

Business Cycle Analysis and VARMA Models

Christian Kascha ^{a,*}, Karel Mertens ^b

^a*Norges Bank, Research Department*

^b*Cornell University, Economics Department*

Abstract

Can long-run identified structural vector autoregressions (SVARs) discriminate between competing models in practice? Several authors have suggested SVARs fail partly because they are finite-order approximations to infinite-order processes. We estimate vector autoregressive moving average (VARMA) and state space models, which are not misspecified, using simulated data and compare true with estimated impulse responses of hours worked to a technology shock. We find little gain from using VARMA models. However, state space algorithms can outperform SVARs. In particular, the CCA subspace method consistently yields lower mean squared errors, although even these estimates remain too imprecise for reliable inference. The qualitative differences for algorithms based on different representations are small. The comparison with estimation methods without specification error suggests that the main problem is not one of working with a VAR approximation. The properties of the processes used in the literature make identification via long-run restrictions difficult for any method.

Key words: Structural VARs, VARMA, State Space Models, Business Cycles
JEL-Codes: E32, C15, C52

1 Introduction

This paper compares structural estimation methods based on different reduced form representations of the same economic model in a simulation study similar to those undertaken by Chari, Kehoe and McGrattan (2005, 2007) and Christiano, Eichenbaum and Vigfusson (2006). Our aim is to assess different

* Corresponding author. Norges Bank, Bankplassen 2, P.O. Box 1179 Sentrum, 0107 Oslo, Norway. Tel: +47 22 31 67 19 ; Fax: +47 22 42 40 62.

Email addresses: christian.kascha@norges-bank.no (Christian Kascha), km426@cornell.edu (Karel Mertens).

algorithms' relative performance and in particular whether the inclusion of moving average terms alone leads to more precise estimates of the structural impulse responses. The fact that algorithms based on different representations yield qualitatively similar results, points out that the main problem with structural identification in this and similar simulation studies is not one of working with a specific model class.

Structural vector autoregressions are a widely used tool in empirical macroeconomics, in particular for the evaluation of dynamic stochastic general equilibrium (DSGE) models. Following Sims's (1989) suggestion, many applied researchers have used SVARs to uncover economic relationships without imposing strong theoretical assumptions. Blanchard and Quah (1989) for example use SVARs to discriminate between supply and demand shocks. King, Plosser, Stock and Watson (1991) look at the effects of permanent changes in the economy on transient economic fluctuations. Christiano and Eichenbaum (1992) investigate the monetary transmission mechanism and Cogley and Nason (1995) analyze output dynamics in real business cycle (RBC) models. The results from SVARs are often viewed as stylized facts that economic models should replicate (see, e.g., Christiano and Eichenbaum, 1999). Stock and Watson (2005) provide a useful overview of structural identification methods.

In this literature, a recent discussion has focussed on the impact of technology shocks on hours worked. In a seminal paper, Gali (1999) identifies productivity innovations using restrictions on the long-run impact matrix of the structural errors. He finds that hours worked fall in response to a positive innovation, which is contrary to the central predictions of the mainstream RBC literature. Many empirical papers have since scrutinized this finding using different data sets and identification schemes. See, for example, the contributions of Christiano, Eichenbaum and Vigfusson (2003); Francis and Ramey (2005a,b) and Gali and Rabanal (2005).

In the context of Gali's (1999) results, there is some debate whether SVARs can in practice discriminate between competing DSGE models and, more generally, whether their sampling properties are good enough to justify their popularity in applied macroeconomics. Chari et al. (2007) and Christiano et al. (2006) investigate the properties of SVARs by simulating artificial data from an RBC model and by comparing true with estimated impulse responses. In order to simulate an empirically relevant DGP, the structural parameters of the underlying RBC model are estimated from the data. According to Chari et al. (2007), long-run identified 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, 2007) is not very realistic. With their preferred parametrization,

Christiano et al. (2006) find that both long-run and short-run identification schemes display only small biases and argue that, on average, the confidence intervals produced by SVARs correctly reflect the degree of sampling uncertainty. Nevertheless, they also find that the estimates obtained via a long-run identification scheme are very imprecise. These results have been further confirmed by Erceg, Guerrieri and Gust (2005). Kehoe (2006) provides an overview of this debate. On the one hand, it is often difficult to even make a correct inference about the *sign* of the structural impulse responses with long-run restrictions, and the question is whether one should use them at all. On the other hand, long-run identification is appealing from a theoretical point of view, since it is usually less model-specific than short-run identification (Chari et al., 2007). In any case, long-run identification constitutes an additional tool of analysis in applied macroeconomic research.

The failure of finite-order SVARs is sometimes attributed to the fact that they are only approximations to VARMA / infinite-order VAR processes or to the possibility that there does not exist a VAR representation at all. King, Plosser, and Rebelo (1988) are among the first to recognize that DSGE models imply a VARMA representation. Cooley and Dwyer (1998) give an example 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, Sargent and Watson (2007), DSGE models generally imply a state space system that has a VARMA and eventually an infinite VAR representation. Christiano et al. (2006) state that “*The specification error involved in using a finite-lag VAR is the reason that in some of our examples, the sum of VAR coefficients is difficult to estimate accurately*”. Most importantly, Chari et al. (2007) argue that a VAR is not able to capture the underlying VARMA process by showing that the *truncation bias*, which is the population bias resulting from applying a finite-order VAR, is the main source of the observed small sample bias in their simulation studies.

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, while imposing minimal theoretical assumptions. We investigate whether estimators based on these alternative representations can outperform SVARs in finite samples. This question is important for several reasons. First, it is useful to find out to what extent one can improve over SVARs by including 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. (2007) and Christiano et al. (2006): We simulate DSGE models and fit different reduced form models to recover the structural shocks using the same long-run identification strategy. As in a closely related study by McGrattan (2006), we then compare the performance of the models by focusing on the estimated contemporaneous impact to a productivity shock. We employ a variety of estimation algorithms for the VARMA and state space representations. One of the findings is that one can indeed perform better by taking the full structure of the DGP into account: While most of the algorithms for VARMA and state space representations do not perform significantly better (and sometimes worse), a subspace algorithm for state space models consistently outperforms SVARs in terms of mean squared error. Unfortunately, we also find that even these alternative estimators are highly variable and are not necessarily much more informative for discriminating between different DSGE models. After all, the qualitative differences between the algorithms are small given a particular parametrization of the DSGE model. The emphasis of many previous studies on truncation bias suggests that the problems of long-run restrictions are somewhat specific to the finite-order VAR approximation. We show that this is not the case. The bad properties of long-run identification are not confined to the finite-order VAR. Instead, we point out some properties of the simulated DGPs that make it hard to identify structural shocks for *any* method. The main problem with long-run restrictions is therefore not one of working with a specific model class.

The rest of the paper is organized as follows. In section 2 we present the RBC model used by Chari et al. (2007) 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 The Data Generating Process

The DGP for the simulations is based on a simple RBC model taken from Chari et al. (2005, 2007). 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 Galí (1999) to assess the role of technology shocks in the business cycle.

