Vector Autoregressions and Reduced Form Representations of Dynamic Stochastic General Equilibrium models

Federico Ravenna*

5 December 2004

Abstract

Dynamic Stochastic General Equilibrium models are often tested against empirical VARs or estimated by minimizing the distance with the VAR impulse response functions. This papers examines under what conditions DSGE models map into VARs representations. We show that DSGE models map into VARMA processes, therefore the empirical impulse response functions from an estimated VAR cannot be generated by the DSGE model except under strict conditions. Comparing inconsistent vector moving average representation is not an appropriate test of the explanatory power of the theoretical model. Using a monetary model with nominal rigidities, we illustrate how the mis-specified VAR returns a largely incorrect estimate of the model's driving shocks. The result does not hinge on mis-identification or the estimator volatility. The paper also proves under what conditions the Kalman-filtered shocks vector returns the true vector of innovations that generated the observable data. This set of conditions is a subset of the assumptions needed to estimate the vector of innovations from the VAR representation of the model.

JEL Classification Numbers: C13; C22; E32; F41

Keywords: Vector Autoreregression; Dynamic Stochastic General Equilibrium Model; Kalman Filter; Business Cycle Shocks

*Department of Economics, University of California, Santa Cruz, CA 95064. Email: fravenna@cats.ucsc.edu. I would like to thank Richard Dennis, Walter Enders, James Hamilton, Bart Hobijn, Oscar Jorda and Yin Wong Cheung for helpful comments and suggestions. Financial suport from the UCSC Social Science Division and the USCS Committee on Research is gratefully acknowledged.

1 Introduction

The explanatory power of dynamic stochastic general equilibrium (DSGE) models is often tested by comparing the model impulse response functions to those obtained from an estimated Vector Autoregression (usually a structural VAR). Christiano, Eichenbaum and Evans (2003), for example, evaluate the performance of a staggered price-setting model of the business cycle by showing that it can account for the dynamic response induced by a monetary policy shock as estimated in a 10-variables VAR. Some researchers estimate model parameters by minimizing the distance between the model's and the estimated VAR impulse response functions (Woodford and Rotemberg, 1998).

These practices rest on the assumption that the DSGE and VAR models share the same reduced form - or map into the same Vector Moving Average (VMA) representation. We show that this hypothesis is valid only under a strict set of conditions, and outline what are the implicit assumptions made when mapping a DSGE model into a VAR. The solution (reduced form) of DSGE models does not generally map into a finite-order VAR representation. This is a point that has been made with reference to particular models by a number of authors. We offer a general proof, showing that unless all state variables are included in the data sample the DSGE model will map into a Vector Autoregression-Moving Average (VARMA) model[1].

In the paper we illustrate the empirical relevance of the VAR mis-specification problem. Using a DSGE model of a small open economy, we estimate the time-series of the monetary policy innovations from the VAR approximation to the true VARMA representation of the model, and compare it to the true shocks vector. Depending on the model parametrization, the mis-specification can lead to extremely large errors. The result does not hinge on identification problems or estimator volatility: it is purely the outcome of neglecting the moving average terms in the estimated system.

An alternative to VARs is estimation of the state-space representation via Maximum Likelihood (as in Cho and Moreno, 2002, Ireland, 2001a, 2001b), Generalized Method of Moments or Simulated Method of Moments (Ruge-Murcia, 2002) or Bayesian estimation (as in Rabanal and Rubio-Ramirez, 2003). Once the state-space representation is estimated, the Kalman Filter algorithm can be used to estimate the vector of shocks generating the data sample. We show under what conditions the Kalman Filter returns the true shocks vector. These conditions are a subset of the assumptions needed to estimate the shocks vector from the VAR representation of the DSGE model.

The paper is organized as follows. Section 2 describes the DSGE model setup. Section 3

[1] See Cooley and Dwyer (1998) and Ingram et al. (1994). As DSGE models have became less stylyzed, and thus more apt to describe the data comovements, a growing literature has focused on the impact of inconsistencies between the VARs underlying assumptions and the DSGE models that are supposed to describe the data-generating process (Chow and Kwan, 1998, Gali and Rabanal, 2004, Erceg, Guerrieri and Gust, 2004). While this literature does not examine the mis-specification arising from neglect of the MA terms, Wallis (1977) and Zellner and Palm (1974) already recognized that the general form of a DSGE model is a VARMA process.

discusses VAR representations of DSGE models. Section 4 proves that the representation of the model solution when state variables are unobservable is a VARMA model, and establishes the order of its AR and MA components. Section 5 shows under what conditions the Kalman filter returns the true vector of innovations that generated the data. Section 6 illustrates the consequences of VAR mis-specification. Section 7 concludes.

2 Linear rational expectations dynamic models

A linear rational expectation stochastic model can be written as a system of stochastic difference equations:

$$\begin{aligned} 0 &= Ax_t + Bx_{t-1} + Cy_t + Dz_t \\ 0 &= E_t[Fx_{t+1} + Gx_t + Hx_{t-1} + Jy_{t+1} + Ky_t + Lz_{t+1} + Mz_t] \\ z_{t+1} &= Nz_t + \varepsilon_{t+1} \end{aligned}$$

where x_t is an $n\times1$ vector of endogenous state variables, z_t is an $m\times1$ vector of exogenous state variables, y_t is an $r \times 1$ vector of endogenous variables, ε_t is a vector white noise stochastic process of dimension $m \times 1$ with variance-covariance matrix Σ and unconditional expectation $E(\varepsilon_t) = 0$. Capital letters denote matrices. It is assumed that the matrix N has only stable eigenvalues. Whenever useful, I will indicate the size of a matrix with an index in the upper-left corner.

The solution to the system is the recursive equilibrium law of motion:

$$\xi_{t+1} = F\xi_t + v_{t+1} \tag{1}$$
$$y_t = H'\xi_t \tag{2}$$

$$^{n+m\times1}\xi_t = \begin{bmatrix} ^{n\times1}\xi_t^1 \\ ^{m\times1}\xi_t^2 \end{bmatrix} = \begin{bmatrix} x_t \\ z_t \end{bmatrix} \tag{3}$$
$$^{n+m\times1}v_t = \begin{bmatrix} ^{n\times1}0 \\ ^{m\times1}\varepsilon_t \end{bmatrix} \tag{4}$$
$$^{n+m\times n+m}F = \begin{bmatrix} ^{n\times n}F^{11} & ^{n\times m}F^{12} \\ ^{m\times n}0 & ^{m\times m}F^{22} \end{bmatrix} \tag{5}$$
$$^{r\times n+m}H' = \begin{bmatrix} ^{r\times n}H'^1 & ^{r\times m}H'^2 \end{bmatrix} \tag{6}$$
$$E(v_t v_t') = {}^{n+m\times n+m}Q = \begin{bmatrix} ^{n\times n}0 & ^{n\times m}0 \\ ^{m\times n}0 & ^{m\times m}\Sigma \end{bmatrix} \tag{7}$$
$$E(v_t v_\tau') = 0 \qquad for\ \tau \neq t \tag{8}$$

In the econometric literature eq. (1) is known as the state equation and eq. (2) as the observation or measurement equation. The assumption in eq. (8) can be dropped and is made only for ease of exposition. The vector v_t can be described by any stationary ARMA process. Without loss of generality, we assume v_t is a white noise stochastic process.

The state-space representation given by eqs. (1) and (2) is common in time series analysis. A large number of models (including ARMA processes) can be written in state-space form and estimated using the Kalman filter (Hamilton, 1994). In the DSGE literature, the exogenous and endogenous state variables are usually defined in distinct vectors, so that the exogenous innovations vector has no zero-element (Uhlig, 1997, Cooley and Dwyer, 1998):

$$\begin{aligned} y_t &= H'^1 x_t + H'^2 z_t \\ x_t &= F^{11} x_{t-1} + F^{12} z_{t-1} \\ z_t &= F^{22} z_{t-1} + \varepsilon_t \end{aligned} \tag{9}$$

3 VAR Representations of DSGE models

To write model (9) as a VAR, re-label x_t as x_{t-1} so that $\xi_t^1 = x_{t-1}$. Then:

$$\begin{aligned} Y_t &= \tilde{A} Y_{t-1} + \tilde{B} z_t \\ z_t &= F^{22} z_{t-1} + \varepsilon_t \\ Y_t &= \begin{bmatrix} x_t \\ y_t \end{bmatrix} \ ; \ \tilde{A} = \begin{bmatrix} F^{11} & 0 \\ H'^1 & 0 \end{bmatrix} \ ; \ \tilde{B} = \begin{bmatrix} F^{12} \\ H'^2 \end{bmatrix} \end{aligned} \tag{10}$$

where the vector $Y_t' = [x_t \ , \ y_t]$ has dimension $1 \times n + r$. Assume the vector z_t has dimension $m = n + r$. This implies that the number $n + r$ of observable variables is equal to the number of shocks, and that the matrix $\tilde{B}$ is square. Since:

$$\begin{aligned} z_t &= \tilde{B}^{-1} Y_t - \tilde{B}^{-1} \tilde{A} Y_{t-1} \\ &= F^{22} [\tilde{B}^{-1} Y_{t-1} - \tilde{B}^{-1} \tilde{A} Y_{t-2}] + \varepsilon_t \end{aligned}$$

we obtain a restricted VAR(2) representation for the system (10):

$$Y_t = (\tilde{A} + \tilde{B} F^{22} \tilde{B}^{-1}) Y_{t-1} - (\tilde{B} F^{22} \tilde{B}^{-1} \tilde{A}) Y_{t-2} + \eta_t \tag{11}$$

where $\eta_t = \tilde{B} \varepsilon_t$. Let $\Sigma = E(\varepsilon_t \varepsilon_t')$ be a diagonal matrix, so that ε_t is the vector of orthogonal innovations[2]. An alternative way to write model (10) is:

$$\begin{aligned} \Lambda_0 Y_t &= \Lambda_1 Y_{t-1} + \Lambda_2 Y_{t-2} + \varepsilon_t \\ \Lambda_0 &= \tilde{B}^{-1} \\ \Lambda_1 &= \tilde{B}^{-1} \tilde{A} + F^{22} \tilde{B}^{-1} \\ \Lambda_2 &= -F^{22} \tilde{B}^{-1} \tilde{A} \end{aligned} \tag{12}$$

[2]This assumption is shared by most DSGE models (though not all, as in the case of a multi-sector model where sectorial productivity shocks are correlated). It is anyway always possible to orthogonalize the vector ε_t.

A large class of structural models can be written in this form. The structural VAR (SVAR) literature studied in detail the problem of recovering the matrices $\Lambda_0, \Lambda_1, \Lambda_2$ from estimation of the model (11) (Leeper et al., 1996, Uhlig, 2004). Restrictions need to be imposed in the estimation of eq. (11) to identify the matrix $\tilde{B}$ and orthogonalize the shocks vector $\eta_t = \tilde{B}\varepsilon_t = \Lambda_0^{-1}\varepsilon_t$. A vast number of papers proposes different SVAR identification strategies, the most popular being the Sims (1980) orthogonalization where Λ_0 is assumed to be lower-diagonal.

In the following we abstract from the identification problem, and restrict our attention to the specification problem. While knowledge of the matrix $\tilde{B}$ is irrelevant when estimating eq. (11), whenever necessary we will assume $\tilde{B}$ is known, so the econometrician can always recover the time series ε_t from a given series η_t.

If $m > n + r$ the matrix $\tilde{B}$ is not invertible and we cannot obtain a VAR representation of the model. If instead $m < n + r$, we still have the option of making $\tilde{B}$ invertible by eliminating some of the observable variables from the system. This case is explored in detail in the following section.

3.1 Reduced Form VAR with Unobserved Variables

If the vector $[x_t \;,\; y_t]'$ were observable, estimating the VAR(2) in eq. (11) or the state-space representation would deliver the same result: both models would generate identical impulse response functions - that is, they would have identical Vector Moving Average (VMA) representation. Estimation of the model (11) requires that the number of unobserved shocks m be equal to the number of observed variables, so that the matrix $\widetilde{B}$ is invertible. Estimation of the model (1), (2) requires that the number of unobserved shocks m be *at least* equal to the number of observable variables - else, the variance-covariance matrix of the model is singular (Ingram, Kocherlakota and Savin, 1994) make this point in the context of an RBC model with one endogenous state variable)[3].

Often the number of variables included in a VAR is substantially smaller than the size of the vector of endogenous state and control variables of the DSGE model the VAR is supposed to map into. There are valid reasons for this. First, econometric practice suggests the use of parsimonious models - since estimated VARs can include many lags, additional variables quickly reduce the available degrees of freedom. Second, some variables which are relevant for the dynamic behaviour of the theoretical model, such as capital, are poorly measured in the data, and therefore excluded from the data sample. Third, DSGE model have typically a smaller number of exogenous states than of observable variables - they are singular, and if they need to be estimated at all some observable variables must be dropped from the data sample.

