Are Spectral Estimators Useful for Implementing Long-Run Restrictions in SVARs?

Elmar Mertens

Working Paper 08.01

Are Spectral Estimators Useful for Implementing Long-Run Restrictions in SVARs?*†

Elmar Mertens‡
Study Center Gerzensee and
University of Lausanne

First Draft: 03 / 07
This Draft: 03 / 08

*This is a draft of my second thesis chapter. An earlier draft of this paper has been written while visiting the Federal Reserve Bank of Minneapolis. I would like to thank the Bank and its staff for their kind hospitality. I would like to thank V.V. Chari, Patrick Kehoe and Ellen McGrattan for discussions about earlier versions of this paper and sharing their computer codes. In addition I am grateful for discussions with Lawrence Christiano, Bill Dupor, Peter Kugler, Mark Watson as well as seminar participants at the University of Basel and the European Economics Association meeting 2007 in Budapest. All remaining errors are of course mine.

†This research has been carried out within the project IP A2 "Macro Risk, systemic risks and international finance" of the National Center of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK). NCCR FINRISK is a research program supported by the Swiss National Science Foundation.

‡For correspondence: Elmar Mertens, Study Center Gerzensee, CH-3115 Gerzensee, Switzerland. email `elmar.mertens@szgerzensee.ch`. Tel.: +41 31 780 3114 Fax: +41 31 780 3100. The paper is available at `http://www.elmarmertens.ch`.

Are Spectral Estimators Useful for Implementing Long-Run Restrictions in SVARs?

Abstract

No, not really. Responding to lingering concerns about the reliability of SVARs, Christiano et al (NBER Macro Annual, 2006, "CEV") propose to combine OLS estimates of a VAR with a spectral estimate of long-run variance. In principle, this could help alleviate specification problems of SVARs in identifying long-run shocks.

But in practice, spectral estimators suffer from small sample biases similar to those from VARs. Moreover, the spectral estimates contain information about serial correlation in VAR residuals and the VAR dynamics must be adjusted accordingly. Otherwise, a naive application of the CEV procedure would misrepresent the data's variance.

JEL Classification: C32, E17

Keywords: Structural VAR, Long-Run Identification, Non-parametric Estimation, Factorization of Spectral Density

1 Introduction

VARs have been criticized for failures in estimating the responses to long-run shocks. A crucial element for long run identification is the spectral density at zero-frequency, also known as "long-run variance". OLS estimates of VAR coefficients are concerned with minimizing forecast error variance, not estimating the long run variance. This has recently motivated Christiano, Eichenbaum, and Vigfusson (2006a, 2006b), henceforth "CEV", to propose a new way of estimating structural VARs using a combination of OLS and a nonparametric estimator. They argue that their estimator virtually eliminates the bias associated with the standard OLS estimator. This paper investigates their procedure in more detail.

Conceptually, there are some pitfalls in combining VAR coefficients with non-parametric estimates of the spectral density. The spectral estimates (correctly) allow for non-*iid* residuals in the finite-order VAR when the underlying model is of infinite order. In what may be called "mixing and matching", the CEV approach plugs this estimate into the standard VAR formula alongside with coefficients from the finite-order VAR. This way, the extra information on omitted lags is used to compute the long-run responses of variables to shocks – but not when mapping these back into impact responses. That would however be necessary to retain the consistency of the data's SVAR representation. Otherwise, the total variance of the data is misrepresented. This inconsistency is shown to be quantitatively relevant. Moreover, the inconsistency makes it impossible to obtain meaningful estimates of the shocks themselves. When the relationship between forecast errors and structural shocks is inverted with the CEV coefficients, one obtains a time series which is identical to the OLS estimates up to a scale factor. These issues show up directly when decomposing the variance of historical data – exercises which the original CEV analysis was not concerned with. Their focus was on impulse responses. But the data's variance is just a convolution of impulse response. If the former is misrepresented, the latter needs rethinking, too. All in all, this is of interest to any researcher wanting to adopt the CEV strategy.

The CEV framework is amended here by recognizing that the non-parametric estimate contains information about omitted lags in the VAR. This misspecification has been stressed by Chari, Kehoe, and McGrattan (2005, henceforth "CKM"), Erceg, Guerrieri, and Gust (2005), Ravenna (2007) and Cooley and Dywer (1998). The adjusted procedure retains the OLS estimates and fills up the omitted lags with a spectral factorization of the spectral density's non-parametric estimate. By construction, this adjusted SVAR – in fact an SVARMA – matches the sample variance of the data just as OLS does.

Empirically, the various procedures are applied to data simulated from the same model economy as in CKM and CEV – but over a wider set of calibrations as in CEV. Four key results emerge:

1. Non-parametric estimates of long-run variance are not much better than OLS. Using the CEV specification for the spectral estimators bandwidth[1], they are even considerably worse.

2. The variance misrepresentation of CEV is substantial – particularly when using their (2006b) estimator.

3. Taking their procedure at face value, it is no panacea at all. Depending on the true process, OLS can have both smaller bias and smaller sampling uncertainty.

4. The spectral factorization proposed here performs almost uniformly better than OLS in terms of bias, however the gains are fairly small and sampling uncertainty is large. But overall the bias is still large and sampling uncertainty is considerable. This is no surprise since it inherits the considerable small sample issues of the spectral density estimates. A major issue for all procedures are the effects of small sample bias, not only misspecification bias.

Conceptually, the amended CEV procedure yields a consistent representation of the data and gets around the truncation issue of a finite-order VAR. But what matters for the empirical performance is that spectral estimates are as much subject to small sample biases as OLS estimates of a VAR. Estimation of the spectral density at zero-frequency is particularly prone to small sample biases, since these are strongest when the data is persistent and the zero-frequency spectrum looks at the most persistent part of the data.[2]

The remainder of this paper is structured as follows: Section 2 compares the standard OLS procedure for long-run identification against the spectral method by CEV. Section 3 points to some conceptual shortcomings in this setup. As a remedy, Section 4 proposes a spectral factorization procedure. Section 5 describes the model economy used to simulate the performance of the various estimation routines. Section 6 presents the Monte Carlo results and Section 7 concludes the paper. Additional results are presented in a separate appendix which is available at the author's website.[3]

2 Long-Run Identification in VARs

An economic model is supposed to specify a VAR representation for a stationary vector of observable variables[4] X_t:

$$X_t = B(L)X_{t-1} + e_t \tag{1}$$

where $B(L)$ is a polynomial in the lag-operator L

$$B(L) = \sum_{k=1}^{\infty} B_k L^{k-1}$$

whose roots lie all outside the unit circle and the innovations are *iid*, $e_t \sim iid(0, \Omega_e)$.

Note that the model prescribes an *infinite* order VAR. When $B_k = 0$ for $k > p$ this is a finite order VAR. But as noted by Cooley and Dywer (1998), many interesting models have only infinite order VAR representations. In the remainder the true VAR representation is always assumed to be of infinite order.

For the identification of structural shocks, there has to be an invertible one-to-one mapping from innovations e_t to the structural shocks ε_t driving the underlying model – such as technology, monetary policy errors, exogenous government spending etc.:

$$e_t = A_0 \, \varepsilon_t \tag{2}$$

where A_0 is square and $|A_0| \neq 0$. Fernàndez-Villaverde et al. (2007) derive conditions when a (linear) dynamic model economy will have an invertible VAR representation (see also Appendix B). This paper considers only cases where these conditions are satisfied, though possibly only in an infinite order VAR representation. The same applies to the situations studied by CKM, CEV as well as Erceg, Guerrieri, and Gust (2005). Fernandez-Villaverde, Rubio-Ramirez, and Sargent (2005) give examples of interesting models where the conditions are satisfied and where not. By excluding the complications arising from non-invertibilities we want to focus on problems stemming solely from finite order approximations of the VAR.[5]

It will be handy to introduce the notation

$$C(L) \equiv (I - B(L)L)^{-1} = \sum_{k=0}^{\infty} C_k L^k \qquad \text{where} \quad C_0 = I \tag{3}$$

for the non-structural moving average (VMA) coefficients of $X_t = C(L)e_t$. The structural moving average representation for X_t is then

$$X_t = A(L)\varepsilon_t \quad \text{with} \quad A(L) = C(L)A_0 \tag{4}$$

Long-Run Restriction

In the spirit of CEV and CKM, we will only be concerned with identifying one of the structural shocks. For concreteness, let it be the first one, denoted ε_t^z, and call it "technology shock". Think of the first element of X_t as being a growth rate (a difference in logs), like the change in labor-productivity (Gali 1999) or output growth (Blanchard and Quah 1989). The identifying assumption is then that only the technology shock has a permanent effect on the level of the first element of X_t. This restricts the following matrix of long-run coefficients, $A(1) = \sum_{i=0}^{\infty} A_i$:

$$\begin{aligned} A(1) &= C(1)A_0 \\ &= \begin{bmatrix} \bar{a}_{11} & 0 & \dots & 0 \\ \# & \# & \dots & \# \end{bmatrix} \quad \text{and} \quad \bar{a}_{11} > 0 \end{aligned} \tag{5}$$

A key object for implementing this constraint is the spectral density of X_t. The spectral density at frequency ω is defined as

$$S_X(\omega) \equiv \sum_{k=-\infty}^{\infty} E(X_t X_{t-k}^T) e^{-i\omega k} = C(e^{-i\omega})\, \Omega\, C(e^{-i\omega})^T \tag{6}$$

where i is the imaginary unit and the transpose is complex conjugate. $A(1)$ factors the spectral density of X_t at frequency zero:

$$A(1)A(1)^T = C(1)\, \Omega\, C(1)^T = S_X(0) \tag{7}$$

One way to compute the first column of A_0 is by recovering $A(1)$ from the Cholesky decomposition of $S_X(0)$. (This is the unique lower triangular factorization of a positive definite matrix.[6]):

$$A(1) = \text{chol}\,\{S_X(0)\}$$

CEV show that the restriction in (5) uniquely pins down the first column of A_0 and the Cholesky factorization is one possible implementation. Its orthogonalization of the remaining columns of $A(1)$ is arbitrary.[7]

The long-run coefficients can then be mapped into the matrix of impact responses using the VAR dynamics encoded in the polynomial of lag coefficients $B(L)$:

$$A_0 = (I - B(1))\,A(1) \tag{8}$$

2.1 OLS: Implementation with Finite-Order VAR

Since the VAR innovations in (1) are assumed to be white noise, they satisfy the OLS normal equations $EX_{t-k}e_t^T = 0$ $(\forall\, k)$. And in principle, the coefficients B_k could be estimated from least squares projections of X_t on its infinite past. An empirical implementation however can only work with a finite lag length. Henceforth $B(L)^{\text{OLS}}$ denotes a lag polynomial of finite order $p < \infty$:

$$\begin{aligned} B(L)^{\text{OLS}} &\equiv \sum_{k=1}^{p} B_k^{\text{OLS}}\,L^k \\ \text{and} \quad v_t^{\text{OLS}} &\equiv X_t - B(L)^{\text{OLS}}\,X_{t-1} \\ \Omega_v^{\text{OLS}} &\equiv E\,[v_t^{\text{OLS}}\,(v_t^{\text{OLS}})^T] \end{aligned} \tag{9}$$

where the normal equations are imposed for all lags $k \leq p$

$$EX_{t-k}(v_t^{\text{OLS}})^T = 0 \tag{10}$$

The associated VMA is $C(L)^{\text{OLS}} \equiv (I - B(L)^{\text{OLS}} L)^{-1}$. Only stable VARs are considered, formally this requires all roots of $C(L)^{\text{OLS}}$ to be outside the unit-circle.

It is standard procedure to assume white noise residuals, v_t^{OLS}. Following Blanchard and Quah (1989), the long run restriction (5) is then implemented based on an estimate of the spectral density at frequency zero constructed from the OLS estimates. Impact coefficients are then computed by plugging these estimates into (8):

$$\begin{aligned} S_X(0)^{\text{OLS}} &= C(1)^{\text{OLS}} \; \Omega_v^{\text{OLS}} \; C(1)^{\text{OLS}\,T} \\ A_0^{\text{OLS}} &= \left(I - B(1)^{\text{OLS}}\right) \text{chol}\left\{S_X(0)^{\text{OLS}}\right\} \end{aligned}$$

This implementation has been criticized for instance by Cooley and Dywer (1998) and CKM on the grounds of interesting models having only infinite order VAR representations and finite order approximations being insufficient. The assumption that the v_t^{OLS} are serially uncorrelated is a good example of what Cooley and Dywer called an “auxiliary” (but not innocuous) assumption.

