Vector Autoregressions and Reduced Form Representations of DSGE models

Federico Ravenna[a*]

[a] Department of Economics, University of California, Santa Cruz, CA 95064

May 2006

Abstract

The performance of Dynamic Stochastic General Equilibrium models is often tested against estimated VARs. This requires that the data-generating process consistent with the DSGE theoretical model has a finite-order VAR representation. This paper discusses the assumptions needed for a finite-order VAR(p) representation of a DSGE model to exist. When a VAR(p) is only an approximation to the true VAR, the truncated VAR(p) may return largely incorrect estimates of the impulse response function. The results do not hinge on small sample bias or on incorrect identification assumptions. But the bias introduced by truncation can lead to bias in the identification of the structural shocks. Identification strategies that are equivalent in the true VAR representation perform differently in the approximating VAR.

JEL classification: C13; C22; E32

---

*Email: fravenna@ucsc.edu. The author thanks Richard Dennis, Walter Enders, Marco del Negro, Bart Hobijn, Oscar Jorda, David Lopez-Salido, Claudio Michelacci, Yin Wong Cheung and an anonymous referee for helpful comments and suggestions, and Juan Rubio-Ramirez for supplying the Matlab code to compute the finite order VAR representation of a state-space model. Financial suport from the UCSC Social Science Division and the UCSC Committee on Research is gratefully acknowledged.

1. Introduction

An important goal of real and monetary business cycle theoretical research is to explain the empirical evidence on the impact of economic shocks on macroeconomic variables. A vast literature is devoted to building Dynamic Stochastic General Equilibrium (DSGE) models able to explain the impact of a monetary policy shock on output and inflation, or the impact of a technology shock on labor hours. The empirical evidence is often obtained from estimating structural Vector Autoregressions' (VAR). In part of the literature the structural parameters of a DSGE model are estimated by minimizing the distance between the model's and the estimated VAR impulse response functions.

A growing number of papers has questioned the ability of estimated VARs to provide reliable guidance to building DSGE models consistent with the data[1]. First, a DSGE model implies restrictions in the mapping between economic shocks and observable variables. In linear models (or in linear approximations) these restrictions are summarised by the Vector Moving Average (VMA) representation. If the VMA representation is not invertible a DSGE model does not admit a VAR representation mapping economic shocks to a vector of observable variables and its lags. Fernandez-Villaverde, Rubio-Ramirez and Sargent (2005) discuss the invertibility problem and provide examples of well-specified DSGE models that lack a VAR representation. Second, even if it exists, the VAR representation of a DSGE model may require an infinite number of lags. Yet macroeconomists work with small data samples and are constrained to estimating truncated VARs which only approximate the true VAR representation. Third, the restrictions used to identify structural shocks from the VAR reduced form innovations may be inconsistent with the DSGE model assumptions, leading to a mis-identification problem.

[1] See Canova and Pina (2005), Chari, Kehoe and McGrattan (2005), Christiano, Eichenbaum and Vigfusson (2006), Cochrane (1998) and Gali and Rabanal (2005). An alternative to the VAR approach is estimation of the state-space form of a DSGE model, as in Rabanal and Rubio-Ramirez (2005).

This paper studies the performance of truncated VAR approximations to the true infinite-order VAR representation of a DSGE model[2]. The model VARMA representation is obtained starting from the equilibrium law of motion, and conditions for a finite-order VAR(p) representation to exist are derived. In the case a VAR(p) representation does not exist, the paper discusses the VAR truncation problem using as a laboratory the finite order VAR(p) approximation to a real business cycle model. Truncation affects the impulse response function through two separate channels: the VAR(p) erroneously constrains to zero some coefficients in the true VAR representation, and the VAR(p) coefficients can lead to mistaken identification of the structural shocks. Depending on the model parametrization, truncation can lead to large errors through one or both channels. In effect, the *truncation bias* can cause an *identification bias* even if the identification strategy is consistent with the theoretical model. Regardless of small sample bias, identification schemes that are equally appropriate in a VAR(∞) perform differently in a truncated VAR.

These findings are related to some recent contributions in the literature. We generalize some results in Chari, Kehoe and McGrattan (2005), who examine a stylized business cycle model, and show that for a standard parametrization the coefficients in the VAR representation converge to zero extremely slowly, making a finite order VAR approximation unsuitable. They find that the impulse response of labor hours to a technology shock identified using long run restrictions in a finite order VAR is a poor approximation to the true magnitude. While we obtain a similar result in a closely related RBC model, we show that the largest part of the approximating error comes from the identification bias. This result is consistent with Christiano, Eichenbaum and Vigfusson (2006), who conclude that when identification is achieved using short run restrictions finite order VARs can achieve a remarkably close approximation to the DSGE model in small sample. Yet, we also find that for some (reason-

[2] The truncation problem has been acknowledged in the literature (as early as in work by Wallis, 1977) but largely neglected in applied work. See Chari, Kehoe and McGrattan (2005), Cooley and Dwyer (1998), Erceg, Guerrieri and Gust (2004), Faust and Leeper (1997) for discussion within specific models.

able) parametrizations of the model, even using the correct theoretical identification matrix and shutting down the identification bias the finite order VAR provides a largely incorrect impulse response function. Erceg, Guerrieri and Gust (2005) study the performance of truncated VAR representations of an RBC model, and conclude that the approximating error stems from the small-sample error impact on the long run identification scheme. In contrast, we show that small sample error is not essential to generate identification bias, and propose a method to measure identification and truncation bias in population.

The paper is organized as follows. Section 2 discusses VAR representations of DSGE models and provides conditions for the VAR representation to be of finite order. Section 3 discusses the performance of truncated VAR and illustrates the impact of truncation and identification bias in an RBC model. Section 4 concludes.

2. VAR representations of DSGE models

A linearized DSGE model can be written as a system of stochastic difference equations. The solution to the system is the recursive equilibrium law of motion:

$$y_t = Px_{t-1} + Qz_t \tag{1}$$

$$x_t = Rx_{t-1} + Sz_t \tag{2}$$

$$Z(L)z_t = \varepsilon_t \tag{3}$$

where x_t is an $n \times 1$ vector of endogenous state variables, z_t is an $m \times 1$ vector of exogenous state variables, y_t is an $r \times 1$ vector of endogenous variables, ε_t is a vector stochastic process of dimension $m \times 1$ such that $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_t') = \Sigma$, $E(\varepsilon_t \varepsilon_\tau') = 0$ for $\tau \neq t$ and Σ is a diagonal matrix. $Z(L)$ is the matrix polynomial $[I - Z_1 L ... - Z_p L^p]$ in the lag operator L defining a stationary vector $AR(p)$ stochastic process. King, Plosser and Rebelo (1988) discuss how to obtain the system in eqs. (1) to (3) as the log-linear approximation to the solution of a DSGE model. The equilibrium law of motion of models with linear transition laws and quadratic objective functions takes the same functional form.

The polynomial $Z(L)$ is typically assumed to be of the first order. For $Z(L) = [I - Z_1 L]$ an alternative way of writing eqs. (1) to (3) is to define the vector $\widetilde{x}_t = [x_{t-1}\ z_t]'$ so that:

$$y_t = \widetilde{P}\widetilde{x}_t + Q\varepsilon_t \tag{4}$$

$$\widetilde{x}_{t+1} = \widetilde{R}\widetilde{x}_t + \widetilde{S}\varepsilon_t \tag{5}$$

$$\widetilde{R} = \begin{bmatrix} R & SZ_1 \\ 0 & Z_1 \end{bmatrix} ;\ \widetilde{S} = \begin{bmatrix} S \\ I \end{bmatrix} ;\ \widetilde{P} = \begin{bmatrix} P & QZ_1 \end{bmatrix} \tag{6}$$

This is the approach followed, for example, in Fernandez-Villaverde, Rubio-Ramirez and Sargent (2005). The results in the paper can be obtained using either of the two equilibrium specifications. The specification in eqs. (1) to (3) offers two advantages. First, the endogenous and exogenous state vectors play a different role in the finite-order VAR representation of the system, and have a different economic interpretation. An economic model is built to explain the dynamics of both y_t and x_t - which typically correspond to observable economic magnitudes. The dynamics of the vector z_t is left unexplained by the model. Second, it will be useful to highlight the role of the matrix $Z(L)$ in subsequent results.