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

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

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

$$C_t + (1 + \tau_x) ((1 + \gamma)K_{t+1} - (1 - \delta)K_t) \leq (1 - \tau_{l,t})w_tL_t + r_tK_t + T_t, \quad (2)$$

for $t = 0, 1, 2, \dots$, where w_t is the wage, r_t is the rental rate of capital, T_t are lump-sum government transfers and $\tau_{l,t}$ 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 depreciation rate $\delta \in (0, 1)$, the population growth rate $\gamma > 0$ and a constant investment tax τ_x . The production technology is

$$Y_t = K_t^\alpha (X_t L_t)^{1-\alpha}, \quad (3)$$

where X_t reflects labor-augmenting technological progress and $\alpha \in (0, 1)$ is the capital income share. Competitive firms maximize $Y_t - w_tL_t - r_tK_t$. Finally, the 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

$$\log X_{t+1} = \mu + \log X_t + \sigma_x \epsilon_{x,t+1}, \quad (4a)$$

$$\tau_{l,t+1} = (1 - \rho)\bar{\tau}_l + \rho\tau_{l,t} + \sigma_l \epsilon_{l,t+1}, \quad (4b)$$

where $\epsilon_{x,t}$ and $\epsilon_{l,t}$ are independent random variables with mean zero and unit standard deviation and $\sigma_x > 0$ and $\sigma_l > 0$ are scalars. $\mu > 0$ 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. (2007) 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 reduced form model for the bivariate, stationary time series

$$\mathbf{y}_t = \begin{bmatrix} \Delta \log(Y_t/L_t) \\ \log(L_t) \end{bmatrix}. \quad (5)$$

Labor 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. (2007) and Christiano et al. (2006).¹ Therefore, the section shows how the structural impulse responses Gali was interested in are related to different statistical models, given the economic model. 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

$$\log k_{t+1} = \phi_1 + \phi_{11} \log k_t - \phi_{11} \log x_t + \phi_{12} \tau_{l,t}, \quad (6a)$$

$$\log y_t - \log L_t = \phi_2 + \phi_{21} \log k_t - \phi_{21} \log x_t + \phi_{22} \tau_{l,t}, \quad (6b)$$

$$\log L_t = \phi_3 + \phi_{31} \log k_t - \phi_{31} \log x_t + \phi_{32} \tau_{l,t}, \quad (6c)$$

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 $x_t = X_t/X_{t-1}$. The ϕ 's are the coefficients of the calculated policy rules. Following Fernández-Villaverde et al. (2007) the system can be written in state space form. The state transition equation is

$$\begin{bmatrix} \log k_{t+1} \\ \tau_{l,t} \end{bmatrix} = K_1 + A \begin{bmatrix} \log k_t \\ \tau_{l,t-1} \end{bmatrix} + B \begin{bmatrix} \epsilon_{x,t} \\ \epsilon_{l,t} \end{bmatrix}, \quad (7)$$

$$\mathbf{x}_{t+1} = K_1 + A\mathbf{x}_t + B\epsilon_t,$$

and the observation equation is

$$\begin{bmatrix} \Delta \log(Y_t/L_t) \\ \log L_t \end{bmatrix} = K_2 + C \begin{bmatrix} \log k_t \\ \tau_{l,t-1} \end{bmatrix} + D \begin{bmatrix} \epsilon_{x,t} \\ \epsilon_{l,t} \end{bmatrix}, \quad (8)$$

$$\mathbf{y}_t = K_2 + C\mathbf{x}_t + D\epsilon_t,$$

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 given by $\mathbf{x}_t = [\log k_t, \tau_{l,t-1}]'$ and the noise vector is $\epsilon_t = [\epsilon_{x,t}, \epsilon_{l,t}]'$. 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.

¹ There are also different information sets that are equally applicable in the present context, e.g. $[\Delta y_t, l_t]'$ which would be more in line with Blanchard et al. (1989). This decision should be based on the statistical properties of the series. Results for this alternative information set can be found in a web appendix to this paper.

The above state space system contains the non-observable state vector and the structural errors. We now show different reduced form representations of the system for \mathbf{y}_t in terms of prediction errors that lead to alternative estimation algorithms. These representations are derived under the assumptions that (i) D , C are invertible and (ii) the eigenvalues of $(A - BD^{-1}C)$ are strictly less than one in modulus (Fernández-Villaverde et al., 2007). We checked that this is the case for all models that are used in the Monte Carlo simulations later. Given these invertibility assumptions, there is an **infinite VAR representation**:

$$\mathbf{y}_t = K_3 + C \left(I - (A - BD^{-1}C)L \right)^{-1} BD^{-1} \mathbf{y}_{t-1} + D\epsilon_t, \quad (9)$$

or

$$\mathbf{y}_t = K_3 + \sum_{i=1}^{\infty} \Pi_i \mathbf{y}_{t-i} + u_t, \quad (10)$$

where K_3 and $\Pi_i, i = 1, 2, \dots$ are constant coefficient matrices, L denotes the lag operator, I denotes an identity matrix of suitable dimensions, $u_t = D\epsilon_t$ and $u_t \sim iid N(0, \Sigma_u)$, $\Sigma_u = DD'$, where Σ_u is the covariance matrix of u_t . 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 representation** in “innovations form”:

$$\begin{aligned} \mathbf{x}_{t+1} &= K_1 + A\mathbf{x}_t + K u_t, \\ \mathbf{y}_t &= K_2 + C\mathbf{x}_t + u_t, \end{aligned} \quad (11)$$

where the innovation, u_t , is defined as above and $K = BD^{-1}$. In contrast to the VAR representation in (9), it is possible to estimate (11) without specification error.

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

$$\begin{aligned} \mathbf{y}_t &= K_4 + CAC^{-1}\mathbf{y}_{t-1} + \left(D + (CB - CAC^{-1}D)L \right) \epsilon_t, \\ \mathbf{y}_t &= K_4 + A_1\mathbf{y}_{t-1} + u_t + M_1u_{t-1}, \end{aligned} \quad (12)$$

where the last equation defines the constant coefficient matrices A_1 , M_1 , K_4 and u_t is defined as above. As with the above state space representation, the VARMA(1,1) representation can also be estimated with no specification error.

Given the conditions stated in Fernández-Villaverde et al. (2007), all three

representations are equivalent. They are just different ways of writing down the same process. However, 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 shock in $\epsilon_{x,t}$ to the technology process. However, the different representations give rise to different estimation algorithms and therefore our study can be regarded as a comparison of different algorithms to estimate the same linear system.

4 The Monte Carlo Experiment

4.1 Monte Carlo Design and Econometric Techniques

To investigate the properties of the various 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, 2007) and is referred to as the CKM-specification, while the second is the one used by Christiano et al. (2006) and is labeled the KP-specification, referring to estimates obtained by Prescott (1986).² The specific parameter values are given in table 1 for the CKM and KP benchmark specifications. 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 are reported in table 2. For the CKM specification, we also consider cases where σ_l assumes a fraction of the original benchmark value. 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.

In the following, we present the long-run identification scheme due to Blanchard et al. (1989). Consider the following infinite moving average representation of \mathbf{y}_t in terms of u_t :

² Both parameterizations 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.

