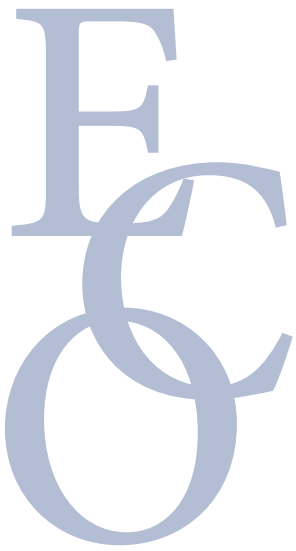
EUI WORKING PAPERS
ECO No. 2006/37



Business Cycle Analysis and VARMA Models

CHRISTIAN KASCHA AND KARL MERTENS

Department of Economics

EUROPEAN UNIVERSITY INSTITUTE
DEPARTMENT OF ECONOMICS

Business Cycle Analysis and VARMA Models

CHRISTIAN KASCHA AND KARL MERTENS

EUI Working Paper **ECO** No. 2006/37

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50016 San Domenico di Fiesole (FI)
Italy

http://www.eui.eu/
http://cadmus.eui.eu/

Business Cycle Analysis and VARMA models*

Christian Kascha, Karel Mertens†

October 2006

Abstract

An important question in empirical macroeconomics is whether structural vector autoregressions (SVARs) can reliably discriminate between competing DSGE models. Several recent papers have suggested that one reason SVARs may fail to do so is because they are finite-order approximations to infinite-order processes. In this context, we investigate the performance of models that do not suffer from this type of misspecification. We estimate VARMA and state space models using simulated data from a standard economic model and compare true with estimated impulse responses. For our examples, we find that one cannot gain much by using algorithms based on a VARMA representation. However, algorithms that are based on the state space representation do outperform VARs. Unfortunately, these alternative estimates remain heavily biased and very imprecise. The findings of this paper suggest that the reason SVARs perform weakly in these types of simulation studies is not because they are simple finite-order approximations. Given the properties of the generated data, their failure seems almost entirely due to the use of small samples.

JEL classification: E32, C15, C52
Keywords: Structural VARs, VARMA, State Space Models, Identification, Business Cycles

*We would like to thank Helmut Lütkepohl, Morten O. Ravn and participants at the "Recent Developments in Econometrics" conference in Florence for valuable comments and discussion.

†European University Institute, Department of Economics, Via della Piazzuola 43, 50133 Firenze, Italy; Contact Author E-mail: christian.kascha@eui.eu.

1 Introduction

Structural VARs are a widely used tool in empirical macroeconomics, particularly for the evaluation of DSGE models.[1] The results from SVARs are often viewed as stylized facts that economic models should replicate. However, there has recently been some debate whether SVARs can actually discriminate between competing DSGE models and whether their sampling properties are good enough to justify their popularity in applied macroeconomics. In response to the seminal paper by Gali (1999), the discussion has focussed on the impact of technology shocks on hours worked, identified using restrictions on the long-run impact matrix of the structural errors. In their contributions, Chari, Kehoe and McGrattan (2005) and Christiano, Eichenbaum and Vigfusson (2006) investigate the properties of the estimators derived from SVARs by simulating an artificial data generating process (DGP) derived from a simple RBC model and by comparing true with estimated impulse responses.

According to Chari et al. (2005), SVARs fail dramatically for both a level and difference specification of hours worked. Even with a correct specification of the integration properties of the series, the SVAR overestimates in most cases the impact of technology on labor and the estimates display high variability. However, Christiano et al. (2006) argue that the parametrization chosen by Chari et al. (2005) is highly unrealistic. With their preferred parametrization, Christiano et al. (2006) find that both long-run and short-run

[1] Examples in the literature are, among many others, Blanchard and Quah (1989), as well as King, Plosser, Stock and Watson (1991), Christiano and Eichenbaum (1992) and Gali (1999).

identification schemes display only small biases and argue that, on average, the confidence intervals produced by SVARs correctly reflect the degree of sampling uncertainty. In addition, they find that short-run identification schemes work much better compared to identification via long-run restrictions. Nevertheless, Christiano et al. (2006) also conclude that the estimates obtained via a long-run identification scheme are extremely imprecise. These results have been further confirmed by Erceg, Guerrieri and Gust (2005). Thus with long-run restrictions, one can often not even make a correct inference about the *sign* of the structural impulse responses. The question is therefore if one should use this type of identification scheme at all. On the other hand, long-run identification is attractive from a theoretical point of view, since it usually requires much weaker assumptions than short-run identification and is in any case a useful additional tool for model evaluation.

The failure of finite-order SVARs is sometimes attributed to the fact that they are only approximations to infinite-order VAR processes or to the possibility that there does not exist a VAR representation at all. For example, Cooley and Dwyer (1998) give an example of an economic model that implies a vector autoregressive moving average (VARMA) representation of the data series and state: "*While VARMA models involve additional estimation and identification issues, these complications do not justify* systematically *ignoring these moving average components, as in the SVAR approach*". As further shown by Fernández Villaverde, Rubio Ramírez and Sargent (2005), DSGE models generally imply a state space system that has a VARMA and eventually an infinite VAR representation. Fernández Villaverde et al. (2005) propose the inclusion of moving average terms if the DSGE model

at hand does not permit an infinite VAR representation. Christiano et al. (2006) state that "*Given our data generating processes, the true VAR of the data has infinite lags. However, the econometrician can only use a finite number of lags in the estimation procedure. The resulting specification error is the reason why in some of our examples the sum of VAR coefficients is difficult to estimate accurately*".

This paper explores the possible advantages of structural VARMA and state space models that capture the full structure of the time series representation implied by DSGE models. We investigate whether estimators based on these alternative models can outperform SVARs in finite samples.[2] Our question is important for several reasons. First, it is useful to find out to what extent the failure of SVARs in these simulation studies is due to the omission of moving average components. Second, whether estimators based on alternative representations of the same DGP have good sampling properties is interesting in itself. Employing these alternatives enables researchers to quantify the robustness of their results by comparing different estimates.

In order to assess whether the inclusion of a moving average component leads to important improvements, we stick to the research design of Chari et al. (2005) and Christiano et al. (2006): We simulate DSGE models and fit different reduced form models to recover the structural shocks. We then compare the performance of these different procedures by focussing on the estimated contemporaneous impact. The estimation procedure employs a

[2]After finishing the paper, we found out that McGrattan's (2006) work is closely related to our paper. In a different setting, McGrattan (2006) also investigated independently whether the application of state space or VARMA models with minimal structural assumptions can uncover statistics of interest. Her work focusses on different business cycle statistics, while we are exclusively concerned with classical structural estimation.

Gauss-Newton algorithm for the VARMA models, and both a prediction error method and a subspace algorithm for the state space models. One of the findings of this exercise is that one could indeed perform slightly better by taking the full structure of the DGP into account: While the algorithm for VARMA models and the prediction error method do not perform significantly better (and sometimes worse), the subspace algorithm for state space models outperforms VARs in terms of mean squared error. Unfortunately, we also find that even these alternative estimators are highly volatile and are also ill-suited to discriminate reliably between different DSGE models. One of the implications is that SVARs do not fail in these simulation studies because they are only finite-order approximations. Given the properties of the data, the poor performance of SVARs is most likely due to the fact that the long-run identification approach is inappropriate with small samples.

The rest of the paper is organized as follows. In section 2 we present the RBC model used by Chari et al. (2005) and Christiano et al. (2006) that serves as the basis for our Monte Carlo simulations. In section 3 we discuss the different statistical representations of the observed data series. In section 4 we present the specification and estimation procedures and the results from the Monte Carlo simulations. Section 5 concludes.

2 A simple RBC model

The DGP for the simulations is based on a very basic RBC model taken from Chari et al. (2005). In the model, a technology shock is the only shock that affects labor productivity in the long-run, which is the crucial identifying

assumption made by Gali (1999) to assess the role of technology shocks in the business cycle. For this reason, Chari et al. (2005) consider the model as a best case scenario for the SVAR approach.