When $Z(L) = [I - Z_1 L]$ the DSGE model can be written as:

$$Y_t = AY_{t-1} + Bz_t \tag{7}$$

$$z_t = Z_1 z_{t-1} + \varepsilon_t$$

$$Y_t = \begin{bmatrix} x_t \\ y_t \end{bmatrix} ;\ A = \begin{bmatrix} R & 0 \\ P & 0 \end{bmatrix} ;\ B = \begin{bmatrix} S \\ Q \end{bmatrix}$$

where the vector $Y_t' = [x_t\ ,\ y_t]$ has dimension $1 \times n + r$. Assume all the components of the vectors x_t and y_t are observable, and the vector z_t has dimension $m = n + r$. Since the number $n+r$ of observable variables is equal to the number of shocks, if B^{-1} exists $z_t = Z_1[B^{-1}Y_{t-1} - B^{-1}AY_{t-2}] + \varepsilon_t$. A restricted VAR(2) representation of the system (7) is[3]:

$$\begin{aligned} Y_t &= (A + BZ_1B^{-1})Y_{t-1} - (BZ_1B^{-1}A)Y_{t-2} + B\varepsilon_t \\ &= \Gamma_1 Y_{t-1} + \Gamma_2 Y_{t-2} + \eta_t \end{aligned} \tag{8}$$

where the VAR innovations $\eta_t = B\varepsilon_t$ are a rotation of the structural shocks vector ε_t.

When $m > n+r$ a VAR representation of the DSGE model may exist. But it will not be possible to map η_t into a higher-dimension vector of orthogonal shocks ε_t. If instead $m < n+r$, as is often the case in DSGE models, the system is singular, preventing likelihood estimation of the VAR. To obtain a non-singular VAR representation of the model (7) some of the observable variables must be dropped from the system so as to satisfy the requirement $n+r=m$. Omitting $r-r_1$ rows of the y_t vector does not affect the VAR(2) representation of any other observable variable. Regardless of which r_1 rows of y_t are included, the VAR(2) representation of any subset $\widehat{Y}_t$ of the vector Y_t is consistent with the DSGE model[4].

When a subset of the components in the x_t vector is unobservable a VAR representation for $\widehat{Y}_t$ cannot be obtained by eliminating rows from the matrices A, B and some of the empty columns of A. Does a finite order VAR representation of the DSGE model still exist? If $n > m$ and $(n-m)$ components of x_t are omitted from the system, the remaining m variables still have a VAR(2) representation, since the omitted variables can be rewritten as a linear combination of lags of the variables included in the VAR. If $n < m$ excluding components of the x_t vector from the list of observable variables implies that a finite order VAR representation for $\widehat{Y}_t$ exists only under the condition stated in the following proposition.

[3] Using the system defined in eqs. (4) and (5) would give a VAR(1) representation.

[4] It is assumed that the VAR representation includes at least m observable variables. Lutkepohl (1993) shows that when the true model is described by the non-singular VAR (8) the data generating process for the observable $gx1$ vector $\widehat{Y}$ where $g < m$ is a VARMA(p,q) with $p \leq 2(n+r)$, $q \leq 2(n+r) - 2$.

Proposition 2.1 *Let the system in eqs. (1), (2), (3) describe the law of motion of the vectors* z_t, x_t, y_t *where* y_t *is a vector of dimension* $r \times 1$, x_t *is a vector of dimension* $n \times 1$ *and* z_t *is a vector of dimension* $m \times 1$. *Assume* $m = r$. *If the vector* $\widehat{Y}_t$ *includes all and only the components of* y_t*:*

1. *the vector* $\widehat{Y}_t$ *has a VARMA*$(n + pm, n + p(m-1))$ *representation;*

2. *a finite order VAR representation for* $\widehat{Y}_t$ *exists if and only if the determinant of* $[|G(L)| + PD_G(L)SQ^{-1}L]$ *is of degree zero in* L, *where* $G(L) = [I - RL]$ *and* $D_G(L)$ *is the adjoint matrix of* $G(L)$

Corollary 2.2 *The necessary and sufficient condition for existence of a finite order VAR representation can also be stated as the requirement that the determinant of* $[I - (R - SQ^{-1}P)L]$ *be of degree zero in* L.

Proposition 2.3 *The results in Proposition 2.1 also obtain in the case the vector* $\widehat{Y}_t$ *includes a subset* $n_1 < n$ *of the vector* x_t *components and a subset* $(r - n_1)$ *of the vector* y_t *components.*

Proof of the results is in the Appendix. Proposition 2.1 through 2.3 provide a guide for the researcher trying to estimate a finite-order non-singular VAR *consistent* with a given DSGE model data-generating process. The VAR estimation assumes either of the two conditions:

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

(a'). The determinant of $[|G(L)| + PD_G(L)SQ^{-1}L]$ is of degree zero in L.

Observability of z_t is irrelevant for a finite order VAR representation to exist. If the sufficient condition (a) is not met the vector $\widehat{Y}_t$ has a finite order VARMA representation. Under certain conditions (Fernandez-Villaverde et al., 2005), the MA component is invertible, and a VAR representation exists. (a') is the necessary and sufficient condition for the VAR representation to be of finite order. The polynomial $Z(L)$ does not enter condition (a').

3. Finite order approximations to the true VAR process

When it exists, the VAR representation for y_t can be written as:

$$\begin{aligned} y_t &= QZ_1Q^{-1}y_{t-1} + ... + QZ_pQ^{-1}y_{t-p} + \\ & -[QZ_1Q^{-1}PL + ... + QZ_pQ^{-1}PL^p - P]\sum_{j=0}^{\infty}(R - SQ^{-1}P)^j L^{j+1}SQ^{-1}y_t + Q\varepsilon_t \end{aligned} \tag{9}$$

Eq. (9) is derived from eq. (27) in the Appendix. When conditions (a) or (a') are not met, a finite order VAR may still be a very good approximation to the true data generating process (9) if the VAR matrix coefficients for longer lags of y_t are close to zero. This requires either the coefficients in the matrix $[QZ_1Q^{-1}PL + ... + QZ_pQ^{-1}PL^p - P]$ to be close to zero, or the matrix $(R - SQ^{-1}P)^j$ to converge to zero fast enough. Asymptotically, the speed at which the VAR matrices converge to zero depends on the largest eigenvalue of $(R - SQ^{-1}P)$. The polynomial $Z(L)$ does not appear in the matrix relevant for the convergence speed.

