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

Elmar Mertens[1]

Board of Governors of the Federal Reserve System, Washington D.C. 20551

Abstract

No, not really. In response to concerns about the reliability of SVARs, one proposal has been to combine OLS estimates of a VAR with non-parametric estimates of the spectral density. But as shown here, spectral estimators are no panacea for implementing long-run restrictions. They can suffer from small sample and misspecification biases just as VARs do. As a novelty, this paper uses a spectral factorization to ensure a correct representation of the data's variance. But this cannot overcome the basic small sample issues, which arise when trying to estimate long-run properties from relatively short samples of time-series data.

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

1. Introduction

In the past, 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". But OLS estimates of VAR coefficients are concerned with minimizing forecast error variance, not estimating the long-run variance. This has motivated Christiano et al. (2006a, 2006b, henceforth "CEV") to propose a new way of estimating structural VARs using a combination of OLS and a non-parametric estimator. They argue that their estimator virtually eliminates the bias associated with the standard OLS estimator. This paper shows that non-parametric estimates of the spectral density, henceforth called "spectral estimators", are no panacea for the implementation of long-run restrictions in finite sample.

Macroeconomic time series display a fair amount of persistence, posing two serious challenges for long-run identification. First, an accurate representation of the true model typically requires a VAR with a high lag order, much higher than what is affordable in a sample of typical length and resulting in a sizable truncation bias — discussed for example by Chari et al. (2008, henceforth "CKM") and Ravenna (2007). Second, there is the small sample bias in estimated coefficients known from Hurwicz (1950) and also discussed by CKM, which becomes ever more

Email address: elmar.mertens@frb.gov (Elmar Mertens)

[1]The views in this paper do not necessarily represent the views of the Federal Reserve Board, or any other person in the Federal Reserve System or the Federal Open Market Committee. Any errors or omissions should be regarded as solely those of the author.

severe the smaller the sample, and the more persistent the data. As will be shown, both issues affect not only VARs in the time domain, but also spectral estimators in the frequency domain.

The conventional VAR technique as well as different combinations with spectral estimators are evaluated here in the context of a simple two-shock RBC model, which has also been used by CEV and CKM. When using the various procedures to estimate the response of hours to technology, or to decompose the variance of fluctuations in output or hours, none of the procedures clearly dominates the others.

Furthermore, CEV do not consider some conceptual pitfalls in combining VAR coefficients with spectral estimates. Non-parametric estimates of the spectral density allow for non-*iid* residuals in the finite-order VAR, which is good since the underlying model is likely of infinite order. In what may be called "mixing and matching", the CEV approach plugs these estimates into the standard VAR formula alongside with coefficients from the finite-order VAR. This approach uses the extra information about omitted lags in the VAR to compute the long-run responses of variables to shocks—but not when mapping these back into impact responses. To retain a consistent representation of the data, that would however be necessary. Otherwise, the total variance of the data is misrepresented. In the simulations reported here, this misrepresentation is quantitatively relevant.

The CEV framework is amended here by recognizing that the non-parametric estimate contains information about omitted lags in the VAR. Truncation bias due to omitted lags has been stressed by CKM, Erceg et al. (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. But, echoing the results of Kascha and Mertens (2009) for SVARMAs, this corrected procedure suffers from the same basic problems as the other long-run identification methods: truncation and small sample bias.

The remainder of this paper is structured as follows: Section 2 describes the model economy against which the various estimation routines will be evaluated. Section 3 describes the various SVAR methods, including a new spectral factorization procedure. Section 4 presents the Monte Carlo results and Section 5 concludes the paper.

2. A Model Economy

This section describes a simple model economy, which will be used to illustrate and quantify the issues associated with various long-run identification schemes. None of the conceptual concerns related to spectral estimates raised in Section 3 will be specific to this setup. The model is the same one-sector RBC economy which has been used by CEV and CKM. Fluctuations are driven by two shocks: First, a unit root shock to technology, z_t. This is the permanent shock to be estimated by the VAR. Second, a transitory non-technology shock, τ_{lt}, which drives a wedge between private household's labor-consumption decision. The calibration is identical to the baseline model of CEV, using parameter values familiar from the business cycle literature. A detailed description of the model and its calibration is given in Appendix A.

