

REDUCED-RANK IDENTIFICATION OF STRUCTURAL SHOCKS IN VARs

Yuriy Gorodnichenko

Department of Economics
University of Michigan

First draft: April 22, 2003
This version: April 20, 2005

Abstract

This paper integrates imposing a factor structure on residuals in vector autoregressions (VARs) into structural VAR analysis. Identification, estimation and testing procedures are discussed. The paper applies this approach to the well-known problem of studying the effects of monetary policy in open economy VAR models. The use of factor structure in identifying structural shocks is shown to resolve three long-standing puzzles in VAR literature. First, the price level does not increase in response to a monetary tightening. Second, the exchange rate appreciates on impact and then gradually depreciates. Hence, no price level and exchange rate puzzles are found. Third, monetary policy shocks are much less volatile than suggested by standard VAR identification schemes. In addition, the paper suggests that the apparent weak contemporaneous cross-variable responses and strong own responses in structural VARs can be an artifact of identifying assumptions and vanish after imposing a factor structure on the shocks.

Keywords: Vector autoregressions, identification, factor structure, monetary policy.

JEL classification: E520, C320

I am grateful to Robert Barsky, Oli Cobion, Atsushi Inoue, Nikolay Iskrev, Hong Li, participants of the Midwest Econometrics Meeting (2004) and Econometric Society Winter Meeting (2005) and especially Lutz Kilian, Serena Ng, and Matthew Shapiro for comments and suggestions. I thank Lutz Kilian for providing his program codes. All remaining errors are mine. Correspondence: ygorodni@umich.edu or Yuriy Gorodnichenko, Department of Economics, University of Michigan, 611 Tappan St, Ann Arbor, MI 48109.

Copyright 2004 by Yuriy Gorodnichenko. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including notice, is given to the source.

1 Introduction

Vector autoregressions (VARs) are often used to provide a set of stylized facts about responses of macroeconomic variables to structural shocks such as innovations in technology and monetary policy. These facts serve as a yardstick in evaluating or calibrating theoretical business cycle models (e.g., Rotemberg and Woodford 1997, Woodford 2003, Christiano, Eichenbaum, and Evans 2005). However, the mapping from VARs to business cycle models is not straightforward.

First, in VARs, whether structural or not structural, the number of structural shocks is equal to the number of variables. For instance, in VARs with 10 variables there are 10 structural shocks while business cycle models tend to have only a few structural shocks such as innovations in technology, preferences, or policy. Hence, VARs can produce too many structural shocks to be economically meaningful.¹ This mismatch in the number of structural shocks is important because identification of key structural shocks—that is, shocks that constitute the primary focus of the analysis—is sensitive to inclusion of other, possibly less important structural shocks (e.g., Faust 1998, Faust and Rogers 2003). For every additional variable, it may be increasingly difficult to find credible identifying restrictions that differentiate the key structural shocks from other shocks. In fact, many VAR models (e.g., Blinder and Bernanke 1992) are silent about unimportant, or “nuisance,” structural shocks thus leaving the reader wondering what these shocks are and how they are different from the key structural shocks.

Second, Sims (1998) observes that weak cross-variable responses in VARs are a robust stylized fact. Indeed, diagonal entries of the covariance matrix of VAR residuals are typically very large relative to off-diagonal entries and VARs tend to have difficulties with matching responses of models with large contemporaneous responses to structural shocks. For example, Eichenbaum and Evans (1995) find that their identified VARs cannot reproduce exchange rate overshooting predicted by the Dornbusch (1976) model. This tendency to find attenuated responses can lead to rejection of

¹ Alternatively, business cycle models can have too few shocks to match the data.

correct theoretical models with large cross-variable responses, motivate models with significant rigidities, and result in the conclusion that the policymaker needs to policy instruments a lot to move target variables a little. Whether stickiness is induced by the identification scheme or is a genuine regularity in the data remains an open question. A related third problem is the high correlation of identified monetary policy shocks with VAR residuals for policy instrument variables such as the fed funds rate. Given the noisiness of VAR residuals, this high correlation is puzzling because it indicates that policymakers often move instruments by large amounts at random.

This paper reconciles these facts by identifying and estimating a VAR with the number of structural shocks being less than the number of variables in the VAR. The key insight is to treat structural shocks as factors so that structural shocks are not contaminated with series-specific noise. The proposed method not only determines the number of structural shocks but also uniquely identifies these shocks using short- and/or long-run restrictions even when the number of the shocks is greater than one. Furthermore, all standard VAR tools are available for dynamic analyses of economic models under this identification method—for example, impulse response functions are easily constructed. This paper uses the generalized method of moments (GMM) framework to provide a unified treatment of the factor structure of innovations in VARs (FSVARs). This paper extends FSVAR framework to analyses other than studies of propagation of shocks across regions and industries (Norrbin and Schlagenhauf 1988, 1996, Altonji and Ham 1990, Clark and Shin 2000, Stock and Watson 2003). Specifically, I argue that series in FSVARs need not have regional or sectoral structure and factor structure in VAR residuals can arise naturally in economic problems.

When compared with structural VARs, FSVARs have several advantages. First, inclusion of additional variables in a VAR does not imply the existence of additional structural shocks and, hence, identifying restrictions for key structural shocks are not modified to distinguish the key structural shocks from other shocks. Second, structural shocks in FSVARs can be directly mapped to structural innovations in business cycle models. Third, the volatility of structural shocks is

disentangled from the volatility of VAR residuals. I contrast FSVARs with structural VARs and show that imposing a factor structure on VAR residuals can resolve the VAR puzzles that I mention above.

This approach is different in several important respects from factor augmented VARs (FAVARs) considered in Bernanke and Boivin (2003) and Bernanke, Boivin and Elias (2005). First, their modification is focused on incorporating large information sets into an otherwise standard VAR model, and factors are extracted from a set of auxiliary variables not included in the VAR. Second, the number of structural shocks in a FAVAR is equal to the number of variables in the FAVAR. Third, factors in FAVARs are identified only up to a rotation and the factors are ordered first in recursive identification. As I show below, these differences can have important implications for the consistency of estimated contemporaneous responses of variables and, thus, impulse responses.²

To illustrate the method, I apply it to the problem of identifying the effects of monetary policy in a model similar to Eichenbaum and Evans (1995). Using US data, I construct impulse responses of key macroeconomic variables to a monetary tightening. 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 falls gradually 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 that have been documented by Sims (1992) and Eichenbaum and Evans (1995). In addition, under my identification the policy shocks are less noisy than under other identification schemes. I argue that because conventional identification schemes do not distinguish structural shocks from

² Clark and Shin (2000) and Stock and Watson (2003) contrast FSVARs with dynamic factor models (Geweke 1977, Quah and Sargent 1993, Stock and Watson 1998, Forni et al 2000, Forni and Lippi 2001, Giannone, Reichlin, and Sala 2004, Forni, Lippi and Reichlin 2003).

series-specific innovations, the latter can attenuate or even reverse contemporaneous responses of macroeconomic variables to innovations in monetary policy.

The rest of the paper is structured as following. In section 2, I present a unified framework for analysis of FSVARs. I discuss identification, estimation, and inference. In section 3, I present an empirical application to the analysis of monetary policy and contrast conventional identification schemes with the proposed identification scheme. I conclude in section 4.

2 The Method

2.1 Setup

Consider the standard VAR model:

$$\mathbf{X}_t = \sum_{i=1}^p \Pi_i \mathbf{X}_{t-i} + \mathbf{u}_t \quad (1)$$

where \mathbf{X}_t is a vector of q variables, \mathbf{u}_t is the vector of i.i.d. reduced-form errors, Π_1, \dots, Π_p are matrices of conformable sizes.³ The bridge between \mathbf{u}_t and structural shocks $\boldsymbol{\varepsilon}_t$ is embodied in the relationship $\mathbf{u}_t = A\boldsymbol{\varepsilon}_t$ where the matrix A , which is not identified unless one imposes enough restrictions on its elements,⁴ summarizes contemporaneous relationships in the economy. This matrix is of central importance in the VAR framework because incorrect identification of A invalidates all subsequent economic analyses, including structural impulse response functions and associated confidence bounds.

In contrast to standard structural VARs that postulate the number of structural shocks be equal to the number of variables in the VAR (i.e., A is a square matrix), FSVARs assume that reduced-form innovations \mathbf{u}_t are a linear combination of the structural shocks $\boldsymbol{\varepsilon}_t$ and of series-specific idiosyncratic noise \mathbf{v}_t . Specifically, I replace $\mathbf{u}_t = A\boldsymbol{\varepsilon}_t$ by

³ In contrast, FAVARs can be represented $\mathbf{X}_t = \Pi(L)\mathbf{X}_{t-1} + \Upsilon(L)\mathbf{F}_{t-1} + \mathbf{u}_t$ where \mathbf{F}_t is a vector of dynamic factors. Again, note that factors in these models are identified only up to a rotation and the factors are ordered first in recursive identification.

