Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 792 (updated version)

March 2005

CAN LONG-RUN RESTRICTIONS IDENTIFY TECHNOLOGY SHOCKS?

Christopher J. Erceg

Luca Guerrieri

Christopher Gust

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to International Finance Discussion Papers (other than an acknowledgement that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at www.federalreserve.gov/pubs/ifdp/.

Can Long-Run Restrictions Identify Technology Shocks?

Christopher J. Erceg, Luca Guerrieri*, and Christopher Gust**

This Draft: March 2005
First Draft: December 2003

Abstract

Galí's innovative approach of imposing long-run restrictions on a vector autoregression (VAR) to identify the effects of a technology shock has become widely utilized. In this paper, we investigate its reliability through Monte Carlo simulations using calibrated business cycle models. We find it encouraging that the impulse responses derived from applying the Galí methodology to the artificial data generally have the same sign and qualitative pattern as the true responses. However, we find considerable estimation uncertainty about the quantitative impact of a technology shock on macroeconomic variables, and little precision in estimating the contribution of technology shocks to business cycle fluctuations. More generally, our analysis emphasizes that the conditions under which the methodology performs well appear considerably more restrictive than implied by the key identifying assumption, and depend on model structure, the nature of the underlying shocks, and variable selection in the VAR. This cautions against interpreting responses derived from this approach as model-independent stylized facts.

Keywords: Technology shocks, vector autoregressions, business cycle models.

* Corresponding author: Board of Governors of the Federal Reserve System, Washington D.C. 20551-0001. Telephone (202) 452 2550. E-mail Luca.Guerrieri@frb.gov.
** The authors are economists at the Board of Governors of the Federal Reserve System. We thank Olivier Blanchard, James Nason, Adrian Pagan, Robert Vigfusson, Chao Wei, and seminar participants at the Federal Reserve Bank of Atlanta, Federal Reserve Board, George Washington University, MIT, the Society for Computational Economics 2004 Meeting, and the Society for Economic Dynamics 2003 Meeting. We especially thank Jon Faust, Roberto Perotti (the editor), and three anonymous referees for their comments. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve System or of any other person associated with the Federal Reserve System.

1 Introduction

The pioneering work of Blanchard and Quah (1989), King, Plosser, Stock, and Watson (1991), and Shapiro and Watson (1988) has stimulated widespread interest in using vector autoregressions (VARs) that impose long-run restrictions to identify the effects of shocks. This methodology has proved appealing because it does not require a fully-articulated structural model or numerous model-specific assumptions.

One important recent application of this approach, introduced by Galí (1999), involves using long-run restrictions to identify the effects of a technology shock. The key identifying assumption in this approach is that only technology innovations can affect labor productivity in the long-run. As discussed in Galí (1999), this assumption holds in a broad class of models under relatively weak assumptions about the form of the production function. Numerous researchers have used this approach to assess how technology shocks affect macroeconomic variables, and to quantify the importance of technology shocks in accounting for output and employment fluctuations.[1]

While the simplicity of Galí's methodology has contributed to its broad appeal, the recent literature has suggested reasons to question whether it is likely to yield reliable inferences about the effects of technology shocks. One reason is that it is difficult to estimate precisely the long-run effects of shocks using a short data sample. Accordingly, as emphasized by Faust and Leeper (1997), structural VARs (SVARs) that achieve identification through long-run restrictions may perform poorly when estimated over the sample periods typically utilized. A second reason, discussed by Cooley and Dwyer (1998) and Lippi and Reichlin (1993), is that a short-ordered VAR may provide a poor approximation of the dynamics of the variables in the VAR if the true data-generating process has a VARMA representation.

In this paper, we critique the reliability of the Galí methodology by using Monte Carlo simulations of reasonably-calibrated dynamic general equilibrium models. In particular, we compare the response of macroeconomic variables to a technology innovation derived from

[1] See, for example, Galí (1999), Francis and Ramey (2003), Christiano, Eichenbaum, and Vigfusson (2003), and Altig, Christiano, Eichenbaum, and Lindé (2003).

applying Galí's identifying scheme with the "true" response implied by our models. We utilize two alternative models of the business cycle as the data-generating process. The first is a standard real business cycle (RBC) model with endogenous capital accumulation that includes shocks to total factor productivity, labor income tax rates, government spending, and labor supply. The second model incorporates some of the dynamic complications that have been identified in the recent literature as playing an important role in accounting for the effects of real and monetary shocks.[2] These features include habit persistence in consumption, costs of changing investment, variable capacity utilization, and nominal price and wage rigidity. The latter model, which we call the sticky price/wage model, provides an alternative perspective on how technology shocks affect the labor market in the short-run, since hours worked decline sharply after a positive innovation in technology rather than exhibit a modest rise as in the RBC model.

We generate Monte Carlo simulations from each model using an empirically-reasonable sample length of 180 quarters. The SVAR that we estimate using the simulated data includes labor productivity growth, the level of hours worked, the ratio of nominal consumption to output, and the ratio of nominal investment to output.[3] One appealing feature of this specification is that a low-ordered VAR (i.e., four lags) provides a close approximation to the true data-generating process in the benchmark parameterizations of each of the models considered.[4] This allows us to interpret the bias in the estimated impulse responses as arising almost exclusively due to the small sample problems emphasized by Faust and Leeper (1997).

Broadly speaking, the shocks derived from application of the Galí methodology to the simulated data "look like" true technology shocks in both of the models we consider. In particular, the mean impulse response functions (IRFs) of output, investment, consumption, and hours worked derived from the Monte Carlo simulations uniformly have the same sign and

[2] See, for example, Christiano, Eichenbaum, and Evans (2001) and Smets and Wouters (2003).

[3] Our inclusion of consumption and investment shares follows Christiano, Eichenbaum, and Vigfusson (2003).

[4] As we show below, our four-variable SVAR with only four lags performs well in recovering the true responses in the benchmark parameterizations of each of the models if the SVAR is estimated using population moments from the DGE model rather than sample moments.

qualitative pattern as the true responses. Moreover, we find that the probability of inferring a response of output, consumption, or investment that has the qualitatively incorrect sign (even for only a few quarters) is generally low.

However, we find that small-sample bias poses quantitative problems for this identifying scheme. There is substantial downward bias in the estimated responses of output, labor productivity, consumption, and investment derived from the Monte Carlo simulations in each of the models. Moreover, given the bias and substantial spread in the distribution of the impulse responses, we find that the probability that a researcher would estimate a response for output that lies uniformly more than 33 percent away from the true response (for the first four quarters following the shock) is about 25 percent in each of the models.

We show that the bias in the estimated impulse responses is dependent on model structure. Within the context of the benchmark models, the bias can be attributed to two related sources. First, the slow adjustment of capital makes it hard to gauge the long-run impact of a technology shock on labor productivity, contributing to downward bias in the estimated impulse responses.[5] Second, the identification procedure has difficulty disentangling technology shocks from other shocks that have highly persistent, even if not permanent, effects on labor productivity (such as labor supply or tax rate shocks).[6] As a result, even in the absence of shocks that would violate Galí's long-run identifying assumption, the estimated technology shock may incorporate a sizeable non-technology component. Accordingly, the bias in the estimated response of a given variable to a technology shock depends on the relative magnitude of technology and non-technology shocks, and on its response to non-technology shocks.

Our results also have implications for a principal application of the Galí methodology,

[5]The fact that slow adjustment of capital creates problems for the identification scheme may seem surprising given the well-known problem emphasized by Cogley and Nason (1995) that standard real business cycle models fail to generate enough endogenous persistence. Cogley and Nason (1995) focus on the inability of these models to generate enough positive autocorrelation in output *growth*, but this is still consistent with slow adjustment in the *level* of labor productivity.

[6]In this respect, our paper shares similarities with an earlier literature emphasizing that the measured Solow residual is contaminated by aggregate demand disturbances. See, for example, Evans (1992) and references therein.

which has involved using estimates derived from SVARs to evaluate the plausibility of alternative models of the business cycle.[7] Interestingly, though there is considerable uncertainty about the estimated response of hours worked, our results suggest that the SVAR approach may provide some basis for discriminating between models that have sufficiently divergent implications about how technology shocks affect the labor market. For instance, we find that the probability of finding an initial decline in hours that persists for two quarters is 93 percent in the model with nominal rigidities, but only 26 percent in the RBC model. Accordingly, a researcher who found that hours worked declined after a positive innovation in technology in the data could reasonably interpret this finding as providing some evidence in favor of the sticky price/wage model. While our results are encouraging on this dimension, our analysis cautions that appropriate specification of the SVAR (i.e., variables included and their transformations) appears to play a key role in allowing tests to have sufficient power to discriminate between alternative models.

By contrast, we find that there is very little precision in estimating the contribution of technology shocks to output fluctuations at business cycle frequencies. For example, the 90 percent confidence intervals for the contribution range between 7 and 90 percent for the benchmark RBC model, and between 7 and 80 percent for the sticky price/wage model.

Our analysis also illustrates how the performance of the Galí procedure may be influenced by the selection of variables in the VAR, the transformations applied, and the nature of the underlying shocks. We find that the performance of the Galí procedure may exhibit noticeable sensitivity to the specification of variables in the VAR. This sensitivity in part reflects that for some variable choices a low-ordered VAR may perform poorly in capturing the VARMA representations implied by our models.[8] We also find that the performance of the Galí methodology

[7]See Galí (1999), Francis and Ramey (2003), Galí and Rabanal (2005), and Christiano, Eichenbaum, and Vigfusson (2003).

[8]In a recent paper, Chari, Kehoe, and McGrattan (2005) find that bivariate SVARs with labor productivity growth and hours (in either levels or differences) perform poorly in the RBC model. Our analysis corroborates their finding in this particular case; however, we consider a broader class of models and SVAR specifications. Overall, we are more sanguine towards the Galí approach because we find specifications (e.g., the four-variable

deteriorates on some dimensions with the inclusion in the RBC model of technology shocks that are stationary but highly persistent.

Overall, Galí's methodology appears to offer a fruitful approach to uncovering the effects of technology shocks, and it is encouraging that our baseline, four-variable SVAR specification performs reasonably well across the alternative models considered. However, our analysis emphasizes that the conditions under which the Galí methodology performs well appear considerably more restrictive than implied by the key identifying restriction, and depend on model structure, the nature of the underlying shocks, and on variable selection in the SVAR. Accordingly, we caution that empirical estimates of the effects of technology shocks should not be regarded as model-independent stylized facts. Instead, the interpretation of results derived from the Galí approach should be informed by the model or class of models that the researcher regards as most plausible, with the model serving as a guidepost about biases likely to arise and the limitations of the approach.

The rest of this paper is organized as follows. Section 2 outlines our baseline RBC model and describes the calibration. Section 3 reviews the Galí identification scheme. Section 4 reports our results for the RBC model, and Section 5 discusses the results for the sticky price/wage model. Section 6 concludes.

2 The RBC Model

We begin by outlining a relatively standard real business cycle model. The model structure is very similar to that analyzed by King, Plosser, and Rebelo (1988), though we include a broader set of shocks.

SVAR) that perform reasonably well across the models we considered.

2.1 Household Behavior

The utility function of the representative household is

$$E_t \sum_{j=0}^{\infty} \beta^j \{\log(C_{t+j}) - \chi_{0t+j} \frac{N_{t+j}^{1+\chi}}{1+\chi}\}, \tag{1}$$

where the discount factor β satisfies $0 < \beta < 1$ and E_t is the expectation operator conditional on information available at time t. The period utility function depends on consumption, C_t, labor, N_t, and a stochastic shock, χ_{0t}, that may be regarded as a shock to labor supply. We assume that this labor supply shock evolves according to:

$$\log(\chi_{0t}) = (1-\rho_\chi)\log(\chi_0) + \rho_\chi \log(\chi_{0t-1}) + \sigma_\chi \epsilon_{\chi t}, \tag{2}$$

where χ_0 denotes the steady state value of χ_{0t} and $\epsilon_{\chi t} \sim N(0,1)$.

The representative household's budget constraint in period t states that its expenditure on consumption and investment goods (I_t) and net purchases of bonds B_{t+1} must equal its after-tax disposable income:

$$\begin{aligned} C_t + I_t + \tfrac{1}{1+r_t} B_{t+1} - B_t = \\ (1-\tau_{Nt}) W_t N_t + \Gamma_t + T_t + (1-\tau_{Kt}) R_{Kt} K_t + \tau_{Kt} \delta K_t. \end{aligned} \tag{3}$$

The household earns after-tax labor income of $(1-\tau_{Nt})W_t N_t$, where τ_{Nt} is a stochastic tax on labor income, and also receives an aliquot share of firm profits Γ_t and a lump-sum government transfer of T_t. The household leases capital services to firms at an after-tax rental rate of $(1-\tau_{Kt})R_{Kt}$, where τ_{Kt} is a stochastic tax on capital income. The household receives a depreciation writeoff of $\tau_{Kt}\delta$ per unit of capital (where δ is the steady state depreciation rate of capital). Purchases of investment goods augment the household's capital stock according to the transition law:

$$K_{t+1} = (1-\delta)K_t + I_t. \tag{4}$$

In every period t, the household maximizes utility (1) with respect to its consumption, labor supply, investment, (end-of-period) capital stock, and real bond holdings, subject to its budget constraint (3), and the transition equation for capital (4).

2.2 Firms

The representative firm uses capital and labor to produce a final output good that can either be consumed or invested. This firm has a constant returns-to-scale Cobb-Douglas production function of the form:

$$Y_t = K_t^{\theta}(Z_t V_t N_t)^{1-\theta}, \tag{5}$$

In the above, Z_t is a unit-root process for technology whose law of motion is governed by:

$$\log(Z_t) - \log(Z_{t-1}) = \mu_z + \sigma_z \epsilon_{zt}, \tag{6}$$

and V_t is a stationary process for technology whose law of motion is governed by:

$$\log(V_t) = \rho_V \log(V_{t-1}) + \sigma_V \epsilon_{Vt}, \tag{7}$$

with ϵ_{zt}, $\epsilon_{Vt} \sim N(0,1)$.

The firm purchases capital services and labor in perfectly competitive factor markets, so that it takes as given the rental price of capital R_{Kt} and the aggregate wage W_t. Since the firm behaves as a price taker in the output market as well as in factor markets, the following efficiency conditions hold for the choice of capital and labor:

$$\frac{W_t}{MPL_t} = \frac{R_{Kt}}{MPK_t} = 1. \tag{8}$$

2.3 Government

Some of the final output good is purchased by the government, so that the market-clearing condition is:

$$Y_t = C_t + I_t + G_t. \tag{9}$$

Government purchases are assumed to have no direct effect on the utility function of the representative household. We also assume that government purchases as a fraction of output, $g_t = G_t/Y_t$, are exogenous and evolve according to:

$$\log(g_t) = (1-\rho_g)\log(g) + \rho_g \log(g_{t-1}) + \sigma_g \epsilon_{gt}, \tag{10}$$

where g denotes the steady state value of g_t and $\epsilon_{gt} \sim N(0,1)$.

The government's budget is balanced every period, so that total taxes – which include both distortionary taxes on labor and capital income – equal the sum of government purchases of the final output good and net lump-sum transfers to households.[9] Hence, the government's budget constraint at date t is:

$$T_t + G_t = \tau_{Nt} W_t N_t + \tau_{Kt}(R_{Kt} - \delta)K_t. \tag{11}$$

The tax rates on capital and labor are assumed to be exogenous and evolve according to:

$$\tau_{it} = (1 - \rho_{\tau_i})\tau_i + \rho_{\tau_i}\tau_{it-1} + \sigma_{\tau_i}\epsilon_{\tau_i t}, \tag{12}$$

where τ_i is the steady state tax rate and $\epsilon_{\tau_i t} \sim N(0,1)$ for $i = K, N$.