2.2 CEV: Combination with Spectral Estimate

CEV propose an alternative estimator for the matrix of impact coefficients. This new estimator uses a mixture of the OLS estimates of $B(1)$ and a nonparametric estimator for $S_X(0)$. The procedure is motivated by the following observation: The OLS projections construct $B(L)^{\text{OLS}}$ such as to minimize the forecast error variance Ω_v^{OLS}. As shown by

Sims (1972), this can be expressed in the frequency domain as[8]

$$\min_{B_1^{\text{OLS}},\ldots,B_p^{\text{OLS}}} \Omega_v^{\text{OLS}} = \Omega_v + \int_{-\pi}^{\pi} \left(B(e^{-i\omega}) - B(e^{-i\omega})^{\text{OLS}}\right) S_X(\omega) \left(B(e^{-i\omega}) - B(e^{-i\omega})^{\text{OLS}}\right)' d\omega \quad (11)$$

Written this way, it is evident that OLS coefficients are constructed in order to minimize the average distance between themselves and the true $B(e^{-i\omega})$, weighted by the spectral density of X_t, which may or may not be large at zero frequency: Based on this objective, $S_X(0)^{\text{OLS}}$ need not be the best possible estimate for the spectral density at frequency zero. OLS will try to set $B(1)^{\text{OLS}}$ close to $B(1)$ only if the data's spectrum is high at the zero frequency.

Instead, CEV construct $A(1)$ from a spectral estimator of $S_X(0)$. In Christiano, Eichenbaum, and Vigfusson (2006a), they consider two estimators, one based on Newey and West (1987) and the other on Andrews and Monahan (1992). Both are based on truncated sums of autocovariance matrices. To ensure positive definiteness, these are weighted by a Bartlett kernel. Where Newey-West uses the (sample) autocovariances of X_t, Andrews-Monahan uses first the VAR to prewhiten the data and then takes the *residual* autocovariances:

$$S_X(0)^{\text{AM}} = C(1)^{\text{OLS}} S_v^{NW}(0) \left(C(1)^{\text{OLS}}\right)^T \quad (12)$$

$$\text{where} \quad S_v(0)^{\text{NW}} = \sum_{k=-b}^{b} \left(1 - \frac{|k|}{b+1}\right) E\left[v_t^{\text{OLS}} (v_{t-k}^{\text{OLS}})^T\right] \quad (13)$$

where b is a truncation parameter, also known as "bandwidth" to be chosen by the researcher. The Newey-West estimator applies (13) to X_t directly. As elsewhere in this section, I have expressed the estimators above in terms of population moments, $E\left[v_t^{\text{OLS}} (v_{t-k}^{\text{OLS}})^T\right]$. In empirical applications, the population moments are replaced by sample moments[9].

In general, other weighting schemes than the Bartlett weights can be used, but as shown by Newey and West (1994), this is of secondary importance compared to the bandwidth choice. For a consistent estimator, b can grow with the sample size but at a smaller rate. Andrews (1991) and Newey and West (1994) propose data dependent schemes of optimal bandwidth selection whereas CEV use a fixed and fairly large value of $b = 150$ in a sample of 180 observations.[10] I will return to this issue in the lab simulations of Section 6, where both automatic selection and fixed bandwidth schemes are evaluated.

The prewhitening of Andrews-Monahan is theoretically appealing since it removes spikes from the spectral density of X_t which make spectral estimation difficult (Priestley 1981, Chapter 7). It is not meant to necessarily eliminate all of the data's serial dependence. Both Andrews and Monahan (1992) and Newey and West (1994) find the pre-whitening to fair better in Monte Carlo studies than the original Newey-West estimator. Christiano, Eichenbaum, and Vigfusson (2006a) find no clearly superior choice between the two and proceed to use only the Newey-West estimator in Christiano, Eichenbaum, and Vigfusson (2006b). For ease of exposition, I will focus my presentation on the Andrews-Monahan estimator. Amongst others, this is appealing since it nests the OLS estimator by setting $b = 0$. (Section 6 presents results for both.)

The new CEV estimator computes the long-run coefficients from the non-parametric density estimate

$$A(1)^{\mathrm{AM}} = \mathrm{chol}\left\{S_X(0)^{\mathrm{AM}}\right\} \tag{14}$$

and combines this with the OLS lag coefficients to obtain the impact coefficients

$$A_0^{\mathrm{CEV}} = \left(I - B(1)^{\mathrm{OLS}}\right) A(1)^{\mathrm{AM}} \tag{15}$$

The impulse responses of CEV are then

$$A(L)^{\text{CEV}} = C(L)^{\text{OLS}} A_0^{\text{CEV}} \tag{16}$$

Analogous formulas hold when using the Newey-West estimator.

3 Problems with the CEV Procedure

The CEV procedure is motivated by dissatisfaction with $B(1)^{\text{OLS}}$, which is needed to construct the long run responses $A(1)$. But when transforming long-run responses into impact coefficients, $B(1)^{\text{OLS}}$ is used again. This leads to some serious problems which are stated here in the form of three remarks. A fourth remark motivates my corrected procedure, presented at the end of this section.

Remark 1 (CEV Shocks are just a rescaling of OLS)**.** *Given v_t^{OLS} and A_0^{CEV} a researcher might want to re-construct the structural shocks based on (2) $\varepsilon_t^{CEV} = \left(A_0^{CEV}\right)^{-1} v_t^{OLS}$ and compare them against $\varepsilon_t^{OLS} = \left(A_0^{OLS}\right)^{-1} v_t^{OLS}$. She will be troubled noticing that the estimated technology shocks are perfectly correlated:*

$$(\varepsilon_t^z)^{CEV} = \frac{\bar{a}_{11}^{OLS}}{\bar{a}_{11}^{CEV}} \cdot (\varepsilon_t^z)^{OLS}$$

(Recall from (5) that $\bar{a}_{11}$ is the top element of $A(1)$.) This holds both for population and sample moments. Actually, it holds for any pair of matrices A_0^1 and A_0^2 constructed from (8) using $B(1)^{OLS}$ and a $A(1)$ satisfying the zero restrictions (5).

Proof. Both CEV and OLS use $B(1)^{\text{OLS}}$ in computing $A_0^{-1} = A(1)^{-1}(I - B(1))^{-1}$ and except for the top left element, the first row of $A(1)^{-1}$ is full of zeros. This follows from the long run restriction (5) which places the same zero restrictions on $A(1)^{-1}$ as it does on

$A(1)$ and applies both to $A(1)^{\text{CEV}}$ and $A(1)^{\text{OLS}}$. Finally, the top left element of $A(1)^{-1}$ equals $1/\bar{a}_{11}$.[11] $\square$

The point of the previous remark is that the top rows of $\left(A(1)^{\text{OLS}}\right)^{-1}$ and $\left(A(1)^{\text{CEV}}\right)^{-1}$ are identical up to a scaling. Since CEV were only concerned with impulses-responses and A_0, the problem does not show up in their analysis. The construction of estimated shocks is however often used by researchers, for instance in order to plot historical decompositions or when identifying several shocks (see for example Altig et al. (2004)). Of course, if something is wrong about A_0^{-1}, this applies also to A_0. Looking at A_0 the problem shows up more subtly.

Remark 2 (Mismatch with OLS Forecast Error Variance)**.** *The CEV procedure is motivated by a dissatisfaction with $S_X(0)^{OLS}$. A researcher adopting their strategy wants $S_X(0)^{AM} \neq S_X(0)^{OLS}$ and thus $S_v(0)^{NW} \neq \Omega^{OLS}$.*[12] *This immediately implies*

$$A_0^{CEV}(A_0^{CEV})^T \neq \Omega_v^{OLS} \tag{17}$$

Implicitly, CEV attribute any difference between spectra estimated from OLS and the non-parametric methods to the VAR's forecast error variance, and not to a misspecification of the dynamics. However, the accuracy of estimating Ω_v^{OLS} has never been doubted. In fact, getting a good estimate for forecast error variance is precisely the objective of OLS projections. This objective is doubted by CEV only when it comes to the zero-frequency spectral density. It is also noteworthy that with (17), their procedure also deviates from the previous literature where identification is defined as a search over the space of matrices A_0 satisfying $A_0A_0^T = \Omega_v^{\text{OLS}}$ (Faust 1998; Canova and de Nicolo 2003; Uhlig 2005).

Remark 3 (Total Variance not matched either)**.** *Given (17), the CEV SVAR cannot match* *the variance of X_t. Their impulse-responses (16) imply the following variance measure*

$$\begin{aligned}
\operatorname{Var} X_t^{\,CEV} &= \sum_{k=0}^{\infty} C_k^{OLS} A_0^{CEV} (A_0^{CEV})^T (C_k^{OLS})^T && (18)\\
&\neq \sum_{k=0}^{\infty} C_k^{OLS} \Omega_v^{OLS} (C_k^{OLS})^T && (19)\\
&= \operatorname{Var} X_t
\end{aligned}$$

Proof. CEV model the data as $X_t^{\text{CEV}} = C(L)^{\text{OLS}} A_0^{\text{CEV}} \varepsilon_t$. The second step follows directly from Remark 2 and the third step holds because of the normal equations (10) and the definition of the VAR(p) in (9), regardless of whether v_t^{OLS} is *iid* or not. (See Appendix B.) □

These remarks hold both for population moments as well as for sample moments[13]. They are unsettling and raise issues about the applicability of the CEV procedure for variance decompositions. To understand what is amiss in the CEV procedure, it is useful to recognize that the OLS residuals v_t^{OLS} are not *iid* and that this is embedded in the long-run coefficients of CEV.

Remark 4 (CEV are concerned about serially-correlated VAR residuals)**.** *The Andrews-Monahan estimator is constructed from autocovariances of the VAR residuals v_t^{OLS}. Rewriting (13) and considering also non-zero-frequencies we have*

$$\begin{aligned}
S_v(\omega)^{NW} &= \Sigma_v^{OLS} + \sum_{k=1}^{b} \kappa(k) \left(\Gamma_k e^{-i\omega k} + \Gamma_k^T e^{i\omega k} \right) && (20)\\
\textit{where} \qquad \Gamma_k &\equiv E\left[v_t^{OLS} (v_{t-k}^{OLS})^T \right] && (21)\\
\textit{and} \qquad \kappa(k) &= 1 - \frac{|k|}{b+1}
\end{aligned}$$

Obviously, $b > 0$ expresses a concern about serially correlated residuals. Implicitly, the

Newey-West estimator of $S_X(\omega)$ also embodies concerns about serially correlated VAR residuals since it implies the following spectrum for v_t^{OLS}, which is generally not constant across frequencies

$$(I - B(e^{-i\omega})^{OLS})S_X(\omega)^{NW}(I - B(e^{-i\omega})^{OLS})^T$$

(As before $S_X(\omega)^{NW}$ is (20) applied to the autocovariances of X_t.)

The CEV procedure is clearly concerned about misspecified dynamics of the VAR(p) when constructing $A(1)$ but not when mapping this back to the short run responses A_0. As argued in the next section, this is the source of the problems listed in Remarks 1, 2 and 3 above.

4 Correct Identification via Spectral Factorization

We need to reconsider the consequences of approximating the infinite order model (1) with a VAR(p). In particular, once we start modeling the serial correlation in v_t^{OLS}, it needs to be done consistently.

OLS projections are still well defined in the sense of satisfying the projection equations (10) for $k \leq p$, but the residuals v_t^{OLS} are not *iid*. In general, they follow a moving average representation:

$$\begin{aligned} v_t^{\text{OLS}} &= D(L)A_0\varepsilon_t \\ D(L) &= (I - B(L)^{\text{OLS}})C(L) \\ &= I + \sum_{k=1}^{\infty} D_k L^k \end{aligned} \tag{22}$$

with spectral density

$$S_v(\omega) = D(e^{-\omega})\Omega_e D(e^{-\omega})^T \tag{23}$$

The results of CKM and CEV on a truncation bias which is hard to detect based on VAR lag-length selection procedures can be read as finding

$$D_i \approx 0 \quad \text{but} \quad D(1) \neq I$$

(This will be confirmed also in the lab economy of Section 5, see Figure 1 there.) For our purposes, an important property of $D(L)$ is its invertibility:

Proposition 1 (Invertibility of D(L))**.** *When the underlying model has a fundamental VAR representation as in (1), and the OLS-VAR is stable, the moving average polynomial $D(L)$ defined in (22) has all its roots outside the unit-circle.*

Proof. The proof is straightforward since $(I - B(L))^{-1} = C(L) = (I - B(L)^{\text{OLS}})^{-1}D(L)$ has all roots outside the unit circle and the same has been assumed for the VMA of the VAR(p), $C(L)^{\text{OLS}} = (I - B(L)^{\text{OLS}} L)^{-1}$. □

Via (23), the spectral estimates of $S_v(\omega)$ contain information on the D_i coefficients. For the time being, I want to abstract from estimation issues such as bandwidth selection and weighting schemes and consider the case where an econometrician is given the population values of $B(L)^{\text{OLS}}$ and $S_v(\omega)$. It is then straightforward to recover $D(L)$ by performing a spectral factorization of $S_v(\omega)$. The "canonical spectral factorization" is a classic theorem in linear quadratic control and assures us of existence and uniqueness of an invertible[14] $D(L)$ and a positive definite Ω_e consistent with (23). Below I adapt its statement from Hannan (1970), see also Whittle (1996, Chapter 13) and Li (2005). For a reference in the context of economics see Hansen and Sargent (2007, 2005).