⁴ For examples of such restrictions see Christiano, Eichenbaum and Evans (1999).

$$\mathbf{u}_t = \begin{bmatrix} u_{1t} \\ \vdots \\ u_{nt} \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nk} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{kt} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ \vdots \\ v_{nt} \end{bmatrix} = A\boldsymbol{\varepsilon}_t + \mathbf{v}_t, \quad (2)$$

where A is an $(n \times k)$ matrix and

$$E(\boldsymbol{\varepsilon}_t) = 0 \quad \forall t, \quad E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = I_k \quad \forall t, \quad E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s') = 0 \quad \forall s \neq t, \quad (3)$$

$$E(\mathbf{v}_t) = 0 \quad \forall t, \quad \Psi \equiv E(\mathbf{v}_t \mathbf{v}_t') = \text{diag}(\sigma_{v1}^2, \sigma_{v2}^2, \dots, \sigma_{vn}^2) \quad \forall t, \quad E(\mathbf{v}_t \mathbf{v}_s') = 0 \quad \forall s \neq t, \quad (4)$$

$$E(\boldsymbol{\varepsilon}_t \mathbf{v}_s') = 0 \quad \forall s, t. \quad (5)$$

In addition to the standard normalization assumption (3), assumption (4) postulates that shocks in \mathbf{v}_t are contemporaneously and intertemporally uncorrelated and may have different variances. Condition (5) requires \mathbf{v}_t to be uncorrelated with $\boldsymbol{\varepsilon}_t$ at all leads and lags. Representation (2) and conditions (3)-(5) essentially replicate assumptions necessary for the existence of a factor representation of the series \mathbf{u}_t (e.g., Anderson and Rubin 1956). Technically, the presence of noise shocks guarantees that the matrices $\Omega \equiv E(\mathbf{u}_t \mathbf{u}_t')$ and, thus, $E(\mathbf{X}_t \mathbf{X}_t')$ have full rank. Because $\text{rank}(A) < n$, I refer to this approach as reduced rank (RR) identification.

In previous applications of FSVARs (e.g., Altonji and Ham 1990, Stock and Watson 2003), $\boldsymbol{\varepsilon}_t$ is typically interpreted as a global/national/economy-wide shock while \mathbf{v}_t is understood as a collection of country/region/industry-specific shocks. Critically, the noise shock \mathbf{v}_t does not affect the contemporaneous cross-variable responses embodied in the matrix A . In contrast, in standard VARs without regional or sectoral patterns, the series-specific shock \mathbf{v}_t may be a shock induced by errors in decision rules arising from expectation errors, partial/incomplete information sets, noise/chartist traders, myopic consumers, etc. To see the point, note that solutions to rational expectations (RE) models can be generically formulated as follows (e.g., Anderson and Moore 1985):

$$\mathbf{D}_t = B_1 \mathbf{S}_t + A \boldsymbol{\varepsilon}_t, \quad (6)$$

$$\mathbf{S}_{t+1} = B_2 \mathbf{S}_t + B_3 \mathbf{D}_t, \quad (7)$$

where \mathbf{D}_t is a vector of jump (non-predetermined) variables (e.g., consumption, output, asset prices), \mathbf{S}_t is a vector of state (predetermined) variables (e.g., capital, habit), $\boldsymbol{\varepsilon}_t$ is a vector of structural innovations (e.g., innovation in technology), A , B_1 , B_2 , B_3 are matrices of conformable sizes. The matrix A embodies contemporaneous responses of variables to structural innovations. Equations in (6) summarize decision rules. Equations in (7) capture the evolution of state variables.

In practice, there can be many forces that prevent agents from acting according to the prescribed decision rules. For example, expectations errors, measurement errors, heterogeneous information sets (e.g., consumers and the central banker have different information sets), myopia and other forms of irrational behavior are possible reasons for why \mathbf{D}_t may depart from optimal responses prescribed by (6).⁵ To accommodate these errors, I follow Sims (2002) by augmenting the decision rule (6) with shocks \mathbf{v}_t so that

$$\mathbf{D}_t = B_1 \mathbf{S}_t + A \boldsymbol{\varepsilon}_t + \mathbf{v}_t. \quad (8)$$

This recognizes the fact that error terms in equilibrium models are more naturally attached to equations than to variables (Sims 2002). After combining (7) and (8), one can find the dynamics of \mathbf{D}_t :

$$\mathbf{D}_t = B_1 (I - B_2 L)^{-1} B_3 \mathbf{D}_{t-1} + A \boldsymbol{\varepsilon}_t + \mathbf{v}_t = \Pi(L) \mathbf{D}_{t-1} + \mathbf{u}_t, \quad (9)$$

where L is the lag operator, $B_1 (I - B_2 L)^{-1} B_3$ is approximated with a lag polynomial $\Pi(L)$ of a suitable order, and $\mathbf{u}_t \equiv A \boldsymbol{\varepsilon}_t + \mathbf{v}_t$ is a vector of reduced-form innovations in \mathbf{D}_t . In this VAR, the lag polynomial $\Pi(L)$ does not depend on the structure of \mathbf{u}_t . In Appendix B, I provide a stylized economic model with imperfect information in the spirit of Lucas (1972) that generates a factor structure in VAR residuals.

Given (2)-(5), the covariance matrix of \mathbf{u}_t can be decomposed as:

⁵ Note that if these errors are correlated across variables then these errors are analogous to genuine structural shocks in the sense that multiple variables are affected by these errors. Thus, I distinguish errors that are uncorrelated across variables from errors that are correlated across variables. The latter are treated as part of $\boldsymbol{\varepsilon}_t$.

$$\Omega = E(A\varepsilon_t\varepsilon_t'A) + E(\mathbf{v}_t\mathbf{v}_t') = AA' + \Psi, \quad (10)$$

where AA' and Ψ measure the portions explained by structural and noise shocks, respectively. Note that, consistent with (9), decomposition of the covariance matrix Ω and estimation of Π_1, \dots, Π_p are disentangled. Thus, 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 (10).

Whenever imposing factor structure on \mathbf{u}_t is appropriate, it has several advantages relative to the standard case with $\dim(\varepsilon_t) = \dim(\mathbf{u}_t)$ and $\mathbf{v}_t \equiv 0$. First, specification (2) can have only a few structural shocks even in VARs with many variables. In contrast, every additional variable in a VAR with a square A requires “inventing” another structural shock to avoid stochastic singularity in \mathbf{X}_t .

Second, the number of structural shocks in VARs with square A can affect identification of key structural shocks such as innovations in monetary policy that are the primary focus of the study. Indeed, identifying restrictions are designed to distinguish the key structural shocks from nuisance structural shocks—i.e., structural shocks that are not the focus of a study—and as the number of nuisance structural shocks increases the researcher has to impose restrictions on *all* structural shocks which may be difficult to justify. For example, to achieve identification, Bernanke and Mihov (1998) assume that asset prices do not respond contemporaneously to innovations in monetary policy. In many cases (e.g., in VARs with a block-recursive A), nuisance structural shocks are not even named or interpreted so that it can be unclear if the identifying restrictions can credibly differentiate the key structural shocks from nuisance structural shocks. Faust (1998) and Faust and Rogers (2003) illustrate this point eloquently. In contrast, FSVARs solve this problem by introducing noise shocks \mathbf{v}_t thus keeping the number of structural shocks as small as necessary.

Third, it is plausible that reduced-form VAR residuals \mathbf{u}_t absorb not only structural shocks ε_t but also idiosyncratic innovations in the series \mathbf{v}_t . For example, financial variables are often found to be well-described by factor models (e.g., Bai and Ng 2002). By pooling structural and idiosyncratic

shocks, conventional VARs can give a misleading picture about the true responses of the economy to structural shocks because these types of shocks have different cross-variable responses (see section 3). In contrast, specification (2) differentiates ε_t and ν_t explicitly. Furthermore, this distinction separates noisiness of VAR residuals from volatility of structural shocks, i.e., changes in a VAR residual are not necessarily due to structural shocks.

2.2 Identification

The fundamental problem of identifying covariance structure (10) with $k \geq 2$ is in the rotational equivalence of solutions to (10). Specifically, 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 that $AA' = BB'$ with $B=AM$. To uniquely (up to sign) identify A , one has to impose restrictions on A so that there is no orthonormal rotation matrix M (unless $M=I$) such that $B=AM$ satisfies the restrictions imposed on A . I focus on zero restrictions on some entries of A .⁶ In the case of orthogonal structural shocks, the necessary condition to identify A is given by Lawley and Maxwell (1974): $r_A \geq \frac{1}{2}k(k-1)$, where r_A denotes the number of imposed zero restrictions. The model is overidentified if $d = \frac{1}{2}[(n-k)^2 - n - k] + r_A - \frac{1}{2}k(k-1) > 0$, where the parameter d denotes degrees of freedom.

For local identification of A, Ψ it is sufficient that the Jacobian of $vech(AA' + \Psi)$ has full rank (Bekker, Merkens and Wansbeek 1994, Corollary 4.2.1, p. 78). On the other hand, global identification requires proving that the only admissible rotation matrix M is the identity matrix. Anderson and Rubin (1956) give sufficient conditions for zero restrictions to uniquely identify A .⁷ For the r^{th} column of an $(n \times k)$ matrix A —i.e., the r^{th} factor—to be uniquely identified three requirements must be satisfied:

⁶ 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 Basilevsky (1994).

⁷ There are other sufficient conditions. See Anderson and Rubin (1956) and Algina (1980).

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 ($k \times k$) matrix, which is similar to recursive identification in VARs. In previous applications of FSVARs, identification is guaranteed because A is lower diagonal (e.g., Stock and Watson 2003) or A has a special structure such as $A = \left[A_0 \mid \text{diag}(a_1, \dots, a_q) \otimes 1_{p \times 1} \right]$ where A_0 is an ($pq \times 1$) matrix that embodies responses to national shocks, a_1, \dots, a_q reflect responses to regional shocks, q is the number of regions, and p is the number of industries in a region (e.g., Altonji and Ham 1990).⁹

Note that conditions R1-R3 allow partially identified 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. In practical work, one can put arbitrary constraints on the entries of unidentified columns (e.g., R4 for unidentified columns) without affecting the identified factors to reach identification.

