

Federal Reserve Bank of Minneapolis
Research Department Staff Report 379

October 2006

How to Advance Theory with Structural VARs: Use the Sims-Cogley-Nason Approach

Patrick J. Kehoe*

Federal Reserve Bank of Minneapolis,
University of Minnesota,
and National Bureau of Economic Research

ABSTRACT

The common approach to evaluating a model in the structural VAR literature is to compare the impulse responses from structural VARs run on the data to the theoretical impulse responses from the model. The Sims-Cogley-Nason approach instead compares the structural VARs run on the data to identical structural VARs run on data from the model of the same length as the actual data. Chari, Kehoe, and McGrattan (2006) argue that the inappropriate comparison made by the common approach is the root of the problems in the SVAR literature. In practice, the problems can be solved simply. Switching from the common approach to the Sims-Cogley-Nason approach basically involves changing a few lines of computer code and a few lines of text. This switch will vastly increase the value of the structural VAR literature for economic theory.

*Forthcoming in the *NBER Macroeconomics Annual 2006*. This work is a response to the comments of Lawrence Christiano, Martin Eichenbaum, and Robert Vigfusson (forthcoming in the *NBER Macroeconomics Annual 2006*) on the critique of structural VARs with long-run restrictions by V.V. Chari, Patrick Kehoe, and Ellen McGrattan. The author thanks numerous economists including his coauthors, Tim Cogley, Jesus Fernandez-Villaverde, Bob Hall, Chris House, Narayana Kocherlakota, Ricardo Lagos, Monika Piazzesi, Juan Rubio-Ramirez, Tom Sargent, Martin Schneider, and Jim Stock for very helpful comments. The author also thanks the NSF for support and Kathy Rolfe and Joan Gieseke for excellent editorial assistance. Any views expressed here are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

Most of the existing structural VAR literature argues that a useful way of advancing theory is to directly compare impulse responses from structural VARs run on the data to theoretical impulse responses from models. The crux of the Chari, Kehoe, and McGrattan (2006) (henceforth, CKM) critique of this *common approach* is that it compares the empirical impulse responses from the data to inappropriate objects in the model. We argue that instead of being compared to the theoretical impulse responses, the empirical impulse responses should be compared to impulse responses from identical structural VARs run on data from the model of the same length as the actual data. We refer to this latter approach as the *Sims-Cogley-Nason approach* since it has been advocated by Sims (1989) and successfully applied by Cogley and Nason (1995).

CKM argue that in making the inappropriate comparison, the common approach makes an error mistake relative to the Sims-Cogley-Nason approach. That error makes the common approach prone to various pitfalls, including small-sample bias and lag-truncation bias. For example, the data length may be so short that the researcher is forced to use a short lag length, and the estimated VAR may be a poor approximation to the model's infinite-order VAR. In contrast, since the Sims-Cogley-Nason approach treats the data from the U.S. economy and the model economy symmetrically, it avoids the potential problems of the common approach.

On purely logical grounds, then, the Sims-Cogley-Nason approach seems to dominate the common approach.¹ How well does the common approach do in practice using SVARs based on long-run restrictions on data from a real business cycle model? CKM show that for data of the relevant length, SVARs do miserably: the bias is large and SVARs are unable to distinguish between models of interest—unless technology shocks account for virtually all the fluctuations in output.

Christiano, Eichenbaum, and Vigfusson (2006) (henceforth, CEV), perhaps the most prominent defenders of the common approach, seem to agree with CKM on the most important matters of substance. Indeed, since there seems to be no dispute that the Sims-Cogley-Nason approach dominates the common approach, there should be little disagreement over how future research in this area should be conducted. Likewise, there seems to be no dispute that when shocks other than technology play a sizable role in output fluctuations, SVARs do miserably. The primary point of disagreement between CEV and CKM is thus a relatively minor one about the likely size of the errors in the past literature that uses the common approach. CEV argue that the errors are small because the evidence is overwhelming that in U.S. data, technology shocks account for virtually all the fluctuations in output. CKM point to both 20 years of business cycle research and simple statistics in the data that all lead to the opposite conclusion about technology shocks and, hence, to the opposite conclusion as to

the size of the errors of the common approach.

CEV also venture beyond the confines of the CKM critique and analyze SVARs with short-run restrictions. They focus on SVARs applied to monetary models which satisfy the same recursive identifying assumptions as their VARs. CEV argue that the error in this application of the common approach is small, and thus the technique can be used broadly to distinguish promising models from the rest. Here the primary problem with their analysis is that it is subject to the Lucas and Stokey critique (Lucas and Stokey 1987): only a tiny subset of existing monetary models in the literature actually satisfies the recursive identifying assumptions. That subset does not include even, for example, the best-known monetary models of Lucas (1972 and 1990). Yet the technique has been used to reject these and other such models. Clearly, comparing impulse responses from SVARs with a set of identifying assumptions to those from models which do not satisfy those assumptions is problematic.

Notice that the Sims-Cogley-Nason approach is immune to the Lucas and Stokey critique. Under this approach, it is entirely coherent to compare impulse responses with a set of identifying assumptions to those from models which do not satisfy these assumptions. Under this approach, the impulse responses are simply statistics with possibly little economic interpretation. Now, those statistics may not be interpretable as being close to the model's theoretical response, but so what? When Kydland and Prescott (1982) compare variances, covariances, and cross-correlations in the model and the data, it does not matter whether these statistics have some deep economic interpretation.

Of course, it is not true that all statistics are equally desirable. What properties lead certain statistics to be more desirable than others? One important property is that the statistics vary across alternative models in such a way that, with samples of the lengths we have, they can be used to point with confidence toward one class of models and away from another. (If no such statistics exist, then the data have little to say about the theories of interest.) A second desirable property is that the statistics depend on key features of theory and not on inessential auxiliary assumptions. An important question for a serious assessment of the SVAR literature is, in what sense are the SVAR statistics more or less desirable than a host of other non-SVAR-related statistics? Regrettably, there seems to be little or no work in the SVAR literature directed to this critical question.

To reiterate: The CKM critique does not apply to all SVAR analyses, only those that use the common approach rather than the Sims-Cogley-Nason approach. Switching to that dominant approach would cost little—changing only a few lines of computer code and a few lines of text. By making such a switch, researchers using the SVAR approach can vastly enhance their role in guiding theory.

