REDUCED-RANK IDENTIFICATION OF STRUCTURAL SHOCKS IN VARS

Yuriy Gorodnichenko

Department of Economics
University of Michigan

First draft: April 22, 2003
This version: June 17, 2004

This paper suggests a novel approach to identify structural shocks in vector autoregressions (VARs). Factor structure is imposed on observed VAR residuals to identify structural shocks thus linking VARs to business cycle models. Identification, estimation and testing procedures are developed. The method is applied to identify the effects of monetary policy (US data). Under this identification, the price level does not increase in response to restrictive policy shocks. Moreover, unlike most VAR estimates, the exchange rate appreciates on impact and then gradually depreciates. Hence, no price level and exchange rate puzzles are found. The paper also explains weak contemporaneous cross-variable impulse responses typically found in conventionally identified VARs.

Keywords: Vector autoregressions, identification, factor structure, monetary policy.
JEL classification: E520, C320

The author is grateful to Robert Barsky, Atsushi Inoue, Lutz Kilian, Serena Ng, and Matthew Shapiro for helpful comments and suggestions. The author thanks Lutz Kilian for providing his program codes. All remaining errors are the author's. Correspondence: ygorodni@umich.edu or Yuriy Gorodnichenko, Department of Economics, University of Michigan, 611 Tappan St, Ann Arbor, MI 48109.

1 Introduction

In vector autoregressions (VARs), whether structural or not structural, the number of structural shocks is equal to the number of variables. In contrast, typical business cycle (BC) models have only very few shocks. In addition, VARs normally find very small contemporaneous cross-variable responses while BC models usually predict sizeable contemporaneous cross-variable responses. This paper reconciles these two important differences between VAR and BC models by identifying and estimating a VAR with the number of structural shocks less than the number of variables in the VAR.

The key insight is to treat structural shocks as factors. The proposed method not only determines the number of structural shocks but also uniquely identifies these shocks using *short* run restrictions even when the number of the shocks is greater than one. Furthermore, all VAR tools are available for dynamic analyses of economic models under this identification method – for example, impulse response functions are easily constructed. Although the method is related to latent index models (e.g. Sargent and Sims, 1977), it is different from factor models in many respects. For example, latent index models can estimate only the number of factors and their total contribution to the variance of analyzed variables but, in general, they cannot be used for economic analyses because factors are not uniquely identified.

To illustrate the method, I apply it to the problem of identifying effects of monetary policy. Using US data, I construct impulse responses of key macroeconomic variables to restrictive monetary policy shocks. In contrast to commonly used VAR identification schemes, my identification leads to responses of the price level and exchange rate consistent with macroeconomic theory. Specifically, in response to policy surprises the price level does not rise on impact and then gradually falls and the exchange rate appreciates on impact and then gradually depreciates. In other words, I do not find the price level and exchange rate puzzles. In addition, under my identification the policy is a lot less noisy than under other identification schemes. I also provide an explanation of why recursive identification

schemes typically find weak contemporaneous responses of macroeconomic variables to innovations in monetary policy.

In the next section I briefly discuss conventional identification methods in VARs and establish when the researcher can restrict the number of structural shocks to be smaller than the number of variables in a VAR. Then I consider identification issues (Section 3). I discuss estimation and inference in Section 4. I show how to estimate the number of structural (fundamental) shocks in Section 5. In Section 6, my identification scheme is compared with other popular schemes. In Section 7, I present an empirical application to the analysis of monetary policy. Section 8 concludes.

2 Conventional and reduced rank VARs

The appealing features of VARs are the minimal set of assumptions necessary for the analysis, simple estimation and inference. Simplicity comes at the cost of underidentification of structural parameters of economic models. To illustrate this point, suppose that the reduced form of an economic model can be written as

$$\boldsymbol{X}_t = \sum_{i=1}^{p} \Pi_i \boldsymbol{X}_{t-i} + A\boldsymbol{\varepsilon}_t \qquad (2.1)$$

where $\boldsymbol{X}_t$ is a vector of q variables, $\boldsymbol{\varepsilon}_t$ is the vector of structural innovations, $A, \Pi_1, ..., \Pi_p$ are matrices of conformable sizes. The matrix A summarizes contemporaneous relationships in the economy. The standard theory of simultaneous equations immediately indicates that the matrix A is not generally identified because there are no excluded variables. Indeed, the VAR, by assumption, includes all relevant variables. For many cases (e.g. forecasting), this inability to recover A is not important. However, if A is incorrectly identified, all subsequent economic analyses, including impulse response functions and associated confidence bounds, can be safely scrapped. As I show below, some puzzles found in the VAR literature may well be a result of wrong identification of the matrix A. Thus, the matrix A is of central importance in the VAR framework.

To identify the matrix A, the researcher must impose extra assumptions on the structure of A and the properties of structural innovations ε_t. In the seminal paper, Sims (1980) postulated that $E\left(\varepsilon_t \varepsilon_t'\right) = I_q$ and A is a lower triangular matrix implying recursive structure of the economic model. Unfortunately, contemporaneous effects in macroeconomics do not typically have a recursive structure. A partial remedy to this problem is to assume that matrix A is block recursive (Keating, 1994). Dissatisfaction with the assumption of recursive structure generated a voluminous literature on so called structural VARs (e.g. Bernanke (1986), Shapiro and Watson (1988), Blanchard and Quah (1989), King et al. (1991), Strongin (1995), Christiano, Eichenbaum and Evans (1996, 1999), Leeper, Sims and Zha (1996), Bernanke and Mihov (1998), Gonzalo and Ng (2003), Rigobon and Sack (2003) and many others). Structural VARs impose identifying restrictions on the matrix A and/or matrices $\Pi_1, ..., \Pi_p$ and/or higher moments of innovations so that A can be recovered completely or partially (semi-structural VARs). Identifying restrictions are justified on the grounds of economic theory (e.g. demand shocks are neutral in the long run as in Blanchard and Quah (1989)) or institutional arrangements (e.g. reserves respond in a particular way as in Strongin (1995)).

It is conventional to assume that the number of shocks is equal to the number of variables in the VAR. In contrast to standard recursively identified VARs, structural VARs have the identified shocks associated with fundamental innovations, i.e. innovations in technology, preferences, or policy. Structural VARs, however, can produce too many structural shocks to be fundamental. For instance, in VARs with 10 variables there are 10 structural shocks.[1] In contrast to VARs, the standard business cycle (BC) model implies that $\boldsymbol{X}$, a vector of q economic variables, evolves according to

$$\boldsymbol{X}_t = \sum_{i=1}^{p} B_i \boldsymbol{X}_{t-i} + D\boldsymbol{\eta}_t \qquad (2.2)$$

where $\boldsymbol{\eta}_t$ is a vector of k shocks to technology, preferences and policy, and $B_1,\ldots,B_p,D$ are matrices of conformable sizes. Typically, $k << q$. If equation (2.2) is taken as the benchmark for economic

analysis, then $\boldsymbol{\varepsilon}_t$ in (2.1) is a linear combination of $\boldsymbol{\eta}_t$ and possibly other shocks. Linking macroeconomic theory and empirics then entails imposing restrictions on the covariance structure of VAR residuals so that the matrix A in (2.1) is consistent with the matrix D in (2.2). Furthermore, structural innovations in conventional VARs are completely attributed to fundamental shocks. It is, however, plausible that reduced form VAR residuals also absorb idiosyncratic innovations in the series. For example, the stock market is highly volatile and it is hardly conceivable that every change in the stock market is due to changes in fundamental factors. By pooling fundamental and idiosyncratic shocks, conventional VARs can give a misleading picture about the true responses of the economy to fundamental shocks. Specifically, as I show below, conventional VARs can underestimate the degree of contemporaneous cross-variable responses.

To formally analyze the problem, suppose, without loss of generality, that $\Pi_1,...,\Pi_p$ are known in (2.1). Also suppose there are n variables and k structural shocks.[2] It is impossible to recover structural shocks with $n<k$. A typical assumption is that $k = n$, as I have discussed, but there is no a priori reason not to have $n>k$. Consider a standard VAR(p) model: $\boldsymbol{X}_t = \sum_{i=1}^{p} \Pi_i \boldsymbol{X}_{t-i} + \boldsymbol{u}_t$. VAR residuals $\boldsymbol{u}_t$ are related to structural shocks $\boldsymbol{\varepsilon}_t$ in (2.1) as follows:

$$\underset{(n\times 1)}{\boldsymbol{u}_t} = \begin{bmatrix} u_{1t} \\ \vdots \\ u_{nt} \end{bmatrix} = \underset{(n\times k)}{\begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nk} \end{bmatrix}} \underset{(k\times 1)}{\begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{kt} \end{bmatrix}} = A\boldsymbol{\varepsilon}_t \tag{2.3}$$

where the matrix A is not necessarily a square matrix. Consider second moments of the error term $\boldsymbol{u}_t$. From (2.3) it is easy to find that

$$\Omega \equiv E\left(\boldsymbol{u}\boldsymbol{u}'\right) = E\left(A\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'A'\right) = AE\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)A' \tag{2.4}$$

If structural shocks $\boldsymbol{\varepsilon}_t$ are uncorrelated and normalized to have unit variance, then $E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right) = I$ and equation (2.4) simplifies to

$$\Omega = AA' \tag{2.5}$$

Equation (2.5) implies that Ω has reduced rank and, therefore, $X'X$ has a stochastic singularity under the null of reduced rank of A. To eliminate the singularity, I augment the model (2.3) with an additive idiosyncratic "noise" shock $\boldsymbol{v}_t$. The noise shock $\boldsymbol{v}_t$ does not have an economic meaning in the sense that it does not affect the interrelationship among $\boldsymbol{u}_t$ embodied in the matrix A and structural shocks $\boldsymbol{\varepsilon}_t$. The noise shock $\boldsymbol{v}_t$ may be a measurement error, a shock induced by noise/chartist traders, myopic consumers, etc.[3] Specifically, I consider

$$\begin{bmatrix} u_{1t} \\ \vdots \\ u_{nt} \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nk} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{kt} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ \vdots \\ v_{nt} \end{bmatrix} = A\boldsymbol{\varepsilon}_t + \boldsymbol{v}_t \tag{2.6}$$

where

$$E(\boldsymbol{\varepsilon}_t) = 0 \ \forall t, \ E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = I_k \ \forall t, E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s') = 0 \forall s \neq t \tag{2.7}$$

$$E(\boldsymbol{v}_t) = 0 \ \ \forall t, \Psi \equiv E(\boldsymbol{v}_t \boldsymbol{v}_t') = diag\left(\sigma_{v1}^2, \sigma_{v2}^2, ..., \sigma_{vn}^2\right) \ \ \forall t, E(\boldsymbol{v}_t \boldsymbol{v}_s') = 0 \ \ \forall s \neq t \tag{2.8}$$

$$E(\boldsymbol{\varepsilon}_t \boldsymbol{v}_s') = 0 \ \ \forall s, t \tag{2.9}$$

In addition to the standard assumption (2.7), assumption (2.8) postulates that shocks $\boldsymbol{v}_t$ are contemporaneously and intertemporally uncorrelated and have different variances. Condition (2.9) requires $\boldsymbol{v}_t$ to be uncorrelated with $\boldsymbol{\varepsilon}_t$ at all leads and lags. Representation (2.6) and conditions (2.7)-(2.9) essentially replicate assumptions necessary for the existence of a factor representation of the series $\boldsymbol{u}_t$.[4]

Given (2.6)-(2.9), consider the second moments of the error term $\boldsymbol{u}_t$:

$$\begin{aligned} \Omega \equiv E(\boldsymbol{u}_t \boldsymbol{u}_t') &= E(A\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' A') + E(\boldsymbol{v}_t \boldsymbol{v}_t') = \\ &= AE(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') A' + \begin{bmatrix} \sigma_{v1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{v2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{vn}^2 \end{bmatrix} = AA' + diag\left(\sigma_{v1}^2, \sigma_{v2}^2, ..., \sigma_{vn}^2\right) = AA' + \Psi \end{aligned} \tag{2.10}$$

This shows that the covariance matrix Ω can be decomposed into the covariance explained by the factor structure of the model (AA') and the covariance explained by idiosyncratic innovations in the data (Ψ). Note that Ω is of full rank as long as Ψ is non-degenerate.[5] Hence, $E(X'X)$ is invertible.