Non-zero restrictions on entries of A can also provide global identification, although identification requirements for non-zero restrictions are much stricter (see Algina 1980). For instance, long-run restrictions as in Shapiro and Watson (1988) and Blanchard and Quah (1989) or

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. Bekker, Merken and Wansbeek (1994) compare necessary and sufficient conditions for covariance structures with the respective conditions for simultaneous equation models.

⁹ In the latter case, identification follows from conditions in Algina (1980) because the restrictions are both zero and non-zero.

short run restrictions as in Bernanke and Mihov (1998) require that linear combinations of the entries of A are equal to known constants.¹⁰

2.3 Estimation and inference

To start with a simple case, suppose Π_1, \dots, Π_p are known. The goal is to decompose \mathbf{u}_t into factors and uncorrelated errors, i.e., find matrices A, Ψ such that $\Omega = AA' + \Psi$. Note that, even in VARs including nonstationary variables, \mathbf{u}_t is stationary so that standard estimation and inference techniques can be applied. To simplify notation, define $F_t(A, \Psi) = \text{vech}(\mathbf{u}_t \mathbf{u}_t' - AA' - \Psi)$ and note that under my assumptions $E[F_t(A, \Psi)] = \text{vech}(\Omega - AA' - \Psi) = 0$. Using this moment condition, I follow Bernanke (1986) and use GMM to estimate $\hat{A}, \hat{\Psi}$ as

$$\hat{A}, \hat{\Psi} = \arg \min_{A, \Psi} \left\{ \left(\frac{1}{T} \sum_{t=1}^T F_t(A, \Psi) \right)' W^{-1} \left(\frac{1}{T} \sum_{t=1}^T F_t(A, \Psi) \right) \right\}, \quad (11)$$

where W is the optimal weighting matrix of conformable size.^{11,12} The resulting estimates are consistent, efficient and asymptotically normally distributed (Anderson 2003). If the model is overidentified, one can use the J -statistic, a byproduct of the GMM estimation, to test overidentifying restrictions imposed on A and Ψ (specification test). The J -statistic does not depend on whether the model is uniquely identified since all optimal solutions to (11) yield the same minimum. I use this fact to determine the rank of A , i.e., the number of factors, even when A is partially identified (see Appendix A).

¹⁰ It should be noted that it is difficult (if possible) to identify factors in dynamic factor models unless one uses long run restrictions like Giannone et al (2004) or has a natural ordering of factors (e.g., global/national/regional/local) like Kose, Otrok and Whiteman (2003).

¹¹ Since (11) entails non-linear optimization in entries of A , numerical optimization of the objective function may require checking different starting values to verify that the solution is global optimum. To verify local identification, I suggest checking if the Jacobian of the objective function has full rank.

¹² Alternatively, one can use maximum likelihood estimator (MLE) to find A and Ψ . In this case, the solution minimizes the following log likelihood: $\log L = -\frac{T}{2} \log |AA' + \Psi| - \frac{1}{2} \sum_{t=1}^T \left[\mathbf{u}_t' (AA' + \Psi)^{-1} \mathbf{u}_t \right]$. Asymptotically, MLE and GMM are both consistent and normally distributed under suitable assumptions (Anderson 2003). Although GMM requires less structure than MLE, in small samples MLE can perform somewhat better than GMM (see Clark 1996). I suggest using GMM to estimate parameters matrices A, Ψ because it conveniently combines estimation and inference.

Of course, $\boldsymbol{\Pi} \equiv \{\Pi_1, \dots, \Pi_p\}$ is not known in applications and only an OLS estimate $\hat{\boldsymbol{\Pi}}$ is typically available. Although this does not affect the consistency of $\hat{A}, \hat{\Psi}$, one generally needs to adjust standard errors of $\hat{A}, \hat{\Psi}$ for the uncertainty introduced at the first stage when $\hat{\boldsymbol{\Pi}}$ is estimated (Pagan 1984). Fortunately, the following proposition shows that the standard errors of $\hat{A}, \hat{\Psi}$ need not be adjusted for asymptotic inference.

Proposition 1

Suppose A, Ψ are locally identified (i.e., $E[\nabla_{A, \Psi} F(A, \Psi)]$ has full rank). Then under regularity conditions, $[\hat{A}(\hat{\boldsymbol{\Pi}}), \hat{\Psi}(\hat{\boldsymbol{\Pi}})]$ and $[\hat{A}(\boldsymbol{\Pi}), \hat{\Psi}(\boldsymbol{\Pi})]$ are asymptotically normal with the same covariance matrices.

Proof: see Appendix C.

This proposition proves that, while making standard asymptotic inference about A , one can ignore the fact that $\hat{\Pi}_1, \dots, \hat{\Pi}_p$ are estimates.¹³ This greatly simplifies estimation and inference for A and Ψ : one only needs to estimate the VAR by LS and in the GMM part treat the estimated residuals $\hat{\boldsymbol{u}}_t$ as population counterparts \boldsymbol{u}_t . In Appendix D, I describe the bootstrap method I use to compute bias-corrected GMM estimates, to improve the accuracy of inference and, thus, to enhance the finite sample properties of the GMM (see Horowitz 1998). 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. Anderson (2003, Section 14.7) shows that structural shocks can be estimated as $\hat{\boldsymbol{\varepsilon}}_t = (I_n + \hat{A}'\hat{\Psi}^{-1}\hat{A})^{-1} \hat{A}'\hat{\Psi}^{-1}\boldsymbol{u}_t$, which can be interpreted as the posterior estimate of $\boldsymbol{\varepsilon}_t$ given VAR residuals \boldsymbol{u}_t . Variance decomposition and impulse response functions are easily constructed from $MA(\infty)$ representation $\boldsymbol{X}_t = \boldsymbol{\Theta}(L)(A\boldsymbol{\varepsilon}_t + \boldsymbol{v}_t)$ where $\boldsymbol{\Theta}(L)$ is a lag polynomial. Confidence bounds

¹³ This result cannot be extended to cases that use long-run restrictions to identify structural shocks. In such cases, one can use bootstrap to find correct standard errors and confidence intervals.

for impulse responses and variance decompositions can be computed using procedures as in Kilian (1999) or Inoue and Kilian (2002).¹⁴

The reduced rank (RR) identification possesses 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, permits partial identification of A , and preserves simple and computationally undemanding estimation and inference. More importantly, the RR identification can rationalize 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 decompositions of Ω result in weak cross-variable responses. On the other hand, business cycle models tend to produce strong contemporaneous responses thus making empirical stickiness a puzzle. In the next section, I use a specific example to demonstrate how stickiness can arise in models where innovations have factor structure.

3 Application: identifying the effects of monetary policy

I apply the reduced rank (RR) identification to the classical problem of determining the effects of unanticipated innovations to monetary policy. The VAR example is based on models used by Eichenbaum and Evans (1995) and Kim and Roubini (2000). I compare the RR identification with the recursive Cholesky identification as a prototype of all identification schemes.

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

¹⁴ For Bayesian approaches see Sims and Zha (1999).

fed funds rate (FFR), i.e., $\mathbf{R}=[FFR]$. I assume that there are two shocks in this economy $\boldsymbol{\varepsilon}_t = [\boldsymbol{\varepsilon}_t^T, \boldsymbol{\varepsilon}_t^P]$ where $\boldsymbol{\varepsilon}_t^P$ is an innovation in monetary policy and $\boldsymbol{\varepsilon}_t^T$ is a technology innovation.¹⁵

To identify the effects of monetary policy, I follow the structural VAR literature (e.g., Bernanke and Blinder 1992, Eichenbaum and Evans 1995, Bernanke and Mihov 1998) and assume a block recursive structure of the matrix A linking unobserved structural shocks to observed VAR residuals.¹⁶ My baseline ordering is $[RGDP, PGDP, CRB, EXRUS, FFR]$. Other orderings, e.g., $[RGDP, PGDP, FFR, CRB, EXRUS]$, produce essentially identical results. For concreteness, consider the following relationship between reduced-form residuals and structural innovations:

$$\mathbf{u}_t = \begin{bmatrix} u_t^{RGDP} \\ u_t^{PGDP} \\ u_t^{CRB} \\ u_t^{EXRUS} \\ u_t^{FFR} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} & a_{34} & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} & 0 \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_t^T \\ \boldsymbol{\varepsilon}_{2t} \\ \boldsymbol{\varepsilon}_{3t} \\ \boldsymbol{\varepsilon}_{4t} \\ \boldsymbol{\varepsilon}_t^P \end{bmatrix} = \begin{bmatrix} A_{YY} & 0 & 0 \\ A_{VY} & A_{VV} & 0 \\ A_{RY} & A_{RV} & A_{RR} \end{bmatrix} \boldsymbol{\varepsilon}_t = \mathbf{A}\boldsymbol{\varepsilon}_t, \quad (12)$$

where A_{YY} , for example, indicates contemporaneous interactions among variables in the \mathbf{Y} -block and A_{RY} indicates the contemporaneous effect of innovations in the \mathbf{Y} -block variables on variables in the \mathbf{R} -block. Since this identification requires equality of the number of shocks and variables, the researcher has to introduce, name and interpret “nuisance” structural shocks $\boldsymbol{\varepsilon}_{2t}, \boldsymbol{\varepsilon}_{3t}, \boldsymbol{\varepsilon}_{4t}$. 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.¹⁷

A recursive structure of A has been typically justified by minimum delay restrictions (e.g., Bernanke and Blinder 1992). Although no response of the \mathbf{Y} -block variables to shocks in the \mathbf{R} -block variables appears to be a reasonable assumption, the assumption that asset prices do not respond to policy changes is dubious. In fact, any block recursiveness of $[\mathbf{V}, \mathbf{R}]$ sub-block can be disputed

¹⁵ Giannone et al (2004) find in the context of factor models that two-structural-shock structure is a good description of the US economy.