Households choose infinite sequences of per capita consumption, labor and capital, $\{C_t,\, L_t,\, K_{t+1}\,\}_{t=0}^{\infty}$, to maximize expected lifetime utility

$$E_0 \sum_{t=0}^{\infty} [\beta(1+\gamma)]^t \left[\log C_t + \psi \frac{(1-L_t)^{1-\sigma}}{1-\sigma} \right],$$

given an initial capital stock K_0, and subject to a set of budget constraints given by

$$C_t + (1+\tau_x)\left((1+\gamma)K_{t+1} - (1-\delta)K_t\right) \leq (1-\tau_{lt})w_t L_t + r_t K_t + T_t,$$

for $t = 0, ..., \infty$, where w_t is the wage, r_t is the rental rate of capital, T_t are lump-sum government transfers and τ_{lt} is an exogenous labor tax. The parameters include the discount factor $\beta \in (0,1)$, the labor supply parameters, $\psi > 0$ and $\sigma > 0$, the deprecation rate $\delta \in (0,1)$, the population growth rate $\gamma > 0$ and a constant investment tax τ_x. The production technology is

$$Y_t = K_t^{\alpha}(X_t L_t)^{1-\alpha},$$

where X_t reflects labor-augmenting technological progress and $\alpha \in (0,1)$ is the capital income share. Competitive firms maximize $Y_t - w_t L_t - r_t K_t$. The economy-wide resource constraint is $Y_t \geq C_t + (1+\gamma)K_{t+1} - (1-\delta)K_t$.

The model contains two exogenous shocks, a technology shock and a tax

shock, which follow the stochastic processes

$$\log X_{t+1} = \mu + \log X_t + \sigma_x \epsilon_{x,t+1},$$
$$\tau_{lt+1} = (1-\rho)\bar{\tau}_l + \rho\tau_{lt} + \sigma_l \epsilon_{l,t+1},$$

where $\epsilon_{x,t}$ and $\epsilon_{l,t}$ are independent random variables with mean zero and unit standard deviation. $\sigma_x > 0$ and $\sigma_l > 0$ are the standard deviations of the two shocks, $\mu > 0$ is the mean growth rate of technology, $\bar{\tau}_l > 0$ is the mean labor tax and $\rho \in (0,1)$ measures the persistence of the tax process. Hence, the model has two independent shocks: a unit root process in technology and a stationary AR(1) process in the labor tax.

3 Statistical Representations

Fernández Villaverde et al. (2005) show how the solution of a detrended, log-linearized DSGE model leads to different statistical representations of the model-generated data. This section presents several alternative ways to write down a statistical model for the bivariate, stationary time series

$$\mathbf{y}_t = \begin{bmatrix} \Delta \log(Y_t/L_t) \\ \log(L_t) \end{bmatrix}.$$

Labour productivity growth $\Delta \log(Y_t/L_t)$ and hours worked $\log(L_t)$ are also the series analyzed by Gali (1999), as well as Chari et al. (2005) and Christiano et al. (2006). The appendix provides more detail on the derivations. Given the log-linearized solution of the RBC model of the previous section,

we can write down the law of motion of the logs

$$\begin{aligned}
\log k_{t+1} &= \phi_1 + \phi_{11} \log k_t - \phi_{11} \log x_t + \phi_{12} \tau_t, \\
\log y_t - \log L_t &= \phi_2 + \phi_{21} \log k_t - \phi_{21} \log x_t + \phi_{22} \tau_t, \\
\log L_t &= \phi_3 + \phi_{31} \log k_t - \phi_{31} \log x_t + \phi_{32} \tau_t,
\end{aligned}$$

where $k_t = K_t / X_{t+1}$ and $y_t = Y_t / X_t$ are capital and output detrended with the unit-root shock and the ϕ's are the coefficients of the calculated policy rules. Following Fernández Villaverde et al. (2005) the system can be written in state space form. The state transition equation is

$$\begin{aligned}
\begin{bmatrix} \log k_{t+1} \\ \tau_t \end{bmatrix} &= K_1 + A \begin{bmatrix} \log k_t \\ \tau_{t-1} \end{bmatrix} + B \begin{bmatrix} \epsilon_{x,t} \\ \epsilon_{lt} \end{bmatrix}, \\
\mathbf{x}_{t+1} &= K_1 + A \mathbf{x}_t + B \epsilon_t
\end{aligned}$$

and the observation equation is

$$\begin{aligned}
\begin{bmatrix} \Delta \log(Y_t / L_t) \\ \log L_t \end{bmatrix} &= K_2 + C \begin{bmatrix} \log k_t \\ \tau_{t-1} \end{bmatrix} + D \begin{bmatrix} \epsilon_{x,t} \\ \epsilon_{lt} \end{bmatrix}, \\
\mathbf{y}_t &= K_2 + C \mathbf{x}_t + D \epsilon_t,
\end{aligned}$$

where K_1, A, B, K_2, C and D are constant matrices that depend on the coefficients of the policy rules and therefore on the "deep" parameters of the model. The state vector is $\mathbf{x}_t = [\log k_t, \ \tau_{t-1}]'$ and the noise vector is $\epsilon_t = [\epsilon_{xt}, \ \epsilon_{lt}]'$. Note that the system has a state vector of dimension two with the logarithm of detrended capital and the tax rate shock as state

components.

The above state space system is still a structural model, since the formulation contains the non-observable state vector and the structural errors. We now show different representations of the system for $\mathbf{y}_t$, which can actually be estimated. Given certain invertibility conditions on the system matrices, A, B, C, D, there is an **infinite VAR representation**:

$$\mathbf{y}_t = K_3 + C\left(I - (A - BD^{-1}C)L\right)^{-1} BD^{-1}\mathbf{y}_{t-1} + D\epsilon_t, \tag{1}$$

or

$$\mathbf{y}_t = K_3 + \sum_{i=1}^{\infty} \Pi_i \mathbf{y}_{t-i} + u_t,$$

where K_3 and $\Pi_i, i = 1, 2, \ldots$ are constant coefficient matrices, L denotes the lag operator, $u_t = D\epsilon_t$ and $u_t \sim N(0, DD')$. Note that a condition for the existence of an infinite VAR representation is that the eigenvalues of $(A - BD^{-1}C)$ are strictly less than one in modulus. In practice, it is only possible to approximate this structure by a finite-order VAR.

Alternatively, the system can be written as a **state space model** in "innovations form":

$$\begin{aligned} \mathbf{x}_{t+1} &= K_1 + A\mathbf{x}_t + Ku_t, \\ \mathbf{y}_t &= K_2 + C\mathbf{x}_t + u_t, \end{aligned} \tag{2}$$

where the innovation, u_t, is defined as above and $K = BD^{-1}$. In contrast to the VAR representation in (1), it is possible to estimate (2) exactly.

Finally, the underlying DGP can be represented by a **VARMA(1,1) representation**:

$$\begin{aligned} \mathbf{y}_t &= K_4 + CAC^{-1}\mathbf{y}_{t-1} + \left(D + (CB - CAC^{-1}D)L\right)\epsilon_t, \qquad (3) \\ \mathbf{y}_t &= K_4 + A_1\mathbf{y}_{t-1} + u_t + M_1 u_{t-1}, \end{aligned}$$

where the last equation defines A_1, M_1 and $u_t = D\epsilon_t$. As with the state space representation representation, the VARMA(1,1) representation can also be estimated exactly.

Given the conditions stated in Fernández Villaverde et al. (2005), all three representations are algebraically equivalent. That is, given the same input sequence $\{\epsilon_t\}$, they produce the same output sequence $\{\mathbf{y}_t\}$. The representations are however not statistically equivalent: the properties of estimators and tests depend on the chosen statistical representation. It should be emphasized that we are always interested in the same process and ultimately in the estimation of the same coefficients, i.e. those associated with the first-period response of $\mathbf{y}_t$ to a unit technology shock, $\epsilon_{x,t}$. However, the different representations require different estimation algorithms and therefore our comparative study can be regarded as a comparison of different algorithms to estimate the same linear system.