In these comments, I begin by carefully describing the difference between the common

approach and the Sims-Cogley-Nason approach. Then I describe four issues of perceived disagreement between CKM and CEV about SVARs with long-run restrictions. Finally, in terms of CEV's analysis with short-run restrictions, I describe two critiques which need to be addressed by researchers who steadfastly refuse to abandon the common approach.

1. Getting Precise

Let me begin with some notation with which I can make the CKM argument precise.

The first step in both SVAR approaches is to run a VAR with p lags on a data set $\{Y_t\}_{t=1}^T$, then apply the identifying assumptions to construct the impulse response matrices $A_i(p, T)$ for $i = 0, 1, \dots$, where i denotes periods after the impact period and the notation emphasizes that the impulse responses depend on the lag length p and the sample size T . In applications using postwar U.S. data, it is common to set $p = 4$ and to have $T = 180$ or numbers similar to these, and I will denote the resulting matrices by $A_i^{US}(p = 4, T = 180)$.

The common approach emphasizes the interpretation of these matrices. For instance, in the standard example, the data consist of a measure of labor productivity and a measure of hours and the theoretical model has two shocks, technology and non-technology shocks. The first column of the impact matrix $A_0^{US}(p = 4, T = 180)$ is interpreted as the impact effect of the technology shock on productivity and hours, while the second column is interpreted as the impact effect of the non-technology shock on productivity and hours. The subsequent matrices are similarly interpreted.

In contrast, CKM and Sims, Cogley, and Nason view these matrices as moments of the data that may be used in discriminating among models of interest.

Now suppose we have a quantitative economic model in which the impulse responses to the technology and non-technology shocks are the matrices $D_i(\theta)$, $i = 0, 1, \dots$, where θ denotes the model parameters. The second step of the common approach compares

$$(1) \quad A_i^{US}(p = 4, T = 180) \text{ to } D_i(\theta).$$

Sometimes this comparison is informal and implicit, as in the work of Galí (1999), Francis and Ramey (2005), and Galí and Rabanal (2005), who find the labor falls after a positive productivity shock and conclude that real business cycles are dead. Sometimes this comparison is formal and explicit, as in the work of Altig et al. (2005), and is used to choose model parameters θ .

The second step of the Sims-Cogley-Nason approach is quite different. To understand this step, let $\bar{A}_i(p, T|\theta)$ denote the mean of impulse responses found by applying the SVAR approach with p lags in the VAR to the many simulations of data of length T generated from the model with parameters θ . The second step of the Sims-Cogley-Nason approach compares

$$(2) \quad A_i^{US}(p = 4, T = 180) \text{ to } \bar{A}_i(p = 4, T = 180|\theta).$$

At a conceptual level, we interpret the Sims-Cogley-Nason approach as advocating comparing the exact small-sample distribution of the estimator of the impulse responses with $p = 4$ and $T = 180$ to the estimated impulse response parameters. We view the simulations involved as a simple way to approximate that small-sample distribution. If it were feasible to analytically work out the small-sample distribution of the estimator, then so much the better.

CKM interpret (2) as the correct comparison, which is firmly grounded in (simulated) method-of-moments theory, and (1) as simply a mistake of the common approach.

The whole point of the CKM work is to quantify when and why these two comparisons will yield different answers, that is, when and why the two objects computed from the model, $\bar{A}_i(p = 4, T = 180|\theta)$ and $D_i(\theta)$, will differ. Part of CKM's analysis focuses on the two-variable case with $Y_t(\alpha) = (\Delta(y_t/l_t), l_t - \alpha l_{t-1})'$, where y_t is the log of output, l_t is the log of hours, and $\alpha \in [0, 1]$ is the quasi-differencing parameter. The specification $Y_t(\alpha)$ nests three cases of interest: $\alpha = 0$, the level SVAR (LSVAR) case; $\alpha = 1$, the differenced SVAR (DSVAR) case; and $\alpha = .99$, the quasi-differenced SVAR (QDSVAR) case.

When θ is such that technology shocks do not account for the vast bulk of fluctuations in output, LSVARs do miserably: the bias is large and the confidence bands are so enormous that the technique is unable to distinguish among most classes of models of interest.

With such a θ , the DSVARs and QDSVARs also fare poorly: the bias is large enough to flip the sign of the impact coefficient of hours on a technology shock. While the confidence bands are large, they don't stop a researcher from rejecting that the simulated data came from a real business cycle model, even though they did. CKM think that this result suggests that researchers who have determined that real business cycle models are dead based on SVAR evidence may have come to that conclusion simply because they were not comparing the appropriate objects in the model and the data.

Note that, at least for the long-run restriction branch, the issue is all about approximation error. If we had an infinite sample of data from a model that satisfies the identifying restrictions and we estimated a VAR with an infinite number of lags, we would have (in the relevant sense of convergence)

$$(3) \quad \bar{A}_i(p = \infty, T = \infty) = D_i(\theta)$$

for both the LSVAR and the QDSVAR cases, where, for simplicity, we have assumed that the identifying assumptions are sufficient as well as necessary. (As we discuss below, Marcet (2005) shows why (3) holds even for the DSVAR case in which hours are "over-differenced.")

With (1)–(3) in mind, note that $\bar{A}_i(p = 4, T = 4) - D_i(\theta)$ can be interpreted as the error in the common approach relative to the Sims-Cogley-Nason approach. CKM decompose

this error into

$$[\bar{A}_i(p = 4, T = 180) - \bar{A}_i(p = 4, T = \infty)] + [\bar{A}_i(p = 4, T = \infty) - D_i(\theta)],$$

where the first term is the Hurwicz-type *small-sample bias* and the second term is the *lag-truncation bias*. It turns out that for both the LSVAR case and the QDSVAR case, most of the error is coming from the lag-truncation bias. Intuitively, this truncation bias arises both because the $p = 4$ specification forced terms to be zero that are not and that the OLS estimator adjusts the estimates of the included lags to compensate for those that have been excluded. CKM develop propositions that give intuition for when the error from the lag-truncation bias will be large.²

2. The Common Approach With Long-Run Restrictions . . .

The SVAR literature with long-run restrictions, in general, and CEV, in particular, claim that the common approach is a state-of-the-art technique which is a useful guide for theory. We disagree. Here I describe three specific points of disagreement relevant to CEV's discussion of long-run restrictions and one point in which CEV seem to think there is disagreement when none really exists. Overall, here my point is that we all agree there exist circumstances under which the errors from using the common approach are small; however, as CKM have shown, these circumstances are not general. Moreover, regardless of the circumstances, this approach is dominated by what we consider the state-of-the-art technique, the Sims-Cogley-Nason approach. This approach is at least as easy to use as the common approach, and it has the advantage of a firm logical and statistical foundation.