Since the sequence $(R - SQ^{-1}P)^j$ for $j = 0, 1...$ converges to zero, a finite order VAR(p) that well approximates the true VAR process always exists for some sufficiently large value of p. The problem facing economists is whether the number of lags p to be included is reasonable given the length of economic time series over which VARs are estimated. When estimating a VAR consistent with DSGE business cycle models it is standard to assume that including few lags is sufficient to provide a reasonable approximation to the true VAR. This assumption can be misleading. Truncation affects the approximating VAR performance through two separate channels. First, the truncated VAR coefficients are biased: a VAR(p) does not describe the true dynamics of the DSGE model, since all coefficients for lags larger than p are restricted to be equal to zero. Second, if the VAR coefficients enter in the computation of the matrix identifying structural shocks from reduced form innovations, truncation results in an identification error. Depending on the model none, one or both of these channels - the *truncation bias* and the *identification bias* - can prejudice the accuracy of the approximating VAR(p). The identification bias does not originate in mistaken identification assumptions. Finally, truncation and identification bias need not depend on small sample bias of the estimator. To illustrate the impact of truncation and identification bias we compute from the approximating finite order VAR(p) representation of an RBC model the impulse response function to an identified technology shock, and the structural shocks vector ε_t. We examine how these magnitudes approximate the true ones. Because the VAR(p) coefficients are population values, any approximation error does not depend on the variance of the estimator.

3.1. A Real Business Cycle Model Example

Consider Hansen's (1985) indivisible labor model with two exogenous shocks: a non-stationary technology shock, and a stationary labor supply shock. The planner's choice for consumption C_t, capital K_t, labor N_t, and output Y_t maximizes the utility function:

$$E_t \sum_{t=1}^{\infty} \beta^t [\ln C_t + AD_t(1 - N_t)] \tag{10}$$

subject to the capital accumulation and production function constraints:

$$\begin{aligned} K_t &= Y_t - C_t + (1-\delta)K_{t-1} && (11) \\ Y_t &= K_{t-1}^{\rho}(Z_t N_t)^{1-\rho} && (12) \end{aligned}$$

The labor-augmenting technology level Z_t and the labor supply shifter D_t follow exogenous stochastic processes:

$$\begin{aligned} \ln Z_t &= \ln Z_{t-1} + \mu_z + \varepsilon_{z_t} && (13) \\ \ln D_t &= (1-\rho_d)\ln \overline{D} + \rho_d \ln D_{t-1} + \varepsilon_{d_t} && (14) \\ \varepsilon_{i_t} &\sim i.i.d.\ N(0, \sigma_i^2) \ \ i = z, d \end{aligned}$$

The first order conditions for the planner's problem are:

$$AD_t = C_t^{-1}(1-\rho)\frac{Y_t}{N_t} \tag{15}$$

$$1 = \beta E_t\left[\frac{C_t}{C_{t+1}}R_{t+1}\right] \tag{16}$$

$$R_t = \rho\frac{Y_t}{K_{t-1}} + (1-\delta) \tag{17}$$

where R_t is the gross real interest rate. Equations (11) to (17) describe the equilibrium of the economy. A technology innovation ε_{z_t} has a permanent impact on the level of Z_t, $Y_t,\ K_t,\ C_t$ but only a transitory impact on the growth rate of these variables. The assumption of a utility function logarithmic in C_t and separable in C_t and N_t implies the steady state level of N_t is independent of the level of technology.

The model defined in terms of $N_t,\ R_t,\ D_t,\ \widehat{K}_t = K_t/Z_t,\ \widehat{Y}_t = Y_t/Z_t,\ \widehat{C}_t = C_t/Z_t,\ \widehat{Z}_t = Z_t/Z_{t-1}$ is stationary, and an approximate solution can be obtained by log-linearizing the equilibrium conditions around the steady state. This yields a linear model cast in the form of eqs. (1) to (3). The model is parametrized following the RBC literature (see Erceg, Guerrieri and Gust, 2005). The capital share ρ is set to 0.35. The quarterly depreciation rate for installed capital δ is assumed equal to 2%. The discount rate β is chosen so that in the steady state the annual real interest rate is equal to 3%. The steady state level of labor is set equal to one third of the available time endowment.

The second moment implications of the model depend on the parametrization of the shock processes Z_t and D_t. The volatility of the technology innovation is set at $\sigma_z = 0.0148$ following the estimation of the Solow residual $S_t = Z_t^{1-\rho} = Y_t/K_t^{\rho}N_t^{1-\rho}$ on US postwar data in Erceg, Guerrieri and Gust (2005). The values for ρ_d and σ_d are calibrated so that the model can match the second moments of US postwar data. As in King, Plosser and Rebelo (1988) the calibration matches the model's implications for the stationary variables $\log(C/Y)$, $\log(I/Y)$ and $\log(N)$ to US data. Table 1 compares the second moments under the assumption that $\rho_d = 0.8$ and $\sigma_d = 0.009$. Even with only two shocks and absent any source of nominal rigidity, the model can account fairly well for the volatility of the aggregate ratios and hours. The model underpredicts the volatility of the consumption-output ratio, though its performance improves considerably when compared to the sample starting in 1980:1. As is common in Real Business Cycle models, the correlation between hours and the aggregate ratios is much higher than in the data.

3.2. Consequences of truncation and the role of identification

To write the model in terms of the P, Q, R, S, Z_1 matrices define the vectors of endogenous control, endogenous and exogenous state variables respectively as $y_t = [n_t, r_t, \widehat{c}_t, \widehat{y}_t]'$, $x_t = [\widehat{k}_t]$, $z_t = [\widehat{z}_t, d_t]'$ where $n_t, r_t, \widehat{c}_t, \widehat{y}_t, \widehat{k}_t, \widehat{z}_t, d_t$ stand for log-deviations from the steady state of $N_t, R_t, \widehat{C}_t, \widehat{Y}_t, \widehat{K}_t, \widehat{Z}_t, D_t$. The results in the previous section show that any VAR(p) including $\widehat{k}_t$ among the observables is a correct representation of the model data-generating process. Consider instead a VAR(p) for the observable variables $X_t = [\Delta \ln Y_t,\ n_t]$ where $\Delta \ln Y_t$ can be obtained as a linear combination of the model's variables: $\Delta \ln Y_t = \widehat{y}_t - \widehat{y}_{t-1} + \widehat{z}_t$. The data-generating process implies the vector X_t has an infinite order VAR representation.

To generate impulse response functions to the structural shocks the econometrician needs an estimate of the matrix identifying the orthogonal shocks vector ε_t from the reduced form innovations η_t. Define the VAR(p) representation $X_t = \widehat{\Gamma}_1 X_{t-1} + ... + \widehat{\Gamma}_p X_{t-p} + \eta_t$ and the associated VMA representation $X_t = \eta_t + \widehat{\Theta}_1 \eta_{t-1} + \widehat{\Theta}_2 \eta_{t-2} + ...$ where $\widehat{\Gamma}_i, \widehat{\Theta}_i$ indicate magnitudes related to the finite order VAR approximation, whereas Γ_i, Θ_i indicate the corresponding magnitudes for the true VAR representation. Let Λ_0 be the identifying matrix such that $\eta_t = \Lambda_0 \varepsilon_t$. Then:

$$
\begin{aligned}
X_t &= \Lambda_0 \varepsilon_t + \widehat{\Theta}_1 \Lambda_0 \varepsilon_{t-1} + \widehat{\Theta}_2 \Lambda_0 \varepsilon_{t-2} + ... \\
&= \Lambda_0 \varepsilon_t + \widehat{\Lambda}_1 \varepsilon_{t-1} + \widehat{\Lambda}_2 \varepsilon_{t-2} + ... \qquad (18)
\end{aligned}
$$

To isolate the impact of truncation and identification bias assume the econometrician sets $\Lambda_0 = \hat{B}$, where the rows of the matrix $\hat{B}$ are such that they map structural shocks into reduced form shocks η_t consistently with the DSGE model in the true VAR representation of the data, as in eq. (8). The Appendix shows that $\hat{B}$ is composed of the rows of the matrix Q corresponding to the observable variables. Because the impact of a component of the shocks vector ε_{i_t} at time t does not depend on the matrices $\widehat{\Theta}_i$, the identifying matrix $\hat{B}$ has the property that the response of any variable at time t to a ε_{i_t} innovation is exactly the one implied by the theoretical model. But since the VMA representation is obtained from a truncated VAR, the coefficients in the VMA polynomial $\widehat{\Theta}(L)$ are biased. This approximation error is generated entirely by the *truncation bias*.