4 Long-run Identification in the RBC Model

4.1 Monte Carlo Design and Econometric Techniques

To investigate the properties of the different estimators, we simulate 1000 samples of the vector series $\mathbf{y}_t$ in linearized form and transform log-deviations to values in log-levels. As in the previous Monte Carlo studies, the sample size is 180 quarters. We use two different sets of parameter values: The first is due to Chari et al. (2005) and is referred to as the CKM-specification, while the second is the one used by Christiano et al. (2006) and is labelled the KP-specification, referring to estimates obtained by Prescott (1986).[3] The specific parameter values are given in Table 1. Christiano et al. (2006) show that the key difference between the specifications is the implied fraction of the variability in hours worked that is due to technology shocks. Table 1 also provides the eigenvalues of the autoregressive and moving average matrices of the VARMA representation, together with the eigenvalues of the Kalman gain K. In terms of these values, the time series properties are very similar and indicate why estimation could be difficult. Note that the moving average part is not of full rank and the associated eigenvalue is close to unity in modulus. Also, the eigenvalues of the autoregressive part are close to one and close to the eigenvalue of the moving average part in modulus. The fact that one eigenvalue of the moving average part is close to one eigenvalue of the autoregressive part could imply that the VARMA(1,1) representation is

[3]Both parametrizations are obtained by maximum likelihood estimation of the theoretical model, using time series on productivity and hours worked in the US. However, because of differences in approach, both papers obtain different estimates and therefore reach different conclusions.

close to being not identified (Klein, Melard and Spreij 2004).

To check the robustness of our results, we also consider variations of the benchmark models. As in Christiano et al. (2006), we consider different values for the preference parameter σ and the standard deviation of the labor tax, σ_l. These variations change the fraction of the business cycle variability that is due to technology shocks. The different values for σ can be seen in Table C.2. For the CKM specification, we also consider cases where σ_l assumes a fraction of the original benchmark value.

Turning to the issue of identification, consider the following infinite moving average representation of $\mathbf{y}_t$

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \Phi_{u,i} u_{t-i} = \Phi_u(L) u_t,$$

where we abstract from the intercept term. Let

$$\Phi_u(1) = I + \Phi_{u,1} + \Phi_{u,2} + \dots$$

be the long-run impact matrix of the reduced form error u_t. Note that the existence of the infinite sum depends on the stationarity of the series. If the stationarity requirement is violated or "nearly" violated, then the long-run identification scheme is not valid or may face difficulties. Also note that the matrix D defined in section 3 gives the first-period impact of a unit shock in ϵ_t. Using the relation $u_t = D\epsilon_t$, we know that $\Phi_\epsilon(1) = \Phi_u(1)D$ and further $\Sigma_u = DD'$, where $\Phi_\epsilon(1)$ is the long-run impact matrix of the underlying structural errors and Σ_u is the covariance matrix of u_t. The identifying

restriction on $\Phi_\epsilon(1)$ is that only the technology shock has a permanent effect on labor productivity. This restriction implies that in our bivariate system the long-run impact matrix is triangular,

$$\Phi_\epsilon(1) = \begin{pmatrix} \Phi_{11} & 0 \\ \Phi_{21} & \Phi_{22} \end{pmatrix},$$

and it is assumed that $\Phi_{11} > 0$. Using $\Phi_\epsilon(1)\Phi'_\epsilon(1) = \Phi_u(1)\Sigma_u\Phi'_u(1)$ we can obtain $\Phi_\epsilon(1)$ from the Cholesky decomposition of $\Phi_u(1)\Sigma_u\Phi'_u(1)$. The contemporaneous impact matrix can be recovered from $D = [\Phi_u(1)]^{-1}\Phi_\epsilon(1)$. Correspondingly, the estimated versions are

$$\begin{aligned} \hat{\Phi}_\epsilon(1) &= \text{chol}[\hat{\Phi}_u(1)\widehat{\Sigma}_u\hat{\Phi}'_u(1)], \\ \hat{D} &= [\hat{\Phi}_u(1)]^{-1}\hat{\Phi}_\epsilon(1). \end{aligned}$$

Only the first column of $\hat{D}$ is identified and is our estimate of the first-period impact of the technology shock.

Next, we comment on the estimation techniques. First, note that for each representation there are several possible estimation methods. We chose algorithms that are both popular in the literature and known to work well in general. Of course, it is possible that there are algorithms that work slightly better for one of the representations in the current setting. However, the aim of this study is primarily to quantify whether the inclusion of the moving average term alone leads to important gains in terms of more precise estimates of the structural parameters.

Vector Autoregressive Models: Structural VARs are well known, so we comment only on a few issues. Fernández Villaverde et al. (2005) show that for the CKM-specification, there exists an infinite VAR representation. We verified that the same is true for the benchmark KP-specification. As in the previous Monte Carlo studies, the VAR lag length is set at four. However, for different sets of parameter values a VAR with different lags may yield slightly better results. We have chosen to stick to the VAR(4) because we want to facilitate comparison with the results of Christiano et al. (2006) and because there was no lag order that performed uniformly better for all DGPs.

State Space Models: There are many ways to estimate a state space model, e.g., the Kalman-based maximum likelihood methods and subspace identification methods such as N4SID of van Overschee and DeMoor (1994) or the CCA method of Larimore (1983). An obvious candidate is maximum likelihood. Therefore, we included a prediction error method that is implemented with the PEM routine in the MATLAB system identification toolbox. However, it is well-known that maximum likelihood methods can face numerical problems that are due to the dependence on starting values, nonlinear optimization or local minima. Indeed, these problems also apply to our setting. Therefore, we also use the CCA subspace algorithm that is asymptotically equivalent to maximum likelihood and was found to be remarkably accurate in small samples. As argued in Bauer (2005), CCA might be the best algorithm for econometric applications. The basic idea of subspace methods is that the state, $\mathbf{x}_t$, summarizes all information of the past

that can be used for mean square prediction. Thus, the center of attention is the state that is estimated in a first step. In a second step the coefficient matrices are estimated by OLS. The different subspace algorithms use the structure of the state space representation in various ways. For general introductions to subspace methods see Bauer (2005) and the appendix.

While implementing the algorithm, we chose the correct dimension of the state vector $n = 2$.[4] To calculate the long-run effect of the prediction errors, it is necessary to solve the state space equations $\mathbf{x}_{t+1} = A\mathbf{x}_t + Ku_t,\ \mathbf{y}_t = C\mathbf{x}_t + u_t$, where the deterministic component is omitted. The transfer function is given by

$$\Phi_u(z) = I + \sum_{j=0}^{\infty} CA^j z^{j+1} K = I + zC(I - zA)^{-1}K,$$

where $z \in \mathbb{C}$. The long-run impact matrix of the reduced form error is given by $\Phi_u(1)$. An estimate can be obtained from the estimated system matrices, $\hat{A}, \hat{C}, \hat{K}$, as $\widehat{\Phi}_u(1) = I + \hat{C}(I - \hat{A})^{-1}\hat{K}$. Henceforth, the estimation of the contemporaneous impact matrix is entirely analogous to long-run identification in a standard VAR setting, that is, we recover $\Phi_\epsilon(1)$ by a Cholesky decomposition and then obtain an estimate of D.

[4]There are two auxiliary parameters in the subspace algorithm, f, $\mathfrak{p}$, which determine the row and column dimension of a Hankel matrix which is estimated in an intermediate step (see Bauer (2005) and the appendix). They have been set to $f = \mathfrak{p} = 8$. These parameters are of no importance asymptotically as long as they increase at certain rates with the sample size. In the literature it has been suggested to set $f = \mathfrak{p} = 2\hat{p}$ where $\hat{p}$ is the order of the chosen autoregressive approximation (Bauer 2005).