¹⁶ Relevant non-recursive identifications can be found in Cushman and Zha (1997) and Kim and Roubini (2000).

¹⁷ Many shocks may satisfy identifying restrictions in (12) and thus these shocks have to be grouped under the rubrics “monetary policy shock” and “technology shock.” Blanchard and Quah (1989) discuss the implications of such a grouping.

because the policymaker can readily observe assets prices and react to them while asset prices are free to adjust in response to policy changes.

In contrast, the RR identification imposes the following structure of the matrix A :

$$\mathbf{u}_t = \begin{bmatrix} u_t^{RGDP} \\ u_t^{PGDP} \\ u_t^{CRB} \\ u_t^{EXRUS} \\ u_t^{FFR} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & 0 \\ \hline a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \end{bmatrix} \begin{bmatrix} \mathcal{E}_t^T \\ \mathcal{E}_t^P \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \\ v_{5t} \end{bmatrix} = \begin{bmatrix} A_{YY} & 0 \\ \hline A_{VY} & A_{VR} \\ \hline A_{RY} & A_{RR} \end{bmatrix} \boldsymbol{\varepsilon}_t + \mathbf{v}_t = A\boldsymbol{\varepsilon}_t + \mathbf{v}_t, \quad (13)$$

where I allow noise \mathbf{v}_t to be in each equation to keep the specification flexible. The matrix A is globally uniquely identified by the condition R4. From (13), one can see that I continue to assume that structural policy shocks cannot contemporaneously affect variables in the \mathbf{Y} -block. However, in contrast to (12), I assume that policy innovations affect variables in the \mathbf{V} -block. I also assume that there are no structural shocks associated with asset prices. This assumption indicates that asset prices respond only to structural innovations (i.e., innovations in policy and technology), idiosyncratic innovation in asset prices (e.g., “bubble” component, noise traders), and do not respond to noise innovations in other variables.

Note that a similar structure can be generated by VARs that order the policy block \mathbf{R} before the asset prices block \mathbf{V} . The similarity is, however, superficial. In those VARs, asset prices respond to *all* innovations in \mathbf{Y} and \mathbf{R} variables. In the RR identification, asset prices respond to innovations in *policy* and *technology* and *do not* respond to *idiosyncratic noise in \mathbf{Y} and \mathbf{R} 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 funds rate series. In other words, the RR framework separates structural shocks from idiosyncratic noise while conventional identification schemes do not.

To see the importance of this point, consider the model in which structural shocks and VAR residuals are linked in population as follows:

$$\mathbf{u} = \begin{bmatrix} u^Y \\ u^R \\ u^V \end{bmatrix} = \begin{bmatrix} a_{YT} & 0 \\ a_{RT} & a_{RP} \\ a_{VT} & a_{VP} \end{bmatrix} \begin{bmatrix} \varepsilon^T \\ \varepsilon^P \end{bmatrix} + \begin{bmatrix} v^Y \\ v^R \\ v^V \end{bmatrix} = A\varepsilon + \mathbf{v}, \quad (14)$$

where Y , V and P denote blocks of variables (one variable per block), \mathbf{u} are reduced-form VAR residuals, ε are structural shocks with unit variance, and \mathbf{v} are idiosyncratic noise in the series with the covariance matrix $\Psi = \text{diag}(\sigma_Y^2, \sigma_V^2, \sigma_P^2)$. I am interested in the contemporaneous response of the variables to an innovation in monetary policy.

By construction, the reduced rank identification has factor structure and, therefore, correctly recovers A in (14). Specifically, the contemporaneous responses of the variables to a unit innovation in ε^T and ε^P are $[a_{YT}, a_{RT}, a_{VT}]'$ and $[0, a_{RP}, a_{VP}]'$, respectively. Note that if $a_{RP}a_{VP}$ is positive (negative), series move in the same (opposite) direction in response to a monetary policy shock ε^P .

Now consider the Cholesky decomposition that ignores the factor structure. Denote the Cholesky-identified structural shocks with $\varepsilon^{T,CH}$ and $\varepsilon^{P,CH}$. The vector of the contemporaneous responses to $\varepsilon^{T,CH}$ is

$$\begin{bmatrix} \text{var}(u^Y)^{1/2} \\ \text{cov}(u^Y, u^R) \cdot \text{var}(u^Y)^{-1/2} \\ \text{cov}(u^Y, u^V) \cdot \text{var}(u^Y)^{-1/2} \end{bmatrix} = \begin{bmatrix} (a_{YT}^2 + \sigma_Y^2)^{1/2} \\ a_{YT}a_{RT} (a_{YT}^2 + \sigma_Y^2)^{-1/2} \\ a_{YT}a_{VT} (a_{YT}^2 + \sigma_Y^2)^{-1/2} \end{bmatrix}.$$

Clearly, the Cholesky decomposition yields correct contemporaneous responses only if $\sigma_Y = 0$. Since generally $\sigma_Y^2 > 0$, the response of the Y -variable to a unit shock in $\varepsilon^{T,CH}$ is overstated while the responses of the R - and V -variables are attenuated towards zero, ceteris paribus. The larger the variance σ_Y^2 is, the larger the attenuation is in the cross-variable responses. Hence, the Cholesky decomposition is biased to finding little contemporaneous comovement in the series, i.e., stickiness.

Because of the ordering, the Cholesky identification correctly requires that the monetary policy $\varepsilon^{P,CH}$ be orthogonal to u^Y . The vector of responses is given by

$$\left[0, \text{var}(\tilde{u}^R)^{1/2}, \text{cov}(\tilde{u}^R, \tilde{u}^V) \cdot \text{var}(\tilde{u}^R)^{-1/2} \right]', \quad (15)$$

where the first entry of the vector is zero by the ordering of the variables, \tilde{u}^R, \tilde{u}^V are found as residuals from regressing u^R, u^V on u^Y (i.e., partial out the effect of the first structural shock under the Cholesky identification), and

$$\text{cov}(\tilde{u}^R, \tilde{u}^V) = a_{RT}a_{VT} \frac{\sigma_Y^2}{a_{YT}^2 + \sigma_Y^2} + a_{RP}a_{VP}, \quad (16)$$

$$\text{var}(\tilde{u}^R) = \frac{a_{RT}^2 (1 + a_{YT}^2) \sigma_Y^4}{(a_{YT}^2 + \sigma_Y^2)^2} + a_{RP}^2 + \sigma_R^2. \quad (17)$$

Residuals u^Y and u^R are contaminated with classical measurement errors v^Y and v^R ; hence, the Cholesky-identified response of the second variable to a unit shock in ε^P is biased up, ceteris paribus, relative to the correct response a_{RP} . Because the Cholesky identification finds little contemporaneous comovement of the R -variable with variables ordered before it and u^R includes the noise shock v^R , recursively identified structural innovations in monetary policy $\varepsilon^{P,CH}$ are often highly correlated with u^R , VAR residuals corresponding to the policy variable equation.

Comovement of responses to $\varepsilon^{P,CH}$ depends on the sign of $\text{cov}(\tilde{u}^R, \tilde{u}^V)$. If $\text{cov}(\tilde{u}^R, \tilde{u}^V) > 0$ ($\text{cov}(\tilde{u}^R, \tilde{u}^V) < 0$), the series respond in the same (opposite) direction to an innovation in $\varepsilon^{P,CH}$. If $\text{sgn}(a_{RT}a_{VT}) \neq \text{sgn}(a_{RP}a_{VP})$, then $\text{cov}(\tilde{u}^R, \tilde{u}^V)$ can be approximately zero even when $a_{RP}a_{VP} \neq 0$. Moreover, it is possible that empirical and theoretical responses have different signs, i.e., $\text{sgn}(\text{cov}(\tilde{u}^R, \tilde{u}^V)) \neq \text{sgn}(a_{RP}a_{VP})$. 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., $\text{sgn}(a_{RT}a_{VT}) \neq \text{sgn}(a_{RP}a_{VP})$ in my example. Thus, the Cholesky identification can produce not only attenuated but also incorrectly signed responses.

Importantly, the true comovement in the series can be recovered if a_{VT} or, especially, a_{RT} are close to zero, that is, asset prices or policy variables do not respond contemporaneously to innovations in technology. This can happen if either a_{VT} and a_{RT} are genuinely zero or the researcher sets them to zero. Thus, structural VARs with $a_{RT}=0$ or $a_{VT}=0$ are likely to find strong cross-variable responses. This explains why, for instance, Cushman and Zha (1997) and Kim and Roubini (2000) who set a_{RT} to zero in their non-recursive identification schemes do find a strong contemporaneous responses of exchange rate to innovations in monetary policy while recursive identification tends to produce weak responses.