If the econometrician is not endowed with knowledge of the matrix $\hat{B}$, the biased polynomial matrices $\widehat{\Gamma}_1, \widehat{\Gamma}_2, ...\widehat{\Gamma}_p$ may also affect the VAR performance through the estimation of the identification matrix Λ_0. Consider the Blanchard and Quah (1989) strategy using long run restrictions to identify the technology innovation ε_{z_t}. Any labor supply innovation ε_{d_t} has no long run impact on the level of $\ln Y_t$, while the opposite is true for a technology innovation ε_{z_t}. Since $\Lambda_j(1,2)$ is the impact of ε_{d_t} on $\Delta \ln Y_t$ after j periods, $\Sigma_{j=0}^{\infty}\Lambda_j(1,2)$ is the long run impact of ε_{d_t} on $\ln Y_t$. The restriction $\Sigma_{j=0}^{\infty}\Lambda_j(1,2) = 0$ can be used to build the identifying matrix Λ_0. It implies that the element $(1,2)$ of the matrix $[\overline{\Theta}\Lambda_0]$ be equal to zero since $\Sigma_{j=0}^{\infty}\Lambda_j = \Sigma_{j=0}^{\infty}\Theta_j\Lambda_0 = \overline{\Theta}\Lambda_0$. Define the shocks vector u_t as the normalized structural shocks vector $u_t = \Sigma^{-1/2}\varepsilon_t$ so that $E(u_t u_t') = I$. Since the covariance matrix of the reduced form innovation $\eta_t = \Lambda_0\varepsilon_t$ is equal to $\Omega = \Lambda_0\Sigma\Lambda_0'$ a Cholesky factorization of $[\overline{\Theta}\Omega\overline{\Theta'}] = [\overline{\Theta}\Lambda_0\Sigma^{1/2}\Sigma^{1/2\prime}\Lambda_0'\overline{\Theta'}]$ provides the lower-triangular matrix $C = \overline{\Theta}\Lambda_0\Sigma^{1/2}$ such that $CC' = [\overline{\Theta}\Omega\overline{\Theta'}]$, implying:

$$\widetilde{\Lambda}_0 = \Lambda_0\Sigma^{1/2} = \overline{\Theta}^{-1}C \tag{19}$$

This is the matrix $\widetilde{\Lambda}_0$ such that the element $(1,2)$ of the matrix $[\overline{\Theta}\Lambda_0\Sigma^{1/2}]$ is zero. If the econometrician estimated the infinite order VAR representation of X_t, the long run identification restriction would ensure $\Lambda_0 = \hat{B}$.

Figure 1 shows the impulse response function of n_t obtained from the VAR(2) representation of the vector X_t when the technology shock is identified using the theoretical matrix $\hat{B}$. The impulse response is constrained to be an exact match to the theoretical one at time $t = 1$ by the matrix $\hat{B}$, and in the long run by the fact that the approximating VAR is stationary, as is the true model. Even so, the VAR(2) impulse response is a very inaccurate approximation of the true one. After 10 quarters the magnitude of the response is more than 60% smaller than the theoretical response, and it drops to zero after about 25 quarters - implying a much less persistent response of hours compared to the model.

By using the Blanchard and Quah identification strategy, the truncation bias also generates an identification bias. The impulse response (figure 1) drops to zero after about 25 quarters, but also predicts at time 1 an increase in n_t about 75% larger than the theoretical response. In a similar model, Chari, Kehoe and McGrattan (2005) obtain an analogous result. This experiment illustrates that the poor performance of the approximating VAR in Chari, Kehoe and McGrattan (2005) can be largely explained by the identification bias. Contrary to our results, Erceg, Guerrieri and Gust (2005) conclude that the truncation bias is negligible in population, and is essentially a small sample issue.

A closer examination sheds light on the role of identification in the VAR performance. The error in the estimate of $\widetilde{\Lambda}_0$ can originate from two sources: error in estimating Ω or in estimating $\overline{\Theta}$. Table 2 shows that the true ε_t and estimated $\widehat{\varepsilon}_t$ shocks (orthogonalized using $\hat{B}$) are remarkably close. Therefore also the vector $\widehat{\eta}_t$ must accurately track the true η_t, and the estimated covariance matrix $\widehat{\Omega}$ must be an accurate approximation to Ω. Figure 2 shows that even if a VAR(p) poorly approximates the true VMA representation, the VAR-estimated shocks vector can still accurately approximate the true shocks. The shocks estimates are calculated using the true data vector X_t, therefore the truncation error is not compounded over time, as is the case for the impulse response functions where the estimated response of X_t depends on its lagged estimates.

Consider next the role of the VAR(p) coefficients. To build intuition for the result, we examine the case of a finite order VAR where the lag order p is large enough to appeal to large sample properties of the OLS estimator. Asymptotically, the matrices $\widehat{\Gamma}_1,\ \widehat{\Gamma}_2$ are consistent estimators of the matrices $\Gamma_1,\ \Gamma_2$ from the infinite order VAR representation[5]. The impulse response function, that is the matrices Θ_i, can be calculated from the recursion:

$$\Theta_i = \sum_{j=1}^{i} \Theta_{i-j}\Gamma_j \tag{20}$$

where $\Gamma_0 = \Theta_0 = I$. Clearly, if the Γ_i matrices are very close to zero, also the Θ_i matrices will be. The matrices $\Gamma_1, \Gamma_2, \Gamma_3$ for example can be easily calculated using eq. (9):

$$\Gamma_1 = \begin{bmatrix} 0.0996 & -0.1933 \\ 0.1327 & 0.6904 \end{bmatrix} ;\ \Gamma_2 = \begin{bmatrix} 0.0958 & -0.0023 \\ 0.1276 & -0.0030 \end{bmatrix} ;\ \Gamma_3 = \begin{bmatrix} 0.0921 & -0.0022 \\ 0.1228 & -0.0029 \end{bmatrix}$$

As i increases the matrices Γ_i are relatively close to zero, but they converge extremely slowly: the largest eigenvalue of the matrix $(R - SQ^{-1}P)$ is $\lambda = 0.962$. Since the long run identification relies on the infinite summation $\sum_{i=0}^{\infty} \Theta_i$, eq. (20) shows that neglecting the terms Γ_i for $i > p$ in the VAR(p) representation implies the identification matrix is subject to a considerable error. Using the correct identification matrix $\widehat{B}$ the truncation only feeds through the mistaken restriction $\Gamma_i = 0$ for $i > p$ in eq. (20). The long run identification compounds this mistake because it also makes use of the quantity $\sum_{i=0}^{\infty} \widehat{\Theta}_i$. Identification restrictions that are more robust to truncation would reduce the approximation error[6].

[5] Convergence in probability of the vector of estimated coefficients in the VAR(p) $[\widehat{\Gamma}_1(p),\ \widehat{\Gamma}_2(p),\ ...\widehat{\Gamma}_m(p)]$ to the vector $[\Gamma_1,\ \Gamma_2,\ ...\Gamma_m]$ when the true data generating process is an infinite order VAR is only assured if $p \to \infty$ as the sample size T goes to infinity, albeit at a much slower speed so that $p^3/T \to 0$ (see Lutkepohl, 1993, p.305).

[6] Christiano, Eichenbaum and Vigfusson (2006), Erceg, Guerrieri and Gust (2005), Faust and Leeper (1997) point out that the difficulty in estimating $\sum_{i=0}^{\infty} \widehat{\Theta}_i$ in small sample adversely affect the performance of long run identification restrictions. Sims (1972) first discussed the fact that the sum of an infinite number of coefficients may be extremely difficult to estimate even if the single coefficients are tightly estimated.