Vector Autoregressive Moving Average Models: The VARMA representation given in (3) implies that we can represent $\mathbf{y}_t$ in terms of the innovations as

$$\mathbf{y}_t = (I - A_1 L)^{-1}(I + M_1 L)u_t = A(L)^{-1}M(L)u_t,$$

where $A(L)$ and $M(L)$ are the autoregressive polynomial and the moving average polynomial, respectively and the intercept term has been omitted. The long-run impact of u_t is given by $\Phi_u(1) = A(1)^{-1}M(1)$ and D can be recovered as before. The representation in (3) is however not the most useful representation in practice, because it is an unrestricted VARMA(1,1) and, furthermore, the matrix multiplying u_{t-1} is singular. It is more useful to choose a specific representation which guarantees that all parameters are identified. For an introduction to the identification problem in VARMA models see Lütkepohl (2005). Here we employ a final moving average (FMA) representation that can be derived analogously to the final equation form (see Dufour and Pelletier 2004). In our case, this results in a VARMA$_{FMA}$(2,1) representation in final moving average form (see appendix).[5]

As in the case of state space models there are many different estimation methods for VARMA models. Examples are the methods developed by Hannan and Rissanen (1982), Koreisha and Pukkila (1990b), Mauricio (1995) or Kapetanios (2003). We report the results for the estimation algorithm

[5]We also experimented with other identified representations such as the final equation representation or the echelon representation. However, the final moving average representation was found to yield the best results.

described by Hannan and Kavalieris (1984*b*). Though it is regression-based, it is a Gauss-Newton procedure for the maximization of the likelihood, conditional on initial values. First, a high-order VAR is fitted to get initial estimates of the innovations. In the second stage these estimates are used to estimate the autoregressive and moving average parameters by least squares. In the third stage the estimated coefficients are used to form new residuals and the coefficient estimates from the second stage are refined (see, e.g., Hannan and Kavalieris (1984*b*), Hannan and Deistler (1988) or Dufour and Pelletier (2004)). We use a VAR with lag length $n_T = 0.5\ T^{1/2}$ for the initial long autoregression.[6]

4.2 Results of the Monte Carlo Study

Table C.2 summarizes the results of the Monte Carlo simulation study. We tabulate Monte Carlo means and standard deviations of the estimates of the contemporaneous impact of a technology shock on productivity and hours worked for the various estimators. We also tabulate the MSE of the different estimators relative to the MSE of the estimator resulting from the benchmark VAR. Figures 1 and 2 depict the average estimated impulse responses of hours worked, together with the true impulse responses for the VAR and the CCA subspace algorithm. In the figures, the bands around the mean lines correspond to the 0.025% and 0.975% quantiles of the estimated impulse responses at each point of time.

[6]We also tried other estimation algorithms. More specifically, we employed full information maximum likelihood maximization as, for example, in Mauricio (1995). However, this procedure was found to be highly non-stable. The algorithm was therefore not considered to be a practical alternative. One reason for these problems is that the roots of the AR and the MA polynomials are all close to the unit circle.

Our SVAR results confirm the findings of both Christiano et al. (2006) and Chari et al. (2005). While the SVAR is unbiased for the KP-specification, the same is not true for the CKM-specification. The associated pictures for both parametrizations show that the 95% bands around the mean impulse responses comprise a very large region ranging from negative values to very high positive values. Also, for the different variations of the benchmark model we find that the SVAR is often severely biased and/or displays high variability.

The PEM routine performs uniformly worse for all sets of parameter values. Although, in contrast to the SVAR, the state space model nests the DSGE model, the small sample performance of this estimation algorithm is worse. The poor accuracy of the PEM routine can be attributed to the near non-stationarity and non-invertibility of the series that cause difficulties for the non-linear optimization. Also local optima are a concern. These problems are not new. The results of the PEM routine illustrate that a formally exact representation of the DGP is not necessarily the most efficient representation for estimation because the associated estimation algorithm may be numerically unreliable or not robust to the near violation of the underlying assumptions.

The results for the CCA algorithm support this claim. We find that the associated MSE of the estimated first-period impulse response is almost uniformly lower for both series and across different specifications. Only in two cases the MSE of the CCA-based estimates exceeds the MSE of the SVAR, and only by a very small amount. In particular, the first-period impact on hours worked is estimated more precisely up to a relative reduction

to 85% in terms of MSE for the KP-specification. Figure 1 shows that the 95% interval is slightly narrower for the estimated state space model, but still very wide. In almost all cases the bias is at least slightly reduced. Generally, the response of hours worked is estimated more precisely, but the performances of the SVAR and the state space model seem to be related. That is, in cases where the VAR fails dramatically, the state space model does so too, although it might still perform relatively better. The advantage of the CCA algorithm over the VAR-based least squares algorithm should be due to the use of the more general state space representation.

The VARMA models perform generally equivalent or worse than a simple VAR approximation. The Hannan-Kavalieris method may give some improvement, but is plagued with numerical difficulties. In fact, the improvements in terms of bias are compromised by increases in variance and a higher MSE. The reason is that we face an ill-conditioned problem that cannot be remedied by rescaling the data because of the presence of eigenvalues close to the unit circle of both the autoregressive and moving average parts. Furthermore, the roots of the moving average part and the autoregressive part imply that the model is close to being not identified. This in turn implies a near singular information matrix of the parameters. In comparison to the SVAR model, the VARMAs perform relatively well in estimating the impact on hours worked, but worse in estimating the response of productivity to a technology shock. While the VARMA model fully nests the underlying DGP, this representation is not very efficient in our context.

A problem common to all algorithms is that the stationarity requirement is nearly violated for the DGPs at hand. As we have seen in section 4.1,

the stationarity assumption lies at the heart of the long-run identification scheme. However, as the eigenvalues in Table C.2 indicate, this assumption is nearly violated for the benchmark models. This problem is independent of the chosen representation and, therefore, does not vanish even when we control for omitted moving average terms. Apart from problems specific to the algorithms, this common problem may explain the poor performance of all algorithms - a problem that could be overcome in larger samples.[7]

5 Conclusions

There has been some debate whether SVARs can actually discriminate between competing DSGE models and whether their sampling properties are good enough to justify their popularity in applied macroeconomics. In particular, several Monte Carlo studies indicate that SVARs based on long-run restrictions could be heavily biased and very imprecise. Some authors suggest that SVARs may fail because they are only approximate representations of the underlying DGPs. Therefore, we replicate the simulation experiments of Chari et al. (2005) and Christiano et al. (2006) and apply more general models to their artificially generated data. In particular, we use algorithms based on SVAR, VARMA and state space models and compare the resulting estimates of the underlying structural model. For our simulations, we found that one can perform only marginally better by taking the full structure of the DGP into account. While our VARMA-based conditional ML algo-

[7] Our estimation results are in line with McGrattan's (2006) findings. However, McGrattan (2006) stresses in her context that unrestricted state space and VARMA models impose too few restrictions to uncover the structural shocks.

rithm and the prediction error algorithm for state space models were found to perform not significantly better and often even worse, the CCA subspace algorithm seems to be at least equivalent to a VAR. Indeed, in several cases the CCA algorithm performed better than the corresponding SVAR. However, the obtained estimates of the structural models display high variability and are biased, regardless of the reduced form model. Furthermore, the performances of the different estimators are strongly correlated. This finding suggests, that SVARs do not fail because they are simple finite-order approximations. Given the properties of the data series, the failure of long-run identification seems almost entirely a small sample problem in these type of simulation studies.

References

Bauer, D. (2005), 'Estimating Linear Dynamical Systems Using Subspace Methods', *Econometric Theory* **21**, 181–211.

Blanchard, O. and Quah, D. (1989), 'The dynamic effects of aggregate demand and supply disturbances', *American Economic Review* **79**(4), 655–673.

Chari, V. V., Kehoe, P. J. and McGrattan, E. R. (2005), 'A Critique of Structural VARs Using Real Business Cycle Theory'. Federal Reserve Bank of Minneapolis, Working Paper 631.

Christiano, L. and Eichenbaum, M. (1992), Identification and the liquidity effect of a monetary policy shock, *in* A. Cukierman, Z. Hercowitz and L. Leiderman, eds, 'Political Economy, Growth and Business Cycles', MIT Press, Cambridge and London, pp. 335–370.

Christiano, L. J., Eichenbaum, M. and Vigfusson, R. (2006), 'Assessing Structural VARs'. Working Paper, March 2006.

Cooley, T. F. and Dwyer, M. (1998), 'Business Cycle Analysis Without Much Theory. A Look at Structural VARs', *Journal of Econometrics* **83**, 57–88.

Dufour, J.-M. and Pelletier, D. (2004), 'Linear Estimation of Weak VARMA Models With a Macroeconomic Application'. Université de Montréal and North Carolina State University, Working Paper.

Durbin, J. (1960), 'The Fitting of Time Series Models', *International Statistical Review* **28**, 233–244.

Erceg, C. J., Guerrieri, L. and Gust, C. (2005), 'Can Long-Run Restrictions Identify Technology Shocks?', (792), March 2005. International Financial Discussion Paper.

Fernández Villaverde, J., Rubio Ramírez, J. and Sargent, T. J. (2005), 'A,B,C's (and D)'s For Understanding VARs', (308), May. NBER Technical Working Paper.

Gali, J. (1999), 'Technology, Employment and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?', *American Economic Review* **89**(1), 249–271.