Consider now the four points.

First, CEV argue that LSVARs are useful in guiding theory about fluctuations in the U.S. economy because in U.S. data, they say, technology shocks account for almost all of the fluctuations in output. We argue that while some reasonable statistics do point to technology shocks playing an overwhelming role, a number of other sensible statistics, as well as much of the literature, strongly suggest that their role is modest.

Second, CEV argue that even if technology shocks do not account for almost all of the fluctuations in output, there is a new estimator of impulse responses that virtually eliminates the bias associated with the standard OLS estimator. We argue that while for some parameter values this new estimator improves on the OLS estimator, for others it does worse. In this sense, the new estimator does not solve all the problems facing this literature.

Third, CEV ignore the DSVAR literature on the grounds, they say, that the DSVAR is misspecified because it incorrectly differences hours. This misspecification, they say, leads to incorrect estimates of impulse responses even with an infinite amount of data. We argue that here, for all practical purposes, CEV are wrong about the DSVAR being misspecified.

Instead the only error in the DSVAR literature is the same as in the LSVAR literature: using the common approach rather than the Sims-Cogley-Nason approach.

Finally, I consider a point on which there is actually no disagreement. CEV argue that when more variables are added to an LSVAR, in special cases it can sometimes usefully distinguish between classes of models. CEV somehow seem to think we disagree here, but we do not. Indeed, part of the point of CKM is to provide a theorem as to when LSVARs can and cannot perform this function. We emphasize, however, that the “can” circumstances are somewhat narrow.

2.1 Do Technology Shocks Account for Virtually All of the Fluctuations in Output?

CKM show that if technology shocks account for virtually all of the fluctuations in output, then the errors associated with the common approach are relatively small. Much of CEV’s work is devoted to arguing that the U.S. data definitively show that technology shocks account for the vast bulk of the movements in output and non-technology shocks, almost none. There is a vast literature on this subject, much of it contradicting that stand.

Let’s take a closer look at the issues at stake. Using the notation of CKM and ignoring means, we can write the stochastic processes for a technology shock, $\log Z_t$, and a non-technology shock, τ_{lt} , for both CEV and CKM, as

$$(4) \quad \log Z_{t+1} = \log Z_t + \log z_{t+1}$$

$$(5) \quad \tau_{lt+1} = \rho\tau_{lt} + \varepsilon_{lt+1},$$

where $\log z_t$ and ε_{lt} are independent, mean zero, i.i.d. normal random variables with variances σ_z^2 and σ_l^2 and ρ is the serial correlation of the non-technology shock. Note that these stochastic processes are determined by three parameters $(\sigma_z^2, \sigma_l^2, \rho)$. CEV estimate these parameters to be $\sigma_z^2 = (.00953)^2$, $\sigma_l^2 = (.0056)^2$, and $\rho = .986$. CKM show that the impulse errors in the SVARs increase with the ratio of the variances of the innovations σ_l^2/σ_z^2 .

CEV’s finding that LSVARs do well with U.S. data rests crucially on their estimate of the variance of non-technology shocks. CKM and CEV agree that LSVARs do miserably when this variance is large. The main disagreement between us here is whether we can confidently assert that, when the U.S. data are viewed through the lens of a real business cycle model, the variance of non-technology shocks is, indeed, small. CEV do not make clear that at a mechanical level, the only source of their disagreement with us is the relevant values of that one parameter σ_l^2 . Here, to demonstrate that point, I set all of the parameters, except σ_l^2 , equal to those of CEV.

The question then is, What is a reasonable value for the variance of non-technology shocks? Before confronting this question formally, recall a well-known fact: In real business

cycle models with unit root technology shocks, the volatility of hours due to technology shocks is tiny. The reason is that the unit root nature of the shocks diminishes the already small intertemporal substitution effects present in real business cycle models with mean-reverting shocks.³ Indeed, based on unfiltered series in both the data and the model along with the CEV estimates for σ_z^2 , we find that

$$(6) \quad \frac{\text{the variance of hours in the model with only technology shocks}}{\text{the variance of hours in the U.S. data}} = 1.8\%$$

(where for the hours series we use the same Prescott and Ueberfeldt series as in CKM). Thus, for the CEV model to reproduce the observed volatility of hours, the non-technology shocks alone must account for over 98% of the volatility in hours. In this sense, the data clearly suggest that non-technology shocks must be very large relative to technology shocks.

How large? To answer that question, in the top panel of Figure 1, I plot the variance of hours in the model and in the data against the variance of the non-technology shocks, holding fixed σ_z^2 and ρ_l at CEV's values. Clearly, under these conditions, as σ_l^2 is increased, the variance of hours in the model rises. This figure shows that at CEV's estimate for σ_l^2 (.0056²), hours are only about a third as volatile in their model as in the data. The figure also shows that for the model to account for the observed variability in hours, σ_l^2 must be substantially larger (about .0098²).⁴

The bottom panel of Figure 1 shows that when the parameter under dispute, σ_l^2 , is chosen to reproduce CEV's estimate, the bias is modest but the confidence bands are large. When this parameter is chosen to reproduce the observed volatility of hours, the LSVAR does miserably: the bias is large and the confidence bands are so enormous that the technique is unable to distinguish among most classes of models of interest.

I should be clear that we do not disagree that there exist some statistics, including some maximum likelihood statistics, that would lead to the conclusion that non-technology shocks are small. CKM find that the maximum likelihood estimates are sensitive to the variables included in the observer equation, especially to investment. Under some specifications, the variance of non-technology shocks is large while in others it is small. 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 that it matters what features of the data the researcher wants to mimic. In such a circumstance, we think it makes sense to use a limited-information technique in which we can choose the moments on which we want the model to do well.

In that vein, in designing a laboratory to test whether the SVAR methodology works, we asked, What would be some desirable features of the data for the model to reproduce? We came up with three answers, all of which contradict the condition necessary for SVARs to work well in practice; that is, all three suggest that non-technology shocks must be large.

One of our answers, which motivates the exercise just conducted, is that if the whole point of the procedure is to decompose the movements in hours, then the model should generate volatility in hours similar to that in the data. As CKM demonstrate, in the context of the CEV model, to do that the model needs large non-technology shocks.