It is proven in factor analysis (e.g. Anderson and Rubin, 1956) that Ψ being positive definite (i.e. non-degenerate in my case) is a necessary condition for the existence of a factor representation, which is my representation (2.6) with the number of structural shocks smaller than the number of variables in the VAR (i.e. A has reduced rank). Thus, Ψ being non-degenerate is a necessary condition for the existence of representation (2.6), which I shall call the structural or fundamental form. Because rank(A)<n, I call my approach the reduced rank (RR) identification.[6]

Note that decomposition of the covariance matrix Ω and estimation of $\Pi_1,...,\Pi_p$ are disentangled. In the spirit of Sims (1980) and Bernanke (1986), I use lags to filter the series, find reduced form shocks, impose short run restrictions on the covariance of the reduced form shocks and identify/estimate the matrices A and Ψ in (2.10). The implicit assumption is that dynamic responses embodied in $\Pi_1,...,\Pi_p$ are the same for structural and noise shocks.

3 Identification

The covariance matrix Ω has $\frac{1}{2}n(n+1)$ unique entries; thus, the maximum number of identifiable parameters is $\frac{1}{2}n(n+1)$. The matrices A and Ψ have nk and n parameters, respectively. Therefore, a necessary condition for identification of the structural form (2.6) is:

$$\frac{1}{2}n(n+1) \geq nk + n = n(k+1) \tag{3.1}$$

However, A is not unique for $k \geq 2$ structural shocks. The classical factor analysis (e.g. Basilevsky, 1994) shows that while the matrix Ψ is uniquely identified, there are infinitely many matrices B and orthonormal rotation matrices M such $AA' = BB'$ with $B=AM$. To uniquely (up to sign) identify A, I have to eliminate the possibility of rotating axes of the factor space. Specifically, one needs at least $\frac{1}{2}k(k-1)$ restrictions on A so that there is no orthonormal rotation matrix M such that $B=AM$ satisfies the restrictions I impose on A. I focus on zero restrictions on some entries of A.[7] Let r_A denote the number of imposed zero restrictions. In the case of orthogonal structural shocks, a typical case in economic applications, the necessary condition to identify A is (Lawley and Maxwell, 1974):

$$r_A \geq \tfrac{1}{2}k\left(k-1\right) \tag{3.2}$$

Given (3.2) holds, identification of $nk+n-r_A$ then requires:

$$d = \tfrac{1}{2}n\left(n+1\right)-n\left(k+1\right)+r_A = \tfrac{1}{2}\left[\left(n-k\right)^2-n-k\right]+r_A-\tfrac{1}{2}k\left(k-1\right)\geq 0 \tag{3.3}$$

The parameter d denotes degrees of freedom and $d>0$ implies overidentification.

Anderson (2003) and Anderson and Rubin (1956) give sufficient conditions for zero restrictions to uniquely identify A. For the r^{th} column of an $\left(n\times k\right)$ matrix A – i.e. the r^{th} factor – to be uniquely identified three requirements must be satisfied:

R1) the column must contain at least k-1 zeros;

R2) the matrix of rows $A^{(r)}$ containing zeros at the r^{th} column with r^{th} column deleted must have rank k-1;

R3) each column of A must contain at least three non-zero entries.[8]

Another sufficient condition to identify the matrix A (Anderson (2003), Anderson and Rubin (1956)), which I will call R4, is that rows and columns of the matrix A can be rearranged so that the first k rows of the rearranged A form a lower triangular $\left(k\times k\right)$ matrix.[9] In general, conditions R4 and R1-R3 are not equivalent.[10]

As an example, consider the following covariance structure

$$\boldsymbol{u}_t = \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \\ u_{5t} \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ 0 & a_{52} \end{bmatrix}}_{A} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \\ v_{5t} \end{bmatrix} \tag{3.4}$$

The matrix A satisfies requirements R1-R3 since all columns have at least one zero entry and three non-zero entries and matrices $A^{(1)}=[a_{52}]$ and $A^{(2)}=[a_{11}]$ have rank one. Condition R4 is also satisfied: the first two rows of A form a lower triangular matrix. The model is overidentified because d=2.

Note that conditions R1-R3 enable me to partially identify A. In other words, if the researcher is interested in identifying the effects of only one factor (i.e. he or she does not need to identify A completely), it is enough to identify the column of A corresponding to the factor of interest. The following example illustrates partial identification:

$$A' = \begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} & a_{51} & a_{61} & a_{71} \\ a_{12} & a_{22} & a_{32} & a_{42} & a_{52} & a_{62} & a_{72} \\ 0 & 0 & a_{33} & a_{43} & a_{53} & a_{63} & a_{73} \end{bmatrix} \tag{3.5}$$

Observe that $A^{(3)}=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ has rank two. Thus, the third factor (the third column of A) is uniquely identified, while the first and the second are identified only up to rotation (no zero entries in the first and second columns). The rows and columns of A cannot be rearranged to form a lower triangular matrix in the first three rows of the rearranged A, i.e. R4 fails. To estimate the identified columns of A, I can put an arbitrary constraint on the entries of the first two columns to reach identification without affecting the identified third factor, e.g. set $a_{12}=0$.

Provided identification requirements are satisfied, I proceed to estimation and inference section.

4 Estimation and inference

I suggest using generalized method of moments (GMM) to estimate parameters matrices A, Ψ because it conveniently combines estimation and inference. The task is to cast factor model (2.6) as a collection of moment restrictions.[11]

To start with a simple case, suppose $\Pi_1, ..., \Pi_p$ are known. The goal is to decompose $\boldsymbol{u}_t$ into factors and uncorrelated errors, i.e. find matrices A, Ψ such that $\Omega = AA' + \Psi$ where Ψ is a diagonal matrix with non-negative entries. This provides the following moment conditions[12]:

$$vech\left(\Omega - AA' - \Psi\right) = 0 \tag{4.1}$$

The sample analogues to (4.1) corresponding to diagonal elements of Ω are

$$\frac{1}{T}\sum_{t=1}^{T} u_{it}^2 - \left(\sigma_{iv}^2 + \sum_{j=1}^{k} a_{ij}^2 \right) = 0 \tag{4.2}$$

and to off-diagonal elements:

$$\frac{1}{T}\sum_{t=1}^{T} u_{it} u_{jt} - \left(\sigma_{iv}^2 + \sum_{p=1}^{k} a_{ip} a_{pj} \right) = 0 \tag{4.3}$$

To simplify notation, define $F_t\left(A,\Psi\right) = vech\left(\boldsymbol{u}_t \boldsymbol{u}_t' - AA' - \Psi\right)$. Then the population parameters A and Ψ satisfy:

$$\left[A,\Psi\right] = \arg\min_{A,\Psi} \left\{ E\left[F_0\left(A,\Psi\right)\right]' \Sigma^{-1} E\left[F_0\left(A,\Psi\right)\right] \right\} \tag{4.4}$$

where Σ is some $\left(n \times n\right)$ positive definite weighting matrix.

Efficient GMM estimates of $\hat{A}, \hat{\Psi}$ are derived as a solution to

$$\min_{A,\Psi} \left\{ \left(\frac{1}{T}\sum_{t=1}^{T} F_t\left(A,\Psi\right) \right)' \hat{\Sigma}^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} F_t\left(A,\Psi\right) \right) \right\} \tag{4.5}$$

where

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} F_t\left(\tilde{A},\tilde{\Psi}\right) F_t\left(\tilde{A},\tilde{\Psi}\right)'$$

$$\left[\tilde{A},\tilde{\Psi}\right] = \arg\min_{A,\Psi} \left\{ \left(\frac{1}{T}\sum_{t=1}^{T} F_t\left(A,\Psi\right) \right)' \left(\frac{1}{T}\sum_{t=1}^{T} F_t\left(A,\Psi\right) \right) \right\}$$

Note that estimation of A and Ψ entails non-linear optimization in entries of A. Also note that the derivative of F with respect to A and Ψ is independent from $\boldsymbol{u}_t$, which is important in deriving the asymptotic distribution of A and Ψ. Significantly, if d in (3.3) is greater than zero, one can use the *J*-statistic, a byproduct of the GMM estimation, to test overidentifying restrictions imposed on A and Ψ. Note that the *J*-statistic does not depend on whether the model is uniquely identified since all optimal

solutions to (4.5) yield the same minimum. Thus, the statistic can be used to determine the rank of A, i.e. the number of factors (see below), even when A is partially identified.

Of course, $\boldsymbol{\Pi} \equiv \{\Pi_1, ..., \Pi_p\}$ is not known in typical applications. Recall that the moments used in estimation of the VAR are

$$E\left(\left[X_t - \Pi_1 X_{t-1} - ... - \Pi_p X_{t-p}\right] X_{t-k}\right) = 0 \qquad k = 1, ..., p \tag{4.6}$$

and note that VAR residuals $\boldsymbol{u}_t$ are functions of $\boldsymbol{\Pi}$. This is an example of two stage estimation. By focusing on (4.2) and (4.3) only and treating $\hat{\Pi}_1, ..., \hat{\Pi}_p$ as population parameters, I generally underestimate the variance of A and Ψ. The straightforward way to resolve this problem is to combine (4.2) and (4.3) with (4.6) and estimate $\boldsymbol{\Pi}, A, \Psi$ simultaneously. However, with large p and n, the number of moments can be overwhelming and the finite-sample properties of GMM can quickly deteriorate. Fortunately, I can circumvent simultaneous estimation as the following proposition suggests.

Proposition 1

Under regularity conditions, $\left[\hat{A}\left(\hat{\boldsymbol{\Pi}}\right), \hat{\Psi}\left(\hat{\boldsymbol{\Pi}}\right)\right]$ and $\left[\hat{A}\left(\boldsymbol{\Pi}\right), \hat{\Psi}\left(\boldsymbol{\Pi}\right)\right]$ have the same asymptotic distribution.

Proof: see appendix.

This proposition proves that while making standard asymptotic inference about A I can ignore the fact that $\hat{\Pi}_1, ..., \hat{\Pi}_p$ are estimates from the first stage and treat $\hat{\Pi}_1, ..., \hat{\Pi}_p$ as if they were the population parameters.[13] This greatly simplifies estimation and inference for A and Ψ: one only needs to estimate the VAR by OLS and in the GMM part treat the estimated residuals $\hat{\boldsymbol{u}}_t$ as population counterparts $\boldsymbol{u}_t$.

In the following proposition I show that $\hat{A}, \hat{\Psi}$ are asymptotically normally distributed.

Proposition 2

Suppose A is locally identified (i.e. $E\left[\nabla_{A,\Psi} F\left(A,\Psi\right)\right]$ is of full rank) and VAR residuals $\{u_t\}_{t=-\infty}^{\infty}$ are iid and $E\left(\boldsymbol{u}_t^4\right) < \infty$. Then $\sqrt{T}\left(\begin{bmatrix} vec\left(\hat{A}\right) \\ diag\left(\hat{\Psi}\right)' \end{bmatrix} - \begin{bmatrix} vec\left(A\right) \\ diag\left(\Psi\right)' \end{bmatrix}\right) \xrightarrow{d} N\left(0,\Phi\right)$.

Proof: see appendix.

In case A is not identified, the Hessian of the objective function in (4.4) is not invertible because the objective function is flat along parameters that are not identified uniquely (i.e. there are infinitely many rotations of A). To fix this problem, I can identify A by imposing further restrictions on A without distorting the estimates of identified factors, e.g. impose zeros as I do for (3.5). In sum, the estimation procedure is as follows:

1. Estimate VAR by OLS and store the residuals $\hat{\boldsymbol{u}}_t$.
2. If the matrix A is partially identified, impose restrictions on A to achieve identification without affecting identified columns of A.
3. Estimate the matrices A and Ψ using the stored residuals and the GMM estimator described above.
4. Use standard asymptotic inference to construct confidence intervals, test hypotheses, etc.