Hannan, E. J. and Deistler, M. (1988), *The Statistical Theory of Linear Systems*, John Wiley, New York.

Hannan, E. J. and Kavalieris, L. (1984*a*), 'A Method for Autoregressive-Moving Average Estimation', *Biometrika* **71**(2), 273–280.

Hannan, E. J. and Kavalieris, L. (1984*b*), 'Multivariate Linear Time Series Models', *Advances in Applied Probability* **16**(3), 492–561.

Hannan, E. J. and Rissanen, J. (1982), 'Recursive Estimation of Mixed ARMA Order', *Biometrika* **69**, 81–94.

Kapetanios, G. (2003), 'A Note on the Iterative Least-Squares Estimation Method for ARMA and VARMA Models', *Economics Letters* **79**, 305–312.

King, R., Plosser, C., Stock, J. and Watson, M. (1991), 'Stochastic trends and economic fluctuations', *American Economic Review* **81**(4), 819–840.

Klein, A., Melard, G. and Spreij, P. (2004), 'On the resultant property of the fisher information matrix of a vector arma process.'. Disscussion Paper 2004/13.

Koreisha, S. and Pukkila, T. (1990b), 'A GLS approach for estimation of ARMA models', *Journal of Time Series Analysis* **11**(2), 139–151.

Larimore, W. E. (1983), 'System Identification, Reduced-Order Filters and Modeling via Canonical Variate Analysis', *In: Proc. 1983 Amer. Control Conference 2 (H.S. Rao and P. Dorato, Eds.)* pp. 445–451.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Mauricio, J. A. (1995), 'Exact Maximum Likelihood Estimation of Stationary Vector ARMA Models', *Journal of the American Statistical Association* **90**, 282–291.

McGrattan, E. R. (2006), 'Measurement with Minimal Theory'. Federal Reserve Bank of Minneapolis, Working Paper 643.

Prescott, E. (1986), 'Theory ahead of business cycle measurement', pp. 9–22. Quarterly Review, Federal Reserve Bank of Minneapolis, issue Fall.

van Overschee, P. and DeMoor, B. (1994), 'N4sid: Subspace algorithms for the identification of combined deterministic-stochastic processes', *Automatica* **30**, 75–93.

A Final MA Equation Form

Consider a standard representation for a stationary and invertible VARMA process

$$A(L)\mathbf{y}_t \quad = \quad M(L)u_t.$$

Recall that $M^{-1}(L) = M^*(L)/|M(L)|$, where $M^*(L)$ denotes the adjoint of $M(L)$ and $|M(L)|$ its determinant. We can multiply the above equation with $M^*(L)$ to get

$$M^*(L)A(L)\mathbf{y}_t \quad = \quad |M(L)|u_t.$$

This representation therefore places restrictions on the moving average polynomial which is required to be a scalar operator, $|M(L)|$. Dufour and Pelletier (2004) show that this restriction leads to an identified representation. More specifically, consider the VARMA(1,1) representation in (3). Since the moving average part is not of full rank we can write the system as

$$\begin{bmatrix} 1-a_{11}L & -a_{12}L \\ -a_{21}L & 1-a_{22}L \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} 1+m_{11}L & \alpha m_{11}L \\ m_{21}L & 1+\alpha m_{21}L \end{bmatrix} u_t,$$

where α is some constant not equal to zero.
Clearly, $\det(M(L)) = 1 + (m_{11} + \alpha m_{21})L$ and we can write

$$\begin{bmatrix} 1+\alpha m_{21}L & -\alpha m_{11}L \\ -m_{21}L & 1+\alpha m_{11}L \end{bmatrix} \begin{bmatrix} 1-a_{11}L & -a_{12}L \\ -a_{21}L & 1-a_{22}L \end{bmatrix} \mathbf{y}_t = [1 + (m_{11} + \alpha m_{21})L]u_t.$$

Because of the reduced rank we end up with a $\text{VARMA}_{FMA}(2,1)$. Note that the moving average part is indeed restricted to be a scalar operator.

B Statistical Representations

This section elaborates on the derivation of the infinite VAR, VARMA and state space representations that result from our DSGE model in order to get an insight into the relationship between the economic model and the implied time series properties.

Consider again the law of motion of the logs

$$\begin{aligned}
\log k_{t+1} &= \phi_1 + \phi_{11} \log k_t - \phi_{11} \log x_t + \phi_{12} \tau_t, \\
\log y_t - \log L_t &= \phi_2 + \phi_{21} \log k_t - \phi_{21} \log x_t + \phi_{22} \tau_t, \\
\log L_t &= \phi_3 + \phi_{31} \log k_t - \phi_{31} \log x_t + \phi_{32} \tau_t,
\end{aligned}$$

and the exogenous states

$$\begin{aligned}
\log x_{t+1} &= \mu + \sigma_x \epsilon_{x,t+1}, \\
\tau_{t+1} &= (1-\rho)\bar{\tau}_l + \rho \tau_t + \sigma_l \epsilon_{l,t+1}.
\end{aligned}$$

From these equations the state space representation can derived as follows. First write down the law of motion of labor productivity in differences:

$$\Delta \log(Y_t/L_t) = \log x_t + \phi_{21} \Delta \log k_t - \phi_{21} \Delta \log x_t + \phi_{22} \Delta \tau_t.$$

Thus the observed series can be expressed as

$$\begin{aligned}\Delta \log(Y_t/L_t) &= \phi_{21} \log k_t - \phi_{21} \log k_{t-1} + (1-\phi_{21}) \log x_t \\ &\quad +\phi_{21} \log x_{t-1} + \phi_{22}\tau_t - \phi_{22}\tau_{t-1}, \\ \log L_t &= \phi_3 + \phi_{31} \log k_t - \phi_{31} \log x_t + \phi_{32}\tau_t.\end{aligned}$$

Next, rewrite the law of motion for capital as

$$\log k_{t-1} = -\phi_{11}^{-1}\phi_1 + \phi_{11}^{-1} \log k_t + \log x_{t-1} - \phi_{11}^{-1}\phi_{12}\tau_{t-1},$$

in order to substitute for capital at time $t-1$:

$$\begin{aligned}\Delta \log(Y_t/L_t) &= \phi_{21}\phi_{11}^{-1}\phi_1 + \phi_{21}(1-\phi_{11}^{-1}) \log k_t \\ &\quad +(1-\phi_{21}) \log x_t + \phi_{22}\tau_t + (\phi_{21}\phi_{11}^{-1}\phi_{12} - \phi_{22})\tau_{t-1}.\end{aligned}$$

Using the laws of motion for the stochastic shock processes, substitute the current exogenous shocks to get

$$\begin{aligned}\Delta \log(Y_t/L_t) &= \left[\phi_{21}\phi_{11}^{-1}\phi_1 + (1-\phi_{21})\mu + \phi_{22}(1-\rho)\bar{\tau}_l\right] + \phi_{21}(1-\phi_{11}^{-1}) \log k_t \\ &\quad +(\phi_{21}\phi_{11}^{-1}\phi_{12} - (1-\rho)\phi_{22})\tau_{t-1} + (1-\phi_{21})\sigma_x\epsilon_{x,t} + \phi_{22}\sigma_l\epsilon_{l,t}, \\ \log L_t &= [\phi_3 - \phi_{31}\mu + \phi_{32}(1-\rho)\bar{\tau}_l] + \phi_{31} \log k_t + \phi_{32}\rho\tau_{t-1} \\ &\quad -\phi_{31}\sigma_x\epsilon_{x,t} + \phi_{32}\sigma_l\epsilon_{l,t}.\end{aligned}$$

Next, consider the law of motion for capital and express future capital in

terms of the current states as

$$\begin{aligned}\log k_{t+1} &= [\phi_1 - \phi_{11}\mu + \phi_{12}(1-\rho)\bar{\tau}_l] + \phi_{11}\log k_t + \phi_{12}\rho\tau_{t-1} \\ &\quad -\phi_{11}\sigma_x\epsilon_{x,t} + \phi_{12}\sigma_l\epsilon_{l,t}.\end{aligned}$$