2.4 Solution and Calibration

To analyze the behavior of the model, we first apply a stationary-inducing transformation to those real variables that share a common trend with the level of technology. This entails detrending real GDP, the GDP expenditure components, and the real wage by Z_t and the capital stock, K_t, by Z_{t-1}. We then compute the solution of the model using the numerical algorithm of Anderson and Moore (1985), which provides an efficient implementation of the solution method proposed by Blanchard and Kahn (1980).

Table 1 summarizes the calibrated values of most of the model's parameters. The model is calibrated at a quarterly frequency so that $\beta = 1.03^{-0.25}$ and $\delta = 0.02$. The utility function parameter χ is set to 1.5 so as to imply a Frisch elasticity of labor supply of 2/3, an elasticity well within the range of most empirical estimates.[10] The capital share parameter θ is set to 0.35, and we normalized $\chi_0 = 1$, as χ_0 does not affect the model's log-linear dynamics.

Using data on the share of government consumption to U.S. GDP, we fit a first order autoregression for g_t (allowing for a linear time trend) and estimated ρ_g and σ_g in equation (10)

[9] The assumption of a balanced budget is not restrictive given the availability of lump-sum taxes or transfers.

[10] See, for example, Pencavel (1986), Killingsworth and Heckman (1986), and Pencavel (2002).

to be 0.98 and 0.003, respectively. We set g so that the ratio of government spending to output is 20% in the model's non-stochastic steady state.

For the parameters governing the two tax rate series, we estimated equation (12) using OLS after constructing these tax rates series based on U.S. data from 1958-2002 following the methodology described in Jones (2002).[11] Our estimates implied $\tau_N = 0.22$, $\rho_{\tau_N} = 0.98$, and $\sigma_{\tau_N} = 0.0052$ for the labor tax rate and $\tau_K = 0.38$, $\rho_{\tau_K} = 0.97$, and $\sigma_{\tau_K} = 0.008$ for the capital tax rate.

For reasons that we discuss below, it is convenient to exclude capital tax rate and temporary technology shocks from our benchmark calibration of the RBC model; thus, we set $\sigma_{\tau_K} = \sigma_V = 0$. In this case, we can obtain a time series for Z_t by defining the Solow residual as:

$$S_t = \frac{Y_t}{K_t^{\theta} N_t^{1-\theta}}, \tag{13}$$

and noting that $Z_t = S_t^{\frac{1}{1-\theta}}$. We then estimate $\mu_z = 0.0037$ and $\sigma_z = 0.0148$. Later, we give special attention to the capital tax rate and temporary technology shocks in an alternative parameterization of the RBC model.

In the absence of labor-supply shocks, our calibrated RBC model would significantly underestimate the volatility in hours worked – a familiar problem in the real business cycle literature. To see this, Table 2 compares the second moments of several key variables that are implied by our model with their sample counterparts based on U.S. data. As shown in the column labelled "$\sigma_\chi = 0$", the model significantly understates the ratio of the standard deviation of HP-filtered hours to the standard deviation of HP-filtered output. For our benchmark calibration, we address this issue by incorporating labor supply shocks.[12] In particular, we set $\rho_\chi = 0.95$ and choose an innovation variance σ_χ that allows the model to match the observed standard deviation of HP-filtered hours relative to the standard deviation of HP-filtered output.

Table 2 shows the selected moments for the benchmark RBC model. A comparison of the model's implications for the volatility of output, investment, and consumption to the

[11] Following Appendix B in Jones (2002), we used quarterly data collected by the Bureau of Economic Analysis.

[12] Others who have followed this approach include Hall (1997), Shapiro and Watson (1988), and Parkin (1988).

corresponding sample moments suggests that this calibrated model performs fairly well on these dimensions, even though it was not calibrated specifically to match these moments.

3 The SVAR Specification

In this section, we outline the estimation procedure that a researcher would follow given a single realization of data. The structural VAR takes the form:

$$A(L)X_t = u_t = A_0^{-1}e_t, \tag{14}$$

where $A(L) = I - A_1L - ... - A_pL^p$, and A_i for $i = 1, 2, ..., p$ is a square matrix of reduced-form parameters; L is the lag operator, and X_t , u_t, and e_t are $4X1$ vectors of endogenous variables, reduced-form innovations, and structural innovations, respectively. The lag length, p, is chosen by using the information criterion in Schwarz (1978), where $p \in \{1, 2, ..., 10\}$.

In our benchmark specification of the VAR, X_t contains the log difference of average labor productivity, the log of hours worked, the log of the consumption-to-output ratio, and the log of the investment-to-output ratio. All variables are expressed as a deviation from the model's nonstochastic steady state, and average labor productivity is defined as Y_t/N_t. The inclusion of average labor productivity growth in X_t is standard in the empirical literature using VARs to identify technology shocks. While the empirical literature is divided on whether hours worked are best included in levels or differences, the former specification is selected, because the DGE model implies that hours are stationary in levels. The ratios of investment and consumption to output are included in the VAR, in part because Christiano, Eichenbaum, and Vigfusson (2003) have found these variables to be important in controlling for omitted-variable bias when using U.S. data.

The identification of the technology shock is achieved in the following way. First, it is assumed that the innovations are orthogonal and have been normalized to unity so that

$$Ee_te_t' = A_0\Sigma A_0' = I, \tag{15}$$

where Σ denotes the variance-covariance matrix of the reduced-form residuals. Denote the first element of e_t as e_{zt}, the technology shock identified by the VAR. Following Galí (1999), a researcher would then impose that the technology shock is the only shock that can affect the level of productivity in the long run, an assumption that is consistent with the models we consider. Thus, letting $R(L) = A(L)^{-1}$, it follows that

$$[R(1)A_0^{-1}]_{1j} = 0 \quad \text{for } \; j \neq 1. \tag{16}$$

Here, R(L) is the reduced-form moving average representation of the VAR given by

$$R(L) = \sum_{i=0}^{\infty} R_i L^i, \tag{17}$$

where R_i is a 4×4 matrix and $R_0 = I$. The restrictions associated with equation (16) are imposed through a Cholesky decomposition after estimating $A(L)$ and Σ using least squares. This decomposition is used to solve for the first column of A_0^{-1} given that $R(1) = A(1)^{-1}$. No attempt is made to identify the non-technology shocks.

In our Monte Carlo study, we generate 10,000 data samples from the relevant DGE model, and apply the estimation strategy discussed above to each sample. Every data sample consists of 180 quarterly observations.[13]

4 Estimation results for the RBC Model

Figure 1 reports the response of labor productivity, hours worked, consumption, investment, and output to a technology shock for the benchmark calibration of the RBC model.[14] In each panel, the solid lines show the true responses from the DGE model. The innovation occurs at date 1 and has been scaled so that the level of labor productivity rises by one percent in the long run.

The dashed lines show the mean of the impulse responses derived from applying our benchmark, four-variable SVAR to the 10,000 artificial data samples (the median response is

[13]In the appendix, we discuss the sensitivity of our results to different sample lengths.

[14]More precisely, the responses shown are the deviations of the log level of each variable from the steady-state growth path.

nearly identical).[15] The dotted lines show the 90 percent pointwise confidence interval of the SVAR's impulse responses.[16]

As shown in Figure 1, the mean responses of labor productivity, consumption, investment, and output have the same sign and qualitative pattern as the true responses. As indicated by the pointwise confidence intervals, the SVAR is likely to give the appropriate sign of the response for these variables. For hours worked, the mean estimate is also qualitatively in line with the true response; however, the confidence interval is wide, indicating that there is a non-negligible probability of a negative estimate.

Quantitatively, the SVAR does not perform as well. As seen in Figure 1, the mean responses of the SVAR systematically underestimate labor productivity, consumption, investment, and output, while overestimating hours worked. To gauge the size of the bias, the top row of Table 3 reports the average absolute percent difference between the mean response and the true response over the first twelve quarters for each of the variables except hours worked.[17] For hours, Table 3 reports the absolute value of the difference between the mean estimated response and the true response (we simply report the difference because the true response is very small). As reported in the first row of Table 3, labor productivity is underestimated by the SVAR by 40% on average over the first 12 quarters after the innovation to technology, while output is underestimated by 25%. We defer our explanation of these results to Section 4.1.

While useful for illustrating the bias associated with the SVAR's estimates, the relative distance measure does not capture the uncertainty that a researcher confined to a single draw of the data would confront. After all, the impulse response derived using a single realization of the data may diverge substantially from the mean. Accordingly, we consider an alternative

[15]We scale up the technology innovation derived from the SVAR by the same constant factor as applied to the true innovation.

[16]These confidence intervals are also constructed from the estimated impulse responses derived from applying the SVAR to the 10,000 artificial data samples from our model.

[17]For variable i, this measure is defined as $rd_i^m = \frac{1}{12}\sum_{l=1}^{12} |rd_{l,i}^m|$ where $rd_{l,i}^m = \frac{\hat{d}_{l,i}^m - d_{l,i}^*}{d_{l,i}^*}$, and $d_{l,i}^*$ and $\hat{d}_{l,i}^m$ denote the DGE model's impulse response and the SVAR's mean response to a technology shock, respectively, at lag l.

measure of how well the SVAR's point estimates of the impulse responses match the truth. For variable i, this measure is defined as

$$\hat{P}_i(\frac{1}{3}) = P(|rd_{l,i}| \geq \frac{1}{3}), \quad \forall l \in \{1, 2..., N\}, \tag{18}$$

where $rd_{l,i} = \frac{\hat{d}_{l,i} - d^*_{l,i}}{d^*_{l,i}}$ and $\hat{d}_{l,i}$ denotes the estimated impulse response for the i^{th} variable at lag l for a given draw of data, and $d^*_{l,i}$ denotes the response from the DGE model. In words, $\hat{P}_i(\frac{1}{3})$ is the probability that the SVAR produces an impulse response that lies at least 33 percent above or below the true response for all lags between 1 and N, which we call a "large" error. Tables 4, 5, and 6 show these probabilities for N equal to two, four, and twelve quarters, respectively (as noted below, we define the measure of a large error for hours worked differently). As shown in the top row of Table 5, the probability of a large error over the first year is 43% for labor productivity and 24% for output. Furthermore, we found that nearly all of the large misses of the SVAR's impulse responses for output and labor productivity were the result of underpredicting the true response. Given the strict criterion that only counts impulse response functions that lie uniformly outside the 33 percent band, our results suggest considerable estimation uncertainty about the quantitative effects of a technology shock.

While the probability of underestimating labor productivity, consumption, output, and investment is substantial, the probability of inferring an incorrect sign for several quarters is very low (not reported). It is also interesting to assess the probability of inferring a response of hours worked that has the incorrect sign in the first few periods, given the significant attention recent research has devoted to this question. Accordingly, for hours worked, Tables 4, 5, and 6 report the probability that the estimated response of hours worked is incorrect (negative in this model) in each of the first 2, 4, and 12 quarters, respectively. As shown in Figure 1, the true response of hours is positive, and there is upward bias in the mean estimated response. Nevertheless, Table 5 shows that there is a 23% chance a researcher would find that hours worked fell for four straight quarters in the year following a technology shock.[18]

[18] This probability may seem surprisingly low given the width of the confidence intervals shown in Figure 1. However, it is important to note that the confidence intervals are pointwise, while the probabilities reported

Galí (1999), Galí (2004), and Christiano, Eichenbaum, and Vigfusson (2003) have employed SVARs with long-run restrictions to estimate the contribution of technology shocks to business cycle fluctuations, and have used these estimates to conclude that technology shocks only play a small role in driving output fluctuations over the business cycle.[19] We use our framework to assess the reliability of these estimates. The top left panel of Figure 2 shows the cumulative distribution function derived from Monte Carlo simulations of our estimator of the contribution of technology shocks to output fluctuations. This contribution is defined as $RC_z = \sigma^2_{y|z}/\sigma^2_y$ where σ^2_y denotes the unconditional variance of HP-filtered output in the model and $\sigma^2_{y|z}$ is the variance of HP-filtered output conditional on only unit-root technology shocks.[20] The distribution function appears close to uniform over the unit interval, so that the 90% confidence bands for the estimator include contributions ranging from 7 to 91 percent (confidence bands are indicated by stars on the x-axis). Therefore, for the benchmark RBC model, this application of the SVAR methodology yields uninformative estimates of the contribution of technology shocks to output fluctuations at business cycle frequencies.

4.1 Interpreting the Bias

In this section, we estimate the contribution of two sources of bias in the impulse responses shown in Figure 1. We then provide an interpretation of the economic mechanisms accounting for the bias in the RBC model.

The first source of bias, which we call "truncation bias," arises because the finite-ordered VAR chosen by our estimation procedure only provides an approximation to the true dynamics

in Tables 4-6 are uniform measures, requiring that hours worked fall in each period for 2, 4 or 12 quarters. Furthermore, the distribution of the estimated responses of hours at a given lag is not uniform.

[19]A notable exception is Fisher (2002), who attempts to discriminate between multi-factor productivity shocks and investment-specific technology shocks.

[20]In order to estimate $\sigma^2_{y|z}$ we did the following: for a given replication of data from the DGE model, we used the point estimates from the SVAR to bootstrap a series of 41,000 observations for output conditional on only the identified technology shocks; we HP-filtered this series after dropping the first 1,000 observations. Similarly, for σ^2_y, we bootstrapped a series for output from the fitted VAR using all the shocks.

implied by the model. In particular, our model produces an exact VARMA(4,5) representation and even though this VARMA process is invertible, a finite-ordered VAR may be misspecified to some degree.[21] Cooley and Dwyer (1998) emphasized that most popular DGE models imply VARMA representations; thus, truncation bias is a fairly general phenomenon. The second source of bias is the small sample bias that arises in all time series work. Faust and Leeper (1997) highlighted small sample imprecision as a problem in this context, although they did not focus on bias in particular.

To measure the truncation bias, we calculate the population limit of a VAR(4) based on our model. In principle, we could estimate this using one very long sample drawn from the model, but we simply use the relevant population moments from the model to derive the VAR(4).[22]

Figure 3 compares the effects of a technology shock derived from the population SVAR with the true model responses. Though the four variables in the VAR have a VARMA(4,5) representation in our benchmark RBC model, it is clear that the truncation bias appears negligible for each of the variables depicted. Thus, for the benchmark calibration of the RBC model, the assumption that a short-ordered VAR provides a good approximation to the true data-generating process seems warranted. This proves attractive heuristically, because we can interpret almost all of the bias as arising from a small sample.

Accordingly, we further decompose the small-sample bias into two parts, and show that most of the small sample bias is attributable to the difficulty in precisely estimating the long-run response of variables to the innovations in the VAR. Noticing that equation (14) can be

[21]We checked numerically that the benchmark RBC model implied a VARMA process that is invertible and thus a fundamental representation. See the appendix for details of these calculations. Hansen and Sargent (2004) and Lippi and Reichlin (1993) analyze the problem in which the moving average component is not invertible so that it is not possible to recover the fundamental shocks from a VAR of any lag-length.