3.3. How model parametrization matters

Consider a model where the labor supply shock is a very persistent process by setting $\rho_d = 0.97$. The impulse response function to a technology shock is not affected by such change. Yet Figure 3 shows that the VAR(2) performance is greatly improved. The impulse response function is remarkably accurate using the theoretical identification matrix.

The improvement in performance can be explained by examining the infinite order VAR matrices Γ_i. For $i = 1, 2, 3$ they are:

$$\Gamma_1 = \begin{bmatrix} 0.0161 & -0.0287 \\ 0.0015 & 0.9490 \end{bmatrix} \Gamma_2 = \begin{bmatrix} 0.0154 & -0.0002 \\ -0.0014 & -0.0001 \end{bmatrix} \Gamma_3 = \begin{bmatrix} 0.0147 & -0.0002 \\ -0.0014 & -0.0001 \end{bmatrix}$$

The elements of the matrices Γ_i are now much closer to zero than in the baseline parametrization. This means that (asymptotically) by restricting Γ_i to be equal to zero for $i > p$ a correctly identified VAR(p) is a fairly accurate approximation to the true VAR. Nevertheless, the summation $\sum_{i=0}^{\infty} \widehat{\Theta}_i$ suffers from a large error. The VAR(p) identified using the long run restriction still tracks poorly the time 1 impact of a technology innovation on hours, though it now implies a very persistent response consistently with the DSGE model.

It may seem puzzling that a change in the parametrization of the labor supply shock that does not affect the dynamics of the model after a technology shock has important implications for the performance of the VAR(2). What is required to the VAR representation for the impulse response to a technology shock to be invariant as ρ_d varies is that the first column of the matrix $\Theta_i \widehat{B} = \Theta_i Q$ does not change. The matrix Θ_i itself gives the impulse response function of X_t to a shock in η_t, that is, to the linear combination $\eta_t = Q\varepsilon_t$ of the innovation vector ε_t. Since the matrix Q changes across different parametrizations, there is no reason for any of the elements in Θ_i to stay constant as ρ_d increases. As a consequence, also all the elements in the matrices Γ_i change together with ρ_d.

3.4. How the number of lags included in the VAR matters

A strategy often used by researchers is to include enough lags in the VAR in the hope that the approximation to the correctly specified infinite order VAR would improve.

In the case of zero identification bias, Figure 4 shows that the impulse response function from a correctly identified VAR(p) with $p = 6$ and $p = 12$ is accurate up to the p^{th} lag (the error depicted in the plot converges to zero as the approximating VAR lag order p becomes large). This behaviour is easily explained using eq. (20) and considering that asymptotically the matrices $\widehat{\Gamma}_p$ are consistent estimators of the matrices Γ_p. The matrices Θ_j in the true VMA representation depend only on the infinite order VAR matrices Γ_i up to $i = j$. Under the conditions for which the matrices $\widehat{\Gamma}_i$ from the VAR(p) converge in probability to Γ_i, the estimated impulse response function will be correct up to the p^{th} lag. Yet even including 12 lags has only a small impact in reducing the long run identification bias (Chari, Kehoe and McGrattan, 2005, investigate this result in a related model).

4. Conclusions

This paper discusses the conditions under which a DSGE model has a finite order VAR representation. These conditions are the very implicit assumptions made by the researcher when comparing a DSGE model impulse response functions to the ones obtained from an estimated VAR. Ordinarily a DSGE model has an infinite order VAR representation, unless the vector of endogenous variables is observable. Observability of the exogenous shocks vector is instead irrelevant.

Economists typically assume that including a small number of lags is enough to provide a reasonable approximation to the true VAR. The paper uses an RBC model to show that this assumption can be misleading. The VAR(p) approximation can provide largely inaccurate estimates of the model impulse response functions. The error in the approximation affects the results through two separate channels: the truncated VAR coefficients are biased, and the truncation error may lead to an identification bias. Depending on the parametrization and the identification strategy none, one or both of these channels will weigh on the accuracy of the approximating VAR(p). This result does not rely on small sample volatility of the estimator, nor on the use of mistaken identification assumptions. Identification strategies which are equally correct in the true VAR representation can perform very differently in the truncated VAR estimate. Even if the impulse response functions can be inaccurate, the VAR(p) can provide a close approximation to the true shocks vector.

These results suggest some caution has to be used by researchers relying on VAR evidence to build DSGE models. VARs have much to tell: they summarize the dynamics of the data with as few restrictions as possible. Compared to alternative econometric procedures, they may be more robust to mis-specification and perform better in small sample. Assuming though that the dynamics VARs describe can always be obtained from the structural models economists are interested in testing is misleading. If economists wish to build DSGE models that can account for the correlations among macroeconomic variables, these models should be tested against model-consistent representations of the data.

References

Blanchard, O., and Quah, D., (1989), 'The Dynamic Effects of Aggregate Demand and Supply Disturbances', American Economic Review 79: 655 -680.

Canova, F., and Pina, J., (2005), 'Monetary policy mis-specification in VAR models', in Diebolt, C. and Krystou C., eds., New Trends In Macroeconomics, Springer Verlag, forthcoming.

Chari, V. V., Kehoe, P., and McGrattan, E., (2005), 'A Critique of Structural VARs Using Business Cycle Theory', Federal Reserve Bank of Minneapolis Staff Report 364.

Christiano, L., Eichenbaum, M., and Vigfusson, R., (2006), 'VARs as a Guide to Estimating Dynamic General Equilibrium Models', mimeo, Northwestern University.

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

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

Erceg, C., Guerrieri, L. and Gust, C., (2005), 'Can long run restrictions identify technology shocks?', Journal of the European Economic Association 3: 1237-1278.

Fernandez-Villaverde, J., Rubio-Ramirez, J., and Sargent, T., (2005), 'A, B, C's (and D's) for understanding VARS', mimeo, New York University.

Faust, J. and Leeper, E., (1997), 'When Do Long Run Identifying Restrictions Give Reliable Results?', Journal of Business and Economic Statistics 15: 345-353.

Gali, J. and Rabanal, P., (2005), 'Technology shocks and aggregate fluctuations: how well does the RBC model fit postwar US data?', in Gertler, M. and Rogoff, K., eds., NBER Macroeconomics Annual, MIT Press.

Hansen, G., (1985), 'Indivisible Labor and the Business Cycle', Journal of Monetary Economics, 16: 281-308.

King, R., Plosser, C. and Rebelo, S., (1988), 'Production, Growth and Business Cycles: New Directions', Journal of Monetary Economics, 21: 309-341.

Lutkepohl, H., (1993), Introduction to Multiple Time Series Analysis, Berlin: Springer-Verlag.

Rabanal, P. and Rubio-Ramirez, J., (2005), " Comparing New Keynesian Models of the Business Cycle : A Bayesian approach ", Journal of Monetary Economics 52: 1151-1166.

Sims, C., (1972), 'The Role of Approximate Prior Restrictions in Distributed Lag Estimation', Journal of the American Statistical Association 67(337): 169-175.

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

5. Appendix

PROOF OF PROPOSITION 2.1 Assume $[I - RL]$ is invertible. Eq. (2) implies $x_{t-1} = [I - RL]^{-1} S L z_t$. Substituting x_t in the control variables equation, and since $\widehat{Y}_t = y_t$:

$$y_t = Q z_t + P G(L)^{-1} S L z_t \tag{21}$$

where $G(L)^{-1} = [I - RL]^{-1}$ is a lag polynomial of potentially infinite order.