[3]For the model (1), (2) to be written as a VAR all that is needed is that the matrix H be invertible. But note that invertibility of H does not guarantee a non-singular variance-covariance matrix. For the model (1), (2) to be written as a non-singular VAR, it must be possible to rewrite it as in eq. (11).

Is a VAR which excludes some of the variables included in the DSGE reference model still able to summarize the dynamics of the theoretical model? In most cases, a VAR will not be the reduced form representation of the time-series process describing the solution of the DSGE model for only a portion of the vector $[x_t \ , \ y_t]'$. In other words, the VAR will be a mis-specified representation of the theoretical model reduced form.

Let's maintain the assumption that the DSGE model (1), (2) is the true data-generating process. Suppose that the number of shocks is smaller than the number of endogenous variables ($m < n + r$). We can estimate the VAR(2) in eq. (11) omitting portions of the y_t vector, either because the variables are unobservable or because we try to obtain an invertible $\hat{B}$ matrix (where $\hat{B}$ is a partition of $\tilde{B}$) by reducing the size r of the y_t vector. Provided m equals the number of observable variables included in the VAR, omitting a portion of the y_t vector will not affect the reduced form dynamics of all the other variables. In fact eq. (10) shows that y_t is a function of its own lags only through its dependence on x_t. The estimated model would be:

$$\begin{aligned} \hat{Y}_t &= (\hat{A} + \hat{B}F^{22}\hat{B}^{-1})\hat{Y}_{t-1} - (\hat{B}F^{22}\hat{B}^{-1}\hat{A})\hat{Y}_{t-2} + \hat{B}\varepsilon_t \\ &= \Gamma_1\hat{Y}_{t-1} + \Gamma_2\hat{Y}_{t-2} + \hat{B}\varepsilon_t \end{aligned} \tag{13}$$

where $\hat{A}$, $\hat{B}$ are appropriate partitions of the matrices $\tilde{A}, \tilde{B}$ (the rows of the matrices $\tilde{A}, \tilde{B}$ corresponding to the observable y_t, and all the non-zero columns) and where $\hat{Y}$ indicates the observable portion of the Y vector[4]. Regardless of which portion of y_t we use, the VAR estimates of $(\hat{A}+\hat{B}F^{22}\hat{B}^{-1})$ and $(\hat{B}F^{22}\hat{B}^{-1}\hat{A})$ will be appropriate partitions of $(\tilde{A}+\tilde{B}F^{22}\tilde{B}^{-1})$ and $(\tilde{B}F^{22}\tilde{B}^{-1}\tilde{A})$, and correctly describe the model dynamics. No information is lost, since the omitted columns of the matrix $\widetilde{A}$ are empty to start with. The omitted rows of the matrix $\widetilde{B}$ are linear combinations of the remaining m rows (the rank of $\widetilde{B}$ is m). This implies that changing one of the components of the y_t vector will not change the estimated impulse response functions of the other components as the VAR is re-estimated. Obviously, if the theoretical model is not the true data-generating process, estimating the VAR using different portions of y_t would return different reduced form estimates.

If all the variables are state variables, the vector y_t is empty, and we must assume that the vector x_t is observable and that $m = n$ (therefore $\widetilde{A} = F^{11}$). This is, for example, the case in many three-variables model used in the optimal monetary policy literature, describing the joint evolution of output, inflation and the interest rate as a function of expected and lagged values of the three variables (Rudebusch and Svensson, 1999)[5].

[4]In this discussion we assume that the researcher will always estimate a VAR for at least m observable variables. If the number of exogenous shocks in the VAR is smaller than in the DSGE model, we cannot obtain impulse response functions for the innovation vector ε_t. Lutkepohl (1993) shows that when the true model is described by eq. (13), the data generating process for the observable $gx1$ vector $\widehat{Y}$ where $g < m$ is a VARMA(p,q) with $p \leq 2(n + r)$, $q \leq 2(n + r) - 2$.

[5]In the case when all variables are endogenous states, Zellner and Palm (1974) illustrate how to further solve the model in eq. (11) to obtain an ARMA representation for each individual component of Y_t (the *final form* equations).

Suppose now that a portion of the x_t vector is unobservable - for example, assume that no variable in x_t belongs to the available data sample. Then $\hat{A} = [0]$ and if we tried to estimate a VAR(2) using m observable variables the coefficient matrices in Γ_1, Γ_2 in eq. (13) would *not* be partitions of $(\tilde{A} + \tilde{B}F^{22}\tilde{B}^{-1})$ and $(\tilde{B}F^{22}\tilde{B}^{-1}\tilde{A})$. Omitting only a portion of the vector x_t means that $\hat{A}$ will include only some of the non-zero columns, and estimation of eq. (13) will return a mis-specified VAR(2). This exercise would instead provide a finite-order approximation of the matrix lag polynomial $\Phi(L)$, where $\Phi(L)\hat{Y}_t = \varepsilon_t$ is the VAR(∞) representation of the model (13).

4 VARMA and VAR Representation of DSGE models with Unobserved State Variables

Provided that: 1) the number of observable variables included in the estimated model is equal to the number of exogenous shocks; 2) the endogenous state vector x_t is observable, a finite order VAR representation of the model (10) exists, regardless of which endogenous control variables we include in the estimation. What form does the true data-generating process take when a portion of the x_t vector is unobservable?

We show that the model (9) has a finite order VARMA representation in terms of the endogenous control variables y_t. If it exists, the model also has an infinite order VAR representation (see Cooley and Dwyer, 1998, for a similar result in the context of a simple RBC model with one endogenous state variable).

Theorem 1 *Let*

$$\begin{aligned} \widetilde{y}_t &= \widetilde{H}'^1 x_t + \widetilde{H}'^2 z_t \\ x_t &= F^{11} x_{t-1} + F^{12} z_{t-1} \\ z_t &= Z(L)\varepsilon_t \\ \\ \Sigma &= E(\varepsilon_t \varepsilon_t') \ ; \ E(\varepsilon_t) = 0 \ ; \ E(\varepsilon_t \varepsilon_\tau') = 0 \qquad for\ \tau \neq t \end{aligned}$$

describe the dynamics of the vectors z_t, x_t, $\widetilde{y}_t$, where Σ is a diagonal matrix, $Z(L)$ is a matrix polynomial in the lag operator L defining a vector AR(p) stochastic process, $\widetilde{y}_t$ is a vector of dimension $r_1 \times 1$, x_t is a vector of dimension $n \times 1$ and z_t is a vector of dimension $m \times 1$. Assume the vector x_t is unobservable, the number of observable variables $y_t \in \widetilde{y}_t$ is r, where $r \leq r_1$, and $m = r$. Then the vector y_t has a VARMA(mn+p,mn) representation.

Proof: See the Appendix.

It is possible to prove that the order of the VARMA process for y_t is considerably reduced in the case $n = m = r$.

Corollary 1: *If* $n = 1$ *or if* $n = m = r$, *the vector* y_t *has a VARMA(m+p,m) representation.*

Proof: see the Appendix.

4.0.1 Discussion

The DSGE model in eq. (9) has a finite order VAR representation in terms of the observable variables $Y_t = [x_t \ , \ y_t]'$ when the full vector $x_t \in Y_t$. This representation id given by eq. ([**?**]). The theorem shows that if only $y_t \in Y_t$ (or some element of the vector x_t is omitted from the set of observable variables) the model has a VARMA representation of order *(nm+1,nm)*. In this case, eq. (13) is a mis-specified model of the system (9). Under certain regularity conditions, an infinite order VAR representation will also exist. Corollary 1 applies to the model analyzed in Cooley and Dwyer (1998), which has only one endogenous unobservable state variable, and to a number of small monetary models used in the optimal monetary policy literature, where all variables are also states and $n = m = r$ (see Walsh, 2003).

If we believe the DSGE model is well-specified, estimation of a finite order VAR when a portion of the x_t vector is unobservable will fail to uncover the true data-generating process, since the MA components are ignored - and therefore the true and estimated process for y_t will have different VMA representations. As a consequence, the impulse response functions may change drastically. The estimated system could still be useful for prediction, but it would be hard to give the impulse response functions a structural interpretation, since the DSGE model solution would be inconsistent with the estimated VAR.

4.1 Implicit Assumptions in estimation of DSGE Models

4.1.1 VAR representation

When is estimation of the DSGE model data-generating process with a finite-order VAR appropriate? The researcher is implicitly assuming that:

(a). The number of shocks is equal to the number of observable variables included in the data sample.

(b). The vector x_t belongs to the set of observable variables included in the data sample.

If some variables are omitted from a VAR model estimation, it must be the case that they are not state variables, else the model will be mis-specified. Omitting part of the y_t vector is instead inconsequential.

When assumptions (a), (b) do not hold, a finite order VAR may still be a very good approximation to the true data generating process. This is an empirical issue that needs to be addressed case by case. Section 6 explores this question in the context of a DSGE monetary model of the business cycle.

The discussion so far neglected the issue that has received the largest attention in the literature - how to identify the structural innovations from reduced-form shocks η_t. To explore the role of model mis-specification, in the simulation exercise we perform we will maintain the assumption that identification is possible by imposing restrictions on the estimated model. Section 6 discusses the identification and estimation issue.

4.1.2 State-space representation

An alternative to VAR estimation of DSGE models has been explored by a number of authors (Ireland, 2001a, 2001b, Rabanal and Rubio-Ramirez, 2003, Smets and Wouters, 2002). The state-space form of the model (1) (2) can be estimated - regardless of whether the vector of endogenous state variables x_t is observable - using Maximum Likelihood or Bayesian methodologies (Hamilton, 1994). For the estimated state-space model to be the correct representation of the DSGE data-generating process we must assume (beyond any necessary identification assumption):

a'. The number of shocks is not smaller than the number of observable variables included in the data sample.

b'. The dimension of the vector x_t is known.

For the state-space representation to be consistent with the DSGE model we need less strict assumptions than for a VAR. The researcher need not be sure that she is including *all* the state variables in the model. If a state variable is unobservable or measured with error, the state-space system can still be estimated and will still be consistent with the underlying DSGE model. As section 6 shows, a mis-specified, correctly identified VAR can instead lead to very large errors in estimating the data-generating process.

5 Estimation of the shocks vector

The result in the previous section has two immediate implications. If a finite-order VAR is not a correct specification of the DSGE model (even if identification is possible): 1) impulse response functions from the DSGE model cannot be compared to those obtained from the VAR, since the two models have inconsistent VMA representations; 2) the VAR cannot be used to estimate the vector of structural innovations consistent with the DSGE model.

The first implication has been investigated in a number of papers with regard to specific models (Cooley and Dwyer, 1998, Linde, 2003). In the following we examine the second implication. The researcher may be interested in estimating the shocks vector for at least two reasons. First, the series of shocks is of interest in its own right, for example for historical decomposition of the shocks driving business cycle fluctuations, or to ascertain deviations of monetary policy from its systematic, endogenous behaviour. Second, using a DSGE models to build counterfactuals - paths of the economy under alternative assumptions or policy

- forces the researcher to try and recover the series of shocks which generated the data observed in the first place. While the first exercise can be conducted in the framework of an atheoretical VAR, the second one necessarily requires the building of *model-consistent* shocks: the shock vector that fed through the model would generate the data observed (see Cole, Ohanian and Leung, 2003, King and Rebelo, 1998 and Smith and Zin, 1997, for applications of the methodology in the context of real business cycle models, while Ravenna, 2002 extends the method to a multiple shock model of the monetary business cycle to build model-consistent counterfactual histories under alternative monetary policies).

5.1 Using a VAR representation

Assume eq. (13) correctly models the true data-generating process. Provided identification is possible, it is straightforward to obtain the vector ε_t from the VAR representation:

$$\varepsilon_t = \hat{B}^{-1}(\hat{Y}_t - \Gamma_1 \hat{Y}_{t-1} - \Gamma_2 \hat{Y}_{t-2}) \tag{14}$$

To calculate ε_t, the following assumptions must be met:

Assumptions:

1. Var *The initial values $\hat{Y}_{t-1}$, $\hat{Y}_{t-2}$ are known*
2. Var *There is no measurement error in the estimated model*
3. Var *The number of observable variables is equal to the number of exogenous shocks*
4. Var *The vector $\hat{Y}$ includes all the state variables x_t*

Assumptions 3 and 4 were discussed earlier. Assumptions 1 states that we need two initial observations of the vector $\hat{Y}$ to start the recursion. Assumption 2 states that if we wish ε_t to be the vector of unobservable structural innovations, there must be no measurement error in the system. Else the recursion would report $\varepsilon_t + w_t$ where w_t is a measurement error vector.

