Journal of Monetary Economics 55 (2008) 1337–1352

Contents lists available at ScienceDirect

Journal of Monetary Economics

journal homepage: www.elsevier.com/locate/jme

Are structural VARs with long-run restrictions useful in developing business cycle theory? ☆

V.V. Chari [a,b], Patrick J. Kehoe [a,b,*], Ellen R. McGrattan [a,b]

[a] *Department of Economics, University of Minnesota, USA*
[b] *Federal Reserve Bank of Minneapolis, USA*

A R T I C L E I N F O

Available online 9 October 2008

JEL classification:
E2
E3
E13
E32
E37
C32
C51

Keywords:
Vector autoregressions
Technology shocks
Real business cycle
Impulse response

A B S T R A C T

No, unless technology shocks account for virtually all of the fluctuations in output.
© 2008 Published by Elsevier B.V.

1. Introduction

The goal of researchers using structural vector autoregressions (SVARs) with long-run restrictions is to guide the development of new business cycle models. This procedure uses a simple time series technique and minimal economic theory to identify the pattern of responses of economic aggregates to a technology shock in the data. The idea is that developers of new models should focus only on ones which can reproduce these patterns.[1] We subject the commonly used version of this procedure to a natural economic test and find that it is not useful unless technology shocks account for virtually all of the fluctuations in output.

The growing interest in such SVARs stems largely from the recent finding of researchers using this procedure that a technology shock leads to a fall in hours. (See, for example, Galí, 1999; Francis and Ramey, 2005; Galí and Rabanal, 2005.) The conclusion drawn from this finding is that only models with this pattern are deemed promising.

☆ The authors thank the editor, Robert King, and the referees for very helpful comments, Kathy Rolfe and Joan Gieseke for excellent editorial assistance, and the National Science Foundation for financial support. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

* Corresponding author at: Federal Reserve Bank of Minneapolis, 90 Hennepin Ave., Minneapolis, MN 55401, USA. Tel.: +1 612 204 5525; fax: +1 612 204 5515.

E-mail address: pkehoe@res.mpls.frb.fed.us (P.J. Kehoe).

[1] Of course, if a researcher has a specific model in hand, there are many more efficient ways to evaluate it than the SVAR procedure. Thus, the main point of the procedure is to help guide the development of new models which have not yet been specified in enough detail that one could use these other methods.

0304-3932/$ - see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.jmoneco.2008.09.010

Since a positive technology shock leads to a rise in hours in most real business cycle models, the researchers argue that their SVAR analyses doom both existing and future real business cycle models and point to other types of models, such as sticky price models which can produce this pattern, as promising. (See Galí, 1999; Francis and Ramey, 2005; Galí and Rabanal, 2005.) In particular, these analyses doom all of the work on second generation business cycle models which augment the early technology-shock-only models with more shocks that help the model better fit key features of the data, especially the volatility of hours.[2] A key feature of these second generation models is that, even though nontechnology shocks account for a sizable fraction of output fluctuations, a positive technology shock leads to a rise in hours.

Here we focus on the SVAR literature that uses the *common approach*. In this approach, researchers run VARs on the actual data, impose some identifying assumptions on the VARs in order to back out empirical impulse responses to various shocks, and then compare those empirical SVAR impulse responses to theoretical responses from various economic models. Models that generate theoretical responses with the same pattern as the SVAR responses are thought to be promising; others are not. The main claim of this literature is that its SVAR procedure is useful because it can confidently and correctly distinguish between promising and unpromising classes of models regardless of the details of nontechnology shocks, other than minimal assumptions like orthogonality.

We evaluate this claim by subjecting the SVAR procedure to a natural economic test. We treat a real business cycle model with *technology shocks* and labor wedge shocks that we refer to as *nontechnology shocks* as the data-generating mechanism, apply the SVAR procedure to the model's data, and see if the procedure can do what is claimed for it. Our model is a stripped-down version of many popular business cycle models that satisfies the two key identifying assumptions of the SVAR literature: technology and nontechnology shocks are orthogonal, and nontechnology shocks have no permanent effect on the level of labor productivity, whereas technology shocks do—a commonly used *long-run restriction*.

In our test, we first generate data from the business cycle model, drawing a large number of sequences of roughly the same length as postwar U.S. data. Then we apply the SVAR procedure with the typical small number of lags on each sequence of model-generated data and compute the means of the impulse responses and confidence bands.[3] Finally, we compare the SVAR impulse responses to those of the theoretical model, to see how well this procedure can reproduce the model's responses.[4]

We begin with the specification used by Galí (1999), Francis and Ramey (2005), and Galí and Rabanal (2005), which has two variables in the VAR: the growth rate of labor productivity and the first difference of hours. We refer to this specification as the *DSVAR*. Because of the limited length of the available time series, the VAR is estimated with a small number of lags, typically four. In investigating this specification, we sidestep one minor technical issue: the existence of an autoregressive representation of the model, by using a *QDSVAR* specification in which hours are quasi-differenced.[5] When the quasi-differencing parameter is close to one, as it is in our test, the impulse responses of the QDSVAR and the DSVAR are indistinguishable.

We find that, in principle, the SVAR claim of not needing to specify the details of nontechnology shocks is correct if the researcher has extremely long time series to work with. With series of the length available in practice, however, the SVAR claim is incorrect. When nontechnology shocks play a nontrivial role in output fluctuations, a researcher who applies the QDSVAR procedure to data generated by our model will confidently but incorrectly conclude that the data are not generated from our model.

Specifically, we find that when nontechnology shocks account for a trivial fraction of output fluctuations, the means of the SVAR impulse responses are close to the model's theoretical impulse responses. But that is not true when nontechnology shocks account for a nontrivial fraction of output fluctuations. Then the estimated impulse response of hours to a technology shock is negative and significantly so. Indeed, when nontechnology shocks account for 50% of output fluctuations, the estimated impact effect of a technology shock on hours coincides with Galí's negative estimate and has confidence bands that rule out our real business cycle model. In the model used to generate the data, of course, the theoretical impulse response of hours to a technology shock is positive. If we followed the logic in the SVAR literature (as in Galí, 1999; Francis and Ramey, 2005; Galí and Rabanal, 2005), we would then infer that the data are inconsistent with our real business cycle model and potentially consistent with a sticky price model. This inference would be incorrect

[2] See, for example, Cooley and Hansen (1989), Benhabib et al. (1991), Bencivenga (1992), Rotemberg and Woodford (1992), Braun (1994), McGrattan (1994), Stockman and Tesar (1995), Hall (1997), Bernanke et al. (1999), Greenwood et al. (2000), and Christiano et al. (2005).

[3] We also conduct a variety of standard lag-length tests and find that these tests do not detect the need for more lags.

[4] We emphasize that our test is a logical analysis of the inferences drawn from the SVAR approach and neither asks nor depends on why productivity in the U.S. data fluctuates. In our test, we use data generated from an economic model because in the model we can take a clear stand on what constitutes a technology shock. Hence, in our test, the question of whether fluctuations in total factor productivity in U.S. data come from changes in technology or from other forces is irrelevant.

[5] One critique of the DSVAR procedure is that in all economic models, the time series hours worked per person is bounded, and therefore, the stochastic process for hours per person cannot literally have a unit root. Hence, according to the critique, the DSVAR procedure is misspecified with respect to all economic models and, thus, is useless for distinguishing among broad classes of models. This critique is simplistic. We are sympathetic to the view expressed in the DSVAR literature that the unit root specification is best viewed as a statistical approximation for variables with high serial correlation. See, for example, Francis and Ramey (2005) for an eloquent defense of this position. See also Marcet (2005) for a defense of differencing in VARs.

because the data are generated by our model. In this sense, our results cast doubt on the conclusion in the SVAR literature that their SVAR analyses doom existing real business cycle models.

Our findings cast doubt on another conclusion as well, that technology shocks account for, at best, a minor fraction of business cycle fluctuations. Galí (1999) reaches this conclusion by combining the SVAR result that hours fall after a technology shock with the observation that hours are procyclical to infer that technology shocks cannot account for more than a minor fraction of business cycle fluctuations. Ironically, we have shown that SVAR results are likely to lead to the wrong inference when technology shocks account for a minor fraction of business cycle fluctuations and to lead to the right inference when technology shocks account for essentially all of business cycle fluctuations.

We also test another specification of the VAR, the *LSVAR*, which replaces the difference in hours with the level of hours. Here we find that when nontechnology shocks play a nontrivial role in business cycle fluctuations, the confidence bands are so wide that the procedure does not allow a researcher to distinguish between promising and unpromising classes of models. In particular, the procedure cannot tell whether the data are generated by a real business cycle model or a sticky price model. Thus, the LSVAR procedure also fails the main claim of the SVAR literature.