Model simulations are performed for different calibrations of the share of output fluctuations explained by technology shocks, henceforth called the "technology share".[2] As a benchmark,

[2]The technology share is affected by varying the relative variances of transitory to permanent shocks. CKM exten-

maximum-likelihood estimates of CEV obtained from fitting the model to U.S. post-war data imply that around two-thirds of the bandpass-filtered variance in output can be attributed to technology shocks.[3]

Data is simulated for samples of length $T = 180$, corresponding to 45 years of quarterly data; identical to the simulations of CEV and CKM. Bivariate VARs are estimated using simulated data of the (log) growth rate of labor productivity and hours worked. For the results reported here, lag length is chosen by minimizing the Schwartz Information Criterion (SIC) in each simulation, typically picking small values close to one. However, results are insensitive to using other information criteria, such as the Akaike criterion (AIC). In general, AIC is known for picking higher values of p compared to SIC.[4] For brevity I have chosen to report only SIC results in the main paper. Results based on AIC lag-selection are available in a separate web-appendix. If anything, using the SIC criterion should give spectral estimators some extra bite in uncovering serial correlation in VAR residuals. In the spirit of the Andrews-Monahan estimator, the VAR would then be used for pre-whitening the data, without necessarily pruning all of its serial correlation. When computing population moments of the truncated procedures, a VAR(1) is used, but results are insensitive to using lag choices which are somewhat larger, say $p = 4$ or 8, but still typical in practical applications using quarterly data. For each calibration, 1,000 samples are simulated.

When looking at data simulated from this model, two statistics are of particular interest: How do hours worked respond to a technology shock? What is the share of fluctuations due to technology shocks? These questions are typically asked by empirical researchers trying to evaluate predictions from business cycle models with SVARs, such as Galì (1999) or Christiano et al. (2004).

3. Long-Run Identification in VARs

The linearized solution to the model described in the previous section is only one example from a wider class of linear dynamic models to which the SVAR methods discussed here can be applied. None of the issues discussed in this section will be specific to the model described above. An economic model from this wider class of dynamic models is supposed to specify a VAR representation for a stationary vector of observable variables: $X_t = B(L)X_{t-1} + e_t$ 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*, $e_t \sim iid(0, \Omega)$.[5] For future reference, it will be handy to denote the non-structural moving average (VMA) coefficients of $X_t = C(L)e_t$ where $C(L) = (I - B(L)L)^{-1} = \sum_{k=0}^{\infty} C_k L^k$ and $C_0 = I$.

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

sively document how different ratios in the variance of transitory to permanent shocks, affect the performance of standard VARs both in population and in small sample. When the relative variance of transitory shocks goes to zero, the true responses can be recovered by a finite order VAR in productivity growth and hours—though the true system does not have a finite-order VAR representation McGrattan (2005).

[3] As elsewhere in the literature, The bandpass filter employed throughout this paper considers only fluctuations with durations between two-and-a-half and eight years.

[4] For this lab economy, AIC has been found to pick lag lengths of up to $p = 6$ with an average of $p = 2$.

[5] For notational convenience, but without loss of generality, X_t represents the demeaned variables, which is equivalent to including a constant in a VAR using the original data.

model described in Section 2 has such an infinite order VAR representation; details are shown in Appendix C.

For the identification of structural shocks, there must exist an invertible one-to-one mapping from innovations e_t to the structural shocks ε_t driving the underlying model $e_t = A_0 \, \varepsilon_t$ where A_0 is square and $|A_0| \neq 0$. This paper considers only cases where such an invertible representation exists, though possibly only in the form of an infinite order VAR.[6] Excluding the complications arising from non-invertibilities allows to focus on problems related to small sample bias and the finite order approximations of the VAR.

In the spirit of CEV and CKM, only one of the structural shocks will be identified. 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 (Galì, 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 matrix of long-run coefficients, $A(1) = \sum_{i=0}^{\infty} A_i$:

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

This restriction holds exactly in the linearized solution to the model described in Section 2.

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

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

and $A(1)$ factors the spectral density of X_t at frequency zero,

$$A(1)A(1)^T = S_X(0) \; .$$

CEV show that the restriction in (1) uniquely pins down the first columns of A_0 and $A(1)$, and the latter can be computed from the Cholesky decomposition of $S_X(0)$.