A second answer is that the laboratory model should reproduce the key statistic that both started off the whole debate and is the main result in the long-run restriction SVAR literature: Galí's (1999) initial drop in hours after a positive technology shock. CKM ask, holding fixed the estimates of the variance of technology shocks and the persistence of non-technology shocks, What must be the variance of the non-technology shocks in order to reproduce Galí's impact coefficient on hours? We find that the variance of non-technology shocks must be large, large enough so that the SVARs do miserably in terms of bias and size of confidence bands.

(Note here that under the Sims-Cogley-Nason approach, whether or not Galí's DSVAR is misspecified is irrelevant. Galí's statistic is just a moment of the data that has happened to receive a lot of attention, with possibly no more interpretation than those in the standard Kydland and Prescott (1982) table of moments.)

A third answer to the question of reproducible features is that if the SVAR procedure works well, then the variance of the shocks should be consistent with the variance decompositions in the SVAR literature itself. Much of this literature, including Christiano, Eichenbaum, and Vigfusson (2003), attributes only a small fraction of the fluctuations to technology shocks. As CKM show, with the parameters set to generate any of these statistics, the SVAR responses are badly biased and have enormous confidence bands.

In sum, contrary to the argument of CEV, the U.S. data do not definitively show that technology shocks account for virtually all of the movements in output. Most of the literature agrees with us, including much of the previous work of CEV, both alone and in concert.

2.2 Does the Mixed OLS–Newey–West Estimator Uniformly Improve on OLS?

Perhaps the most interesting part of CEV's work is their proposed estimator of impulse responses with long-run restrictions. They argue that this estimator, which splices together the OLS estimator and a Newey and West (1987) estimator, "virtually eliminates the bias" (CEV 2006, p. 3) associated with the standard OLS estimator and thus makes the errors of their approach tiny. In this sense, CEV argue that it does not matter whether technology shocks account for almost all of the fluctuations in output because their new estimator takes care of the bias problem.

We disagree. The results of Mertens (2006) show that actually the new estimator does not even uniformly improve on the standard OLS estimator. Unfortunately, the new

estimator is thus not a comprehensive solution for the problems with long-run restrictions.

To understand these issues, use the notation of Mertens (2006) to write the standard OLS estimator of the impact coefficient matrix A_0 as

$$A_0^{OLS} = \left(I - \sum_i B_i^{OLS} \right) Chol(S_X(0)^{OLS}),$$

where B_i^{OLS} denotes the regression coefficient matrices from the VAR and $Chol(S_X(0)^{OLS})$ denotes the Cholesky decomposition of the OLS estimate of the spectral density matrix $S_X(0)^{OLS}$ of the variables in the VAR at frequency zero. Here

$$S_X(0)^{OLS} = \left(I - \sum_i B_i^{OLS} \right)^{-1} \Omega^{OLS} \left(I - \sum_i B_i^{OLS} \right)^{\prime-1},$$

where Ω^{OLS} is the OLS estimate of the covariance matrix of residuals from the VAR.

CEV propose replacing $S_X(0)^{OLS}$ with a spectral density estimator along the lines of Newey and West (1987), with a Bartlett weighting scheme given by

$$S_X(0)^{NW} = \sum_{k=-b}^b \left(1 - \frac{|k|}{b+1} \right) E_T X_t X_{t-k}',$$

where X_t is the data, T is the sample length, E_T is the sample moments operator, and b is a truncation parameter.⁵

Figure 2, taken from Mertens (2006), displays the impact errors resulting from the use of the OLS estimator and the mixed OLS–Newey–West estimator, with four lags in the VAR, $b = 150$, $T = 180$, various values of σ_l^2/σ_z^2 , and the rest of the parameters set as in CEV. The figure shows that when non-technology shocks are small, the CEV estimator has a larger bias than does the OLS estimator. As Mertens shows, if non-technology shocks are large enough, the positions eventually reverse. Clearly, the mixed OLS–Newey–West estimator is not uniformly better than the OLS estimator. (For more details, see Mertens 2006.)

2.3 Are DSVARs Misspecified?

It is somewhat of a puzzle to me why, in their broad assessment of SVARs, CEV focus on the LSVAR literature, which does not have economic results and has garnered neither much attention nor many publications, and ignore the DSVAR literature, which both does and has. (See CKM’s discussion of Fernald 2005, Gambetti 2006, and the LSVAR literature for details supporting this assertion.)

Both CKM and CEV argue that there are problems with the DSVAR literature, but we disagree as to what they are. CKM argue that the only mistake in the DSVAR literature is that it uses the common approach rather than the Sims–Cogley–Nason approach; that is, this literature compares empirical SVARs to inappropriate objects in the model. In this

comparison the lag-truncation bias is severe enough that it flips the sign of the estimated impulse response. CEV argue that the DSVAR literature makes a different mistake. In particular, CEV argue that the procedure of differencing hours has “an avoidable specification error” (CEV, p. 26). They seem to conclude that, even with an infinite amount of data, the DSVAR impulse responses will not coincide with the model’s impulse responses. We disagree: CKM address the issue of misspecification directly and argue that the DSVAR procedure has no specification error of importance.

CKM argue this result in two steps. The first step in our argument is that with a QDSVAR, with α close to 1, say, .99, Galí (1999) would have obtained impulse responses virtually indistinguishable from the ones he actually obtains in his DSVAR in which he sets α equal to 1. In this sense, for all practical purposes, we can think of Galí as having run a QDSVAR. The second step is that, for any $\alpha < 1$ and a long enough data set, the QDSVAR will get exactly the right answer. That is, with the lag length chosen to be suitably increasing with sample size, the sample impulse responses in the QDSVAR procedure will converge in the relevant sense to the model’s impulse response, that is, $\bar{A}_i(p = \infty, T = \infty) = D_i(\theta)$. In this precise sense, contrary to what CEV claim, this procedure has no specification error of importance.

Marcet (2005) shows something subtler. He shows that with the DSVAR procedure in which α equals 1, the sample impulse responses from a procedure in which the lag length increases appropriately with sample size converge in the relevant sense to the model’s impulse response. Marcet notes that his Proposition 1 seems to directly contradict the, at least implicit, claims of Christiano, Eichenbaum, and Vigfusson (2003).⁶

So, with large samples, researchers have no a priori reason to prefer the LSVAR procedure to the QDSVAR procedure, and with α close to 1 in samples of length typical to that in postwar data, the QDSVAR is indistinguishable from the DSVAR. Beyond that, small-sample issues do lead one specification to be preferred. Quasi-differencing lessens the amount of Hurwicz-type small-sample bias in estimating the parameters of a highly correlated series like per capita hours. Thus, at least a priori, the QDSVAR seems to be preferable to the LSVAR.