Theorem 1 (Spectral Factorization, (Hannan 1970)). *Given a spectral density*

$$S_v(\omega) \equiv \sum_{k=-q}^{q} \Gamma_k e^{-ik\omega} \quad \forall \quad \omega \in [-\pi, \pi]$$

which is non-singular at each frequency ($|S_v(\omega)| \neq 0 \ \forall \omega$) and where $\Gamma_k = (\Gamma_{-k})^T$ are autocovariance matrices as in (21), there is a factorization of $S_v(\omega)$ into

$$S_v(\omega) = D(e^{-ik\omega}) \, \Omega_e \, D(e^{-ik\omega})^T$$

This factorization is unique and Ω_e is positive definite. $D(z)$ is a q'th order polynomial

$$D(z) = I + \sum_{k=1}^{q} D_k z^k$$

which has all its roots on or outside the unit circle. (The transposes are complex conjugate).

The theorem factors a spectrum constructed from a finite number of autocovariances into a finite-order MA. As will be seen below, a finite q has of course to be chosen for an empirical application. But when applying the spectral factorization to the population objects of the true model (1), we need to consider that in general the true $D(L)$ is an MA(∞). However, since the processes for X_t and v_t^{OLS} are stationary, their autocovariances and MA-coefficients vanish for large lags[15]. A spectral factorization with an arbitrarily large but finite q can arbitrarily well approximate the true spectrum and true $D(L)$. (This is analogous to Sims (1972).) Alternatively we can think of the true $D(L)$ being the limit of applying Theorem 1 to an ever increasing sequence of q's.

Correct Identification

Of course, knowing $B(L)^{\text{OLS}}$ and $D(L)$ is equivalent to knowing the fundamental VMA $C(L)$. Expressed in terms of the former the correct impact coefficients from (8) can be rewritten as

$$A(1) = \text{chol}\,\{(I - B(1)^{\text{OLS}})^{-1} S_v(0)(I - B(1)^{\text{OLS}})^{-T}\} \tag{24}$$

$$A_0 = D(1)^{-1} \underbrace{(I - B(1)^{\text{OLS}})A(1)}_{A_0^{\text{CEV}}} \tag{25}$$

CEV construct $A(1)^{\text{AM}}$ by plugging into (24) a spectral estimator for $S_v(0)$ (see (13)). But residual dynamics are ignored when mapping $A(1)^{\text{AM}}$ back into the impact coefficients. Such a practice errs in treating $D(1) = I$ for a given $A(1)$. The point of Remarks 2 and 4 is however, that $A(1)^{\text{AM}}$ includes an estimate of $D(1)$ which is not identical to the identity matrix. This is the source of the variance misrepresentation noted in Remark 3.

Many moving averages are observationally equivalent with a given spectrum, but only one of them is invertible. Proposition 1 tells us to look exactly for this fundamental representation of the data. Theorem 1 tells us that the spectral factorization gives us exactly the right $D(L)$ for that purpose.

Implementation

Theorem 1 can also be implemented empirically based on a spectral estimate like $S_v(\omega)^{\text{NW}}$ in (20), which is constructed as the truncated sum of b autocovariances. The factorization will then yield a unique and invertible MA(b), denoted $D(L)^{\text{LSF}}$, and an innovations variance matrix Ω_e^{LSF}. The superscript "LSF" indicates that these are calculated from a spectral factorization of sample moments from the least-squares residuals. Sayed and Kailath (2001) survey a number of different algorithms. I use a reliable and efficient algorithm from Li (2005), which is based on a state space representation of the moving average process of v_t^{OLS}. Details are given in Appendix A. Based on the spectral factorization, I propose the following procedure:

1. Estimate a VAR(p) to capture the main dynamics of the data. (Lag-length selection is chosen as usual, for example based on information criteria.)

2. Construct a spectral estimate $S_v(\omega)^{\mathrm{NW}}$ from the sample autocovariances of the VAR residuals (Bandwidth q can either be fixed or data dependent.)

3. Construct long run coefficients as in (14):

$$A(1)^{\mathrm{AM}} = \mathrm{chol}\left\{C(1)^{\mathrm{OLS}}\, S_v(0)^{\mathrm{NW}}\, C(1)^{\mathrm{OLS}\,T}\right\}$$

4. Factorize this spectral estimate into a MA(b) denoted $D(L)^{\mathrm{LSF}}$ with innovation variance Ω_v^{LSF}.

5. Short run coefficients are then

$$\begin{aligned} A_0^{\mathrm{LSF}} &= \left(D(1)^{\mathrm{LSF}}\right)^{-1}(I - B(1)^{\mathrm{OLS}})A(1)^{\mathrm{AM}} \\ &= \left(D(1)^{\mathrm{LSF}}\right)^{-1} A_0^{\mathrm{CEV}} \end{aligned}$$

The first three steps are identical to the CEV procedure, which is correct in its construction of $A(1)^{\mathrm{AM}}$. The spectral factorization is needed to obtain an estimate of $D(1)$ which corrects their impact coefficients and impulse responses. In addition to the VAR's lag length, the bandwidth b is a free parameter here. The lab simulations reported in Section 6 use both fixed bandwidth schemes, as CEV do, and the optimal, data-dependent selection scheme of Newey and West (1994).

Using population values of $B(L)^{\mathrm{OLS}}$ and $S_v(\omega)$, the spectral factorization correctly represents the true VMA and thus also the variance of X_t. The latter also applies to the sample variance of X_t when using sample estimates of the VAR and a spectrum like $S_v(\omega)^{\mathrm{NW}}$ constructed as a weighted and truncated sum of sample autocovariances:

Proposition 2 (OLS and Spectral Factorization correctly represent Variance of Data)**.** *Estimates of* A_0^{LSF} *and* $D(L)^{LSF}$ *are consistent with the sample variance of the VAR residuals*

$$\begin{aligned}\hat{\Omega}_v^{OLS} &\equiv \frac{1}{T}\sum_{t=1}^{T} v_t^{OLS}(v_t^{OLS})^T \\ &= \int_{-\pi}^{\pi} \hat{S}_v(\omega)^{NW} d\omega \end{aligned} \tag{26}$$

and thus consistent with the VAR's sample variance

$$\widehat{\mathrm{Var}\, X_t}^{OLS} = \sum_{k=0}^{\infty} \hat{C}_k^{OLS} \hat{\Omega}_v^{OLS} (\hat{C}_l^{OLS})^T \tag{27}$$

Despite the serial correlation of v_t^{OLS}*, this is the correct variance measure because of the normal equations which are enforced by OLS in sample.*[16]

Proof. (26) follows from the construction of the Newey-West estimate which is (20) evaluated at sample autocovariances $\hat{\Gamma}_k = \frac{1}{T}\sum_{t=k}^{T} v_t^{\mathrm{OLS}}(v_{t-k}^{\mathrm{OLS}})^T$ and since $\int_{-\pi}^{\pi} e^{-i\omega k} d\omega = 0$. (27) is the sample analogue to (19). □

5 Lab Economy

The previous section described various procedures for implementing long-run identifications in a VAR. The next section will assess their effectiveness with data simulated from a model economy, where the true coefficients are known. This model economy is described here. It is identical to the two-shock model used by CKM and CEV.

The model is a common one-sector RBC model driven by two shocks: First, a unit root shock to technology. This is the permanent shock to be estimated by the VAR. Second, a transitory shock which drives a wedge between private household's labor-consumption decision.

The representative household maximizes his lifetime utility over (per-capita) consumption, c_t, and labor services, l_t:

$$\max_{\{c,l\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} (\beta(1+\gamma))^t u(c_t, l_t)$$

and faces the budget constraint

$$c_t + (1+\gamma)k_{t+1} - (1-\delta)k_t = (1-\tau_{lt})w_t l_t + r_t k_t + T_t$$

where k_t is the per-capita stock of capital, w_t the wage rate, r_t the rental rate of capital, T_t are lump sum taxes, γ is the growth rate of population, δ the depreciation rate of capital ($\gamma > 0$, $0 \leq \delta \leq 1$ and $\beta < 1$).

τ_{lt} is an exogenous labor tax. As discussed by CKM, it need not be literally interpreted as a tax levy, but stands in for the effects of a variety of non-technology shocks introduced into second-generation RBC models. Mechanically, it distorts the first-order condition for consumption and labor. It works similar to a stochastic preference shock to the Frisch elasticity of labor supply. Chari, Kehoe, and McGrattan (2006) show how this labor "wedge" can be understood more generally as the *reduced form* process for more elaborate distortions, such as sticky wages.

The production function $F(k_t, Z_t l_t)$ is constant returns to scale, where Z_t is labor-augmenting technological progress. Firms are static and maximize profits

$$F(k_t, Z_t l_t) - w_t l_t - r_t k_t$$

Per-capita output equals production, $y_t = F(k_t, Z_t l_t)$, and the economy's resource constraint is

$$y_t = c_t + (1+\gamma)k_{t+1} - (1-\delta)k_t$$

The exogenous drivers follow linear stochastic processes:

$$\log Z_t = \mu_z + \log Z_{t-1} + \sigma_z \varepsilon_t^Z \tag{28}$$

$$\log \tau_{l,t+1} = (1-\rho_l)\bar{\tau}_l + \rho_l \log \tau_{l,t-1} + \sigma_l \varepsilon_t^l \tag{29}$$

where ε_t^Z and ε_t^l are *iid* standard-normal random variables. They are the technology shock, respectively labor shock. ρ_l measures the persistence of the transitory labor tax. The scale factors σ_z and σ_l determine their relative importance in the model. (μ_z is the drift in log-technology and $\bar{\tau}_t$ is the average tax rate.)

The calibration is identical to the baseline model of CKM and CEV, which uses parameter values familiar from the business cycle literature. Utility is specified as $u(c,l) = \log c + \psi \log(1-l)$ (consistent with balanced growth) and the production function is Cobb-Douglas $F(k,l) = k^\theta l^{1-\theta}$ with a capital share of $\theta = 0.33$. The labor preference parameter is set to $\psi = 2.5$. On an annualized basis, the calibration sets the depreciation rate to 6%, the rate of time preferences to 2% and population growth to 1%.[17]

The model economy is calibrated over different ratios in the variance of transitory to permanent shocks, σ_l^2/σ_z^2.[18] As a benchmark, note that the maximum-likelihood estimates of CEV imply a variance ratio of 0.345.[19] Following CEV, the transitory shock is calibrated as an AR(1) with persistence $\rho_l = 0.986$. For $\sigma_l^2/\sigma_z^2 = 0$, this is a one-shock RBC model and CKM show how our bivariate VAR will recover the correct impact coefficients in this case[20]

Appendix B shows how the linearized solution of the model can generally be represented by an infinite order VAR[21] In this VAR representation of the model, the technology shock satisfies the identifying assumptions made in (5) above.

Following CEV and CKM, the analysis looks at bivariate VARs in the growth rate of labor productivity and hours worked.[22]

$$X_t = \begin{bmatrix} \Delta y_t - \Delta l_t \\ l_t \end{bmatrix}$$

and the matrix of impact coefficients is now 2×2

$$A_0 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{30}$$

The impact responses of hours are thus

$$e_t^l = a_{21}\varepsilon_t^Z + a_{22}\varepsilon_t^l \tag{31}$$

where a_{21} is the immediate response of hours in response to a current shock in technology.

Mimicking the empirical literature, a small lag length is specified. Lag length, sample size and Newey-West truncation are as follows (identical to CEV and CKM)[23]:

$$p = 4 \qquad\qquad T = 180 \qquad\qquad b = 150$$

For each calibration, 10,000 samples are simulated.

[Figure 1 about here.]