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 $B(L)$:

$$A_0 = (I - B(1)) \, A(1) \tag{2}$$

3.1. OLS Implementation with Finite-Order VAR

Since the innovations of the true VAR representation are assumed to be white noise, they satisfy the OLS normal equations $EX_{t-k}e_t^T = 0 \; (\forall \, k)$. In principle, the coefficients B_k could be estimated from least squares projections of X_t on its infinite past. In practice, an empirical implementation can only work with a finite lag length. Henceforth $B(L)^{\text{OLS}}$ denotes a lag polynomial of finite order.[7] The residuals from such a finite-order VAR will be denoted

$$v_t^{\text{OLS}} \equiv X_t - B(L)^{\text{OLS}} X_{t-1} \tag{3}$$

[6] Fernàndez-Villaverde et al. (2007) derive conditions when a linear dynamic model has an invertible VAR representation, which are also summarized in Appendix C.

[7] The OLS residuals are identfied from the normal equations of the VAR, $EX_{t-k}(v_t^{\text{OLS}})^T = 0$ for all lags $k \leq p$.

with $\Omega_v^{\text{OLS}} = E\, v_t^{\text{OLS}}(v_t^{\text{OLS}})^T$. The associated VMA is $C(L)^{\text{OLS}} \equiv (I - B(L)^{\text{OLS}}L)^{-1}$. Only stable VARs are considered.

Since Blanchard and Quah (1989) the standard procedure is to proceed as if the residuals v_t^{OLS} were uncorrelated. For the long-run restriction (1) the spectral density at frequency zero can then be constructed from the OLS estimates, $S_X(0)^{\text{OLS}} = C(1)^{\text{OLS}}\, \Omega_v^{\text{OLS}}\, \left(C(1)^{\text{OLS}}\right)^T$. Impact coefficients are then computed by plugging these estimates into (2):

$$A_0^{\text{OLS}} = \left(I - B(1)^{\text{OLS}}\right) \text{chol}\left(S_X(0)^{\text{OLS}}\right) \tag{4}$$

But when the data has been generated from an infinite order process, a truncation bias arises from using a finite-order VAR. In this case, the OLS assumption of uncorrelated forecast errors v_t^{OLS} is violated, which is an example of what Cooley and Dywer (1998) criticized as an "auxiliary" (but not innocuous) assumption. This truncation bias arises even when the true population moments of the data generating process were known. Applied to data generated from a business cycle model the truncation bias in SVARs can be substantial, as shown by Cooley and Dywer (1998), Erceg et al. (2005), Ravenna (2007) or CKM. The truncation bias is also sizable for data from the model described above as will be seen in Figure 2 below.

3.2. CEV: Combining OLS 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 non-parametric estimator for $S_X(0)$. The procedure is motivated by observing that OLS projections construct $B(L)^{\text{OLS}}$ not necessarily with regard to $B(1)$ but in order to minimize the forecast error variance Ω_v^{OLS}. As argued by Sims (1972), the least-squares objective of OLS seeks coefficients which 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 the zero frequency. Accordingly, 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 and $S_X(0)^{\text{OLS}}$ need not be the best possible estimate for the spectral density at frequency zero.

Instead of using $S_x(0)^{\text{OLS}}$, CEV employ a spectral estimator of $S_X(0)$ to construct $A(1)$. In Christiano et al. (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 sums over the (sample) autocovariances of X_t, Andrews-Monahan uses first the VAR to prewhiten the data and then sums over the *residual* autocovariances:

$$\begin{aligned} S_X(0)^{\text{NW}} &= \sum_{k=-b}^{b}\left(1 - \frac{|k|}{b+1}\right) E\left[X_t X_{t-k}^T\right] \\ S_X(0)^{\text{AM}} &= C(1)^{\text{OLS}} S_v^{NW}(0) \left(C(1)^{\text{OLS}}\right)^T \\ \text{where} \quad S_v(0)^{\text{NW}} &= \sum_{k=-b}^{b}\left(1 - \frac{|k|}{b+1}\right) E\left[v_t^{\text{OLS}}(v_{t-k}^{\text{OLS}})^T\right] \end{aligned} \tag{5}$$

The truncation parameter b is also known as "bandwidth", and its selection will be discussed in more detail below. The Andrews-Monahan estimator nests the OLS case when $b = 0$.[8]

[8] As elsewhere in this section, estimators have been written in terms of population moments. In empirical applications, the population moments are supposed to be replaced by sample moments, and this is also done in the simulation reported below.