5.2 Using a state-space representation

The state-space representation in eqs. (1), (2) can be inverted to obtain the vector ξ_t and ε_t given the vector of observable variables y_t[6].

As in the case of the VAR representation, this is possible only under certain conditions imposed on the matrices F, H. The structure of the system defined by equations (1) and (2) implies that once we have estimated the vector ε_{t-1}, or equivalently the vector ξ^2_{t-1}, the

[6]This specification does not preclude any or all the components of the vector x_t to be observable. Simply relabel the observable variables vector as $\widetilde{y}_t = [y_t, x_t]'$.

vector ξ_t^1 can be computed using the matrices F_{11} and F_{12}. Thus we can use the vector of observable variables y_t to compute the $m \times 1$ vector ξ_t^2 by solving the matrix equation:

$$y_t = H'^1 \xi_t^1 + H'^2 \xi_t^2 \tag{15}$$

The matrix H'^2 has dimension $r \times m$. Economic models will usually have a number r of observable variables larger then the number m of exogenous forcing terms. In this situation the matrix H'^2 is non-invertible, and the problem of solving for the vector ξ_t is not well defined[7].

There are three ways to solve the non-invertibility problem.

I: **increase the number m of exogenous state variables until $m = r$.** This method is problematic in most models unless the number of endogenous variables r is constrained to a minimum. It is obviously undesirable to introduce in a model a large number of additional driving forces without economic justification.

II: **assume that the observations of the vector y_t are affected by a measurement error.** Eq. (2) is then written as:

$$y_t = H' \xi_t + w_t \tag{16}$$

where w_t is an $r \times 1$ measurement error vector. It is assumed w_t follows a vector stochastic process with variance-covariance matrix given by:

$$E(w_t w_t') = R \tag{17}$$
$$E(w_t w_\tau') = 0 \qquad for\, \tau \neq t \tag{18}$$

and that $E(v_t w_\tau') = 0$ for any τ. Under these assumptions and given the sample $[y_1 ... y_T]$ the vector ξ_t can be estimated from the system defined by equations (1) and (16) using the Kalman filter algorithm. This method has the advantage of providing the best unbiased estimate of the vector ξ_t using the full structure of the model and observations on all the endogenous variables in the vector y_t. The drawback of using the Kalman filter is that we are forced to introduce a measurement error vector in the model representation. The state vector estimate (as well as its variance) would then be a function of the matrix R. The larger the variance of w_t relative to the variance of ε_t, the closer to the initial value ξ_0 (or to $E(\xi_0)$ if the initial value is unknown) will the estimate of ξ_t be.

[7] Ingram et al. (1994) discuss the consequences of the non-invertibility for the estimates of total factor productivity shocks, and observe that most business cycle models' representation is in fact singular.

III**: use only a portion of the observable variables vector y_t to recover the vector ξ_t^2.** We can assume, as we did when deriving the VAR representation in section 2, that only m out of the r variables in y_t can be measured. Eq. (15) becomes:

$$^{m\times 1}y_t =^{m\times n} H'^1 \xi_t^1 +^{m\times m} H'^2 \xi_t^2$$

which can be solved for ξ_t^2 provided the matrix H'^2 is non-singular, as will generally be the case[8]. Given an initial value for the state vector ξ_0, the solution for the vector ξ_1 is:

$$\xi_1^1 = F^{11}\xi_0^1 + F^{12}\xi_0^2 \quad (19)$$

$$\xi_1^2 = (H'^2)^{-1}[y_1 - H'^1\xi_1^1] \quad (20)$$

In general, the unknown state vector ξ at time t is given by the recursion:

$$\xi_{t|T}^1 = F^{11}\xi_{t-1|T}^1 + F^{12}\xi_{t-1|T}^2 = F^{11}\xi_{t-1}^1 + F^{12}\xi_{t-1}^2 = \xi_t^1 \quad (21)$$

$$\xi_{t|T}^2 = (H'^2)^{-1}[y_t - H'^1\xi_{t|T}^1] = (H'^2)^{-1}[y_t - H'^1\xi_t^1] = \xi_t^2 \quad (22)$$

where $\xi_{t|T}$ indicates the estimate of ξ_t (in this case an exact solution) conditional on the observation sample $[y_1...y_T]$. While this solution method does not make efficient use of all the structure of the model, it is in fact the most commonly used. For example, Ingram et al. (1994) note that whenever the econometrician selects the production function equation of a Real Business Cycle model to estimate the technology shock, she is neglecting all the other first order conditions in the model which are as well functions of the observable variables and the technology shock and could be used for estimation. As showed earlier, also a VAR representation of a DSGE reduced form implicitly assumes that the theoretical model is the true data-generating process - therefore omission of a portion of the y_t vector does not lead to mis-specification of the observable variables dynamics.

5.3 The Kalman filter algorithm and invertible state space representations

The Kalman Filter algorithm was devised to estimate the unobservable state vector ξ in systems that have a state-space representation. In the following we prove that under assumptions $1, 2, 3$ established for the VAR representation the Kalman filter smoothed estimate of the vector ξ_t is equal to eqs. (21) and (22) - the solution of the system in eqs. (1) and (2) obtained by using only a portion of the observable variables vector y_t to recover the vector ξ_t^2. Once we know the vector ξ_t it is straightforward to use eq. (1) to obtain the vector of i.i.d. innovations ε_t.

Assume the vector y_t is observed from time 1 to T. For the system:

[8]Note that if only m variables of the vector y_t are observable, the $(m-r)$ variables which are not observable do not need to enter the vector ξ_t since their dynamics does not contribute to the dynamics of the other variables in the state vector ξ_t.

$$\begin{aligned}\xi_{t+1} &= F\xi_t + v_{t+1}\\ y_t &= H^{'}\xi_t + w_t\end{aligned}$$

and given the assumptions in equations (3) to (8), (17), (18) the *Kalman filtered estimate* of the vector ξ_t is defined by the recursion (Hamilton, 1994):

$$\begin{aligned}\xi_{t|t} &= \xi_{t|t-1} + P_{t|t-1}H(H'P_{t|t-1}H+R)^{-1}(y_t - H'\xi_{t|t-1})\\ \xi_{t+1|t} &= F\xi_{t|t}\\ P_{t|t} &= P_{t|t-1} - P_{t|t-1}H(H'P_{t|t-1}H+R)^{-1}H'P_{t|t-1}\\ P_{t+1|t} &= FP_{t|t}F' + Q\end{aligned}$$

where $\xi_{t+1|t}$ denotes the linear projection of ξ_{t+1} on the sample $[y_1...y_t]$ and a constant:

$$\xi_{t+1|t} \equiv \hat{E}(\xi_t|y_1...y_t)$$

and $P_{t+1|t}$ is the mean squared error (MSE) matrix associated with the forecast $\xi_{t+1|t}$:

$$P_{t+1|t} \equiv E[(\xi_{t+1} - \xi_{t+1|t})(\xi_{t+1} - \xi_{t+1|t})')]$$

The *Kalman smoothed estimate* $\xi_{t|T} \equiv \hat{E}(\xi_t|y_1...y_T)$ of the vector ξ_t is based on the full sample of observable variables contained in the vector y_t. It is defined by the recursion:

$$\begin{aligned}\xi_{t|T} &= \xi_{t|t} + J_t(\xi_{t+1|T} - \xi_{t+1|t})\\ J_t &= P_{t|t}F'P^{-1}_{t+1|t}\\ P_{t|T} &= P_{t|t} + J_t(P_{t+1|T} - P_{t+1|t})J_t'\end{aligned}$$

5.3.1 Assumptions

1. Kalman *The initial value ξ_0 of the vector ξ_t is known:*

$$\xi_{0|0} = \xi_0$$

2. Kalman *The measurement error vector is zero at all dates:*

$$w_t = \overrightarrow{0}\ \forall\ t$$

3. Kalman *The number of observable variables is equal to the number of non-zero elements of the vector v_t:*

$$r = m$$

Assumption 1 states that the first n elements of the vector ξ_1 are known with certainty, therefore the upper-left $n \times n$ sub-matrix of $P_{1|0}$ is the zero matrix. It is easy to see that this assumption is the same as assumption 1 Var. Assumption 1 could in fact be replaced by $\xi^1_{1|0} = \xi^1_1$, since the vector ξ^2_0 is needed to prove the theorem only because knowledge of ξ^1_0 and ξ^2_0 implies knowledge of ξ^1_1 with certainty[9]. The VAR model already assumes that ξ^1_1 belongs to the set of observable variables $\widehat{Y}_t$. Assumption 2 states that the only exogenous variable driving the evolution of the vector y_t is the vector of unobserved states ξ_t - as assumption 2 Var. Assumption 3 makes sure that the system in eqs. (1) and (2) is non-singular so that under assumptions 1 and 2 there exist a solution to the Kalman filter recursion, and is equal to assumption 3 Var.

5.3.2 Theorem

Theorem 2 *Let eqs. (1), (2) and eqs. (3) to (8), (17), (18) describe the dynamics of the vectors y_t and ξ_t. Under assumptions $1, 2, 3$ the Kalman smoothed estimate of the vector ξ_t is given by eqs. (21) and (22):*

$$\xi^1_{t|T} = F^{11}\xi^1_{t-1|T} + F^{12}\xi^2_{t-1|T} = \xi^1_t$$
$$\xi^2_{t|T} = (H'^2)^{-1}[y_t - H'^1\xi^1_{t|T}] = \xi^2_t$$

Proof: see the Appendix.

5.3.3 Discussion

The central result of the theorem is intuitive: if the vector of state variables which solves the system in eq. (1) and (2) is unique, the linear projection of the state vector is unique too and is equal to the system's solution. The theorem makes explicit the assumptions that are needed for the system in eqs. (1) and (2) to have a unique solution.

First, the measurement error vector w_t must be the zero vector at all dates. Without this assumption, the variance-covariance matrix R of the vector w_t would be non-empty. Neither equation (45) or (49) in the Appendix would hold. Since then the solution ξ_t would not be unique, the variance of the estimate $\xi_{t|T}$ would be positive.

Second, if $w_t = 0 \ \forall \ t$, the number of observable variables must be no larger than the dimension of the white noise innovations vector v_t. Otherwise the Kalman filtered estimate is not defined. In fact, if $r > m$ it is still true that:

$$^{r\times n+m}\left(H'P_{1|0}\right) = \left[\begin{array}{cc} ^{r\times n}0 & ^{r\times m}(H'^2)(P^{22}_{1|0}) \end{array}\right]$$

But is easy to check that the matrix:

$$^{r\times r}(H'P_{1|0}H + R) = \left(H'P_{1|0}H\right) = [H'^2 P^{22}_{1|0} H^2]$$

[9]Note that knowing only ξ^2_1 is not sufficient to start the recursion as shown in the theorem, unless the matrix $H'^1$is $m \times m$, in which case the equation $y_1 = H'^1\xi^1_1 + H'^2\xi^2_1$ can be solved for ξ^1_1.

will have two proportional rows and will thus be non-invertible. Note that this would not be true if $w_t \neq 0$ *or* if there is uncertainty about the initial vector ξ_0. Therefore, even in the case of a singular system, it is still possible to derive an estimate of ξ_t provided we allow for uncertainty in the initial value. Of course in this instance the vector ξ_t is not unique anymore.

If $r < m$, the solution ξ_t is not unique. The system in eqs. (1) and (2) has then an infinite number of solutions. The Kalman filter will provide the best linear unbiased estimate of the unknown vector given the information on the unconditional distribution of the stochastic vector v_t.

Third, if ξ_1^1 is unknown the vector $P_{1|0}^{11}$ is a non-zero diagonal matrix. Then $P_{1|1} \neq 0$, $P_{2|1} \neq P_{1|0}$ and $P_{t|t}$ will be non-zero for all subsequent periods. In this case the variance of the estimate $\xi_{t|T}$ would be positive.

6 Can VARs estimate model-consistent monetary policy shocks?

Macroeconomists estimating structural models have devoted increasing attention to assuming VAR identifying restrictions consistent with the DSGE reference model (Pagan, 2003, Linde', 2003). For example, Christiano et. al. (2003) estimate a DSGE model by minimizing the distance between the impulse response function to a policy shock in the model, and to an identified policy shock in an estimated VAR. The VAR identification strategy adopted relies on the same informational lags built into the model. The authors fail though to use a reduced-form representation consistent with the theoretical model. The estimated VAR does not include all the model's state variables. The previous sections showed that in this case the correct data-generating process has a VARMA representation.

Whether a mis-specified VAR is a good enough approximation to the true data generating process is an empirical issue: including enough lags in the estimation may approximate the VAR(∞) representation of the system[10].

In this section we show that this assumption is misleading. Using a DSGE model of a small open economy, we estimate the vector ε_t from the VAR representation (14) of the model, and compare it to the one estimated from the state-space model (21), (22).