[22]Of course, if one had a long sample, one could choose a longer lag length. We use this population VAR(4) only to measure the bias that is due to the inherent inability of a VAR(4) to approximate the VARMA(4,5) structure of the model.

expressed as:

$$X_t = A(L)^{-1} A_0^{-1} e_t = R(L) A_0^{-1} e_t, \tag{19}$$

it is evident that the response of X_t to the underlying innovations, e_t, is influenced both by the reduced-form moving average terms, $R(L)$, and by the identifying restrictions as reflected in A_0^{-1}. Therefore, we can think of one part of the bias as reflecting the small-sample error in estimating the reduced-form moving average terms, which we call the "R bias". The second part reflects the error associated with transforming the reduced form into its structural form by imposing the long-run restriction. This latter error occurs because small imprecision in estimating $A(L)$ is exacerbated by the nonlinear mapping involved with imposing the long-run restriction. As a result, estimates of A_0^{-1} can be biased in small samples. We call the error associated with the transformation of the reduced form to the structural form "A bias."[23]

Returning to the lower right panel of panel of Figure 1, we provide a decomposition of the overall bias in the mean response of labor productivity. The overall bias is represented by the solid line labelled "total bias", and is simply the difference between the mean estimated response of labor productivity to a technology innovation and the true response. The dotted line labelled "T bias" for truncation bias shows the bias introduced by assuming that the variables in the VAR can be represented by a VAR with only four lags. As suggested by Figure 3, this source of bias comprises only a tiny fraction of the bias in the mean response of labor productivity. From the dashed-dotted line labelled "A bias", it is clear that most of the small-sample bias initially is attributable to the error in transforming the reduced form into its structural form using the long-run restriction.[24] Eventually, however, imprecision in estimating the long-run responses

[23]Our decomposition is discussed in greater detail in Appendix 7.3, where we also provide more explicit definitions of the alternative types of bias. As noted in the appendix, our "A bias" reflects not only the error associated with transforming the reduced-form to structural, but also the error associated with estimating Σ, the variance/covariance matrix for e_t. We found that this latter source of error was small.

[24]In our analysis, there appears to be a connection between the type of imprecision emphasized by Faust and Leeper (1997) and the weak instrument problem discussed by Pagan and Robertson (1998). In particular, we find that when we estimate the SVAR using the instrumental variable approach of Shapiro and Watson (1988), parameter values of the RBC model that implied the "A bias" was large corresponded to situations where there

has a roughly commensurate effect on each component, so that the R bias contributes about as much to the bias as the A bias.

We now use the benchmark RBC model to provide an economic interpretation of the small sample bias that illustrates how it depends on model structure. This bias can be attributed largely to two related factors in our RBC model. First, the slow adjustment of capital makes it hard to estimate the long-run impact of a technology shock on labor productivity, which serves as a source of downward bias in the estimated impulse responses. Second, the SVAR has difficulty disentangling technology shocks from highly persistent non-technology shocks, so that the estimated technology shock may incorporate a sizeable non-technology component. The second source of bias has more pronounced effects on the estimated responses to a technology shock as the relative magnitude of non-technology shocks rises, and as the non-technology shocks become more persistent.

We conduct two experiments to show that the small sample bias is greatly reduced when the exogenous and endogenous sources of persistence in the model are decreased. First, as seen in the rows of Tables 3 to 6 labelled "with lower persistence", we analyze the effects of halving all of the AR(1) parameters of the non-technology shocks from their benchmark values. Table 3 shows that the (percentage) distance between the mean and the true response narrows for all variables and especially for labor productivity, and Tables 4 to 6 indicate that there are sizeable declines in the frequencies of large misses for all the variables we consider. Our second experiment combines the lower persistence of non-technology shocks with an increase in the depreciation rate of capital from $\delta = 0.02$ to $\delta = 0.9$. In this case, labor productivity adjusts more quickly in response to both technology and non-technology shocks. Table 3 shows that the mean bias falls below 10% for all the variables.[25,26]

were also weak instruments.

[25]If we only increase δ and leave the persistence of the exogenous shocks at their benchmark values, then there is only a small reduction in the bias for most variables with the exception of labor productivity. For example, the average bias in labor productivity over the first 12 quarters declines from 40% in our benchmark RBC model to 23%, but the average bias for output only declines from 25% to 22%.

[26]With less exogenous and endogenous persistence, the SVAR's ability to estimate the contribution of unit-root technology shocks to output fluctuations at business cycle frequencies improves noticeably, though the

Our final experiment in this section illustrates the important influence that the non-technology shocks may have on the SVAR's estimated responses. We reduce the innovation variance of the technology shock to 0.0049, or one-third of its benchmark value, thus effectively increasing the relative size of the non-technology shocks. The mean estimated responses and true responses to a technology shock under this alternative parameterization are depicted in Figure 4 (and reported in Table 3 in the row labelled "with $\sigma_z = 1/3X$"). With this increase in the relative size of the non-technology shocks, the estimated responses look more like the effects that arise from labor supply shocks (the dominant non-technology shock in the benchmark calibration). To see this, we also plot the true responses to a labor supply shock in the same figure. Observe that relative to their effects on labor productivity, labor supply shocks have much larger effects on hours worked and investment than a true technology shock. Given that estimates derived from the SVAR approach confound labor supply with true technology innovations, the former shocks are a source of upward bias in the estimated responses of hours worked and investment to a technology shock. Thus, with the increased importance of labor supply shocks in this alternative calibration, the upward bias in the mean response of hours worked is much more pronounced than under our benchmark calibration, and the bias in investment shifts from negative to noticeably positive.

4.2 Sensitivity Analysis

We next use sensitivity analysis to illustrate how the performance of the Galí procedure may be influenced by the selection of variables in the VAR, the transformations applied, and the inclusion of a wider array of shocks.

Figure 3 shows the responses derived from a four-variable VAR that is modified to include hours in differences rather than levels. As above, it is convenient to begin by abstracting from

confidence interval is quite wide. For example, Figure 2 shows that the 90% confidence bounds range from contributions of 38 to 90 percent for this alternative parameterization of the RBC model. It is only when the number of observations are increased by several multiples that the confidence bands become reasonably tight (as illustrated for the case of 1000 observations using this alternative parameterization).

small-sample issues, and hence replace sample moments with the model's population moments in estimating the VAR (again we use four lags in the VAR). Our model implies that hours worked are stationary so that it might be expected that differencing hours would impair the ability of a short-ordered VAR to recover the true responses.[27] However, it still does very well in capturing the quantitative effects of a technology shock for the other variables, even though the SVAR implies some *upward* bias in the response of hours. Turning to the small sample results in Tables 3-6, there is only modest evidence of a deterioration in performance.

Figure 5 shows responses derived from alternative specifications of bivariate SVARs that include labor productivity growth and either the level of hours worked (the dashed lines) or the first difference of hours worked (the dash-dotted line). These specifications have often been utilized in the empirical literature applying the Galí methodology. The upper panel uses the population moments to derive each of the VARs (using four lags), while the lower panel reports the mean impulses derived from the Monte Carlo simulations (as in section 3, the Schwartz criterion is used to select lag length). It is clear from the upper panel that the two-variable specifications perform less adeptly than our four-variable specification in recovering the true responses: there is upward bias in the hours in levels specification, while there is pronounced downward bias for the hours in differences specification. The lower panel shows that the truncation bias is reflected in the mean bias observed in small samples.

Our results for these bivariate SVARs are quite similar to those reported by Chari, Kehoe, and McGrattan (2005). These authors also highlight (what we term) truncation bias as a significant problem for the two-variable specification, and provide an insightful discussion about the origin of this bias in the RBC model: namely, it occurs primarily because capital is omitted from the SVAR, and this variable has a significant and long-lasting influence on the dynamic responses of labor productivity growth and hours worked. They proceed to argue that the poor performance of the bivariate SVARs (especially with hours in differences) makes this approach

[27]We found that the VARMA process for the four variables in the VAR with hours in differences has a root on the unit circle so that the VARMA process is non-invertible but remains fundamental (this is also true for the two-variable specification with hours in differences considered below).

wholly unsuited to evaluating the plausibility of the RBC model relative to various alternatives, and thus provide a rebuttal to a substantial literature that has utilized SVARs to question the empirical relevance of that model. Finally, insofar as they regard the RBC model as representing a best case for the SVAR methodology, they surmise that application of the SVAR approach to other models would also be likely to yield highly imprecise estimates, and lead to mistaken inferences.

While we defer a more detailed response to the critique of Chari, Kehoe, and McGrattan (2005) to Section 5.4, we believe that our analysis that considers both a wider set of SVAR specifications and models provides a broader perspective for evaluating the SVAR approach. Clearly, our analysis of the four-variable SVAR specifications above (with hours in levels and differences) suggests that certain specifications can perform reasonably well in eliciting the true responses in the RBC model: the problem of truncation bias highlighted by Chari, Kehoe, and McGrattan (2005) is minimized in our benchmark RBC model insofar as the consumption and investment shares help proxy for the omitted capital stock.[28] Although Chari, Kehoe, and McGrattan (2005) may be justified in criticizing some of the specific empirical research critiquing RBC models, we believe the SVAR approach has much greater potential to yield informative estimates than these authors suggest.

Finally, while we show below that our four-variable VAR also performs reasonably well in the model with nominal rigidities, we caution that variable selection should be tailored to the model(s) that the researcher wishes to evaluate. In particular, our analysis suggests that if shocks other than the unit root shock to technology have a large impact on labor productivity, the ability of a low-ordered VAR to approximate the underlying VARMA process

[28]For a more detailed discussion of this issue, see Appendix 7.4. For the benchmark RBC model, we show that the R^2 statistic for a regression of the capital stock scaled by Z_t on the variables in our benchmark four-variable VAR is near one. By contrast, the R^2 for a regression of scaled capital on the variables in the bivariate VAR is very low, suggesting it is unable to capture the dynamic influence of the omitted capital stock. In interpreting these results, we caution that while the four-variable VAR performs well in the benchmark RBC model, the exogenous shocks are also model state variables, and their omission from the VAR can lead to the poor identification of technology shocks. We illustrate this possibility in Figure 6 discussed below.

may deteriorate markedly, even using our four-variable VAR specification: while the VARMA process in this case is invertible, the additional shocks contribute to a very slowly-decaying moving average component. This potential sensitivity is illustrated in Figure 6, which reports responses from a four-variable SVAR that has four lags and is derived using population moments from an alternative calibration of the RBC model that includes capital tax rate and temporary technology shocks.[29] There is a sizeable deterioration in the performance of the population SVAR in this case, with most of the divergence attributable to the temporary technology shocks.[30]

5 Sticky Price/Wage Model

In this section, we examine the robustness of our results by modifying the real business model to include nominal and real frictions that have been found useful in accounting for the observed behavior of aggregate data. These frictions include sticky wages and prices, variable capacity utilization, costs of adjustment for investment, and habit persistence in consumption. As noted above, one of the principal differences between this model and the RBC model is that hours worked decline initially in response to a technology shock rather than rise as in the RBC model. Since our sticky price/wage model is similar to Christiano, Eichenbaum, and Evans (2001) and Smets and Wouters (2003), we provide only a brief account of how it can be derived by modifying the RBC model discussed above.

[29] In this alternative calibration, the temporary technology shock contributes 50 percent of the variation to the growth rate of the Solow residual, while the parameters of the capital tax rate process are estimated using historical data (see Tables 1 and 2 for parameter estimates and selected second moments).

[30] Perhaps surprisingly, the small sample bias appears to decline noticeably relative to the benchmark RBC model, as seen in Table 3. This reflects that the upward bias in the response of labor productivity and hours evident in the population SVARs appears to be roughly offset by the small sample bias discussed in the previous section. However, the various sources of bias could reinforce each other in alternative models, contributing to a considerably more pronounced deterioration in performance than reported here.

5.1 Model Description

We assume that nominal wages and prices are set in Calvo-style staggered contracts in a framework similar to that discussed in Erceg, Henderson, and Levin (2000). The wage and price contracts have a mean duration of four quarters, and we set the wage and price markups both equal to 1/3. The inclusion of nominal rigidities into the model requires us to specify a monetary policy rule. We assume that the central bank adjusts the quarterly nominal interest rate (expressed at an annual rate) in response to the four-quarter inflation rate and to the four-quarter rate of growth of output:

$$i_t = \gamma_i i_{t-1} + \gamma_\pi \pi_t^{(4)} + \gamma_y \Delta y_t^{(4)} + \sigma_m \epsilon_{it}, \tag{20}$$

where $\pi_t^{(4)} = log(P_t/P_{t-4})$, P_t is the aggregate price level, $\Delta y_t^{(4)} = log(Y_t/Y_{t-4})$, and the monetary policy innovation, ϵ_{it}. (Note that constant terms involving the inflation target and the steady-state real interest rate have been suppressed for simplicity). Using U.S. quarterly data for the period 1983:1-2002:4, we estimated values of γ_i, γ_π, γ_y and σ_m to be 0.80, 0.60, 0.28, and 0.006, respectively.[31]

We introduce habit persistence in consumption by modifying the utility function of the representative household in the following way:[32]

$$E_t \sum_{j=0}^{\infty} \beta^j \left\{ \log\left(C_{t+j} - \phi_c \bar{C}_{t+j-1}\right) - \chi_{0t+j} \frac{N_{t+j}^{1+\chi}}{1+\chi} \right\}. \tag{21}$$

Our approach follows Smets and Wouters (2003), among others, by assuming that an individual cares about his consumption relative to the lagged value of aggregate consumption, $\bar{C}_t$. We set $\phi_c = 0.6$, close to the mean estimate of Smets and Wouters (2003).

We incorporate variable capacity utilization into the sticky price/wage model so that variation in the Solow residual reflects both changes in technology and movements in the unob-

[31] We estimated equation (20) using instrumental variables where our instruments included lags of output growth and inflation.

[32] For simplicity, we suppress that our utility function depends on real money balances in a separable fashion. With monetary policy specified by an interest rate rule and money separable in utility, the equilibrium dynamics of our model can be determined independently of the quantity of money.

served level of capacity utilization in response to all of the underlying shocks. The production function modified to include variable capacity utilization, u_t, is given by:

$$Y_t = (u_t K_t)^{\theta}((Z_t V_t)N_t)^{1-\theta}, \tag{22}$$

where Z_t and V_t are the unit-root and temporary shocks to technology described earlier.

In our decentralized economy, households rent capital services ($u_t K_t$) to firms, and choose how intensively the capital is utilized. We follow Christiano, Eichenbaum, and Evans (2001) and assume that households pay a cost to varying u_t in units of the consumption good. These adjustment costs alter the budget constraint of the representative household as follows:

$$\begin{gathered} C_t + I_t + \tfrac{1}{1+r_t}B_{t+1} - B_t = \\ (1-\tau_{Nt})W_t N_t + \Gamma_t + T_t + (1-\tau_{Kt})R_{Kt}u_t K_t + \tau_{Kt}\delta K_t - \tfrac{\phi_i}{2}\tfrac{(I_t - I_{t-1})^2}{I_{t-1}^2} - \nu_0 \tfrac{u_t^{1+\nu}}{1+\nu}. \end{gathered} \tag{23}$$

In the above, the term, $\nu_0 \frac{u_t^{1+\nu}}{1+\nu}$, reflects the cost of adjusting the utilization rate, where ν_0 is normalized so that $u_t = 1$ in non-stochastic steady state and ν is set to 0.01, as in Christiano, Eichenbaum, and Evans (2001). Equation (23) also reflects the addition of adjustment costs for investment, and in our calibration, we set $\phi_i = 2$.[33]

As in our benchmark calibration of the RBC model, our benchmark calibration of the sticky price/wage model abstracts from capital tax rate and temporary technology shocks by setting $\sigma_{\tau_K} = \sigma_V = 0$. We used the method of moments to estimate the innovation variances of the permanent technology shock (0.0152) and the labor supply shock (0.069) by exactly matching the model's implications for the volatility of the Solow residual growth rate and the standard deviation of (HP-filtered) hours worked relative to output to their sample counterparts. For the other model parameters, shown in Table 1, we used the same values as for the RBC model.

[33] This is lower than the value of around 4 for ϕ_i used by Christiano, Eichenbaum, and Evans (2001), who estimated ϕ_i based on the response of investment to a monetary shock. However, we found that low values of ϕ_i (less than one) were necessary for our sticky price and wage model to account for the unconditional volatility of investment relative to output. Our choice of ϕ_i is an intermediate one between the values implied by these calibration procedures.