The new CEV estimator computes the long-run coefficients from the non-parametric density estimate. Combined with the OLS lag coefficients, CEV obtain their impact coefficients as $A_0^{\text{CEV-AM}} = \left(I - B(1)^{\text{OLS}}\right) A(1)^{\text{AM}}$ where $A(1)^{\text{AM}} = \text{chol}\left(S_X(0)^{\text{AM}}\right)$; and the impulse responses are $A(L)^{\text{CEV-AM}} = C(L)^{\text{OLS}} A_0^{\text{CEV-AM}}$.

Theoretically, the prewhitening of Andrews-Monahan is appealing since it removes spikes from the spectral density of X_t which make spectral estimation difficult. Andrews and Monahan (1992) and Newey and West (1994) find the prewhitening to fare better in Monte Carlo studies than the original Newey-West estimator. CEV find no clearly superior choice between the two in their 2006a paper and their 2006b study proceeds to use only the Newey-West estimator. For brevity, the remainder of this section will mostly refer to the Andrews-Monahan estimator, with similar arguments holding for the Newey-West estimator.[9] Section 4 presents simulations using both estimators.

Choosing the bandwidth b is critical in estimating spectra, akin to choosing the lag order of a VAR. Bandwidth choice has been shown to be more important than using other weighting schemes as the Bartlett kernel (Newey and West, 1994).[10]

The bandwidth selection schemes of Andrews (1991) and Newey and West (1994) minimize the mean-squared error (MSE) of estimated spectra, but an MSE optimal estimator of the spectrum does not necessarily translate into an MSE optimal estimate of coefficients like $A_0^{\text{CEV-AM}}$ or $A_0^{\text{CEV-NW}}$. Their MSE depends not solely on the MSE of $S_X(0)^{\text{AM}}$ but—amongst others—on bias and standard error of the spectrum in ways which are specific to the data generating process.[11]

The simulations reported below use the large bandwidth choice of CEV ($b = 150$). A separate web-appendix documents results derived from the optimal bandwidth selection scheme of Newey and West (1994), which tends to pick fairly small bandwidths.[12] In the case of the Andrew-Monahan estimator, this is a natural result, since even the conservative SIC lag-length selection of the VAR has pruned most of the serial correlation of the data. Consequently, automatic bandwidth selection with the Andrews-Monahan estimator does not deliver results much different from OLS. In simulations not reported here, spectral estimates with intermediate bandwidth choices displayed performance characteristics which were intermediate between what is shown here and the web-appendix.

3.3. Identification with Spectral Factorization

In an earlier working paper version, Mertens (2011), I document two shortcomings of the CEV procedure: First, the impact coefficients of CEV do not reproduce the forecast error variance of the VAR, $A_0^{\text{CEV-AM}}(A_0^{\text{CEV-AM}})^T \neq \Omega_v^{\text{OLS}}$. Consequently, their impulse responses do not match the unconditional variance of the data either. Secondly, when estimates of the realized technology shocks based on $A_0^{\text{CEV-AM}}$ are perfectly correlated with the conventional, OLS-based, estimates.

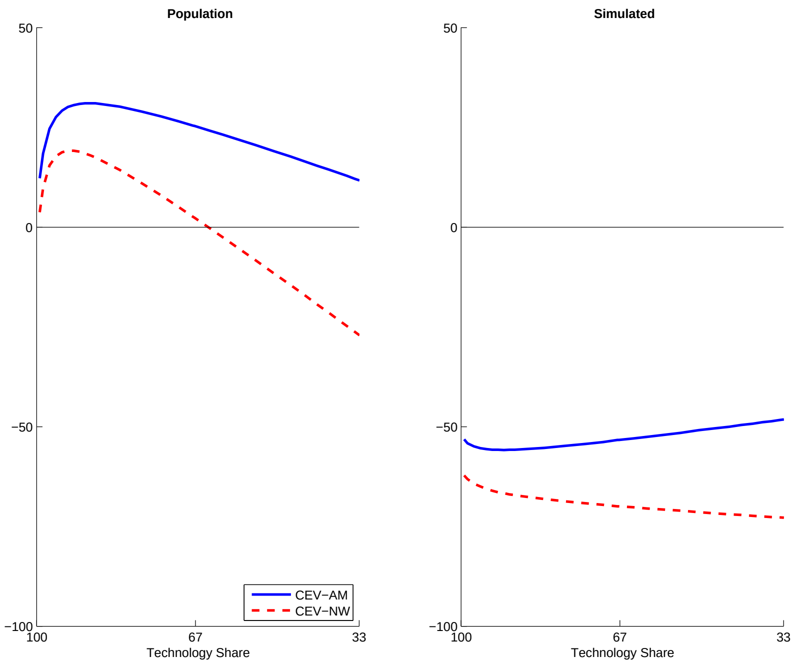
[9] Using the Newey-West estimator, impact coefficients are $A_0^{\text{CEV-NW}} = \left(I - B(1)^{\text{OLS}}\right) \text{chol}\left(S_X(0)^{\text{NW}}\right)$.