So far, I have focused on asymptotic inference and testing. Alternatively, one could use bootstrap methods to improve the finite sample properties of the GMM. For instance, Horowitz (1998) suggests using bootstrap procedures to compute bias-corrected GMM estimates for covariance structures and to improve the accuracy of inference.[14] Since moment conditions in my application are non-linear, I suggest using the k-step bootstrap[15] developed by Davidson and McKinnon (1999) and adapted for testing by Andrews (2002) to reduce the computational burden of the bootstrap.

Although asymptotically $\hat{\Pi}_1,...,\hat{\Pi}_p$ and $\hat{A},\hat{\Psi}$ are uncorrelated (Proposition 1), I cannot expect this condition to hold in a finite sample. Thus, in making inference about A and Ψ I cannot ignore sampling uncertainty induced by $\hat{\Pi}_1,...,\hat{\Pi}_p$. The dimension of $\Gamma = \left[vec[\boldsymbol{\Pi}]' \; vec[A]' \; diag(\Psi)' \right]$ is likely to be very large and, hence, bootstrap based on simultaneous estimation of Γ, which I call simultaneous bootstrap, may perform poorly for the same reasons why the simultaneous estimation of Γ performs poorly. An attractive alternative is to use sequential estimation of Γ for each bootstrap replication, i.e. for each bootstrap replication estimate $\boldsymbol{\Pi}$ and then estimate A,Ψ using $\hat{\boldsymbol{\Pi}}$. This approach, which I call sequential bootstrap, is very fast: the first stage is computed with OLS and the second one is computed with *k*-step bootstrap. Monte Carlo simulations (not reported) show that sequential bootstrap outperforms simultaneous bootstrap in terms of mean square error by a huge margin even in the smallest VAR compatible with reduced rank structure.

In sum, I use the following procedure for bootstrap:

1. Estimate VAR by OLS and store the residuals $\hat{\boldsymbol{u}}_t$.
2. Estimate the matrices A and Ψ using the stored residuals and the GMM estimator described above.
3. Resample residuals $\hat{\boldsymbol{u}}_t$ and create new series $\boldsymbol{u}_t^{(B)}$.[16]
4. Create new series $\boldsymbol{X}_t^{(B)}$ using $\hat{\boldsymbol{\Pi}}$ from step 1 and resampled $\boldsymbol{u}_t^{(B)}$ from step 3.
5. Estimate VAR using $\boldsymbol{X}_t^{(B)}$ and compute residuals.
6. Estimate the matrices A and Ψ using residuals from step 5.[17]
7. Repeat steps 3-6 sufficiently many times.
8. Compute bias, construct confidence intervals, etc.

To illustrate bootstrap and standard asymptotics procedures, I apply both procedures in my empirical example.

Once parameters are estimated, I can use standard VAR tools to analyze the properties of the model. First, I can recover structural shocks. Factor analysis (e.g. Anderson, 2003, Section 14.7) provides a simple formula to *estimate* the factors, i.e. structural shocks in my context:

$$\hat{\boldsymbol{\varepsilon}}_t = \left(I_n + \hat{A}'\hat{\Psi}^{-1}\hat{A} \right)^{-1} \hat{A}'\hat{\Psi}^{-1}\boldsymbol{u}_t$$

The estimate of $\hat{\boldsymbol{\varepsilon}}_t$ can be interpreted as the posterior estimate of ε_t given VAR residuals $\boldsymbol{u}_t$.[18] Note that if A is not fully identified, structural shocks cannot be recovered.

Second, variance decomposition and impulse response functions are easily constructed. Suppose the VAR in (2.1) has moving average representation $\boldsymbol{X}_t = \boldsymbol{\Theta}(L)\left(A\boldsymbol{\varepsilon}_t + \boldsymbol{v}_t\right)$ where $\boldsymbol{\Theta}(L)$ is a lag polynomial. Then the impulse response to the structural shock i is $\left\{\Theta_s A_i\right\}_{s=0}^{\infty}$ where A_i is the i^{th} column of the matrix A and Θ_s is the s^{th} term in $\boldsymbol{\Theta}(L)$. Likewise, the impulse response to idiosyncratic noise v_{it} is $\left\{\Theta_s e_i\right\}_{s=0}^{\infty}$ where e_i is the selection vector, i.e. the i^{th} column of I_n. Confidence bounds for impulse responses can be computed using Kilian's (1998) procedure.

The share of the variance of variable j at horizon h attributed to structural shock i is:

$$V_{j,i}^{\varepsilon} = \frac{\sum_{s=0}^{h} e_j'\Theta_s A^{(i)} A^{(i)\prime}\Theta_s' e_j}{\sum_{s=0}^{h} e_j'\Theta_s \left(AA' + \Psi\right)\Theta_s' e_j} \tag{4.7}$$

where $A^{(i)}$ is the i^{th} column of the matrix A corresponding to the i^{th} structural shock. For the idiosyncratic shock i the share is

$$V_{j,i}^{v} = \frac{\sum_{s=0}^{h} e_j'\Theta_s e_i \Psi e_i'\Theta_s' e_j}{\sum_{s=0}^{h} e_j'\Theta_s \left(AA' + \Psi\right)\Theta_s' e_j} \tag{4.8}$$

where e_i is the i^{th} selection vector. By inspecting the numerator in (4.7) one can note that if the factor i (the i^{th} column of A) is not identified, then its variance decomposition depends on orthonormal

rotations of A. Hence, impulse responses and variance decompositions are meaningful only for identified factors. For variance decomposition, however, one can note that the share of variance attributed to idiosyncratic noise is uniquely identified because Ψ is uniquely identified and, consequently, (4.8) is immune to rotations of A.

So far I have been assuming that I know the rank of A. Now I turn to cases when the number of factors (i.e. rank of A) is unknown.

5 How to determine the rank of A?

Determining the rank of A is not a trivial question. Factor analysis typically uses likelihood ratio tests to check if eigenvalues of certain matrices fall below some threshold so that contribution of would-be factors is indistinguishable from idiosyncratic noise.[19] Cragg and Donald (1997) suggest a simple way to estimate the rank of A in the framework of GLS estimators. I modify their procedure to paste it into my GMM framework.

Note that if the number of factors is less than the true number of factors, say k_0, the J-statistic would diverge to infinity as the sample size increases by consistency of J-tests. On the other hand, if $k>k_0$ then $J(k)\leq J(k_0)$, i.e. one can match moments better with more free parameters.[20] Consider the following generic criterion:

$$S(k) = f(T)^{-1} J(k, A, \Psi) - g(k)$$

where k is the rank of A, $J(k, A, \Psi)$ is the J-statistic for rank k given population parameters A and Ψ, $g(k)$ is a function strictly decreasing in k, and $f(T)$ is a function of sample size T such that $\lim_{T\to\infty} f(T) = \infty$ and $\lim_{T\to\infty} T^{-1} f(T) = 0$.[21] Under these conditions I can prove the following proposition.

Proposition 3

The rank k minimizing $S(k)$ – denoted by $\hat{k}$ – is a consistent estimate of k_0, i.e. $\lim_{T\to\infty} \Pr\left(\hat{k} > k_0\right) = \lim_{T\to\infty} \Pr\left(\hat{k} < k_0\right) = 0$.

Proof: see the appendix.

Note that to determine the number of factors one does not need the uniqueness of A because the value of the objective function – and consequently the value of the J-statistic – is the same for any rotation of A and, therefore, arbitrary identification constraints can be used to determine k. For example, one can impose zero restrictions that form upper triangular $k \times k$ submatrix of zeros in A.[22]

There are many functions $f(N)$ and $g(k)$ that satisfy the required properties. Popular choices are $f(N)=1$ and $g(k)=-2nk$, which corresponds to the Akaike information criterion (AIC), and $f(N)=\log(N)$ and $g(k)=nk$, which corresponds to the Schwarz information criterion (SIC).[23]

Note that AIC does not satisfy $\lim_{N\to\infty} f\left(N\right) = \infty$ thus it would tend to overestimate rank k. However, Monte Carlo simulations suggest that for moderately large samples AIC often outperforms SIC that imposes too heavy a penalty on overfitting.[24] Because the covariance matrix of the VAR residuals $\boldsymbol{u}_t$ is likely to have off-diagonal terms close to zero, the eigenvalues of AA' may be relatively small. In such a case, the above criteria can perform poorly because the penalty is too heavy (Cragg and Donald (1997)). I use Monte Carlo simulations to calibrate the penalty functions.

This completes the description of the reduced rank identification. To show its merits, I briefly compare it in the next section with other popular identification schemes and macroeconometric tools.

6 Discussion

The reduced rank (RR) identification preserves useful properties of structural VARs and improves upon them in several ways. The RR identification easily links structural shocks in VARs with shocks in theoretical business cycle models. Note that, like (semi-)structural VARs, the RR identification permits partial identification of A. Under the proposed identification scheme, VAR residuals are linear combinations of structural and noise shocks. Also the RR identification imposes

short run restrictions and therefore is not sensitive to a chosen VAR order. Thus, estimation and inference are simple and computationally undemanding.

More importantly, the RR identification can avoid Sims's (1998) "stickiness", a stylized fact that contemporaneous cross-variable responses tend to be very weak in conventionally identified VARs. Indeed, the covariance matrix of reduced form VAR residuals Ω typically has off-diagonal entries small relative to diagonal ones and, therefore, popular (e.g. Cholesky) decompositions of Ω result in weak cross-variable responses. In contrast, business cycle models tend to produce strong contemporaneous responses thus making empirical stickiness a puzzle. Now I show how the RR identification can address this puzzle.

To clearly present the difference between recursive identification schemes and the reduced rank identification, I assume that the system of equations linking structural shocks to VAR residuals is described in population as follows:

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = A\boldsymbol{\varepsilon} + \boldsymbol{v} \tag{6.1}$$

where $\boldsymbol{u}$ are reduced-form VAR residuals, $\boldsymbol{\varepsilon}$ are structural shocks with unit variance, and $\boldsymbol{v}$ are idiosyncratic noise in the series with covariance matrix $\Psi = diag\left(\sigma_1^2, \sigma_2^2, \sigma_3^2\right)$. Without loss of generality, I assume that 1) the structural shocks ε_2 represent innovations in monetary policy and 2) the second variable is the fed funds rate. I am interested in the contemporaneous response of the variables to an innovation in monetary policy. I will compare the reduced rank identification with the Cholesky decomposition as a prototype of all recursive identification schemes.

By construction, the reduced rank identification has factor structure and, therefore, it correctly recovers the model in (6.1). Specifically, the contemporaneous response of the variables to a unit innovation in ε_2 is $\left[0, a_{22}, a_{32}\right]$. Comovement of the series is determined by the sign of $a_{22}a_{32}$. If

$a_{22}a_{32}$ is positive (negative), series move in the same (opposite) direction in response to a monetary policy shock ε_2.