5.2 Estimation Results

Figure 7 exhibits the response of labor productivity, hours worked, consumption, investment, and output to a technology shock for the benchmark sticky price/wage model. In each panel the solid lines show the true responses from the DGE model. In the same panels, the dashed lines show the mean responses from the SVAR derived from Monte Carlo simulations (as described in Section 3). As in the case of the benchmark RBC model, the mean response of each of these variables has the same sign and qualitative pattern as the true response. Moreover, as suggested by the pointwise confidence intervals, the SVAR is likely to correctly imply a rise in labor productivity, consumption, and output in response to the technology shock. The SVAR is also likely to capture the initial decrease in hours worked that occurs following a technology shock.[34] Both the mean response and the 90% confidence intervals fall below zero in the two periods following the shock, in line with the model's response.

As in the case of the RBC model, the SVAR does not perform as well quantitatively. The mean responses underestimate the true responses of labor productivity, output, consumption, and investment by roughly 30-35 percent (see Table 3). This downward bias helps account for the substantial probability of making large errors in estimating these variables, as shown in Tables 4-6. Overall, the probability of making a large error in estimating most of the variables seems commensurate with that of the RBC model, with the exception of investment. Interestingly, we found that while the probability of estimating a response of labor productivity, output, or consumption that was uniformly positive for four quarters following the shock exceeded 90%, there was only a 63% chance of estimating a uniformly positive response of investment. Thus, in this model, there appears to be considerably more qualitative uncertainty about the effects of a technology shock on investment.

The bottom left panel of Figure 2 shows the cumulative distribution function derived from Monte Carlo simulations of the estimator of the contribution of technology shocks to

[34] As in Vigfusson (2004) and Francis and Ramey (2003), the real frictions play an important role in accounting for the model's implication of a fall in hours. Thus, the initial fall in hours in the sticky price/wage model occurs for a fairly wide set of reasonable monetary policy rules.

explaining variations in HP-filtered output. The 90% confidence bands for the estimator include contributions ranging from 7 to 80 percent. Therefore, as in the benchmark RBC model, the Galí identification scheme provides little information about the importance of technology shocks in explaining output fluctuations at business cycle frequencies.

We next examine the sources of bias in the mean responses using the same analytical framework that was applied to the RBC model. The dashed lines in the top panel of Figure 8 show the responses to a technology shock derived from a SVAR with four lags that uses the model's population moments. While these responses diverge slightly from the true responses, it is clear that a short-ordered population VAR performs well in approximating the true VARMA process. Accordingly, as in the benchmark RBC model, most of the bias in the estimated impulse responses is attributable to the small-sample problems emphasized by Faust and Leeper (1997).

The small-sample bias in this model depends on many of the same model characteristics as identified using the RBC model. In particular, the bias arises because the identification scheme has difficulty disentangling unit root technology shocks from other shocks that may have highly persistent effects on labor productivity, and because of slow capital adjustment (the latter plays less of a role in accounting for bias, since variable capacity utilization induces labor productivity to rise more quickly to its long-run level). As shown in Table 3, the bias is reduced when we decrease the persistence of the non-technology shocks and accelerate capital adjustment by setting $\delta = 0.9$; however, the change in the bias is somewhat less dramatic than in the RBC model.

5.3 Sensitivity Analysis in the Sticky Price/Wage Model

We next investigate the sensitivity of our results to including a different set of variables in the SVAR, to differencing hours worked, and to adding capital tax rate and temporary technology shocks.[35]

[35] For all these experiments, we checked that the VARMA process implied by the variables in the SVAR was a fundamental representation.

The bottom panels of Figure 8 show results for the two bivariate population SVARs considered in Section 4.2 (i.e., each SVAR has four lags and is derived by replacing sample moments with corresponding population moments). The dashed lines show the responses for the SVAR with labor productivity growth and hours in levels, while the dash-dotted lines show the responses of the alternative specification with hours in differences. Notably, in stark contrast with their performance in the RBC model, each specification does very well in accounting for the short-run response of hours worked. The bivariate SVARs also perform quite well in small samples. For example, Figure 9 illustrates the responses derived from estimating the bivariate specification with hours in differences. The mean response of hours lies very close to the true response in the short-run and the confidence intervals are somewhat narrower than in the four-variable specification with hours in levels (see Figure 7).

Thus, the model with nominal rigidities provides an interesting example of a model that could rationalize the use of the bivariate SVAR in estimation (as often employed in the empirical literature). These results should help dispel the presumption of Chari, Kehoe, and McGrattan (2005) that the RBC model presents a best case for the SVAR approach. The considerably improved performance of the bivariate VAR in this model may indeed seem surprising, insofar as it includes a much larger set of endogenous state variables than the RBC model (e.g., lagged consumption, investment, and the real wage). However, it turns out that the dynamics of labor productivity growth and hours worked in this model are simple enough that they can be more easily approximated by a short-ordered VAR. This reflects the inclusion of variable capacity utilization in the sticky price/wage model, which allows firms to vary their effective capital stock, $u_t K_t$, in response to shocks. As a result, the capital stock, K_t, has a diminished influence on the dynamics of labor productivity and hours relative to the RBC model (this is particularly evident in the response of labor productivity, which reaches its long-run level much more quickly than in the RBC model). It also reflects that the additional state variables such as consumption and investment exert only a small influence on the dynamics of labor productivity, and their effect on hours is fairly transient.[36]

[36]This is evident from analyzing the numerical state-space solution for the log-linearized model.

Overall, we find relatively little sensitivity of our results derived from this model to the transformation applied to hours worked in the SVAR (levels vs. differences). This applies both to our four-variable specification (shown in Figure 8 using population moments, with small sample results in Tables 3-6) and to the bivariate SVARs. Thus, the pronounced sensitivity to the transformation of hours evident in the bivariate SVARs derived from the RBC model appears exceptional among the cases we consider; and to the extent we do see some sensitivity, there is no clear pattern of bias in the hours worked variable (e.g., in the sticky price/wage model, differencing hours implies upward bias in the hours response for both the two and four-variable SVAR specifications, in sharp contrast with the RBC model, where hours is downward biased in the bivariate specification). To the extent that empirical results appear noticeably more sensitive to the transformation of hours, this may reflect other factors not captured in either of our models, e.g. demographic shifts or other shocks to labor force participation.[37]

We also find that the sticky price/wage model is somewhat less sensitive to the inclusion of the additional shocks than the RBC model.[38] Using the population moments of the sticky price/wage model with these additional shocks, the four-variable SVAR with four lags displays a deterioration (not shown) in its ability to approximate the true dynamics; however, this deterioration is less pronounced than in the RBC model. Similarly to the RBC model, there is no evident deterioration in small sample performance in this case (see Tables 3-6).

Taking stock of our sensitivity analysis across the two models, our results suggest that specification choice should be suited to the particular use or interpretation to be attached to the results, and to the researcher's beliefs about the plausibility of alternative models. Thus,

[37]See Francis and Ramey (2004), who showed that the standard measure of hours per capita used in the empirical literature is significantly affected by low frequency demographic and institutional trends, which are not accounted for in most business cycle models. They constructed a revised measure of hours per capita, which they argue is better suited to these theoretical models, and using bivariate SVARs showed that a positive technology shock leads to a fall in hours irrespective of whether hours per capita are specified in levels or differences.

[38]We calibrated these two additional shocks following the same approach discussed above for the RBC model. Thus, the innovation variance of the stationary technology shock accounts for 50 percent of the variation in the growth rate of the Solow residual (see Tables 1 and 2 for details).

a researcher exclusively interested in estimating the effects of technology shocks who had high confidence in the model with nominal rigidities might find it desirable to use a bivariate VAR specification with hours in differences, even if this specification performed relatively poorly in the RBC model. By contrast, a researcher interested in evaluating the plausibility of alternative models would presumably want to adopt a SVAR specification that performed well enough across models to help differentiate between them. As one might conjecture and as we verify below, the four-variable SVAR would appear to offer a useful basis for discriminating between the implications of the two benchmark models we have examined.

5.4 Discriminating Between Models Based on the Response of Hours

A key objective of Galí's (1999) seminal paper applying the SVAR approach to technology shocks was to differentiate between alternative business cycle models. In this vein, Galí interpreted his result that hours worked fell in response to a technology shock as contravening the RBC paradigm, and suggested that his findings might be more consistent with a model that incorporated nominal rigidities. Galí's provocative conclusion generated considerable subsequent empirical research examining the robustness of inferences drawn from SVARs to changes in data and specification. This research has included papers by Francis and Ramey (2003) and Galí and Rabanal (2005), that broadly lent support to Galí's original conclusion, while Christiano, Eichenbaum, and Vigfusson (2003) favored a VAR specification that implies a rise in hours worked following a technology shock. But notwithstanding the lively debate that has emerged about specification issues, these papers share the common thread that they regard the SVAR approach as a useful methodological approach in helping to discriminate between alternative business cycle models.

The recent paper by Chari, Kehoe, and McGrattan (2005) diverges from this literature in its complete rejection of the use of the SVAR approach to evaluate the plausibility of alternative models. As noted above, these authors (hence CKM) concluded based on the performance of bivariate SVARs estimated using artificial data from an RBC model that inferences from the SVAR approach are likely to be uninformative or misleading.

We are certainly sympathetic with CKM's specific point that the SVAR methodology may perform poorly even if the data-generating process satisfies Galí's headline identifying assumptions. Complementary to our own analysis, the CKM results should provide strong caution against interpreting the results of SVARs in a model-independent fashion: inferences about technology shocks, and tests of alternative models using the SVAR approach, are invariably more model-specific than has been recognized in the literature. As applied to the literature, we concur with CKM that at least some of the evidence brought to bear against the RBC model, such as that derived from a two-variable SVAR with hours in differences, probably can be dismissed.

But do we agree with CKM that the SVAR methodology is completely ill-suited to discriminating between alternative models of the business cycle? Our answer is an emphatic "no." To the contrary, we interpret our results as suggesting that the SVAR approach may indeed be a useful tool in this regard, provided that the models have sufficiently divergent implications about the effects of technology shocks on the labor market, and that the SVAR performs reasonably well in each model.

We illustrate this by assessing the ability of the SVAR to discriminate between our two benchmark models based on the response of hours worked. We use a four-variable SVAR with hours in levels, since we have shown that it performs reasonably well in both the RBC model, and in the model with nominal rigidities. The upper panel of Figure 10 shows the probabilities that the estimated response of hours is uniformly negative in the first two and four quarters, respectively. The probability of finding an initial decline in hours that persists for two quarters is 93 percent in the model with nominal rigidities, but only 26 percent in the RBC model. Accordingly, a researcher who found that hours worked declined after a positive innovation in technology in the data could reasonably interpret this finding as providing some evidence in favor of the sticky price/wage model. By contrast, a researcher who found that hours worked rose after a technology shock could regard this finding as offering evidence in support of the RBC model: as shown in the lower panel, the probability of finding an initial rise in hours that persists for two quarters is 71 percent in the RBC model, but less than 1 percent in the sticky price/wage model.

Interestingly, fairly similar results obtain for a two-variable VAR specification in which hours is specified in levels. For example, for this specification we find that the probability of finding an initial decline in hours that persists for two quarters is 87 percent in the model with nominal rigidities, but only 23 percent in the RBC model. Ironically, while there is admittedly some upward bias in the two-variable SVAR when applied to the RBC model – a feature which CKM highlight in pejorative terms – such bias may actually facilitate differentiating between models. This is because the bias induces a larger wedge between the implications of these models for hours worked, making it less probable that an observed decline in hours would be consistent with the RBC model.

Several recent papers have in fact examined the implications of bivariate SVARs with hours in levels, including Galí (2004), Francis and Ramey (2004), and Christiano, Eichenbaum, and Vigfusson (2003). The empirical results appear quite sensitive to measurement of the hours worked variable, with no consensus on the most appropriate measure (e.g., Galí and Rabanal and Francis and Ramey find that hours decline for most of their measures, while Christiano, Eichenbaum, and Vigfusson report a rise in hours for their preferred measure). While identifying the proper empirical counterpart to the hours concept in our theoretical model is an issue beyond the scope of the present paper, our analysis does provide a rationale for using such a SVAR specification; and it would appear to offer a much better test of the RBC model than SVAR specifications involving hours in differences.

From a more general perspective, our analysis suggests that the SVAR methodology may offer a plausible means of differentiating between alternative business-cycle models, as envisioned in Galí's original article and in most of the subsequent empirical literature. However, it is important that a researcher is aware of the limitations of alternative SVAR specifications in evaluating the models considered, and that he adopt a specification that is well-suited to differentiate between the particular models of interest.

6 Conclusion

While identifying technology shocks and their effects is a difficult task, our analysis suggests that Galí's methodology is a useful tool. We find it encouraging that our four-variable VAR specification performs reasonably well across the RBC and sticky price/wage models in characterizing the qualitative effects of a technology shock on a range of macro variables. But our analysis highlights that the conditions under which the Galí methodology performs well appear considerably more restrictive than implied by the key identifying restriction. Accordingly, it will be useful in future research to delineate further the class of models for which this methodology works well, and also to examine empirically realistic conditions that might exacerbate some of the problems we have identified in our analysis (e.g., stationary technology shocks). Moreover, it will be beneficial to identify VAR specifications that appear to be robust across a class of plausible models, insofar as this would enhance the latitude to use this methodology in discriminating across models.

References

Altig, D., L. J. Christiano, M. Eichenbaum, and J. Lindé (2003). Technology Shocks and Aggregate Fluctuations. Mimeo, Northwestern University.

Anderson, G. and G. Moore (1985). A Linear Algebraic Procedure for Solving Linear Perfect Foresight Models. *Economic Letters 17*, 247–52.

Blanchard, O. J. and C. M. Kahn (1980). The Solution of Linear Difference Models under Rational Expectations. *Econometrica 48*(5), 1305–1312.

Blanchard, O. J. and D. Quah (1989). The Dynamic Effects of Aggregated Demand and Supply Disturbances. *American Economic Review 79*(4), 655–673.

Chari, V., P. J. Kehoe, and E. R. McGrattan (2005). A Critique of Structural VARs Using Real Business Cycle Theory. Federal Reserve Bank of Minneapolis, Working Paper No 631.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2001). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. NBER Working Paper No. w8403.

Christiano, L. J., M. Eichenbaum, and R. Vigfusson (2003). What Happens after a Technology Shock. NBER Working Paper No.w9819.

Christiano, L. J. and J. D. Fisher (1995). Tobin's q and Asset Returns: Implications for Business Cycle Analysis. NBER Working Paper No.5292.

Cogley, T. and J. M. Nason (1995). Output Dynamics in Real-Business-Cycle Models. *American Economic Review 85*(3), 492–462.

Cooley, T. and M. Dwyer (1998). Business Cycle Analysis without much Theory: A Look at Structural VARs. *Journal of Econometrics 83*, 57–88.

Erceg, C., L. Guerrieri, and C. Gust (2004). Can Long-Run Restrictions Identify Technology Shocks? International Finance Discussion Paper 792.

Erceg, C. J., D. W. Henderson, and A. T. Levin (2000). Optimal Monetary Policy with Staggered Wage and Price Contracts. *Journal of Monetary Economics 46*, 281–313.

Evans, C. L. (1992). Productivity Shocks and Real Business Cycles. *Journal of Monetary Economics 29*, 191–208.

Faust, J. and E. M. Leeper (1997). When Do Long-Run Identifying Restrictions Give Reliable Results? *Journal of Business and Economic Statistics 15*(3), 345–353.

Fisher, J. D. M. (2002). Technology Shocks Matter. Mimeo, Federal Reserve Bank of Chicago.

Francis, N. and V. A. Ramey (2003). Is the Technology-Driven Real Business Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited. Mimeo, University of California at San Diego.

Francis, N. and V. A. Ramey (2004). A New Measure of Hours Per Capita with Implications for the Technology-Hours Debate. Mimeo, University of California at San Diego.

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

Galí, J. (2004). On the Role of Technology Shocks as a Source of Business Cycles: Some New Evidence. *Journal of the European Economic Association 2*, 372–380.