We obtain intuition for our findings from two propositions. One shows that the VAR on observables from our model has an infinite-order representation. With our parameter values, the coefficients in this representation decay very slowly. With series of the length available in practice, the estimated impulse responses are not close to the theoretical impulse responses when the nontechnology shock is not trivial. Our analysis indicates that this poor performance is due to *lag-truncation bias.* That is, the small number of lags in the estimated VAR dictated by available data lengths makes the estimated VAR a poor approximation to the infinite-order VAR of the observables from the model. Another proposition shows that a VAR with a small number of lags is a good approximation to the model's infinite-order VAR when nontechnology shocks play a minor role. This result helps explain why theoretical and empirical responses are close when nontechnology shocks play a small role.

We then ask, are the only models worth developing ones in which technology shocks account for virtually all of the fluctuations in output? We argue that both the literature and the data say the answer is no.

Our critique challenges the dramatic recent result from the SVAR literature, which implies the death of the real business cycle model. The common SVAR approach with long-run restrictions is not a useful tool for making such judgments. The root of its problem is that the procedure compares two very different sets of statistics: empirical and theoretical impulse responses. As statistics of the data, empirical impulse responses are entirely unobjectionable. The comparison between the two sets of statistics, however, is inappropriate because it is prone to various pitfalls, especially lag-truncation bias.

Not all SVAR procedures make such inappropriate comparisons. A preferable alternative to the common approach is one that compares empirical impulse responses based on the data to impulse responses from identical SVARs run on data from the model of the same length as the actual data. We call this the *Sims–Cogley–Nason approach* because it has been advocated by Sims (1989) and successfully applied by Cogley and Nason (1995). On purely logical grounds, the Sims–Cogley–Nason approach is superior to the approach we scrutinize here; it treats the data from the U.S. economy and the model economy symmetrically, thereby avoiding the problems of the common approach.

Our work here is related to a theoretical literature that discusses estimation and inference problems in models with an infinite number of parameters, most prominently, the work of Sims (1971, 1972). Faust and Leeper (1997) build on Sims' arguments to discuss inference problems in infinite-dimensional VARs that underlie the SVAR approach. They argue that "unless strong restrictions are applied, conventional inferences regarding impulse responses will be badly biased in all sample sizes" (p. 345). For some related applied critiques of the SVAR literature, see Cooley and Dwyer (1998) and Erceg et al. (2004). (For a discussion of the literature, see Chari et al., 2007a henceforth, CKM.)

2. Tools for testing

Let us start our critique of the common SVAR approach with long-run restrictions by briefly describing the two basic tools needed to apply our natural economic test: an SVAR procedure and a business cycle model.

2.1. An SVAR procedure

The VAR procedure we evaluate is a version of Blanchard and Quah's (1989) procedure used recently by Galí (1999), Francis and Ramey (2005), and Galí and Rabanal (2005). The idea is to uncover the parameters of a structural model from running VARs and using several identifying assumptions. The structural model is given by

$$Y_t = A_0\varepsilon_t + A_1\varepsilon_{t-1} + A_2\varepsilon_{t-2} + \cdots, \tag{1}$$

where the vector Y_t is given by $(y_{1t}, y_{2t})'$, where $y_{1t} = \Delta\log(y_t/l_t)$ is the first difference of the log of labor productivity, $y_{2t} = \log l_t - \alpha\log l_{t-1}$, and l_t is a measure of the labor input. We consider two specifications: in the *differenced* specification (DSVAR), $\alpha = 1$, so y_{2t} is the first difference in the log of the labor input; in the *level* specification (LSVAR), $\alpha = 0$, so y_{2t} is simply the log of the labor input. Here the A's are the structural coefficients and the $\varepsilon_t = (\varepsilon_t^z, \varepsilon_t^d)'$ represent the structural *technology* and *nontechnology* shocks, with $E\varepsilon_t\varepsilon_t' = \Sigma$ and $E\varepsilon_t\varepsilon_s' = 0$ for $s \neq t$. The response of Y_t in period $t+i$ to a shock in period t is given by A_i. From these responses, the impulse responses for y_t/l_t and l_t can be determined.

The associated VAR is of the form

$$Y_t = B_1 Y_{t-1} + \cdots + B_p Y_{t-p} + v_t. \tag{2}$$

This VAR, as it stands, can be thought of as a reduced form of an economic model. Specifically, the reduced-form error terms v_t have no structural interpretation. Inverting the VAR is convenient in order to express it in its equivalent moving-average form

$$Y_t = C_0 v_t + C_1 v_{t-1} + C_2 v_{t-2} + \cdots, \tag{3}$$

where the moving-average coefficients, the C's, are related to the B's in the standard way.

The idea behind the SVAR procedure is to use the reduced-form model equation (3), together with the bare minimum of economic theory, to back out structural shocks and the responses to those shocks.

To identify the structural parameters, the A's, from the reduced-form parameters, the B's, some assumptions are needed. The SVAR procedure we are evaluating uses two identifying assumptions together with a sign restriction and an auxiliary assumption. One identifying assumption is that technology and nontechnology shocks are *orthogonal*. The other identifying assumption is a *long-run restriction*, the assumption that $\sum_{i=0}^{\infty} A_i(1,2) = 0$ and $\sum_{i=0}^{\infty} A_i(1,1) \neq 0$, where $A_i(j,k)$ is the element in the jth row and the kth column of the matrix A_i. This assumption captures the idea that nontechnology shocks do not affect the level of labor productivity in the very long run, but technology shocks do. The sign restriction is that a technology shock is called *positive* if it raises the level of labor productivity in the long run.[6]

An *auxiliary assumption*, typically not emphasized by researchers, is also needed, namely, that

$$A(L)^{-1} \text{ exists and is equal to } I - \sum_{i=1}^{p} B_i L^i, \tag{4}$$

where $A(L) = A_0 + A_1 L + \cdots$ and where L is the lag operator. (See CKM for details.)

Our analysis of the problems with the common approach rests crucially on an analysis of the auxiliary assumption equation (4). In all of our versions of the baseline business cycle model, the auxiliary assumption is satisfied for an infinite number of lags ($p = \infty$). In practice, however, with existing data lengths, researchers are forced to run VARs with a much smaller number of lags, typically four. This lag truncation introduces a bias into the impulse responses computed using the common approach. The point of our analysis is to quantify how this lag-truncation bias varies with parameters. We also point out special circumstances under which, even though the VAR is truncated, the impulse responses to a technology shock have no lag-truncation bias.

2.2. A business cycle model

To test the claim made for the common SVAR approach with long-run restrictions, we will use a business cycle model with two shocks: changes in technology Z_t, which have a unit root, and an orthogonal tax on labor τ_{lt}. The model satisfies the two key identifying assumptions of the SVAR procedure we are evaluating, that technology and nontechnology shocks are orthogonal and that nontechnology shocks do not permanently affect the level of labor productivity, whereas technology shocks do.

We think of our model as a natural one with which to conduct our test. It is a variant of widely used real business cycle models that have been extensively studied in the literature. The multiple shocks in our model are motivated in part by the inability of the early technology-shock-only models, such as those in Kydland and Prescott (1982) and King et al. (1988), to generate the volatility of hours observed in the data. (See the references in footnote 1.) A key feature of the multiple-shock models is that, in them, nontechnology shocks account for a nontrivial fraction of output fluctuations.

Our choice of the tax on labor as the nontechnology shock is motivated by our earlier work (Chari et al., 2007b). In that work we show that many of the models in the literature are equivalent to a prototype business cycle model with a labor wedge that resembles a stochastic tax on labor. We also show that the labor wedge and the productivity shock account for the bulk of fluctuations in U.S. data. As we show in Chari et al. (forthcoming), this labor wedge can be thought of as arising from fluctuations in government spending, money, and the elasticity of demand of final goods producers in a New Keynesian model. (For early work making similar connections, see, among others, Rotemberg and Woodford, 1992; Goodfriend and King, 1997.)

In our model, consumers' utility functions are given by $E_0 \sum_{t=0}^{\infty} [\beta(1+\gamma)]^t U(c_t, l_t)$ defined over per capita consumption c_t and per capita labor l_t, where β is the discount factor and γ the growth rate of the population. Consumers maximize utility subject to the budget constraint

$$c_t + (1+\tau_x)[(1+\gamma)k_{t+1} - (1-\delta)k_t] = (1-\tau_{lt})w_t l_t + r_t k_t + T_t, \tag{5}$$

[6] In some of the VAR literature, sign restrictions are viewed as convenient normalizations with no economic content. Our sign restriction, in contrast, is a restriction implied by a large class of economic models, including the business cycle models considered below. It is similar in spirit to the long-run restriction. Both restrictions use the idea that while economic models may have very different implications for short-run dynamics, they often have very similar implications for long-run behavior.

where τ_x denotes a constant tax on investment, k_t denotes the per capita capital stock, δ the depreciation rate of capital, w_t the wage rate, r_t the rental rate on capital, and T_t lump-sum transfers, and where $\beta<1$, $\gamma\geqslant 0$, and $0\leqslant\delta\leqslant 1$.