Although theoretically possible, it is unlikely that $\text{cov}(\tilde{u}^R, \tilde{u}^V) \cdot \text{var}(\tilde{u}^R)^{-1/2}$, the estimated size of the contemporaneous response of the V -variable to a unit shock in $\varepsilon^{P,CH}$, is greater in absolute value than a_{VP} , the true response to a unit shock in monetary policy ε^P , because $\sigma_Y^2, \sigma_R^2, \sigma_V^2$ dominate entries of AA' in typical applications (that is, diagonal entries of Ω are considerably larger than off-diagonal entries of Ω). As one moves to variables ordered last the estimated covariances become increasingly contaminated with terms like $a_{RT}a_{VT}\sigma_Y^2 / (a_{YT}^2 + \sigma_Y^2)$ in (16) and, thus, estimated responses can fail to reveal the true comovement of the variables in response to structural shocks if the data generating process for \mathbf{u}_t has a factor structure. This is particularly important if many variables are ordered before the \mathbf{R} -block, e.g., in cases where the \mathbf{R} -block is ordered last.

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 (14). Overstated responses to own shocks and understated cross-variable responses can results in the wrong conclusion that the policymaker needs to change his instruments a lot to move target variables a little. The same conclusion follows for non-

recursively identified VARs although details depend on specific assumptions about the structure of the matrix A . In contrast to conventional identification schemes, the reduced rank identification takes out noise from series and, consequently, finds large comovement of series in response to structural shocks. Hence, stickiness of responses can be an artifact of applied identification.

Now I contrast these two approaches to identification in my empirical application. 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 (see Appendix A) 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 (13), and their asymptotic standard errors and bootstrap confidence intervals.¹⁸

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 in the Cholesky identification—e.g., $[RGDP,PGDP,FFR,CRB,EXRUS]$ —produce qualitatively identical

¹⁸ 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).

results, i.e., on impact *CRB* and *EXRUS* respond very weakly to innovations in *FFR*, thus confirming Sims's (1998) stickiness observation.

Impulse response functions¹⁹ (IRFs) for the Cholesky identification and reduced rank identification are reported in Figure 1. 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 the index of commodity prices falls on impact. Then the index 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 it starts depreciating. Eichenbaum and Evans (1995) label this dynamics as “the exchange rate puzzle.”²⁰ 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 and to decrease

¹⁹ I focus on point estimates and do not present confidence bounds for the sake of clarity.

²⁰ Faust and Rogers (2003) find that the exchange rate puzzle can be sensitive to alternative identifications of monetary policy shocks, orderings of variables and the size of the VAR model (recall that identification of key structural shocks can depend on inclusion of unimportant structural shocks).

after 20 or so months. Since there is no theory behind these dynamics, the phenomenon has been called “the price level puzzle” (e.g., Sims 1992). Hanson (2004) reports that the price level puzzle survives across subsamples and specifications although including information variables such as the index of commodity prices attenuates (but does not eliminate) the puzzle. Consistent with previous results, my point estimate of the price level response under the Cholesky identification is positive in the first 20 months and only after 20 months the price level begins to fall.²¹ 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 it starts to fall.

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 months, GDP stabilizes in the RR case and continues to fall in the 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. In my discussion of the conventional and RR identification schemes, I have noted that the Cholesky identification tends to produce attenuated and, possibly, incorrectly signed responses to

²¹ Although consistent with Hanson (2004), 95% confidence intervals for months with positive point estimates include zero.

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. FFR and $EXRUS$ increase on impact in response to ε^P but move in opposite directions in response to ε^T . FFR and CRB co-move in response to ε^T but move in opposite directions in response to ε^P . Consistent with my theoretical predictions, the Cholesky identification even with ordering $[RGDP, PGDP, FFR, CRB, EXRUS]$ finds 1) a weak response of $EXRUS$ to monetary policy shocks ($a_{EXRUS, FFR} = 0.381$, s.e.=0.263) and 2) the theoretically wrong positive sign for response of CRB to identified shocks to monetary policy ($a_{CRB, FFR} = 0.979$, s.e.=0.351).²² In summary, because the conventional identification scheme (in contrast to the RR identification) does not separate structural shocks from series-specific innovations, the conventional scheme finds weak and incorrectly signed contemporaneous responses.

Moreover, monetary policy shocks under the RR identification are less correlated with innovations in FFR than monetary policy shocks under the Cholesky identification.²³ The reason is that the variances of idiosyncratic errors ν 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. Specifically, in the Cholesky identification, almost 95% of the variance in u_{FFR} is attributed to innovations in monetary policy so that the Cholesky-identified monetary policy innovations and the estimated residuals in the FFR equation are virtually identical. In contrast, the correlation is only $26\% = a_{52}^2 / \sigma_{u_{FFR}}^2$ in the reduced rank case. Hence, the RR

²² The responses of CRB and $EXRUS$ to shocks in FFR are computed after partialing out the effect of innovations in $RGDP$ and $PGDP$.

²³ 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.

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 the *FFR* residual, the RR identification appears more plausible than the Cholesky identification.

The Cholesky and RR identification schemes have sharply different predictions about the contribution of innovations in monetary policy to variation in the variables (Figure 2). The Cholesky identification shows that monetary policy has essentially zero effect on exchange rates at short (1 to 12 months) horizons and the importance of monetary policy gradually increases for longer horizons. On the other hand, the RR identification suggests that monetary policy can explain 10% of variation in real exchange rate over short horizons and approximately 13% over long horizons. Again, the RR identification is consistent with the overshooting model of Dornbusch (1976) while the Cholesky is not. Likewise, the RR identification attributes a significantly larger variation in commodity prices to variation in monetary policy.

Variance decomposition for price level is particularly interesting. The Cholesky identification predicts a 2.5% spike at approximately 12 months, then a gradual fall over next 20 months and afterwards a steady increase to 2% at long horizons. In contrast, the RR assigns practically zero contribution for monetary policy in explaining variation in GDP deflator at short (up to 12 months) horizons and then the contribution monotonically increases to 6% at 10 years. Thus, RR identification attributes a significantly more important role to monetary policy than the Cholesky identification does. The shape under the RR identification is consistent with the sticky prices hypothesis, while the shape under the Cholesky identification appears to be controversial at best.

The schemes also differ in explaining the variation in GDP. The Cholesky identification produces a hump shape in the importance of the monetary policy, while the RR identification yields a monotonic increase. Both schemes produce similar magnitudes of the contribution (10%) at long horizons.

Finally, the Cholesky identification predicts that variation of the fed funds rate at short horizons is almost exclusively determined by innovations in monetary policy. In contrast, the RR scheme suggests that at most 30% of the variation at short horizons is attributed to monetary policy and the rest is attributed to noise and technology shocks (this includes endogenous responses of the policy variable to technology shocks, e.g., the Taylor rule). This is not surprising because the Cholesky identification overstates the response of variables to the structural shocks associated with these variables. For longer horizons, the contribution of monetary policy to variation in FFR falls to 20% for the Cholesky identification and to 7% for the RR identification.

In summary, the reduced rank identification appears to produce more reasonable impulse responses and contribution of monetary policy to the variation in the VAR variables than the recursive identification. Specifically, the 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 these results is the identified factor structure of VAR residuals.

4 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 structural shocks (i.e., technology, preferences, policy) in business cycle models. Specifically, identification of structural shocks in a VAR is not invariant to the number of variables included in the VAR. 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 and, thus, it does not depend on the number of variables in the VAR. This identification, which outgrows from FSVARs, 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, the reduced rank approach uniquely identifies structural innovations and therefore it can be used for economic analyses. Finally, the paper suggests that the apparent weak contemporaneous cross-variable responses and strong own responses in structural VARs can be an artifact induced by the fact that conventional identification schemes do not separate structural shocks from series-specific innovations. This artifact vanishes after imposing a factor structure on the shocks as in the reduced rank identification. In a similar vein, the reduced rank identification can rationalize noisiness of structural policy shocks regularly found in conventionally identified VARs.

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.

5 References

- Algina, James, 1980. "A Note on Identification in the Oblique and Orthogonal Factor Analysis Models," *Psychometrika* 45(3), 393-396.
- Altonji, J.G. and J.C. Ham, 1990. "Variation in Employment Growth in Canada: The Role of External, National, Regional, and Industrial Factors," *Journal of Labor Economics* 8(1), S198-S236.
- Anderson, Gary and George Moore, 1985. "A Linear Algebraic Procedure for Solving Linear Perfect Foresight Models," *Economics Letters* 17, 247-252
- Anderson, T.W. and Herman Rubin, 1956. "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.
- Anderson, T.W., 2003. *An Introduction to Multivariate Statistical Analysis*. Wiley.
- Andrews, Donald W.K., 2002. "Higher Order Improvements of a Computationally Attractive k-Step Bootstrap for Extremum Estimators," *Econometrica* 70(1), 119-162
- Bai, Jushan and Serena Ng, 2002. "Determining the Number of Factors in Approximate Factor Models," *Econometrica* 70(1), 191-221.
- Basilevsky, Alexander, 1994. *Statistical Factor Analysis and Related Methods: Theory and Applications*. Wiley.
- Bekker, Paul A., Arjen Merkens, Tom J. Wansbeek. 1994. *Identification, Equivalent models and Computer Algebra*. Academic Press.
- Bernanke, Ben, 1986. "Alternative explanations of the money-income correlation," *Carnegie-Rochester Conference Series on Public Policy* 25, 49-99.
- Bernanke, Ben and Alan Blinder, 1992. "The federal Funds Rate and the Channels of monetary Transmission," *American Economic Review* 82, 901-921.
- Bernanke, Ben and Ilian Mihov, 1998. "Measuring Monetary Policy," *Quarterly Journal of Economics* 113, 869-902.
- Bernanke, Ben, and Jean Boivin, 2002. "Monetary Policy in a Data-Rich Environment," *Journal of Monetary Economics* 50, 525-546.
- Bernanke, Ben, Jean Boivin and Piotr Elias, 2005. "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics* forthcoming.
- Blanchard, O. J. and D. Quah, 1989. "The Dynamic Effects of aggregate Demand and Supply Disturbances," *American Economic Review* 79, 655-673.
- Christiano, L., M. Eichenbaum and C.L. Evans, 1999. "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.
- Christiano, L., M. Eichenbaum and C.L. Evans, 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy* 113(1), 1-45.
- Clark, Todd E. and Kwanho Shin, 2000. "The Sources of Fluctuations within and across Countries," in *Intranational Macroeconomics*, Gregory D. Hess and Eric van Wincoop, eds. Cambridge: Cambridge University Press.