Collecting the above equations, the system can be written in state space form according to Fernández Villaverde et al. (2005). The state transition equation is

$$\begin{bmatrix} \log k_{t+1} \\ \tau_t \end{bmatrix} = K_1 + A \begin{bmatrix} \log k_t \\ \tau_{t-1} \end{bmatrix} + B \begin{bmatrix} \epsilon_{x,t} \\ \epsilon_{lt} \end{bmatrix},$$

where the system matrices are given by

$$K_1 = \begin{bmatrix} \phi_1 - \phi_{11}\mu + \phi_{12}(1-\rho)\bar{\tau}_l \\ (1-\rho)\bar{\tau} \end{bmatrix},$$

$$A = \begin{bmatrix} \phi_{11} & \phi_{12}\rho \\ 0 & \rho \end{bmatrix},$$

and

$$B = \begin{bmatrix} -\phi_{11}\sigma_x & \phi_{12}\sigma_l \\ 0 & \sigma_l \end{bmatrix}.$$

The observation equation is

$$\begin{bmatrix} \Delta\log(Y_t/L_t) \\ \log L_t \end{bmatrix} = K_2 + C \begin{bmatrix} \log k_t \\ \tau_{t-1} \end{bmatrix} + D \begin{bmatrix} \epsilon_{x,t} \\ \epsilon_{lt} \end{bmatrix},$$

with system matrices

$$K_2 = \begin{bmatrix} \phi_{21}\phi_{11}^{-1}\phi_1 + (1-\phi_{21})\mu + \phi_{22}(1-\rho)\bar{\tau}_l \\ \phi_3 - \phi_{31}\mu + \phi_{32}(1-\rho)\bar{\tau}_l \end{bmatrix},$$

$$C = \begin{bmatrix} \phi_{21}(1-\phi_{11}^{-1}) & \phi_{21}\phi_{11}^{-1}\phi_{12} - (1-\rho)\phi_{22} \\ \phi_{31} & \phi_{32}\rho \end{bmatrix},$$

and

$$D = \begin{bmatrix} (1-\phi_{21})\sigma_x & \phi_{22}\sigma_l \\ -\phi_{31}\sigma_x & \phi_{32}\sigma_l \end{bmatrix}.$$

This representation permits us to derive the infinite VAR and VARMA representation in compact form.

Let $\mathbf{y}_t$ denote the vector of observables, $\mathbf{x}_t$ the vector of states, and ϵ the white noise shocks. Then we have as above

$$\begin{aligned} \mathbf{x}_{t+1} &= K_1 + A\mathbf{x}_t + B\epsilon_t, \\ \mathbf{y}_t &= K_2 + C\mathbf{x}_t + D\epsilon_t. \end{aligned}$$

If D is invertible, it is possible to use

$$\epsilon_t = D^{-1}\left(\mathbf{y}_t - K_2 - C\mathbf{x}_t\right)$$

in the transition equation to obtain

$$\begin{aligned} \mathbf{x}_{t+1} &= K_1 + A\mathbf{x}_t + BD^{-1}(\mathbf{y}_t - K_2 - C\mathbf{x}_t), \\ (I - (A - BD^{-1}C)L)\mathbf{x}_{t+1} &= [K_1 - BD^{-1}K_2] + BD^{-1}\mathbf{y}_t. \end{aligned}$$

If the eigenvalues of $(A - BD^{-1}C)$ are strictly less than one in modulus we can solve for $\mathbf{x}_{t+1}$:

$$\mathbf{x}_{t+1} = \left(I - (A - BD^{-1}C)L\right)^{-1} \left([K_1 - BD^{-1}K_2] + BD^{-1}\mathbf{y}_t\right).$$

Using this relation in the observation equation yields the infinite VAR representation for $\mathbf{y}_t$:

$$\mathbf{y}_t = K_2 + C\left(I - (A - BD^{-1}C)L\right)^{-1} \left([K_1 - BD^{-1}K_2] + BD^{-1}\mathbf{y}_{t-1}\right) + D\epsilon_t,$$

$$\mathbf{y}_t = K_3 + C\left(I - (A - BD^{-1}C)L\right)^{-1} BD^{-1}\mathbf{y}_{t-1} + D\epsilon_t.$$

Note that the condition for the existence of an infinite VAR-representation is that $I - (A - BD^{-1}C)$ is invertible. If this condition does not hold, impulse responses from a VAR are unlikely to match up those from the model.

If C is invertible, it is possible to rewrite the state as

$$\mathbf{x}_t = C^{-1}\left(\mathbf{y}_t - K_2 - D\epsilon_t\right)$$

and use it in the observation equation:

$$C^{-1}\left(\mathbf{y}_{t+1} - K_2 - D\epsilon_{t+1}\right) = K_1 + AC^{-1}\left(\mathbf{y}_t - K_2 - D\epsilon_t\right) + B\epsilon_t,$$

$$\mathbf{y}_{t+1} - CAC^{-1}\mathbf{y}_t = CK_1 + K_2 - CAC^{-1}K_2 + (CB - CAC^{-1}D)\epsilon_t + D\epsilon_{t+1}.$$

Therefore, we obtain a VARMA(1,1) representation of $\mathbf{y}_t$:

$$\mathbf{y}_t = K_4 + CAC^{-1}\mathbf{y}_{t-1} + \left(I + (CBD^{-1} - CAC^{-1})L\right) u_t.$$

with $u_t \sim N(0, DD')$.

C Estimation Algorithms

C.1 Hannan-Kavalieris Method

This method goes originally back to Durbin (1960) and has been introduced by Hannan and Kavalieris (1984*a*) for multivariate processes.[8] It is a Gauss-Newton procedure to maximize the likelihood function conditional on $y_t = 0,\ u_t = 0$ for $t \leq 0$, but its first iteration has been sometimes interpreted as a three-stage least squares procedure (Dufour and Pelletier (2004)). The method is computationally very easy to implement because of its recursive nature.

We discuss this method in the framework of a standard VARMA (p, q) representation

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t + M_1 u_{t-1} + \ldots M_q u_{t-q}.$$

Usually additional restrictions need to be imposed on the coefficient matrices to ensure identification of the parameters. For this purpose, we use the

[8] See also Hannan and Deistler (1988), sections 6.5, 6.7, for an extensive discussion.

following notation. Let $\beta = \text{vec}[A_1,\ldots, A_p, M_1, \ldots M_q]$ denote the vector of all parameters. The vector of free parameters, γ, can be defined by introducing restriction matrices R and r such that the vectors are related by $\beta = R\gamma + r$.

Given invertibility of the process, there exists an infinite VAR representation $y_t = \sum_{i=1}^{\infty} \Pi_i y_{t-i} + u_t$. In the first step of the algorithm, this representation is approximated by a "long" VAR to get an estimate of the residuals. More precisely, the following regression equation is used

$$y_t = \sum_{i}^{n_T} \Pi_i y_{t-i} + u_t,$$

where n_T is large and goes to infinity as the sample size grows. Given an estimate of the residuals, $\hat{u}_t$, we might obtain starting values for future iterations by performing a (restricted) regression in

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t + M_1 \hat{u}_{t-1} + \ldots M_q \hat{u}_{t-q}.$$

Denote the estimated coefficient matrices by $\tilde{A}_1, \tilde{A}_2, \ldots$ and $\tilde{M}_1, \tilde{M}_2, \ldots$. The first iteration of the conditional maximum likelihood algorithm can be expressed in a simple regression framework. One forms new residuals, ε_t,

and new matrices, ξ_t, η_t and $\hat{X}_t$, according to

$$\begin{aligned}
\varepsilon_t &= y_t - \sum_{j=1}^{p} \tilde{A}_j y_{t-j} - \sum_{j=1}^{q} \tilde{M}_j \varepsilon_{t-j}, \\
\xi_t &= -\sum_{j=1}^{q} \tilde{M}_j \xi_{t-j} + \varepsilon_t, \\
\eta_t &= -\sum_{j=1}^{q} \tilde{M}_j \eta_{t-j} + y_t, \\
\hat{X}_t &= -\sum_{j=1}^{q} \tilde{M}_j \hat{X}_{t-j} + (Y_t' \otimes I_K) R,
\end{aligned}$$

for $t = 1, 2, \ldots, T$, $Y_t = [y_t', \ldots, y_{t-p+1}', \hat{u}_t', \ldots, \hat{u}_{t-q+1}']$ and $y_t = \varepsilon_t = \xi_t = \eta_t = 0$ and $\hat{X}_t = 0$ for $t \leq 0$. The final estimate is

$$\hat{\gamma} = \left(\sum_{m+1}^{T} \hat{X}_{t-1}' \tilde{\Sigma}_t^{-1} \hat{X}_{t-1} \right)^{-1} \left(\sum_{m+1}^{T} \hat{X}_{t-1} \tilde{\Sigma}^{-1} (\varepsilon_t + \eta_t - \xi_t) \right),$$

where $\tilde{\Sigma} = T^{-1} \sum \varepsilon_t \varepsilon_t'$, $m = \max\{p, q\}$. This procedure is asymptotically efficient under certain conditions (see Lütkepohl (2005)).