Galí, J. and P. Rabanal (2005). Technology Shocks and Aggregate Fluctuations: How Well Does the Real Business Cycle Model Fit Postwar U.S. Data? In M. Gertler and K. Rogoff (Eds.), *NBER Macroeconomics Annual 2004*, pp. 225–288. MIT Press.

Hall, R. E. (1997). Macroeconomic Fluctuations and the Allocation of Time. *Journal of Labor Economics 15*(1), S223–S250.

Hansen, L. and T. Sargent (2004). Recursive Models of Dynamic Linear Economies. Mimeo, New York University.

Jones, J. B. (2002). Has Fiscal Policy Helped Stabilize the Postwar U.S. Economy? *Journal of Monetary Economics 49*, 709–746.

Killingsworth, M. and J. Heckman (1986). Female Labor Supply: A Survey. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics*, Volume 1 of *Handbooks in Economics Series*, pp. 103–204. Elsevier Science B.V.

King, R. G., C. I. Plosser, and S. Rebelo (1988). Production, Growth and Business Cycles. *Journal of Monetary Economics 21*, 195–232.

King, R. G., C. I. Plosser, J. H. Stock, and M. W. Watson (1991). Stochastic Trends and Economic Fluctuations. *American Economic Review 81*, 819–840.

Lippi, M. and L. Reichlin (1993). The Dynamic Effects of Aggregate Supply and Demand Disturbances: Comment. *American Economic Review 83*, 644–652.

Lutkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.

Pagan, A. and J. Robertson (1998). Structural Models of the Liquidity Effect. *Review of Economics and Statistics 80*, 202–217.

Parkin, M. (1988). A Method for Determining Whether Parameters in Aggregative Models are Structural. *Carnegie-Rochester Conference Series on Public Policy 29*, 215–252.

Pencavel, J. (1986). Labor Supply of Men: A Survey. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics*, Volume 1 of *Handbooks in Economics Series*, pp. 3–102. Elsevier Science B.V.

Pencavel, J. (2002). A Cohort Analysis of the Association Between Work and Wages Among Men. *The Journal of Human Resources 37*(2), 251–274.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics 6*(2), 461–464.

Shapiro, M. and M. Watson (1988). Sources of Business Cycle Fluctuations. *NBER Macroeconomics Annual*, 111–148.

Smets, F. and R. Wouters (2003). An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association 1*(5), 1124–1175.

Vigfusson, R. J. (2004). The Delayed Response To A Technology Shock. A Flexible Price Explanation. International Finance Discussion Papers, Number 810, Board of Governors of the Federal Reserve System.

Table 1: Parameters Values Common Across Calibrated Versions of Model*

$\beta = 1.03^{-0.25}$	$\tau_K = 0.38$
$\chi_0 = 1$	$\rho_g = 0.98$
$\chi = 1.5$	$\sigma_g = 0.003$
$\delta = 0.02$	$\rho_{\tau_N} = 0.98$
$\theta = 0.35$	$\sigma_{\tau_N} = 0.0052$
$\mu_z = 0.0037$	$\rho_\chi = 0.95$
$g/y = 0.20$	$\rho_{\tau_K} = 0.97$
$\tau_N = 0.22$	$\rho_V = 0.95$

* g/y denotes the steady state value of the ratio of government consumption to output.

Table 2: Selected Moments and Parameter Values of Calibrated Versions of Model[a]

Moment	U.S. Data[b]	Real Business Cycle			Sticky Prices and Wages	
		$\sigma_\chi = 0$	Benchmark	with Additional Shocks	Benchmark	with Additional Shocks
σ_y	2.17	1.38	1.72	1.67	2.00	1.82
σ_h/σ_y	0.80	0.28	0.80	0.80	0.80	0.80
σ_c/σ_y	0.47	0.73	0.61	0.56	0.78	0.72
σ_i/σ_y	2.91	1.98	2.26	2.55	2.35	2.53
$\sigma_{\Delta S}$	0.96	0.96	0.96	0.96	0.96	0.96
Parameter Values						
σ_z		0.0148	0.0148	0.0104	0.0152	0.0103
σ_χ		0	0.024	0.0198	0.0619	0.0335
σ_{τ_K}		0	0	0.008	0	0.008
σ_V		0	0	0.0103	0	0.0102

[a] All moments except $\sigma_{\Delta S}$ were computed by first transforming the data using the HP-filter (with $\lambda = 1600$). $\sigma_{\Delta S}$ refers to the standard deviation of the growth rate of the Solow residual.
[b] σ_y and σ_h were computed using BLS data on nonfarm business sector output and hours from 1958-2002. σ_c/σ_y and σ_i/σ_y were taken from Christiano and Fisher (1995) who used DRI data from 1947-1995.

Table 3: Distance Between Mean Estimates and True Impulse Responses[a]

Experiment	Labor Productivity	Output	Hours	Consumption	Investment
RBC Model	0.40	0.25	0.09	0.28	0.19
with $\sigma_z = 1/3X$	0.49	0.02	0.34	0.11	0.21
with lower persistence[b]	0.17	0.16	0.01	0.16	0.15
with lower persistence[b] and $\delta = 0.9$	0.10	0.10	0.00	0.09	0.10
with hours differenced	0.24	0.33	0.08	0.30	0.34
with additional shocks[c]	0.22	0.02	0.14	0.13	0.15
Sticky Price/Wage Model	0.34	0.33	0.05	0.34	0.32
with lower persistence[b] and $\delta = 0.9$	0.19	0.20	0.03	0.20	0.30
with hours differenced	0.38	0.37	0.06	0.37	0.37
with additional shocks[c]	0.29	0.29	0.05	0.29	0.29

[a] Absolute value of percent difference between mean estimated response and true model response averaged over first twelve periods. For hours worked, we report the absolute value of the difference from the true model response.
[b] Lower persistence refers to the case where AR(1) parameters of non-technology shocks are set to half the benchmark values.
[c] The additional shocks are capital tax rate and temporary technology shocks.

Table 4: Probability that Estimated Response is Uniformly Far From True Response Over First Two Quarters[a]

Experiment	Labor Productivity	Output	Hours	Consumption	Investment
RBC Model	0.48	0.27	0.26	0.31	0.35
with $\sigma_z = 1/3X$	0.61	0.67	0.28	0.52	0.78
with lower persistence[b]	0.10	0.08	0.16	0.04	0.15
with lower persistence[b] and $\delta = 0.9$	0.03	0.08	0.37	0.07	0.10
with hours differenced	0.27	0.39	0.42	0.34	0.44
with additional shocks[c]	0.34	0.40	0.23	0.26	0.61
Sticky Price/Wage Model	0.35	0.31	0.02	0.36	0.79
with lower persistence[b] and $\delta = 0.9$	0.12	0.16	0.00	0.15	0.70
with hours differenced	0.41	0.35	0.03	0.41	0.80
with additional shocks[c]	0.38	0.35	0.10	0.30	0.86

[a] For all variables except hours worked, the probability that the estimated response lies at least 33% above or below the true response for the first two quarters. For hours worked, the probability that the sign of the estimated response is incorrect in each of the first two quarters.
[b] Lower persistence refers to the case where AR(1) parameters of non-technology shocks are set to half the benchmark values.
[c] The additional shocks are capital tax rate and temporary technology shocks.

Table 5: Probability that Estimated Response is Uniformly Far From True Response Over First Four Quarters[a]

Experiment	Labor Productivity	Output	Hours	Consumption	Investment
RBC Model	0.43	0.24	0.23	0.28	0.28
with $\sigma_z = 1/3X$	0.54	0.58	0.25	0.44	0.70
with lower persistence[b]	0.05	0.03	0.04	0.03	0.05
with lower persistence[b] and $\delta = 0.9$	0.02	0.04	0.20	0.04	0.05
with hours differenced	0.22	0.35	0.38	0.30	0.39
with additional shocks[c]	0.30	0.31	0.21	0.22	0.51
Sticky Price/Wage Model	0.31	0.26	0.02	0.32	0.71
with lower persistence[b] and $\delta = 0.9$	0.10	0.12	0.00	0.13	0.30
with hours differenced	0.37	0.30	0.03	0.38	0.71
with additional shocks[c]	0.34	0.30	0.07	0.28	0.78

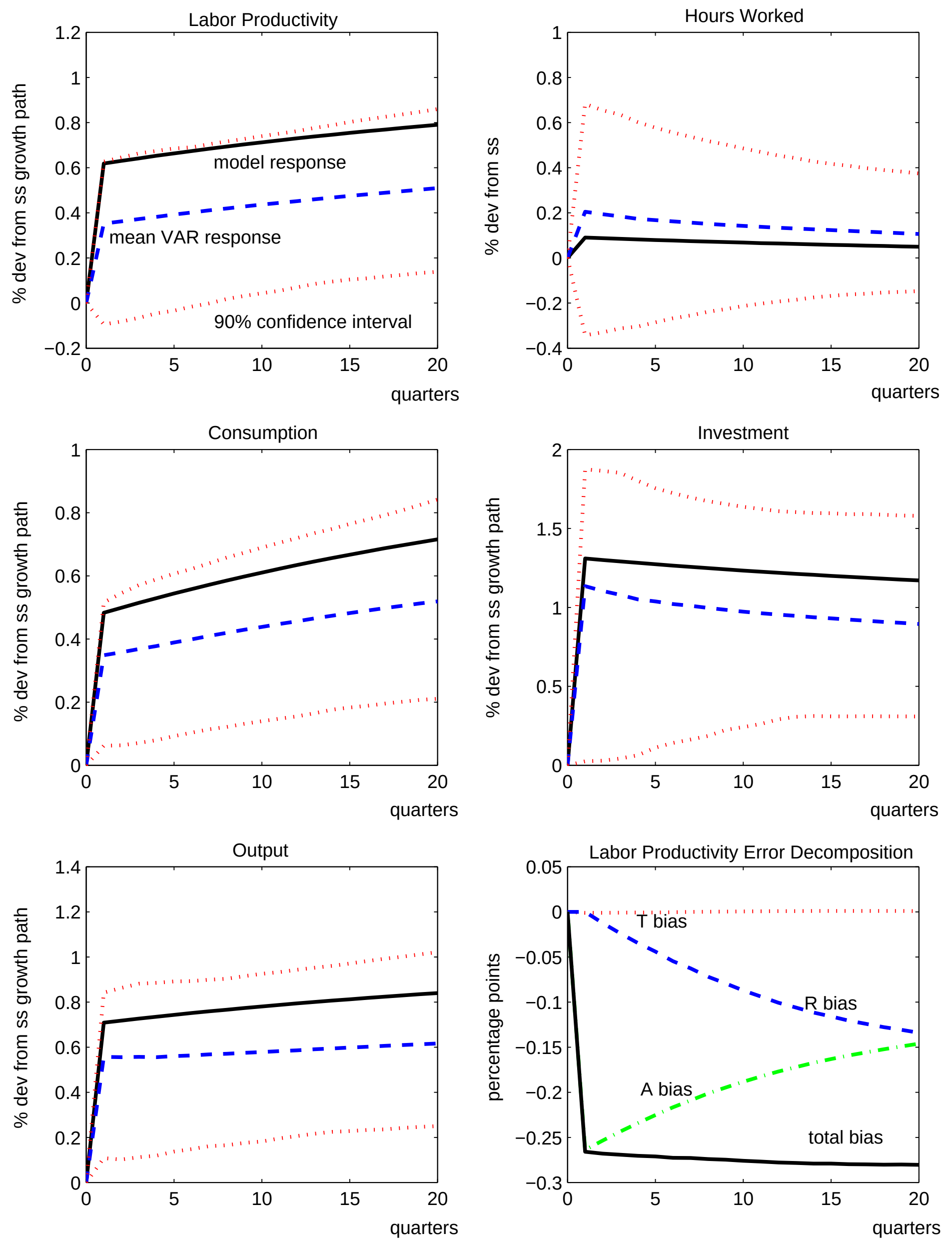
[a] For all variables except hours worked, the probability that the estimated response lies at least 33% above or below the true response for the first four quarters. For hours worked, the probability that the sign of the estimated response is incorrect in each of the first four quarters.
[b] Lower persistence refers to the case where AR(1) parameters of non-technology shocks are set to half the benchmark values.
[c] The additional shocks are capital tax rate and temporary technology shocks.

Table 6: Probability that Estimated Response is Uniformly Far From True Response Over First Twelve Quarters[a]

Experiment	Labor Productivity	Output	Hours	Consumption	Investment
RBC Model	0.36	0.17	0.16	0.22	0.16
with $\sigma_z = 1/3X$	0.40	0.36	0.19	0.32	0.46
with lower persistence[b]	0.02	0.01	0.00	0.01	0.01
with lower persistence[b] and $\delta = 0.9$	0.01	0.01	0.06	0.01	0.01
with hours differenced	0.13	0.29	0.32	0.25	0.31
with additional shocks[c]	0.22	0.19	0.17	0.15	0.30
Sticky Price/Wage Model	0.25	0.23	NA	0.25	0.61
with lower persistence[b] and $\delta = 0.9$	0.05	0.07	NA	0.07	0.15
with hours differenced	0.30	0.26	NA	0.30	0.61
with additional shocks[c]	0.24	0.25	NA	0.22	0.63

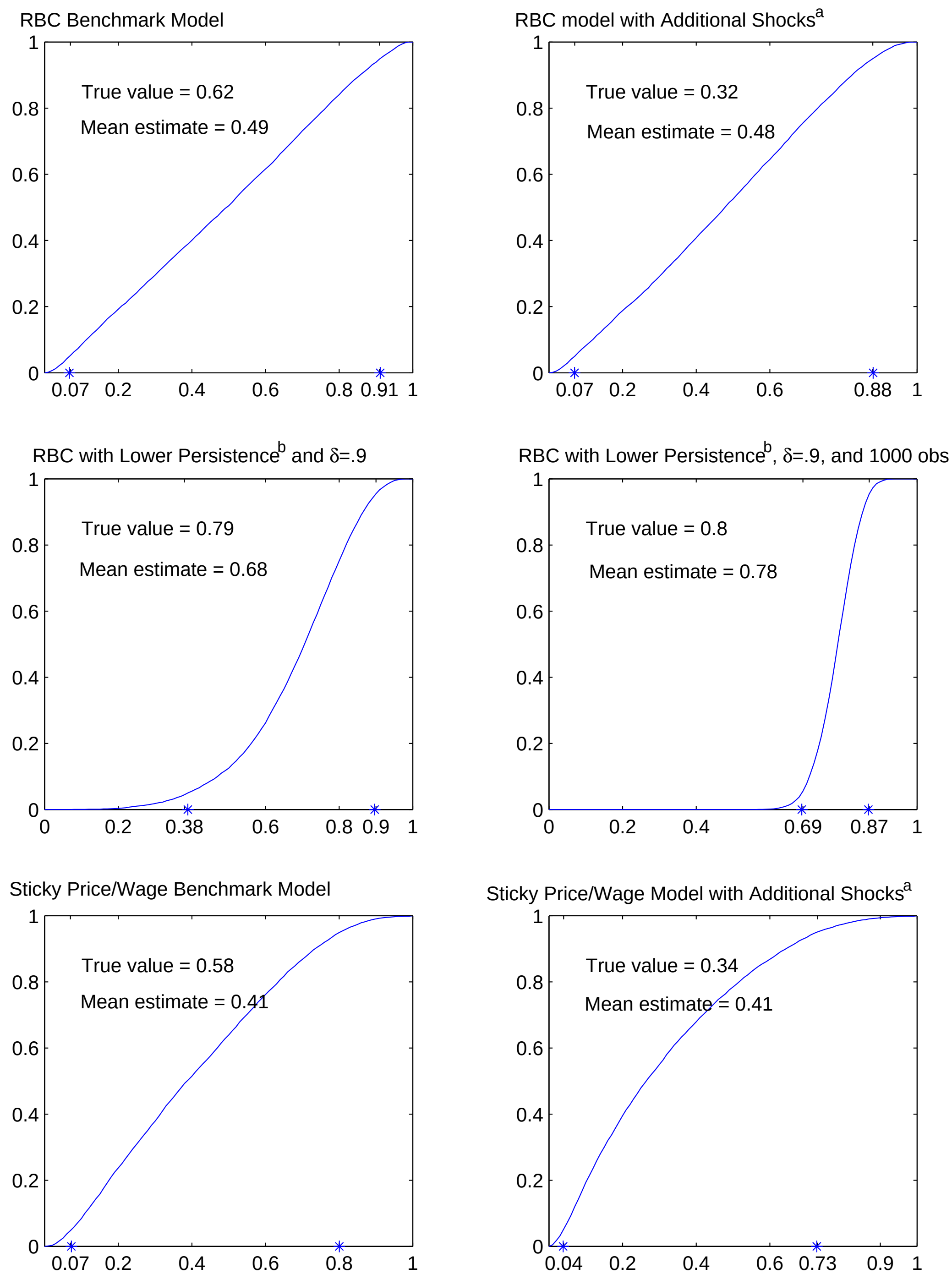
[a] For all variables except hours worked, the probability that the estimated response lies at least 33% above or below the true response for the first twelve quarters. For hours worked, the probability that the sign of the estimated response is incorrect in each of the first twelve quarters. In the sticky price/wage model, this probability is not reported as the model response changes its sign after five quarters.
[b] Lower persistence refers to the case where AR(1) parameters of non-technology shocks are set to half the benchmark values.
[c] The additional shocks are capital tax rate and temporary technology shocks.