In the model, firms have a constant returns to scale production function, $F(k_t, Z_t l_t)$, where Z_t is labor-augmenting technical progress. Firms maximize $F(k_t, Z_t l_t) - r_t k_t - w_t l_t$. The resource constraint is $c_t + (1+\gamma)k_{t+1} = y_t + (1-\delta)k_t$, where y_t denotes per capita output.

The stochastic process for the two shocks, $\log Z_t$ and τ_{lt}, which we refer to as the *technology* and *nontechnology shocks*, is

$$\log Z_{t+1} = \mu_z + \log Z_t + \log z_{t+1}, \tag{6}$$

$$\tau_{lt+1} = (1-\rho_l)\bar{\tau}_l + \rho_l \tau_{lt} + \varepsilon_{lt+1}, \tag{7}$$

where $\log z_t$ and ε_{lt} are mean zero normal random variables with standard deviations σ_z and σ_l. We let $\varepsilon_t = (\log z_t, \varepsilon_{lt})$, where these variables are independent of each other and i.i.d. over time. The constant $\mu_z \geqslant 0$ is the drift term in the random walk for technology, the parameter ρ_l is the persistence parameter for the labor tax, and $\bar{\tau}_l$ is the mean of the labor tax.

Our model satisfies the two key identifying assumptions of the SVAR approach using long-run restrictions. By construction, the two types of shocks are orthogonal. And in the model's steady state, the level of labor productivity is not affected by labor tax rates but is affected by technology levels. Thus, regardless of the persistence of the stochastic process on labor taxes, a shock to labor taxes has no effect on labor productivity in the long run.

In making our model quantitative, we use functional forms and parameter values familiar from the business cycle literature, and we assume that the time period is one quarter. We assume that the utility function has the form $U(c,l) = \log c + \phi \log(1-l)$ and the production function, the form $F(k,l) = k^{\theta} l^{1-\theta}$. We choose the time allocation parameter $\phi = 1.6$, the capital share $\theta = .33$, the serial correlation of the nontechnology shock $\rho_l = .95$, and the mean tax labor tax $\bar{\tau}_l = .4$. We choose the depreciation rate, the discount factor, and the growth rates so that, on an annualized basis, depreciation is 6%, the rate of time preference 2%, the population growth rate 1%, and the technology growth rate 2%.

In this model, a 1% positive technology shock leads to a persistent increase in hours worked. For our parameter values, on impact hours increase by .42%, and the half-life of the impulse response is about 18 quarters.

It is easy to show that the model's theoretical impulse response is independent of the persistence parameter ρ_l and the variances of the innovations σ_z^2 and σ_l^2. As we will see, when the SVAR procedures are applied to data from this model, their empirical impulse responses will vary with these parameters.

3. The natural economic test

We test the main claim of the common SVAR approach with long-run restrictions by comparing the business cycle model's impulse responses to those obtained by applying the SVAR procedure to data from that model, the *SVAR impulse responses*. Proponents of this procedure claim that it can confidently distinguish between promising and unpromising classes of models without the researchers having to specify the details of nontechnology shocks. We show that this claim is false.

3.1. An inessential technical issue

Before describing our test, we dispense with a technical issue. The common SVAR approach assumes that an autoregressive representation of the variables $(\Delta \log(y_t/l_t), l_t - \alpha l_{t-1})$ exists for the models to be evaluated, in the sense that the auxiliary assumption is satisfied for some, possibly infinite, number of lags p. For the LSVAR specification ($\alpha = 0$), the variables have an autoregressive representation. The DSVAR specification ($\alpha = 1$), however, overdifferences hours and introduces a root of 1 in the moving-average representation, which is at the edge of the noninvertibility region of roots. Hence, no autoregressive representation for the DSVAR exists.

This technical issue is not essential to our findings. We demonstrate that by considering, instead, a QDSVAR specification with α close to 1. When α is less than 1, these variables have an autoregressive representation. When α is close to 1, the impulse responses of the QDSVAR and the DSVAR are so close as to be indistinguishable. In our quantitative analyses, we will set the quasi-differencing parameter α equal to .99. With the QDSVAR specification and an infinitely long data series, the SVAR recovers the model's impulse response. Hence, there is no issue of misspecification with the QDSVAR.

3.2. Which specification is preferable?

We also ask which specification a researcher would prefer, the QDSVAR or the LSVAR, on a priori grounds. The time series of hours worked in our model is highly serially correlated, and we find that standard unit root tests do not reject the hypothesis that the hours series has a unit root. At least since Hurwicz (1950), we have known that autoregressions on highly serially correlated variables are biased in small samples and that quasi-differencing such variables may diminish that bias. Since both the QDSVAR and the LSVAR specifications have desirable asymptotic properties, on a priori grounds the QDSVAR seems preferable.

3.3. Evaluation of the SVAR claim

In our evaluation, we treat the business cycle model as the data-generating process and draw from it 1,000 data sequences of roughly the same length as our postwar U.S. data, which is 180 quarters. We run the SVAR procedure for each of the two specifications on each sequence of model data and report on the SVAR impulse responses of hours worked to technology shocks. We repeat this procedure for a wide range of parameter values for the stochastic processes and find that the SVAR procedure cannot do what is claimed for it.

In this test, we are conducting a simple thought experiment. Suppose that the data-generating mechanism is our real business cycle model at a particular parameter setting, and suppose that a researcher applies the SVAR procedure to these data without knowing that they were generated by our model. What would this researcher conclude? Note that as we vary the parameter values underlying our data-generating mechanism, we are simply asking how the conclusions of the researcher would change, not which parameter values of our model fit the data for some given country in some given time period.

In terms of our evaluation, we study the impulse response of hours worked to a technology shock and focus on a simple statistic designed to capture the difference between the impulse responses of the business cycle model and the SVARs. That statistic is the *impact error*, defined as the percentage difference between the mean across sequences of the SVAR impact coefficient and the model's impact coefficient.

We compute the impact error for a range of values of the relative variability of the two shocks: the ratio of the innovation variance of the nontechnology shock to that of the technology shock (σ_l^2/σ_z^2). To help interpret this relative variability, we translate it into more familiar units: the percentage of output variability due to technology shocks. We compute this percentage from the ratio of the variance of HP-filtered output with the technology shock alone relative to the variance of HP-filtered output with both shocks. We compute these variances from simulations of length 100,000.

In Fig. 1, panels A and B, the lines labeled *small-sample mean* plot the impact errors of the QDSVAR and LSVAR specifications against the percentage of the output variability due to technology shocks. The dashed lines represent the mean of the 95% bootstrapped confidence bands across the same 1,000 sequences used in computing the small-sample mean.

Notice that the impact errors for the QDSVAR specification are all negative, whereas those for the LSVAR specification are essentially all positive. Note that an error of -100% implies that the SVAR impact coefficient is zero (instead of .42), whereas any error more negative than -100% implies that the SVAR impact coefficient is negative. The figure reveals that when the percentage of output variability due to technology shocks is large, so that the variance of nontechnology shocks is small relative to that of technology shocks, the impact error is small in both specifications. As the percentage of output variability due to technology shocks decreases, the absolute value of the impact error increases.

Fig. 1 supports our main finding: the claim of the SVAR literature that this approach can confidently distinguish between promising and unpromising classes of models regardless of the details of the other shocks is incorrect. For the QDSVAR, we see that except when the technology shock accounts for more than 80% of the variability of output, the QDSVAR confidently gets the wrong answer on impact, in the sense that the confidence bands do not include 0% error. Moreover, unless technology shocks account for the bulk of output variability, say, more than 70%, the mean impact coefficient is negative, since the impact error is more negative than -100%.

For the LSVAR, we see that except when the technology shock accounts for virtually all of the variability of output, the confidence bands in the LSVAR are so wide that this procedure cannot distinguish between most models of interest. Here, unless the technology shock accounts for much more than 90% of output fluctuations, the confidence bands include negative values for the impact coefficient (that is, values for which the impact error is below -100%). Hence, as long as technology shocks account for less than 90% of output fluctuations, the LSVAR cannot distinguish between a class of models that predict a negative impact (like sticky price models) and a class of models that predict a positive impact (like real business cycle models). In terms of the impact error, note that when technology shocks account for less than 45% of the variability of output, the mean impact error is greater than 100%. Note also that the confidence bands for the LSVAR are wider than those for the QDSVAR.