C.2 Subspace Algorithms

Subspace algorithms rely on the state space representation of a linear system. The CCA algorithm is originally due to Larimore (1983). The basic idea behind subspace algorithms lies in the fact that if we knew the unobserved state, x_t, we could estimate the *system matrices*, A, K, C, by linear

regressions as can be seen from the basic equations

$$\begin{aligned} x_{t+1} &= Ax_t + Ku_t, \\ y_t &= Cx_t + u_t. \end{aligned}$$

Given the state and the observations, $\hat{C}$ and $\hat{u}_t$ are obtained by a regression of y_t on x_t and $\hat{A}$ and $\hat{K}$ are obtained by a regression of x_{t+1} on x_t and $\hat{u}_t$. Therefore, the problem is to obtain in a first step an estimate of the n-dimensional state, $\hat{x}_t$. This is analogous to the idea of a long autoregression in VARMA models that estimates the unobserved residuals in a first step which is followed by a least squares regression. For this purpose, one expresses the state as a function of the history of y_t and an initial state as shown below. Solving the state space equations for positive integers j and $\mathfrak{p}$, we get

$$y_{t+j} = CA^j x_t + \sum_{i=0}^{j-1} CA^i K u_{t+j-i-1} + u_{t+j}, \tag{4}$$

$$\begin{aligned} x_t &= (A-KC)^{\mathfrak{p}} x_{t-\mathfrak{p}} + \sum_{i=0}^{\mathfrak{p}-1} (A-KC)^i K y_{t-i-1} \\ &= (A-KC)^{\mathfrak{p}} x_{t-\mathfrak{p}} + \mathcal{K}_{\mathfrak{p}} Y_{t,\mathfrak{p}}^{-}, \end{aligned} \tag{5}$$

where $\mathcal{K}_{\mathfrak{p}} = [K, (A-KC)K, \ldots, (A-KC)^{\mathfrak{p}-1}K]$ and $Y_{t,\mathfrak{p}}^{-} = [y'_{t-1}, \ldots, y'_{t-p}]'$. The first equation states that the best predictor of y_{t+j} is a function of the state only. The state summarizes in this sense all available information in the past up to time t. The second equation states that the state can be recovered as a combination of an initial state and past observations. Define

$Y_{t,f}^{+} = [y_t', \ldots, y_{t+f-1}']'$ for some integer $f > 0$ and formulate equation (4) for all observations contained in $Y_{t,f}^{+}$ simultaneously. Combine these equations with (5) in order to obtain

$$Y_{t,f}^{+} \quad = \quad \mathcal{O}_f \mathcal{K}_{\mathfrak{p}} Y_{t,\mathfrak{p}}^{-} + \mathcal{O}_f (A - BC)^{\mathfrak{p}} x_{t-\mathfrak{p}} + \mathcal{E}_f E_{t,f}^{+},$$

where $\mathcal{O}_f = [C', A'C', \ldots, (A^{f-1})'C']'$, $E_{t,f}^{+} = [u_t', \ldots, u_{t+f-1}']'$ and $\mathcal{E}_f$ is a function of the system matrices. The above equation is central for most subspace algorithms. Note that if the maximum eigenvalue of $(A - BC)$ is less than one in absolute value, we have $(A - BC)^{\mathfrak{p}} \approx 0$ for large $\mathfrak{p}$. This condition is satisfied for stationary and invertible processes. This reasoning motivates an approximation of the above equation given by

$$Y_{t,f}^{+} \quad = \quad \beta Y_{t,\mathfrak{p}}^{-} + N_{t,f}^{+}, \tag{6}$$

where $\beta = \mathcal{O}_f \mathcal{K}_{\mathfrak{p}}$ and $N_{t,f}^{+}$ is defined by the equation. Most popular subspace algorithms use this equation to obtain an estimate of β that is decomposed into $\mathcal{O}_f$ and $\mathcal{K}_{\mathfrak{p}}$. The identification problem is solved implicitly during this step.

To be precise, the employed algorithm consists of the following steps for given integers $n, \mathfrak{p}, f$:

1. Perform OLS on (6) using the avaible data to get an estimate $\hat{\beta}_{f,\mathfrak{p}}$.

2. Compute

$$\hat{\Gamma}_f^{+} = \frac{1}{T_{f,\mathfrak{p}}} \sum_{t=\mathfrak{p}+1}^{T-f+1} Y_{t,f}^{+} (Y_{t,f}^{+})' \ , \ \hat{\Gamma}_{\mathfrak{p}}^{-} = \frac{1}{T_{f,\mathfrak{p}}} \sum_{t=\mathfrak{p}+1}^{T-f+1} Y_{t,\mathfrak{p}}^{-} (Y_{t,\mathfrak{p}}^{-})',$$

where $T_{f,\mathfrak{p}} = T - f - p + 1$.

3. Given the dimension of the state, n, perform a singular value decomposition

$$(\hat{\Gamma}_f^+)^{-1/2}\hat{\beta}_{f,\mathfrak{p}}(\hat{\Gamma}_{\mathfrak{p}}^-)^{1/2} \quad = \quad \hat{U}\hat{\Sigma}\hat{V}'.$$

Since $\hat{\beta}_{f,\mathfrak{p}}$ is of reduced rank one considers only the first n singular values of $\hat{\Sigma}$ and the corresponding vectors in $\hat{U}$ and $\hat{V}$. That is, one uses an approximation, $(\hat{\Gamma}_f^+)^{-1/2}\hat{\beta}_{f,\mathfrak{p}}(\hat{\Gamma}_{\mathfrak{p}}^-)^{1/2} = \hat{U}_n\hat{\Sigma}_n\hat{V}_n' + \hat{R}_n$. The reduced rank matrices are obtained as

$$\hat{\mathcal{O}}_f\hat{\mathcal{K}}_{\mathfrak{p}} \quad = \quad [(\hat{W}_f^+)^{-1}\hat{U}_n\hat{\Sigma}_n^{1/2}][\hat{\Sigma}_n^{1/2}\hat{V}_n'(\hat{W}_{\mathfrak{p}}^-)^{-1}].$$

4. Estimate the state as $\hat{x}_t = \hat{\mathcal{K}}_{\mathfrak{p}}Y_{t,\mathfrak{p}}^-$ and estimate the system matrices using linear regressions as described above.

Although the algorithm looks quite complicated at first sight, it is actually very simple and is regarded to lead to numerically stable and accurate estimates. There are certain parameters which have to be determined, namely the dimension of the state and the integers f and $\mathfrak{p}$. For the asymptotic consequences of various choices see Bauer (2005).

Table 1: Benchmark calibrations and resulting time series properties of the RBC model.