In contrast, the Cholesky decomposition ignores possible factor structure. Because of the ordering, the Cholesky identification correctly requires that the monetary policy ε_2 be orthogonal to u_1. Define $\tilde{u}_2, \tilde{u}_3$ as VAR residuals net of u_1, i.e. $\tilde{u}_2, \tilde{u}_3$ are found as residuals from regressing u_2, u_3 on u_1. Note that u_1 is contaminated with classical measurement error v_1. Hence, estimated α_{12}, α_{13} are attenuated towards zero. The contemporaneous response to a unit innovation in ε_2 is represented by the following vector

$$\left[0, \operatorname{var}\left(\tilde{u}_1\right)^{1/2}, \operatorname{cov}\left(\tilde{u}_1, \tilde{u}_2\right) \cdot \operatorname{var}\left(\tilde{u}_1\right)^{-1/2}\right] \tag{6.2}$$

where the first entry of the vector is zero by the ordering of the variables. Comovement of responses to ε_2 depends on the sign of $\operatorname{cov}\left(\tilde{u}_2, \tilde{u}_3\right)$. If $\operatorname{cov}\left(\tilde{u}_2, \tilde{u}_3\right) > 0$ ($\operatorname{cov}\left(\tilde{u}_2, \tilde{u}_3\right) < 0$), the series respond in the same (opposite) direction to an innovation in ε_2. In this simple case, it can be easily shown that

$$\operatorname{cov}\left(\tilde{u}_2, \tilde{u}_3\right) = a_{21}a_{31}\frac{\sigma_1^2}{a_{11}^2 + \sigma_1^2} + a_{22}a_{32} \tag{6.3}$$

$$\operatorname{var}\left(\tilde{u}_2\right) = \frac{a_{21}^2\left(1 + a_{11}^2\right)\sigma_1^4}{\left(a_{11}^2 + \sigma_1^2\right)^2} + a_{22}^2 + \sigma_2^2 \tag{6.4}$$

$$\operatorname{var}\left(\tilde{u}_2\right) = \frac{a_{31}^2\left(1 + a_{11}^2\right)\sigma_1^4}{\left(a_{11}^2 + \sigma_1^2\right)^2} + a_{32}^2 + \sigma_3^2 \tag{6.5}$$

It follows from (6.3)-(6.5) that estimates of the covariance and variances are in general biased and inconsistent. Specifically, variances are unambiguously overestimated. From (6.2) I then find that $\operatorname{cov}\left(\tilde{u}_1, \tilde{u}_2\right) \cdot \operatorname{var}\left(\tilde{u}_1\right)^{-1/2}$ is attenuated towards zero, ceteris paribus, and, consequently, the third variable weakly responds to structural innovation ε_2 under the Cholesky identification. Moreover,

note that even when $\sigma_1 = 0$, the Cholesky identification predicts that the contemporaneous response of the variables to ε_2 is $\left[0, \left(a_{22}^2 + \sigma_2^2\right)^{1/2}, a_{22}a_{32}\left(a_{22}^2 + \sigma_2^2\right)^{-1/2}\right]$. Since the variances are generally inflated by $\sigma_1^2, \sigma_2^2, \sigma_3^2$, the Cholesky identification finds a weak response of the third variable to ε_2. For models more general than (6.1), I can conclude that the Cholesky decomposition is biased to finding little contemporaneous comovement in the series, i.e. stickiness. Because the Cholesky identification finds little contemporaneous comovement of fed funds rate with variables ordered before fed funds rate, this can also explain why recursively identified structural innovations in monetary policy are often highly correlated with VAR residuals corresponding to the fed fund rate equation.

In general, I cannot unambiguously sign the bias of the Cholesky identification for the estimate of $\mathrm{cov}\left(\tilde{u}_2, \tilde{u}_3\right)$ because it depends on the sign of $a_{21}a_{31}$. Note, however, that if $\mathrm{sgn}\left(a_{21}a_{31}\right) \neq \mathrm{sgn}\left(a_{22}a_{32}\right)$, then $\mathrm{cov}\left(\tilde{u}_2, \tilde{u}_3\right)$ can be approximately zero even when $a_{22}a_{32} \neq 0$. Moreover, it is possible that empirical and theoretical responses can have different signs, i.e. $\mathrm{sgn}\left(\mathrm{cov}\left(\tilde{u}_2, \tilde{u}_3\right)\right) \neq \mathrm{sgn}\left(a_{22}a_{32}\right)$. Specifically, the necessary condition for this to happen is that variables co-move in response to some structural shocks and move in opposite directions in response to other structural shock, i.e. $\mathrm{sgn}\left(a_{21}a_{31}\right) \neq \mathrm{sgn}\left(a_{22}a_{32}\right)$ in our example. Thus, the Cholesky identification can produce not only attenuated but also incorrectly signed responses.

Overall this simple exercise shows that the Cholesky-type identification is likely to find little and, possibly, incorrectly signed contemporaneous comovement of the series in response to identified structural shocks in a model like (6.1). In contrast to the Cholesky-type identification, the reduced rank identification takes out noise, i.e. the matrix Ψ, from the diagonal of the covariance matrix Ω and, consequently, finds large comovement of series in response to structural shocks. Hence, stickiness of responses can be an artifact of inappropriate identification.

In addition to VARs, macroeconomic analyses have used (approximate) factor models (Geweke (1977), Sargent and Sims (1977), Quah and Sargent (1993), Stock and Watson (1988, 1998), Forni et al (2000), Forni and Lippi (2001), Giannone et al (2002), Kose et al (2003) among many others). This approach can, at least partially, address a call by Granger (2001) to expand the list of variables in macroeconometric models. A typical factor model (FM) can be written as (e.g. Forni et al, 2000):

$$\boldsymbol{X}_t = \Gamma(L) \cdot \boldsymbol{f}_t + \boldsymbol{v}_t \tag{6.6}$$

$$\Theta(L)\, \boldsymbol{f}_t = \boldsymbol{\xi}_t \tag{6.7}$$

where $\boldsymbol{X}_t$ is a vector of q variables, Γ, Θ are lag polynomials (Γ is often called factor loadings), $\boldsymbol{f}_t$ is a $(k \times 1)$ vector of factors and $\boldsymbol{\xi}_t, \boldsymbol{v}_t$ are iid innovations. Like the RR identification, FMs typically have $k \ll q$. FMs usually impose few restrictions in (6.6) and (6.7). This is, however, a double sided sword. Most importantly, the factors are identified only up to rotation unless further restrictions on Γ, Θ are imposed or there is only one factor (i.e. k=1) like in, for example, Altug (1989), Otrok and Whiteman (1998), Forni et al (2000). I am not aware of FMs with two or more uniquely identified factors.[25] In typical applications of the FM the dimension of $\boldsymbol{X}$ is large. Hence, I shall have to impose enormously many restrictions to uniquely identify factors, thus making estimation of (6.6) and (6.7) computationally infeasible. It may also be very difficult to justify restrictions in borderline cases. The inability to "name" (i.e. uniquely identify) factors greatly limits the use of FMs in macroeconomic analyses because if the causes (factors) are not identified the effects are not identified either.

In contrast, the VAR model with a reduced rank identification utilizes 1) parametric modeling of dynamics of $\boldsymbol{X}$ and 2) factor analysis to recover structural shocks from (approximately) serially uncorrelated VAR residuals. A few plausible restrictions on A can break the possibility of rotating factor basis and, consequently, factors can be uniquely identified. Moreover, a partial identification of A and a small number of variables free the researcher from imposing dubious restrictions on factor

loadings (columns of A) to identify structural shocks.[26] Computationally, the reduced rank identification is far less demanding than factor models.

7 Application: identifying the effects of monetary policy

I apply my method to the classical problem of identifying the effects of unanticipated innovations to monetary policy. The analyzed VAR specification is a simplified version of the specification widely used in the literature (e.g. Kim and Roubini (2000)). I choose the recursive Cholesky identification as a benchmark for comparison.

Let $\boldsymbol{Y}$ be a vector of macroeconomic variables with sluggish adjustment (e.g. real GDP), $\boldsymbol{V}$ – vector of asset prices (e.g. S&P 500) and monetary aggregates, $\boldsymbol{R}$ – vector of policy variables (e.g. fed funds rate). Hence, $\boldsymbol{X} \equiv [\boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{R}]$. In my model, two macroeconomic variables with sluggish adjustment are real GDP (RGDP) and GDP deflator (PGDP) so that $\boldsymbol{Y}$=[*RGDP*, *PGDP*].[27] The vector of asset price variables includes the index of commodity prices (CRB) and the real exchange rate (EXRUS); hence, $\boldsymbol{V}$=[*CRB*, *EXRUS*]. The set of policy variables is exhausted by the fed funds rate (FFR), i.e. $\boldsymbol{R}$=[*FFR*]. My baseline ordering is [*RGDP*,*PGDP*,*CRB*,*EXRUS*,*FFR*].[28]

I assume that there are two shocks in this economy $\boldsymbol{\varepsilon}_t = \left[\varepsilon_t^{TECH}, \varepsilon_t^{POL}\right]$ where ε_t^{POL} is an innovation in policy and ε_t^{TECH} is a technology innovation.[29]

To identify the effects of monetary policy, many (e.g. Bernanke and Blinder (1992), Bernanke and Mihov (1998)), but not all (e.g. Sims and Zha (1996), Cushman and Zha (1997)), of the VAR identification schemes assume a block recursive structure of the matrix A linking unobserved structural shocks to observed VAR residuals. For example, the Cholesky decomposition resulting in a unique lower triangular matrix A is a special case of the block recursive structure. For concreteness, consider the following relationship between reduced-form residuals and structural innovations:

$$\boldsymbol{u}_t = \begin{bmatrix} u_t^{RGDP} \\ u_t^{PGDP} \\ u_t^{CRB} \\ u_t^{EXRUS} \\ u_t^{FFR} \end{bmatrix} = \left[\begin{array}{cc:cc:c} a_{11} & a_{12} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 \\ \hdashline a_{31} & a_{32} & a_{33} & a_{34} & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} & 0 \\ \hdashline a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right] \begin{bmatrix} \varepsilon_t^{TECH} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \\ \varepsilon_t^{POL} \end{bmatrix} = \left[\begin{array}{c:c:c} A_{YY} & 0 & 0 \\ \hdashline A_{VY} & A_{VV} & 0 \\ \hdashline A_{RY} & A_{RV} & A_{RR} \end{array}\right] \boldsymbol{\varepsilon}_t = A\boldsymbol{\varepsilon}_t \qquad (7.1)$$

where A_{YY}, for example, indicates contemporaneous interactions among variables in the ***Y***-block and A_{RY} indicates the contemporaneous effect of innovations in the ***Y***-block variables on variables in the ***R***-block. Since this identification requires equality of the number of shocks and variables, the researcher has to introduce, name and interpret structural shocks $\varepsilon_{2t}, \varepsilon_{3t}, \varepsilon_{4t}$. Hence, the matrix A is a square 5×5 matrix. Since I am interested in the effects of monetary policy, I can, without loss of generality, set matrices A_{YY} and A_{VV} to be lower triangular.

A recursive structure of A is justified on the grounds of Wold causal chain and minimum delay restrictions. (For instance, the policy cannot affect contemporaneously variables in ***Y*** and ***V*** blocks.) No response of the ***Y***-block variables to shocks in the ***R***-block variables appears to be a reasonable assumption because the variables in the ***Y***-block are slow to adjust. However, there is essentially no delay in the reaction of continuously clearing asset prices to changes in policy variables. Thus, any identification scheme that relies on a recursive structure in the [***V***,***R***] sub-block of the matrix A is potentially erroneous. In more general setups, even procedures that focus on the matrix A_{RR}, for example semistructural identification of monetary policy, can produce misleading results because identification of A_{RR} is based on residuals from regressing VAR errors in the ***R*** block on errors in the ***Y*** and ***V*** blocks and, thus, contemporaneous effects of ***R*** on ***V*** are ignored (i.e. A_{VR} is assumed to be zero matrix).

In contrast, my reduced rank (RR) identification imposes the following structure of the matrix A:

$$\boldsymbol{u}_t = \begin{bmatrix} u_t^{RGDP} \\ u_t^{PGDP} \\ u_t^{CRB} \\ u_t^{EXRUS} \\ u_t^{FFR} \end{bmatrix} = \left[\begin{array}{c|c} a_{11} & 0 \\ a_{21} & 0 \\ \hline a_{31} & a_{32} \\ a_{41} & a_{42} \\ \hline a_{51} & a_{52} \end{array}\right] \begin{bmatrix} \varepsilon_t^{TECH} \\ \varepsilon_t^{POL} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \\ v_{5t} \end{bmatrix} = \left[\begin{array}{c|c} A_{YY} & 0 \\ \hline A_{VY} & A_{VR} \\ \hline A_{RY} & A_{RR} \end{array}\right] \boldsymbol{\varepsilon}_t + \boldsymbol{v}_t = A\boldsymbol{\varepsilon}_t + \boldsymbol{v}_t \qquad (7.2)$$

The second column of A in (7.2) summarizes the contemporaneous effects of an innovation in monetary policy. Note that A is identified because the condition R4 is satisfied. To keep the specification flexible, I allow noise $\boldsymbol{v}_t$ to be in each equation.