Figure 1: Responses to Technology Shocks in the Benchmark RBC Model*



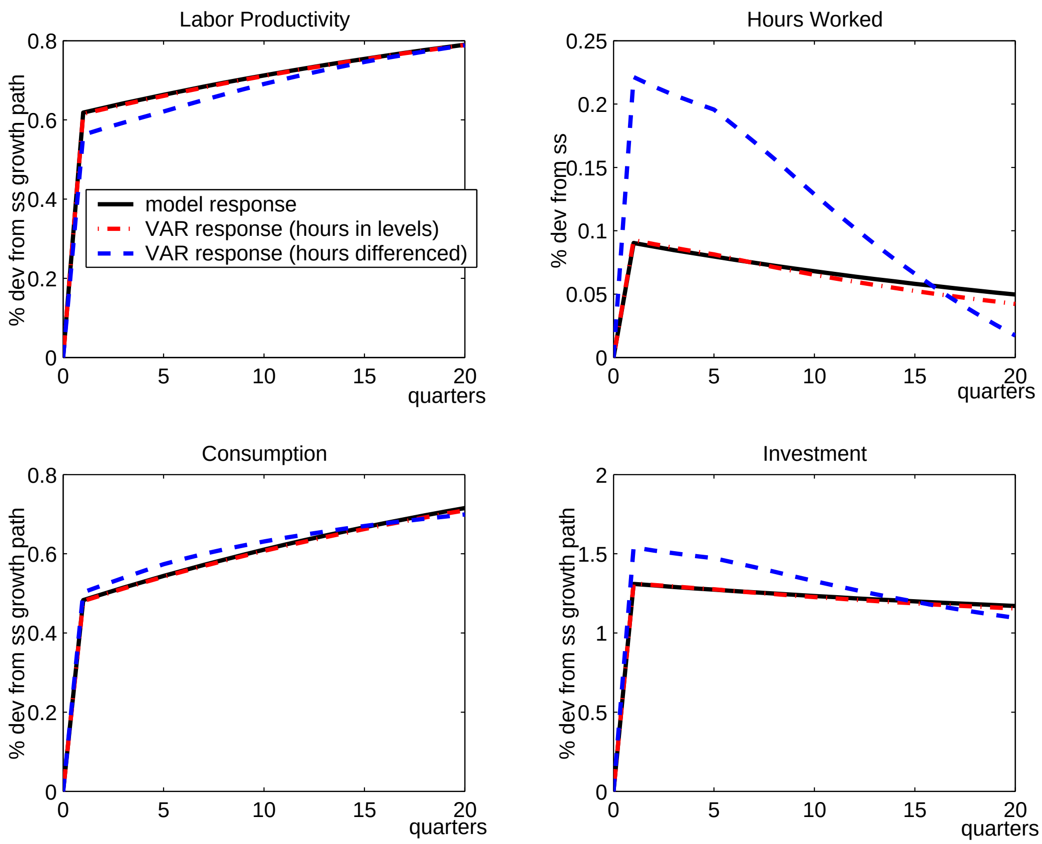
* VAR results based on 10,000 samples of 180 quarterly observations. In the lower right panel, T bias refers to bias that arises from approximating the true VARMA process with a VAR of order 4. The R bias reflects small-sample bias from estimating the reduced-form VAR. The A bias reflects small-sample bias associated with the transformation of the reduced-form to the structural form.

Figure 2: Estimated Cumulative Distribution Functions for the Contribution of Unit-Root Technology Shocks to HP-Filtered Output Variation*



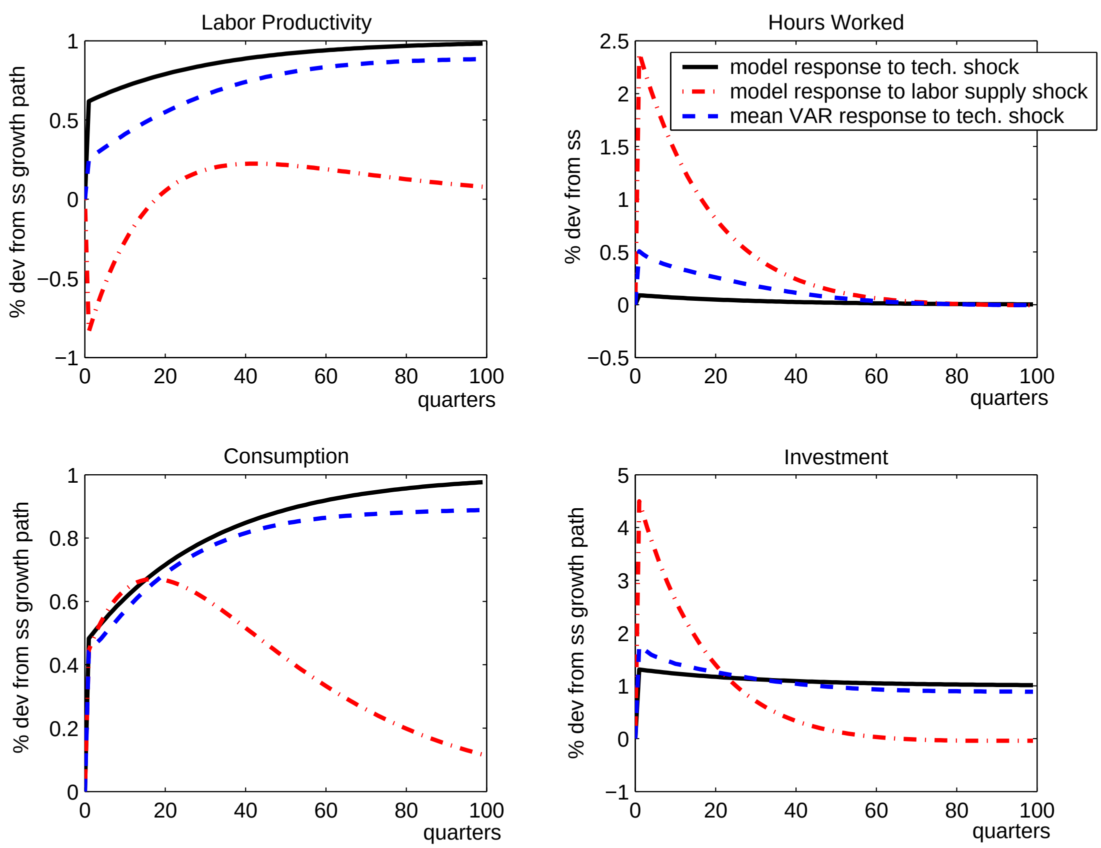
* The star symbols on the charts' abscissae denote the bounds of the 90% confidence intervals.
[a] The additional shocks are capital tax rate and temporary technology shocks.
[b] Lower persistence refers to the case where AR(1) parameters of non-technology shocks are set to half the benchmark values.

Figure 3: Responses to a Technology Shock in the Benchmark RBC Model Using Population Moments*



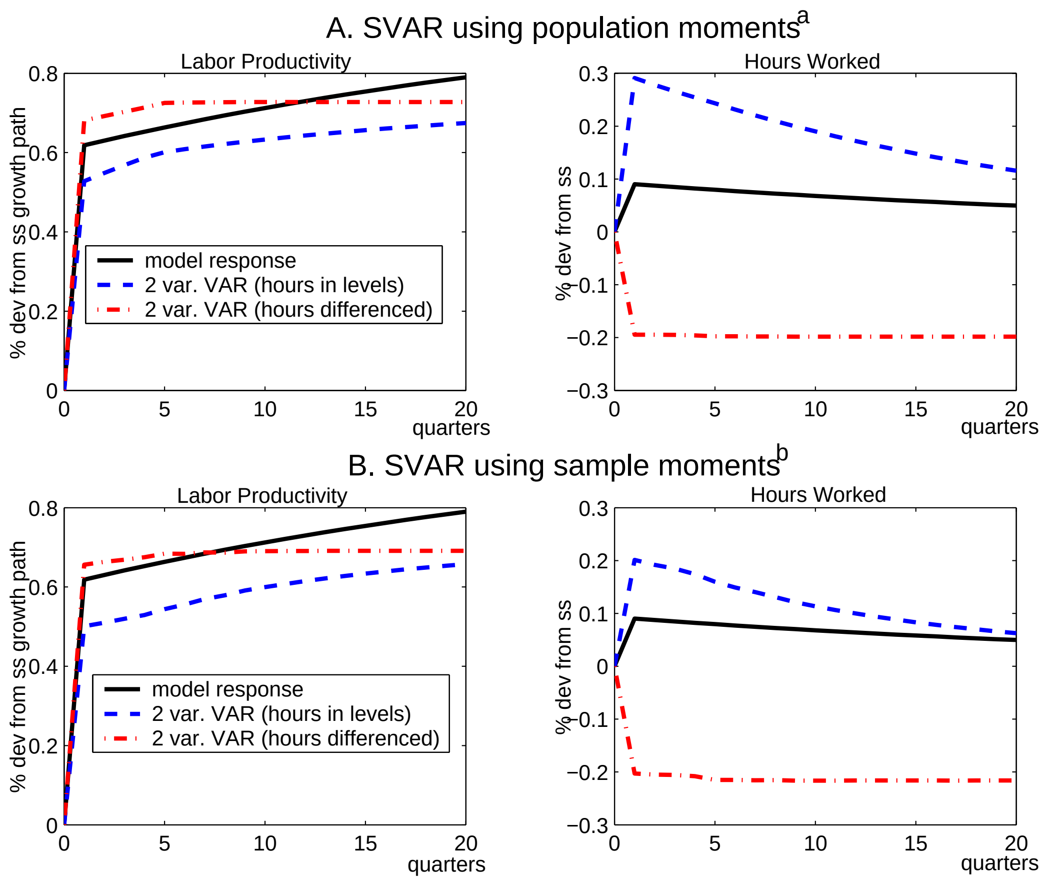
* Results based on VARs of order 4 estimated with population moments.

Figure 4: Responses to Technology and Labor Supply Shocks in the RBC Model*



* VAR results based on 10,000 samples of 180 quarterly observations using the RBC model with smaller technology shocks ($\sigma_z = 0.0049$).

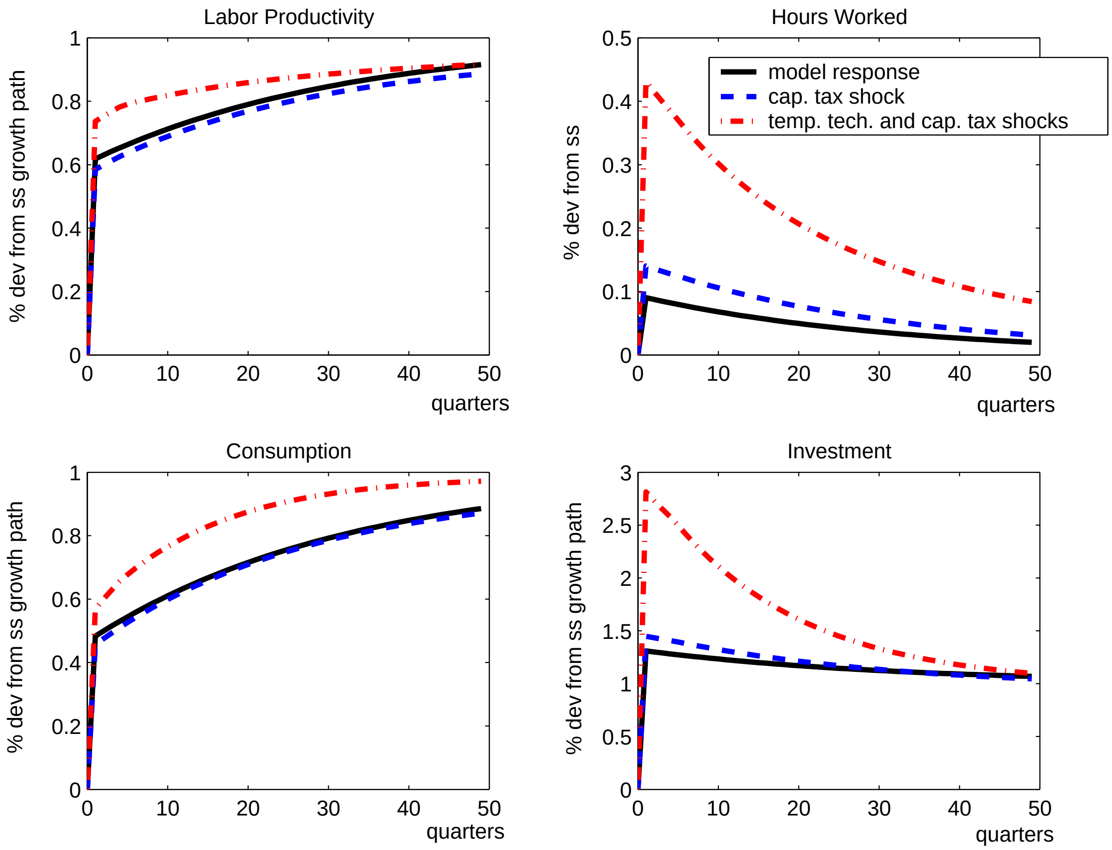
Figure 5: Responses to a Technology Shock in Benchmark RBC Model for Bivariate VAR Specifications*



[a] Results based on VARs of order 4 estimated with model's population moments.

