | \mathbf{F}_{t+1}, \mathbf{F}_{t+2}, ..., \mathbf{F}_T, \mathbf{X}_T, \theta) = p(\mathbf{F}_t | \mathbf{F}_{t+1}, \mathbf{X}_t, \theta)$.

Because the state-space model (A.3)-(A.4) is linear and Gaussian, we have

$$\text{(A.6)} \qquad \begin{aligned} &\mathbf{F}_T | \widetilde{\mathbf{X}}_T, \theta \sim N(\mathbf{F}_{T|T}, \mathbf{P}_{T|T}) \\ &\mathbf{F}_t | \mathbf{F}_{t+1}, \widetilde{\mathbf{X}}_t, \theta \sim N(\mathbf{F}_{t|t,\mathbf{F}_{t+1}}, \mathbf{P}_{t|t,\mathbf{F}_{t+1}}) \qquad t = T-1, ..., 1 \end{aligned}$$

where

$$\text{(A.7)} \qquad \begin{aligned} \mathbf{F}_{T|T} &= E(\mathbf{F}_T | \widetilde{\mathbf{X}}_T, \theta) \\ \mathbf{P}_{T|T} &= Cov(\mathbf{F}_T | \widetilde{\mathbf{X}}_T, \theta) \\ \mathbf{F}_{t|t,\mathbf{F}_{t+1}} &= E(\mathbf{F}_t | \mathbf{F}_{t+1}, \widetilde{\mathbf{X}}_t, \theta) = E(\mathbf{F}_t | \mathbf{F}_{t+1}, \mathbf{F}_{t|t}, \theta) \\ \mathbf{P}_{t|t,\mathbf{F}_{t+1}} &= Cov(\mathbf{F}_t | \mathbf{F}_{t+1}, \widetilde{\mathbf{X}}_t, \theta) = Cov(\mathbf{F}_t | \mathbf{F}_{t+1}, \mathbf{F}_{t|t}, \theta) \end{aligned}$$

where the notation $\mathbf{F}_{t|t}$ refers to the expectation of $\mathbf{F}_t$ conditional on information dated t or earlier. To obtain these, we first calculate $\mathbf{F}_{t|t}$ and $\mathbf{P}_{t|t}$, $t = 1,2,...T$, by Kalman filter, conditional on θ and the data through period t, $\widetilde{\mathbf{X}}_t$, with starting values of zeros for the factors and the identity matrix for the covariance matrix. (Hamilton, 1994). The last iteration of the filter yields $\mathbf{F}_{T|T}$ and $\mathbf{P}_{T|T}$, which together with the first line of (A.6) allows us to draw a value for $\mathbf{F}_T$. Treating this drawn value as extra information, we can move "backwards in time" through the sample, using the Kalman filter to obtain updated values of $\mathbf{F}_{T-1|T-1,\mathbf{F}_T}$ and $\mathbf{P}_{T-1|T-1,\mathbf{F}_T}$; drawing a value of $\mathbf{F}_{T-1}$ using the second line of (A.6); and continuing in similar manner to draw values for $\mathbf{F}_t$, $t = T-2, T-3, \ldots, 1$.

If the order d of $\Phi(L)$ exceeds one, as it does in our applications, then lags of the factors appear in the state vector $\mathbf{F}_t$ and Q is singular, as is $\mathbf{P}_{t|t,\mathbf{F}_{t+1}}$ for any $t < T$. (The singularity of these two covariance matrices follows from the fact that, in this case, $\mathbf{F}_t$ and $\mathbf{F}_{t+1}$ have common components.) In this case we cannot condition on the full vector $\mathbf{F}_{t+1}$ when drawing $\mathbf{F}_t$, but only on the first d elements of $\mathbf{F}_{t+1}$. Kim and Nelson (1999, p. 194-6) show how to modify the Kalman filter algorithm in this case.