From (7.2), one can see that I continue to assume that structural policy shocks cannot contemporaneously affect variables in the $\boldsymbol{Y}$ block. However, in contrast to (7.1), I assume that policy innovation affect variables in the $\boldsymbol{V}$ block. I also assume that there are no structural shocks associated with asset prices. This assumption indicates that asset prices respond only to fundamental innovations (i.e. innovations in policy and technology) and idiosyncratic innovation in asset prices (e.g. "bubble" component, noise traders). In this interpretation, A_{VR} reflects equilibrium responses of asset prices to policy shocks. One could note that a similar structure can be generated by VARs that order the policy block $\boldsymbol{R}$ before the asset prices block $\boldsymbol{V}$. The similarity is, however, superficial. In those VARs, asset prices respond to *all* innovations in *policy variables*. In my identification, asset prices respond to innovations in *policy* and *do not* respond to *idiosyncratic noise in policy variables*. For example, asset prices respond to fed funds rate shocks corresponding to "change in policy" but do not respond to fed funds rate shocks corresponding to noise in fed fund rate series. In other words, my framework separates innovations in policy variables into innovations in policy and idiosyncratic innovations while standard and structural VARs do not.

Now I contrast these two approaches to identification. I estimate the VAR in levels after taking logs of *RGDP*, *PGDP*, *CRB*, and *EXRUS*. All series are monthly. *RGDP* and *PGDP* series are from Bernanke and Mihov (1998). The number of lags in the VAR is selected to be 13 to eliminate serial correlation in the residuals and to match lags in Bernanke and Mihov (1998). My sample covers 1965-

1996 but I exclude 1979-1982 when a different monetary regime was in place (Bernanke and Mihov, 1998).

My AIC selection criterion suggests that k=2 is the appropriate number of factors. Overidentifying restrictions test cannot reject the hypothesis of two structural shocks at any reasonable significance level (p-value is 0.7). Table 1 presents the standard and bias corrected estimates of the matrix A in (7.2), and their asymptotic standard errors and bootstrap confidence intervals.[30]

The estimated coefficients of the second column of A show the impact responses of *CRB*, *EXRUS* and *FFR* variables to a innovation in monetary policy. Note that commodity prices and real exchange rate respond to a shock in monetary policy as predicted by macroeconomic theory: commodity prices fall on impact ($a_{32}<0$) while the real exchange rate rises (USD appreciates; $a_{42}>0$). These responses are economically and statistically different from zero. In contrast, the Cholesky identification with ordering [*RGDP*,*PGDP*,*CRB*,*EXRUS*,*FFR*] implies that *CRB* and *EXRUS* do not respond contemporaneously. This is, however, not an artifact of ordering. Other orderings – e.g. [*RGDP*,*PGDP*,*FFR*,*CRB*,*EXRUS*] – in the Cholesky identification produce qualitatively very similar results, i.e. on impact *CRB* and *EXRUS* respond very weakly to innovations in *FFR*, thus confirming Sims's (1998) stickiness observation.

Interestingly, in the Cholesky identification, almost 95% of the variance in u_{FFR} is attributed to innovations in monetary policy, while in the reduced rank case it is only 26%=$a_{52}^2 / \sigma_{U_{FFR}}^2$. The remaining noise in the fed funds rate due to other structural and idiosyncratic innovations (e.g. trading frictions, weather, etc.) move the rate but not the policy while in the Cholesky identification much of this variance is labeled as policy. Hence, the RR identification indicates that a large portion of innovation in the *FFR* is not associated with policy changes. This makes sense if the Fed does not move *FFR* at random. Given the noisiness of *FFR* series, the RR identification appears more plausible than the Cholesky identification.

Impulse response functions[31] (IRFs) for the conventional Cholesky identification and reduced rank identification are reported in Figure 1.[32] For each identification scheme, the innovation is a unit shock to monetary policy. The time path of *FFR* is similar for the Cholesky and reduced rank (RR) identifications.

As I have already noted from Table 1, the index of commodity prices (*CRB*) and the real exchange rate (*EXRUS*) contemporaneously respond to changes in monetary policy under the RR identification. The RR identification predicts that commodity prices fall on impact. Then it continues falling, after 20 months it levels off and gradually recovers. In contrast, under the Cholesky identification, *CRB* increases in the first months after the shock and only then it starts to fall and generally replicate the shape of the response under the RR identification.

EXRUS appreciates on impact and then gradually depreciates under the RR identification. This response is consistent with the prediction of macroeconomic models with sticky prices (e.g. Dornbusch, 1976): tight monetary policy should lead to an appreciation of the real exchange rate on impact and its subsequent depreciation. Under the Cholesky identification, *EXRUS* peaks only after 30 months and then starts depreciating, which is "the exchange rate puzzle" (e.g. Eichenbaum and Evans (1995), Kim and Roubini (2000) and Faust and Rogers (2003)). As I have mentioned above, the response does not qualitatively change if I put *FFR* before asset prices. Apparently, the RR identification matches the theoretical responses much better than the Cholesky identification.

The differences in the response of the price level to changes in monetary policy are of special interest. It is a well known observation that in VARs with a block recursive identification the price level tends to rise in the first periods after "restrictive" monetary policy shocks. Since there is no theory behind these dynamics, the phenomenon has been called "the price level puzzle" (e.g. Sims (1992)). It has been conjectured that once the effect of asset prices as leading indicators is taken into account the puzzle would disappear. Indeed, zero is within confidence bounds around the spike in the price level, yet the point estimate of the price level response remains embarrassingly positive.

According to the RR identification, there is no price level puzzle after a "restrictive" monetary shock. The price level remains essentially zero for six months and then starts to fall. On the other hand, the Cholesky identification produces the puzzle as expected.

The response of GDP is also quite different. In the first months after the shock, the behavior of IRFs according to the Cholesky and RR identification is very similar. However, after six month, GDP stabilizes in the RR case and continues to fall in Cholesky case. The RR identification does not produce a hump shaped response typical for recursive identification schemes. A partial answer to this is in the behavior of commodity prices and the real exchange rate. In contrast to the impulse responses produced by the Cholesky identification, the RR identification has a gradual increase in the index of commodity prices and gradual depreciation of the real exchange rate at approximately 20 months. These two effects move real GDP in different directions: higher commodity prices slow the growth of real GDP while the depreciation of the real exchange rate stimulates it.

The dynamics of real GDP and the price level provide an interesting perspective. On impact, the price level is sticky and all adjustment is happening through real quantities. After approximately six months, the price level begins to adjust. This is consistent with the New Keynesian explanation of how monetary policy affects output and prices.

Impulse responses produced by the Cholesky and reduced rank identification are strikingly different. What drives this result? In our discussion of the Cholesky and RR identification schemes, I have noted that the Cholesky identification tends to produce attenuated and, possibly, incorrectly signed responses to structural shocks. Specifically, I have derived that if the pattern of comovement in responses is different across structural shocks (i.e. variables co-move in response to some structural shock and move in different directions in response to another structural shock), then the Cholesky identification can produce incorrectly signed responses. Estimates of the matrix A presented in Table 1 show that this is precisely the case. For example, *FFR* and *EXRUS* increase on impact in response to ε^{POL} but move in opposite directions in response to ε^{TECH}. Also *FFR* and *CRB* co-move in response to

ε^{TECH} but move in opposite directions in response to ε^{POL}. Consistent with our theoretical predictions, the Cholesky identification even with ordering [*RGDP*,*PGDP*,*FFR*,*CRB*,*EXRUS*] finds 1) a very weak correlation between monetary policy shocks and VAR innovations in *EXRUS* and 2) the theoretically wrong positive sign for correlation between identified shocks to monetary policy and VAR innovations in *CRB* (Figure 2).[33]

Moreover, as anticipated, monetary policy shocks under the RR identification are less correlated with innovations in fed funds rate FFR than monetary policy shocks under the Cholesky identification (Figure 3).[34] The reason is that the variances of idiosyncratic errors $\boldsymbol{\nu}$ dwarf off-diagonal entries of the covariance matrix of VAR residuals relative to its diagonal entries and, therefore, the Cholesky identification finds little comovement in the series thus implying a high correlation between VAR innovations in *FFR* and the Cholesky-identified monetary policy shocks.

In sum, reduced rank identification appears to produce more reasonable impulse responses than recursive identification. Specifically, reduced rank identification 1) resolves two well-known puzzles, 2) explains why the policy instrument is volatile in the recursive identification schemes, and 3) addresses stickiness of contemporaneous cross-variable response. The cornerstone of this success is, of course, the identified factor structure of VAR residuals.

8 Conclusion

There is often a disconnection between theoretical business cycle models and empirical VAR models as the number of structural shocks in VARs typically is not equal to the number of fundamental (i.e. technology, preferences, policy) shocks in business cycle models. This paper shows how the reduced rank identification that imposes a factor structure on VAR residuals rectifies this problem. Specifically, the reduced rank identification sets the number of structural shocks to be less than the number of variables in a VAR. This novel approach to identification has simple estimation and inference and it is no more computationally demanding than typical structural identification

schemes. The paper also presents formal statistical criteria to determine the number of structural shocks. Importantly, unlike dynamic factor models, the reduced rank approach uniquely identifies structural innovations and therefore it can be used for economic analyses. Finally, the reduced rank identification can explain stickiness (weak contemporaneous cross-variable response) regularly found in conventionally identified VARs. In empirical applications, the key is to find economically plausible restrictions on the matrix A linking observed VAR residuals with structural innovations.

In my monetary policy example I show that the reduced rank identification produces results qualitatively different from and more theoretically plausible than those of conventional identification schemes. In particular, the reduced rank identification finds no price level and exchange rate puzzles, the curse of all recursive identification exercises. The reduced rank identification also finds a small contribution of policy innovations to the variance of fed funds rate innovations thus implying that the policy is not as noisy as recursive identification schemes often suggest.

Of course, some issues remain unresolved. For example, reduced rank identification does not specify how many and which variables should be included in VAR. Potentially, more signals can improve the estimate of A but too many signals can easily make VAR infeasible. Where to draw the line is left for future research.

9 References

Abowd, John M. and David Card, "Intertemporal Labor Supply and Long-Term Employment Contracts," *American Economic Review* 77:1 (1987), 50-68.

---------- "On the Covariance Structure of Earnings and Hours Changes," *Econometrica* 57:2 (1989), 411-445.

Altug, Semru, "Time-to-Build and Aggregate Fluctuations: Some New Evidence," *International Economic Review* 30:4 (1989), 889-920.

Anderson, T.W. and Herman Rubin, "Statistical Inference in Factor Analysis," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Jerzy Neyman, ed., (Berkeley, CA: University of California Press, 1956).

Anderson, T.W., *An Introduction to Multivariate Statistical Analysis*. (Wiley, 2003).

Andrews, Donald W.K., "Higher Order Improvements of a Computationally Attractive k-Step Bootstrap for Extremum Estimators," *Econometrica* 70:1 (2002), 119-162

Bai, Jushan and Serena Ng, "Determining the Number of Factors in Approximate Factor Models," *Econometrica* 70:1 (2002), 191-221.

Basilevsky, Alexander, *Statistical Factor Analysis and Related Methods: Theory and Applications*. (Wiley, 1994)

Bernanke, Ben and Alan Blinder, "The federal Funds Rate and the Channels of monetary Transmission," *American Economic Review* 82 (1992), 901-921.

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

Bernanke, Ben S., "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on public Policy* 26 (1986), 49-100.

Bernanke, Ben, and Jean Boivin, "Monetary Policy in a Data-Rich Environment," *Journal of Monetary Economics* 50 (2002), 525-546.

Bernanke, Ben, Jean Boivin and Piotr Eliasz, "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," NBER Working paper #10220 (2004).

Blanchard, O. J. and D. Quah, "The Dynamic Effects of aggregate Demand and Supply Disturbances," *American Economic Review* 79 (1989), 655-673.

Christiano, L., M. Eichenbaum and C.L. Evans, "Monetary Policy Shocks: What We Have Learned and to What End?" in *Handbook of Macroeconomics*, M. Woodford and J. Taylor, eds., Vol. 1A, (Amsterdam: North Holland, 1999).

---------- "The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds," *Review of Economics and Statistics* 78 (1996), 16-34.