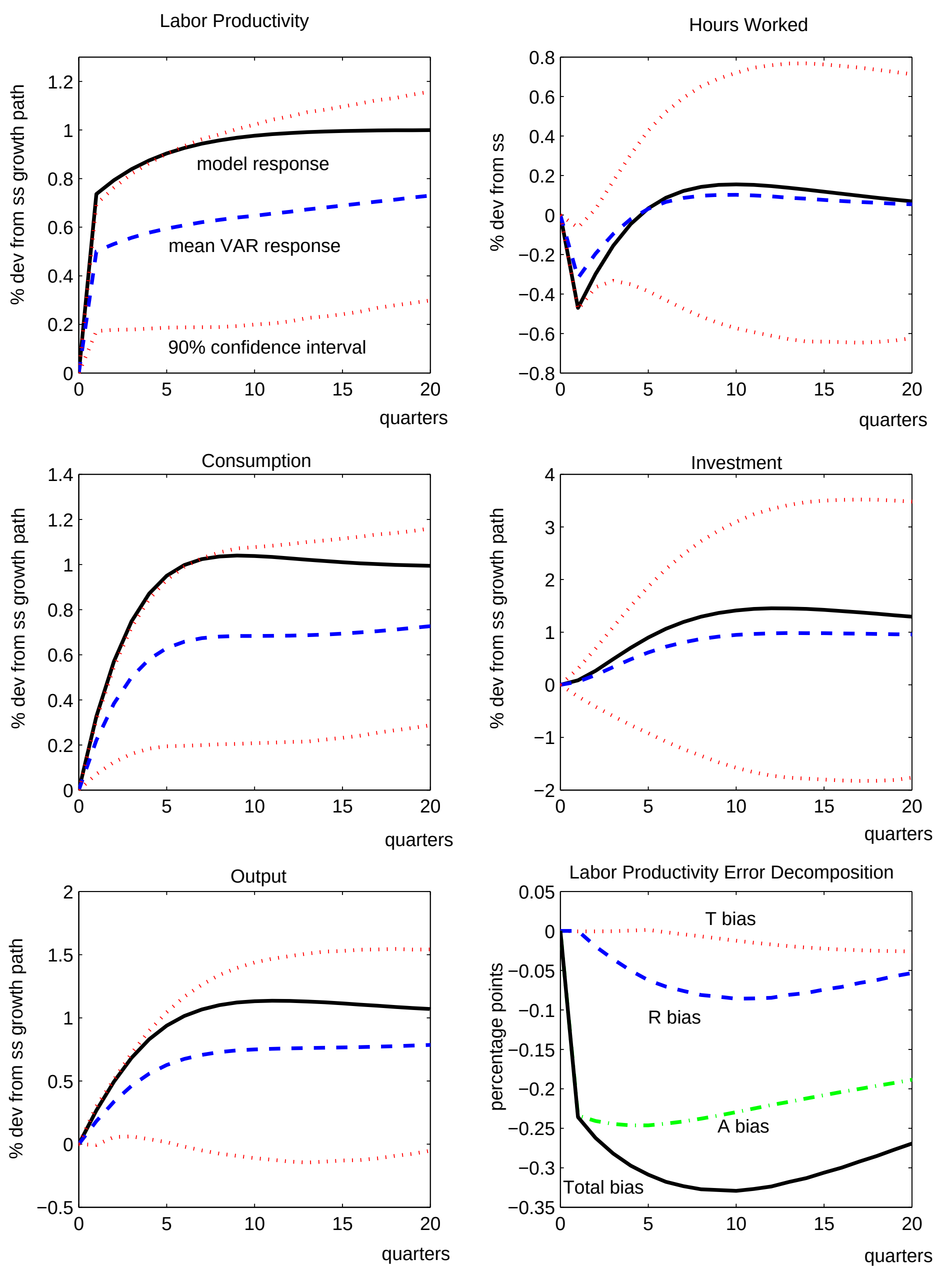
[b] Mean results based on 10,000 samples of 180 quarterly observations.

Figure 6: Responses to a Technology Shock in the RBC Model with Additional Shocks Using Population Moments*



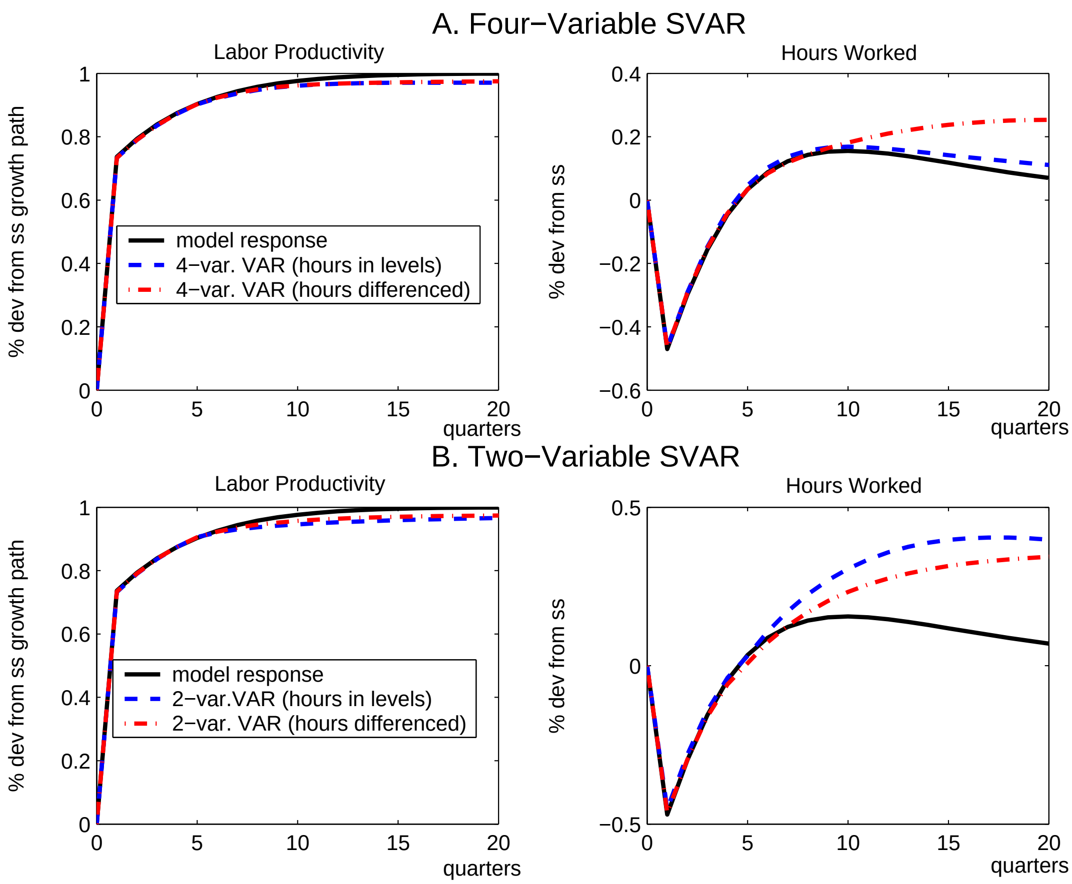
* Results based on VARs of order 4 estimated with model's population moments. VAR response for "cap. tax shock" refers to the case where the data-generating process is the benchmark RBC model augmented to include capital tax rate shocks (with $\sigma_{\tau_K} = 0.008$). VAR response for "temp. tech. and cap. tax shocks" refers to the case where the data-generating process is the RBC model augmented to include both capital tax rate and temporary technology shocks.

Figure 7: The Effects of Technology Shocks in Benchmark Sticky Price/Wage Model*



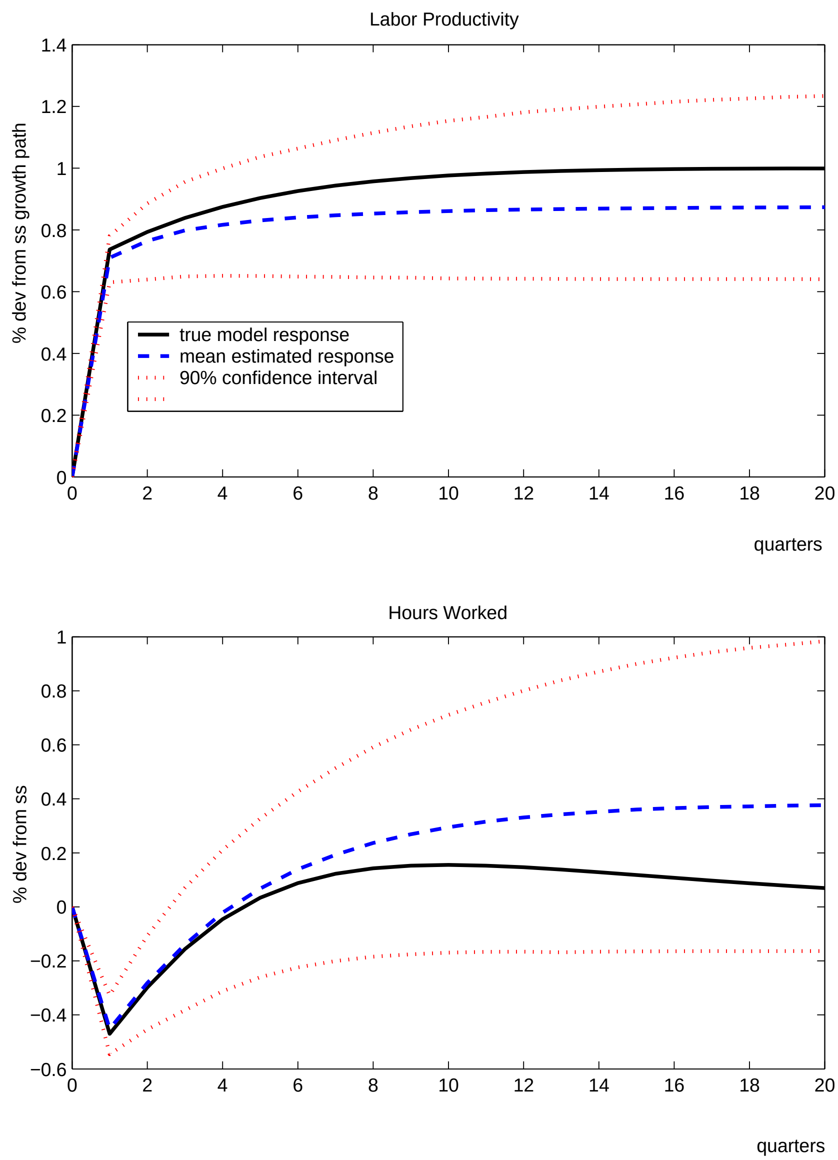
* VAR results based on 10,000 samples of 180 quarterly observations. In the lower right panel, T bias refers to bias that arises from approximating the true VARMA process with a VAR of order 4. The R bias reflects small-sample bias from estimating the reduced-form VAR. The A bias reflects small-sample bias associated with the transformation of the reduced-form to the structural form.

Figure 8: Responses to a Technology Shock in Sticky Price/Wage Model Using Population Moments*



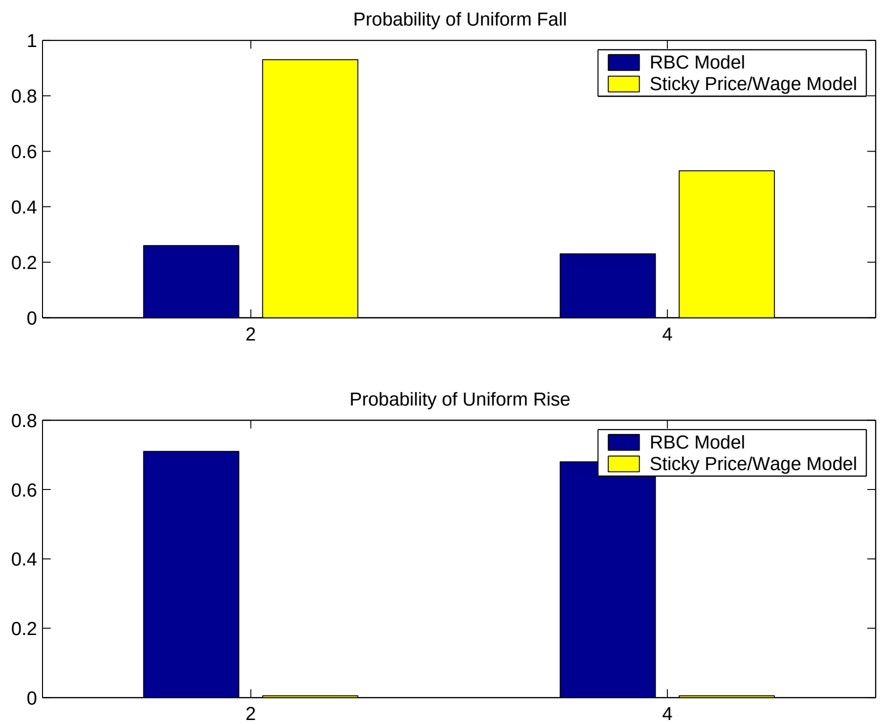
* Results based on VARs of order 4 estimated with model's population moments.

Figure 9: Responses to a Technology Shock in Sticky Price/Wage Model Using a Bivariate SVAR with Hours Differenced*



* VAR results based on 10,000 samples of 180 quarterly observations.

Figure 10: The Response of Hours in Each of the Benchmark Models*



* VAR results based on 10,000 samples from each model of 180 quarterly observations. Probability uniformly negative (positive) refers to the likelihood that the estimated response of hours is negative (positive) in each of the first two and first four quarters. Because we use uniform probabilities, the probabilities of positive and negative responses do not necessarily sum to one.

7 Appendix

This appendix is divided into three sections. In the first, we show results for the SVAR with different sample lengths and different fixed lag-lengths. In the second, we discuss how the log-linear solution of our RBC model can be written as a VARMA(4,5). Finally, we describe the decomposition of the sources of error in the SVAR impulse responses.

7.1 Results for Different Sample Lengths and Fixed Lag Lengths

Table A documents the performance of the SVAR using different sample lengths of data generated under the benchmark RBC calibration. In practice, researchers might be limited to samples shorter than 180 quarterly observations, or might choose to work with a smaller sample due to structural breaks. In the row labelled “120”, which corresponds to 30 years of quarterly data, we report the probabilities of large misses over the first four quarters following the shock. Not surprisingly, our results suggest that the problems documented so far are compounded by reducing the length of the estimation sample.

We investigated how large a sample we would need to ameliorate the small-sample problems documented so far. Table A shows that even with 100 years of data there would still be a sizeable chance of making large errors. For instance, the probability that the response of labor productivity would be estimated uniformly outside a 33% band around the true response remains as high as 19%. Only when the estimation sample includes 1000 quarterly observations do most of the probabilities of large misses drop below 10%. The exception is hours worked. This reflects that the model’s response is close to zero, and we use an alternative criterion that gives the probability of a uniformly negative response. However, the probability of a negative response for hours diminishes when we increase the number of observations further, as might be expected from our estimates using population moments.

Table B investigates how the performance of the SVAR depends on the number of lags included; thus, rather than using the Schwarz criterion to determine the lag length for each Monte Carlo draw, in these experiments we simply fix the lag length at a constant value p (we

Table A. Varying the Sample Size for the Benchmark RBC Calibration: Uniform Probability that Estimated Response is Far From True Response Over First Four Quarters[a]

Number of Quarters	Labor Productivity	Output	Hours	Consumption	Investment
120 (10 years less)	0.63	0.43	0.23	0.51	0.38
180 (benchmark length)	0.44	0.25	0.23	0.27	0.28
260 (20 years more)	0.32	0.16	0.22	0.18	0.24
400 (100 years)	0.22	0.11	0.22	0.11	0.22
1000 (250 years)	0.05	0.03	0.20	0.04	0.12

[a] For all variables except hours worked, the probability that the estimated response lies at least 33% above or below the true response for the first four quarters. For hours worked, the probability that the sign of the estimated response is incorrect in each of the first four quarters.

Table B. Varying the VAR Lag Structure for the Benchmark RBC Model: Probability that Estimated Response is Uniformly Far From True Response Over First Four Quarters[a]

Experiment	Labor Productivity	Output	Hours	Consumption	Investment
Lag length = 2	0.40	0.19	0.21	0.21	0.26
Lag length = 3	0.40	0.20	0.21	0.22	0.26
Lag length = 4	0.41	0.22	0.21	0.24	0.26
Lag length = 5	0.42	0.23	0.22	0.26	0.26
Lag length = 6	0.44	0.25	0.22	0.28	0.27
Lag length = 9	0.49	0.32	0.23	0.35	0.31
Lag length = 10	0.51	0.34	0.24	0.38	0.32
BIC	0.44	0.25	0.23	0.27	0.28

[a] For all variables except hours worked, the probability that the estimated response lies at least 33% above or below the true response for the first four quarters. For hours worked, the probability that the sign of the estimated response is incorrect in each of the first four quarters.

use a sample length of 180 quarterly observations). The table reports the probabilities of large errors over the first four quarters for different lag lengths. There is some modest improvement in the fit of the SVAR for smaller values of p. Still, the probability of a large miss for labor productivity is above 40 percent, and there is over a 20 percent chance of concluding that hours worked fall when in truth it rises.