- Clark, Todd E., 1996. "Small-Sample Properties of Estimators of Nonlinear Models of Covariance Structure," *Journal of Business and Economic Statistics* 14(3), 367-373.
- Cragg, G. John and Stephen G. Donald, 1997. "Inferring the Rank of a Matrix," *Journal of Econometrics* 76, 223-250.
- Cushman, D.O. and T. Zha, 1997. "Identifying Monetary Policy in a Small Open Economy under Flexible Exchange Rates," *Journal of Monetary Economics* 39, 433-448.
- Davidson, Russell and James G. McKinnon, 1999. "Bootstrap Testing in Non-linear models," *International Economic Review* 40(2), 487-508.
- Dornbusch, Rudiger, 1976. "Expectations and Exchange Rate Dynamics," *Journal of Political Economy* 84, 1161-1176.
- Efron, Bradley, and Robert J Tibshirani, 1993. *An Introduction to the Bootstrap*. Chapman and Hall/CRC.
- Eichenbaum, Martin and Charles L. Evans, 1995. "Some Empirical Evidence on the Effects of Shocks to Monetary Policy on Exchange Rates," *Quarterly Journal of Economics* 110(4), 975-1009.
- Faust, J., 1998. "The robustness of identified VAR conclusions about money," *Carnegie-Rochester Conference Series on Public Policy* 49, 207-244.
- Faust, Jon, and John H. Rogers, 2003. "Monetary Policy's Role in Exchange Rate Behavior," *Journal of Monetary Economics* 50, 1403-1424.
- Fischer, Stanley, 1974. "Money and the Production Function," *Economic Inquiry* 12(4), 517-533.
- Forni, M. Hallin, M. Lippi, and Reichlin, L., 2000. "The Generalized Dynamic Factor Model: Identification and Estimation," *Review of Economic and Statistics* 82, 540-554.
- Forni, M., M. Lippi, and L. Reichlin, 2003. Opening the black box: identifying shocks and propagation mechanisms in VAR and factor models. CEPR working paper #4133.
- Forni, Mario, and Marco Lippi, 2001. "The Generalized Dynamic Factor Model: Representation Theory," *Econometric Theory* 17, 1113-1141
- Geweke, J., 1977. "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.
- Giannone, D., L. Reichlin, and L. Sala, 2004. Monetary policy in real time. mimeo.
- Hall, Peter, and Joel Horowitz, 1996. "Bootstrap Critical Values for Testing Based on Generalized-Methods-of-Moment Estimators," *Econometrica* 64(4), 891-916.
- Hanson, Michael S., 2004. "The "price puzzle" reconsidered," *Journal of Monetary Economics* 51, 1385-1413.
- Horowitz, Joel L., 1998. "Bootstrap Methods for Covariance Structures," *Journal of Human Resources* 33(1), 39-61.
- Inoue, Atsushi and Lutz Kilian, 2002. "Bootstrapping Autoregressive Processes with Possible Unit Roots," *Econometrica* 70(1), 377-391.
- Kilian, Lutz, 1998. "Small-sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics* 80(2), 218-230.
- Kilian, Lutz, 1999. "Finite-Sample Properties of Percentile and Percentile-t Bootstrap Confidence Intervals for Impulse Response," *Review of Economics and Statistics* 81(4), 652-660.

- Kim, Soyoung and Nouriel Roubini, 2000. "Exchange Rate Anomalies in the Industrial Countries" A Solution with a Structural VAR Approach," *Journal of Monetary Economics* 45, 561-586.
- Kose, M Ayhan, Christopher Otrok, and Charles H. Whiteman, 2003. "International Business Cycles: World, Region, and Country-Specific Factors," *American Economic Review* 93(4), 1216-1239.
- Lawley, D.N. and A.E. Maxwell, 1974. *Factor Analysis as a Statistical Method*. London: Butterworths.
- Lucas, Robert E., Jr., 1972. "Expectations and the Neutrality of Money," *Journal of Economic Theory* 4(2), 321-334
- Norrbin, S.C. and D.E. Schlagenhauf, 1988. "An Inquiry into the Source of Macroeconomic Fluctuations," *Journal of Monetary Economics* 22(1), 43-70.
- Norrbin, S.C. and D.E. Schlagenhauf, 1996. "The Role of International Factors in the Business Cycle," *Journal of International Economics* 40(1-2), 84-104.
- Pagan, Adrian, 1984. "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review* 25(1), 221-247.
- Quah, D., and T.J. Sargent, 1993. "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.
- Rotemberg, Julio J., Woodford, Michael, 1997. "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," in Bernanke, Ben S. and Rotemberg, Julio J., eds. *NBER macroeconomics annual*. Cambridge and London: MIT Press, 297-346.
- Shapiro, Matthew D. and Mark W. Watson, 1988. "Sources of Business Cycle Fluctuations," *NBER Macroeconomics Annual*. Cambridge: MIT Press.
- Sims, Christopher and Tao Zha, 1999. "Error Bands for Impulse Responses," *Econometrica* 67(5), 1113-1155.
- Sims, Christopher, 1980. "Macroeconomics and Reality," *Econometrica* 48, 1-48.
- Sims, Christopher, 1992. "Interpreting the Macroeconomics Time Series Facts: the Effects of Monetary Policy," *European Economic Review* 36, 975-1011.
- Sims, Christopher, 1998. "Stickiness," *Carnegie-Rochester Conference Series on Public Policy* 49, 317-356.
- Sims, Christopher, 2002. "Solving Linear Rational Expectations Models," *Computational Economics*, October 20(1-2), 1-20.
- Stock, James H. and Mark W. Watson, 1998. "Diffusion Indexes," NBER Working paper #6702.
- Stock, James H. and Mark W. Watson, 2003. "Understanding Changes in International Business Cycle Dynamics," NBER Working paper #9859.
- Woodford, Michael, 2003. *Interest and Prices*. Princeton, NJ: Princeton University Press.
- Wooldridge, Jeffrey M., 2003. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

6 Appendix A. Selection criterion.

In this appendix, I present a criterion to consistently select the number the number of factors in VAR residuals, or equivalently the rank of A . 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. Specifically, the number of factors k can be chosen using the criterion suggested in the following proposition.

Proposition 2

Define criterion functions $S(k) = f(T)^{-1} J(k, A, \Psi) - g(k)$, where $J(k, A, \Psi)$ 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 $\lim_{T \rightarrow \infty} f(T) = \infty$ and $\lim_{T \rightarrow \infty} T^{-1} f(T) = 0$. Then \hat{k} minimizing $S(k)$ is a consistent estimate of the true number of factors k_0 , i.e., $\lim_{T \rightarrow \infty} \Pr(\hat{k} > k_0) = \lim_{T \rightarrow \infty} \Pr(\hat{k} < k_0) = 0$.

Proof: see the Appendix C.

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 . Popular choices of functions f and g 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).²⁴ Note that AIC does not satisfy $\lim_{N \rightarrow \infty} f(N) = \infty$ 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.²⁵ Because the covariance matrix of the VAR residuals \mathbf{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

²⁴ 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 .

²⁵ An alternative strategy is to sequentially test if k factors account for observed covariance until the 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.

because the penalty is too heavy (Cragg and Donald 1997). I use Monte Carlo simulations (available upon request) to calibrate the penalty functions.

Alternatively, one can use less formal methods such as scree plots to determine the rank of A .

7 Appendix B. An Economic model with factor structure of innovations.

To illustrate how (9) can arise in economic settings, consider a stylized model with money in production function (see Fischer 1974 for a discussion). Suppose that the central banker chooses real money balances M_t to provide economy with sufficient liquidity. The production function is $Y_t = \exp(a_t) K_t^\alpha M_t^{1-\alpha}$ where Y_t is output, K_t is capital stock, $a_t \sim niid(0, \sigma_a^2)$ is an exogenously given technology shock. The representative consumer sets consumption to track output, that is, $C_t = Y_t$. Capital stock K_t evolves according to $K_{t+1} = (1 - \delta) K_t + Y_t - C_t + \bar{I}$ where \bar{I} is autonomous non-stochastic investment and δ is depreciation rate. The model is closed by equilibrium condition $M_t = Y_t$.