We also report on the distribution of the impact error across simulations. For brevity's sake, we investigate this distribution for a particular ratio of the variance of the nontechnology shock to that of the technology shock that we refer to as the *Galí parameters*. Briefly, this ratio is set so that the mean impact coefficient equals $-.33$, the impact coefficient consistent with Galí's (1999) bivariate DSVAR.[7]

Fig. 2 reports on these distributions for the QDSVAR and the LSVAR. In panel A we see the distribution for the QDSVAR: almost all of the mass lies below the theoretical impact coefficient, and the distribution is fairly symmetric around its mean value. In panel B we see a different pattern for the LSVAR distribution: the mass is spread out over an extremely wide range,

[7] Galí (1999) reports that on impact, a one standard deviation technology shock leads to a $-.38\%$ change in hours. We convert this statistic to the response to a 1% technology shock, z, by dividing his statistic by the standard deviation of the technology shock. We use Prescott's (1986) measure of the standard deviation of an innovation to total factor productivity σ_{TFP} to construct the standard deviation of the technology shock σ_z. The relationship between these standard deviations is $\sigma_z = \sigma_{\mathrm{TFP}}/(1-\theta)$. Prescott measures σ_{TFP} to be .763, and our capital share is $\theta = .33$, so that after conversion Galí's statistic becomes $-.33$ $(= -.38/\sigma_z)$.

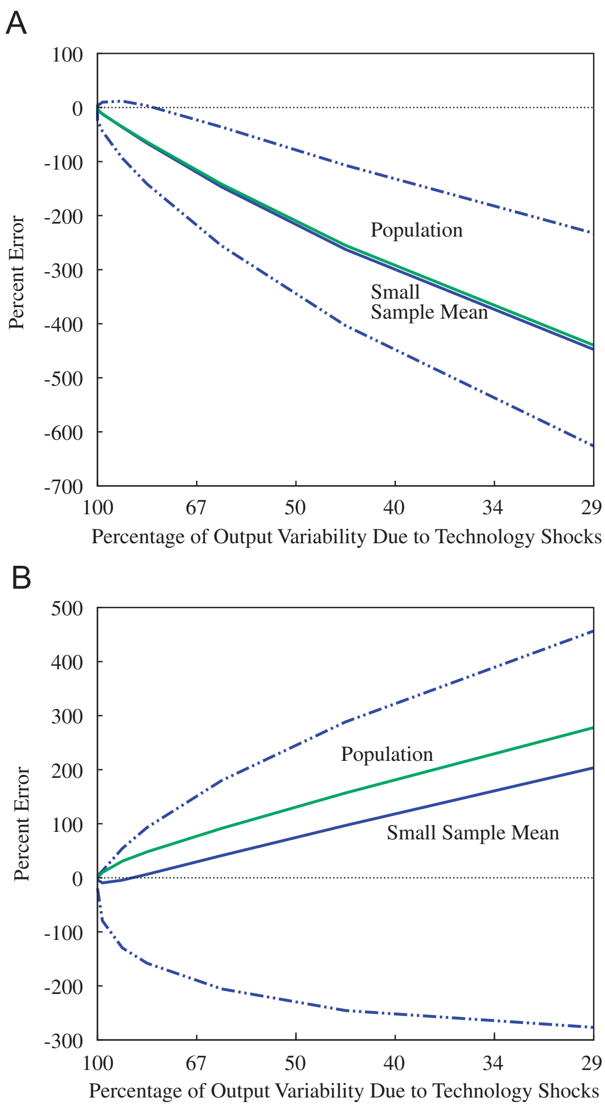

Fig. 1. Impact errors and confidence bands of the SVAR procedures: QDSVAR (A), LSVAR (B). *Note*: In both panels, small-sample impact errors are mean errors in the impact coefficient of hours from 1,000 applications of the four-lag SVAR procedures with $\rho_1 = .95$ applied to model simulations of length 180. Dashed lines are 95% confidence bands. Population errors are also shown.

which includes the model's impact coefficient. In CKM we analyze properties of the impulse responses in periods other than in the impact period and obtain similar results.

We also ask which specification a researcher would prefer, the QDSVAR or the LSVAR, on a priori grounds. The time series of hours worked in our model is highly serially correlated, and we find that standard unit root tests do not reject the hypothesis that the hours series has a unit root. At least since Hurwicz (1950), we have known that autoregressions on highly serially correlated variables are biased in small samples and that quasi-differencing such variables may diminish that bias. Since both the QDSVAR and the LSVAR specifications have desirable asymptotic properties, on a priori grounds the QDSVAR seems preferable.

So far we have simply assumed that researchers must choose either the QDSVAR specification or the LSVAR specification for all samples. In practice, researchers often conduct tests to determine which specification is preferable for their particular samples. Typically, they conduct unit root tests to determine whether hours should be specified in the VAR as levels or as first differences. We also investigated whether our findings are robust to a procedure that mimics the procedures conducted in practice, and we found that they are robust.

Researchers often conduct lag-length tests to determine the appropriate number of lags to include in their VARs. In an attempt to mimic a variant of the common approach that uses both lag-length tests and unit root tests, we experiment with variants of the SVAR procedure. For the QDSVAR specification, we retained only sequences that passed both the unit root

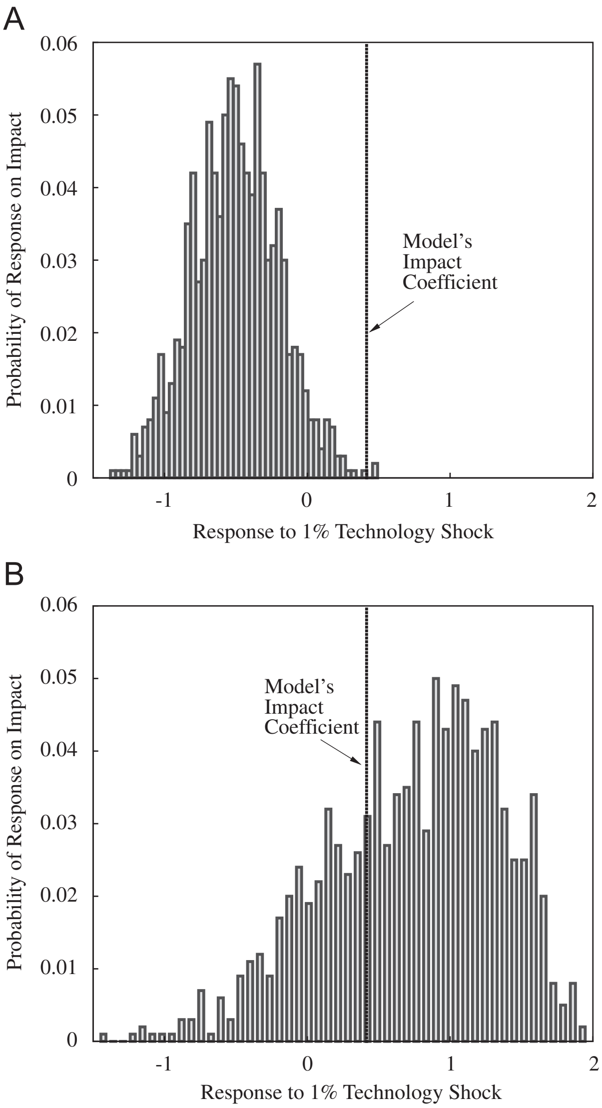

Fig. 2. Histogram of errors for Galí parameters: QDSVAR (A), LSVAR (B). *Note*: In both panels, histograms show the impact coefficients of hours from 1,000 applications of the four-lag SVAR procedures to model simulations of length 180 using Galí (1999) parameters.

test and the standard lag-length tests. We also allowed the lag length for each sequence to be determined by the lag-length tests. Our findings are virtually identical to those reported above. For the LSVAR specification, we retained only sequences that passed the lag-length test, and we allowed the lag length for each sequence to be determined by the lag-length tests. Here also, our results are virtually identical to those we have reported.

Considering the results from all our quantitative analysis, we conclude that for both specifications, the main claim of the SVAR literature is not correct.