[10] Alternative weighting schemes are for example discussed by Phillips et al. (2006).

[11] To be specific, $A_0^{\text{CEV-AM}}$ and $A_0^{\text{CEV-NW}}$ are functions of a spectral estimate and OLS estimates of the VAR. Analogously to arguments employed by Sun et al. (2008) in the context of constructing confidence intervals, the MSEs of $A_0^{\text{CEV-AM}}$ and $A_0^{\text{CEV-NW}}$ could be approximated to a second order—holding the OLS estimates fixed—by a linear combination of bias and variance of the spectral estimate whose weights depend on the true values of the VAR. Furthermore, the MSEs of $A_0^{\text{CEV-AM}}$ and $A_0^{\text{CEV-NW}}$ will also depend on the covariance between $S_X(0)^{\text{AM}}$, respectively $S_X(0)^{\text{NW}}$, and the OLS estimates.

[12] For the various calibrations of the model economy considered here, the optimal bandwidth of the Newey-West estimator is typically close to ten, and it is close to four for the Andrews-Monahan estimator.

Figure 1: Percentage Errors in Variance of Output Growth Computed from CEV Procedure



Note: Percentage errors of (unfiltered) variance in output growth, relative to true moments (left panel, "Population") or OLS small sample moments from mean of from 1,000 simulations (right panel, "Simulated"). Errors are computed for different technology shares in true model. Variances denoted "CEV-NW" and "CEV-AM" are computed from the impulse responses of CEV's SVAR procedure using the Newey-West or Andrews-Monahan estimators respectively.

The variance mismatch is illustrated in Figure 1 for the variance of output growth in the model economy described in Section 2. For this figure, the CEV procedure is applied to population moments as well as to moments from small sample simulations. In small sample, the variances implied by the CEV procedure are only about half as big as the OLS sample moments. As can be seen in the right panel of Figure 1, this occurs both when using the Newey-West or the Andrews-Monahan variant of the procedure and regardless of the share of fluctuations explained by technology shocks. As depicted in the left panel of the figure, the mismatch is qualitatively different, but also sizable when applying the procedure to the population moments of the model while using a lag length of $p = 1$ and spectral bandwidth of $b = 150$.

These issues can easily be resolved by using a spectral factorization procedure, which transforms non-parametric estimates of the spectral density into a VMA for the VAR residuals. In general — when the true VAR representation is of infinite order — the residuals v_t^{OLS} are not *iid*, instead they follow a moving average representation,

$$v_t^{\text{OLS}} = e_t + D_1 e_{t-1} + D_2 e_{t-2} + \ldots = D(L) e_t$$

where $D(L) = (I - B(L)^{\text{OLS}} L) C(L)$ and with spectral density

$$S_v(\omega) = D(e^{-\omega}) \Omega D(e^{-\omega})^T \tag{6}$$

CKM and CEV discuss a truncation bias which is hard to detect based on VAR lag-length selection procedures. In terms of the moving average $D(L)$, their results can be read as finding $D_i \approx 0$ but $D(1) \neq I$, which is further illustrated in Appendix A.

For a given estimate of the spectral density of the residuals, such as the one used in the Andrews-Monahan estimator (5), a spectral factorization yields $D(L)$ as the only moving average representation, which is invertible and matches the spectrum of the residuals.[13] The computations reported here use a reliable and efficient algorithm from Li (2005). Details are described in Appendix B.

The true impact coefficients (2) and VMA can be written in terms of $B(L)^{\text{OLS}}$ and $D(L)$ as

$$A(1) = \text{chol}\left((I - B(1)^{\text{OLS}})^{-1} S_v(0) (I - B(1)^{\text{OLS}})^{-T} \right) \tag{7}$$

$$A_0 = D(1)^{-1} (I - B(1)^{\text{OLS}}) A(1) \tag{8}$$

$$C(L) = (I - B(L)^{\text{OLS}}) D(L) \tag{9}$$

As usual, impulse response can be computed from the product of $C(L)$ and A_0.