Nevertheless, the QDSVAR turns out to actually do worse than the LSVAR. When CKM decompose the mean impulse response error into small-sample bias and lag-truncation bias, we find that even though the QDSVAR has smaller Hurwicz-type bias, it has a much larger lag-truncation bias for reasonable parameters; the QDSVAR does worse. That is a quantitative result, however. We are not sure that it holds in a large class of models with a wide variety of parameters.

2.4 Does Adding More Variables to the SVARs Help?

CEV argue that, even though for a wide variety of circumstances, SVARs with long-run restrictions are uninformative, they can be informative in special cases—for example, when more variables are added to an LSVAR. Contrary to the impression we get from CEV, there is no disagreement on this point. Indeed, part of the point of CKM is to prove analytically exactly what the special cases are.

A commonly cited example of an economy in which SVARs with long-run restrictions work well is Fisher’s (2006) model. (See Fernandez-Villaverde, Rubio-Ramirez, and Sargent 2005.) CKM show that Fisher’s model is a special case of our Proposition 4 of when LSVARs can be informative. In this sense, we obviously agree with CEV about the validity of our proposition. We do not think, however, that an approach that works only in special cases has much to offer researchers seeking a reliable, generally applicable tool.

3. . . . And With Short-Run Restrictions

The use of the common approach on SVARs with long-run restrictions thus has little to recommend it. What about using it on SVARs with short-run restrictions? CEV claim that with this type of SVAR, their approach is a state-of-the art technique that is useful for guiding theory. They focus on short-run restrictions that are satisfied in models which satisfy certain timing assumptions, often referred to as *recursive* assumptions. CEV claim to show that when a model satisfies such an assumption, SVARs with short-run restrictions perform remarkably well in small samples. And CEV imply that, because of this finding, this technique can be used broadly to distinguish promising models from the rest.

Since the CKM work has nothing to do with short-run restrictions, I have not studied the details of CEV’s claims about how well the short-run restrictions work in practice with small samples and therefore have nothing to disagree with on these small-sample claims. Nevertheless, I do disagree with CEV’s main message with respect to short-run restrictions in this area. As other researchers do, CEV ignore some important critiques which seem to be widely thought of as devastating for much of the literature that uses SVARs with short-run restrictions. It is important to emphasize that these critiques are of a theoretical nature, not about some problems with small samples. These critiques imply that, regardless of how well the short-run restrictions work with small samples, they are of little value in guiding the development of a broad class of monetary research. Hence, these critiques need to be addressed with a precise theoretical argument, not with some small-sample results.

The main critique of the SVAR literature with short-run restrictions is the *Lucas and Stokey critique* of Lucas and Stokey (1987). The point of this critique is that the particular class of short-run identifying assumptions made by CEV and related work in the short-run

SVAR literature do not apply to a broad class of models and hence are of little use in guiding the development of a broad class of research.

The upshot of this critique is that some of the prominent work in the short-run SVAR literature has drastically overreached the conclusions of their studies. The short-run identifying assumptions in their work apply to only a tiny subset of monetary models, but the SVAR results have been used to rule out models not in that tiny subset. This mismatch between assumptions and models is a serious problem for this work.

A simple way for researchers in the short-run literature using the common approach to inoculate themselves from the Lucas and Stokey critique is to include in an appendix a list of the papers in the existing literature that satisfy their proposed identifying assumptions. Unfortunately, for most of the identifying schemes that I have seen, that list would be extremely short and would exclude most of the famous monetary models that constitute the core of theoretical monetary economics. If researchers are able to invent new identifying schemes for which this list is both broad and long, then this literature would have a much greater impact on guiding the development of monetary theory than it currently does. Doing so would constitute progress.

To understand my claim that the current literature is subject to the Lucas and Stokey critique, let us begin with the recursiveness assumption itself. As Christiano, Eichenbaum, and Evans (1998, p. 68) explain, “The economic content of the recursiveness assumption is that the time t variables in the Fed’s information set do not respond to the time t realizations of the monetary policy shock.” To see how this assumption might be satisfied in a model, note that if the monetary authority at time t sets its policy as a function of time t variables, including output, consumption, and investment, as it does in Christiano, Eichenbaum, and Evans (2005), then the model must have peculiar timing assumptions in which, in a quarterly model, after a monetary shock is realized, private agents cannot adjust their output, consumption, and investment decisions during the remainder of the quarter. Whether or not one agrees that this timing assumption is peculiar, it is irrefutable that this timing assumption is not satisfied in the primary models in the monetary literature. I illustrate this point in Figure 3, which depicts the typical classes of models studied in monetary economics. (Technically, for all the models in the large rectangle, the impulse responses from the SVAR procedure do not converge in the relevant sense to the impulse responses in the model, so that $\bar{A}_i(p = \infty, T = \infty) \neq D_i(\theta)$.)

As an illustration of the claim that some of the short-run literature overreaches, consider the exposition by Christiano and Eichenbaum (1999) of the research agenda of the monetary SVAR literature. This exposition draws on the well-cited comprehensive survey by Christiano, Eichenbaum, and Evans (1998) of the short-run SVAR literature, which is the

clearest statement of the research agenda of the monetary SVAR literature that I could find.

Christiano and Eichenbaum (1999) start with a summary of the so-called facts and a brief note that some identifying assumptions have been used to establish them:

In a series of papers, we have argued that the key consequences of a contractionary monetary policy shock are as follows: (i) interest rates, unemployment and inventories rise; (ii) real wages fall, though by a small amount; (iii) the price level falls by a small amount, after a substantial delay; (iv) there is a persistent decline in profits and the growth rate of various monetary aggregates; (v) there is a hump-shaped decline in consumption and output; and (vi) the US exchange rate appreciates and there is an increase in the differential between US and foreign interest rates. See CEE [Christiano, Eichenbaum, and Evans] (1998) for a discussion of the literature and the role of identifying assumptions that lie at the core of these claims.

Christiano and Eichenbaum (1999) then go on to reject some models that, they say, are not consistent with those facts. In particular, based on their SVAR-established facts, they reject both Lucas' (1972) island model and Lucas' (1990) liquidity model. These claims are clearly overreaching. Since Lucas' two models in particular do not satisfy the peculiar timing assumptions needed to justify the recursive identifying assumption in the SVAR, I do not see how it is logically coherent to reject them based on the SVAR-established facts.