To allow a fair comparison, we make a number of simplifying assumptions. First, we assume the matrices $\hat{B}$ and F are known. Therefore identification of the orthogonal shocks vector is possible. While this assumption lets us identify *all* shocks, we focus on the monetary policy shock, which is the one for which informational lags allow the most popular identification procedure in the VAR literature, Cholesky decomposition. Second, we use data simulated by the DSGE model. Using real data makes the results sensitive to which

[10] Lutkepohl (1993) discuss in detail the implications of reducing the dimension of the estimated system (i.e. estimation of a $g-$dimensional VARMA process when the true data generating process is $m-$dimensional, $m > g$) and fitting finite order VAR models to infinite order processes.

observable variable is included in the vector y_t. Only if the theoretical model is the correct data generating process the solution ξ_t^2 will be independent of the $r - m$ variables chosen to be excluded from the vector y_t. Third, we estimate the VAR over a sample of 10,000,000 data points. Such a long time series assures that we are looking at asymptotic results, and that any error in the estimate of ε_t is not due to variance in the estimator.

Under assumptions VAR $1, 2, 3$ and 4, specified in section 5, the estimated VAR(2) recovers the correct vector ε_t. As a consequence, adding further lags to the VAR does not improve the accuracy of the estimate. Similarly, the estimate of the state-space model converges to the true model (1), (2). Therefore the recursion (21), (22) recovers the true shocks ε_t. This will be true regardless of the observable variables included (neglecting any estimation issues). The same is not true for the VAR representation.

6.1 A DSGE small open economy model

Below we report the loglinearized equations describing a micro-founded small open economy general equilibrium model with nominal rigidities. The model is akin to many recent staggered-price adjustment open economy models (see Devereux, 2001, Gali and Monacelli, 2002, Natalucci and Ravenna, 2002, Walsh, 2003). Complete derivation of the model and parametrization on the Canadian economy can be found in Ravenna (2003).

The model lends itself to our purposes since it features seven exogenous stochastic shocks and five state variables. It is easy to envision situations in which the econometrician estimating a seven-variables VAR omits one or more of the state variables. Monetary policy is described by a Taylor rule, an assumption often used in empirical VARs to identify the policy shock.

Household sector:

$$c_t = (1-\gamma)c_{H,t} + \gamma c_{F,t} \tag{23}$$

$$muc_t = \kappa_1 d_t - \kappa_2 E_t d_{t+1} - \kappa_3 c_t + \kappa_4 c_{t-1} + \kappa_5 E_t c_{t+1} \tag{24}$$

$$c_{H,t} - c_{F,t} = \rho s_t \tag{25}$$

$$mrs_t = \eta n_t - muc_t = \zeta_t \tag{26}$$

$$muc_t = r_t + E_t muc_{t+1} \tag{27}$$

$$i_t = i_t^* + E_t \Delta e_{t+1} \tag{28}$$

$$r_t = i_t - E_t \pi_{t+1} \tag{29}$$

Production sector:

$$y_t = a_t + n_t \tag{30}$$

$$\zeta_t = \zeta_{t-1} + \xi_t - \pi_t \tag{31}$$

$$mc_t = \zeta_t + n_t - y_t + \pi_t - \pi_{H,t} + \gamma s_{t-1} + u_t \tag{32}$$

$$\pi_{H,t} = \lambda_1 mc_t + \beta E_t \pi_{H,t+1} \tag{33}$$

Foreign sector:

$$c^*_{H,t} = \rho s_t + c^*_t \quad (34)$$
$$s_t - s_{t-1} = \Delta e_t + p^*_{F,t} - p^*_{F,t-1} - \pi_{H,t} \quad (35)$$
$$\pi_t = \pi_{H,t} + \gamma(s_t - s_{t-1}) \quad (36)$$

Market clearing:

$$y_t = \frac{C_H}{Y} c_{H,t} + \frac{C^*_H}{Y} c^*_{H,t} \quad (37)$$
$$b_t = \mu_1 b_{t-1} + \mu_2 c^*_{H,t} - \mu_3 (s_t + c_{F,t}) \quad (38)$$

Capitale letters denote steady state values. All variables are measured in percent deviation from the zero net foreign asset position steady state. Eq. (23) defines the consumption index c aggregating home (c_H) and foreign (c_F) produced good consumption baskets. Eq. (24) defines the marginal utility of consumption muc with habit-persistent preferences, where d_t is a preference shock. Setting $b = 0$ returns log-linear preferences over the consumption basket. $\kappa_1, \kappa_2, \kappa_3, \kappa_4, \kappa_5$ are functions of the model's parameters and steady state values. Eq. (25) gives the intratemporal choice between home and foreign goods, as a function of the terms of trade s. Eq. (26) equates the marginal rate of substitution (mrs) between labor hours (n) and consumption to the real wage ζ. Eqs. (27), (28), (29) are obtained from the household's first order conditions over the real, nominal and foreign-currency denominated assets, where r_t is the consumption-based real interest rate, i_t is the nominal interest rate, i^*_t is the foreign nominal interest rate, Δe_t is the nominal exchange rate depreciation, and π_t is the CPI inflation rate.

In the production sector, eq. (30) is the production function for domestic output y, where a is a productivity shock. Eq. (31) is the definition of the real wage ζ_t, where ξ is wage inflation. Eq. (32) defines the real marginal cost mc, where π_H is domestic-produced good inflation, s_t is the terms of trade and u_t is a cost-push shock. Eq. (33) describes the domestic inflation process and is derived from the firms' staggered price adjustment optimality condition.

Eq. (34) gives the foreign demand for home goods c^*_H, as a function of the exogenously given total foreign consumption c^*. Eq. (35) defines the terms of trade depreciation, where $p^*_{F,t}$ is the exogenously given price of foreign goods in terms of foreign currency. Eq. (36) is the relationship between domestic (π_H) and CPI (π) inflation. Eq. (37) is the economy-wide resource constraint, and eq. (38) gives the foreign asset accumulation equation for the foreign issued bonds b_t, where μ_1, μ_2, μ_3 are functions of the model's parameters and steady state values.

The monetary policy rule for the baseline specification is:

$$i_t = \omega_p E_t \pi_{t+1} + \omega_y y_t + \omega_e \Delta e_t + v_t$$

where we set $\omega_p = 2, \omega_y = 1.5, \omega_e = 0.5$. The rule includes an autocorrelated, unexpected policy shock v_t. The exogenous stochastic processes for the preference shifter, the technology shock, the cost-push shock, the world interest rate, the imports' price and the aggregate foreign consumption demand follow an AR(1) specification:

$$
\begin{aligned}
d_t &= \rho_d d_{t-1} + \varepsilon_{d,t} \\
a_t &= \rho_a a_{t-1} + \varepsilon_{a,t} \\
u_t &= \rho_u u_{t-1} + \varepsilon_{u,t} \\
v_t &= \rho_v v_{t-1} + \varepsilon_{v,t} \\
i_t^* &= \rho_{i^*} + \varepsilon_{i^*,t} \\
p_{F,t}^* &= \rho_p p_{F,t-1}^* + \varepsilon_{p,t} \\
c_t^* &= \rho_c c_{t-1}^* + \varepsilon_{c,t}
\end{aligned}
$$

where $\rho_a = \rho_v = \rho_{i^*} = 0.9$, $\rho_d = \rho_u = \rho_p = \rho_c = 0.8$ and the components of the vector ε are i.i.d. stochastic processes. The innovation vector ε volatility is set to $\sigma_a = 0.8$, $\sigma_d = 2.4$, $\sigma_u = 0.15$, $\sigma_v = 0.3$, $\sigma_p = 0.8$, $\sigma_c = 0.8$, $\sigma_{i^*} = 0.3$. These are values in line with calibrated and estimated staggered price adjustment models (Natalucci and Ravenna, 2002, Rabanal and Rubio-Ramirez, 2003).

6.2 Finite order VAR approximation to the DSGE model with unobserved state variables

The estimated VAR(2) includes the observable variables $\iota_t, y_t, \pi_t, c_t^*, \Delta e_t, mc_t, \pi_{H,t}$. These are variables commonly included in open-economy VARs (see Favero, 2001). The VAR does not include any of the model's state variables: $\zeta_t, s_t, b_t, p_{F,t}^*, c_t$.

Table 1 shows the root mean square distance (RMSE) between the true innovation time series for the policy shock ε_v (equal to the series estimated with a state-space model) and the series estimated with the VAR(2).

Table 1

Relative Root Mean Square Error - Baseline model

Innovation	Relative RMSE
policy	**696%**
productivity	722%
demand	76%
cost-push	6200%
export demand	227%
import price	301%
foreign interest rate	92%

Note: Root mean square distance between the VAR(2)-estimated vector ε and the true vector calculated over 250,000 40-quarters simulated time series. Data are generated by the DSGE model. The RMSE is scaled by the standard deviation of the corresponding shock.

The performance of the VAR is lacking: the root mean square error for ε_v is about 7 times larger than the standard deviation of the shock itself. Figure 1 plots a 10-year sample path to show the impact of VAR mis-specification.

While usually the econometrician does not have available enough restrictions to be able to identify all shocks, we assume that the matrix $\hat{B}$ is known, and therefore we can examine the impact of the mis-specification on all innovation estimates. Table 1 shows that the foreign interest rate and the demand shock are estimated with higher precision than the other shocks. Figure 2 shows a 10 year sample path for six shocks. Clearly the VAR estimated innovation shocks ε_d and ε_{i^*}, while not completely accurate, follow remarkably closely the true shocks. This result can be readily explained by examining the list of variables included in the VAR. The variables i_t, i_t^*, $E_t\Delta e_{t+1}$ are linked by the uncovered interest parity equations. Both the domestic and foreign interest rate are highly autocorrelated. Therefore ε_{i^*} can be tightly estimated once the three variables are included in the model. Intuitively, the VAR performs better when we include variables which are linked by an equilibrium relationship and at the same time are highly correlated with the state variables.

6.3 Are all models equal?

The data-generating process has a large impact on how well a finite order VAR approximates the true process. The policy rule in the baseline model assumed autocorrelated ε_v shocks, as in Rudebusch (2002). A more general policy rule includes explicit interest rate smoothing. Let policy be described by:

$$\begin{aligned} i_t &= (1-\chi)[\omega_p E_t\pi_{t+1} + \omega_y y_t + \omega_e \Delta e_t] + \chi i_{t-1} + v_t \\ v_t &= \rho_v v_{t-1} + \varepsilon_{v,t} \end{aligned}$$

Table 2 reports RMSE values under different policy regimes. Policy 1 includes an interest rate smoothing objective: the central bank adjusts the policy instruments only slowly towards the target. As a consequence i_t becomes a state variable and the estimate of the policy shocks is more accurate. A large smoothing coefficient χ implies that following any shock, the interest rate i_t is very persistent. Then a large part of the variance of i_t is explained by i_{t-1}, so it comes as no surprise that this model fares much better than the baseline. Simply doubling the variance of the policy shock in the baseline model (policy 2) lowers the RMSE by nearly half. Including a contemporaneous inflation target (policy 4) rather than a target in terms of expected inflation allows a very large improvement in the estimate's precision. But this is not a sufficient condition for accurate estimates: changing the target from π to domestic inflation π_H, in policy 5, yields a RMSE nearly as large as in the baseline model.

Table 2

Relative Root Mean Square Error of ε_i - comparison across policy rules

Policy Rule	Relative RMSE
baseline: $\chi = 0$, $\omega_p = 2$, $\omega_y = 1.5$, $\omega_e = 0.5$, $\rho_v = 0.9$	**696%**
1: $\chi = 0.9$, $\omega_p = 2$, $\omega_y = 1.5$, $\omega_e = 0.5$, $\rho_v = 0$	225%
2: $\chi = 0$, $\omega_p = 2$, $\omega_y = 1.5$, $\omega_e = 0.5$, $\rho_v = 0.9$; $\sigma_v = 0.6$	353%
4: contemporaneous inflation target	141%
5: contemporaneous domestic inflation target	667%

Note: Root mean square distance between the VAR(2)-estimated series ε_v and the true series calculated over 250,000 40-quarters simulated time series. Data are generated by the DSGE model. The RMSE is scaled by the standard deviation of ε_v

The baseline model labor market can be modified to include staggered wage adjustment, as in Erceg et al. (1999). Then the labor supply equilibrium condition $mrs_t = \zeta_t$ is substituted by the forward-looking wage adjustment equation:

$$\xi_t = \lambda_2[mrs_t - \zeta_t] + \beta E_t \xi_{t+1}$$

This modification makes ζ_t very persistent. The real wage now carries much more additional information in the VAR, since in the baseline flexible wage model the correlation of ζ_t with a variable already included in the system, y_t, is very high. Table 3 reports the resulting large improvement in accuracy. Adding an interest rate smoothing policy makes the VAR estimate of ε_v nearly 100% accurate, as is clear from Figures 3 and 4.