As discussed in Section 4 above, it is a salient feature of the VAR(p) approximation that $D_i \approx 0$ ($\forall i > 0$) whereas $D(1) \neq I$. This holds also for our VAR(4) in this economy as can be seen in Figure 1. For the "CEV calibration" with $\sigma_l^2/\sigma_z^2 = 0.345$, it plots the population values of the cumulated sums $\sum_{k=0}^{K} D_k$. At each lag, the increments are small and close to zero[24], but summing over many lags we clearly have $D(1) \neq I$.

6 Performance in Simulated Economies

This section presents results from applying the various estimation procedures discussed in Section 2 to data simulated from the lab economy described in Section 5. The following questions are addressed:

1. Are the non-parametric estimates of the zero-frequency spectrum really better than OLS? What are the effects of bandwidth selection?

2. Is the misrepresentation of sample variance by CEV stated in Remark 3 quantitatively important?

3. How do OLS, CEV and LSF compare in terms of bias and sampling uncertainty of impact coefficients A_0?

4. Finally, the accuracy of variance decompositions is compared.

[Figure 2 about here.]

Goodness of Spectral Estimates

CEV's initial motivation is that the non-parametric estimates of Newey and West (1987) and Andrews and Monahan (1992) should yield better estimates of the spectral density at zero frequency than OLS. Two things are known from this literature, when it comes to the spectra of persistent data: First, estimation is improved by prewhitening with a VAR as with the Andrews-Monahan formula. Second, there are substantial small sample biases when persistence is high (Newey and West 1994; Andrews 1991). These results are confirmed by my simulations.

Figure 2 shows the median percentage errors for each element of the two-by-two matrix

$S_X(0)$. Percentage errors are defined as

$$\frac{\hat{S}_X(0) - S_X(0)}{S_X(0)} \cdot 100\%$$

where the division is elementwise. $\hat{S}_X(0)$ is the median of simulated estimates for a given estimator and $S_X(0)$ is the true spectrum from the model's population values, see Appendix B for details on computing population values. Since the spectrum is symmetric, the top right panel reports trace$(\hat{S}_X(0)\,(S_X(0))^{-1})/2$ as a joint measure of closeness. If the estimates were equal to the true vales, this would be equal to one.

First, biases are large in an absolute sense with estimates being off by about 100%. Comparing their relative performance, the Andrews-Monahan is generally doing better than Newey-West. This confirms the results of Newey and West (1994). What is striking, is that with an optimal bandwidth selection, the Andrews-Monahan spectrum is not substantially different from OLS. The bandwidths chosen for $S_v(\omega)^{\mathrm{NW}}$ vary between $b = 1$ and $b = 3$ and are not picking up any substantial serial dependence amongst estimated v_t^{OLS}.[25] Actually, this is no wonder, since the VAR's lag length has been chosen to whiten v_t^{OLS} as good as possible already. Using CEV's fixed and large bandwidth of $b = 150$, both Andrews-Monahan and Newey-West are doing substantially worse than OLS. Overall, OLS appears to yield amongst the best estimates of the zero-frequency spectrum. A similar picture emerges when looking at the corresponding percentage errors of $A(1) = \mathrm{chol}\,\{S_X(0)\}$.

These results are not encouraging for including spectral estimates in SVAR analysis. The optimal bandwidth procedures are very close to OLS also in terms of $A(1)$ and A_0. Henceforth, results will only be reported for CEV's fixed bandwidth selection of $b = 150$.

Variance Measures

Figure 3 reports measures for the variance of hours, $\mathrm{Var}(l_t)$ derived from the various procedures both in population and in sample. CEV variances are computed from (18) and OLS variance from (19), respectively from their sample analogues.[26] The population values are calculated from applying the estimation formulas to population moments. (Please recall that in population, OLS variance equals the true value by construction.) In particular, the spectral estimates are still calculated from the truncated and weighted summation of equation (20) with $b = 150$. The thought experiment is to isolate specification bias from small sample bias, not to consider what a researcher would see if he had an infinite amount of data. For the sample measures, medians are reported over 10,000 Monte Carlo draws.

In sample, the deviation of CEV from sample variance (OLS) is substantial. Both the A_0^{CEV} constructed from Andrews-Monahan and Newey-West understate total variance by at least half of the OLS variance, which again is approximately equal to the sample moments, (see Proposition 2). In population, Andrews-Monahan is quite close to the true value since the residual autocovariances are close to zero[27]

Another striking effect is visible in Figure 3: There is a large bias in the sample estimates when compared against the population. This is due to small sample bias (Hurwicz 1950) which is very active in these calibrations with $\rho_l = 0.986$. It is well known that autoregressive parameters are downward biased when the true process is close to unit root. Forecast error variance is however estimated quite well. As a corollary, the sample variance is understated as well.[28]

[Figure 3 about here.]

Bias and Uncertainty in Hours Impact

CEV claim that A_0^{CEV} "virtually eliminates bias" in estimated impact coefficients. Following CEV and CKM, I focus on the impact of technology on hours[29]. Median percentage errors are computed as

$$\frac{\hat{a}_{21} - a_{21}}{a_{21}} \cdot 100\%$$

where $\hat{a}_{21}$ is the median over 10,000 simulated estimations for a given estimator. The median errors are shown in Figure 4.

Ignoring the preceding discussions and taking A_0^{CEV} at face value, it is not even a panacea for estimating impact coefficients. The Newey-West version reported in Christiano, Eichenbaum, and Vigfusson (2006b) has actually larger biases than OLS for low to intermediate ratios between the variances of transitory and permanent shocks, σ_l^2/σ_z^2. This includes the preferred calibration of CEV.[30]

As can be anticipated from the preceding discussions, the Andrews-Monahan estimator for A_0^{CEV} is much closer to OLS. It has an almost uniformly lower bias than OLS, though. The LSF estimator behaves similarly, and has a somewhat smaller absolute bias for most calibrations. Again, a key difference between the LSF estimator and A_0^{CEV} is also that it is fully consistent with the OLS sample variance, whereas A_0^{CEV} is not.

[Figure 4 about here.]

The simulated distributions of these impact errors are shown in Figure 5. The spread in simulated estimates is huge – swamping even the considerable size of the biases. Even the 68% confidence intervals regularly span errors exceeding minus 100%, which means that they include estimates of a_{21} having the wrong sign. This is the case for OLS as well as the various spectral methods. The Newey-West A_0^{CEV} has a considerably tighter distribution of errors. Based on the preceding discussion, this estimator however appears to be the least useful. However, even this finding is not robust to changes in the model's calibration. For a model with less persistence – $\rho_l = 0.5$ – OLS errors are more tightly distributed than CEV. This is shown in Figure 1 of the webappendix.

[Figure 5 about here.]

As discussed before in the context of Figure 2, OLS spectra have almost uniformly better bias properties than the spectral estimators. So how is it possible that for some of these calibrations, A_0^{CEV} has a lower bias than OLS? Figure 6 decomposes the biases in A_0^{OLS} and A_0^{CEV} into effects from small sample and misspecification issues. Across calibrations of σ_l^2/σ_z^2, the small sample bias in A_0^{CEV} is almost constant at around -10% and variations in the performance of A_0^{CEV} are caused by variations in its truncation bias which is steadily rising.[31]

[Figure 6 about here.]

Variance Decompositions

Apart from impulse response analysis, an important application of SVARs are variance decompositions. They ask "What share of total variance is explained by technology shocks?" For the innovation in hours (31), the share of variance explained by technology equals

$$\frac{a_{21}^2}{(a_{21}^2 + a_{22}^2)}$$

[Figure 7 about here.]

The distributions of OLS, CEV and LSF for this measure are shown in Figure 7. The figure also displays the population estimates as well as the true variance share. A pertinent feature of the underlying model is that hours do not respond much to permanent shocks[32]

So apart from calibrations where technology is almost the only driving force, the true variance rapidly drops to values below 10%.

The results are sobering again: All procedures overstate the variance share by 10 to 20 percentage points (medians) and the 68% confidence intervals easily span values between 10% and 60% (or wider). If anything, OLS is doing a better job than the spectral estimates – both in terms of a somewhat lower bias and tighter confidence bands. Again, the CEV procedure is not a panacea. And neither is LSF when looking at variance decompositions. The high persistence of the underlying model makes estimation generally harder. For example, when $\rho_l = 0.5$ OLS performs much better than CEV as is shown in the webappendix.

7 Conclusions

Using non-parametric estimators to learn about dynamics missed by a VAR sounds appealing. But when combining these two estimations, we need a consistent account of the fluctuations in the data. When a VAR(p) is used to approximate what is truly a VAR(∞), then its residuals will be serially correlated $v_t^{\text{OLS}} = D(L)e_t$. However, OLS computes spectra and and short run responses as if they were *iid*. But even a misspecified VAR will correctly represent the total variance of the data. Combining the VAR with spectral estimates requires an adjustment in order to retain this property and to compute the correct impact coefficients. This can be achieved in sample with a spectral factorization of the non-parametric estimates.

The long-run responses of CEV allow for non-zero moving average terms in v_t^{OLS} which have permanent effects, $D(1) \neq I$. Once long run responses are constructed from the non-parametric estimates, these permanent effectsare disregarded by the CEV procedure when computing the short-run responses A_0^{CEV}. The correct responses are however

$$A_0 = D(1)^{-1} A_0^{\text{CEV}} \tag{32}$$

Using simulations from the lab economy used by CEV and CKM, I demonstrate that the total variance is seriously misrepresented by treating $D(1) = I$ above. This is particularly so when using the Newey-West spectrum as in Christiano, Eichenbaum, and Vigfusson (2006b). The Andrews-Monahan estimates used in Christiano, Eichenbaum, and Vigfusson (2006a) inherit a lot more structure from the VAR and thus the problem is less prevalent. Related to this inconsistency is that the estimated time series of shocks will be a mere rescaling of the OLS estimates (see Remark 1), even though impulse responses are not. This is of practical importance to researchers interested in adopting their strategy. These issues are resolved with the spectral factorization presented in this paper.

A deeper question is whether and how spectral estimates can actually help to overcome the biases associated with OLS. After all they are calculated from sample moments of the data, just as the VAR and its lag-length selection criteria. Erceg, Guerrieri, and Gust (2005) and Chari, Kehoe, and McGrattan (2005) have already highlighted that there are two kinds of biases: The truncation bias arising from a misspecified VAR and the Hurwicz-type bias in coefficients estimated from small samples of data with high persistence (Hurwicz 1950). Whilst the spectral estimates may offer a way around the VAR's misspecification, they are subject to similar small sample biases. Indeed my simulation results paint a sobering picture:

- Based on optimal bandwidth selection procedures, the spectral estimates do not deviate much from OLS. This is simply because the VAR's lag-length has already been chosen to whiten the residuals as good as possible. OLS estimates *of the spectra* are generally much better, having mostly smaller biases, than the high-bandwidth spectral estimates used by CEV.

- Looking at the impact coefficients A_0, the large and fixed bandwidth advocated by CEV ($b = 150$ in a sample of $T = 180$) can improve on OLS, but it is no panacea either. Depending on the true process it can have larger biases and larger sampling uncertainty. The spectral factorization almost uniformly improves upon OLS while providing a correct account of total variation – the gains appear to be small however.

- When CEV sometimes outperforms OLS, it is not because of better spectral estimates, but because of canceling biases in coefficients from OLS and non-parametric estimates *and* the "freedom" to deviate from correctly modeling the data's variance in the sample.

The corrected procedure yields a VARMA representation of the data where the MA-process is orthogonal to the lagged variables in the VAR. My results complement other studies looking at the performance of conventional VARMA specifications, notably McGrattan (2006) and Kascha and Mertens (2006) (not related to me). Their specifications enjoy the benefit of being chosen to match exactly the underlying model, whereas my procedure is fairly agnostic in its specification of lags in the VAR and MA component. All in all, their results as well as mine point to small sample biases and not just specification issues being a key factor in the estimation of permanent shocks and their effects on the business cycle. One way to handle this small sample bias is to compare the estimated small sample moments with those simulated from a model economy as done by Cogley and Nason (1995) and advocated by Kehoe (2006a) and Dupaigne, Feve, and Matheron (2007) in the context of the SVAR discussion.