VARMA representation for $Z(L) = I$ If $z_t = \varepsilon_t$ eq. (21) is a VMA representation of y_t. If Q is invertible eq. (21) can be written as $y_t = \eta_t + P G(L)^{-1} S Q^{-1} L \eta_t$ where $\eta_t = Q \varepsilon_t = Q z_t$ is the reduced form innovations vector. The matrix Q is the theoretical identifying matrix needed to map ε_t into η_t in the true VMA representation of the model. The inverse of $G(L)$ can be written in terms of its determinant $|G(L)|$, of order n in the lag operator L, and the adjoint matrix $D_G(L)$ of order $(n-1)$ in L: $G(L)^{-1} = D_G(L)|G(L)|^{-1}$. Therefore:

$$|G(L)| y_t = |G(L)| \eta_t + P D_G(L) S Q^{-1} L \eta_t = G^*(L) \eta_t \tag{22}$$

Eq. (22) is a VARMA(n,n). The system (22) is written in final equations form: each component of the vector y_t depends only on its own lags. Since the matrix for L of order zero in both lag polynomials $|G(L)|$ and $G^*(L)$ is the identity matrix, the VARMA representation is unique.

VAR representation for $Z(L) = I$ If $G^*(L)$ is invertible, a VAR representation for y_t is:

$$|G(L)| G^*(L)^{-1} y_t = \eta_t \tag{23}$$

The necessary and sufficient condition for eq. (22) to be a finite order VAR is that the invertible univariate operator $|G^*(L)|$ be of degree zero in L. If this is the case, $G^*(L)$ is a unimodular lag operator and $G^*(L)^{-1}$ is of finite order (Lutkephol, 1993, p. 245). This property follows from the fact that the inverse of $G^*(L)$ can be expressed as $G^*(L)^{-1} = D_{G^*}(L)|G^*(L)|^{-1}$. The adjoint matrix $D_{G^*}(L)$ is a finite order lag operator, while the inverse of the univariate operator $|G^*(L)|$ is of infinite order, unless $|G^*(L)|$ is a constant. The only case in which the product $|G(L)|G^*(L)^{-1}$ in eq. (23) would be of finite order when $G^*(L)$ is not a unimodular operator occurs when $G^*(L) = D_G(L)$. But this equality will be true only when all the variables included in the system belong to the state vector. Then $|G(L)|y_t = D_G(L)SLz_t$ and $G(L)y_t = SLz_t$ since $D_G(L)^{-1}|G(L)| = G(L)$ (where we assumed, WLOG, that $m = n$). Similarly, if all the state variables are included in the system, together with endogenous variables, the system can be rewritten as $[I - AL]Y_t = Bz_t$ where $Y_t,\ A,\ B$ are defined in eq. (7). This process has a VARMA representation $|G(L)|Y_t = D_G(L)Bz_t$ with $G(L) = [I - AL]$, and also in this case it obtains $|G(L)|G^*(L)^{-1} = |G(L)|D_G(L)^{-1} = G(L)$ (where we assumed, WLOG, that $r + m = n$).

VARMA representation for $Z(L) \neq I$ Write Eq. (22) as $|G(L)|y_t = G^*(L)QZ(L)^{-1}\varepsilon_t$. Define $QZ(L)Q^{-1} = [I - QZ_1Q^{-1}L ... - QZ_pQ^{-1}L^p] = \widetilde{Z}(L)$. Then $QZ(L)^{-1}\varepsilon_t = \widetilde{Z}(L)^{-1}\eta_t$ and $|G(L)|y_t = G^*(L)D_{\widetilde{Z}}(L)|\widetilde{Z}(L)|^{-1}\eta_t$ where $|G^*(L)|$ is of order nm in L, $D_{G^*}(L)$ is of order $n(m-1)$ in L. Therefore y_t is described by:

$$|\widetilde{Z}(L)||G(L)|y_t = G^*(L)D_{\widetilde{Z}}(L)\eta_t \tag{24}$$

Since $|\widetilde{Z}(L)|$ is of order pm in L, eq. (24) is a VARMA(n+pm,n+p(m-1)) process.

VAR representation for $Z(L) \neq I$ A VAR representation for y_t is given by:

$$\widetilde{Z}(L)G^*(L)^{-1}|G(L)|y_t = \eta_t \tag{25}$$

which will not be of finite order unless the conditions for the VAR defined in eq. (23) to be of finite order are met. ■

Proof of Corollary 2.2 The equilibrium law of motion implies $z_{t-1} = Q^{-1}(y_{t-1} - Px_{t-1})$. Substituting in eq. (2) it holds that $H(L)x_t = SQ^{-1}y_{t-1}$ where $H(L) = [I - (R -$

$SQ^{-1}P)L]$. If $H(L)$ is invertible:

$$y_t = Qz_t + P[H(L)^{-1}SQ^{-1}]Ly_t \tag{26}$$

or $y_t = P\sum_{j=0}^{\infty}[R - SQ^{-1}P]^j SQ^{-1}L^{j+1}y_t + \eta_t$. This is the derivation obtained in Fernandez-Villaverde et al. (2005). To obtain a finite order VAR $H(L)$ must be a unimodular operator. This will happen when $|H(L)|$ is of degree zero in L. Since eq. (23) and eq. (26) define the same VAR process, this condition is equivalent to the condition for a unimodular operator established in terms of $|G^*(L)|$. Since the VAR representation (25) can be rewritten using eq. (26) as:

$$\widetilde{Z}(L)\{I - P[H(L)^{-1}SQ^{-1}]L\}y_t = \eta_t \tag{27}$$

it follows that the condition for the existence of a finite order VAR representation when $Z(L) \neq I$ can be expressed as the requirement that $|H(L)|$ be of degree zero in L. ■

Proof of Proposition 2.3 WLOG assume that only the first $n-1$ components of x_t are observable, and the n^{th} component $\widehat{x}_t$ is unobservable. Define the vector of observable variables $\widehat{Y}_t = [\overline{x}_t \; y_t]'$ where y_t is an $r \times 1$ vector of endogenous variables, $x_t = [\overline{x}_t \; \widehat{x}_t]'$ and $(n-1)+r = m$. Then $\widehat{x}_{t-1} = Tx_{t-1} = T[I - RL]^{-1}SLz_t$ where $T = [0...0 \; 1]$ is a $1 \times n$ row vector where the first $n-1$ components are equal to zero. Partition the matrix P so that $P = [\overline{P} \quad \widehat{P}]$ where $\overline{P}$ is an $r \times (n-1)$ matrix and $\widehat{P}$ is an $r \times 1$ matrix. The vectors y_t and $\overline{x}_t$ can be written as:

$$\begin{aligned} y_t &= \overline{P}\overline{x}_{t-1} + \widehat{P}\widehat{x}_{t-1} + Qz_t \\ \overline{x}_t &= \overline{R}\overline{x}_{t-1} + \widehat{R}\widehat{x}_{t-1} + \overline{S}z_t \end{aligned}$$

$\overline{R}$ is the matrix composed of the first $n-1$ rows and columns of the matrix R, $\widehat{R}$ is a vector containing the first $n-1$ components of the last column of R, and $\overline{S}$ contains the first $n-1$ rows and all the m columns of the matrix S. We can then write the process for $\widehat{Y}_t$ as:

$$
\begin{aligned}
\widehat{Y}_t &= \begin{bmatrix} \overline{R} & {}^{(n-1)\times r}0 \\ \overline{P} & {}^{r\times r}0 \end{bmatrix} \widehat{Y}_{t-1} + \begin{bmatrix} \overline{S} \\ Q \end{bmatrix} z_t + \begin{bmatrix} \widehat{R} \\ \widehat{P} \end{bmatrix} T[I-RL]^{-1}SLz_t \\
&= \overline{\Gamma_1}\widehat{Y}_{t-1} + \overline{\Gamma_2}z_t + \overline{\Gamma_3}G(L)^{-1}SLz_t \qquad (28)
\end{aligned}
$$