CEV construct $A(1)^{\text{AM}}$ according to (7) while using the spectral estimate $S_v(0)^{\text{NW}}$. But they ignore the residual dynamics captured by $D(1)$ in (8) when mapping $A(1)^{\text{AM}}$ back into the impact coefficients. Ignoring the residual captured by $D(L)$ is the source of the variance misrepresentation discussed in the previous subsection.

To combine VAR coefficients and spectral estimates in an internally consistent fashion, a spectral factorization must be used. The spectral factorization of $S_v(\omega)^{\text{NW}}$ yields a unique and invertible MA(b), denoted $D(L)^{\text{SF-AM}}$, and an innovations variance matrix $\Omega^{\text{SF-AM}}$. The superscript "SF-AM" indicates that these are calculated from the residual spectrum employed by the Andrews-Monahan estimator $S_X(\omega)^{\text{AM}}$. VMA and impact coefficients and are then

$$C(L)^{\text{SF-AM}} = (I - B(L)^{\text{OLS}}) D(L)^{\text{SF-AM}}$$

$$A_0^{\text{SF-AM}} = \left(D(1)^{\text{SF-AM}} \right)^{-1} A_0^{\text{CEV-AM}}$$

By construction, the spectral factorization is consistent with the variance of the data, in sample as well as in population, such that $A_0^{\text{SF-AM}}$ and $D(L)^{\text{SF-AM}}$ reproduce the variance of the VAR residuals as well as the unconditional variance of X_t.[14]

A spectral factorization can also be applied directly to the Newey-West estimate of the data's spectrum, $S_X(\omega)^{\text{NW}}$, yielding coefficients for the VMA of X_t, $C(L)^{\text{SF-NW}}$ and innovation variance $\Omega^{\text{SF-NW}}$. Following (2), impact coefficients can then be computed as

$$A_0^{\text{SF-NW}} = \left(C(1)^{\text{SF-NW}} \right)^{-1} A(1)^{\text{NW}}$$

The impulse responses $C(L)^{\text{SF-NW}} A_0^{\text{SF-NW}}$ do not involve any VAR coefficients.[15]

[13] Invertibility of the true $D(L)$ follows from noting that $(I - B(L))^{-1} = C(L) = (I - B(L)^{\text{OLS}})^{-1} D(L)$ has all roots outside the unit circle and the same has been assumed for the VMA of the VAR(p), $C(L)^{\text{OLS}} = (I - B(L)^{\text{OLS}} L)^{-1}$.

[14] The variance of the VAR residuals can be written as $\Omega_v^{\text{OLS}} = \int_{-\pi}^{\pi} S_v(\omega)^{\text{NW}} d\omega$. And the rest follows from $S_v(\omega)^{\text{NW}} = \Gamma_0 + \sum_{k=1}^{b} \left(1 - \frac{|k|}{b+1}\right) \left(\Gamma_k e^{-i\omega k} + (\Gamma_k)^T e^{i\omega k} \right)$ and $\int_{-\pi}^{\pi} e^{-i\omega k} d\omega = 0$.

[15] In a similar spirit, Dupor and Kiefer (2007) construct impulse responses independently from a formal VAR representation, based on local projections as proposed by Jordá (2005).

4. SVARs applied to Data from Lab Economy

The previous section described several schemes for imposing the long-run restriction (1) on the data. The conventional method, going back to Blanchard and Quah (1989), uses OLS estimates of a VAR. The recently proposed procedure of CEV combines this with a non-parametric estimate of the spectral density at frequency zero. This procedure has been criticized above for its lack of internal consistency. Finally, a new method has been proposed, which provides an internally consistent combination of OLS and spectral estimators. This method relies on a spectral factorization ("SF") to uncover the dynamics implied by the non-parametric spectral estimators.

These procedures are applied here to data simulated from the model economy described in Section 2. The same data generating process has also been used by CEV and CKM.[16] For the CEV and SF methods, there are two variants depending on whether the spectral estimators of Newey and West (1987) or Andrews and Monahan (1992) are used. This section reports results for both.