Appendix

A Spectral Factorization Method

Spectral factorization has a long tradition for instance in linear quadratic control, robust estimation and control, see for example Whittle (1996). For a reference in the context of economics see Hansen and Sargent (2007, 2005). Theorem 1 above is adapted from Hannan (1970, p. 66). A similar version is also stated by Li (2005). It assures us of the existence and uniqueness of an MA(q) polynomial $D(L)$, based on an autocovariance function with q elements. Invertibility of the MA(q) follows immediately when excluding the case of zero power of the spectral density at zero-frequency:

Corollary. *Suppose that $S(0) \neq 0$. Since Ω is positive definite, it follows that $D(1) \neq 0$. All roots of $D(z)$ are thus outside the unit circle and $D(L)$ is an invertible MA(q).*

In the context of this paper, $S_v(\omega)$ will be the spectral density of $v_t^{\text{OLS}} = D(L)e_t$ where $Ee_te_t' = \Omega = A_0A_0'$. We will be using non-parametric estimates of $S_v(\omega)$ based on weighted sums of the *sample* autocovariance function as described in Section 2.2. The sample autocovariances are however not to be confused with the $\Gamma(z)$ above. In the above theorem $\Gamma(z)$ is the inverse Fourier transform of the spectral density, and thus a smoothed version of the sample autocovariance.

Theorem 1 requires $S_v(\omega)$ to be non-singular. This can be understood as requiring that the autocovariances need to decay sufficiently fast in relation to the number of MA lags. For example, in the scalar case and with $q = 1$, the first-order autocorrelation to be matched with a MA(1) cannot be larger than 0.5 in absolute value.[33]

Algorithms for implementing the factorization have also a long tradition, see for example Whittle (1963) or Sayed and Kailath (2001) for a recent survey. I use the implementation of

Li (2005) which is based on a state space representation of v_t and performs very reliably in my simulations. The algorithm of Li (2005) is described in the remainder of this appendix.

Suppose v_t follows an MA(q) as above. To represent it in a state space system, define the state vector

$$x_t = E\left\{ \begin{bmatrix} v_t & v_{t+1} & \dots & v_{t+q-1} \end{bmatrix}' \middle| v^{t-1} \right\} \tag{33}$$

where v^{t-1} is the entire history of realizations of the v_t process up to time $t-1$. Li then constructs the following state space system

$$x_{t+1} = Ax_t + Be_t \tag{34}$$

$$v_t = Cx_t + e_t \tag{35}$$

whose system matrices are equal to

$$A = \begin{bmatrix} 0_m & I_m & 0_m & \dots & & 0_m \\ 0_m & 0_m & I_m & 0_m & \dots & 0_m \\ \vdots & & & \ddots & & \vdots \\ 0_m & \dots & & & 0_m & I_m \\ 0_m & \dots & & & 0_m & 0_m \end{bmatrix} \tag{36}$$

$$C = \begin{bmatrix} I_m & 0_m & \dots & 0_m \end{bmatrix} \tag{37}$$

$$D = \begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ D_q \end{bmatrix} \tag{38}$$

where I_m and 0_m are the $m \times m$ identity matrix, respectively the $n \times n$ zero matrix.

We need a mapping from autocovariances Γ_k to the state space objects. Our objects of interest are the matrix D containing the stacked MA coefficients D_i as well as the variance Ω_e of the innovations process. For this mapping, stack the autocovariances into the matrix

$$M = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \vdots \\ \Gamma_q \end{bmatrix}$$

Li (2005, Theorem 2) shows that variance matrix of the states

$$\Psi \equiv E x_t x_t'$$

solves the following Riccati equation

$$\Psi = A\Psi A' + (M - A\Psi C')(\Gamma_0 - C\Psi C')^{-1}(M - A\Psi C')' \tag{39}$$

which allows the following mapping to our MA(q) coefficients:

$$\Omega_e = \Gamma_0 - C\Psi C'$$
$$D = (M - A\Psi C')(\Gamma_0 - C\Psi C')^{-1}$$

This yields coefficients D such that $D(L)$ has all roots outside or on the unit circle.

The above Riccati equation can be solved using standard routines. Li shows that it can also be solved recursively starting from $\Psi^{(0)} = 0$ and iterating over

$$\Psi^{(n+1)} = A\Psi^{(n)}A' + (M - A\Psi^{(n)}C')(\Gamma_0 - C\Psi^{(n)}C')^{-1}(M - A\Psi^{(n)}C')'$$

where $\Psi^{(n+1)} \geq \Psi^{(n)}$ and $\Psi = \lim_{n\to\infty} \Psi^{(n)}$.[34]

At the end of each factorization, I check that the factorization produces an invertible MA(q) polynomial and that it matches the original spectral density. In all simulations, this holds up to machine accuracy.

The paper of Li also shows how to reduce the number of iterations by stacking the MA(q) into first order form, however this comes at the cost of iterating over inverting larger matrices in the Riccati iterations which I found to yield numerically less stable solutions.

B VAR's Implied by Lab Economy

This section outlines how to derive the following:

1. The true values of A_0, $A(1)$, $B(1)$, and the autocovariances of X_t in the lab economy

2. Population coefficients of finite-order VAR's implied by the lab economy

For this specific two-shock economy, details of these derivations can be found in McGrattan (2005). For general state space models details can be found in Fernandez-Villaverde, Rubio-Ramirez, and Sargent (2005). To simplify the VAR notation, X_t has been demeaned prior to the analysis.

The linearized solution to the lab economy described in Section 5 yields a state space model for labor productivity growth and hours

$$X_t = \begin{bmatrix} \Delta y_t - \Delta l_t \\ l_t \end{bmatrix} \tag{40}$$

$$= \mathcal{C} Z_t$$

$$= \mathcal{C}\mathcal{A} Z_{t-1} + \mathcal{C}\mathcal{B}\varepsilon_t$$

$$Z_t = \mathcal{A} Z_{t-1} + \mathcal{B}\varepsilon_t \tag{41}$$

State vector and shock vector are:

$$Z_t = \begin{bmatrix} \hat{k}_t \\ \varepsilon_t^z \\ \tau_{l,t}^z \\ \hat{k}_{t-1} \\ \varepsilon_{t-1}^z \\ \tau_{l,t-1}^z \end{bmatrix} \qquad \varepsilon_t = \begin{bmatrix} \varepsilon_t^z \\ \varepsilon_t^l \end{bmatrix} \tag{42}$$

where $\hat{k}_t$ is the log-deviation of detrended capital from its stead state state, $\tau_{l,t}$ and ε_t^z are the labor wedge and the growth rate in technology. (Z_t includes also lagged variables due to the presence of labor productivity *growth* in X_t.)

The computation of the matrices $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ is straightforward, please see CKM for a detailed presentation.

True VAR objects

The decomposition in section 6 uses the following objects of the true process: A_0, $A(1)$, $B(1)$ as well as the autocovariances of X_t. Their computation from the state space is straightforward. True impulse responses and spectrum of X_t are given by

$$A(L) = \mathcal{C}\left(I - \mathcal{A}L\right)^{-1}\mathcal{B}$$
$$S_X(\omega) = A(e^{-i\omega})A(e^{-i\omega})^T$$

From (41), it is apparent how the structural shocks are linearly related to the forecast errors of X_t:

$$\mathcal{C}\mathcal{B}\varepsilon_t = e_t$$
$$\Rightarrow \qquad \mathcal{C}\mathcal{B} = A_0$$

In order to map forecast errors into structural shocks, A_0 must obviously be square and invertible. Furthermore, Fernandez-Villaverde, Rubio-Ramirez, and Sargent (2005) show that invertibility requires the eigenvalues of $\mathcal{A} - \mathcal{B}(\mathcal{C}\mathcal{B})^{-1}\mathcal{C}\mathcal{A}$ to be strictly less than one in modulus, which is satisfied for all calibrations considered here.

Recall from equation (2) that this also ties down the covariance matrix of the forecast errors $\Omega = \mathcal{C}\mathcal{B}\mathcal{B}\mathcal{C}^T$. Applying (4), the non-structural moving average representation is then simply

$$X_t = A(L)A_0^{-1}e_t = C(L)e_t$$
$$\Rightarrow \quad B(1) = I - C(1)^{-1}$$

The autocovariances $EX_tX_{t-k}^T$ can be directly computed from the state space model, see for instance the textbook of Sargent and Ljungqvist (2004). The covariance matrix of the states $EZ_tZ_t^T \equiv \Omega$ is obtained as the solution to a discrete Lyapunov equation:

$$\Omega = \mathcal{A}\Omega\,\mathcal{A}^T + \mathcal{B}\mathcal{B}^T$$

The autocovariances of X_t are then

$$EX_tX_{t-k}^T = \mathcal{C}\mathcal{A}^k\Omega\,\mathcal{C}^T$$

VAR(p) coefficients in population

Chari, Kehoe, and McGrattan (2005, Proposition 1) show that the VAR representation of X_t in the model is of infinite order. Still, finite-order VAR(p) can be computed as projections of X_t on a finite number of its past values, $X_{t-1}\ldots X_{t-p}$. Their residuals will however not be martingales. In line with the notation of the main text, population coefficients of a VAR(p) are denoted with a superscript "OLS".

$$X_t = B(L)^{\mathrm{OLS}}\,X_{t-1} + v_t^{\mathrm{OLS}}$$

The coefficients of the lag polynomial $B(L)^{\mathrm{OLS}} = \sum_{i=0}^{p-1} B_i^{\mathrm{OLS}}\,L^i$ solve the OLS normal equations

$$E\underbrace{\left(X_t - \sum_{i=0}^{p-1} B_i^{\mathrm{OLS}}\,X_{t-1-i}\right)}_{=v_t^{\mathrm{OLS}}} X_{t-j}^T = 0 \quad \forall\, j = 1\ldots p$$

which are evaluated using the autocovariance matrices of X_t whose computations are described in the preceding paragraph. For instance if $p = 1$, $B_1^{\mathrm{OLS}} = \left(EX_tX_{t-1}^T\right)\left(EX_tX_t^T\right)^{-1}$. (Detailed formulas for higher VAR's can be found in Fernandez-Villaverde, Rubio-Ramirez, and Sargent (2005).)

Notice that by construction, the projection residuals v_t^{OLS} are orthogonal to $X_{t-1}, \ldots,$ X_{t-p}, but they are not orthogonal to the entire history of X_t, because of the truncation bias in the VAR(p). Their moving average representation $v_t^{\text{OLS}} = D(L)e_t$ is easily constructed using

$$D(L) = \left(I - B(L)^{\text{OLS}}\right)(I - B(L))^{-1} \tag{43}$$

Variance equation

Even though the VAR(p) residuals v_t^{OLS} are not *iid*, the usual variance equation is still applicable. For notational convenience, take $p = 1$: $X_t = B_1^{\text{OLS}} X_{t-1} + v_t^{\text{OLS}}$. The normal equations imply

$$\begin{aligned} \operatorname{Var} X_t &= B_1^{\text{OLS}} \, (\operatorname{Var} X_t) \, (B_1^{\text{OLS}})^T + \Omega_v^{\text{OLS}} \\ &= \sum_{k=0}^{\infty} (B_1^{\text{OLS}})^k \, \Omega_v^{\text{OLS}} \, ((B_1^{\text{OLS}})^k)^T \\ &= \sum_{k=0}^{\infty} C_k^{\text{OLS}} \, \Omega_v^{\text{OLS}} \, (C_k^{\text{OLS}})^T \end{aligned} \tag{44}$$

To see that the last line holds for general VAR(p), rewrite the VAR in companion form and derive its variance analogously to (44).

Notes

1. Comparable to lag length in a time series model, bandwidth is a key parameter in spectral estimation. See Section 2.2 for details.

2. For a demonstration, see the web-appendix of this paper at `http://www.elmarmertens.ch`.

3. `http://www.elmarmertens.ch`

4. For notational convenience, but without loss of generality, X_t represents the demeaned variables. As an application of the Frisch-Waugh-Lowell Theorem, this is both theoretically and numerically equivalent to including a constant in a VAR using the original data.

5. See for example Giannone and Reichlin (2006) on the non-invertibility problem.

6. The spectral density $S_X(0) = C(1)\Omega C(1)^T$ is strictly positive definite when the forecast errors e_t are linearly independent, which implies that their variance covariance matrix Ω is nonsingular. $S_X(0)$ inherits positive definiteness from Ω since $C(1)$ is nonsingular. $I - B(1) = C(1)^{-1}$ exists because of the assumed stationarity of the VAR process.

7. In general, $A(1)$ is described by

$$A(1) = \text{chol}\left\{S_X(0)\right\} \begin{bmatrix} 1 & 0 \\ 0 & W \end{bmatrix} \quad \text{where } WW^T = I$$

In the lab economy described later, the VAR will be bivariate and the forecast errors e_t are a linear combination of only two shocks. Knowing the technology shock will then also identify the second shock up to its sign, $|W| = 1$.

8. To derive this apply the definition of spectral density (6) to v_t^{OLS} and recall that the variance equals the integral under the associated spectral density. Finally note that we can write $B(L)^{\text{OLS}} = B(L) + (B(L)^{\text{OLS}} - B(L))$.