Defining $\overline{Y}_t = [I - \overline{\Gamma_1}L]\widehat{Y}_t$ obtain $\overline{Y}_t = \overline{\Gamma_2}z_t + \overline{\Gamma_3}G(L)^{-1}SLz_t$ which has the same functional form as eq. (21). The steps of the proof of Proposition 2.1 follow through unchanged for the observable variable $\overline{Y}_t$. The variable $\widehat{Y}_t$ will have a VARMA(n+pm+1,n+p(m-1)) representation since the observable $\overline{x}_t$ vector introduces an extra lag. ■

Table 1

Second Moments: Real Business Cycle model and US data

	Variable	Std. Dev.	Relative Std. Dev.	Cross-correlation with *log(N)*
Model				
	log(C/Y)	3.17	0.75	−0.96
	log(I/Y)	9.25	2.18	0.96
	log(N)	4.22	1	1
Data: 1955:1-2006:1				
	log(C/Y)	4.72	1.08	−0.77
	log(I/Y)	9.57	2.19	0.73
	log(N)	4.36	1	1
Data: 1980:1-2006:1				
	log(C/Y)	2.91	0.65	−0.85
	log(I/Y)	10.03	2.26	0.68
	log(N)	4.42	1	

Note: Standard deviation measured in percent. Relative standard deviation is ratio to standard deviation of $\log(N)$. Sample moments for US data are obtained from quarterly per capita values of Y, C, I, N. Y is measured as real GDP net of government consumption expenditures. C is real personal consumption expenditures of non-durables and services. I is real gross private fixed investment. The measure for total per capita labor hours N of all workers is equal to average weekly hours for private industries multiplied by the ratio between the total number of workers employed in the non-farm sector and the civilian non-institutional population. The average weekly hours series starts in 1964:1. All series are seasonally adjusted and obtained from the US Bureau of Labor Statistics.

Table 2

VAR(2) performance - Estimated Identified Shocks Vector $\widehat{\varepsilon}_t$

Shock	Technology	Labor supply
Correlation with true shock	0.99	0.98
Relative Root Mean Square Error	3.99%	16.04%

Note: Root mean square distance between the VAR(2)-estimated vector $\widehat{\varepsilon}_t$ and the true vector calculated over 1.5 million observations. The shocks vector $\widehat{\varepsilon}_t$ is obtained from reduced form innovations η_t using the theoretical identification matrix. Data are generated by the DSGE model with labor supply shock autocorrelation $\rho_d = 0.8$. The RMSE is scaled by the standard deviation of the corresponding shock.

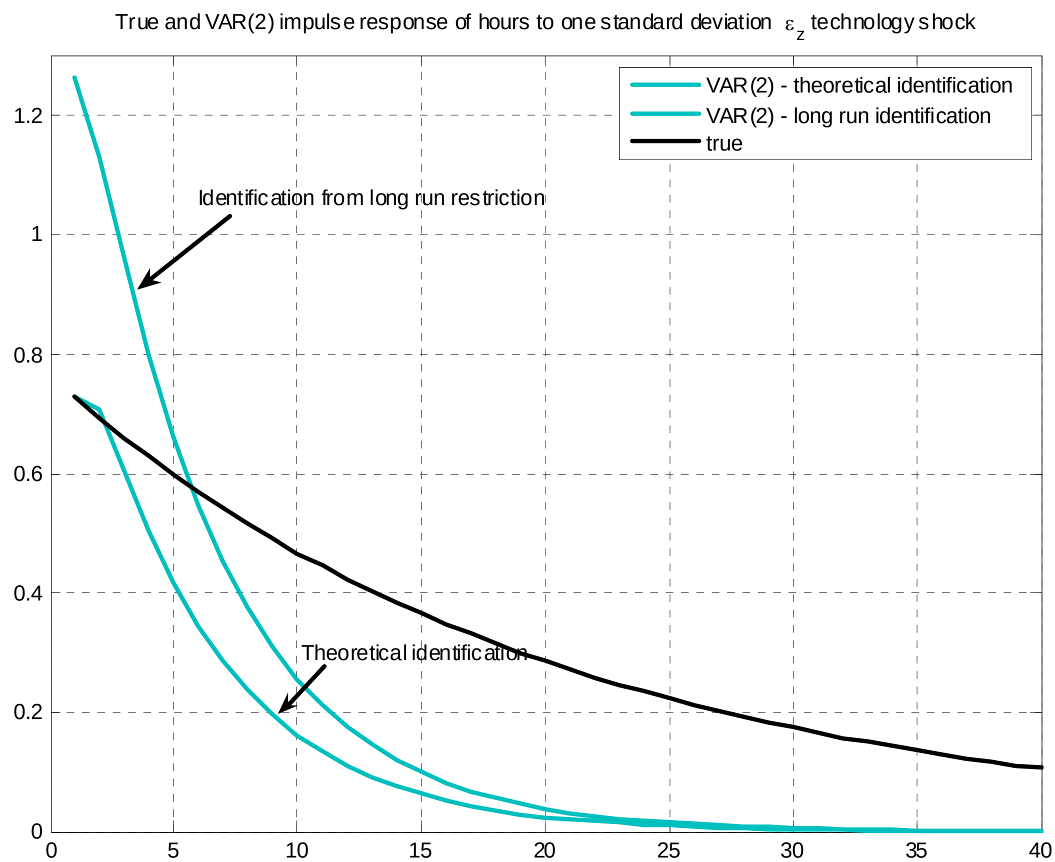

Figure 1: Impulse response to technology shock ε_z in correct and approximating VAR. The VAR(2) coefficients are computed from population orthogonality conditions. Scaling is in percentage points. Preference shock autocorrelation $\rho_d = 0.8$.