Table 3

Relative Root Mean Square Error of ε_i - comparison across models

Model	Relative RMSE
baseline	**696%**
sticky wages	121%
sticky wages and interest rate smoothing	12%

Note: Root mean square distance between the VAR(2)-estimated series ε_v and the true series calculated over 250,000 40-quarters simulated time series. Data are generated by the DSGE model. The RMSE is scaled by the standard deviation of ε_v

6.4 The choice of observable variables

The choice of observable variables to include in the VAR plays an important role, and will depend on the question the researcher is investigating. Table 4 shows the consequences of excluding i_t from the set of observable variables, replacing it with labor hours n_t. The econometrician interested in the policy innovation series would hardly make this choice. The RMSE of ε_v increases by a factor of 2.5 compared to Table 1. But once both n_t and y_t are included in the VAR, the production function implies that the shock ε_a can be recovered with nearly complete accuracy (Fig. 5 and 6)

Table 4

Relative Root Mean Square Error - Baseline model - i_t not included among observables

Innovation	Relative RMSE
policy	**1746%**
productivity	0.65%
demand	785%
cost-push	17200%
export demand	227%
import price	301%
foreign interest rate	1095%

Note: Root mean square distance between the VAR(2)-estimated vector ε and the true vector calculated over 250,000 40-quarters simulated time series. Data are generated by the DSGE model. The RMSE is scaled by the standard deviation of the respective shock. VAR(2) data includes $n_t, y_t, \pi_t, c_t^*, \Delta e_t, mc_t, \pi_{H,t}$

6.5 VAR lags and accuracy of approximation

Does increasing the number of lags lead to a rapidly increasing accuracy of the estimated shock vector? In the case of the baseline model, Table 5 shows that moving from a VAR(2) to a VAR(6) only marginally reduces the root mean square distance between the estimated and true ε_v.

Table 5

Relative Root Mean Square Error of ε_v - VAR(n)

Lags included in VAR	Relative RMSE
2	696%
6	695%

Note: Root mean square distance between the VAR(2) and VAR(6)-estimated series ε_v and the true series calculated over 250,000 40-quarters simulated time series. Data are generated by the DSGE model. The RMSE is scaled by the standard deviation of ε_v. VAR(2) data includes $i_t, y_t, \pi_t, c_t^*, \Delta e_t, mc_t, \pi_{H,t}$

6.6 Identification, mis-specification, and system estimation

The poor performance of the VAR approximation to the DSGE model derives entirely from the estimated model's mis-specification. Identification issues play no role in the result. In fact we endow the econometrician with knowledge of the matrix $\widetilde{B}$ to disentangle the structural innovations ε_t from the reduced form shocks η_t.

The mis-specification problem has been recognized in the literature - for example, Canova (2004) notes that the VAR(n) representation of a DSGE model is transformed into a VARMA whenever a variable is omitted (see also Cooley and Dwyer, 1998). Nevertheless, its empirical relevance has been neglected. The vast majority of researchers implicitly assumes that either all the state variables from the reference DSGE model have been included in the estimated VAR (and therefore the model has a VAR(n) representation, or a VAR(1) representation if the shocks z_t are *i.i.d.*), or that enough lags are included in the VAR so that any remaining MA component is negligible.

The literature on VAR modeling focused instead on the alternative identification schemes. Researchers base their identifying assumptions on different theories, and therefore will reach different conclusions on what portion of the observable variables' volatilities can be attributed to each of the identified shocks (see Canova, 1995, and Uhlig, 2004). The so called 'price puzzle' - the result that in estimated VARs on postwar US data a contractionary monetary policy shock leads to a persistent price *increase* - is often labeled as a 'failure to identify correctly the policy shock'. The price puzzle is truly a mis-specification problem. Sims (1992) discusses how it could be due to a missing element in the policy rule included in the VAR, and suggests adding commodity prices to the system. The fact that adding commodity prices solves the price puzzle in estimated VARs implies that commodity prices are a model's state variable. Were this not the case, Theorem 1 shows that the VAR would not be mis-specified.

A number of papers discusses the consequences of imposing VAR identifying restrictions inconsistent with the reference DSGE model, assumed to be the true data-generating process. Canova and Pina (1999) compare the theoretical VAR representation of a DSGE model with a 4-variables VAR estimated under a variety of (model-inconsistent) identifying assumptions. Since they include in the VAR all state variables, the distance between the true and estimated VMA representation can only be ascribed to mis-identification (see also Cochrane, 1998).

Researchers are increasingly aware of the risks from mis-identification when comparing DSGE and VAR models. Christiano et al. (2003) take great care of using identifying restrictions for the estimated VAR consistent with the theoretical model. But they fail to include capital (a state variable) in the estimated VAR. Whether this mis-specification leads to a severe bias in their results is difficult to assess.

Is then Maximum Likelihood estimation of the state space model a sure way to avoid mis-specifications? Clearly, the state-space representation allows much more flexibility - it is even possible to add extra unobserved state variables, which (asymptotically) will not affect the model's estimate. But all along we have avoided the problems which can arise in estimation over limited samples. As VARs, also state-space model need a number of restrictions to be identified, and in practice the identification requirements of the state-space representation may be more taxing than in a VAR (see Hamilton, 1994, and Canova, 1992, for an overview of the problems connected to the state-space estimation approach). One recognized advantage of the Cholesky identification scheme in SVARs is that it allows the econometrician to estimate the impact to an orthogonalized policy shock, without having to take a stand on all the other shocks. Essentially, the VAR result on the policy shock impulse response function depends on correct identification of only a column of the matrix $\widetilde{B}$.

7 Conclusions

Increasingly the explanatory power of DSGE models is tested by comparing the model impulse response functions to those obtained from an estimated VAR. Some researchers estimate model parameters by minimizing the distance between the model's and the estimated VAR impulse response functions.

This paper showed that testing or estimating DSGE models using the VAR representation is only appropriate when a number of conditions is met. We provided a general proof of the reduced form representation in which DSGE model map, and showed that unless all state variables are included in the system the model takes a VARMA representation. We showed that neglecting the MA terms can have serious implications for estimating the vector of structural innovation, regardless of identification or estimation problems.

The state-space representation can accommodate the DSGE model dynamics under a less strict set of conditions than a VAR. The paper proves under what conditions the Kalman-filtered shocks vector will return the true vector of innovations that generated the observable data sample. The state-space model advantage lies in the fact that we do not need to assume all the model's state variables are observable to have a correctly specified model.

Structural VAR are not necessarily inappropriate models for the data. *Atheoretical* VARs have much to tell: they summarize the dynamics of the data with as few restrictions as possible. Assuming though that the dynamics they describe can always be obtained from the structural model we are interested in testing is misleading. If we wish to take DSGE models seriously - that is, expect that they can account for the correlations in the data - then we should compare them to *consistent* representations of the data. The empirical relevance of the critique of VAR models we provide has to be tested case by case. Once small sample estimation problems are taken into considerations, VARs may still provide a more accurate representation of the data-generating process.

References

[1] Canova, F., (1995), 'VAR: specification, estimation, testing and forecasting', in Pesaran, H. and Wickens, M., eds., *Handbook of applied econometrics* : 31-65.

[2] - , (2002), "Validating monetary DSGE models though VARs', CEPR Discussion Paper 3442.

[3] - , (2004), 'Methods for applied research: VAR models', mimeo, IGIER-Universita' Bocconi.

[4] - , and Pina, J., (1999), 'Monetary policy mis-specification in VAR models', CEPR Discussion Paper 2333.

[5] Cho, S. and Moreno, A., (2002), 'A structural estimation and interpretation of the New Keynesian macro model', mimeo, Columbia University.

[6] Chow, G. and Kwan, Y., (1998), 'How the basic RBC model fails to explain US time series', *Journal of Monetary Economics* 41: 301-319.

[7] Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans, (2003), "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," mimeo and NBER Working Paper No. 8403, 2001.

[8] Cochrane, J., (1998), 'What do VARs mean? MEasuring the output effect of monetary policy', *Journal of Monetary Economics* 41: 277-299.

[9] Cole, H., Ohanian, L. and Leung, R., 'Deflation, real; wages and the international great depression: a productivity puzzle', Federal Reserve Bank of Minneapolis Staff Report, forthcoming.

[10] Cooley, Thomas and Dwyer, Mark, (1998), 'Business cycle analysis without much theory. A look at structural VARs', Journal of Econometrics 83: 57-88.

[11] Devereux, M., (2001), 'Monetary policy, exchange rate flexibility and exchange rate pass-through', in Revisiting the Case for Flexible Exchange Rate, Bank of Canada.

[12] Erceg, C., Guerrieri, L. and Gust, C., (2004), 'Can long run restrictions identify technology shocks?', mimeo, Federal Reserve Board

[13] Erceg, C., Henderson, D. and Levin, A., (1999), 'Optimal monetary policy with staggered wage and price contracts', Journal of Monetary Economics 46: 281-313.

[14] Favero, C., (2001), *Applied Macroeconometrics*, Oxford: Oxford University Press.

[15] Gali, J. and Monacelli, T., (2002), 'Monetary policy and exchange rate volatility in a small open economy', NBER WP 8905.

[16] Gali, J. and Rabanal, P., (2004), 'Technology shocks and aggregate fluctuations: how well does the RBC model fit postwar US data?', NBER Working Paper 10636.

[17] Hamilton, J., (1994), *Time Series Analysis*, Princeton: Princeton University Press.

[18] King, R. and Rebelo, S., (1998) 'Resuscitating real business cycle', in Woodford, M. and Taylor, J., eds., Handbook of Macroeconomics, Amsterdam: North-Holland.

[19] Ingram, B., Kocherlakota, N. and Savin, N., (1994), 'Explaining business cycles: a multiple-shock approach', Journal of Monetary Economics 34: 415-28.

[20] Ireland, P., (2001a), 'Money's Role in the Monetary Business Cycle', NBER Working Paper 8115.

[21] -, (2001b), 'Sticky-Price Models of the Business Cycle: Specification and Stability', *Journal of Monetary Economics* 47(1): 3-18.

[22] Leeper, E., Sims, C. and Zha, T., (1996), 'What Does Monetary Policy Do?', *Brookings Papers on Economic Activity*, 0(2): 1-63.

[23] Linde', J., (2003), 'Taking models to the data', mimeo, Research Department, Sveriges Riksbank.

[24] Lutkepohl, H., (1993), *Intyroduction to Multiple Time Series Analysis*, Berlin: Springer-Verlag.

[25] Natalucci, F. and Ravenna, F., (2002), 'The road to adopting the Euro: monetary policy and exchange rate regimes for EU candidate countries', Board of Governors of the Federal Reserve, International Finance Discussion Paper 741.

[26] Pagan, A., (2003), 'An examination of some tools for macro-econometric model building', mimeo, Australian National University.

[27] Rabanal, P. and Rubio-Ramirez, J, (2003), "Comparing New Keynesian Models in the Euro Area: A Bayesian Approach ", FRB of Atlanta Working Paper No. 2003-30.

[28] Ravenna, F., (2002), 'The gains from credible inflation targeting', mimeo, University of California - Santa Cruz.

[29] Rudebusch, G., (2002), 'Term Structure Evidence on Interest Rate Smoothing and Monetary Policy Inertia', *Journal of Monetary Economics*, 49(6): 1161-87.

[30] Rudebusch, G. and Svensson, L., (1999), 'Policy Rules for Inflation Targeting', in Taylor, John, ed., *Monetary policy rules*, NBER Conference Report series, Chicago and London: University of Chicago Press.

[31] Ruge-Murcia, Francisco, (2002), 'Methods to Estimate Dynamic Stochastic General Equilibrium Models', University of California at San Diego, Economics Working Paper Series 2002.

[32] Sims, C., (1980), 'Macroeconomics and Reality', *Econometrica*, 48(1): 1-48.

[33] - , (1992), 'Interpreting the macroeconomic time-series facts: the effects of monetary policy', *European Economic Review* 36: 975-1011.

[34] Smets, F. and Wouters, R., (2002), 'An estimated Dynamic Stochastic General Equilibrium model of the Euro area', ECB Working Paper 171.

[35] Smith, G. and Zin, S., (1997), 'Real Business-Cycle realizations', *Carnegie-Rochester Conference Series on Public Policy* 27: 243-287.

[36] Uhlig, Harald, (1995), 'A toolkit for analyzing nonlinear dynamic stochastic models easily', Tilburg University, Center for Economic Research, Discussion Paper: 97.

[37] – , (2004), 'What are the effects of monetary policy on output? Results from an agnostic identification procedure', mimeo, Humboldt-Universität zu Berlin.