9. For some variable Z_t, the sample moment is $E_T\, Z_t \equiv 1/T \sum_{t=1}^{T} Z_t$.

10. The discussion of CEV suggests this practice to be compatible with consistency. Watson (2006) regards it as a practically untruncated and inconsistent estimator.

11. This is a standard result for inverting partitioned matrices, see for example Magnus and Neudecker (1988, p. 11).

12. Likewise she wants $S_X(0)^{\mathrm{NW}} \neq S_X(0)^{\mathrm{OLS}}$ and thus

$$(I - B(1)^{\mathrm{OLS}})S_X(0)^{\mathrm{NW}}(I - B(1)^{\mathrm{OLS}})^T \neq \Omega^{\mathrm{OLS}}$$

13. See Proposition 2 below for details.

14. The theorem assures us of a $D(L)$ which has no roots inside the unit circle. Excluding zero spectra, all roots have to be outside the unit circle.

15. See for example Hamilton (1994, Chapter 3.A) or Hayashi (2000, Chapter 6).

16. Up to the treatment of initial observations this is also identical to the sample variance

$$\widehat{\operatorname{Var} X_t}^{\mathrm{OLS}} \approx \frac{1}{T}\sum_{t=1}^{T} X_t X_t^T$$

Take a first order VAR and index the data from $-(p-1), \ldots, 0, \ldots T$. The approximation holds to the extent that

$$\left(\frac{1}{T}\sum_{t=1}^{T} X_t^2\right) \approx \left(\frac{1}{T}\sum_{t=-p+1}^{T-p} X_t^2\right)$$

The exact relationship enforced by the OLS normal equations is

$$\operatorname{Var}_T X_t \equiv \frac{1}{T}\sum_{t=1}^{T} X_t^2 = B_1^{\mathrm{OLS}}\left(\frac{1}{T}\sum_{t=-p+1}^{T-p} X_t^2\right)(B_1^{\mathrm{OLS}})^T + \frac{1}{T}\sum_{t=1}^{T}\left(v_t^{\mathrm{OLS}}\right)^2$$

(Please recall that the data is demeaned.)

17. The drift in technology is set to 0.4% and the average "labor tax" is set to 24.2% per quarter.

18. CKM extensively document how different ratios in the variance of transitory to permanent shocks, σ_l^2/σ_z^2, affect the performance of standard VARs both in population and in small sample. McGrattan (2005) shows that (only) in the limit, $\sigma_l/\sigma_z \to 0$, a finite

order VAR (even a $p = 1$) recovers the true responses – though the true system is not a finite-order VAR. The OLS error thus shrinks to zero with the variance ratio. This can also be seen in the results below.

19. CEV estimate $\sigma_l = 0.0056^2$, $\sigma_z = 0.00953$.

20. An important aspect of this result is that there are more observables than shocks in this case.

21. For all calibrations, the model satisfies the condition of Fernandez-Villaverde, Rubio-Ramirez, and Sargent (2005) for an invertible mapping from structural shocks to forecast errors.

22. In addition to this “LSVAR” specification, CKM run also VARs with quasi-differenced hours. This replaces the second VAR element l_t with $(1 - \alpha L)\ l_t$ ($\alpha \in \{0; 0.999\}$). Depending on α, this captures popular (but also contested) specifications: On the one hand the “LSVAR” with hours in levels and $\alpha = 0$ and on the other hand the “QDSVAR” with $\alpha = 0.999$, which approximates a VAR with differenced hours without introducing a unit MA root. Sensitivity and sensibility of results to these choices are discussed amongst others by Gali and Rabanal (2004) and Christiano, Eichenbaum, and Vigfusson (2003). The quasi-differencing is discussed in more detail by CKM.

23. Instead of a fixing the lag length, it could also be chosen for each simulation based an information criteria. But for the simulations considered here, this does not affect results substantially.

24. They would only be barely visible on the graph.

25. Bandwidths selected for $S_X(\omega)^{\mathrm{NW}}$ vary between $b = 6$ and $b = 10$.

26. The variance of hours is the bottom left element of those matrices.

27. This is a corollary of $D_i \approx 0$ as documented in Figure 1.

28. For example, consider a simple AR(1), $x_t = \rho x_{t-1} + \sigma_e \varepsilon_t$, whose variance is decreasing in ρ: $\operatorname{Var} x_t = \sigma_e^2/(1-\rho^2)$. A numerical demonstration for this example can be found at `http://www.elmarmertens.ch`.

29. Cursory inspections suggest that results are qualitatively similar for the impact of output.

30. Curiously, previous versions of CKM were calibrated to higher values of this ratio where A_0^{CEV} fairs better.

31. The webappendix shows that constancy of the small sample bias is the effect of canceling biases in $B(1)^{\text{OLS}}$ and $A(1)^{\text{AM}}$ (respectively $A(1)^{\text{NW}}$), see Figure 5 there. The performance of A_0^{CEV} is not (solely) determined by the quality of the $A(1)$ estimates, but by the interaction of various biases arising from truncation and small effects, as well as canceling effects from $B(1)^{\text{OLS}}$ and the spectral estimate.

32. This is a direct consequence of the log-log preferences with canceling substitution and income effects. See also Kehoe (2006b)'s discussion in this context.

33. Given a covariance γ_0 and first-order autocovariance γ_1, the spectrum equals $s(\omega) = \gamma_0 \cdot (1 + 2\gamma_1 \cos(\omega))$. And $|s(\omega)| \neq 0$ requires $|\gamma_1/\gamma_0| < 0.5$.

34. Alternatively, a standard Riccati solver such as Matlab's `dare` could be used. I found both routines to operate numerically equally well. The relative performance of the alternatives routines can differ substantially but it also depends on the size of the problem and the underlying operating system. A clearly superior choice did not emerge.

References

Acemoglu, Daron, Kenneth Rogoff, and Michael Woodford, eds. 2006. *NBER Macroeconomics Annual 2006.* Cambridge, MA: MIT Press.

Altig, David, Lawrence J. Christiano, Martin Eichenbaum, and Jesper Linde. 2004, December. "Firm-Specific Capital, Nominal Rigidities and the Business Cycle." Northwestern University, *mimeo*.

Andrews, Donald W K. 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59 (3): 817–58 (May).

Andrews, Donald W K, and J Christopher Monahan. 1992. "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator." *Econometrica* 60 (4): 953–66 (July).

Blanchard, Olivier Jean, and Danny Quah. 1989. "The Dynamic Effects of Aggregate Demand and Supply Disturbances." *The American Economic Review* 79 (4): 655–673 (September).

Canova, Fabio, and Gianni de Nicolo. 2003. "On the sources of business cycles in the G-7." *Journal of International Economics* 59 (1): 77–100 (January).

Chari, V. V., Patrick J. Kehoe, and Ellen R. McGrattan. 2005, July. "A critique of structural VARs using business cycle theory." Staff Report 364, Federal Reserve Bank of Minneapolis.

———. 2006. "Business cycle accounting." *Forthcoming Econometrica*.

Christiano, Lawrence J., Martin Eichenbaum, and Robert Vigfusson. 2003, July. "What Happens After a Technology Shock?" Working Paper 9819, NBER.

———. 2006a. "Alternative Procedures for Estimating Vector Autoregressions identified with Long-Run Restrictions." *Journal of the European Economic Association* 2 (2–3): 475–483 (April–May).

———. 2006b. "Assessing Structural VARs." In Acemoglu, Rogoff, and Woodford 2006, Chapter 1.

Cogley, Timothy, and James M Nason. 1995. "Output Dynamics in Real-Business-Cycle Models." *The American Economic Review* 85 (3): 492–511.

Cooley, Thomas F., and Mark Dywer. 1998. "Business cycle analysis without much theory A look at structural VARs." *Journal of Econometrics* 83 (1-2): 57–88 (March-April).

Dupaigne, Martial, Patrick Feve, and Julien Matheron. 2007. "Avoiding pitfalls in using structural VARs to estimate economic models." *Review of Economic Dynamics* 10 (2): 238–255 (April).

Erceg, Christopher J., Luca Guerrieri, and Christopher Gust. 2005. "Can Long-Run Restrictions Identify Technology Shocks?" *Journal of the European Economic Association* 3 (6): 1237–1278 (December).

Faust, Jon. 1998. "The robustness of identified VAR conclusions about money." *Carnegie-Rochester Conference Series on Public Policy* 49 (December): 207–244.

Fernandez-Villaverde, Jesus, Juan Rubio-Ramirez, and Thomas J. Sargent. 2005, June. "A, B, C's (and D)'s for Understanding VARs." Technical Working Paper 0308, NBER.

Fernàndez-Villaverde, Jesús, Juan F. Rubio-Ramirez, Thomas J. Sargent, and Mark W. Watson. 2007. "ABCs (and Ds) of Understanding VARs." *American Economic Review* 97 (3): 1021–1026 (June).

Gali, Jordi. 1999. "Technology, Employment, and the Business Cycle: Do Technology Shocks explain Aggregate Fluctuations?" *The American Economic Review* 89 (1): 249–271.

Gali, Jordi, and Pau Rabanal. 2004, July. "Technology Shocks and Aggregate Fluctuations: How well does the RBC Model fit postwar U.S. data?" NBER Working Paper 10636.

Giannone, Domenico, and Lucrezia Reichlin. 2006. "Does information help recovering structural shocks from past observations?" *Journal of the European Economic Association* 4 (2-3): 455–465 (April-May).

Hamilton, James D. 1994. *Time-Series Analysis.* Princeton, NJ: Princeton University Press.

Hannan, E. J. 1970, June. *Multiple Time Series.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons.

Hansen, Lars P., and Thomas J. Sargent. 2005, September. "Recursive Models of Dynamic Linear Economies." unpublished Manuscript.

———. 2007, January. "Robustness." unpublished Manuscript.

Hayashi, Fumio. 2000. *Econometrics.* Princeton, NJ: Princeton University Press.

Hurwicz, Leonid. 1950. "Least-Squares Bias in Time Series." Chapter XV of *Statistical Inference in Dynamic Economic Models*, edited by Tjalling C. Koopmans, Volume 10 of *Cowles Commission for Resarch in Economics*, 365–383. New York: John Wiley & Sons, Inc.

Kascha, Christian, and Karel Mertens. 2006. "Business Cycle Analysis and VARMA models." Economics Working Papers 37, European University Institute.

Kehoe, Patrick J. 2006a. "How to advance theory with structural VARs: use the Sims-Cogley-Nason approach." Staff Report 379, Federal Reserve Bank of Minneapolis.

———. 2006b. "NBER Macroeconomics Annual 2006." In Acemoglu, Rogoff, and Woodford 2006, Chapter 1.

Li, Lei M. 2005. "Factorization of moving-average spectral densities by state-space representations and stacking." *Journal of Multivariate Analysis* 96 (2): 425–438 (October).

Magnus, Jan R., and Heinz Neudecker. 1988. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley Series in Probability and Statistics. John Wiley & Sons.

McGrattan, Ellen. 2005, May. "Appendix: A Critique of Structural VARs Using Real Business Cycle Theory." *mimeo.*

McGrattan, Ellen R. 2006. "Measurement with minimal theory." Working Papers 643, Federal Reserve Bank of Minneapolis.

Newey, Whitney K, and Kenneth D West. 1994. "Automatic Lag Selection in Covariance Matrix Estimation." *Review of Economic Studies* 61 (4): 631–53.

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

Priestley, M.B. 1981. *Spectral Analysis and Time Series.* Probability and Mathematical Statistics. Amsterdam: Elsevier Academic Press.

Ravenna, Federico. 2007. "Vector autoregressions and reduced form representations of DSGE models." *Journal of Monetary Economics* 54 (7): 2048–2064 (October).

Sargent, Thomas J., and Lars Ljungqvist. 2004. *Recursive Macroeconomic Theory.* 2nd. Cambridge, MA: The MIT Press.

Sayed, A.H., and T. Kailath. 2001. "A Survey of Spectral Factorization Methods." *Numerical Linear Algebra with Applications* 8:467–496.

Sims, Christopher A. 1972. "The Role of Approximate Prior Restrictions in Distributed Lag Estimation." *Journal of the American Statistical Association* 67 (337): 169–175 (March).