4. Analyzing the SVAR's impulse response error

Here we investigate why the SVAR procedure fails our natural economic test. We determine that the problem with the procedure rests crucially on the auxiliary assumption equation (4), that Y_t has an autoregressive representation well approximated with a small number of lags. The impact error is large in our test when the business cycle model does not satisfy this assumption and small when it does. In all of the versions of our business cycle model, the auxiliary assumption is satisfied with an infinite number of lags ($p = \infty$). In practice, however, researchers are forced by the existing data lengths to run SVARs with a small number of lags, typically four. This lag truncation introduces a bias into the SVAR impulse responses. We here quantify how the lag-truncation bias varies with parameters and point out that, even though the VAR is truncated, when nontechnology shocks play a trivial role in output fluctuations, the impulse responses to a technology shock have no such bias.

4.1. Decomposition of the impact error: two biases

As we discuss below, if a VAR with an infinite number of lags were to be estimated on an infinite amount of data, then the impulse responses from the common approach would converge, in the usual sense, to the theoretical impulse responses. This discussion implies a natural decomposition of the impact error into that due to small-sample bias and that due to lag-truncation bias. We do this decomposition here and find that the SVAR error is primarily due to the lag-truncation bias.

Let $\bar{A}_0(p,T)$ denote the mean of the small-sample distribution of the matrix of impact coefficients of the SVAR impulse response when the VAR has p lags and the length of the sample is T. In practice, this mean is approximated as the mean across a large number of simulations. Since the observed variables have an autoregressive representation, it follows that $\bar{A}_0(p=\infty, T=\infty)$ coincides with the model's theoretical impulse response. That convergence implies that the (level of the) impact error associated with our implementation of the common approach is $\bar{A}_0(p=4, T=180) - \bar{A}_0(p=\infty, T=\infty)$. We can decompose this error into two parts:

$$[\bar{A}_0(p=4, T=180) - \bar{A}_0(p=4, T=\infty)] + [\bar{A}_0(p=4, T=\infty) - \bar{A}_0(p=\infty, T=\infty)].$$

The term in the first brackets is the *small-sample bias*, the difference between the mean of the SVAR impulse response over simulations of length 180 when the VAR has four lags and the SVAR population impulse response when the VAR has four lags. The term in the second brackets is the *lag-truncation bias*, the difference between the SVAR population impulse response when the VAR has four lags and the model's theoretical impulse response.

That VARs have small-sample biases has been known at least since Hurwicz (1950): even when the true model has a VAR with four lags, the estimated coefficients are biased in small samples. This type of bias is small for our model: for the QDSVAR specification, it is very small, and for the LSVAR specification, it is small compared to the lag-truncation bias. These findings can be seen in panels A and B of Fig. 1, which display the population estimates and means of those in the small sample. Note in the figure that the small-sample bias does not vary much with the relative variance of the nontechnology shock.

These findings lead us to focus on the lag-truncation bias. As we will show, with a sufficiently large number of lags, the lag-truncation bias becomes arbitrarily small. We ask how many lags are needed here for the lag-truncation bias to be small. The answer, we find, is too many.

Panel A of Fig. 3 displays the QDSVAR responses to a technology shock for lag lengths p ranging from 4 to 300. Notice that even with 20 lags, the lag-truncation bias of the QDSVAR specification is large. Note, too, that the convergence to the model's impulse response function is not monotonic. Finally, note that more than 200 lags are needed for the lag-truncation bias of the QDSVAR to be small.

Panel B of Fig. 3 shows the LSVAR impulse responses to the same shock for lag lengths p ranging from 4 to 100. Here, as with the QDSVAR, we see that the impulse response is a good approximation to the model's impulse response only for an extremely large number of lags. In practice, of course, accurately estimating VARs with so many lags is not feasible.

4.2. Intuition for the lag-truncation bias

Here we discuss two propositions that provide intuition for when the SVAR procedure performs poorly and when it performs well.

The first proposition shows that if both shocks have positive variance, then the observed variables have an infinite-order VAR representation. Therefore, any finite-order VAR suffers from truncation bias. The second proposition shows that if the variance of nontechnology shocks is small, then so is the truncation bias. Thus, here the VAR procedure with a small number of lags performs well.

Consider a state space system, of the form $X_{t+1} = AX_t + B\varepsilon_t$, and the *observer equation*, of the form $Y_t = CX_t + D\varepsilon_t$, with the same number of observables as shocks, so that the matrix D is square. Standard arguments (as in Fernández-Villaverde et al., 2007) lead to the following result:

Proposition 1 (*Existence of an infinite-order autoregressive representation*). *Consider any state space system in which D is invertible, the eigenvalues of A are less than* 1, *and the eigenvalues of $A - BD^{-1}C$ are strictly less than* 1. *Then the observed variables Y_t have an infinite-order autoregressive representation given by*

$$Y_t = B_1 Y_{t-1} + MB_1 Y_{t-2} + M^2 B_1 Y_{t-3} + \cdots + D\varepsilon_t, \tag{8}$$

where the decay matrix M is given by $M = C[A - BD^{-1}C]C^{-1}$ and $B_1 = CBD^{-1}$.

In CKM we show that for a wide range of parameters, the sufficient conditions in Proposition 1 are satisfied in our business cycle model for both the QDSVAR and LSVAR specifications. Thus, our model has a VAR

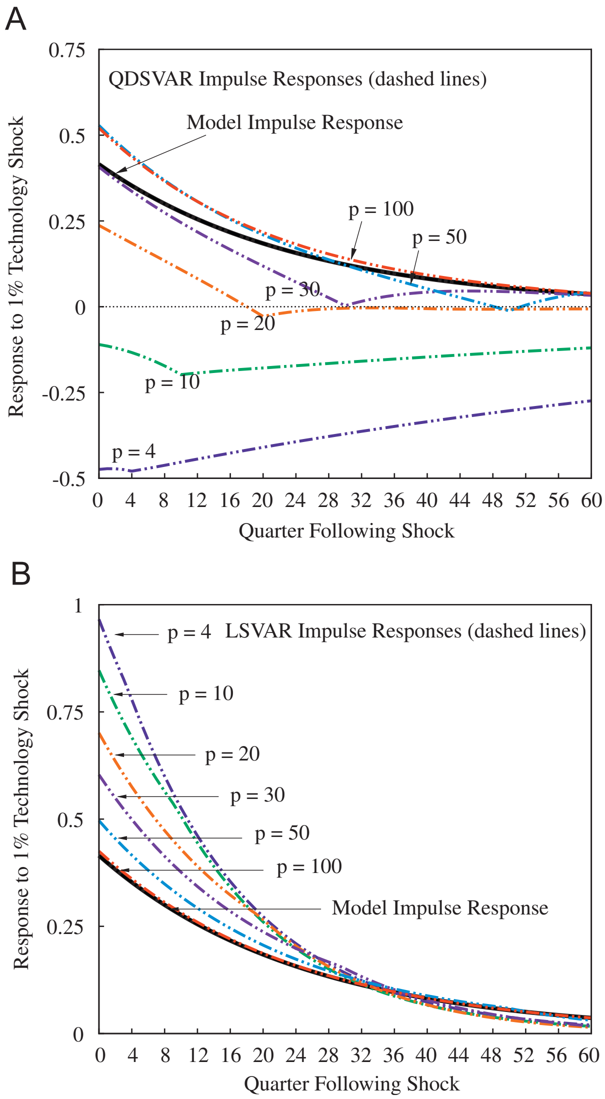

Fig. 3. Model and SVAR population responses of hours: QDSVAR (A), LSVAR (B). *Note*: In both panels, dashed lines are population impulse responses of hours to a technology shock using Galí (1999) parameters and varying the lag length p in the SVAR procedures.

representation with $p = \infty$. Since our model also satisfies the two key identifying assumptions of the SVAR procedures, we have that if a VAR with an infinite number of lags were run on an infinitely long sample of data generated by our model, then the impulse responses from these two SVAR specifications would coincide exactly (in the relevant sense of convergence) with those of the model. We emphasize that neither specification suffers from the invertibility problems discussed by Hansen and Sargent (1991) and Fernández-Villaverde et al. (2007). Moreover, neither specification suffers from issues of identification, overdifferencing, or specification error. Without more detailed quantitative analyses, theory provides no guidance as to which specification is preferable.

Given our model parameters, for the QDSVAR specification (including the quasi-differencing parameter $\alpha = .99$), the eigenvalues for M are $\lambda_1 = .99$ and $\lambda_2 = .96$, whereas for the LSVAR specification, they are $\lambda_1 = 0$ and $\lambda_2 = .96$. At our model parameters, for both specifications, the largest eigenvalue is close to 1. Since the rate of decay is, at least asymptotically, determined by the largest eigenvalue, these eigenvalues suggest that an autoregression with a small number of lags is a poor approximation to the infinite-order autoregression. It is not surprising, then, that the SVAR procedure performs poorly when both shocks have nontrivial variances.