Mimicking conditions faced by empirical researchers, "small" samples with 180 observations are simulated. In small sample, two distinct issues arise. First, there is truncation bias in VARs and spectral estimators arising from the need to specify a finite lag length p, respectively a finite bandwidth b. As discussed in Section 2, lag length is determined individually for each draw with an information criterion and spectral bandwidth is fixed at 150. (A separate web-appendix shows results based on automatic bandwidth selection procedures.[17])

Second, there is the small sample bias in estimated parameters known from Hurwicz (1950).[18] To isolate the pure truncation effects from the Hurwicz bias, the identification procedures are not only applied to simulated data, but also to VARs and spectral estimates constructed from the model's true population moments. To distinguish the truncation bias from sampling uncertainty, population moments are computed for fixed VAR lag-length and bandwidth choice of the spectral estimators. [19]

The procedures are evaluated in terms of of their capability to uncover two statistics typically of interest to applied researchers. Following CKM and CEV, estimated impact responses of hours to a technology shock are computed. In addition, the share of fluctuations in output and hours due to technology shocks is estimated. These variance shares also reflect estimation of the entire impulse responses, not only the impact coefficients, derived from each method. As it is typical in the business cycle literature, these shares are computed after filtering out any fluctuations which do not correspond to cycles with a duration between two-and-a-half and eight years; details are

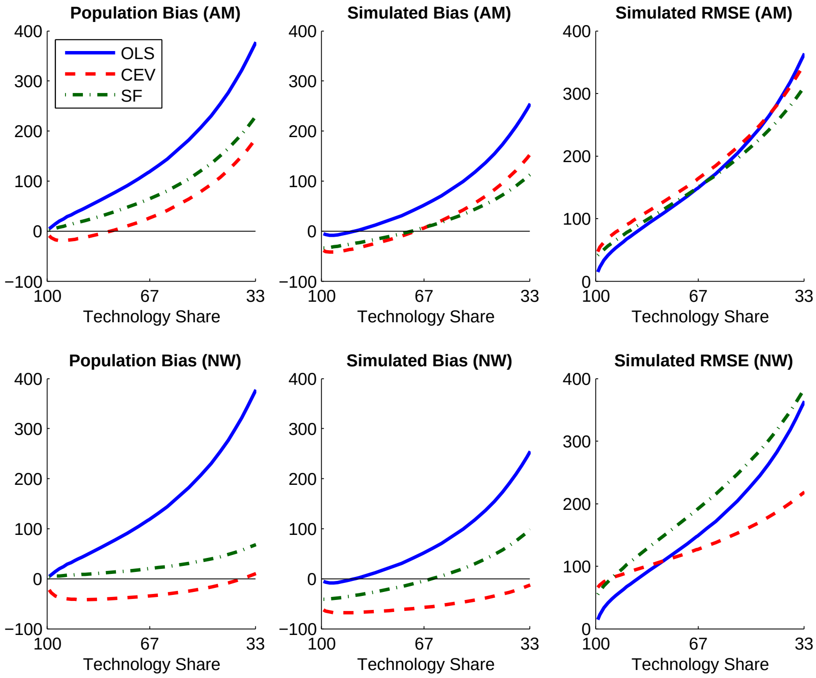
[16] As discussed in Section 2, the calibrations employed here are identical to the setting of CEV—except for considering a wider range of the technology share in output fluctuations. As discussed above as well, the CKM experiments differ slightly in their choice of ψ and ρ_l.

[17] Automatic bandwidth selection for the Andrews-Monahan estimator yields bandwidths close to zero, such that the results are mostly indistinguishable from the OLS estimates. For Newey-West spectra with optimal bandwidth selection results are qualitatively similar to what is shown here.

[18] This bias is particularly acute the smaller the sample and the higher the persistence of the data. It is pertinent in this example, since calibrating the model to match salient features of U.S. data requires a high degree of persistence in the non-technology shock, ρ_l.

[19] In the case of the spectral estimators, this means evaluating their truncated sums at true population moments, instead of sample autocovariances, while keeping the Bartlett weights and the truncation at the chosen bandwidth. The computation of VARs from population moments is described in Appendix C. This does not mean, that low-order lags and bandwidth choices should be made, when sufficient data were available, such that sample moments could arbitrarily well converge to population, and when higher-order VAR lags and spectral bandwidths could (and should) be implemented. The point is merely to isolate truncation bias from sampling uncertainty.