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

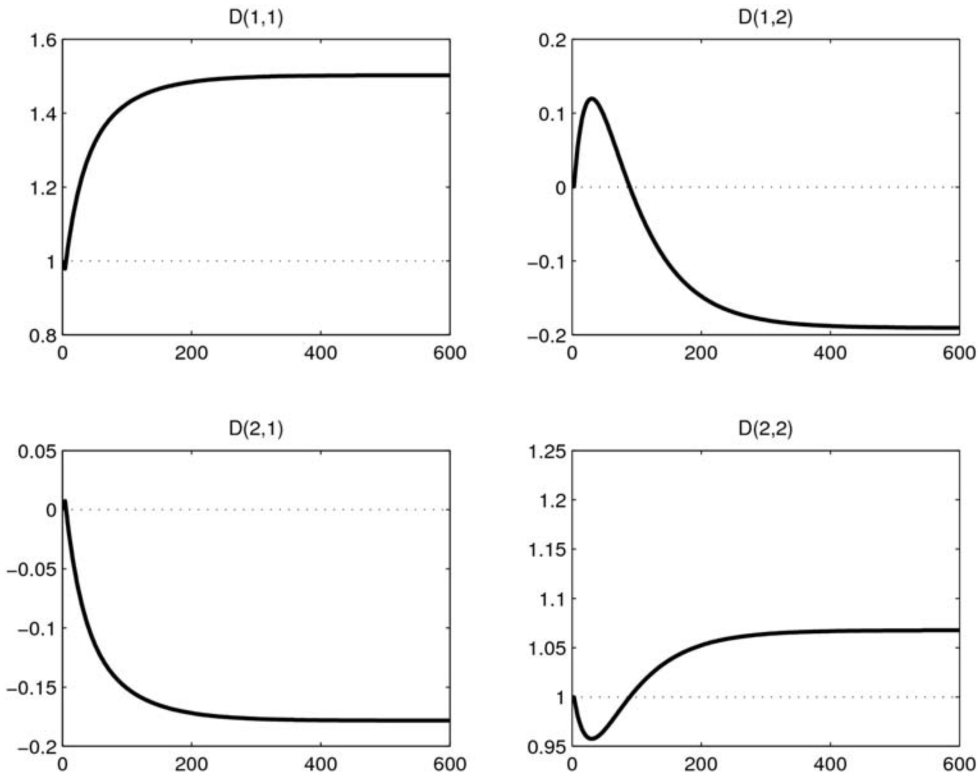
Watson, Mark W. 2006. "Comment on Christiano, Eichenbaum, and Vigfusson's "Assessing Structural VARs"." In Acemoglu, Rogoff, and Woodford 2006, Chapter 1.

Whittle, P. 1963. "On the Fitting of Multivariate Autoregressions, and the Approximate Canonical Factorization of a Spectral Density Matrix." *Biometrika* 50 (1/2): 129–134 (June).

Whittle, Peter. 1996. *Optimal Control, Basics and Beyond.* Wiley-Interscience Series in Systems and Optimization. Chichester: John Wiley & Sons.

List of Figures

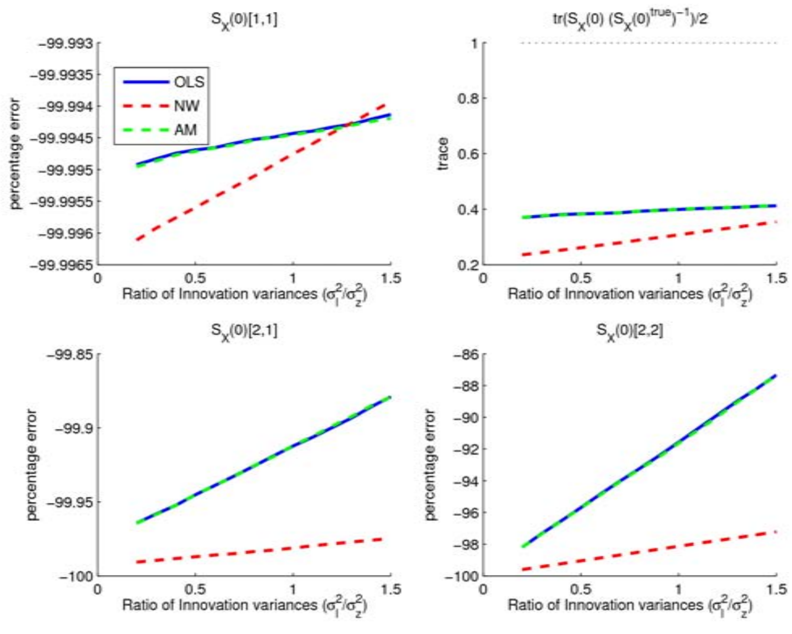
Figure 1: $D(1) \neq I$



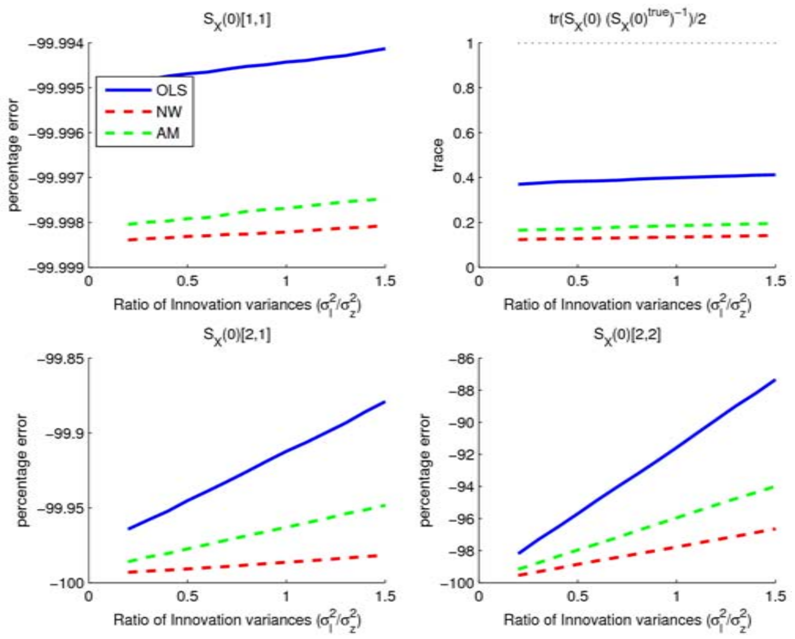
Note: Each panel shows an element of the cumulated sum $\sum_{k=0}^{K} D_k$, which is a two-by-two matrix, for the bivariate VAR in the lab economy of Section 5. "CEV calibration" with $\rho_l = 0.986$ and $\sigma_l^2/\sigma_z^2 = 0.345$. Lags K on the x-axis.

Figure 2: Median Errors in Spectral Estimates

(a) b automatically selected

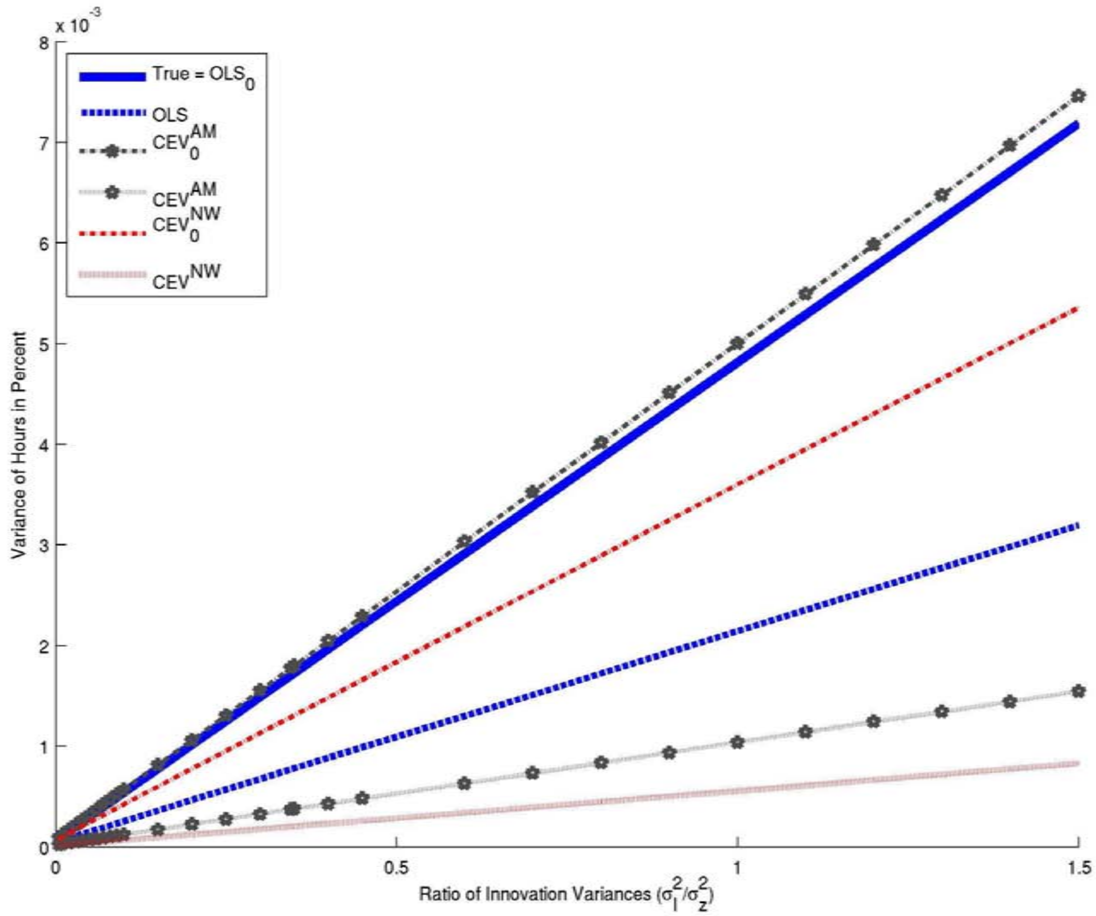


(b) $b = 150$



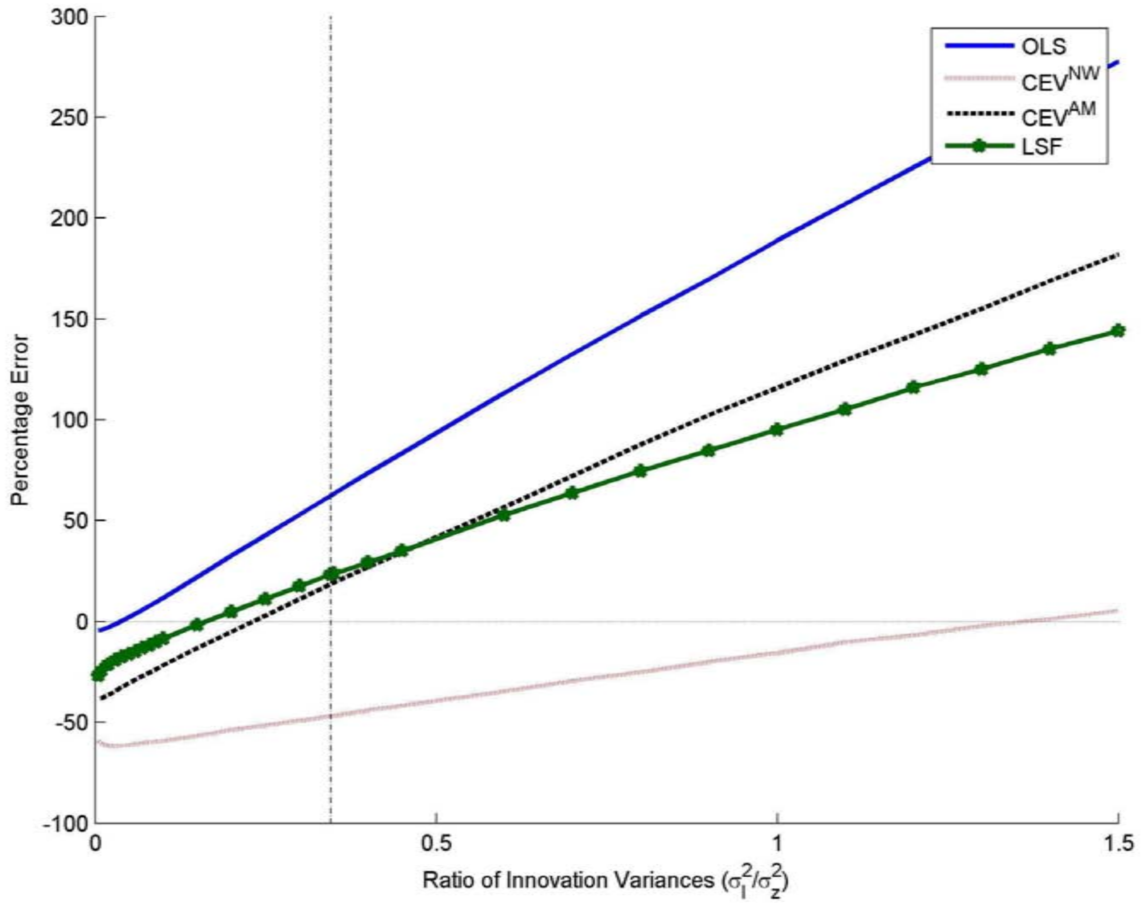
Note: Median percentage errors of spectral estimates over 10,000 simulations with $\rho_l = 0.986$. Automatic bandwidth selection in Panel a) uses scheme of Newey and West (1994).