3. Drawing from the conditional distribution $p(\theta|\widetilde{\mathbf{X}}_T, \widetilde{\mathbf{F}}_T)$.

Conditional on the observed data and the estimated factors from the previous iteration, a new iteration is begun by drawing a new value of the parameters θ. With known factors, (A.3) and (A.4) amount to standard regression equations, with (A.3) specifying the distribution of Λ and R, and (A.4) the distribution of $vec(\Phi)$ and Q. Consider (A.3) first. Because the errors are uncorrelated, we can apply OLS to (A.3) equation by equation to obtain $\hat{\Lambda}$ and $\hat{e}$. Let $\hat{R}_{ii} = \hat{e}_i' \hat{e}_i / (T - (K_i + M_i))$, where $K_i + M_i$ is the number of regressors in equation i, and set $R_{ij} = 0, i \neq j$. If we assume an uninformative prior, by standard Bayesian results we have

$$R_{ii}|\widetilde{\mathbf{X}}_T, \widetilde{\mathbf{F}}_T \sim inverse\,\chi^2(T - K_i - M_i, \hat{R}_{ii})$$

Having drawn R_{ii}, we can then draw values for the coefficients of the i-th equation, Λ_i, from $N(\hat{\Lambda}_i, R_{ii}[\widetilde{\mathbf{F}}_T^{(i)\prime}\widetilde{\mathbf{F}}_T^{(i)}]^{-1})$, where $\widetilde{\mathbf{F}}_T^{(i)}$ corresponds to the regressors of the i-th equation.

Turning to (A.4), we see that this system has a standard VAR form and can thus can also be estimated equation by equation, to obtain $vec(\hat{\Phi}), \hat{Q}$. Imposing a flat prior on $\log|Q|$, we draw Q from the $InverseWishart((T-d)\hat{Q}^{-1}, T-(K+M)d-1)$. Conditional on the sampled Q, we then draw $\{\Phi^{ij\prime}\}$ from the conditional normal according to

$$vec(\Phi) \sim N(vec(\hat{\Phi}), Q * (\widetilde{\mathbf{F}}_T{}'\widetilde{\mathbf{F}}_T)^{-1})$$

where $vec(\Phi)$ is the rows of Φ stacked in a column vector of length $d(K+M)^2$ and "*" denotes the Kronecker product. This completes the sampling of the parameters θ conditional on the estimated factors from the previous iteration and the observed data.

Steps 2 and 3 are repeated for each iteration s. Inference is based on the distribution of $(\widetilde{\mathbf{F}}_T^s, \theta^s)$, for $s \geq B$, with B large enough to guarantee convergence of the algorithm. As noted, the empirical distribution from the sampling procedure should well approximate the joint posterior or normalized joint likelihood. Calculating medians and quantiles of $(\widetilde{\mathbf{F}}_T^s, \theta^s)$ for $s = B,...,S$ provides estimates of the values of the factors and the model parameters and the associated confidence regions. Note that the Gibbs-sampling algorithm is guaranteed to closely approximate the shape of the likelihood, especially around its peak, even if the likelihood is rather irregular and complicated, as is typically the case in the large models considered in this paper.

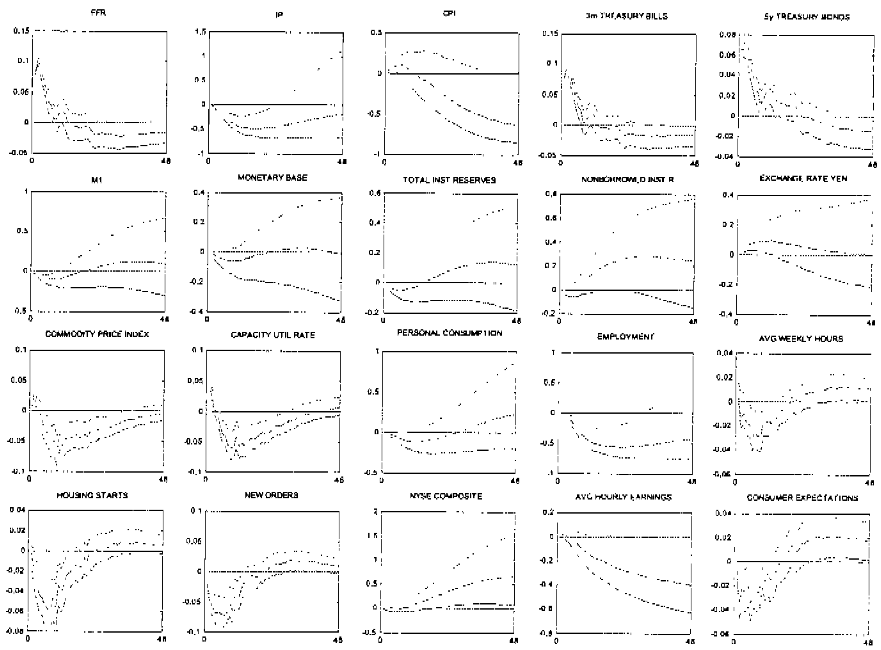

Figure 1. Impulse responses generated from FAVAR with 3 factors and FFR estimated by principal components with 2 step bootstrap.