We now give intuition for why the SVAR procedures perform well when the variance of the nontechnology shock is small relative to that of the technology shock. In CKM we show that, when the nontechnology shock is zero, for both specifications, the observed variables have an AR1 representation. By continuity we have the following proposition:

Proposition 2 (*Small truncation bias when nontechnology shocks small*). *In both the QDSVAR and the LSVAR specifications of the observed variables Y_t, as the variance of nontechnology shocks converges to zero, the truncation bias in an SVAR procedure that uses a VAR of order* 1 *converges to zero.*

In CKM we also show that a version of Proposition 2 applies for more general systems: loosely, if we have as many variables as shocks, then the truncation bias converges to zero as the variance of one of the nontechnology shocks converges to zero. This version of Proposition 2 sheds light on a literature that argues that sometimes SVARs with long-run restrictions work well. For example, Fernández-Villaverde et al. (2005) show that in Fisher's (2006) model, the population estimates from an SVAR procedure with one lag closely approximate the model's impulse responses. In CKM we show that Fisher's VAR system is a special case of a general version of Proposition 2.

5. Does adding variables and shocks help?

So far we have focused on an SVAR with just two variables—the log difference of labor productivity and a measure of the labor input—and two shocks—one to technology and one to nontechnology. In the SVAR literature, researchers often check how their results change when they add one or more variables and shocks to the SVAR. Would such an alteration to the SVAR we have been testing help improve its performance with the data from our business cycle model? We find that typically it does not.

In the data, the covariance of observed variables is almost always nonsingular, so that models that generate singular covariance matrices are uninteresting. If we simply add variables to the VAR without adding shocks to a business cycle model, the covariance matrix in the resulting VAR is often singular. To ensure that the covariance matrix in the VAR is nonsingular when we add more variables, we add more shocks.

The obvious candidate for a variable to include is some measure of capital. In practice, most researchers prefer using the investment–output ratio as a capital-like variable rather than the capital stock directly because they think the capital stock is poorly measured.

Let us see what happens with this ratio included in the VAR. Consider an SVAR with three variables and three shocks. The third variable is the log of the investment–output ratio. Here, in addition to the growth of labor productivity and the measure of labor, Y_t includes the investment–output ratio. In terms of additional shocks, we let the investment tax be the third shock. We assume that taxes on investment follow the autoregressive process

$$\tau_{xt+1} = (1-\rho_x)\bar{\tau}_x + \rho_x \tau_{xt} + \varepsilon^x_{t+1}, \tag{9}$$

where ε^x_t, together with our earlier shocks ε^z_t and ε^d_t, are jointly normal, independent of each other, and i.i.d. over time. The standard deviation of ε^x_t is σ_x. As we show in Chari et al. (2007b), taxes on investment in a prototype model arise either from financial frictions as in Bernanke et al. (1999) or from fluctuations in the technology for producing investment goods as in Greenwood et al. (2000).

We examine a quantitative version of our three-shock model with a three-variable LSVAR with $Y_t = (\Delta \log(y_t/l_t), l_t, x_t/y_t)$ and show that the impact errors are large even for small variances of the investment tax shock. To do so, we use Galí parameters for the labor tax shock and set $\rho_x = .95$ for the investment tax shock. We find, for example, when the investment tax accounts for 30% of output variability, both specifications have impact errors over 100% as well as extremely wide confidence bands. (For details, see CKM.)

Proposition 2 might be interpreted as suggesting that the SVAR procedure will approximately uncover the model's impulse response as long as a relatively small number of shocks (or factors) account for the bulk of fluctuations in the data. Our results here suggest that, in practice, such an interpretation should be treated with caution.

6. Are we confident that nontechnology shocks are trivial?

We have disproved the main claim of the literature on SVARs with long-run restrictions: that the SVAR procedure can confidently distinguish between promising and unpromising classes of models regardless of the details of nontechnology shocks, other than minimal assumptions like orthogonality. We have also shown that both specifications do reasonably well only when nontechnology shocks account for a trivial fraction of the variability in output.

Should a researcher be confident that nontechnology shocks play a trivial role for all cross-country data sets and all time periods? Presumably not. Should a researcher be confident that nontechnology shocks play a trivial role even in postwar U.S. data? We argue no. Indeed, many macroeconomists are interested in developing models in which nontechnology shocks play a substantial role in output fluctuations. And evidence from our model suggests that nontechnology shocks may actually play a nontrivial role. We also respond to our critics who have argued that macroeconomists are confident that nontechnology shocks have played a trivial role in output fluctuations in the postwar U.S. data.

6.1. Evidence from the literature

Here we argue that the business cycle literature and the SVAR literature provide evidence that macroeconomists are unwilling to confidently declare that nontechnology shocks are trivial. Indeed, the emerging consensus is that nontechnology shocks are significant.

6.1.1. The business cycle literature

The business cycle literature contains a wide range of procedures for estimating the percentage of output variability due to technology shocks and a wide range of resulting estimates. Prescott (1986) estimates this percentage at 76%. Eichenbaum (1991) estimates a related fraction and finds that it can range between 5% and 200%.[8] McGrattan (1994) obtains a point estimate of 41% with a standard error of 46%. Thus, a plausible case can be made that in the U.S. data, technology shocks account for essentially any value between 0% and 100% of output variance. Put differently, when the U.S. data are viewed through the lens of the growth model, dismissing any estimate in this range is unreasonable.

The emerging consensus in the business cycle literature, in fact, is that nontechnology shocks play an important role in output variability. This emerging consensus can be seen in the second generation of the real business cycle literature, which emphasizes that, in order for models to mimic the data, nontechnology shocks must account for a significant fraction of output volatility. (See, for example, the references in footnote 1.)

6.1.2. The SVAR literature

The SVAR literature also provides direct evidence on the relative contribution of technology shocks to the variability in output. Galí and Rabanal (2005), using an LSVAR specification, estimate that the percentage of output variability due to technology shocks ranges from 3% to 37%. (For the DSVAR specification, this percentage ranges from 6% to 31%.) Other SVAR studies, such as the one by Christiano et al. (2003), find estimates that range from 2% to 63%, depending on the specification.

6.2. Evidence from our model

We can also use our model to shed light on whether researchers should be confident that the only models worth developing are ones in which nontechnology shocks play a trivial role. As we argue in Chari et al. (2007b), in terms of aggregates many detailed models are observationally equivalent to a prototype model like the business cycle model we use here. Each detailed model implies a particular stochastic process for the shocks in our prototype model. We show here that the evidence from our prototype model suggests a substantial role for nontechnology shocks. Thus, using the logic of Chari et al. (2007b) on the mapping between prototype and detailed models, we argue that researchers should be interested in developing detailed models in which nontechnology shocks play a nontrivial role.

Here we discuss three pieces of evidence that suggest that developing such models is worthwhile. This evidence is based on the SVAR procedure's central finding, on the volatility of hours in the data, and on a variety of specifications of maximum likelihood estimation. With one exception, all of this evidence suggests that nontechnology shocks play a substantial role in output fluctuations. The exception is the maximum likelihood specification, which uses the growth of output and the level of hours as observables. This exception does not lead us to be confident that nontechnology shocks play a trivial role, however, because even within the maximum likelihood estimation context, other specifications are preferable.

6.2.1. Based on the SVAR central finding

One type of evidence on the relative size of the two shocks is based on the central result of the SVAR literature: Galí's (1999) widely noted finding that a positive technology shock drives down hours worked on impact. For our SVAR model to generate that finding, technology shocks must account for only a modest fraction of output variability.

We demonstrate that in Fig. 4 and Table 1. Fig. 4 adds to Fig. 1 details on the values of impact errors and confidence bands when the parameters of nontechnology shocks are set to the *Galí parameters*, that is, when the mean impact coefficient equals $-.33$, the impact coefficient consistent with Galí's (1999) bivariate DSVAR. At this value of the impact coefficient, the variance of output due to a technology shock is roughly 50% for our model, so that nontechnology shocks play an important role. (See the first row of Table 1.)

6.2.2. Based on the volatility of hours in U.S. data

Next, we ask how large must nontechnology shocks be if our business cycle model is to reproduce the volatility of U.S. hours? To answer that question, we set the standard deviation of the technology shock, σ_z, to reproduce Prescott's (1986)

[8] In a summary of the evidence on this fraction, Eichenbaum eloquently states, "What the data are actually telling us is that, while technology shocks almost certainly play some role in generating the business cycle, there is simply an enormous amount of uncertainty about just what percent of aggregate fluctuations they actually do account for. The answer could be 70% as Kydland and Prescott (1991) claim, but the data contain almost no evidence against either the view that the answer is really 5% or that the answer is really 200%" (Eichenbaum, 1991, p. 608).