Parameters	Common	CKM Benchmark	KP Benchmark
α	0.33		
β	$0.98^{1/4}$		
σ	1		
δ	$1-(1-0.6)^{1/4}$		
ψ	2.5		
γ	$1.01^{1/4}-1$		
μ	0.00516		
$\bar{L}$	1		
$\bar{\tau}_l$	0.243		
τ_x	0.3		
ρ		0.94	0.993
σ_τ		0.008	0.0066
σ_x		0.00568	0.011738
Selected time series properties			
eig(A_1)		0.9573, 0.9400	0.9573, 0.9930
eig(M_1)		-0.9557, 0	-0.9505, 0
eig(K)		$-1.7779 \pm 0.51i$	$-2.0298 \pm 0.35i$

Table 2: Simulation Results

		VAR (4)			PEM (2)			SS (2,8,8)			VARMA$_{FMA}$ (2,1)		
Variable	True Value	Mean	Std.	MSE	Mean	Std.	MSE	Mean	Std.	MSE	Mean	Std.	MSE
KP Benchmark													
Prod.	0.69	0.55	0.19	1	0.53	0.23	1.44	0.57	0.18	0.86	0.54	0.18	1.02
Hours	0.28	0.31	0.43	1	0.32	0.50	1.33	0.31	0.40	0.85	0.33	0.42	0.92
KP, $\sigma = 0$ (Indivisible Labor)													
Prod.	0.65	0.48	0.23	1	0.45	0.28	1.49	0.50	0.22	0.92	0.47	0.22	1.02
Hours	0.43	0.56	0.56	1	0.58	0.67	1.44	0.52	0.54	0.92	0.57	0.54	0.93
KP, $\sigma = 6$ (Frisch elasticity=0.63)													
Prod.	0.75	0.61	0.16	1	0.57	0.24	2.03	0.63	0.14	0.80	0.59	0.16	1.10
Hours	0.11	0.10	0.19	1	0.10	0.24	1.52	0.10	0.18	0.89	0.10	0.19	0.98
CKM Benchmark													
Prod.	0.34	0.10	0.17	1	0.13	0.20	1.03	0.11	0.18	1.00	0.09	0.16	1.07
Hours	0.14	0.65	0.39	1	0.61	0.53	1.21	0.62	0.40	0.95	0.67	0.37	1.02
CKM, $\sigma = 0$ (Indivisible Labor)													
Prod.	0.31	-0.12	0.21	1	-0.07	0.30	1.02	-0.12	0.23	1.05	-0.15	0.19	1.09
Hours	0.21	1.26	0.49	1	1.16	0.84	1.20	1.24	0.54	1.01	1.31	0.45	1.06
CKM, $\sigma = 6$ (Frisch elasticity=0.63)													
Prod.	0.36	0.30	0.08	1	0.29	0.11	1.67	0.31	0.08	0.95	0.29	0.08	1.04
Hours	0.05	0.12	0.17	1	0.12	0.20	1.35	0.11	0.17	0.92	0.13	0.17	0.97
CKM, $\sigma_l/2$													
Prod.	0.34	0.25	0.10	1	0.25	0.12	1.45	0.26	0.10	0.93	0.25	0.09	1.06
Hours	0.14	0.26	0.22	1	0.26	0.26	1.35	0.24	0.21	0.88	0.27	0.21	0.97
CKM, $\sigma_l/3$													
Prod.	0.34	0.28	0.07	1	0.28	0.09	1.63	0.29	0.07	0.89	0.28	0.07	1.08
Hours	0.14	0.18	0.15	1	0.18	0.18	1.44	0.17	0.14	0.87	0.19	0.14	0.96

Percent contemporaneous impact on productivity and hours of one standard deviation shock to technology.
The entries are Monte Carlo means and standard deviations. MSEs are relative to the MSE of the SVAR.

KP-Specification

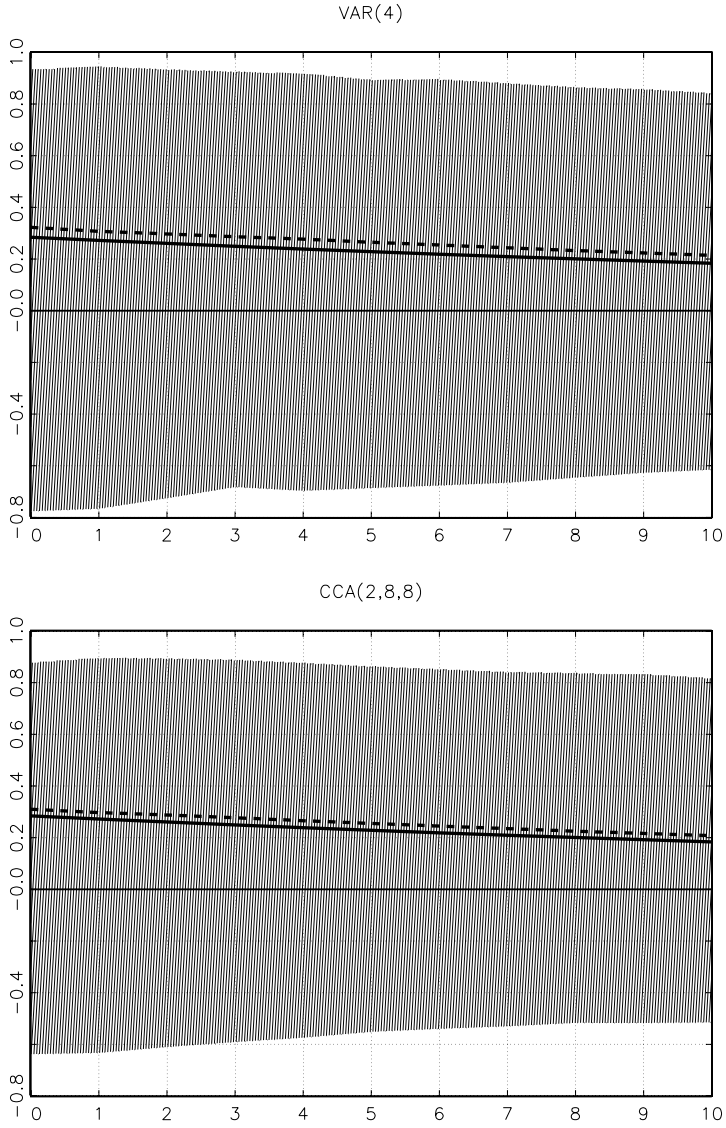


Figure 1: Mean impulse response (- -), true impulse response (–) and 95% intervals of hours worked to one standard deviation shock to technology for the VAR and the CCA subspace algorithm.

CKM-Specification

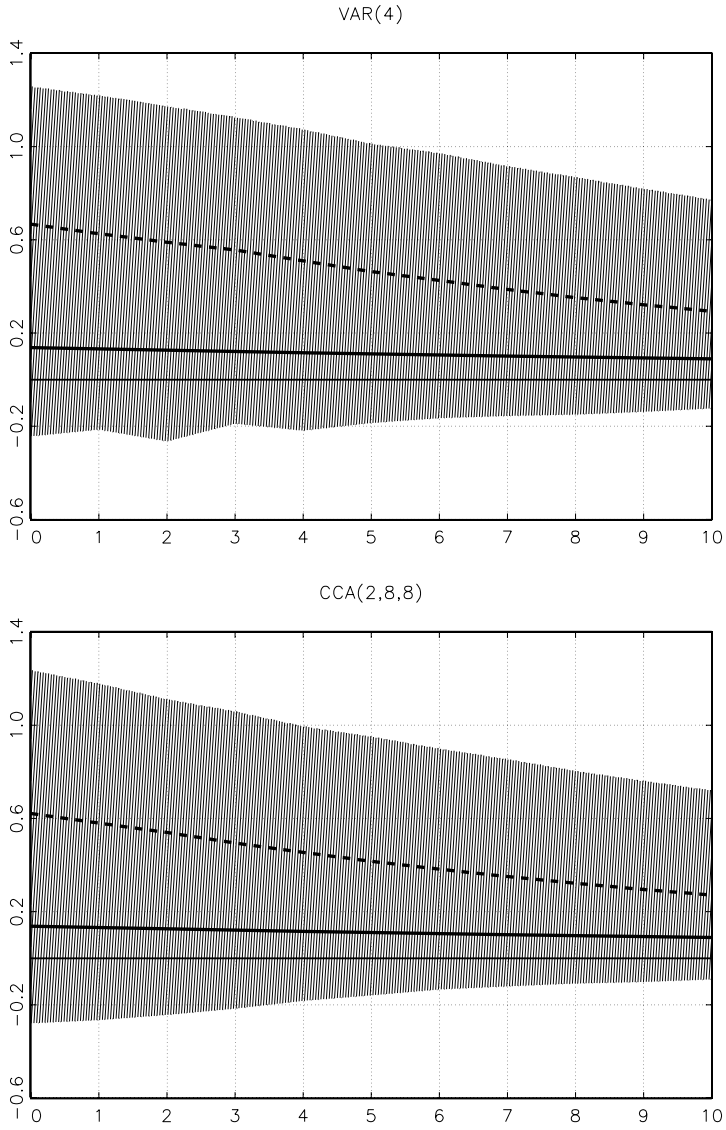


Figure 2: Mean impulse response (- -), true impulse response (–) and 95% intervals of hours worked to one standard deviation shock to technology for the VAR and the CCA subspace algorithm.