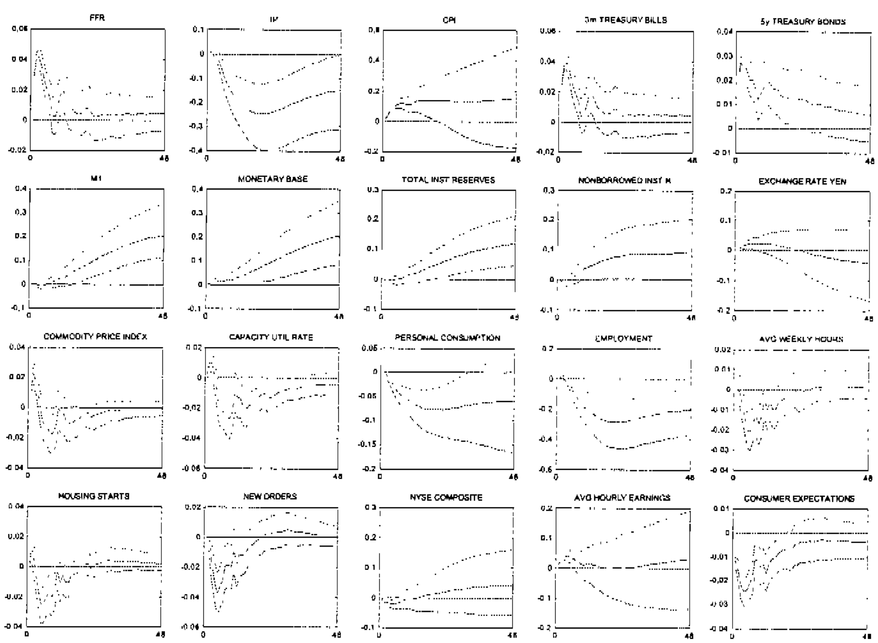

Figure 2. Impulse responses generated from FAVAR with 3 factors and FFR estimated by Gibbs sampling.

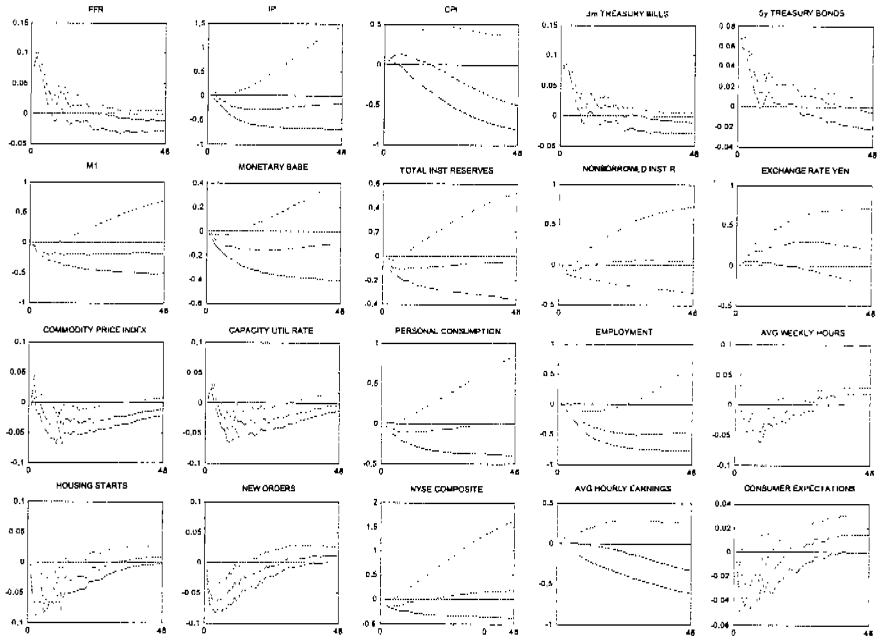

Figure 3. Impulse responses generated from FAVAR with 5 factors and FFR estimated by principal components with 2 step bootstrap.