[38] Wallis, Kenneth, (1977), 'Multiple Time Series analysis and the final form of econometric models', *Econometrica* 45:1481-1492.

[39] Walsh, Carl E., (2003), *Monetary Theory and Policy*, 2nd. ed., Cambridge, MA: MIT Press.

[40] Woodford, M. and Rotemberg, J., (1998), 'An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy: Expanded Version', NBER Technical Working Paper 233.

[41] Zellner, A. and Palm, F., (1974), 'Time Series Analysis and Simultaneous Equation Econometric Models', *Journal of Econometrics* 2(1): 17-54

8 Appendix

Proof of Theorem 1: Let $y_t = H'^1 x_t + H'^2 z_t$ where H'^1, H'^2 include all and only the rows of $\widetilde{H}'^1, \widetilde{H}'^2$ corresponding to the observable variables in $\widetilde{y}_t$. Assume for the time being that $m \geq r$ (if the number of observable variables r is larger than the number of exogenous shocks m the variance-covariance matrix of y_t will be singular, preventing model estimation). To eliminate x_t from the model, rewrite the endogenous state equation as:

$$\begin{aligned}[I - F^{11}L]x_t &= F^{12}Lz_t \\ x_t &= [I - F^{11}L]^{-1}F^{12}Lz_t\end{aligned}$$

Substituting x_t in the control variables equation:

$$\begin{aligned}y_t &= H'^2 z_t + H'^1[I - F^{11}L]^{-1}F^{12}Lz_t \\ &= H'^2 z_t + H'^1 G(L)^{-1}F^{12}Lz_t\end{aligned} \tag{39}$$

where $G(L)^{-1}$ is an infinite-order lag polynomial. Assume that z_t is white noise. Then eq. (39) gives the VMA(∞) representation of the process. We can express the inverse of $G(L)$ in terms of its determinant $|G(L)|$, of order n in the lag operator L, and the cofactor matrix $D_G(L)$ of order $(n-1)$ in L: $G(L)^{-1} = D_G(L)/|G(L)|$. Therefore:

$$|G(L)|y_t = |G(L)|H'^2 z_t + H'^1 D_G(L)F^{12}Lz_t = G^*(L)z_t \tag{40}$$

Eq. (40) is a VARMA(n,n). The system (40) is written in final equations form: each component of the vector y_t depends only on its own lags.

We will assume in the following that $m = r$. If $G^*(L)$ is invertible, a VAR representation of infinite order for y_t is given by:

$$|G(L)|G^*(L)^{-1}y_t = z_t$$

Assume now that z_t is a vector AR(p) process: $z_t = Z(L)\varepsilon_t$. Eq. (40) then gives:

$$\begin{aligned}|G(L)|y_t &= G^*(L)Z(L)^{-1}\varepsilon_t \\ &= |G^*(L)|D_{G^*}(L)^{-1}Z(L)^{-1}\varepsilon_t\end{aligned}$$

where $G^*(L)^{-1} = D_{G^*}(L)/|G^*(L)|$, $|G^*(L)|$ is of order nm in L, $D_{G^*}(L)$ is of order $n(m-1)$ in L. Therefore y_t is described by:

$$|G(L)|Z(L)D_{G^*}(L)y_t = |G^*(L)|\varepsilon_t \tag{41}$$

which is a VARMA(mn+p,mn) process. A VAR(∞) representation is given by:

$$\frac{|G(L)|}{|G^*(L)|}Z(L)D_{G^*}(L)y_t = \varepsilon_t$$

■

Proof of Corollary 1: Starting from eq. (39), compute the inverse of $H'^1[I - F^{11}L]^{-1}F^{12}$. First, we need to assume that the number of observable variables r is equal to the number m of exogenous state variables. Otherwise, the matrix $H'^1[I - F^{11}L]^{-1}F^{12}$ will not be square. Then note that:

$$H'^1[I - F^{11}L]^{-1}F^{12} =$$

$$\begin{aligned} &= \lim_{i\to\infty} H'^1[A(L)]F^{12} \\ &= \lim_{i\to\infty} H'^1[I + F^{11}L + F^{11^2}L^2 + ... + F^{11^i}L^i]F^{12} \qquad (42) \\ &= \lim_{i\to\infty}[H'^1IF^{12} + H'^1F^{11}F^{12}L + H'^1F^{11^2}F^{12}L^2 + .. + H'^1F^{11^i}F^{12}L^i] \end{aligned}$$

Since H'^1IF^{12} is a square matrix of size $r \times m = r \times r$, we can write:

$$(H'^1IF^{12})^{-1}\lim_{i\to\infty}[H'^1IF^{12} + H'^1F^{11}F^{12}L + H'^1F^{11^2}F^{12}L^2 + .. + H'^1F^{11^i}F^{12}L^i] =$$
$$\lim_{i\to\infty}[I + (H'^1IF^{12})^{-1}H'^1F^{11}F^{12}L + (H'^1IF^{12})^{-1}H'^1F^{11^2}F^{12}L^2 + .. + (H'^1IF^{12})^{-1}H'^1F^{11^i}F^{12}L^i] =$$

if $(H'^1IF^{12})^{-1}$ exists. Note that if $n < r, m$ then $rank(H'^1IF^{12}) < r$ and the matrix H'^1IF^{12} is not invertible. Then the lag polynomial $H'^1[A(L)]F^{12}$ cannot be mapped into its inverse $B(L)$. Suppose that $n \geq r, m$. If a $B(L)$ of order one in L exists, then it must be true that:

$$[(H'^1IF^{12})^{-1}H'^1F^{11}F^{12}]^i = (H'^1IF^{12})^{-1}H'^1F^{11^i}F^{12}$$

If $n = r$, we have that:

$$\begin{aligned} &[(H'^1IF^{12})^{-1}H'^1F^{11}F^{12}]^i \\ &= [(F^{12})^{-1}(H'^1)^{-1}H'^1F^{11}F^{12}]^i \\ &= [(F^{12})^{-1}F^{11}F^{12}][(F^{12})^{-1}F^{11}F^{12}][(F^{12})^{-1}F^{11}F^{12}]..... \\ &= [(F^{12})^{-1}F^{11^i}F^{12}] \end{aligned}$$

It is straightforward to show that if $n > r$ this result will not obtain. Assume then $n = r$. The above result implies we can write:

$$\begin{aligned} &(H'^1IF^{12})^{-1}\lim_{i\to\infty}[H'^1IF^{12} + H'^1F^{11}F^{12}L + H'^1F^{11^2}F^{12}L^2 + .. + H'^1F^{11^i}F^{12}L^i] \\ &= \lim_{i\to\infty}[I + PL + P^2L^2 + .. + P^iL^i] \\ P &= (H'^1IF^{12})^{-1}H'^1F^{11}F^{12} = (F^{12})^{-1}F^{11}F^{12} \end{aligned}$$

and

$$
\begin{aligned}
& H'^1[I - F^{11}L]^{-1}F^{12} \\
= & \lim_{i\to\infty}[H'^1 I F^{12} + H'^1 F^{11} F^{12} L + H'^1 F^{11^2} F^{12} L^2 + .. + H'^1 F^{11^i} F^{12} L^i] \\
= & (H'^1 I F^{12})(H'^1 I F^{12})^{-1} \lim_{i\to\infty}[H'^1 I F^{12} + H'^1 F^{11} F^{12} L + H'^1 F^{11^2} F^{12} L^2 + .. + H'^1 F^{11^i} F^{12} L^i] \\
= & (H'^1 I F^{12}) \lim_{i\to\infty}[I + PL + P^2 L^2 + .. + P^i L^i] \\
= & (H'^1 I F^{12})[I - PL]^{-1}
\end{aligned}
$$

where $P = (H'^1 I F^{12})^{-1} H'^1 F^{11} F^{12}$. The VMA($\infty$) representation for y_t can then be mapped into:

$$
\begin{aligned}
[I - PL](H'^1 I F^{12})^{-1} y_t &= [I - PL](H'^1 I F^{12})^{-1} H'^2 z_t + L z_t \qquad (43) \\
y_t &= H'^1 F^{11} H'^{1^{-1}} y_{t-1} + H'^2 z_t + (H'^1 F^{12} - H'^1 F^{11} H'^{1^{-1}} H'^2) z_{t-1}
\end{aligned}
$$

which is a VARMA(1,1) if z_t is white noise[11]. If z_t is a vector AR(p) process we have:

$$
\begin{aligned}
[I - PL](H'^1 I F^{12})^{-1} y_t &= [I - PL](H'^1 I F^{12})^{-1} H'^2 z_t + L z_t \\
&= C(L) z_t = C(L) Z(L)^{-1} \varepsilon_t
\end{aligned}
$$

where $z_t = Z(L)^{-1}\varepsilon_t$ and $Z(L)$ is a lag polynomial of order p. Express the inverse of $C(L)$ in terms of its determinant $|C(L)|$, of order m in L, and the cofactor matrix $D_C(L)$ of order $m-1$ in L: $C(L)^{-1} = D_C(L)/|C(L)|$. Then:

$$
\begin{aligned}
[I - PL](H'^1 I F^{12})^{-1} y_t &= |C(L)| D_C(L)^{-1} Z(L)^{-1} \varepsilon_t \\
Z(L) D_C(L) [I - PL](H'^1 I F^{12})^{-1} y_t &= |C(L)| \varepsilon_t
\end{aligned}
$$

which is the VARMA (m+p,m) representation of y_t. A VARMA (m+p,m) representation also exists if $n = 1$. In this case, $[I - F^{11}L]^{-1}$ in eq. (39) is a scalar. Each component of y_t will depend on its own lags only and eq. (43) becomes:

$$
\begin{aligned}
(1 - F^{11}L) y_t &= H'^2 z_t + (H'^1 F^{12} - F^{11} H'^2) z_{t-1} \\
y_t &= F^{11} y_{t-1} + H'^2 z_t + (H'^1 F^{12} - F^{11} H'^2) z_{t-1}
\end{aligned}
$$

■

[11] Note that if $n = m = r$ the matrix H'^1 is square, and by simply inverting $H'^1[I - F^{11}L]$ in eq. (39) we obtain eq.(43).

Proof of Theorem 2: Start the Kalman filter recursion by computing the estimate of ξ_1:

$$\xi_{1|1} = \xi_{1|0} + P_{1|0}H(H'P_{1|0}H + R)^{-1}(y_1 - H'\xi_{1|0}) \tag{44}$$

Because of assumption 1:

$$\begin{aligned}
\xi_{1|0} &= E(\xi_1) = \begin{bmatrix} E(\xi^1) \\ E(\xi^2) \end{bmatrix} = \begin{bmatrix} F^{11}\xi^1_{0|0} + F^{12}\xi^2_{0|0} \\ E(\xi^2_1) \end{bmatrix} = \begin{bmatrix} F^{11}\xi^1_0 + F^{12}\xi^2_0 \\ F^{22}\xi^2_0 + E(v_1) \end{bmatrix} \\
P_{1|0} &= E\left(\begin{bmatrix} 0 \\ v_1 \end{bmatrix} \begin{bmatrix} 0 & v_1 \end{bmatrix} \right) = \begin{bmatrix} {}^{n\times n}P^{11}_{1|0} & {}^{n\times m}P^{12}_{1|0} \\ {}^{m\times n}P^{21}_{1|0} & {}^{m\times m}P^{22}_{1|0} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & P^{22}_{1|0} \end{bmatrix}
\end{aligned}$$

$${}^{m\times m}P^{22}_{1|0} = \begin{bmatrix} \sigma^2_{\varepsilon_1} & 0 & ... & 0 \\ 0 & \sigma^2_{\varepsilon_2} & ... & 0 \\ .... & ... & ... & ... \\ 0 & 0 & ... & \sigma^2_{\varepsilon_m} \end{bmatrix}$$

Using the above equations:

$$\begin{aligned}
{}^{m\times n+m}\left(H'P_{1|0}\right) &= \begin{bmatrix} H'^1 & H'^2 \end{bmatrix} P_{1|0} \\
&= \begin{bmatrix} 0 & H'^2 P^{22}_{1|0} \end{bmatrix}
\end{aligned}$$

Assumption 2 implies:

$${}^{m\times m}(H'P_{1|0}H + R) = \left(H'P_{1|0}H\right) = [H'^2 P^{22}_{1|0} H^2]$$

Because of assumption 3 the matrices H'^2 and H^2 are invertible. Since $P^{22}_{1|0}$ is diagonal:

$$(H'P_{1|0}H + R)^{-1} = (H^2)^{-1}(P^{22}_{1|0})^{-1}(H'^2)^{-1} \tag{45}$$

Since:

$${}^{n+m\times m}(P_{1|0}H) = \begin{bmatrix} 0 \\ P^{22}_{1|0}H^2 \end{bmatrix}$$

the following holds:

$$\begin{aligned}
P_{1|0}H(H'P_{1|0}H + R)^{-1} &= \begin{bmatrix} 0 \\ P^{22}_{1|0}H^2(H^2)^{-1}(P^{22}_{1|0})^{-1}(H'^2)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} {}^{n\times m}0 \\ {}^{m\times m}(H'^2)^{-1} \end{bmatrix}
\end{aligned}$$

Therefore:

$$P_{1|0}H(H'P_{1|0}H + R)^{-1}(y_1 - H'\xi_{1|0}) = \begin{bmatrix} {}^{n\times 1}0 \\ (H'^2)^{-1}(y_1 - H'\xi_{1|0}) \end{bmatrix} \tag{46}$$

Equation (46) together with eq. (44) implies:

$$\xi^1_{1|1} = \xi^1_{1|0} + [0] = F^{11}\xi^1_0 + F^{12}\xi^2_0 \tag{47}$$

which is equal to the result obtained in eq. (19). Equations (46) and (44) also imply:

$$\begin{aligned}
\xi^2_{1|1} &= F^{22}\xi^2_0 + (H'^2)^{-1}y_1 - (H'^2)^{-1}H'\xi_{1|0} \\
&= F^{22}\xi^2_0 + (H'^2)^{-1}y_1 - \left[\ (H'^2)^{-1}H'^1 \quad I\ \right]\xi_{1|0} \\
&= F^{22}\xi^2_0 + (H'^2)^{-1}y_1 - (H'^2)^{-1}H'^1\xi^1_{1|0} - \xi^2_{1|0} \\
&= F^{22}\xi^2_0 + (H'^2)^{-1}y_1 - (H'^2)^{-1}H'^1\xi^1_1 - F^{22}\xi^2_0 \\
&= (H'^2)^{-1}\left[y_1 - H'^1\xi^1_1\right]
\end{aligned} \tag{48}$$

which is equal to the result obtained in eq. (20).