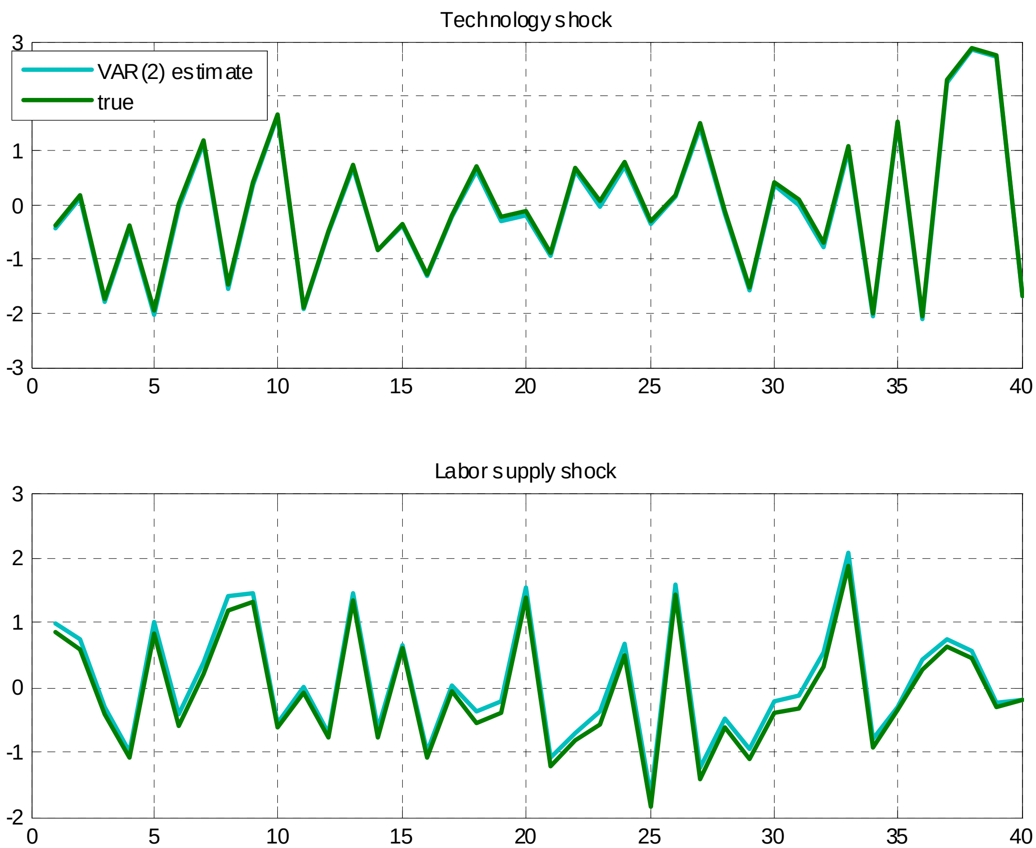

Figure 2: Ten year sample path of VAR(2)-estimated series of the shocks vector ε and true series, $\rho_d = 0.8$. The theoretical matrix from the DSGE model identifies the structural shocks. Scaling is in percentage points.

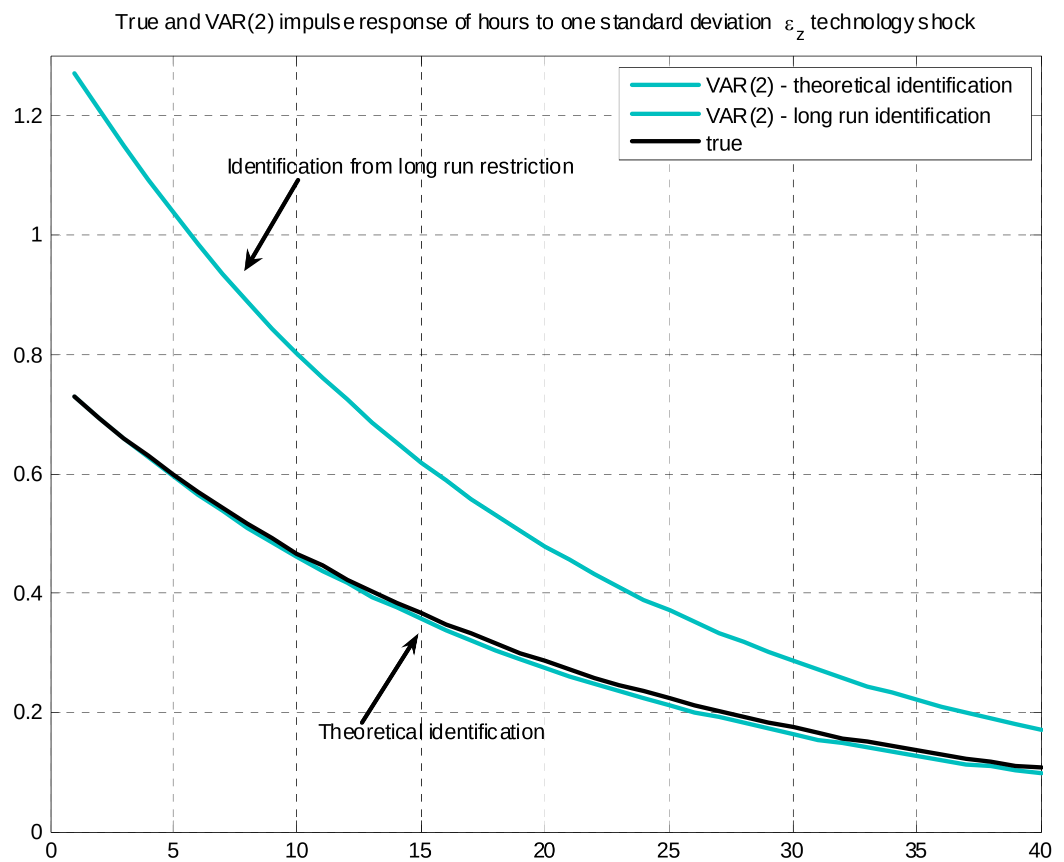

Figure 3: Impulse response to technology shock ε_z in correct and approximating VAR. The VAR(2) coefficients are computed from population orthogonality conditions. Scaling is in percentage points. Preference shock autocorrelation $\rho_d = 0.97$.

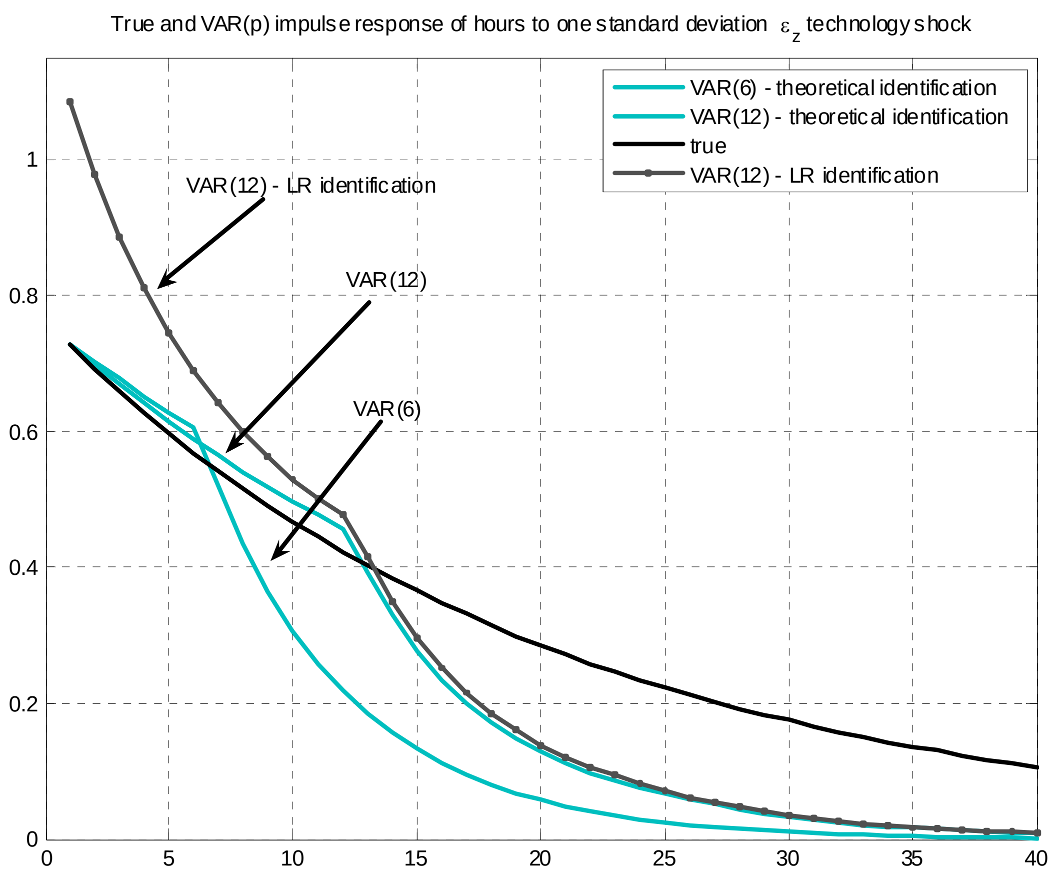

Figure 4: Impulse response to technology shock ε_z in correct and approximating VAR. The VAR(p) coefficients are computed from population orthogonality conditions. Scaling is in percentage points. Preference shock autocorrelation $\rho_d = 0.8$.