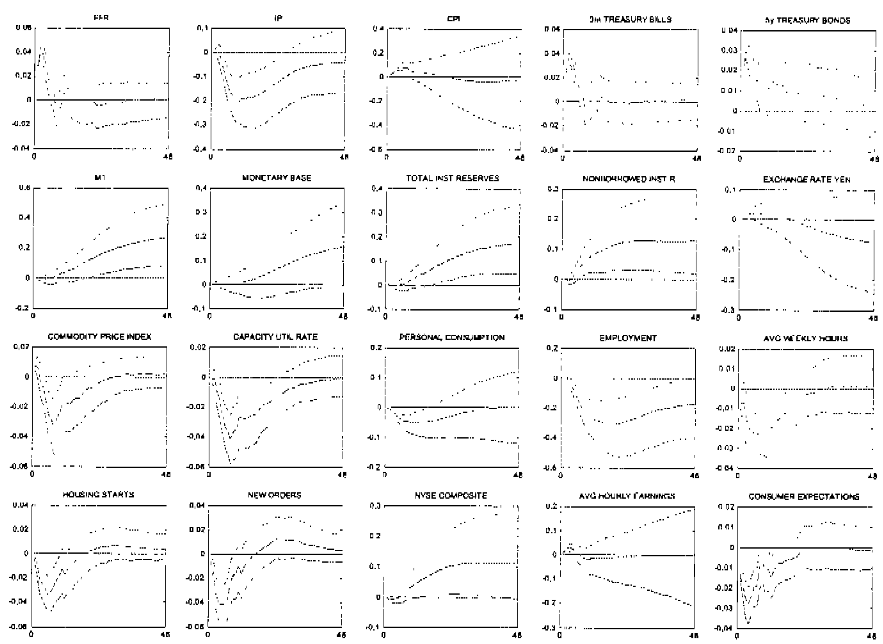

Figure 4. . Impulse responses generated from FAVAR with 5 factors and FFR estimated by Gibbs sampling.

References

Bai, Jushan, "Estimating Cross-Section Common Stochastic Trends in Non-Stationary Panel Data", Boston College, unpublished paper, 2002.

Balke, Nathan and Kenneth Emery, "Understanding the Price Puzzle", Federal Reserve Bank of Dallas, Economic Review, Fourth Quarter 1994, 15-26.

Bernanke, Ben and Alan Blinder, "The Federal Funds Rate and the Channels of Monetary Transmission", American Economic Review 82, September 1992, 901-21.

Bernanke, Ben and Jean Boivin, "Monetary Policy in a Data-Rich Environment", Journal of Monetary Economics 50:3, April 2003, 525-546.

Bernanke, Ben, and Mark Gertler, "Inside the Black Box: The Credit Channel of Monetary Transmission", Journal of Economic Perspectives 9,. Fall 1995, 27-48.

Bernanke, Ben, Mark Gertler, and Mark Watson, "Systematic Monetary Policy and the Effects of Oil Price Shocks", Brookings Papers on Economic Activity, 1997:1, 91-142.

Bernanke, Ben and Ilian Mihov, "Measuring Monetary Policy", Quarterly Journal of Economics 113, August 1998, 869-902.

Bernanke, Ben and Ilian Mihov, "The Liquidity Effect and Long-Run Neutrality", in B. McCallum and C. Plosser, eds., Carnegie Rochester Conference Series on Public Policy 49, 1998, 149-94.

Boivin, Jean, and Marc Giannoni, "Has Monetary Policy Become More Effective?" NBER Working Paper no. 9459, 2003.

Carter, C.K. and P. Kohn, "On Gibbs Sampling for State Space Models", Biometrika 81, 1994, 541-53.

Christiano, Lawrence, "Modeling the Liquidity Effect of a Money Shock", Federal Reserve Bank of Minneapolis, Quarterly Review 15, 1991, 3-34.

Christiano, Lawrence and Martin Eichenbaum, "Identification and the Liquidity Effect of a Monetary Policy Shock", in A. Cukierman, Z. Hercowitz, and L. Leiderman, eds., Political Economy, Growth, and Business Cycles, Cambridge: MIT Press, 1992, 335-370.

Christiano, Laurence, Martin Eichenbaum, and Charles Evans, "Monetary Policy Shocks: What Have We Learned and to What End?", in J. Taylor and M.Woodford, eds., Handbook of Macroeconomics, Amsterdam: North-Holland 2000.

Cochrane, John, "Shocks", Carnegie Rochester Conference on Public Policy, vol. 41, 1994, 295-34.

Cochrane, John, "What Do the VARs Mean? Measuring the Output Effects of Monetary Policy", unpublished, University of Chicago, August 1996.