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \Phi_{u,i} u_{t-i} = \Phi_u(L) u_t, \quad (13)$$

where we abstract from the intercept term and $\Phi_u(L)$ is a lag polynomial, $\Phi_u(L) = \sum_{i=0}^{\infty} \Phi_{u,i} L^i$. Analogously, we can represent \mathbf{y}_t in terms of the structural errors using the relation $u_t = D\epsilon_t$:

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \Phi_{u,i} D\epsilon_{t-i} = \Phi_\epsilon(L)\epsilon_t, \quad (14)$$

where $\Phi_\epsilon(L) = \sum_{i=0}^{\infty} \Phi_{u,i} DL^i$. The former lag polynomial, evaluated at one,

$$\Phi_u(1) = I + \Phi_{u,1} + \Phi_{u,2} + \dots \quad (15)$$

is the long-run impact matrix of the reduced form error u_t . Note that the existence of this 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 shocks in ϵ_t . Using the above relations, 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. 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,

$$\Phi_\epsilon(1) = \begin{bmatrix} \Phi_{11} & 0 \\ \Phi_{21} & \Phi_{22} \end{bmatrix}, \quad (16)$$

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

$$\hat{\Phi}_\epsilon(1) = \text{chol}[\hat{\Phi}_u(1)\hat{\Sigma}_u\hat{\Phi}_u'(1)], \quad (17a)$$

$$\hat{D} = [\hat{\Phi}_u(1)]^{-1}\hat{\Phi}_\epsilon(1). \quad (17b)$$

Only the first column of \hat{D} is identified and is our estimate of the first-period impact of the technology shock.³

³ Alternatively, one could solve for \hat{D} directly using the three restrictions implied by $\Sigma_u = DD'$ and the long-run identifying restriction (Blanchard et al., 1989), since the Cholesky decomposition can occasionally produce an ill conditioned matrix. In

Next, we comment on the estimation techniques. First, note that for each representation there is more than one reasonable estimation method. We tried several algorithms for all representations but chose to present only the results for the algorithms that worked best for each representation.⁴ Of course, it is still 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. For all methods described below, we ensure that stationary and invertible models are obtained.

Vector Autoregressive Models: VARs are well known, so we comment only on a few issues. As in the previous Monte Carlo studies, the lag length is set at four and the VAR is estimated by OLS. However, for different sets of parameter values a VAR with a different number of 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.⁵ Enforcing stationarity of the estimated model improves the VAR results by some degree.

State Space Models: There are many ways to estimate a state space model, e.g., maximum likelihood methods based on the Kalman filter or subspace identification methods such as N4SID of Van Overschee and De Moor (1994) or the CCA method of Larimore (1983). We use the CCA subspace algorithm that was previously found to be remarkably accurate in small samples. As argued by Bauer (2005a), CCA might be the best algorithm for econometric applications. It is also asymptotically equivalent to maximum likelihood (Bauer, 2005b).⁶ The 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 least squares. The different subspace algorithms use the structure of the state space representation in various ways. See Bauer (2005a) for a more general introduction to subspace

the present context, however, the results from both strategies are identical.

⁴ Additional results and programs may be obtained from the authors.

⁵ Data dependent criteria such as AIC are unfortunately not very helpful for these DGPs. Results for the VAR with AIC selection are presented in a web appendix to this paper. See also Chari et al. (2007).

⁶ We also investigated a maximum likelihood approach using the PEM routine in MATLAB. The results (not reported) were not satisfactory due to reasons discussed below.

methods and the appendix for a detailed description of the algorithm that is employed in this paper.

While implementing the algorithm, we chose the correct dimension of the state vector, $n = 2$.⁷ 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 lag polynomial of the infinite moving average representation is given by

$$\Phi_u(L) = I + \sum_{j=0}^{\infty} CA^j L^{j+1} K = I + LC(I - LA)^{-1}K. \quad (18)$$

An estimate of the long-run impact matrix $\Phi_u(1)$ can be obtained from the estimated system matrices, $\hat{\Phi}_u(1) = I + \hat{C}(I - \hat{A})^{-1}\hat{K}$. 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 .

Vector Autoregressive Moving Average Models: The VARMA representation in (12) implies that we can represent \mathbf{y}_t in terms of the innovations as

$$\mathbf{y}_t = (I - A_1L)^{-1}(I + M_1L)u_t = A(L)^{-1}M(L)u_t, \quad (19)$$

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 the innovations can be estimated by $\hat{\Phi}_u(1) = \hat{A}(1)^{-1}\hat{M}(1)$ and an estimate of the first column of D can be obtained as before. Instead of estimating the VARMA(1,1) representation in (12) we chose a specific representation which guarantees that all parameters are identified and the number of moving average parameters is minimal. 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 (2,1) representation in final moving average form (see

⁷ There are two auxiliary parameters in the subspace algorithm, f , \mathbf{p} , which determine the row and column dimension of a Hankel matrix which is estimated in an intermediate step (see Bauer (2005a) and the appendix). They have been set to $f = \mathbf{p} = 8$. 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 $f = \mathbf{p} = 2\hat{p}$ where \hat{p} is the order of the chosen autoregressive approximation (Bauer, 2005a).

appendix).⁸

As in the case of state space models there are many different estimation methods for VARMA models. Examples are the methods developed by Durbin (1960), Hannan and Rissanen (1982), the generalized least-squares algorithm (Koreisha and Pukkila, 1990), full information maximum likelihood (Mauricio, 1997) or Kapetanios's (2003) iterative least-squares algorithm. We tried the mentioned algorithms but report results for the best performing method which is a simple two-stage least squares algorithm also known as Hannan-Rissanen method. The method starts with an initial "long" autoregression in order to estimate the unobserved residuals. The estimated residuals are then treated as observed and a (generalized) least squares regression is performed. We use a VAR with lag length $n_T = 0.5 \sqrt{T}$ for the initial long autoregression.⁹

4.2 Results of the Monte Carlo Study

Tables 2 - 3 summarize the results of the Monte Carlo simulation study. Table 2 displays Monte Carlo means and standard deviations of the estimates of the contemporaneous impact of a technology shock on productivity and hours. That is, the estimates of 100 times the first column of D , $100 \cdot D[., 1]$. Likewise, table 3 shows Monte Carlo means and standard deviations of the different structural estimators for the percent long-run effect, $100 \cdot \Phi_\epsilon[., 1]$. We chose to compute means and standard deviations based on a trimmed sample of estimates for the long-run effects. Trimming of replications associated with the most extreme upper and lower estimates is beneficial because of a few outliers.¹⁰ In the tables, VAR(4) stands for the SVAR with four lags estimated by OLS. SS(2) denotes the CCA algorithm based on the state space representation with state dimension two. VARMA(2,1) stands for the 2SLS algorithm based on the VARMA in final moving average form with two autoregressive lags and one moving average lag. For all entries, we also tabulate the MSE of the different estimators relative to the MSE of the estimator of the same entry resulting from the VAR(4).

For the estimates of the contemporaneous impact, our SVAR results confirm the findings of both Christiano et al. (2006) and Chari et al. (2005). While the

⁸ We experimented with other identified representations such as the final equation representation or the Echelon representation. However, the final moving average representation yielded the best results.

⁹ In particular, we also tried full information maximum likelihood maximization as formulated in Mauricio (1997). However, this procedure proved to be highly unstable and was therefore not considered to be a practical alternative. One reason is that the roots of the AR and the MA polynomials are all close to the unit circle.

¹⁰ We present result for a sample of estimates trimmed by 4 %. That is, we disregard the 20 lowest and the 20 highest estimates out of 1000 simulations.

SVAR is approximately unbiased for the KP-specification (first row in table 2), the same is not true for the CKM-specification (fourth row in table 2). Also, for the different variations of the benchmark model we find that the SVAR is often biased and/or displays high variability. As can be seen from row 2, 3, 5 and 6 in table 2, both the biases and standard deviations are larger for the models with higher Frisch elasticities of labor supply (lower σ), as in the model this decreases the proportion of the variation in hours worked that is due to the technology shock. From row 7 and 8 it is clear that reducing the relative importance of the tax shock by lowering σ_l by 1/2 and 1/3 reduces the bias and the standard deviations.