Cragg, G. John and Stephen G. Donald, "Factor analysis under More General Conditions with Reference to Heteroscedasticity of Unknown Form," in *Advances in Economics and Quantitative Economics*, G.S. Maddala, P.C.B. Phillips, T.N. Srinivasan, eds. (Cambridge, MA: Basil Blackwell, 1995) .

---------- "Inferring the Rank of a Matrix," *Journal of Econometrics* 76 (1997), 223-250.

Cushman, D.O. and T. Zha, "Identifying Monetary Policy in a Small Open Economy under Flexible Exchange Rates," *Journal of Monetary Economics* 39 (1997), 433-448.

Davidson , Russell and James G. McKinnon, "Bootstrap Testing in Non-linear models," *International Economic Review* 40:2 (1999), 487-508.

Dornbusch, Rudiger, "Expectations and Exchange Rate Dynamics," *Journal of Political Economy* 84 (1976), 1161-1176.

Efron, Bradley, and Robert J Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall/CRC, 1993).

Eichenbaum, Martin and Charles L. Evans, "Some Empirical Evidence on the Effects of Shocks to Monetary Policy on Exchange Rates," *Quarterly Journal of Economics* 110:4 (1995), 975-1009.

Faust, Jon and Eric M. Leeper, "When Do Long-Run Identifying Restrictions Give Reliable Results," *Journal of Business and Economic Statistics* 15:3 (1997), 345-353.

Faust, Jon, and John H. Rogers, "Monetary Policy's Role in Exchange Rate Behavior," *Journal of Monetary Economics* 50 (2003), 1403-1424.

Forni, M. Hallin, M. Lippi, and Reichlin, L., "The Generalized Dynamic Factor Model: Identification and Estimation," *The Review of Economic and Statistics* 82 (2000), 540-554.

---------- "Opening the black box: identifying shocks and propagation mechanisms in VAR and factor models." CEPR working paper #4133 (2003).

Forni, Mario, and Marco Lippi, "The Generalized Dynamic Factor Model: Representation Theory," *Econometric Theory* 17 (2001), 1113-1141

Geweke, J., "The Dynamic Factor Analysis of Economic Time Series," in D.J. Aigner and A.S. Goldberger (eds.), *Latent Variables in Socio-Economic Models* (Amsterdam: North-Holland, 1977).

Giannone, D., L. Reichlin, and L. Sala, Monetary policy in real time. mimeo (2004).

Gonzalo, Jesus and Serena Ng, "A Systematic Framework for Analyzing the Dynamic Effects of Permanent and Transitory Shocks," *Journal of Economic Dynamics and Control* 25 (2003), 1527-1546.

Granger, C.W.J., "Macroeconometrics – Past and future," *Journal of Econometrics* 100 (2001), 17-19.

Hall, Peter, and Joel Horowitz, "Bootstrap Critical Values for Testing Based on Generalized-Methods-of-Moment Estimators," *Econometrica* 64:4 (1996), 891-916.

Hall, Robert and Frederic S. Mishkin, "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households," *Econometrica* 50 (1982), 1029-1054.

Hausman, Jerry A. and William E. Taylor, "Identification in Linear Simultaneous Equations Models with Covariance Restrictions: An Instrumental Variables Interpretation," *Econometrica* 51:5 (1983), 1527-1550.

Hausman, Jerry A., Whitney K. Newey, and William E. Taylor, "Efficient Estimation and Identification of Simultaneous Equation Models with Covariance Restrictions," *Econometrica* 55:4 (1987), 849-874.

Horowitz, Joel L., "Bootstrap Methods for Covariance Structures," *Journal of Human Resources* 33:1 (1998.), 39-61.

Joreskog, Karl G. and Dag Sorbom, *Advances in Factor Analysis and Structural Equation Models* (Cambridge, MA: Abt Books, 1979).

Keating, John W, "Structural Information in Recursive VAR Orderings," *Journal of Economic Dynamics and Control* 20 (1996), 1557-1580

Kilian, Lutz, "Small-sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics* 80:2 (1998), 218-230.

Kim, Soyoung and Nouriel Roubini, "Exchange Rate Anomalies in the Industrial Countries" A Solution with a Structural VAR Approach," *Journal of Monetary Economics* 45 (2000), 561-586.

King, Robert G., Charles I. Plosser, James H. Stock and Mark W. Watson, "Stochastic Trends and Economic Fluctuations," *American Economic Review* 81:4 (1991), 819-840.

Kose, M Ayhan, Christopher Otrok, and Charles H. Whiteman, "International Business Cycles: World, Region, and Country-Specific Factors," *American Economic Review* 93:4 (2003), 1216-1239.

Lawley, D.N., and A.E. Maxwell, *Factor Analysis as a Statistical Method* (London: Butterworths,1974).

Leeper, Eric M., Christopher A. Sims and Tao Zha, "What Does Monetary Policy Do?" *Brookings Papers on Economics Activity* 1996:2 (1996), 1-63.

Otrok, Christopher and Charles H. Whiteman, "Bayesian Leading Indicators: Measuring and Predicting Economic Conditions in Iowa," *International Economic Review* 34:4 (1998), 997-1014.

Quah, D., and T.J. Sargent, "A Dynamic Index Model for Large Cross Sections," in J.H. Stock and M.W. Watson (eds.), *Business Cycles, Indicators, and Forecasting* (Chicago: University of Chicago Press, 1993).

Rigobon, Roberto and Sack Brian, "Measuring the Reaction of Monetary Policy to the Stock Market," *Quarterly Journal of Economics* 118:2 (2003), 639-669.

Sargent, T.J. and C.A. Sims. 1977. "Business Cycle Modeling without Pretending to Have Too Much a priori economic theory," in C. Sims et al. (eds), *New Methods in Business Cycle Research*,

Shapiro, Matthew D. and Mark W. Watson, "Sources of Business Cycle Fluctuations," *NBER Macroeconomics Annual* (Cambridge: MIT Press, 1988).

Sims, Christopher A. and Tao Zha, "Does Monetary Policy Generate recessions?" mimeo (1996).

Sims, Christopher, "Macroeconomics and Reality," *Econometrica* 48 (1980), 1-48.

---------- "Interpreting the Macroeconomics Time Series Facts: the Effects of Monetary Policy," *European Economic Review* 36 (1992), 975-1011.

---------- "Stickiness," *Carnegie-Rochester Conference Series on Public Policy* 49 (1998), 317-356.

Stock, J. H. and M. W. Watson, ‘‘New Indexes of Coincident and Leading Economic Indicators,’’ *NBER Macroeconomic Annual* (Cambridge: MIT Press, 1989).

---------- "Diffusion Indexes," NBER Working paper #6702 (1998).

Strongin, Steven, "The Identification of Monetary Policy Disturbances. Explaining the Liquidity Puzzle," *Journal of Monetary Economics* 35 (1995), 463-497.

White, Halbert, *Estimation, Inference and Specification Analysis* (Cambridge University Press: 1994).

Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data* (Cambridge, MA: MIT Press: 2003).

10 Appendix

PROOF OF PROPOSITION 1.

The proof follows Wooldridge (2002, Section 12.4). The objective function of the GMM estimator is

$$L\left(A\left(\hat{\boldsymbol{\Pi}}\right),\Psi\left(\hat{\boldsymbol{\Pi}}\right)\right)=\left\{\left(\tfrac{1}{T}\sum_{t=1}^{T}\hat{F}_t\left(A,\Psi\right)\right)'\Sigma^{-1}\left(\tfrac{1}{T}\sum_{t=1}^{T}\hat{F}_t\left(A,\Psi\right)\right)\right\} \tag{8.1}$$

The score function of (8.1) is

$$s_t\left(A,\Psi,\hat{\boldsymbol{\Pi}}\right)=2\nabla_{A,\Psi}\hat{F}_t\left(A,\Psi\right)\Sigma^{-1}\left(\tfrac{1}{T}\sum_{t=1}^{T}\hat{F}_t\left(A,\Psi\right)\right) \tag{8.2}$$

To show that $\left[\hat{A}\left(\hat{\boldsymbol{\Pi}}\right),\hat{\Psi}\left(\hat{\boldsymbol{\Pi}}\right)\right]$ and $\left[\hat{A}\left(\boldsymbol{\Pi}\right),\hat{\Psi}\left(\boldsymbol{\Pi}\right)\right]$ have the same asymptotic distribution, I need to prove that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_t\left(A,\Psi,\hat{\boldsymbol{\Pi}}\right)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_t\left(A,\Psi,\boldsymbol{\Pi}\right)+o_p\left(1\right) \tag{8.3}$$

Note that by construction of the moments $\nabla_{A,\Psi}\hat{F}_t\left(A,\Psi\right)$ does not depend on $\boldsymbol{X}_t$ and $\boldsymbol{\Pi}$. Given this observation, I take mean value expansion of (8.3) around $\boldsymbol{\Pi}$:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_t\left(A,\Psi,\hat{\boldsymbol{\Pi}}\right)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_t\left(A,\Psi,\boldsymbol{\Pi}\right)+E\left(\nabla_{\Pi}s\left(A,\Psi,\boldsymbol{\Pi}\right)\right)\sqrt{T}vec\left(\hat{\boldsymbol{\Pi}}-\boldsymbol{\Pi}\right)+o_p\left(1\right)$$

Hence to show equivalency of asymptotic distributions, it is sufficient to show that $E\left(\nabla_{\Pi}s\left(A,\Psi,\boldsymbol{\Pi}\right)\right)=0$. To prove this, plug $u_{it}=x_{it}-\sum_{m=1}^{n}\sum_{s=1}^{p}\pi_{s,m}x_{m,t-s}$ into (8.2). Because $\nabla_{A,\Psi}\hat{F}_t\left(A,\Psi\right)$ and Σ do not depend on $\boldsymbol{X}_t$ and $\boldsymbol{\Pi}$, observe that for elements of $\nabla_{\boldsymbol{\Pi}}F\left(A,\Psi\right)$ corresponding to diagonal elements of Ω I have:

$$E\left[\frac{\partial}{\partial\pi_{s,m}}\left\{\tfrac{1}{T}\sum_{t=1}^{T}\left(x_{it}-\sum_{m=1}^{n}\sum_{s=1}^{p}\pi_{s,m}x_{m,t-s}\right)^2-\left(\sigma_{vi}^2+\sum_{j=1}^{k}a_{ij}^2\right)\right\}\right]=$$

$$=E\left[\tfrac{1}{T}\sum_{t=1}^{T}2\left(x_{it}-\sum_{m=1}^{n}\sum_{s=1}^{p}\pi_{s,m}x_{m,t-s}\right)x_{s,m}\right]=0 \qquad \forall s,m,i,$$

by the condition $E\left(\left[X_t-\Pi_1X_{t-1}-\ldots-\Pi_pX_{t-p}\right]X_{t-k}\right)=0 \quad k=1,\ldots,p$ and the same result holds for all off-diagonal entries of Ω. Hence, $E\left(\nabla_{\boldsymbol{\Pi}}F\left(A,\Psi\right)\right)=0$ implying $E\left(\nabla_{\Pi}s\left(A,\Psi,\boldsymbol{\Pi}\right)\right)=0$.

PROOF OF PROPOSITION 2.