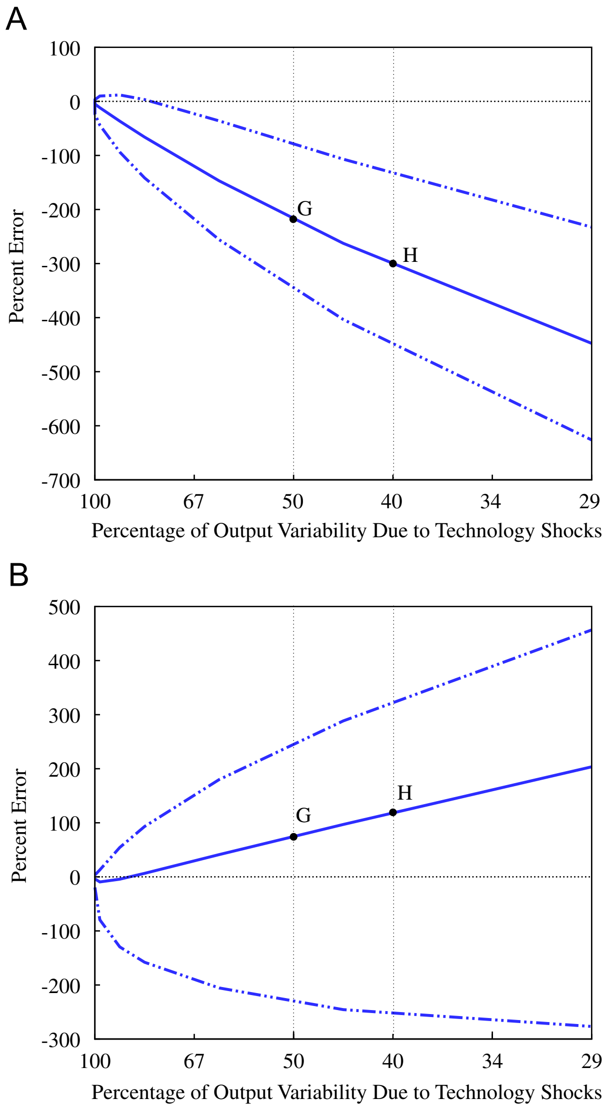

Fig. 4. Innovation variance ratio implied by Galí parameters (G) and U.S. hours (H): QDSVAR (A), LSVAR (B). *Note*: In both panels, the solid line is the small-sample mean impact error in the coefficient of hours from 1,000 applications of the four-lag SVAR procedures with $\rho_1 = .95$ applied to model simulations of length 180. Dashed lines are 95% confidence bands. Point G corresponds to the Galí (1999) parameters, while point H corresponds to parameters which reproduce the variance of U.S. hours.

measure of the standard deviation of an innovation to total factor productivity σ_{TFP}. We then ask, more precisely, what must be the volatility (standard deviation) of the nontechnology shock, σ_l, in order to reproduce the observed volatility in hours? We find that at this level of nontechnology volatility, technology shocks account for roughly 40% of the observed volatility in output.

Now, returning to Fig. 4 and Table 1, we can examine the performance of the SVAR procedure with the QDSVAR and the LSVAR specifications at this setting of nontechnology and technology shocks. We see that the impact error for the QDSVAR specification is -300% (Table 1) and that this specification confidently rejects the possibility that the impact coefficient is positive (Fig. 4). At this level of volatility, the impact error for the LSVAR is 118%, but the confidence bands for this specification are so wide that it cannot distinguish between models of interest.

6.2.3. Based on maximum likelihood estimation

A last type of evidence on the relative sizes of the two shocks is the results of maximum likelihood estimation. In our estimation procedure, we fix all the parameters of the model except for those of the stochastic processes. We then use a standard maximum likelihood procedure to estimate the parameters of the vector AR1 process, Eqs. (6) and (7), using several specifications for the observed variables.

Table 1
Parameter estimates and statistics of interest for the model with taxes on labor

Evidence	Parameter estimates			Statistics of interest[a]		
	ρ_l	σ_z	σ_l	%var(y)	Impact error	
					QDSVAR	LSVAR
Galí VAR response	.950	.0114	.0073	50	−220 (−344, −79)	76 (−230, 245)
Hours volatility	.950	.0114	.0088	40	−300 (−448, −132)	118 (−252, 322)
Maximum likelihood[b]						
Hours specification	.995 (.0093)	.0114 (.0006)	.0050 (.0005)	76	−86 (−171, −5)	3 (−219, 123)
Investment specification	.942 (.0076)	.0178 (.0016)	.0173 (.0013)	30	−438 (−616, −226)	190 (−270, 442)

[a] The first statistic is the variance of output due to the technology shock, reported as a percentage. The last two are the mean impact errors for the QDSVAR and LSVAR specifications. The values in parentheses are means of the upper and lower means of 95% confidence bands across 1,000 applications of the SVAR procedures.

[b] For the maximum likelihood parameter estimates, the values in parentheses are standard errors. The hours specification uses observations on output and labor; the investment specification, observations on output and investment.

Table 2
Monte Carlo analysis of maximum likelihood estimation for two sets of observables in the model with taxes on labor

Estimates	Hours specification[a]			Investment specification[a]		
	ρ_l	σ_z	σ_l	ρ_l	σ_z	σ_l
True estimates	.990	.0100	.0100	.990	.0100	.0100
Monte Carlo estimates						
Mean	.980	.0101	.0096	.990	.0100	.0100
% Standard deviation	1.83	.053	.084	.076	.053	.083

[a] The hours specification uses observations on output and labor; the investment specification, observations on output and investment.

In the *hours specification*, we let the observed variables be $Y_t = (\Delta \log y_t, \log l_t)'$. In the *investment specification*, we let $Y_t = (\Delta \log y_t, \Delta \log x_t)'$. In both specifications, we impose an upper bound of .995 on the persistence parameter ρ_{l}. Table 1 displays the results of the estimation. In the hours specification, the variability of output due to technology is fairly large, 76%; the impact error for the QDSVAR is −86%; and the impact error for the LSVAR is 3%. In the investment specification, the variability of output due to technology is more modest, 30%; the impact error for the QDSVAR is −438%; and the impact error for the LSVAR is 190%. Clearly, the impact error for both of these specifications depends sensitively on the specification of observed variables. But note that the confidence bands for all of these impact errors are huge.

We then ask which specification is preferable, in the sense that it leads to more accurate estimates of the key parameters of the stochastic process. To answer this question, we conduct Monte Carlo experiments for our business cycle model. We set the key parameters at $\rho_{\mathrm{l}} = .99$, $\sigma_{\mathrm{l}} = 1\%$, and $\sigma_z = 1\%$. We generate 1,000 simulations of the same length as the actual data. For each simulation, we estimate the parameters of the stochastic process with maximum likelihood, using the specifications of the observed variables. We impose the same bound on ρ_{l} of .995 as in our estimation using actual data.

From Table 2 we see that the investment specification clearly yields more accurate estimates of the model parameters than does the hours specification. We repeated this exercise using higher values of ρ_{l} and found that the investment specification continues to yield more accurate estimates of the model parameters. These findings lead us to prefer the investment specification for estimating the model's parameters.

Clearly, the variability of output due to technology shocks associated with the maximum likelihood estimates is sensitive to the variables included in the observer equation, especially investment. The reason for this sensitivity is that a stripped-down model like ours cannot mimic well all of the comovements in U.S. data, so it matters what features of the data the researcher is primarily interested in. Full information methods like maximum likelihood turn out to be sensitive to details such as which variables are included in the estimation. Our Monte Carlo experiments lead us to prefer the investment specification. And this specification leads to a large impact error for the LSVAR.

6.3. *Response to our critics*

Christiano et al. (2007) have criticized our critique of SVARs. We agree with them that if researchers are only interested in developing models in which nontechnology shocks play a trivial role, then both the QDSVAR and the LSVAR procedures are useful guides. We have argued, however, that the vast majority of researchers are interested in developing models in which nontechnology shocks play a nontrivial role and that for such researchers both the QDSVAR and the LSVAR procedures using the common approach are likely to be useless.

They have also pointed to one piece of evidence that nontechnology shocks play at best a modest role in postwar U.S. data. This piece essentially consists of our maximum likelihood estimate for the hours specification. While this piece of evidence, taken in isolation, does buttress that case, the bulk of the evidence from the literature and our model argues for assuming nontechnology shocks play a nontrivial role.

We have also argued that structural VARs may be useful in developing business cycle theory if researchers use the Sims–Cogley–Nason approach rather than the common approach. We are puzzled why our critics seem so reluctant to use the Sims–Cogley–Nason approach. This approach seems straightforward: researchers simply run the identical VAR in the model that they run in the data. In our notation this means that researchers compute $\bar{A}_i(p = 4, T = 180)$ for their model and compare it to the impulse response in the data. Doing so will eliminate the problems associated with both the lag-truncation bias and the small-sample bias.