The picture is mixed when it comes to the estimation of the long-run effects of technology shocks. First, the SVAR estimator of the effect on productivity displays a much smaller bias and lower standard deviation compared to the estimator of the long-run effect on hours. Relative to the true values, the estimator is slightly less biased in case of the KP parameterizations (rows 1 -3) but the corresponding standard deviations are high relative to those in the case of the CKM parameterizations (rows 4 - 8). The picture changes for the SVAR estimator of the long-run effects on hours. For all parameterizations the volatility is enormous. Also the estimator is much more volatile in case of the KP parameterizations even though the bias is sometimes smaller relative to the true values. Overall, the long-run effect on hours is not estimated accurately.

The CCA algorithm, based on the state space representation, performs quite differently. For the estimates of the contemporaneous impact, we find that the MSE of the CCA subspace algorithm estimator is almost uniformly lower for both series and across different specifications. Only for two parameterizations (fourth and fifth row) does the MSE of the CCA-based estimates exceed the MSE of the SVAR, and only by a relatively small amount. In particular, the first-period impact on hours worked is estimated more precisely up to a relative reduction to 87% in terms of MSE for the KP-specification. In almost all cases the bias is at least slightly reduced. Second, although the response of hours worked is usually estimated more precisely, the performances of the subspace algorithm and the SVAR are clearly related: in cases where the SVAR does poorly, the state space model does so too. For example, both algorithms do relatively well for the KP parametrization but fail dramatically for the CKM parametrization with indivisible labor. Third, we also note that the CCA algorithm is most advantageous relative to the VAR when the VAR is most precise, that is, for the KP parameterizations. Fourth, even though the CCA algorithm can be more precise, the structural estimators are still highly variable and not necessarily much more useful in a qualitative sense.

For the estimates of the long-run effects, the findings are similar. The CCA algorithm is doing better than the SVAR for most parameterizations. Notable exceptions are the results for the long-run effect on productivity for two CKM

parameterizations (fourth and fifth row). Again, the algorithm is outperforming the VAR for the KP parameterizations and its performance is positively correlated with the performance of the SVAR over different parameterizations. The CCA estimator of the long-run effects of a technology shock can be much more precise than the corresponding SVAR estimator. For example, the standard deviation is dramatically reduced in case of the KP parameterizations. Still, even this estimator is not resolving the essential problem, i.e. the standard deviations are far too large to make a reliable qualitative judgement.

The results for the VARMA algorithm are either similar to or worse than those for the VAR approximation. In contrast to the CCA algorithm, we do not observe any visible improvements from the VARMA-based estimator of the contemporaneous impact. In most of the cases the mean bias of the VARMA estimators is rather higher than the bias resulting from the VAR, while the standard deviation might be slightly reduced. Again, we observe that the performance of the VARMA algorithm is highly correlated with the performance of the VAR over different parameterizations of the model. Also, we see that the VARMA gets worse in the most difficult cases (rows four and five). This finding mirrors the results for the CCA algorithm. While the VARMA model fully nests the underlying DGP, the associated algorithm is not very efficient in our context.

Again, the results for the long-run effect of a technology shock are similar although it seems that the 2SLS algorithm is doing better in estimating these effects than in estimating the contemporaneous impact. Also, the performance of this estimator is highly positively correlated with the performance of the SVAR estimator. Therefore, as in the case of the CCA algorithm based on the state space representation, also this estimator is essentially uninformative.

We summarize the findings for all three algorithms as follows:

- The precision of the structural estimators differs more over the different parameterizations of the benchmark model than between different estimators given the same parametrization.
- While the the CCA algorithm appears superior to the VAR in the simulations, the performances of all reported algorithms do not differ too much in a qualitative sense given a particular parametrization.
- For all examples considered, the standard deviations of all estimators of the contemporaneous and long-run effect on hours are quite large, making the estimates uninformative.

These results make clear that the question what is the empirically relevant parametrization or economic model is quite important. From the work of Christiano et al. (2006) it seems safe to conclude that the differences in bias between different parameterizations is mostly due to variation of the relative

importance of technology shocks for the fluctuations in hours worked. Note, however, that the mean of the estimator is not a good summary of its small sample behavior because the variance is so large that almost no weight is attached to values close to the mean. In case of the KP benchmark parametrization the effect on hours is estimated with a standard deviation of 0.43 given a mean of 0.32. Here, a more relevant loss function is the MSE or some other measure that takes more than the first moment into account. In terms of MSE, long-run restrictions perform poorly also in this case.

Two questions arise: Why are all estimators performing so poorly and why do simple methods (e.g., VAR vs. VARMA) perform generally better in the simulations? Using the VARMA representation we can point out three problems with the simulated DGPs. The processes are nearly non-stationary, nearly non-invertible and the correct VARMA representation is close to being not identified. Estimators based on the state space or VARMA representation are more sophisticated and less robust to the near violation of the assumptions on which they are built. This disadvantage seems to compensate to some extent for the advantage of nesting the DSGE model.

We use a general VARMA(p, q) representation for a K -dimensional process to point out the features of the simulated DGPs:

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

where the constant has been omitted. $A(L) = I - A_1L - \dots - LA_pL^p$ is the autoregressive and $M(L) = I + M_1L + \dots + LM_qL^q$ is the moving polynomial with corresponding eigenvalues λ_i^{ar} , λ_i^{ma} , $i = 1, 2, \dots$ which are the inverse roots of $\det A(z)$ and $\det M(z)$, $z \in \mathbb{C}$, respectively. Now, the process is stationary and invertible if and only if all eigenvalues are less than one in modulus (Lütkepohl, 2005). In our case $|\lambda_i^{ar}| < 1$, $|\lambda_i^{ma}| < 1$ for $i = 1, 2$. Table 1 provides these eigenvalues for the benchmark specifications. For example, for the CKM parametrization these are $\lambda_1^{ar} = 0.9573$, $\lambda_2^{ar} = 0.94$, $\lambda_1^{ma} = -0.9557$ and $\lambda_2^{ma} = 0$. Note that the moving average part is not of full rank. These values are very similar for all other parameterizations. That is, all these processes, are nearly non-stationary and non-invertible.¹¹

The fact that one eigenvalue of the moving average part is very close to one eigenvalue of the autoregressive part in modulus is again not confined to the CKM parametrization. It is true for all processes. This point suggests that the VARMA(1,1) representation, though formally correct, is close to being not identified (Klein, Mélard and Spreij, 2005). We know that a VARMA representation is identified if and only if the corresponding Fisher Information

¹¹ The near non-stationarity has also been noticed by other authors such as Chari et al. (2007).