The log-linearized model is

$$c_t = y_t, \quad (\text{consumption rule})$$

$$m_t = y_t, \quad (\text{monetary policy rule})$$

$$y_t = a_t + \alpha k_t + (1 - \alpha) m_t, \quad (\text{production function})$$

$$k_{t+1} = (1 - \delta) k_t + y_t - c_t, \quad (\text{capital accumulation function})$$

where small letters denote log deviations of the respective variables from their stochastic steady states.

I assume that at the time when the consumer and the central banker choose c_t and m_t they observe $a_t^C = a_t + \varepsilon_t^C$ and $a_t^B = a_t + \varepsilon_t^B$, respectively, with $\varepsilon_t^C, \varepsilon_t^B \sim niid(0, \sigma_\varepsilon^2)$. This is similar to information structure in the island model of Lucas (1972). Capital stock is known to both parties

when they choose c_t and m_t . Each party solves the model given his information sets and chooses his

optimal action. Under these assumptions, $c_t = E(y_t | a_t^C) = k_t + \alpha^{-1}\gamma a_t + \alpha^{-1}\gamma \varepsilon_t^C$,

$m_t = E(y_t | a_t^B) = k_t + \alpha^{-1}\gamma a_t + \alpha^{-1}\gamma \varepsilon_t^B$, $y_t = k_t + \alpha^{-1}(\alpha + \gamma(1 - \alpha))a_t + \alpha^{-1}(1 - \alpha)\gamma \varepsilon_t^B$, where

$$\gamma = (\sigma_a^2 + \sigma_\varepsilon^2)^{-1} \sigma_a^2.$$

Now consider an empirical model that includes consumption and output and use (9):

$$\begin{aligned} \begin{bmatrix} c_t \\ y_t \end{bmatrix} &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} k_t + \begin{bmatrix} \alpha^{-1}\gamma \\ \alpha^{-1}(\alpha + \gamma(1 - \alpha)) \end{bmatrix} a_t + \begin{bmatrix} \alpha^{-1}\gamma \varepsilon_t^C \\ \alpha^{-1}(1 - \alpha)\gamma \varepsilon_t^B \end{bmatrix} = \\ &= (1 - (1 - \delta)L)^{-1} \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} c_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \alpha^{-1}\gamma \\ \alpha^{-1}(\alpha + \gamma(1 - \alpha)) \end{bmatrix} a_t + \begin{bmatrix} \alpha^{-1}\gamma \varepsilon_t^C \\ \alpha^{-1}(1 - \alpha)\gamma \varepsilon_t^B \end{bmatrix} \approx \\ &\approx \Pi(L) \begin{bmatrix} c_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \alpha^{-1}\gamma \\ \alpha^{-1}(\alpha + \gamma(1 - \alpha)) \end{bmatrix} a_t + \begin{bmatrix} \tilde{\varepsilon}_t^C \\ \tilde{\varepsilon}_t^B \end{bmatrix}, \end{aligned}$$

where $\tilde{\varepsilon}_t^C \equiv \alpha^{-1}\gamma \varepsilon_t^C$ and $\tilde{\varepsilon}_t^B \equiv \alpha^{-1}(1 - \alpha)\gamma \varepsilon_t^B$. Reduced-form residuals in this VAR have a factor structure where a_t is a common factor and $\tilde{\varepsilon}_t^C, \tilde{\varepsilon}_t^B$ are idiosyncratic errors.

8 Appendix C. Proofs.

PROOF OF PROPOSITION 1 (INFERENCE FOR TWO-STEP ESTIMATION)

The proof follows Wooldridge (2002, Section 12.4). Suppose that \mathbf{u}_t is a sequence of i.i.d. random vectors with finite fourth moments and $E(\mathbf{u}_t \mathbf{X}'_{t-s}) = 0$ for $s > 0$. The objective function of the GMM estimator is

$$L(A(\hat{\boldsymbol{\Pi}}), \Psi(\hat{\boldsymbol{\Pi}})) = \left\{ \left(\frac{1}{T} \sum_{t=1}^T \hat{F}_t(A, \Psi) \right)' W^{-1} \left(\frac{1}{T} \sum_{t=1}^T \hat{F}_t(A, \Psi) \right) \right\} \quad (\text{C.1})$$

The score function of (C.1) is

$$s_t(A, \Psi, \hat{\boldsymbol{\Pi}}) = 2 \nabla_{A, \Psi} \hat{F}_t(A, \Psi) W^{-1} \left(\frac{1}{T} \sum_{t=1}^T \hat{F}_t(A, \Psi) \right) \quad (\text{C.2})$$

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

need to prove that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T s_t(A, \Psi, \hat{\boldsymbol{\Pi}}) = \frac{1}{\sqrt{T}} \sum_{t=1}^T s_t(A, \Psi, \boldsymbol{\Pi}) + o_p(1) \quad (\text{C.3})$$

Note that by construction of the moments $\nabla_{A,\Psi} \hat{F}_t(A, \Psi)$ does not depend on X_t and Π .

Given this observation, I take a mean value expansion of (C.3) around Π :

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

Hence to show equivalency of asymptotic distributions, it is sufficient to show that

$E(\nabla_{\Pi} s(A, \Psi, \Pi)) = 0$. To prove this, plug $u_{it} = x_{it} - \sum_{m=1}^n \sum_{s=1}^p \pi_{s,m} x_{m,t-s}$ into (C.2). Because

$\nabla_{A,\Psi} \hat{F}_t(A, \Psi)$ and W do not depend on X_t and Π , observe that for elements of $\nabla_{\Pi} F(A, \Psi)$

corresponding to diagonal elements of Ω I have:

$$\begin{aligned} E \left[\frac{\partial}{\partial \pi_{s,m}} \left\{ \frac{1}{T} \sum_{t=1}^T \left(x_{it} - \sum_{m=1}^n \sum_{s=1}^p \pi_{s,m} x_{m,t-s} \right)^2 - \left(\sigma_{vi}^2 + \sum_{j=1}^k a_{ij}^2 \right) \right\} \right] = \\ = E \left[\frac{1}{T} \sum_{t=1}^T 2 \left(x_{it} - \sum_{m=1}^n \sum_{s=1}^p \pi_{s,m} x_{m,t-s} \right) x_{s,m} \right] = 0 \quad \forall s, m, i, \end{aligned}$$

by the condition $E\left(\left[X_t - \Pi_1 X_{t-1} - \dots - \Pi_p X_{t-p}\right] X'_{t-k}\right) = 0 \quad k = 1, \dots, p$ and the same result holds

for all off-diagonal entries of Ω . Hence, $E(\nabla_{\Pi} F(A, \Psi)) = 0$ implying $E(\nabla_{\Pi} s(A, \Psi, \Pi)) = 0$.

QED

PROOF OF PROPOSITION 2 (CONSISTENCY OF THE CRITERION).

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

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

$\Pr(S(k_*) \leq S(k_0)) \leq \Pr(J(k_0, A, \Psi) \geq f(T) [g(k_0) - g(k_*)]) \xrightarrow{T \rightarrow \infty} 0$ because

$f(T) \rightarrow \infty$, $g(k_*) < g(k_0)$ and $J(k_0, A, \Psi)$ is distributed asymptotically as

$\chi^2(\frac{1}{2}n(n+1) - nk - k)$. It follows that $\lim_{T \rightarrow \infty} \Pr(\hat{k} > k_0) = 0$. For $\hat{k} < k_0$, observe that

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

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

QED

9 Appendix D. Bootstrap procedure.

Since moment conditions in my application are non-linear, I suggest using the k -step bootstrap developed by Davidson and McKinnon (1999) and adapted for testing by Andrews (2002) to reduce the computational burden of the bootstrap.²⁶ Although there could be many ways of implementing this bootstrap procedure, I suggest that for each bootstrap replication one estimate Π first and then estimates A and Ψ conditional on $\hat{\Pi}$. This approach performs well in Monte Carlo simulations and it is very fast: the first stage is computed by LS and the second stage is computed with the k -step bootstrap procedure.²⁷

In summary, I use the following procedure:

1. Estimate VAR in (1) by LS and store the estimated residuals $\hat{\mathbf{u}}_t = \mathbf{X}_t - \sum_{i=1}^p \hat{\Pi}_i \mathbf{X}_{t-i}$.
2. Estimate the matrices A and Ψ using the stored residuals $\hat{\mathbf{u}}_t$ and the GMM estimator described above.
3. Resample the residuals $\hat{\mathbf{u}}_t$ (with replacement) and create new series $\mathbf{u}_t^{(B)}$.²⁸
4. Create a new series $\mathbf{X}_t^{(B)}$ using equation (1), $\hat{\Pi}$ from step 1 and $\mathbf{u}_t^{(B)}$ from step 3.
5. Estimate the VAR using $\mathbf{X}_t^{(B)}$ and compute the residuals.
6. Estimate the matrices A and Ψ using the residuals from step 5.²⁹
7. Repeat steps 3-6 sufficiently many times to approximate the distribution of the statistic of interest.

Standard asymptotic inference only requires the first two steps.

²⁶ 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.

²⁷ An alternative is to estimate to Π, A, Ψ simultaneously by combining LS moments for Π and GMM moments for A, Ψ . Performance of this alternative is poor because too many moments are involved. Results of Monte Carlo experiments are available upon request.

²⁸ Note that VAR of a sufficiently high order makes reduced error term \mathbf{u}_t approximately serially uncorrelated and, thus, one can use the residual-based bootstrap algorithm for stationary (Kilian 1998, 1999) and nonstationary (Inoue and Kilian 2002) autoregressions.