A potential objection to the Lucas and Stokey critique is that the SVAR literature is not overreaching because some of the models that violate the recursiveness assumption satisfy some other identifying assumption that researchers have made, and for these other assumptions, SVAR researchers have found similar qualitative patterns. The second main critique of the short-run literature, the *Uhlig critique* of Uhlig (2005), dismisses this objection. The Uhlig critique is that the atheoretical SVAR specification searches are circular: "the literature just gets out what has been stuck in, albeit more polished and with numbers attached" (Uhlig 2005, p. 383). Uhlig argues that the reason the other identifying assumptions find similar answers is that the answers are essentially built into the search algorithm. Uhlig suggests that the algorithm used to find some SVAR results is, perhaps unconsciously, to pick a pattern of qualitative results and then do an atheoretical search over patterns of zeros, lists of variables to include, and periods of time to study, so that the resulting SVAR impulse responses reproduce the desired qualitative results. If this description is accurate, then I am sympathetic with Uhlig's conclusion that not much is to be learned from this branch of the short-run SVAR literature.

Note, again, that neither of these critiques would apply if, when comparing models and data, researchers simply followed the Sims-Cogley-Nason approach. Under that approach, the issue of whether the identifying assumptions of an SVAR hold in a model doesn't come up. The impulse responses from the SVAR on the data simply define some sample statistics that are coherently compared to the analogous statistics from the model. That is, now letting $A_i^{US}(p = 4, T = 180)$ and $\bar{A}_i(p = 4, T = 180|\theta)$ denote the impulse responses obtained from an SVAR with short-run restrictions, using standard (simulated) method-of-moments logic, it makes perfect sense to compare these two even though $\bar{A}_i(p = \infty, T = \infty) \neq D_i(\theta)$.

In sum, if the SVAR literature with short-run restrictions followed the research agenda advocated by Sims (1989) and applied by Cogley and Nason (1995), then it would be on firm statistical and logical grounds.

4. Concluding Remarks

Let me be clear about what I am advocating in practice. For researchers willing to make a quantitative comparison between a model and the data, all I am advocating basically is changing several lines of computer code—replacing the theoretical impulse responses, $D_i(\theta)$, with the more relevant empirical responses derived from applying the SVAR procedure to the model, $\bar{A}_i(p = 4, T = 180|\theta)$, in the relevant spots where the comparison between model and data is being made. For researchers who just want to run SVARs in the data and chat about what it means for a model, all I am advocating is a change in claims. Replace the claim about having robustly discovered what happens after a particular type of shock with a more precise claim about having documented what type of impulse responses should arise in a model when an SVAR with 4 lags and 180 observations is run on the data from it. Changing these several lines of code or text will vastly increase the intellectual impact of the approach.

It is puzzling to me that CEV and CKM agree on most of the relevant facts; yet we somehow disagree on their primary implication.

We agree on these two facts about the common approach:

- The common approach sometimes makes large errors relative to the Sims-Cogley-Nason approach. In particular, with long-run restrictions, SVARs do miserably unless technology shocks account for virtually all of the fluctuations in output.
- The common approach sometimes excludes most models of interest while the Sims-Cogley-Nason approach does not. For example, with short-run restrictions, the recursive identifying assumptions apply to only a tiny subset of the existing monetary models.

We disagree on one significant fact about interpreting the U.S. data:

CEV argue that the evidence definitively implies that technology shocks account for virtually all of the fluctuations in output. CKM argue that while one can find statistics supporting this view, 20 years of macroeconomic research and some simple statistics show that shocks other than technology shocks play a sizable role in generating fluctuations in output and other variables.

And we disagree on the overriding implication of these facts:

- CEV argue that the common approach is a state-of-the-art technique that can be saved with a mechanical fix and analyst restraint:
 - For the long-run restriction branch of the SVAR literature, CEV argue that a mixed OLS–Newey–West estimator essentially eliminates the errors of the common approach.
 - For the short-run restriction branch, CEV think that, in order to avoid the overreaching of some of the recent work in the area, researchers in this literature should be much more careful in delineating exactly to which work they claim their analyses apply. (This view is implicit in their conference discussions of early drafts of CKM’s work.)
- CKM argue, to the contrary, that the common approach is not a state-of-the-art technique and that it should be abandoned in favor of the approach used by Sims, Cogley, and Nason. The Sims–Cogley–Nason approach has firm statistical and theoretical foundations and thus avoids the type of statistical and logical errors present in the SVAR literature that uses the common approach.

My bottom line is simple: Let’s stop worrying about how large are the errors in the SVAR papers written in the past and start moving forward with the more promising Sims–Cogley–Nason approach, which hopefully will become the dominant approach in the future.

Notes

¹The idea of the Sims-Cogley-Nason approach is to compare the exact small-sample distribution of the estimator from the model (with short lags) to the small-sample estimate (with short lags) from the data rather than the common approach, which makes no attempt to deal with any of the issues that arise with a small sample. At a logical level, as long as the small-sample distribution is approximated well, either by hand, which is exceedingly difficult, or by a computer, which is easy, it seems that it must dominate the common approach.

²Note that, at least for the environment considered by CKM, since the Hurwicz-type small-sample bias is small, the comparison of

$$A_i^{US}(p = 4, T = 180) \text{ to } \bar{A}_i(p = 4, T = \infty|\theta)$$

would eliminate most of the error in the common approach and would allow the researcher to use standard asymptotic formulas. We view this comparison as a rough-and-ready approximation to the one in the Sims-Cogley-Nason approach.

³This reasoning helps explain why the bulk of the real business cycle literature has not adopted the unit root specification. In this sense, technically, the SVAR results really have little to say about this literature. But that point has already been forcefully made by McGrattan (2005).

⁴Note that, as other parameters in the model shift, so does the size of σ_l^2 needed to produce a certain volatility in hours. In this sense, whether a certain value of σ_l^2 is small or not should be judged by whether or not the model with this parameter can produce the observed volatility of hours in the data.

⁵The spectral density at frequency zero is defined as $S_X(0) = \sum_{k=-\infty}^{\infty} EX_t X'_{t-k}$. The estimator of Newey and West (1987) is a truncated version of this sum that replaces population moments with sample moments and weights these sample moments to ensure positive definiteness.