matrix (FIM) is non-singular. Formally, the FIM is the negative expected second derivative of the likelihood function with respect to the parameter vector. Klein et al. (2005) prove that the FIM is singular if and only if it is the case that $\lambda_i^{ar} = -\lambda_i^{ma}$ for at least one i . According to Klein et al. (2005), singularity of the FIM is equivalent to singularity of the tensor Sylvester matrix set forth by Gohberg and Lerer (1976)

$$S^{\otimes}(-M, A) := \begin{pmatrix} (-I_K) \otimes I_K & (-M_1) \otimes I_K & \dots & (-M_q) \otimes I_K & 0 & \dots & 0 \\ 0 & \ddots & \ddots & & & \ddots & \vdots \\ \vdots & & \ddots & \ddots & & \ddots & 0 \\ 0 & & \dots & (-I_K) \otimes I_K & (-M_2) \otimes I_K & \dots & (-M_q) \otimes I_K \\ I_K \otimes I_K & I_K \otimes (-A_1) & \dots & I_K \otimes (-A_p) & 0 & \dots & 0 \\ 0 & \ddots & \ddots & & & \ddots & \vdots \\ \vdots & & \ddots & \ddots & & \ddots & 0 \\ 0 & \dots & 0 & I_K \otimes I_K & I_K \otimes (-A_1) & \dots & I_K \otimes (-A_p) \end{pmatrix},$$

where 0 denotes here the null matrix of dimension $(K^2 \times K^2)$. Klein et al. (2005) propose to check the singularity of this matrix instead of checking the singularity of the FIM directly for numerical reasons. For example, for the CKM benchmark the determinant of the tensor Sylvester matrix is 0.000276. We can perturb the process by changing slightly the eigenvalue of the moving average matrix from -0.9557 to -0.9573. The determinant of the tensor Sylvester matrix jumps to -6.41e-019.¹² That is, even though the DSGE model implies a VARMA(1,1), the process is hard to distinguish from a lower dimensional process. We think that this feature is the most likely explanation why Chari et al. (2007) find that the usual VAR lag-selection criteria are almost always suggesting a VAR(1).¹³

It is clear that a potential lack of identification can be a severe problem for the estimation of the impulse responses in general. In addition it is well documented that near non-identification is especially problematic for the estimation of (vector) ARMA models. See, e.g., the introduction in Mélard, Roy and Saidi (2006) or Ansley and Newbold (1980) for an early documentation. It is also known in the literature on VARMA estimation that processes which are close to being non-invertible are difficult to estimate. Again Ansley et al. (1980) is an early account of this problem as well as Davidson (1981) for pure moving-average models. Additionally, the stationarity assumption is at the heart of the long-run identification strategy. While we ensure that the esti-

¹² Formally, we compute the eigenvalue decomposition $M_1 = V\Lambda V^{-1}$ and change the corresponding entry in Λ . The “perturbed” moving average matrix is then $\tilde{M}_1 = V\tilde{\Lambda}V^{-1}$ and the corresponding process is $y_t = A_1 y_{t-1} + u_t + \tilde{M}_1 u_{t-1}$. Calculations were done with double precision.

¹³ Unfortunately, estimating a lower-dimensional processes does not yield a uniform improvement either.

mated model is stable, the high autoregressive roots will induce small sample bias. These problems are faced by all representations and explain why the observed poor performance is not specific to the VAR methodology. We also hesitate to base any strong recommendation towards a particular class of algorithms, because of the special nature of the simulated processes. However, a sensible strategy might be to consider several estimators at the same time such as a VAR and the CCA method and to aggregate the results in some way as suggested by a thick modeling approach (Granger and Jeon, 2004).

How do these results relate to other results in the literature? First, we think that our results are broadly confirmed by the studies of McGrattan (2006) and Mertens (2007). Mertens (2007) uses spectral methods, proposed by Christiano et al. (2006), to estimate technology shocks in a similar setting. He finds that methods based on the frequency domain, though correctly specified, do poorly and concludes that the observed bias is a result of the small sample size used. Since two of the algorithms used in this paper nest the DSGE model and therefore are also correctly specified, one would attribute the errors to the limited sample size as well. On the other hand, Chari et al. (2005, 2007) and Ravenna (2007) stress that the bias in the SVAR estimates are due to the finite-order truncation used and *not* to small sample problems. These different conclusions are largely due to different terminology, since these authors are referring to the so-called *Hurwicz-type small-sample bias* (Hurwicz, 1950). That is, the difference in mean between a SVAR(4) estimated on a finite sample and a SVAR(4) estimated on an infinite sample. If the lag length is viewed as a function of the sample size $p(T)$ when it comes to approximating infinite VAR processes, then the bias is simply due to T being small. Our study suggests, however, that when the true DGP induces a large truncation bias in the VAR estimates, estimation of other representations is equally difficult. We believe attention should be shifted away from the evaluation of a particular model class and towards the study of the statistical processes one is confronted with and, as in Christiano et al. (2006), of the question whether the usual bootstrap inference is reliable.

5 Conclusions

There has been some debate whether long-run identified SVARs can in practice discriminate between competing DSGE models and whether their sampling properties are good enough to justify their widespread use. Several Monte Carlo studies indicate that SVARs based on long-run restrictions are often biased and usually imprecise. Some authors have suggested that SVARs do poorly because they are only approximate representations of the underlying DGPs. Therefore, we replicate the simulation experiments of Chari et al. (2007) and Christiano et al. (2006) and apply more general models to

their simulated data. In particular, we use algorithms based on VARMA and state space representations of the data and compare the resulting estimates of the underlying structural model. For our simulations, we found that one can do better by taking the full structure of the DGP into account. While our VARMA-based estimation algorithms and some algorithms for state space models were not found to do significantly better and often even worse, the CCA subspace algorithm seems to consistently outperform the VAR approximation. However, the estimators display high variability and are often biased, regardless of the reduced form model used. Furthermore, the performances of the different estimators are strongly correlated. The comparison with estimation methods without specification error suggests that the the main problem is not one of working with a VAR approximation insofar as the properties of the processes used in the literature make identification via long-run restrictions difficult for any method.

Acknowledgement

We would like to thank the editor and two anonymous referees, Anindya Banerjee, Wouter Denhaan, Helmut Herwartz, Helmut Lütkepohl, Morten O. Ravn, Pentti Saikkonen and participants at the “Recent Developments in Econometrics” conference in Florence, as well as seminar participants at the Bank of Spain and Norges Bank for valuable comments and discussion. The research for this paper was conducted at the European University Institute. The views expressed in this paper are our own and do not necessarily reflect the views of Norges Bank.

Appendix A: Final MA Equation Form

Consider a standard representation for a stationary and invertible VARMA (p, q) process

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

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

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

This representation therefore places restrictions on the moving average polynomial which is required to be a scalar operator, $|M(L)|$. Dufour et al. (2004) show that this restriction leads to an identified representation. More specifically, consider the VARMA(1,1) representation in (12). 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 and the intercept is omitted. Clearly, $\det(M(L)) = 1 + (m_{11} + \alpha m_{21})L$ and therefore

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

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

Appendix 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. The derivation follows Fernández-Villaverde et al. (2007). An alternative way to derive a state space system for the purpose of maximum likelihood estimation can be found in Ireland (2001).

Consider again the law of motion of the logs

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

and the exogenous states

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

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

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

Thus the observed series can be expressed as

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

Next, rewrite the law of motion for capital as

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

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

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

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

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

Next, consider the law of motion for capital and express future capital in terms of the current states as

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

Collecting the above equations, the system can be written in state space form according to Fernández-Villaverde et al. (2007). The state transition equation is

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

where the system matrices are given by

$$K_1 = \begin{bmatrix} \phi_1 - \phi_{11}\mu + \phi_{12}(1-\rho)\bar{\tau}_l \\ (1-\rho)\bar{\tau} \end{bmatrix}, \quad A = \begin{bmatrix} \phi_{11} & \phi_{12}\rho \\ 0 & \rho \end{bmatrix}$$

and

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

The observation equation is

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

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_t the white noise shocks. Then we have

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

If D is invertible, it is possible to use $\epsilon_t = D^{-1}(\mathbf{y}_t - K_2 - C\mathbf{x}_t)$ 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 :

$$\begin{aligned}\mathbf{y}_t &= K_2 + C \left(I - (A - BD^{-1}C)L\right)^{-1} \left([K_1 - BD^{-1}K_2] + BD^{-1}\mathbf{y}_{t-1}\right) + D\epsilon_t, \\ \mathbf{y}_t &= K_3 + C \left(I - (A - BD^{-1}C)L\right)^{-1} BD^{-1}\mathbf{y}_{t-1} + D\epsilon_t,\end{aligned}$$

where K_3 is defined by the last equation. 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

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

and use it in the transition equation:

$$\begin{aligned}C^{-1}(\mathbf{y}_{t+1} - K_2 - D\epsilon_{t+1}) &= K_1 + AC^{-1}(\mathbf{y}_t - K_2 - D\epsilon_t) + B\epsilon_t, \\ \mathbf{y}_{t+1} - CAC^{-1}\mathbf{y}_t &= CK_1 + K_2 - CAC^{-1}K_2 + (CB - CAC^{-1}D)\epsilon_t + D\epsilon_{t+1}.\end{aligned}$$

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

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

where K_4 is defined by the equation.