Computing $\bar{A}_i(p = 4, T = 180)$ typically involves Monte-Carlo simulations. For researchers who want to avoid such simulations, an alternative approach is to compare $\bar{A}_i(p = 4, T = \infty)$ to the impulse response in the data. This comparison eliminates the problems associated with the lag-truncation bias but not those associated with the small-sample bias. In this sense this approach is superior to the common approach.

Can the Sims–Cogley–Nason approach be used when, as is often the case in practice, researchers want to add more variables to the VAR beyond the two or three considered here? Yes. Of course, in order to avoid singularity when running the VAR on model-generated data, researchers must add at least as many shocks to the model as the number of variables added to the VAR. The reluctance to use this approach might stem from an unwillingness to model the additional shocks in detail. This reluctance is often justified by the idea in the common approach literature that this procedure works well regardless of the specification of the additional shocks. As we have shown, this idea is false.

In sum, we see no coherent argument for the use of the common approach over the Sims–Cogley–Nason approach or the alternative approach described above.

7. Conclusion

The central finding of the recent structural vector autoregression (SVAR) literature with a differenced specification of hours is that technology shocks lead to a fall in hours. Researchers have used this finding to argue that real business cycle models are unpromising. We subjected this SVAR specification to a natural economic test by showing that when applied to data generated from a multiple-shock business cycle model, the procedure incorrectly concludes that the model could not have generated the data as long as nontechnology shocks play a nontrivial role. We also tested another popular specification, which uses the level of hours, and showed that with nontrivial nontechnology shocks, it cannot distinguish between real business cycle models and sticky price models. The crux of the problem for both SVAR specifications is that available data necessitate a VAR with a small number of lags and, when nontechnology shocks play a nontrivial role, such a VAR is a poor approximation to the model's VAR.

Elsewhere (in Chari et al., 2007b), we have argued for the usefulness of another approach to developing business cycle theory: business cycle accounting. This approach has the same goal as the SVAR approach—to quickly shed light on which class of models is promising—but business cycle accounting suffers from fewer shortcomings.

References

Bencivenga, V.R., 1992. An econometric study of hours and output variation with preference shocks. International Economic Review 33, 449–471.

Benhabib, J., Rogerson, R., Wright, R., 1991. Homework in macroeconomics: household production and aggregate fluctuations. Journal of Political Economy 99, 1166–1187.

Bernanke, B.S., Gertler, M., Gilchrist, S., 1999. The financial accelerator in a quantitative business cycle framework. In: Taylor, J.B., Woodford, M. (Eds.), Handbook of Macroeconomics, vol. 1C. North-Holland, Elsevier, Amsterdam, pp. 1341–1393.

Blanchard, O.J., Quah, D., 1989. The dynamic effects of aggregate demand and supply disturbances. American Economic Review 79, 655–673.

Braun, R.A., 1994. Tax disturbances and real economic activity in the postwar United States. Journal of Monetary Economics 33, 441–462.

Chari, V.V., Kehoe, P.J., McGrattan, E.R., forthcoming. New Keynesian models: not yet useful for monetary policy analysis. American Economic Journal: Macroeconomics.

Chari, V.V., Kehoe, P.J., McGrattan, E.R., 2007a. Are structural VARs with long-run restrictions useful in developing business cycle theory? Staff Report 364, Federal Reserve Bank of Minneapolis.

Chari, V.V., Kehoe, P.J., McGrattan, E.R., 2007b. Business cycle accounting. Econometrica 75, 781–836.

Christiano, L.J., Eichenbaum, M., Vigfusson, R., 2003. What happens after a technology shock? NBER Working Paper 9819.

Christiano, L.J., Eichenbaum, M., Evans, C.L., 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. Journal of Political Economy 113, 1–45.

Christiano, L.J., Eichenbaum, M., Vigfusson, R., 2007. Assessing structural VARs. In: Acemoglu, D., Rogoff, K., Woodford, M. (Eds.), NBER Macroeconomics Annual 2006. MIT Press, Cambridge, pp. 1–106.

Cogley, T., Nason, J.M., 1995. Output dynamics in real-business-cycle models. American Economic Review 85, 492–511.

Cooley, T.F., Dwyer, M., 1998. Business cycle analysis without much theory: a look at structural VARs. Journal of Econometrics 83, 57–88.

Cooley, T.F., Hansen, G.D., 1989. The inflation tax in a real business cycle model. American Economic Review 79, 733–748.

Eichenbaum, M., 1991. Real business-cycle theory: wisdom or whimsy? Journal of Economic Dynamics and Control 15, 607–626.

Erceg, C.J., Guerrieri, L., Gust, C., 2004. Can long-run restrictions identify technology shocks? International Finance Discussion Paper 792, Board of Governors of the Federal Reserve System.
Faust, J., Leeper, E.M., 1997. When do long-run identifying restrictions give reliable results? Journal of Business and Economic Statistics 15, 345–353.
Fernández-Villaverde, J., Rubio-Ramírez, J.F., Sargent, T.J., 2005. A, B, C's, (and D's) for understanding VARs. NBER Technical Working Paper T0308.
Fernández-Villaverde, J., Rubio-Ramírez, J.F., Sargent, T.J., Watson, M., 2007. A, B, C's, (and D's) for understanding VARs. American Economic Review 97, 1021–1026.
Fisher, J.D.M., 2006. The dynamic effects of neutral and investment-specific technology shocks. Journal of Political Economy 114, 413–451.
Francis, N., Ramey, V.A., 2005. Is the technology-driven real business cycle hypothesis dead? Shocks and aggregate fluctuations revisited. Journal of Monetary Economics 52, 1379–1399.
Galí, J., 1999. Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations? American Economic Review 89, 249–271.
Galí, J., Rabanal, P., 2005. Technology shocks and aggregate fluctuations: how well does the RBC model fit postwar U.S. data? In: Gertler, M., Rogoff, K. (Eds.), NBER Macroeconomics Annual 2004, vol. 19. MIT Press, Cambridge, pp. 225–288.
Goodfriend, M., King, R., 1997. The new neoclassical synthesis and the role of monetary policy. In: Bernanke, B., Rotemberg, J. (Eds.), NBER Macroeconomics Annual 1997, vol. 12. MIT Press, Cambridge, pp. 231–283.
Greenwood, J., Hercowitz, Z., Krusell, P., 2000. The role of investment-specific technological change in the business cycle. European Economic Review 44, 91–115.
Hall, R.E., 1997. Macroeconomic fluctuations and the allocation of time. Journal of Labor Economics 15, S223–S250.
Hansen, L.P., Sargent, T., 1991. Two difficulties in interpreting vector autoregressions. In: Rational Expectations Econometrics. Westview Press, Boulder, pp. 77–119.
Hurwicz, L., 1950. Least squares bias in time series. In: Koopmans, T.C. (Ed.), Statistical Inference in Dynamic Economic Models. Wiley, New York, pp. 365–383.
King, R., Plosser, C., Rebelo, S., 1988. Production, growth and business cycles: I. The basic neoclassical model. Journal of Monetary Economics 21, 195–232.
Kydland, F.E., Prescott, E.C., 1982. Time to build and aggregate fluctuations. Econometrica 50, 1345–1370.
Kydland, F.E., Prescott, E.C., 1991. Hours and employment variation in business cycle theory. Economic Theory 1, 61–81.
Marcet, A., 2005. Overdifferencing VARs is OK. Manuscript, Universitat Pompeu Fabra.
McGrattan, E.R., 1994. The macroeconomic effects of distortionary taxation. Journal of Monetary Economics 33, 573–601.
Prescott, E.C., 1986. Theory ahead of business cycle measurement. Federal Reserve Bank of Minneapolis Quarterly Review 10, 9–22.
Rotemberg, J.J., Woodford, M., 1992. Oligopolistic pricing and the effects of aggregate demand on economic activity. Journal of Political Economy 100, 1153–1207.
Sims, C.A., 1971. Distributed lag estimation when the parameter space is explicitly infinite-dimensional. Annals of Mathematical Statistics 42, 1622–1636.
Sims, C.A., 1972. The role of approximate prior restrictions in distributed lag estimation. Journal of the American Statistical Association 67, 169–175.
Sims, C.A., 1989. Models and their uses. American Journal of Agricultural Economics 71, 489–494.
Stockman, A.C., Tesar, L.L., 1995. Tastes and technology in a two-country model of the business cycle: explaining international comovements. American Economic Review 85, 168–185.