In proposition 1 I show that $s_t\left(A,\Psi,\hat{\boldsymbol{\Pi}}\right)=s_t\left(A,\Psi,\boldsymbol{\Pi}\right)+o_p\left(1\right)$ thus I will treat $\boldsymbol{\Pi}$ as known. To simplify notation, I define $\beta=\left[vec\left(A\right)'\quad diag\left(\Psi\right)'\right]'$ so that I can write $s_t\left(A,\Psi,\boldsymbol{\Pi}\right)\equiv s_t\left(\beta,\boldsymbol{\Pi}\right)$. The estimate $\hat{\beta}$ solves the first order condition for minimizing GMM objective function, i.e.

$$\sum_{t=1}^{T}s_t\left(\hat{\beta},\boldsymbol{\Pi}\right)=0 \tag{9.1}$$

Since by construction $s_t\left(\beta,\boldsymbol{\Pi}\right)$ is a polynomial function of β, $s_t\left(\beta,\boldsymbol{\Pi}\right)$ is twice continuously differentiable in β. Hence I can expand (9.1) around β:

$$\sum_{t=1}^{T}s_t\left(\hat{\beta},\boldsymbol{\Pi}\right)=\sum_{t=1}^{T}s_t\left(\beta,\boldsymbol{\Pi}\right)+\left(\sum_{t=1}^{T}\ddot{\boldsymbol{H}}_t\right)\left(\hat{\beta}-\beta\right) \tag{9.2}$$

where $\ddot{\boldsymbol{H}}_t$ is the Hessian of the objective function with respect to β, and each $\ddot{\boldsymbol{H}}_t$ is evaluated at a different mean value. Since by consistency of GMM $p\lim\hat{\beta}=\beta$, I have $p\lim\frac{1}{T}\sum_{t=1}^{T}\ddot{\boldsymbol{H}}_t=E\left(\boldsymbol{H}\right)$ implying by the Slutsky theorem that

$$\left(\frac{1}{T}\sum_{t=1}^{T}\ddot{\boldsymbol{H}}_t\right)^{-1}=\left[E\left(\boldsymbol{H}\right)\right]^{-1}+o_p\left(1\right) \tag{9.3}$$

provided $E\left(\boldsymbol{H}\right)$ is positive definite ($\Leftrightarrow$ A,Ψ is locally identified).

From (9.1) and (9.2), I then have $\sqrt{T}\left(\hat{\beta}-\beta\right)=\left(\frac{1}{T}\sum_{t=1}^{T}\ddot{\boldsymbol{H}}_t\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_t\left(\beta,\boldsymbol{\Pi}\right)$. By the central limit theorem,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_t\left(\beta,\boldsymbol{\Pi}\right)\xrightarrow{d}N\left(0,\Xi\right) \tag{9.4}$$

provided $E\left(s\left(\beta,\boldsymbol{\Pi}\right)\right)=0$ (i.e. model is correctly specified) and $s\left(\beta,\boldsymbol{\Pi}\right)$ has finite second moment (e.g. $E\left(\boldsymbol{u}_t^4\right)<\infty$).

Let $\Phi = \left[E\left(\boldsymbol{H}\right)\right]^{-1} \Xi \left[E\left(\boldsymbol{H}\right)\right]^{-1}$. Then by combining (9.3) and (9.4) I have by the continuous mapping theorem that

$$\sqrt{T}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \left[E\left(\boldsymbol{H}\right)\right]^{-1} \Xi \left[E\left(\boldsymbol{H}\right)\right]^{-1}\right)$$

If the optimal weighting matrix is used in the objective function, i.e. $\Sigma = E\left\{F_t\left(A,\Psi\right)F_t\left(A,\Psi\right)'\right\}$, covariance matrix Φ simplifies to $\Phi = \left(D'\Sigma^{-1}D\right)^{-1}$ where

$D = E\left(\nabla_{A,\Psi} F_t\left(A,\Psi\right)\right) = -\dfrac{\partial vech\left(AA' + \Psi\right)}{\partial\left[vec\left(A\right)' \;\; diag\left(\Psi\right)\right]'}$ is an $\left(\frac{1}{2}n\left(n+1\right)\right) \times \left(k\left(n+1\right) - r_A\right)$ matrix.

Note that $D'\Sigma^{-1}D$ is invertible if each column of *D* has a non-negative entry. Put differently, if there are indeed *k* factors so that each column of *D* has a non-zero entry, $D'\Sigma^{-1}D$ is invertible.

QED.

PROOF OF PROPOSITION 3 (CONSITENCY OF THE CRITERION).

Observe that if $\hat{k} > k_0$, then $\exists k_* > k_0 : S\left(k_*\right) \le S\left(k_0\right)$. Hence, $\Pr\left(\hat{k} > k_0\right) \le \sum_{k=k_*}^{n} \Pr\left(S\left(k\right) \le S\left(k_0\right)\right)$.

Note that $S\left(k_0\right) - S\left(k_*\right) = f\left(T\right)^{-1}\left[J\left(k_0, A, \Psi\right) - J\left(k_*, A, \Psi\right)\right] + g\left(k_*\right) - g\left(k_0\right)$ and, thus,

$\Pr\left(S\left(k_*\right) \le S\left(k_0\right)\right) \le \Pr\left(J\left(k_0, A, \Psi\right) \ge f\left(T\right)\left[g\left(k_0\right) - g\left(k_*\right)\right]\right) \xrightarrow{T\to\infty} 0$

because $f\left(T\right) \to \infty, g\left(k_*\right) < g\left(k_0\right)$ and $J\left(k_0, A, \Psi\right)$ is distributed asymptotically as $\chi^2\left(\frac{1}{2}n\left(n+1\right) - nk - k\right)$. It follows that $\lim_{T\to\infty} \Pr\left(\hat{k} > k_0\right) = 0$.

For $\hat{k} < k_0$, observe that

$$\Pr\left(S\left(\hat{k}\right) \le S\left(k_0\right)\right) = \Pr\left(f\left(T\right)^{-1}\left[J\left(\hat{k}, A, \Psi\right) - J\left(k_0, A, \Psi\right)\right] - \left[g\left(k_0\right) - g\left(\hat{k}\right)\right] \le 0\right) \xrightarrow{T\to\infty} 0$$

since $J\left(\hat{k}, A, \Psi\right)$ is bounded away from zero and $\left[g\left(k_0\right) - g\left(\hat{k}\right)\right] < 0$. Thus, $\lim_{T\to\infty} \Pr\left(\hat{k} < k_0\right) = 0$

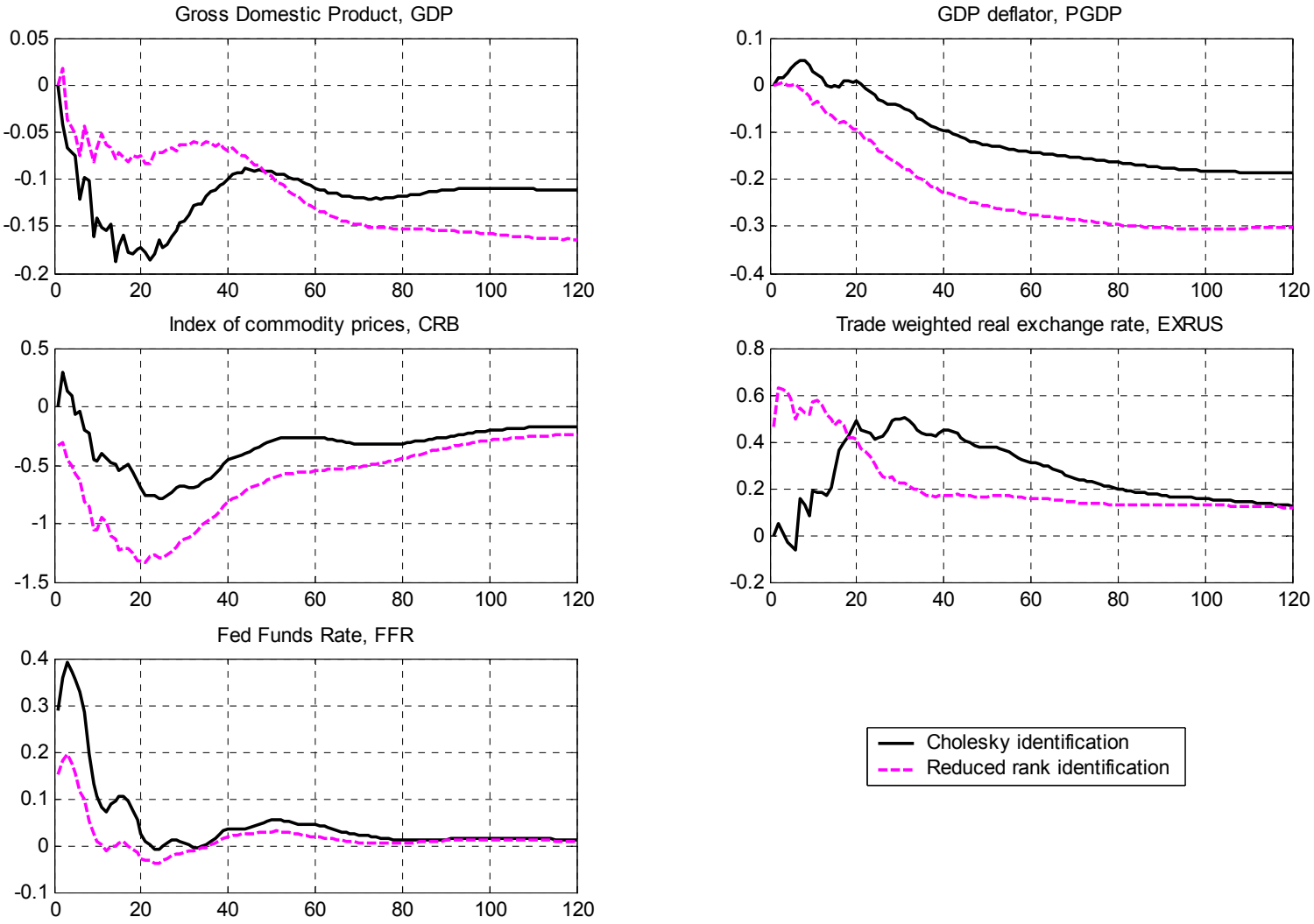
QED

Table 1. Parameter estimates of contemporaneous responses under the reduced rank identification.

variable	coefficient	Standard asymptotics		Bootstrap	
		estimate	Standard error	Bias corrected estimate	95% confidence interval
Technological shock ε^{TECH} : the first column of A in (7.2)					
RGDP	a_{11}	0.098	0.010	0.105	[0.032, 0.268]
PGDP	a_{21}	0.007	0.001	0.009	[-0.016, 0.032]
CRB	a_{31}	0.511	0.053	0.552	[0.367, 0.660]
EXRUS	a_{41}	-0.136	0.016	-0.150	[-0.331, 0.028]
FFR	a_{51}	0.217	0.022	0.254	[0.186, 0.283]
Policy shock ε^{POL} : the second column of A in (7.2)					
CRB	a_{32}	-0.257	0.043	-0.305	[-0.419, -0.104]
EXRUS	a_{42}	0.369	0.114	0.425	[0.222, 0.566]
FFR	a_{52}	0.150	0.062	0.161	[0.096, 0.239]