Appendix C: Estimation Algorithms

Two-Stage Least Squares

This simple estimator uses VAR modeling in a first step to estimate the unknown residuals. In the second step these are used to replace the true innovations in the VARMA equations and the coefficient matrices are estimated by least squares. The procedure is easy to implement and is sometimes called the Hannan-Rissanen method or Durbin's method (Durbin, 1960; Hannan et al., 1982). The resulting estimators are not asymptotically efficient (Hannan and Deistler, 1988, chapter 6). We discuss the method in the framework of a standard VARMA (p, q) representation

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

Usually additional restrictions need to be imposed on the coefficient matrices to ensure identification of the parameters.

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

$$y_t = \sum_{i=1}^{n_T} \Pi_i y_{t-i} + u_t,$$

where n_T is large and goes to infinity as the sample size grows. For the choice of n_T data dependent methods such as AIC or BIC (Hannan and Kavalieris, 1984) or deterministic rules such as $0.5\sqrt{T}$, \sqrt{T} (Koreisha et al., 1990) have been suggested. The estimated residuals are denoted by $\hat{u}_t^{(0)}$. Given these estimates, we might obtain estimates of the parameter matrices by performing a (restricted) regression in

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t + M_1 \hat{u}_{t-1}^{(0)} + \dots + M_q \hat{u}_{t-q}^{(0)}.$$

Write the above equation compactly as

$$y_t = [A_1, \dots, A_p, M_1, \dots, M_q] Y_{t-1}^{(0)} + u_t,$$

where $Y_{t-1}^{(0)} := [y'_t, y'_{t-1}, \dots, y'_{t-p}, (\hat{u}_{t-1}^{(0)})', \dots, (\hat{u}_{t-q}^{(0)})']'$. Collecting all observations we get

$$Y = [A_1, \dots, A_p, M_1, \dots, M_q]X^{(0)} + U, \quad (\text{C-1})$$

where $Y := [y_{n_T+m+1}, \dots, y_T]$, $U := [u_{n_T+m+1}, \dots, u_T]$ is the matrix of regression errors, $X^{(0)} := [Y_{n_T+m}^{(0)}, \dots, Y_{T-1}^{(0)}]$ and $m := \max\{p, q\}$. Thus, the regression is started at $n_T + m + 1$. Denote the vector of parameters by $\beta = \text{vec}[A_1, \dots, A_p, M_1, \dots, M_q]$. In order to impose the zero restrictions of the FMA form on β we introduce a restriction matrix R that relates β to the vector of free parameters γ by $\beta = R\gamma$ (Lütkepohl, 2005). Vectorizing equation (C-1) yields

$$\text{vec}(Y) = (X^{(0)'} \otimes I_K)R\gamma + \text{vec}(U),$$

where \otimes is the Kronecker product. The estimator is given by

$$\tilde{\gamma} = [R'(X^{(0)}X^{(0)'} \otimes \hat{\Sigma}^{-1})R]^{-1}R'(X^{(0)} \otimes \hat{\Sigma}^{-1})\text{vec}(Y).$$

where $\hat{\Sigma}$ is the covariance matrix estimator based on the residuals $\hat{u}_t^{(0)}$.

Subspace Algorithms

Subspace algorithms rely on the state space representation of a linear system. The CCA algorithm is originally due to Larimore (1983). The estimator's theoretical properties, including consistency and asymptotic normality, have been developed in a number of papers. Deistler, Peternell and Scherrer (1995) state conditions for the consistency of the transfer function estimates and Bauer (2005b) shows the equivalence of the system matrix estimators to the pseudo maximum likelihood approach. The paper of Bauer (2005a) provides a comprehensive overview. Moreover, the algorithm is also applicable in the unit root context (Bauer and Wagner, 2002).

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 could be obtained by a regression of y_t on x_t and \hat{A} and \hat{K} could be 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.

Solving the state space equations, one can express the state as a function of past observations of y_t and an initial state for some integer $\mathbf{p} > 0$ as

$$\begin{aligned} x_t &= (A - KC)^{\mathbf{p}} x_{t-\mathbf{p}} + \sum_{i=0}^{\mathbf{p}-1} (A - KC)^i K y_{t-i-1}, \\ &= (A - KC)^{\mathbf{p}} x_{t-\mathbf{p}} + \mathcal{K}_{\mathbf{p}} Y_{t,\mathbf{p}}^-, \end{aligned} \quad (\text{C-2})$$

where $\mathcal{K}_{\mathbf{p}} = [K, (A - KC)K, \dots, (A - KC)^{\mathbf{p}-1}K]$ and $Y_{t,\mathbf{p}}^- = [y'_{t-1}, \dots, y'_{t-\mathbf{p}}]'$. On the other hand, one can express future observations as a function of the current state and future noise as

$$y_{t+j} = CA^j x_t + \sum_{i=0}^{j-1} CA^i K u_{t+j-i-1} + u_{t+j}, \quad (\text{C-3})$$

for $j = 1, 2, \dots$. Therefore, at each t , the best predictor of y_{t+j} is a function of the current state only, $CA^j x_t$, and thus the state summarizes in this sense all relevant information in the past up to time t .

Define $Y_{t,f}^+ = [y'_t, \dots, y'_{t+f-1}]'$ for some integer $f > 0$ and formulate equation (C-3) for all observations contained in $Y_{t,f}^+$ simultaneously. Combine these equations with (C-2) in order to obtain

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

where $\mathcal{O}_f = [C', A'C', \dots, (A^{f-1})'C']'$, $E_{t,f}^+ = [u'_t, \dots, u'_{t+f-1}]'$ 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 $(A - KC)$ is less than one in absolute value, we have $(A - KC)^{\mathbf{p}} \approx 0$ for large \mathbf{p} . This condition is satisfied for stationary and invertible processes. This reasoning motivates an approximation of the above equation by

$$Y_{t,f}^+ = \beta Y_{t,\mathbf{p}}^- + N_{t,f}^+, \quad (\text{C-4})$$

where $\beta = \mathcal{O}_f \mathcal{K}_{\mathbf{p}}$ 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 $\mathcal{K}_{\mathbf{p}}$. The identification problem is solved implicitly during this step.