Figure 3: Measures for Hours Variance



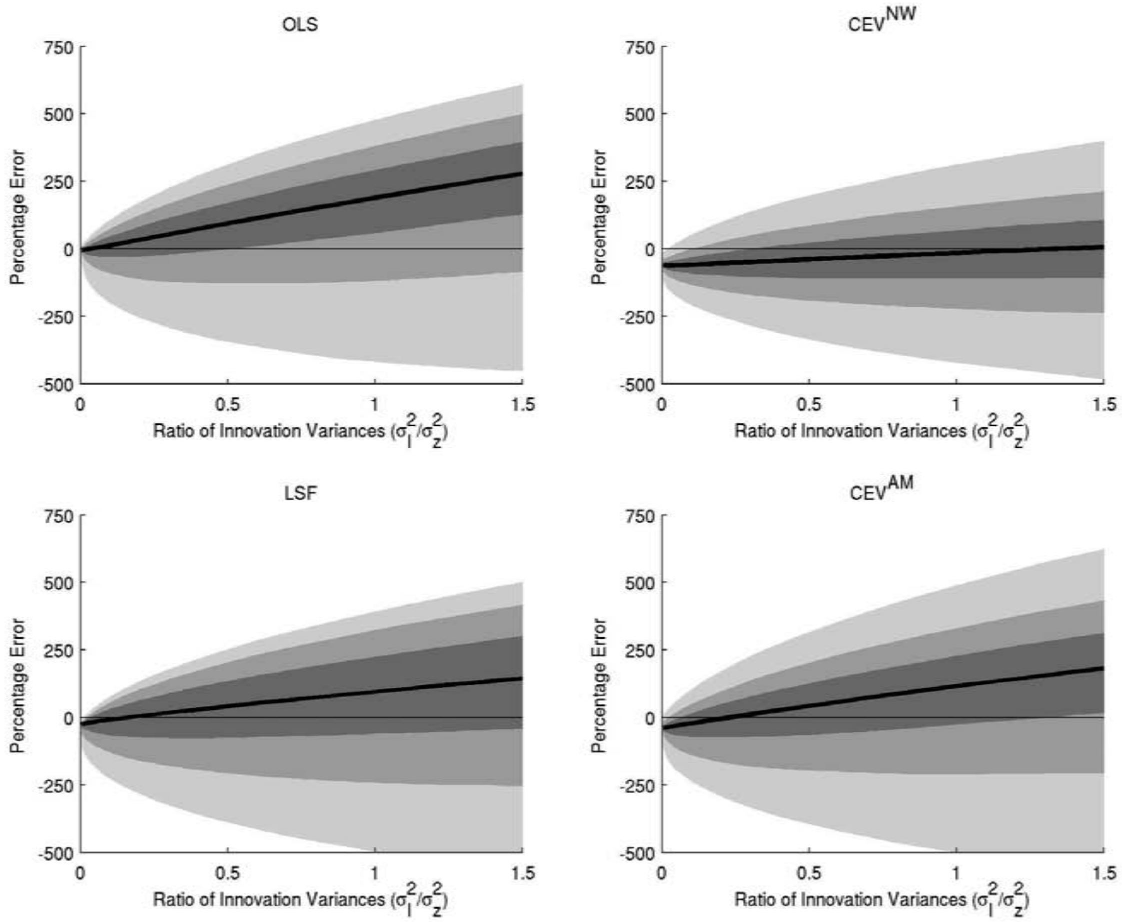
Note: Population values and medians of simulated variances of hours, $\mathrm{Var}(l_t)$. 10,000 simulations with $\rho_l = 0.986$. Population values are computed from the true autocovariances and denoted OLS_0, CEV_0^{NW} and CEV_0^{AM}. The latter two use the truncated and weighted summation of equation (20) with $b = 150$. (Please recall that in population, OLS variance equals the true value by construction.)

Figure 4: Bias in Hours Impact



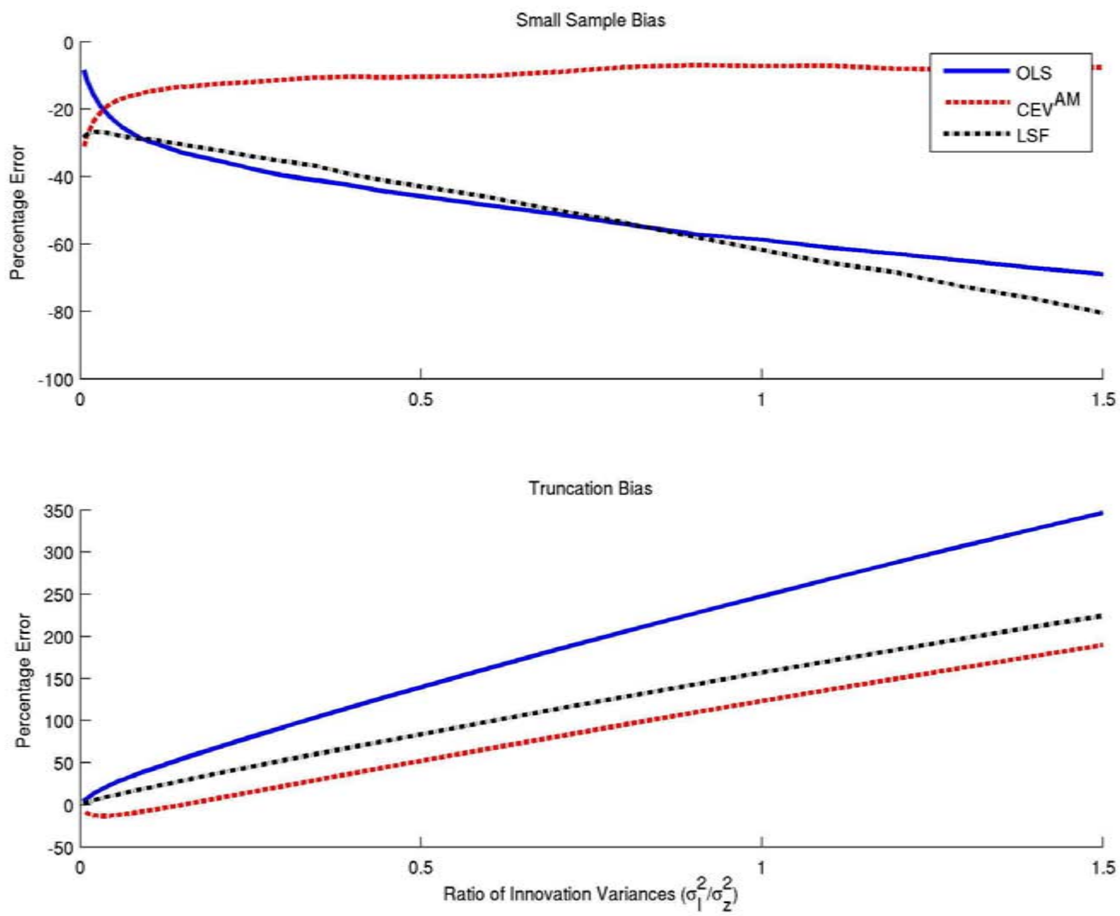
Note: Median percentage errors of the impact response of hours to a technology shock, a_{21}, over 10,000 simulations with $\rho_l = 0.986$. Spectral estimates using $b = 150$. Dashed vertical line denotes CEV's preferred calibration with $\sigma_l^2/\sigma_z^2 = 0.345$.

Figure 5: Distribution of Estimated Impact Coefficients



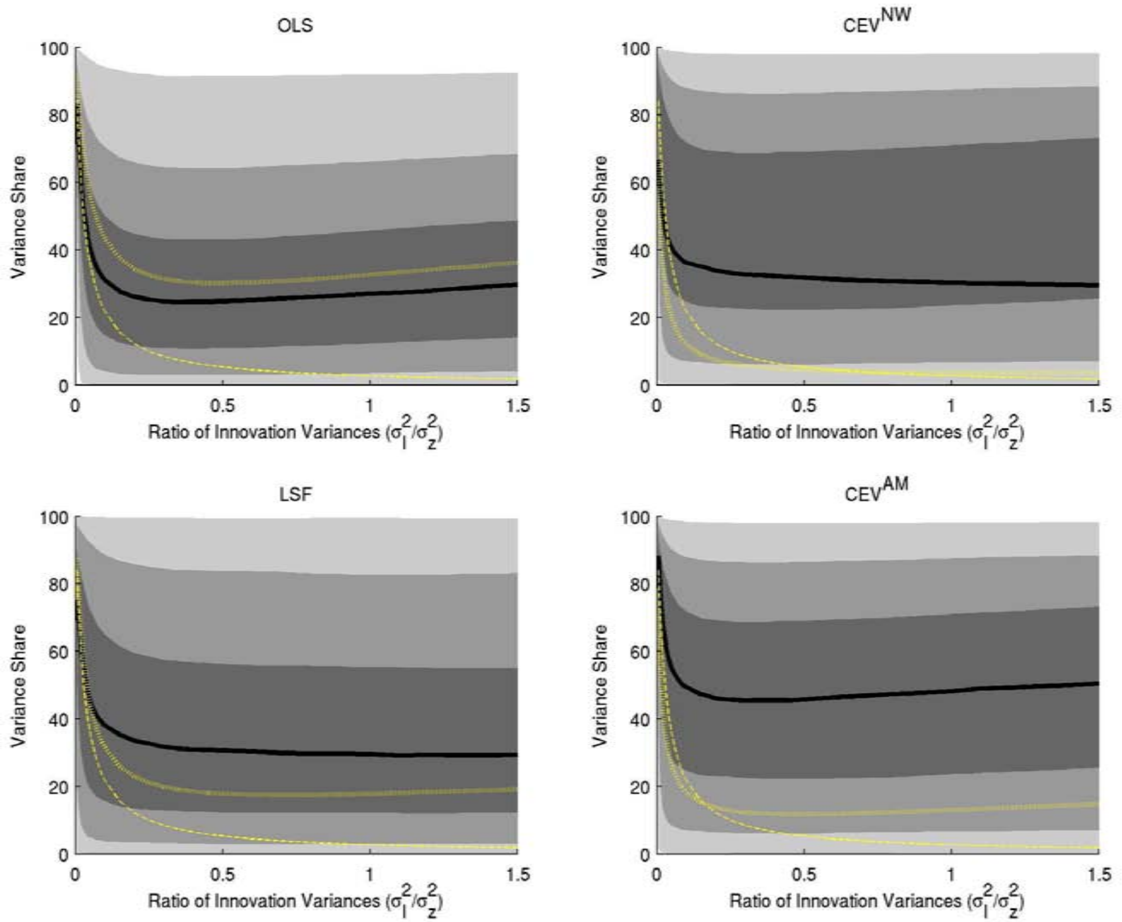
Note: Distribution of percentage errors in the impact response of hours to a technology shock, a_{21}, over 10,000 simulations with $\rho_l = 0.986$. Spectral estimates using $b = 150$. Percentiles are shaded: 95% (light), 68% (middle) and 38% (very dark). Black line is median of bootstraps.

Figure 6: Decomposed Bias in Hours Impact



Note: Small sample and truncation bias in the median percentage errors of the hours impact to technology, a_{21}. Small sample bias computed from $(\widehat{A_0^{\mathrm{OLS}}} - A_0^{\mathrm{OLS}})/A_0$ and truncation bias is $(A_0^{\mathrm{OLS}} - A_0)/A_0$, where $\widehat{A_0^{\mathrm{OLS}}}$ is the median of the OLS estimates over 10,000 simulation with $\rho_l = 0.986$. CEV computed from Andrews-Monahan spectral estimate with bandwidth $b = 150$. (For Newey-West, please see Figure 4 in the webappendix.)

Figure 7: Variance Decompositions



Note: Share of innovation variance in hours explained by technology shocks $a_{21}^2/(a_{21}^2 + a_{22}^2)$. Simulated estimates using 10,000 draws and $\rho_l = 0.986$. Spectral estimates using $b = 150$. Percentiles are shaded: 95% (light), 68% (middle) and 38% (very dark). Black line is median of bootstraps. Yellow dotted line is population value (computed from applying estimation procedures to true autocovariances of model) and yellow dashed line is true value.