Croushore, Dean, and Charles Evans, "Data Revisions and the Identification of Monetary Policy Shocks", manuscript, Federal Reserve Bank of Philadelphia, 1999.

Eichenbaum, Martin, "Comment on 'Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy'", European Economic Review 36, 1992, 1001-1011.

Eliasz, Piotr, "Likelihood-Based Inference in Large Dynamic Factor Models Using Gibbs Sampling," Princeton University, unpublished, 2002.

Faust, Jon and Eric Leeper, "When Do Long-Run Identifying Restrictions Give Reliable Results?", Journal of Business and Economic Statistics 15, 1997, 345-53.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin, "The Generalized Dynamic Factor Model: Identification and Estimation", Review of Economics and Statistics 82, November 2000, 540-54.

Forni, M. and L. Reichlin, "Dynamic Common Factors in Large Cross-Sections", Empirical Economics 21, 1998, 27-42.

Forni, M. and L. Reichlin, "Let's Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycles", Review of Economic Studies, 65, 1998, 453-74.

Gali, Jordi, "How Well Does the IS-LM Model Fit Postwar Data?", Quarterly Journal of Economics 107, 1992, 709-38.

Gelman, A. and D. B. Rubin, "A Single Sequence from the Gibbs Sampler Gives a False Sense of Security", in J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds., Bayesian Statistics, Oxford: Oxford University Press, 1992.

Geman, S. and D. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 1984, 721-41.

Gerlach, Stefan and Frank Smets, "The Monetary Transmission Mechanism: Evidence from the G-7 Countries", Centre for Economic Policy Research discussion paper 1219, 1995.

Giannone D., Reichlin L. and Sala L. (2002), "Tracking Greenspan: Systematic and Unsystematic Monetary Policy Revisited".

Gordon, David and Eric Leeper, "The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification", Journal of Political Economy 102, 1994, 1228-1247.

Hanson, Michael, "The 'Price Puzzle' Reconsidered", working paper, Wesleyan University, 2000.

Kim, Chang-Jin and Charles R. Nelson, State-Space Models with Regime Switching, Cambridge MA: MIT Press, 1999.

Kozicki, Sharon, "Why Do Central Banks Monitor So Many Inflation Indicators?", Federal Reserve Bank of Kansas City, Economic Review 86, Third Quarter 2001, 5-42.

Lastrapes, William and George Slegin, "The Liquidity Effect: Identifying Short-Run Interest Rate Dynamics Using Long-Run Restrictions", Journal of Macroeconomics 17, 1995, 387-404.

Leeper, Eric, Christopher Sims, and Tao Zha, "What Does Monetary Policy Do?", Brookings Papers on Economic Activity 1996: 2, 1-63.

Leeper, Eric and Tao Zha, "Assessing Simple Policy Rules: A View from a Complete Macroeconomic Model", Federal Reserve Bank of St. Louis, July/August 2001, 83-110.

Orphanides, Athanasios, "Monetary Policy Rules Based on Real-Time Data", American Economic Review 91, September 2001, 964-85.

Pagan, Adrian, and J. C. Robertson, "Structural Models of the Liquidity Effect", Review of Economics and Statistics 80, 1998, 202-17.

Quah, Danny and Thomas Sargent, "A Dynamic Index Model for Large Cross Sections", in J. Stock and M. Watson, eds., Business Cycles, Indicators, and Forecasting, Chicago: University of Chicago Press for NBER, 1993, 285-306

Rudebusch, Glenn, "Do Measures of Monetary Policy in a VAR Make Sense?", International Economic Review 39, November 1998, 907-31.

Sims, Christopher, "Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy", European Economic Review 36, 975-1000, 1992.

Sims, Christopher and Tao Zha, "Does Monetary Policy Generate Recessions?", Federal Reserve Bank of Atlanta, July 1998.

Stock, James, and Mark Watson, "New Indexes of Coincident and Leading Economic Indicators", NBER Macroeconomics Annual (4), Cambridge MA: MIT Press, 1989.

_______, "Diffusion Indexes", NBER working paper 6702, 1998.

_______, "Forecasting Inflation", Journal of Monetary Economics, 44, 1999, 293-335.

_______, "Macroeconomic Forecasting Using Diffusion Indexes", Journal of Business Economics and Statistics, 20:2, 2002, 147–162.

Strongin, Steven, "The Identification of Monetary Policy Disturbances: Explaining the Liquidity Puzzle", Journal of Monetary Economics 34, 1995, 463-97.