⁶Part of the disagreement in this regard may come from a failure to precisely distinguish between two types of noninvertibility problems. The type considered by Fernandez-Villaverde, Rubio-Ramirez, and Sargent (2005) are nontrivial and difficult to deal with without using a detailed economic theory. As Fernandez-Villaverde, Rubio-Ramirez, and Sargent (2005) discuss and Marcet (2005) and CKM show, however, the type of knife-edge invertibility issues that come from differencing a stationary series are much more trivial and are easy to deal with.

References

- Altig, D., L. J. Christiano, M. Eichenbaum, and J. Linde (2005). Firm-specific capital, nominal rigidities, and the business cycle. NBER Working Paper 11034.
- Barro, R. (1977). Unanticipated money growth and unemployment in the United States. *American Economic Review* 67(2): 101–115.
- Bernanke, B. S., M. Gertler, and S. Gilchrist (1999). The financial accelerator in a quantitative business cycle framework. In *Handbook of Macroeconomics*, Vol. 1C, J. Taylor and M. Woodford (eds). Amsterdam: North-Holland.
- Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2002). Can sticky price models generate volatile and persistent real exchange rates? *Review of Economic Studies* 69(3): 533–563.
- (2006). Are structural VARs with long-run restrictions useful in developing business cycle theories? Research Department Staff Report 364, Federal Reserve Bank of Minneapolis.
- Christiano, L. J., and M. Eichenbaum (1999). The research agenda: Larry Christiano and Martin Eichenbaum write about their current research program on the monetary transmission mechanism. *EconomicDynamics Newsletter* 1(1). Society for Economic Dynamics, <http://www.economicdynamics.org/>.
- Christiano, L. J., M. Eichenbaum, and C. L. Evans (1998). Monetary policy shocks: What have we learned and to what end? NBER Working Paper 6400.
- (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy* 113(1): 1–45.
- Christiano, L. J., M. Eichenbaum, and R. Vigfusson (2003). What happens after a technology shock? NBER Working Paper 9819.
- (2006). Assessing structural VARs. Northwestern University. Mimeo.
- Cogley, T., and J. M. Nason (1995). Output dynamics in real-business-cycle models. *American Economic Review* 85(3): 492–511.
- Cooley, T. F. and G. D. Hansen (1989). The inflation tax in a real business cycle model. *American Economic Review* 79(4): 733–748.
- Fernald, J. G. (2005). Trend breaks, long-run restrictions, and the contractionary effects of technology improvements. Working Paper 2005-21, Federal Reserve Bank of San Francisco.
- Fernandez-Villaverde, J., J. F. Rubio-Ramirez, and T. J. Sargent (2005). A, B, C’s (and D’s) for understanding VARs. NBER Technical Working Paper 0308.

Fisher, J. D. M. (2006). The dynamic effects of neutral and investment-specific technology shocks. *Journal of Political Economy* 114(3): 413–451.

Francis, N., and V. A. Ramey (2005). Is the technology-driven real business cycle hypothesis dead? Shocks and aggregate fluctuations revisited. *Journal of Monetary Economics* 52(8): 1379–1399.

Fuerst, T. S. (1992). Liquidity, loanable funds, and real activity. *Journal of Monetary Economics* 29(1): 3–24.

Galí, J. (1999). Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations? *American Economic Review* 89(1): 249–271.

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 *NBER Macroeconomics Annual 2004*, M. Gertler and K. Rogoff (eds.). Cambridge, MA: MIT Press, pp. 225–288.

Gambetti, L. (2006). Technology shocks and the response of hours worked: Time-varying dynamics matter. Universitat Pompeu Fabra. Mimeo.

Kydland, F. E., and E. C. Prescott (1982). Time to build and aggregate fluctuations. *Econometrica* 50(6): 1345–1370.

Lagos, R., and R. Wright (2005). A unified framework for monetary theory and policy analysis. *Journal of Political Economy* 113(3): 463–484.

Lucas, R. E., Jr. (1972). Expectations and the neutrality of money. *Journal of Economic Theory* 4(2): 103–124.

——— (1990). Liquidity and interest rates. *Journal of Economic Theory* 50(2): 237–264.

Lucas, R. E., Jr., and N. L. Stokey (1987). Money and interest in a cash-in-advance economy. *Econometrica* 55(3): 491–513.

Marcet, A. (2005). Overdifferencing VAR’s is OK. Universitat Pompeu Fabra. Mimeo.

McGrattan, E. R. (2005). Comment on Galí and Rabanal’s “Technology Shocks and Aggregate Fluctuations: How well does the RBC model fit postwar U.S. data?” *NBER Macroeconomics Annual 2004*, M. Gertler and K. Rogoff (eds.). Cambridge, MA: MIT Press, pp. 289–308.

Mertens, E. (2006). Shocks to the long run: A note on mixing VAR methods with non-parametric methods. Study Center Gerzensee. Mimeo.

Newey, W. K., and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3): 703–708.

Rotemberg, J. J., and M. Woodford (1997). An optimization-based econometric framework for the evaluation of monetary policy. In *NBER Macroeconomics Annual 1997*, B. S. Bernanke and J. J. Rotemberg (eds.). Cambridge, MA: MIT Press, pp. 297–346.

Shi, Shouyong (1997). A divisible search model of fiat money. *Econometrica* 67(1): 75–102.

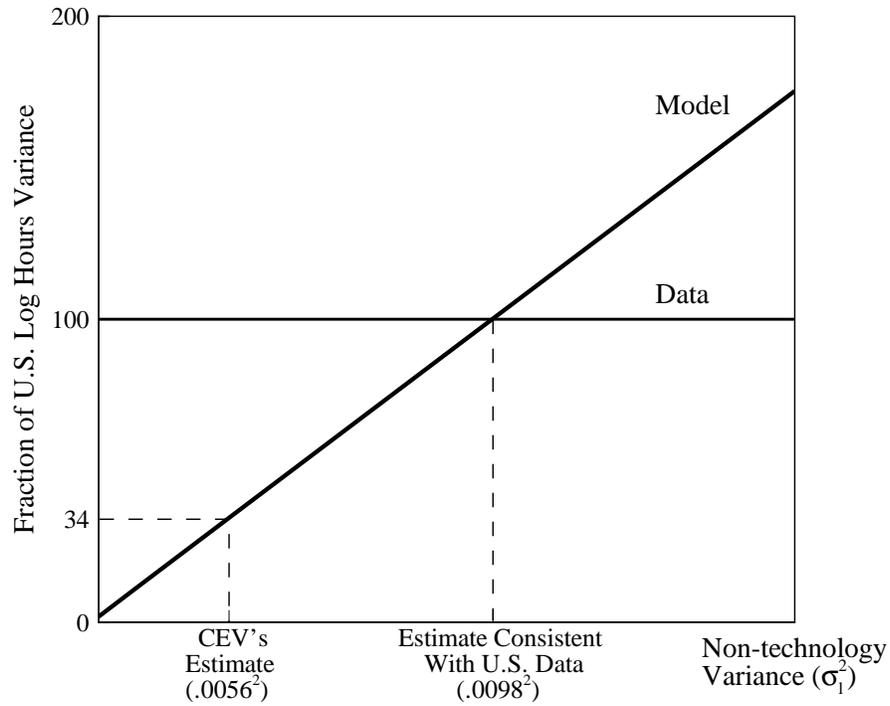
Sims, C. (1989). Models and their uses. *American Journal of Agricultural Economics* 71(2): 489–494.

Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics* 52(2): 381–419.

Yun, T. (1996). Nominal price rigidity, money supply endogeneity, and business cycles. *Journal of Monetary Economics* 37(2): 345–370.

FIGURE 1. THE DATA SAY NON-TECHNOLOGY SHOCKS MUST BE LARGE ...

Fraction of U.S. Hours Variance Generated by Model
vs. Variance of Non-technology Shocks



... AND WHEN THEY ARE, SO ARE THE BIAS AND CONFIDENCE BANDS

Impact Error and 95% Bootstrapped Confidence Bands

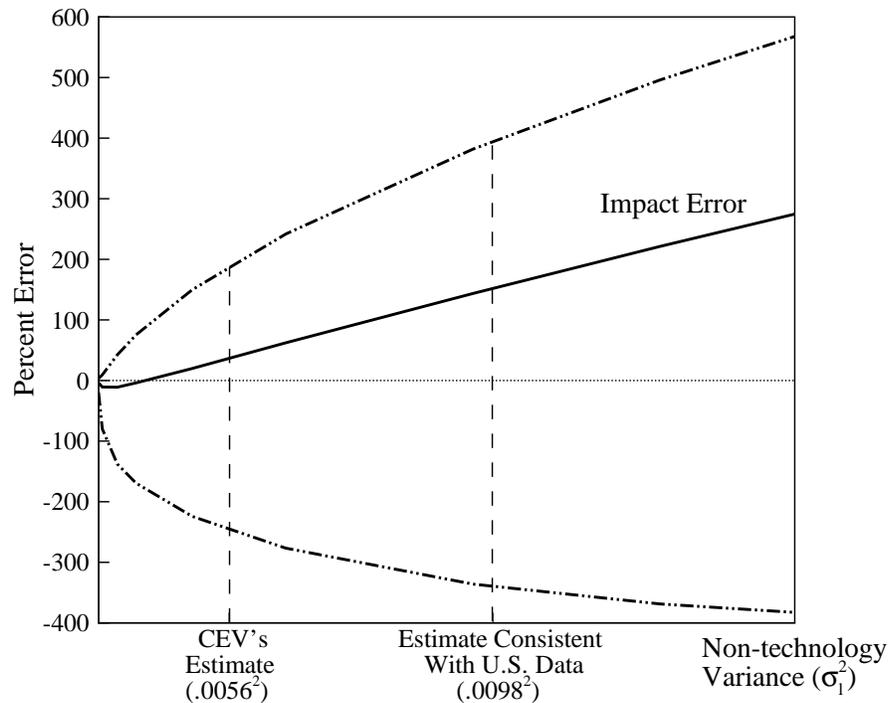
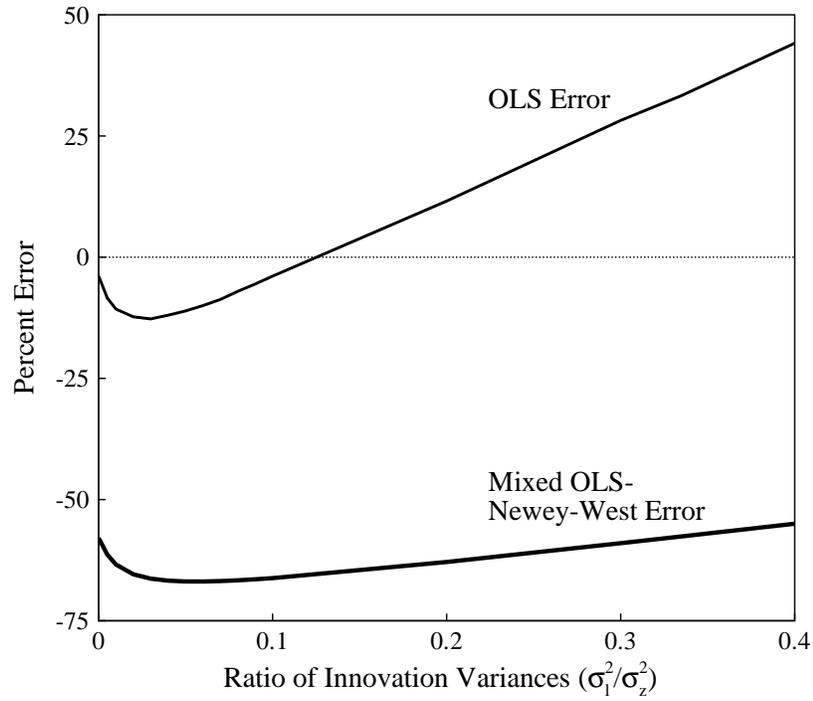


FIGURE 2. THE MIXED OLS-NW ESTIMATOR IS NOT UNIFORMLY BETTER

Impact Errors Using the OLS and Mixed OLS-NW Estimators



Source: Mertens (2006)

FIGURE 3. CEV'S RECURSIVENESS ASSUMPTION DOES NOT APPLY TO MOST MONETARY MODELS

Representative Classes of Existing Monetary Models, Grouped by Whether They Violate or Satisfy CEV's Recursiveness Assumption

Monetary Models

Models that Violate the Recursiveness Assumption

Misperceptions Models

Lucas 1972
Barro 1977

Cash-Credit Models

Lucas & Stokey 1987
Cooley & Hansen 1989

Liquidity Models

Lucas 1990
Fuerst 1992

Sticky-Price/Wage Models

Yun 1996
Chari, Kehoe, & McGrattan 2002

Search Models

Shi 1997
Lagos & Wright 2005

Financial Frictions Models

Bernanke, Gertler, & Gilchrist 1999

Models that Don't

Rotemberg & Woodford 1997
Altig et al. 2005