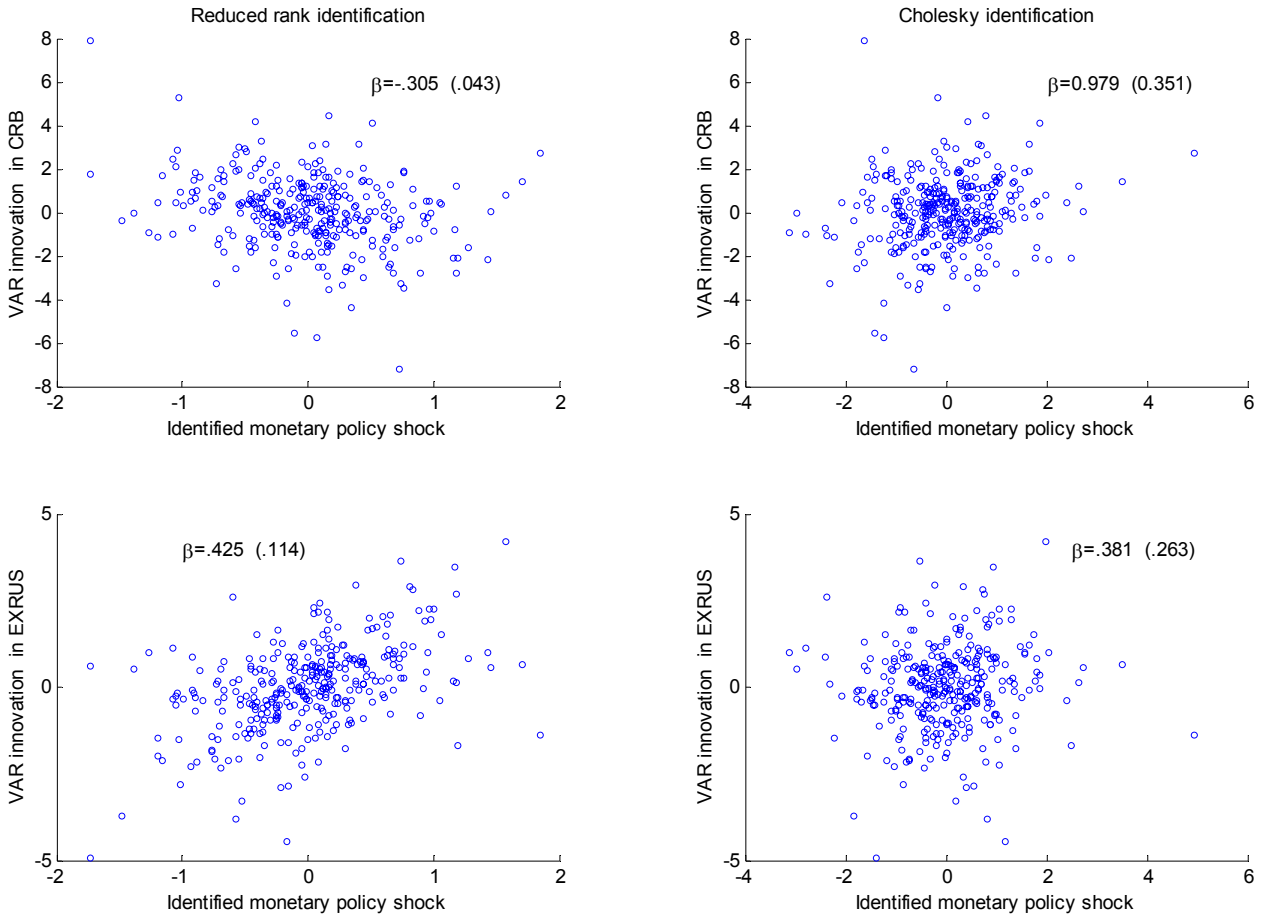
Note: The table presents estimates of the matrix A in (7.2). Bootstrap procedure is described in the text. BC_α bootstrap with re-centering bootstrap moments (2000 replications) is used.

Figure 1. Impulse response functions.



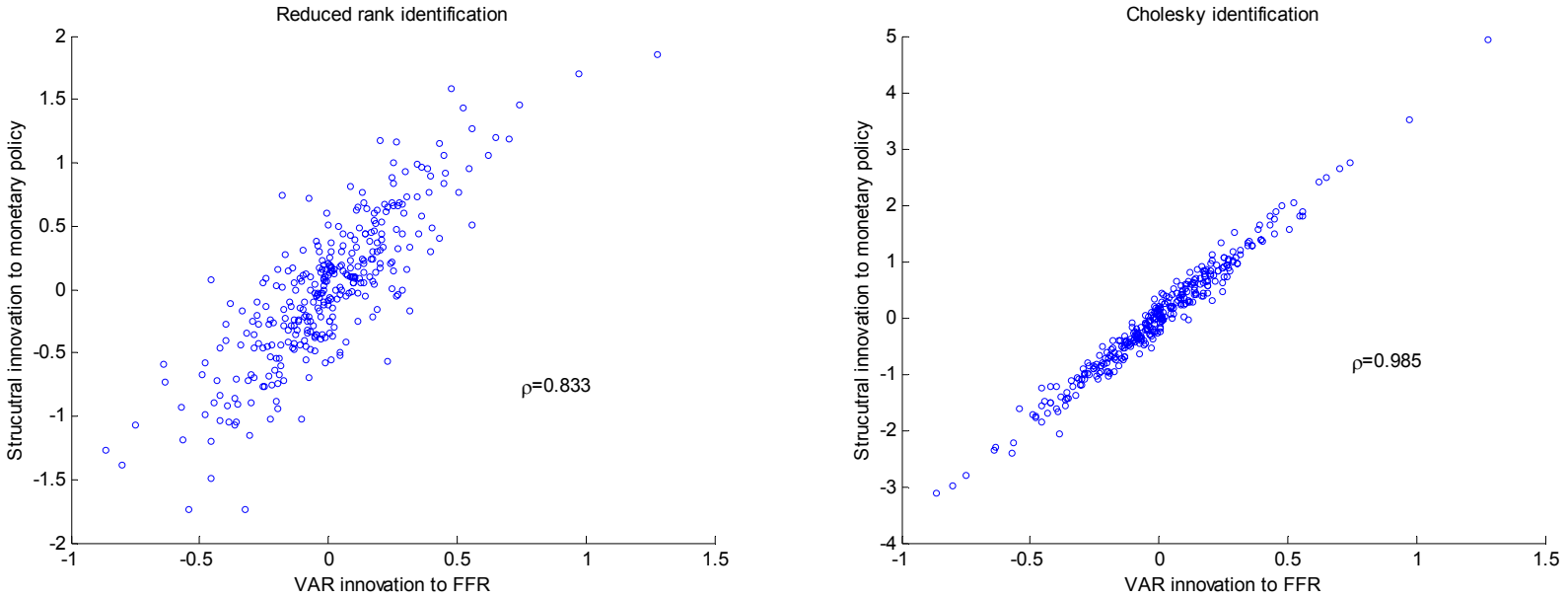
Note: The figure presents impulse response functions to a one-standard-deviation structural innovation to monetary policy. Time (the horizontal axis) is in months. The ordering in the Cholesky factorization is [RGDP,PGDP,CRB,EXRUS,FFR].

Figure 2. Response of EXRUS and CRB to identified innovation in monetary policy.



Note: β in each panel shows the estimated coefficient and associated standard error (in parentheses) from regression of a VAR residual on identified shocks to monetary policy. The ordering of variables for the Cholesky identification is [*RGDP*,*PGDP*,*FFR*,*CRB*,*EXRUS*].

Figure 3.



Note: ρ in each panel shows the estimated correlation coefficient between VAR innovations in *FFR* equation and identified monetary policy shocks. The Cholesky identification has the baseline ordering [*RGDP*,*PGDP*,*CRB*,*EXRUS*,*FFR*].

[1] If a VAR model has a few financial variables, there can be only one common innovation in the series and other shocks in these series are idiosyncratic. Factor models (e.g. Bai and Ng, 2002) typically find that this is a fairly good description of reality.

[2] Henceforth, terms "fundamental" and "structural" are used interchangeably provided there is no space for misinterpretation.

[3] For other interpretations of shocks $\boldsymbol{v}_t$ consult Gianonne et al (2004) and references therein.

[4] If I abstract from slopes in the VAR and focus on the covariance structure of $\boldsymbol{u}_t$, I consider a model similar to the error component models in Abowd and Card (1987, 1989).

[5] Note that it is possible for Ψ to have zeros on the diagonal and Ω to still be positive definite.

[6] The reduced rank identification is similar to reduced rank regressions (including cointegration), but the similarity is limited to imposing reduced rank on certain matrices. Reduced rank regressions impose reduced rank on slope matrices in VARs while the reduced rank identification imposes approximately reduced rank on the covariance matrix of innovations.

[7] Non-zero restrictions on A can also be imposed. Identification requirements for non-zero restrictions are much stricter (see Joreskog and Sorbom, 1979, Chapter 3), yet they can be satisfied in certain applications. Placing zero restrictions on the entries of the matrix A is often called exploratory, confirmatory or simple structure factor analysis. For more details see Lawley and Maxwell (1971) and Joreskog and Sorbom (1979).

[8] Condition R3 is mandated by the simple fact that any factor is identified only if at least three signals are available, otherwise one cannot separate common factor from idiosyncratic noise in the variables.

[9] For example, if k=2, it is enough to fix one of the factor axes and then the second is fixed by orthogonality to the first.

[10] It is instructive to compare (3.2) and R1-R3/R4 with identification conditions in simultaneous equations. Condition (3.2) corresponds to the order condition and conditions R1-R3/R4 correspond to the rank condition.

[11] Maximum likelihood estimator (MLE) is a popular alternative estimator. GMM and MLE are equally popular in the structural VAR literature. In contrast, factor models are typically estimated by MLE although minimum distance estimators, which include GMM, possess the same asymptotic properties (Anderson (2003) and Cragg and Donald (1995)). In comparison to MLE, GMM requires less structure but it can be sensitive to normalization.

[12] Restrictions on covariance structures can imply that certain residuals can be used as instruments to estimate parameters of the matrix A. For details on this interpretation, see Hausman et al (1983) and Hausman and Taylor (1987), and for empirical implementation Shapiro and Watson (1988). Historically, covariance structures have been used mostly for panel micro level data, e.g. Hall and Mishkin (1982) and Abowd and Card (1988, 1989).

[13] It is easy to show the same result for quasi-maximum likelihood estimator. See theorem 6.12 in White (1994).

[14] For distributions with considerable skewness and/or kurtosis, he suggests using a truncated estimator of variance, i.e. observations above or below certain threshold are excluded from the analysis.

[15] The idea of k-step bootstrap is simple and powerful. Instead of iterating optimization routines (like Gauss-Newton) until convergence, one can make only k iterations and incur $o_p(n^{-k})$ difference between parameter estimates from converged optimization and parameter estimates from k iterations of the optimization routines. The starting values in both cases are parameter estimates from initial estimation. The parameter k can be as small as 5 for most applications.

[16] Note that VAR of a sufficiently high order makes reduced error term $\boldsymbol{u}_t$ approximately serially uncorrelated. Thus, one can use bootstrap for independently distributed observations instead of block bootstraps. See Kilian (1998) and Bose (1988) for details.

[17] Note that if the model is overidentified, one has to recenter bootstrap moments as in Hall and Horowitz (1996).

[18] See section 6.8 in Basilevsky (1994) for details.

[19] Informal methods like scree plots are also often used.

[20] This may be violated sometimes since different weighting matrices are used for different number of factors. However, this is easily corrected by using the same matrix when comparing two options for k.

[21] Approximate factor models can utilize $N \to \infty$ to find the number of factors (e.g. Bai and Ng, 2002). I have fixed N equal to the number of variables in a given VAR.

[22] Degrees of freedom must be adjusted accordingly, i.e. decreased by $.5k(k\text{-}1)$.

[23] The term nk in the penalty function g is different from degrees of freedom or free parameters by a constant independent from the number of factors k.

[24] An alternative strategy is to sequentially test if k factors account for observed covariance until J-test cannot reject the null of k-factor structure or d becomes negative. A drawback of such a procedure is that there always is a possibility of type I error and it accumulates as k increases. See Cragg and Donald (1997) for details.

[25] With panel data, one can get limited identification as in Kose et al (2003). Also, Giannone et al (2004) and Forni et al (2003) using long run restrictions can identify shocks with permanent effects and shocks with transitory effects. Identification and inference through long run restrictions, however, can be very sensitive to choice of lags. In addition, economic interpretation of the shocks is limited. See Faust and Leeper (1997) for details.

[26] There have been other attempts to merge factor models and VARs (Bernanke and Boivin (2003) and Bernanke et al (2004)). These are, however, focused on controlling for larger information sets than those represented by the lags of endogenous variables in a VAR.

[27] These two series are from Bernanke and Mihov (1998).

[28] Other orderings, e.g. [*RGDP*,*PGDP*,*FFR*,*CRB*, *EXRUS*], produce essentially identical results.

[29] Giannone et al (2004) find in the context of factor models that two-fundamental-shock structure is a good description of the US economy .

[30] I do 2000 bootstrap replications with re-centering bootstrap moments. I use BC_{α} –method to construct confidence intervals to have good coverage properties and have transformation respecting confidence intervals (see Efron and Tibshirani (1993) for details).

[31] I focus on point estimates and do not present confidence bounds for the sake of clarity.

[32] Structural shocks are normalized to have unit variance.

[33] This conclusion is confirmed by unreported Monte Carlo simulations when the Cholesky identification is applied to data generating process of $\boldsymbol{u}_t$ characterized by reduced rank structure.

[34] To emphasize the point, I use the baseline ordering [RGDP,PGDP,CRB,EXRUS,FFR] that minimizes correlation between VAR residuals in FFR equation and the Cholesky-identified structural shocks in monetary policy.