7.2 Writing the RBC Model as a VARMA(4,5)

We first obtain a log-linear solution of the RBC model around its non-stochastic steady state. This allows us to express the log-linear decision rule for the economy's scaled capital stock, $\hat{k}_{t+1} = K_{t+1}/Z_t$, as a function of lagged capital, $\hat{k}_t$, and a vector of the four exogenous shocks, $S_t = (\tilde{\mu}_{zt}, \tilde{\tau}_{Nt}, \tilde{g}_t, \tilde{\chi}_{0t})'$ in the benchmark calibration, (where the tilde denotes that the variable is expressed in log deviation from its steady state value). Also, for convenience, we have defined $\mu_{zt} = log(Z_t) - log(Z_{t-1})$ and rewritten equation (6) more generally as

$$\mu_{zt} = (1 - \rho_z)\mu_z + \rho_z \mu_{zt-1} + \sigma_z \epsilon_{zt}, \tag{24}$$

even though $\rho_z = 0$.

The log-linear decision rule for the scaled capital stock can then be expressed as:

$$\tilde{k}_{t+1} = a_{kk}\tilde{k}_t + b_{ks}S_t, \tag{25}$$

where a_{kk} is a scalar and b_{ks} is a 4x1 vector of coefficients. We can also write hours worked, the consumption-to-output ratio, and investment-to-output ratio as a function of $\tilde{k}_t$ and S_t, while the growth rate of labor productivity is a function of $\tilde{k}_t$, $\tilde{k}_{t-1}$, S_t, and S_{t-1}. Therefore, the model's dynamics for X_t, the vector containing the variables in our VAR, can be expressed as:

$$\tilde{X}_t = C_1\tilde{k}_t + C_2\tilde{k}_{t-1} + D_1S_t + D_2S_{t-1}, \tag{26}$$

where C_1 and C_2 are 4x1 vectors and D_1 and D_2 are 4x4 matrices.

Using the log-linear decision rule for k_{t+1} to substitute the scaled capital stock out of the linear decision rules for labor productivity growth, hours, and the ratios of consumption and investment to output, we can express the linear dynamics of X_t as:

$$\begin{aligned} X_t &= a_{kk}X_{t-1} + (B_0 + B_1L + B_2L^2)S_t \\ S_t &= \rho S_{t-1} + \sigma\epsilon_t \end{aligned} \tag{27}$$

where $B_0 = D_1$, $B_1 = C_1B_{ks} - a_{kk}D_1 + D_2$, and $B_2 = C_2B_{ks} - a_{kk}D_2$; ρ and σ are diagonal 4x4 matrices whose respective elements contain the AR(1) coefficients and standard deviations of the innovations. Finally, $\epsilon_t = (\epsilon_{zt}, \epsilon_{\tau_N,t}, \epsilon_{\chi t}, \epsilon_{gt})'$.

It is convenient to rewrite the first equation in (27) as:

$$(I - a_{kk}L)X_t = \sum_{j=1}^{4}(B_{0,c(j)} + B_{1,c(j)}L + B_{2,c(j)}L^2)S_{jt}, \tag{28}$$

where $B_{0,c(j)}$ denotes the j^{th} column of B_0, and S_{jt} is the j^{th} shock in S_t. Because ρ and σ are diagonal matrices, we denote the j^{th} element along the diagonal of these matrices as ρ_j and σ_j, respectively. Using these diagonal matrices, we can substitute out S_t from equation (28) to write

$$\begin{aligned} &\prod_{i=2}^{4}(1-\rho_iL)(I - a_{kk}L)X_t = \\ &\prod_{i=2,i\neq j}^{4}\sum_{j=1}^{4}(1-\rho_i)(B_{0,c(j)} + B_{1,c(j)}L + B_{2,c(j)}L^2)\epsilon_{jt}, \end{aligned}$$

or

$$a(L)X_t = b(L)\epsilon_t, \tag{29}$$

with $a(L) = \sum_{i=0}^{4} a_iL^i$ and $b(L) = \sum_{i=0}^{5} b_iL^i$. In the above, $a_0 = I_4$ and a_i for $i = 1, 2, 3, 4$ are 4x4 matrices that depend on a_{kk} and ρ_j for $j = 2, 3, 4$. Also, $b_0 = B_0$ and b_i for $i = 1, 2, 3, 4, 5$ are 4x4 matrices that depend on the elements of B_0, B_1, and B_2 and ρ_j for $j = 2, 3, 4$. Note that $a(L)$ and $b(L)$ do not depend on ρ_1 since $\rho_z = \rho_1 = 0$.

Lippi and Reichlin (1993) make the point that researchers fitting a VAR to the data would not be able to recover the underlying shocks, if the data generating process had a non-fundamental representation. Therefore, for our benchmark calibrations, we checked that our model implied a fundamental representation by verifying numerically that the polynomial

$\det(b_0 + b_1 z + ... + b_5 z^5)$ has all roots strictly outside the unit circle. This condition ensures that the VARMA process in equation (29) is invertible and is a fundamental representation for X_t (see page 222 and page 456 of Lutkepohl (1991)).

7.3 Error Decomposition

In this section, we decompose the error in estimating the response to a technology shock for a given Monte Carlo draw into two sources. The first source arises because the VAR we estimate is an imperfect approximation of the VARMA process implied by our models. The second source is due to small-sample imprecision.

For a given Monte Carlo draw, let $\hat{d}_{l,i}$ denote the estimated impulse response for i^{th} variable, at lag l for a particular; let $d^*_{l,i}$ denote the impulse response from the true model, and let $d_{l,i}$ be the estimate of the SVAR's impulse response using the model's population moments.[39] Accordingly, $\hat{d}_{l,i} - d^*_{l,i}$ is the error in the estimate of the response to a technology shock for ith variable at lag l. We can rewrite this error as:

$$\hat{d}_{l,i} - d^*_{l,i} = (d_{l,i} - d^*_{l,i}) + (\hat{d}_{l,i} - d_{l,i}). \tag{30}$$

The first source of error $(d_{l,i} - d^*_{l,i})$ due to approximating a VARMA process with a VAR. The second source of error $(\hat{d}_{l,i} - d_{l,i})$ arises in all time series work because of limited sample size.

We now proceed to decompose the small-sample error into two parts: one arising from estimating the reduced form and another from transforming the reduced form to structural. Using the notation from equation (19), we begin by noting that

$$\hat{d}_{l,i} = \hat{R}_{l,r(i)}\hat{\alpha}, \tag{31}$$

where $\hat{\alpha}$ denotes the finite-sample estimate of the first column of A_0, $\hat{R}_l$ is the finite-sample estimate of R_l, and the subscript $r(i)$ denotes the i^{th} row of this matrix. The term $\hat{\alpha}$ maps the reduced-form impulse responses into structural ones. It is important to recognize that $\hat{\alpha}$ is

[39]We compute $d_{l,i}$ by using the log-linear solution of the DGE model to find the population estimates of A_j, $j = 1, 2, ..., p$, in equation (14) and use those estimates along with equation (16) to determine A_0.

implicitly a function of $\hat{R}(1)$ through equation (16), where $\hat{R}(1)$ determines the long-run response of the variables in the VAR to unidentified innovations.[40] As discussed in Faust and Leeper (1997) small imprecision in estimating $A(L)$, the reduced-form VAR parameters, can result in large errors in $\hat{R}(1)$, and this error affects all the identified impulse responses through $\hat{\alpha}$.

We decompose the small sample error of estimating the impulse response of variable i at lag l as

$$\hat{d}_{l,i} - d_{l,i} = (\hat{R}_{l,r(i)} - R_{l,r(i)})\tilde{\alpha} + \tilde{R}_{l,r(i)}(\hat{\alpha} - \alpha), \tag{32}$$

where the matrices, $\tilde{\alpha} = \frac{1}{2}(\hat{\alpha} + \alpha)$ and $\tilde{R}_{l,r(i)} = \frac{1}{2}(\hat{R}_{l,r(i)} + R_{l,r(i)})$ are defined to lie halfway between the finite-sample estimates and the population estimates of the SVAR.[41] Equation (32) shows the two parts of our decomposition: the first emphasizes the error in estimating the reduced-form moving average term, $R_{l,r(i)}$, and the second emphasizes the error in estimating $R(1)$ through the α term. Finally, we compute the R and A biases reported in Figures 1 and 7 by averaging these two parts of the small-sample error over the 10,000 Monte Carlo replications.

7.4 Additional Variable Selection Analysis Using the Benchmark RBC Model

In this section, we conduct some additional analysis regarding variable selection and discuss why the benchmark, four-variable SVAR performs better in the RBC model than the bivariate SVAR with hours in levels. We begin by documenting that the three variable SVAR that includes labor productivity growth, hours worked, and the scaled capital stock, K_{t+1}/Z_t, can perform well when the benchmark RBC model is used as the data-generating process.

[40] We define $\hat{\alpha} = \alpha(\hat{R}(1), \hat{\Sigma})$ where $\hat{R}(1)$ and $\hat{\Sigma}$ are the VAR's estimates of $R(1)$ and of the reduced form variance-covariance matrix Σ, respectively. Our decomposition does not parse out the error from estimating the variance-covariance matrix from estimating $R(1)$. However, for both of the benchmark models, we checked that the error from having to estimate $\alpha(R(1), \hat{\Sigma})$ was small and most of the error was due to estimating $\alpha(\hat{R}(1), \Sigma)$.

[41] We thank Jon Faust for suggesting this decomposition of the small-sample error.

This result is shown in Figure A, which shows the responses of labor productivity and hours worked for the three-variable SVAR using four lags and the model's population moments. Comparing this to the results of the bivariate SVAR in Figure 5, it is clear that the performance of the short-ordered SVAR improves considerably if we augment the state space to include the scaled capital stock.

In practice, an obvious difficulty with the above three-variable SVAR is that the scaled capital stock is unobservable. However, in the RBC model, there are several observable variables that are highly correlated with it and can help "proxy" for it. One natural candidate is the capital-to-output ratio, K_{t+1}/Y_t. Although we do not show it here, a three-variable SVAR that includes this variable performs as well as the three-variable SVAR with K_{t+1}/Z_t.

There are also other variables in the RBC model that are correlated with the scaled capital stock and can improve the VAR's performance. One useful way of summarizing such variables is to fit the following equation:

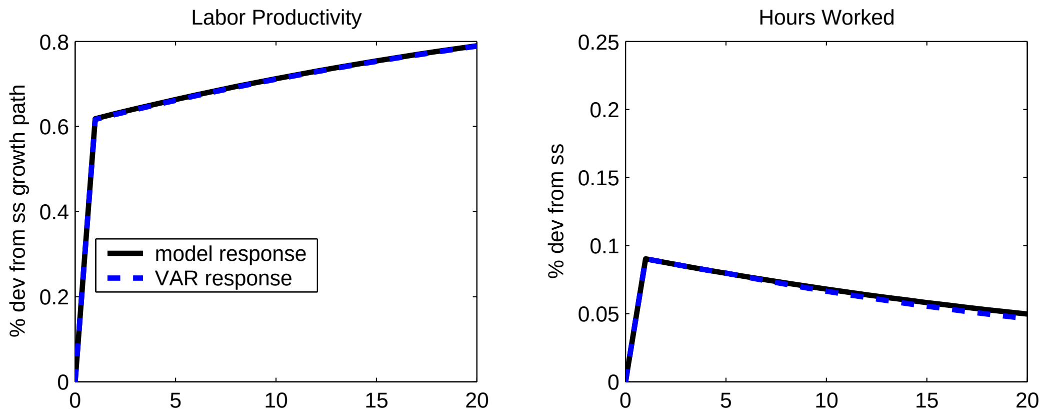
$$\hat{k}_{t+1} = \Theta_0 X_t + \Theta_1 X_{t-1} + \Theta_2 X_{t-2} + ... + \Theta_p X_{t-p} + \varepsilon_t, \tag{33}$$

by choosing Θ_0,...,Θ_p to minimize $E(\varepsilon_t^2)$. In the above, $\hat{k}_{t+1}$ is the log-deviation of the scaled capital stock from its steady state value, and X_t is a vector of variables (in log-deviation from steady state) that includes labor productivity growth and hours worked and possibly other observable variables that are presumed to provide additional explanatory variable for the scaled capital stock.

Table C shows the R^2 statistic from this regression where X_t contains only labor productivity growth and hours worked. In this case, if the lag length is four, R^2 is 0.34. As we increase the lag length to 100, R^2 rises to 0.92. In contrast, the R^2 is always close to one if we include the capital-to-output ratio in the regression.

Table C also suggests that including the ratios of consumption and investment to output would be good additions to the bivariate SVAR, as confirmed in our analysis. The inclusion of these variables in the regression appears to be preferable to including consumption and investment in differences, and not surprisingly, the short-ordered, four-variable SVAR with $\Delta\hat{C}_t$ and $\Delta\hat{I}_t$ (not shown) does not perform as well as the SVAR with $\hat{C}_t - \hat{Y}_t$ and $\hat{I}_t - \hat{Y}_t$.

Figure A. The Response to a Technology Shock in the Benchmark RBC Model Using Population Moments for a 3-Variable SVAR*



* Results based on fourth-ordered VAR that includes labor productivity growth, hours worked, and the scaled capital stock.

Table C. R-Squareds from Scaled Capital Stock Equation*

Independent Variables (X_t)	$p=0$	$p=1$	$p=4$	$p=20$	$p=100$
$X_t = (\Delta(\hat{Y}_t - \hat{N}_t), \hat{N}_t)'$	0.08	0.16	0.34	0.79	0.92
$X_t = (\Delta(\hat{Y}_t - \hat{N}_t), \hat{N}_t, \hat{K}_{t+1} - \hat{Y}_t)'$	≈ 1	1	1	1	1
$X_t = (\Delta(\hat{Y}_t - \hat{N}_t), \hat{N}_t, \Delta\hat{K}_{t+1})'$	0.13	0.89	0.89	0.91	0.95
$X_t = (\Delta(\hat{Y}_t - \hat{N}_t), \hat{N}_t, \hat{C}_t - \hat{Y}_t)'$	0.53	0.59	0.71	0.95	≈ 1
$X_t = (\Delta(\hat{Y}_t - \hat{N}_t), \hat{N}_t, \hat{I}_t - \hat{Y}_t)'$	0.59	0.61	0.66	0.84	≈ 1
$X_t = (\Delta(\hat{Y}_t - \hat{N}_t), \hat{N}_t, \Delta\hat{C}_t)'$	0.17	0.36	0.63	0.87	0.94
$X_t = (\Delta(\hat{Y}_t - \hat{N}_t), \hat{N}_t, \Delta\hat{I}_t)'$	0.10	0.18	0.36	0.80	0.94
$X_t = (\Delta(\hat{Y}_t - \hat{N}_t), \hat{N}_t, \hat{C}_t - \hat{Y}_t, \hat{I}_t - \hat{Y}_t)'$	0.99	0.99	0.99	≈ 1	≈ 1
$X_t = (\Delta(\hat{Y}_t - \hat{N}_t), \hat{N}_t, \Delta\hat{C}_t, \Delta\hat{I}_t)'$	0.22	0.78	0.87	0.91	0.95

* The regression equation is $\hat{k}_{t+1} = \Theta_0 X_t + \Theta_1 X_{t-1} + \Theta_2 X_{t-2} + ... + \Theta_p X_{t-p} + \varepsilon_t$, where p denotes the regression's lag length.