For given integers, n , \mathbf{p} , f , the employed algorithm consists of the following steps:

- (1) Set up $Y_{t,f}^+$ and $Y_{t,\mathbf{p}}^-$ and perform OLS in (C-4) using the available data to get an estimate $\hat{\beta}_{f,\mathbf{p}}$.
- (2) Compute the sample covariances

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

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

- (3) Given the dimension of the state, n , compute the singular value decomposition

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

where $\hat{\Sigma}_n$ is a diagonal matrix that contains the n largest singular values and \hat{U}_n and \hat{V}_n are the corresponding singular vectors. The remaining singular values are neglected and the approximation error is \hat{R}_n . The reduced rank matrices are obtained as

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

- (4) Estimate the state as $\hat{x}_t = \hat{\mathcal{K}}_{\mathbf{p}} Y_{t,\mathbf{p}}^-$ 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 prior to estimation, namely the dimension of the state and the integers f and \mathbf{p} . For the asymptotic consequences of various choices see Bauer (2005a).

References

- Ansley, C.F., Newbold, P., 1980. Finite Sample Properties of Estimators for Autoregressive Moving Average Models. *Journal of Econometrics* 13, 159–183.
- Bauer, D., 2005a. Estimating Linear Dynamical Systems Using Subspace Methods. *Econometric Theory*, 21(1), 181–211.
- Bauer, D., 2005b. Comparing the CCA Subspace Method to Pseudo Maximum Likelihood Methods in the Case of no Exogenous Inputs. *Journal of Time Series Analysis*, 26(5), 631–668.
- Bauer, D., Wagner, M., 2002. Estimating Cointegrated Systems Using Subspace Algorithms. *Journal of Econometrics*, 111, 47–84.

- Blanchard, O.J., Quah, D., 1989. The Dynamic Effects of Aggregate Demand and Supply Disturbances. *The American Economic Review* 79(4), 655–673.
- Chari, V.V., Kehoe, P.J., McGrattan, E.R., 2005. A Critique of Structural VARs Using Real Business Cycle Theory. Federal Reserve Bank of Minneapolis, Working Paper 631.
- Chari, V.V., Kehoe, P.J., McGrattan, E.R., 2007. Are Structural VARs with Long-Run Restrictions Useful in Developing Business Cycle Theory? Federal Reserve Bank of Minneapolis, Staff Report 364.
- Christiano, L.J., Eichenbaum, M., 1992. Identification and the Liquidity Effect of a Monetary Policy Shock, in: Cukierman, A., Hercowitz, Z., Leiderman, L. (Eds.), *Political Economy, Growth and Business Cycles*, Cambridge: The MIT Press.
- Christiano, L.J., Eichenbaum, M., 1992. The research agenda: Larry Christiano and Martin Eichenbaum write about their current research program on the monetary transmission mechanism. *Economic Dynamics Newsletter* 1(1). Society for Economic Dynamics, <http://www.economicdynamics.org>.
- Christiano, L.J., Eichenbaum, M., Vigfusson, R., 2003. What Happens After a Technology Shock? NBER working paper 9819.
- Christiano, L.J., Eichenbaum, M., Vigfusson, R., 2006. Assessing Structural VARs, in: Acemoglu, D., Rogoff, K. and Woodford, M. (Eds.), *NBER Macroeconomics Annual 2006*, Cambridge: The MIT Press.
- Cogley, T., Nason, J.M., 1995. Output Dynamics in Real Business Cycle Models. *The American Economic Review* 85(3), 492–511.
- Cooley, T.F., Dwyer, M., 1998. Business Cycle Analysis Without Much Theory. A Look at Structural VARs. *Journal of Econometrics* 83(1-2), 57–88.
- Davidson, J.E.H., 1981. Problems with the Estimation of Moving Average Processes. *Journal of Econometrics* 16, 295–310.
- Deistler, M., Peternell, K., Scherrer, W., 1995. Consistency and Relative Efficiency of Subspace Methods. *Automatica* 31, 1865–1875.
- Dufour, J.-M., Pelletier D., 2002. Linear Estimation of Weak VARMA Models With a Macroeconomic Application, Université de Montréal and North Carolina State University, working paper. Version: March 2004.
- Durbin, J., 1960. The Fitting of Time-Series Models. *Revue de l'Institut International de Statistique/Review of the International Statistical Institute* 28(3), 233-244.
- Erceg, C.J., Guerrieri, L., Gust, C.J., 2005. Can Long-Run Restrictions Identify Technology Shocks? *Journal of the European Economic Association* 3(4), 1237–1278.
- Fernández-Villaverde, J., Rubio-Ramírez, J.F., Sargent, T.J., Watson, M.W., 2007. A,B,C's (and D)'s For Understanding VARs. *The American Economic Review* 97(3), 1021–1026.
- Francis, N., Ramey, V. A., 2005. Is the Technology-Driven Real Business Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited. *Journal of Monetary Economics* 52(8), 1379–99.
- Francis, N., Ramey, V. A., 2005. Measures of Per-Capita Hours and Their Implications for the Technology-Hours Debate. NBER working paper 11694.
- Gali, J., 1999. Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations? *The American Economic Review* 89(1),

- 249–271.
- Gali, J., Rabanal, P., 2005. Technology Shocks and Aggregate Fluctuations: How well does the RBC model fit postwar U.S. Data? IMF working paper 04/234.
- Gohberg, I., Lerer, L., 1976. Resultants of Matrix Polynomials. *Bulletin of the American Mathematical Society* 82, 565–567.
- Granger, C. W. J., Jeon, Y., 2004. Thick Modeling. *Economic Modelling* 21, 323–343.
- Hannan, E.J., Deistler, M., 1988. *The Statistical Theory of Linear Systems*. New York: Wiley.
- Hannan, E.J., Kavalieris, L., 1984. A Method for Autoregressive-Moving Average Estimation. *Biometrika* 71(2), 273–280.
- Hannan, E.J., Rissanen, J., 1982. Recursive Estimation of Mixed Autoregressive-Moving Average Order. *Biometrika* 69(1), 81–94.
- Hurwicz, L., 1950. Least Squares Bias in Time Series. In: Tjalling C. Koopmans (eds.), *Statistical Inference in Dynamic Economic Models*, 365–383. New York: Wiley.
- Ireland, P.N., 2001. Technology Shocks and the Business Cycle: An empirical investigation. *Journal of Economic Dynamics & Control* 25(5), 703–719.
- Kapetanios, G., 2003. A Note on the Iterative Least-Squares Estimation Method for ARMA and VARMA Models. *Economics Letters* 79(3), 305–312.
- Kehoe, P.J., 2006. How to Advance Theory with Structural VARs: Use the Sims-Cogley-Nason Approach. Federal Reserve Bank of Minneapolis, Research Department Staff Report 379.
- King, R.G., Plosser, C.I., Rebelo, S., 1988. Production Growth and Business Cycles I. The Basic Neoclassical Model. *Journal of Monetary Economics* 21(2-3), 195–232.
- King, R.G., Plosser, C.I., Stock, J.H., Watson, M.W. 1991. Stochastic Trends and Economic Fluctuations. *The American Economic Review* 81(4), 819–840.
- Klein, A., Mélard, G., Spreij, P., 2005. On the Resultant Property of the Fisher Information Matrix of a Vector ARMA Process. *Linear Algebra and its Applications* 403(July), 291–313.
- Koreisha, S., Pukkila, T., 1990. A Generalized Least Squares Approach for Estimation of Autoregressive Moving Average Models. *Journal of Time Series Analysis* 11(2), 139–151. .
- Larimore, W. H., 1983. System Identification, Reduced-Order Filters and Modeling via Canonical Variate Analysis, in: Rao, H.S., Dorato, P. (Eds.), *Proceedings of the 1983 American Control Conference*, New York: Piscataway, 445–451.
- Lütkepohl, H., 2005. *New Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.
- Mauricio, J.A., 1997. Exact Maximum Likelihood Estimation of Stationary Vector ARMA Models. *Journal of the American Statistical Association* 90(429), 282–291.
- McGrattan, E.R., 2006. Measurement with Minimal Theory. Federal Reserve Bank of Minneapolis Working Paper 643.
- Mélard, G., Roy, R., Saidi, A., 2006. Exact Maximum Likelihood Estimation of Structured or Unit Root Multivariate Time Series Models. *Computational Statistics & Data Analysis* 50, 2958–2986.
- Mertens, E., 2007. Are Spectral Estimators Useful for Implementing Long-Run Re-