The MSE of the forecast in eqs. (47) and (48) is:

$$P_{1|1} = P_{1|0} - P_{1|0}H(H'P_{1|0}H + R)^{-1}H'P_{1|0}$$

Using assumption 2:

$$\begin{aligned}
P_{1|1} &= P_{1|0} - \begin{bmatrix} 0 \\ (H'^2)^{-1} \end{bmatrix} \begin{bmatrix} 0 & H'^2 P^{22}_{1|0} \end{bmatrix} \\
&= P_{1|0} - \begin{bmatrix} 0 & 0 \\ 0 & (H'^2)^{-1}H'^2P^{22}_{1|0} \end{bmatrix} \\
&= P_{1|0} - P_{1|0} \\
&= 0
\end{aligned} \tag{49}$$

Equation (49) together with eqs. (47) and (48) implies:

$$\begin{aligned}
\xi^1_{1|1} &= \xi^1_1 \\
\xi^2_{1|1} &= \xi^2_1
\end{aligned}$$

The period 2 recursion is identical to the period 1 recursion since ξ_1 is known with certainty:

$$\xi_{2|1} = F\xi_{1|1} = F\xi_1 = \begin{bmatrix} F^{11}\xi^1_1 + F^{12}\xi^2_1 \\ F^{22}\xi^2_1 \end{bmatrix}$$

and

$$P_{2|1} = FP_{1|1}F' + Q = Q = P_{1|0}$$

In all subsequent periods, $\xi_{t|t-1} = F\xi_t$, $P_{t|t-1} = P_{1|0}$ and $P_{t|t} = 0$. Therefore the results in eqs. (47) and (48) hold in any period t:

$$\begin{aligned}
\xi^1_{t|t-1} &= F^{11}\xi^1_{t-1|t-2} + F^{12}\xi^2_{t-1|t-2} = \xi^1_t \\
\xi^2_{t|t-1} &= (H'^2)^{-1}[y_t - H'^1\xi^1_{t|t-1}] = \xi^2_t
\end{aligned}$$

It is straightforward to show that the Kalman smoothed estimate is still equal to the above equations. In fact, given the Kalman filtered estimates have zero variance, there is no reason to expect that conditioning the estimate on the whole sample will provide a reduction in the estimates' variance. Equation (49) implies:

$$J_1 = P_{1|1}F'P^{-1}_{2|1} = 0$$

Since $P_{t|t} = 0 \; \forall \; t$:

$$J_t = 0 \; \forall \; t$$

The Kalman smoothed estimate is:

$$\xi_{t|T} = \xi_{t|t} + J_t(\xi_{t+1|T} - \xi_{t+1|t}) = \xi_{t|t} = \xi_t$$

Therefore:

$$\begin{aligned}
\xi^1_{t|T} &= F^{11}\xi^1_{t-1|T} + F^{12}\xi^2_{t-1|T} = \xi^1_t \\
\xi^2_{t|T} &= (H'^2)^{-1}[y_t - H'^1\xi^1_{t|T}] = \xi^2_t
\end{aligned}$$

■

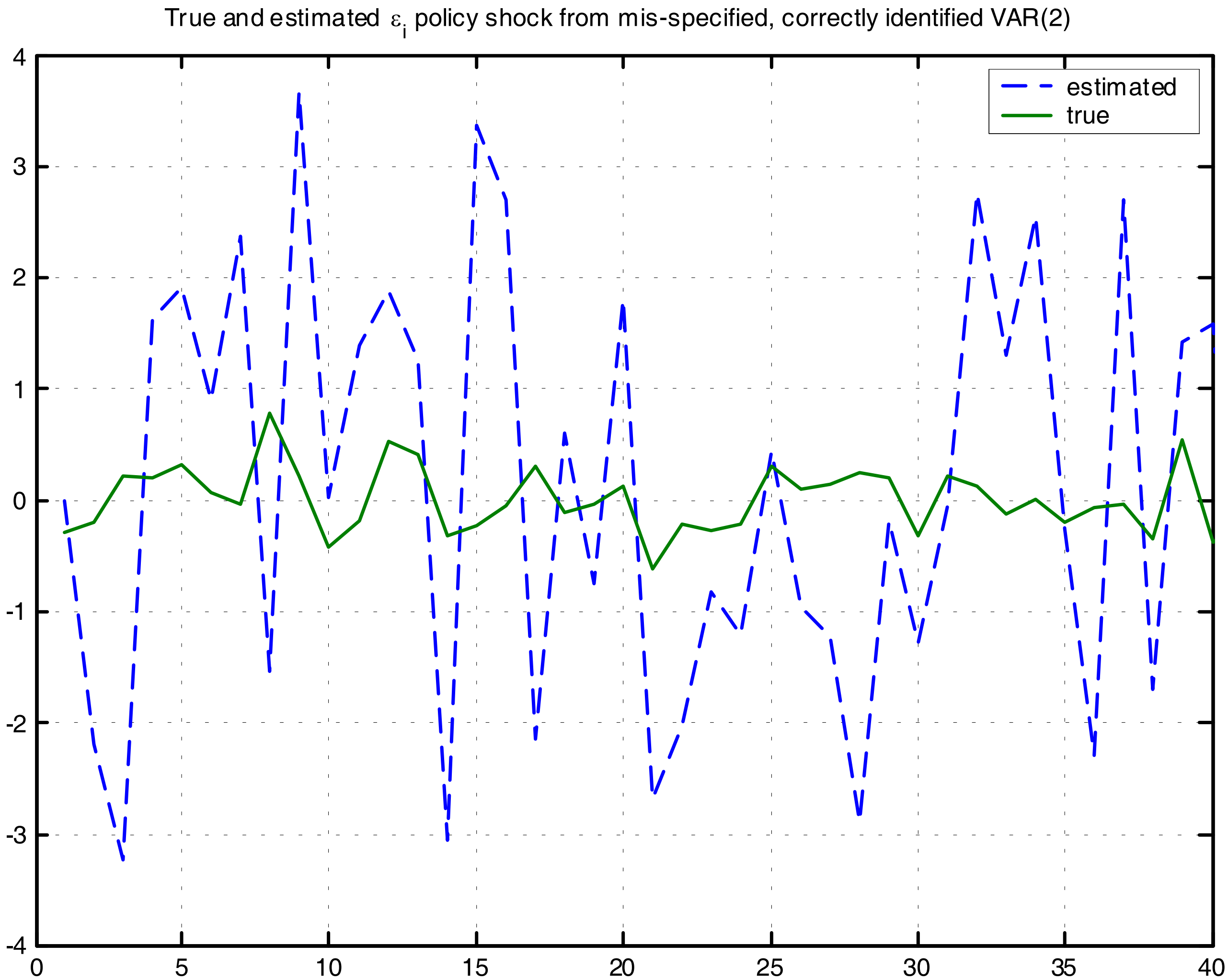

Figure 1: 10 year sample path of VAR(2)-estimated series of policy shocks ε_v and true series. VAR(2) is estimated over 250,000 40-quarters simulated series. Scaling is in percentage points.

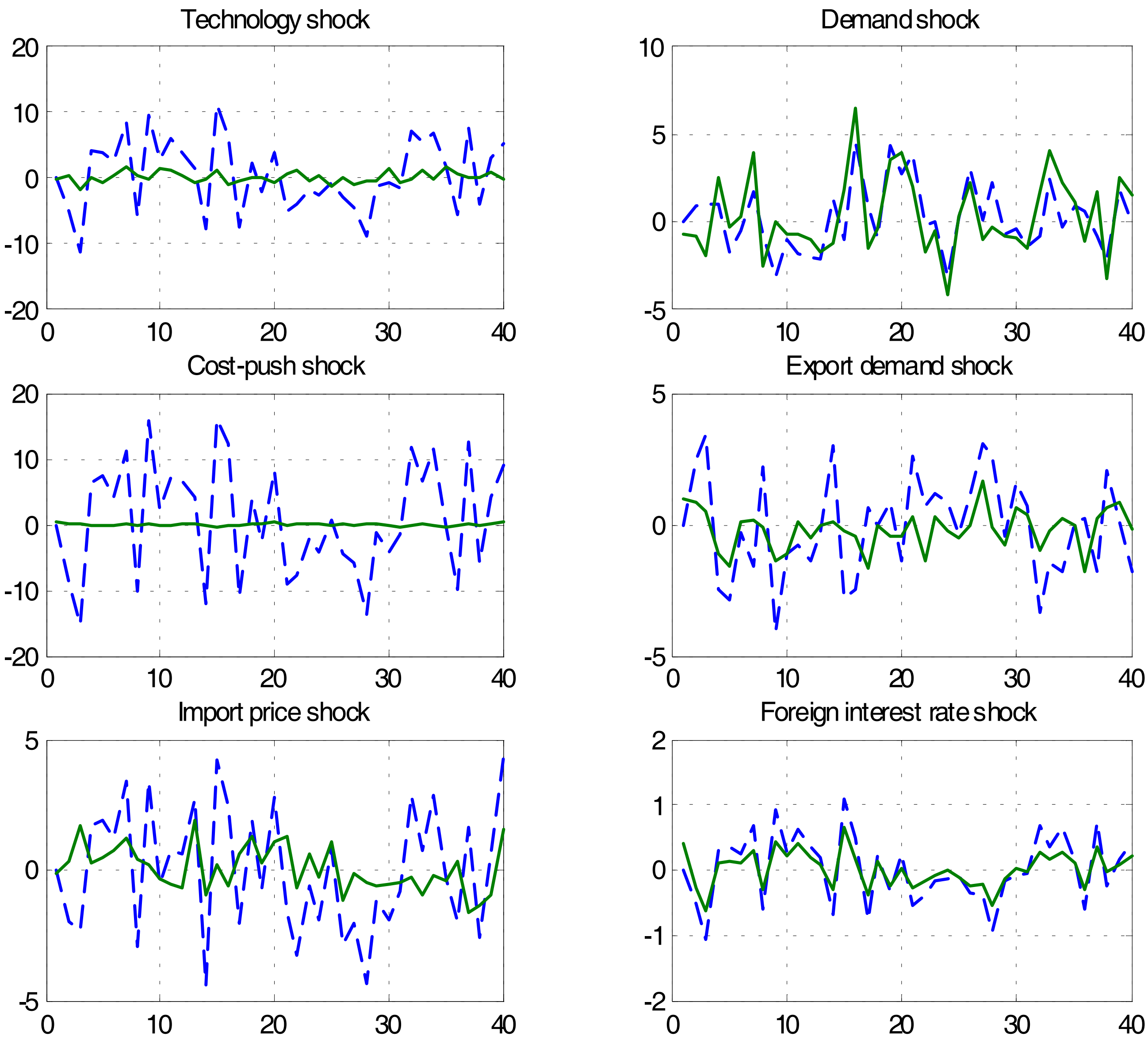

Figure 2: 10 year sample path of VAR(2)-estimated series of the shock vector ε. VAR(2) is estimated over 250,000 40-quarters simulated series. Scaling is in percentage points.