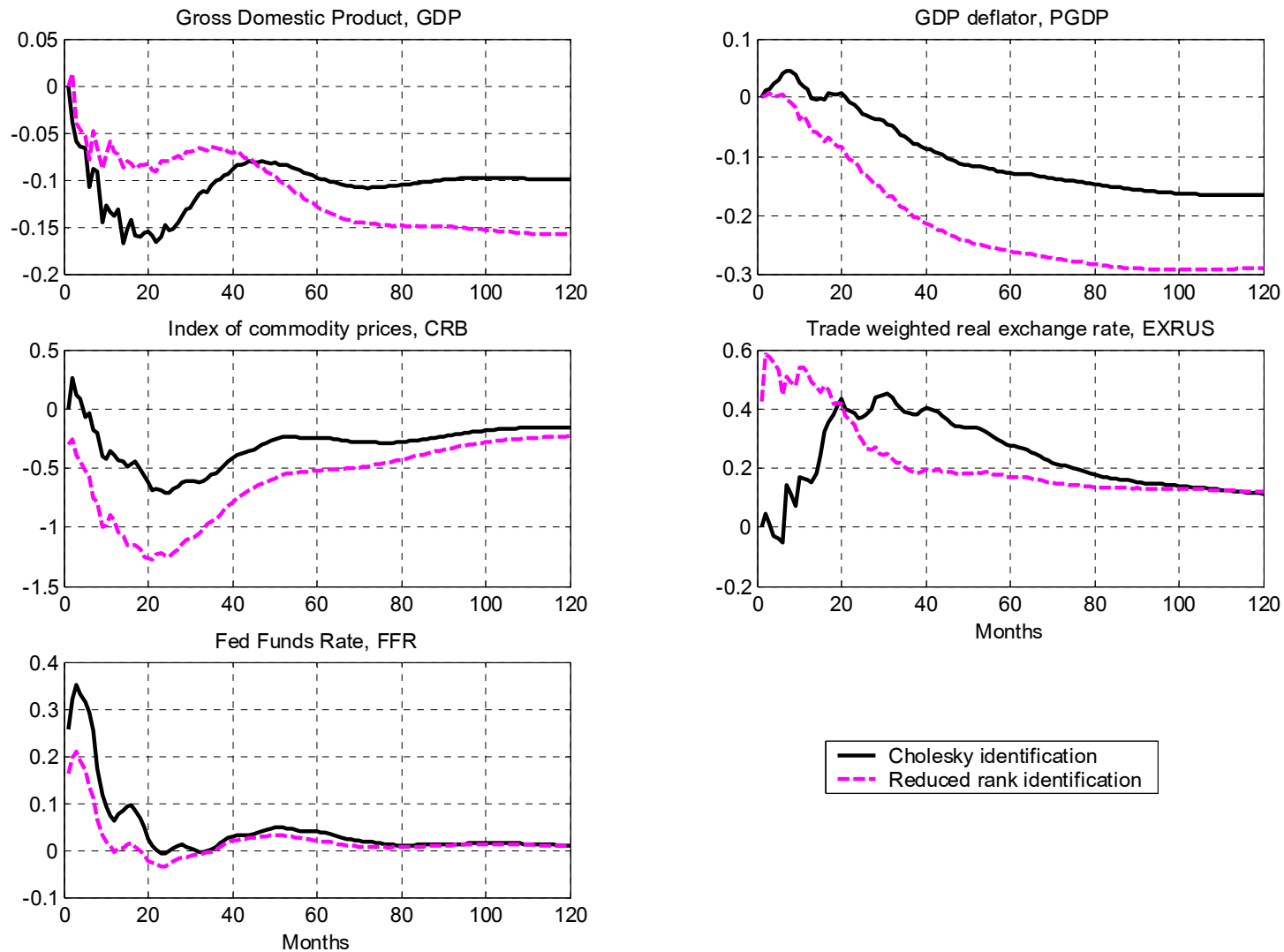
²⁹ If the model is overidentified, bootstrap moments must be re-centered as in Hall and Horowitz (1996).

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^T : the first column of A in (13) | | | | | |
| <i>RGDP</i> | a_{11} | 0.098 | 0.010 | 0.105 | [0.032, 0.268] |
| <i>PGDP</i> | a_{21} | 0.007 | 0.001 | 0.009 | [-0.016, 0.032] |
| <i>CRB</i> | a_{31} | 0.511 | 0.053 | 0.552 | [0.367, 0.660] |
| <i>EXRUS</i> | a_{41} | -0.136 | 0.016 | -0.150 | [-0.331, 0.028] |
| <i>FFR</i> | a_{51} | 0.217 | 0.022 | 0.254 | [0.186, 0.283] |
| Policy shock ε^P : the second column of A in (13) | | | | | |
| <i>CRB</i> | a_{32} | -0.257 | 0.043 | -0.305 | [-0.419, -0.104] |
| <i>EXRUS</i> | a_{42} | 0.369 | 0.114 | 0.425 | [0.222, 0.566] |
| <i>FFR</i> | a_{52} | 0.150 | 0.062 | 0.161 | [0.096, 0.239] |

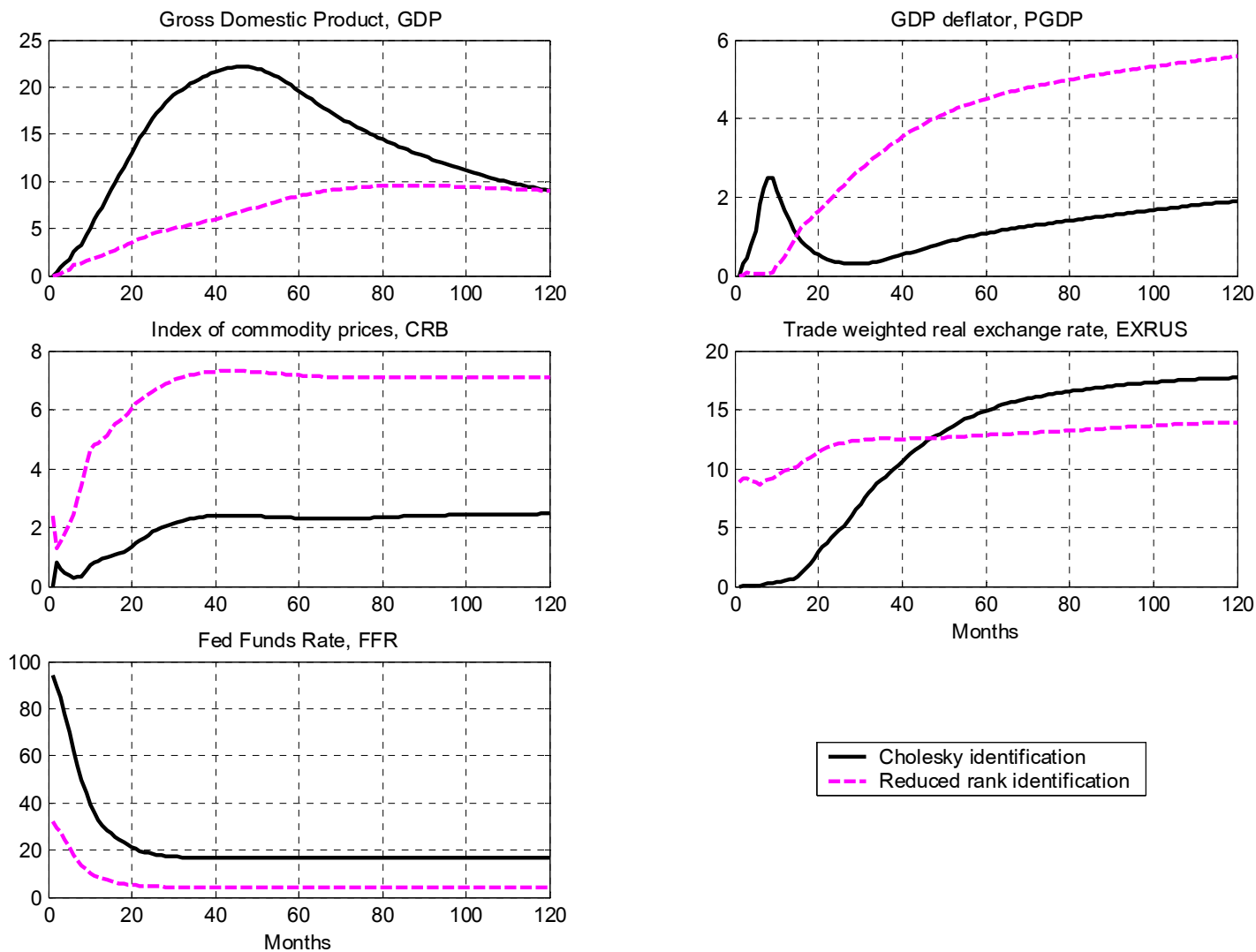
Note: The table presents estimates of the matrix A in (13). Bootstrap procedure is described in the text. BC_α bootstrap procedure with re-centering bootstrap moments (2000 replications) is used. See Appendix D for details on the bootstrap procedure. *RGDP* is real GDP, *PGDP* is GDP deflator, *CRB* is the index of commodity prices, *EXRUS* is trade-weighted real exchange rate, *FFR* is fed funds rate.

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. Variance decomposition. The contribution of innovations in monetary policy (percent).



Note: The figure presents the share of variation in indicated variables due to innovations in monetary policy. Time (the horizontal axis) is in months. The ordering in the Cholesky factorization is [RGDP,PGDP,CRB,EXRUS,FFR].