- restrictions in SVARs? Study Center Gerzensee and University of Lausanne, Working Paper, Version: November 2007.
- Prescott, E.C., 1986. Theory Ahead of Business Cycle Measurement. Federal Reserve Bank of Minneapolis Quarterly Review 10(4), 9–22.
- Ravenna, F., 2007. Vector Autoregressions and Reduced Form Representations of DSGE models. *Journal of Monetary Economics* 54(7), 2048–2064.
- Sims, C.A., 1989. Models and Their Uses. *American Journal of Agricultural Economics* 71(2), 489–494.
- Stock, J. H., Watson, M. W., 2005. Implications of Dynamic Factor Models for VAR Analysis. Working paper.
- Van Overschee, P., De Moor, B., 1994. N4sid: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Processes. *Automatica* 30(1), 75–93.

Tables and Figures

Table 1
Benchmark Calibrations and Time Series Properties

	Common	CKM	KP
Parameters		Benchmark	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			
$\text{eig}(A_1)$		0.9573, 0.9400	0.9573, 0.9930
$\text{eig}(M_1)$		-0.9557, 0	-0.9505, 0

Parameter values of the CKM and KP benchmark calibrations. In the last two rows $\text{eig}(A_1)$ and $\text{eig}(M_1)$ denote the eigenvalues of the autoregressive and the moving average matrix of the associated VARMA representations.

Table 2
Simulation Results - Estimated Contemporaneous Impact

Variable	True Value	VAR(4)			SS(2)			VARMA(2,1)		
		Mean	Std.	MSE	Mean	Std.	MSE	Mean	Std.	MSE
KP Benchmark										
Prod.	0.69	0.55	0.19	1.00	0.58	0.17	0.77	0.55	0.19	1.00
Hours	0.28	0.32	0.43	1.00	0.29	0.41	0.87	0.33	0.43	1.01
KP, $\sigma = 0$ (Indivisible Labor)										
Prod.	0.65	0.47	0.23	1.00	0.50	0.22	0.87	0.47	0.23	0.98
Hours	0.43	0.58	0.56	1.00	0.53	0.52	0.85	0.61	0.55	1.01
KP, $\sigma = 6$ (Frisch Elasticity=0.63)										
Prod.	0.75	0.61	0.15	1.00	0.64	0.13	0.65	0.61	0.15	0.97
Hours	0.11	0.10	0.19	1.00	0.10	0.17	0.82	0.11	0.19	1.02
CKM Benchmark										
Prod.	0.34	0.11	0.16	1.00	0.11	0.18	1.05	0.09	0.16	1.07
Hours	0.14	0.65	0.38	1.00	0.62	0.40	0.98	0.69	0.36	1.09
CKM, $\sigma = 0$ (Indivisible Labor)										
Prod.	0.31	-0.11	0.21	1.00	-0.14	0.24	1.17	-0.15	0.20	1.11
Hours	0.21	1.25	0.49	1.00	1.27	0.55	1.08	1.35	0.46	1.12
CKM, $\sigma = 6$ (Frisch Elasticity=0.63)										
Prod.	0.36	0.30	0.08	1.00	0.31	0.08	0.92	0.30	0.08	0.98
Hours	0.05	0.12	0.17	1.00	0.10	0.17	0.90	0.13	0.17	1.03
CKM, $\sigma_l/2$										
Prod.	0.34	0.26	0.09	1.00	0.26	0.09	0.94	0.26	0.09	1.00
Hours	0.14	0.25	0.21	1.00	0.24	0.21	0.89	0.27	0.21	1.03
CKM, $\sigma_l/3$										
Prod.	0.34	0.28	0.06	1.00	0.29	0.07	0.91	0.29	0.06	0.96
Hours	0.14	0.19	0.14	1.00	0.17	0.14	0.89	0.19	0.14	1.01

Percent contemporaneous impact on productivity and hours of one standard deviation shock to technology. The entries are Monte Carlo means and standard deviations (Std) for the estimates based on the vector autoregressive (VAR), state space (CCA) and vector autoregressive moving average (VARMA) representation. Mean squared errors (MSE) are relative to the MSE of the (structural) VAR estimates.

Table 3
Simulation Results - Estimated Long-Run Effect

Variable	True Value	VAR(4)			SS(2)			VARMA(2,1)		
		Mean	Std.	MSE	Mean	Std.	MSE	Mean	Std.	MSE
KP Benchmark										
Prod.	1.17	1,01	0,39	1,00	0,91	0,23	0,68	1,01	0,39	0,99
Hours	6.66	12,38	22,40	1,00	8,92	14,58	0,41	12,93	22,49	1,02
KP, $\sigma = 0$ (Indivisible Labor)										
Prod.	1.17	1,04	0,39	1,00	0,95	0,28	0,75	1,05	0,40	1,04
Hours	8.71	19,88	26,50	1,00	15,28	18,47	0,46	20,80	26,91	1,05
KP, $\sigma = 6$ (Frisch Elasticity=0.63)										
Prod.	1.17	1,01	0,39	1,00	0,91	0,23	0,67	1,01	0,40	1,04
Hours	3.10	5,15	12,57	1,00	3,37	6,81	0,29	5,10	12,42	0,98
CKM Benchmark										
Prod.	0.57	0,47	0,11	1,00	0,46	0,11	1,12	0,48	0,10	0,91
Hours	3.23	8,26	6,00	1,00	7,80	5,70	0,87	8,66	5,92	1,05
CKM, $\sigma = 0$ (Indivisible Labor)										
Prod.	0.57	0,53	0,13	1,00	0,53	0,15	1,33	0,54	0,13	0,91
Hours	4.22	14,42	8,69	1,00	14,68	8,51	1,01	15,32	8,55	1,09
CKM, $\sigma = 6$ (Frisch Elasticity=0.63)										
Prod.	0.57	0,42	0,08	1,00	0,41	0,06	1,03	0,42	0,07	0,96
Hours	1.50	1,64	2,48	1,00	1,45	2,30	0,86	1,73	2,42	0,96
CKM, $\sigma_l/2$										
Prod.	0.57	0,43	0,09	1,00	0,42	0,08	1,03	0,43	0,08	0,93
Hours	3.23	3,37	2,98	1,00	3,01	2,86	0,92	3,48	2,94	0,98
CKM, $\sigma_l/3$										
Prod.	0.57	0,43	0,10	1,00	0,42	0,08	1,03	0,43	0,09	0,95
Hours	3.23	2,65	2,26	1,00	2,31	2,03	0,92	2,69	2,22	0,97

Percent long-run effect on productivity and hours of one standard deviation shock to technology. The entries are Monte Carlo means and standard deviations (Std) for the estimates based on the vector autoregressive (VAR), state space (SS) and vector autoregressive moving average (VARMA) representation. All statistics are computed over 4% trimmed estimation results. See the text for explanation. Mean squared errors (MSE) are relative to the MSE of the (structural) VAR estimates.