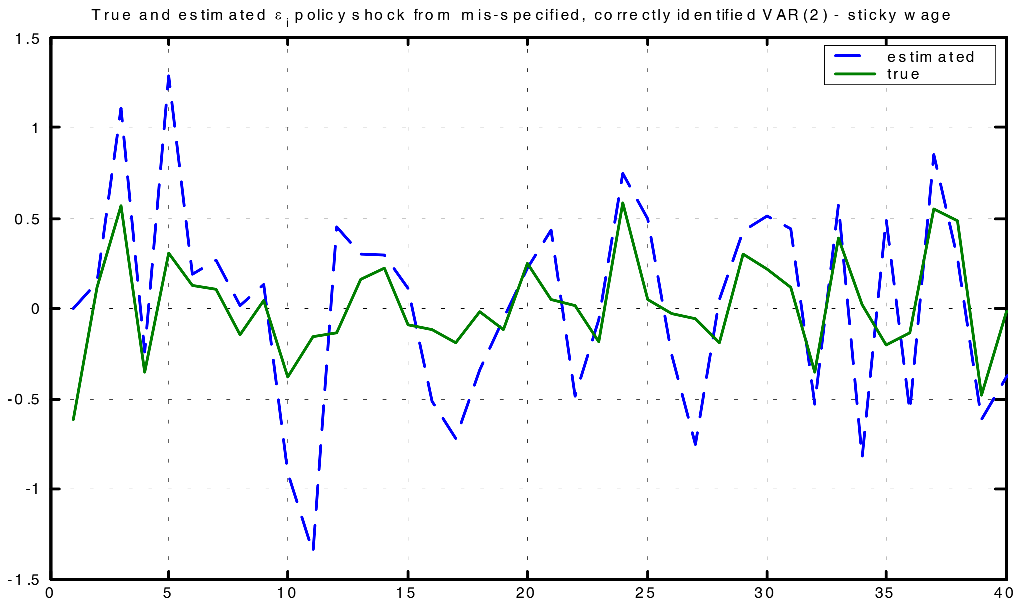

Figure 3: 10 year sample path of VAR(2)-estimated series of the policy shock ε_v. VAR(2) is estimated over 250,000 40-quarters simulated series. Scaling is in percentage points. Policy is as in the baseline model.

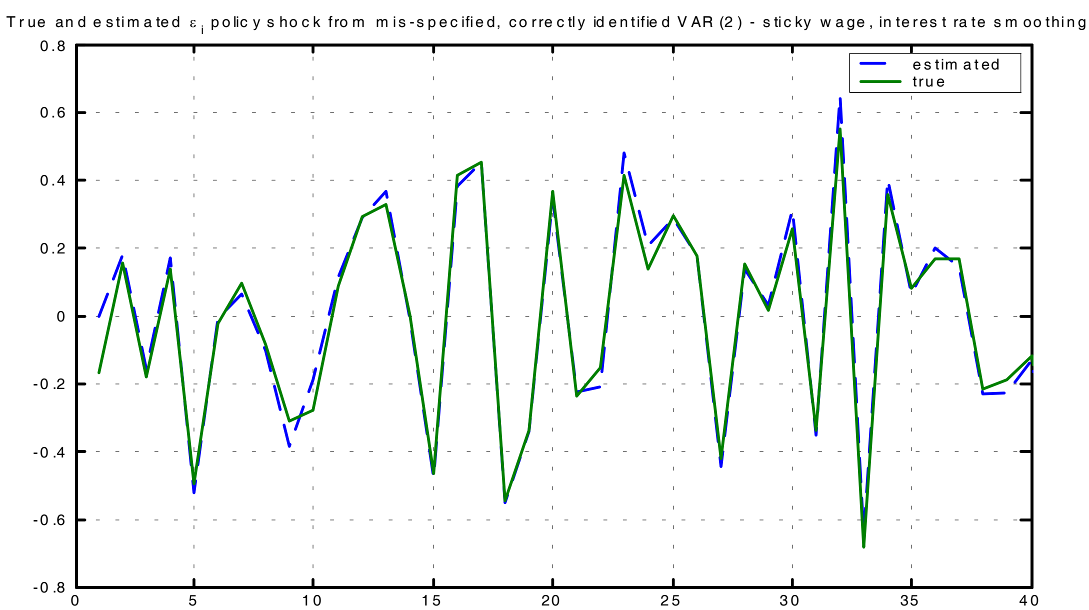

Figure 4: 10 year sample path of VAR(2)-estimated series of the policy shock ε_v. Policy parameters: $\chi = 0.9$, $\rho_v = 0$

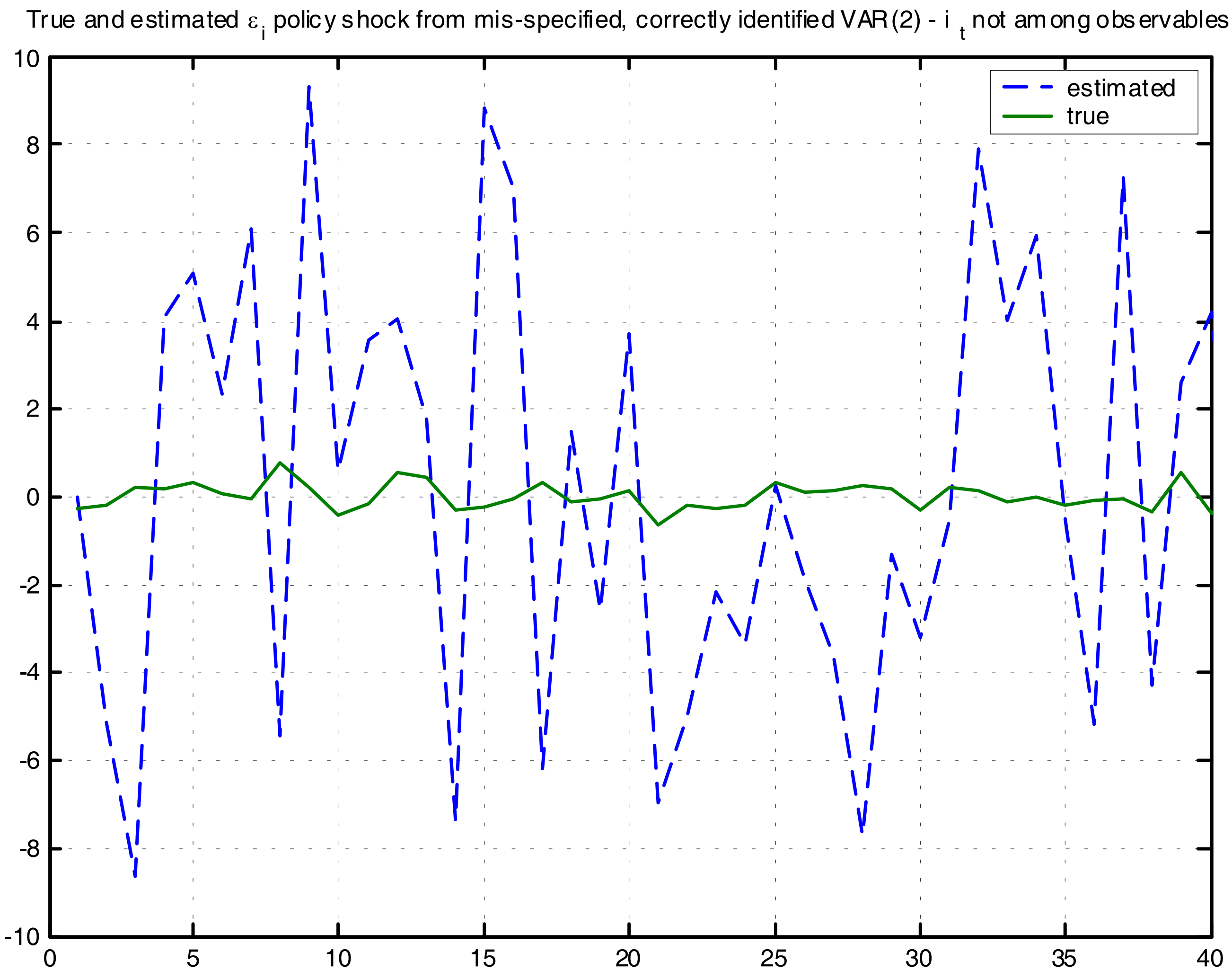

Figure 5: VAR(2)-estimated series of the shock ε_v. VAR(2) is estimated over 250,000 40-quarters simulated series. Scaling is in percentage points. VAR(2) data series includes $n_t, y_t, \pi_t, c_t^*, \Delta e_t, mc_t, \pi_{H,t}$

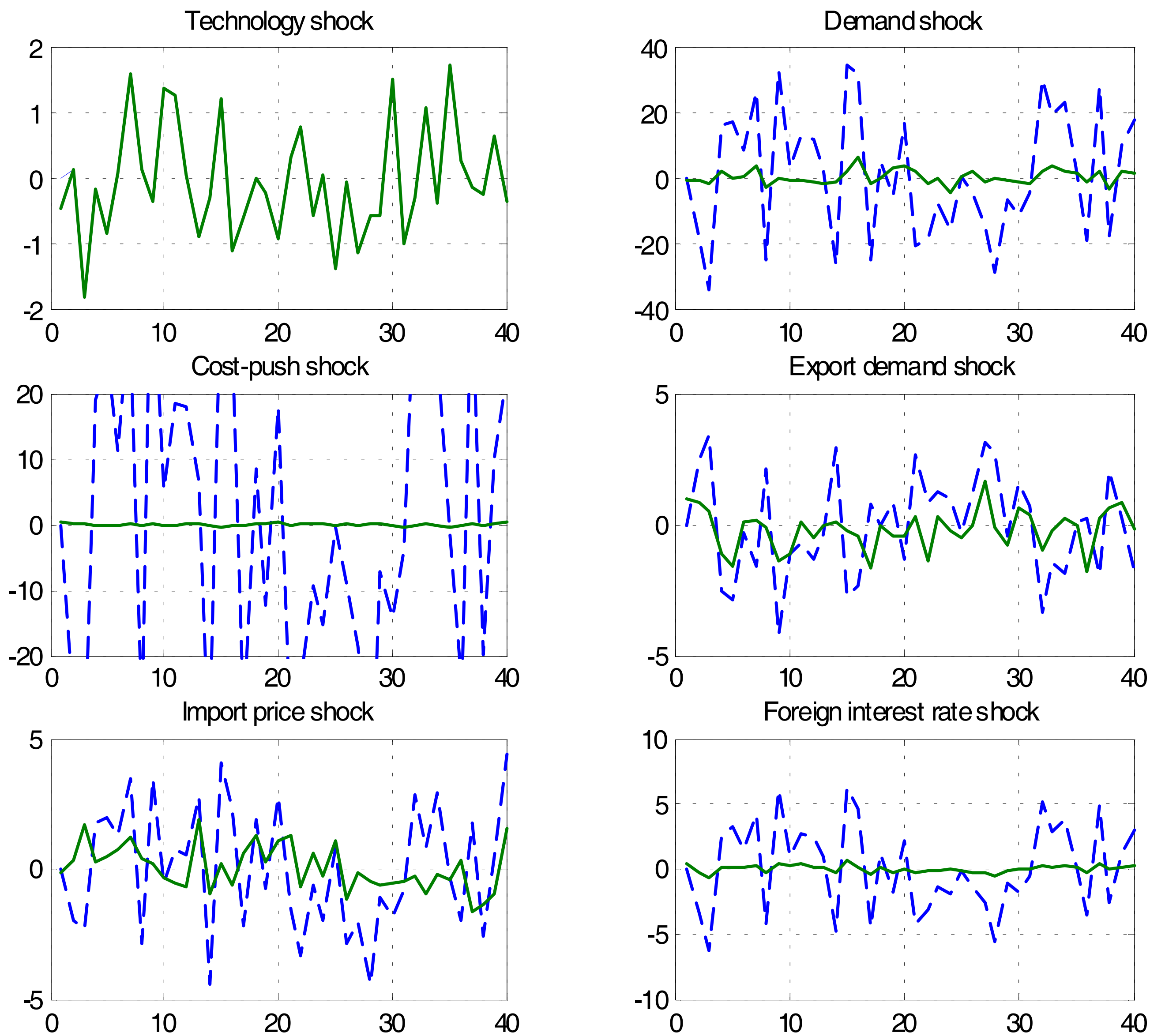

Figure 6: VAR(2)-estimated series of the shock vector ε. VAR(2) is estimated over 250,000 40-quarters simulated series. Scaling is in percentage points. VAR(2) data series includes $n_t, y_t, \pi_t, c_t^*, \Delta e_t, mc_t, \pi_{H,t}$