Figure 2: Impact Response of Hours to Technology Shock



Note: Percentage points relative the model's true impact response of hours to a technology shock. Top row based on Andrews-Monahan estimator for CEV and SF, bottom row using Newey-West estimators; both with a fixed bandwidth of $b = 150$. "Technology share" on the x-axis is the true percentage of bandpass-filtered output variability due to technology shocks in the model economy.

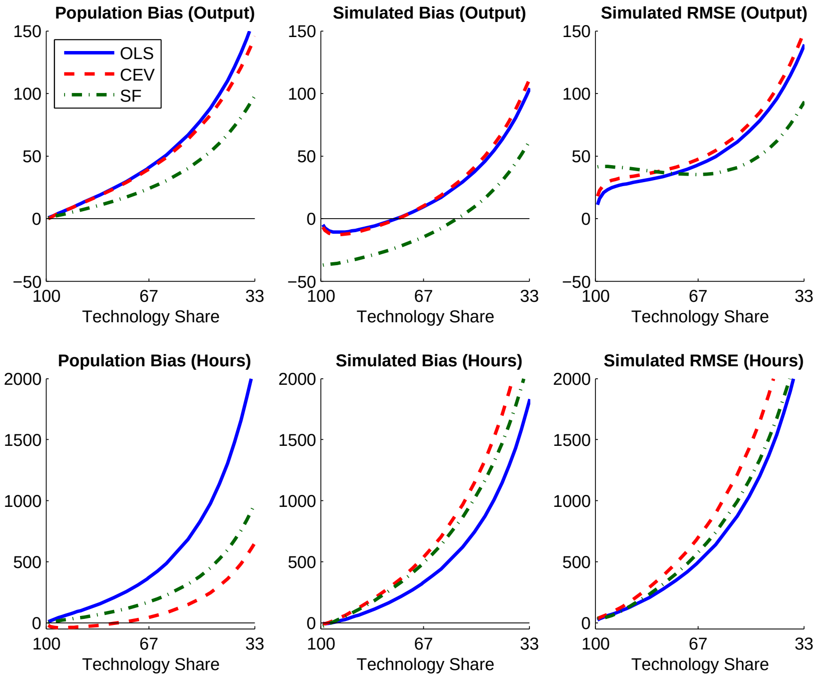
described in Appendix D. Two criteria are reported to assess the goodness of estimates: Bias and Root Mean Square Error (RMSE), both expressed as percentages relative to the true value known from the model.[20]

The results show that all procedures are subject to substantial truncation and small biases and none works like a panacea. Different methods display different strengths and weaknesses. The claims by Christiano et al. (2006a) of "smaller bias, smaller means square error" associated with their procedure do neither generalize to a wider range of model calibrations nor do they extend from the estimation of impact responses to variance shares.

Effects from the truncation and the small sample bias can offset each other. This is the case when estimating the impact of technology on hours. The left column in Figure 2 shows how impact responses are overestimated in population whereas the simulated bias shown in the middle column of the figure is lower (more negative). This simulated bias displays the total

[20]Denoting the estimated parameter as θ and its estimate as $\hat{\theta}$, relative bias is computed as $E(\hat{\theta}-\theta)/\theta \cdot 100\%$. Similarly, relative RMSE is computed from RMSE$/\theta \cdot 100\%$. In both cases, expectations are computed from the arithmetic average over 1,000 simulations.

Figure 3: Technology Shares in Fluctuations of Output and Hours



Note: Percentage points relative the model's true technology share. Top row reports bias and RMSE for variance decomposition of output, bottom row for hours. Andrews-Monahan estimators of the spectral density used for CEV and SF with a fixed bandwidth of $b = 150$. "Technology share" on the x-axis is the true percentage of bandpass-filtered output variability due to technology shocks in the model economy.

effect from truncation and Hurwicz bias. The OLS method has the largest population bias and it is only partially offset by the Hurwicz bias. The two spectral methods suffer from substantially smaller truncation bias, and depending on the simulated importance of technology shocks, the total bias can be either negative or positive. Incidentally, the upwards bias in SF-AM and SF-NW is exactly offset around technology shares of about two thirds, corresponding to the range of MLE estimates of CEV and CKM for U.S. data. (Similarly for CEV-AM, but not CEV-NW.) However, results are different for other calibrations of the technology share, which cautions strongly against extrapolating from a particular result to different data sets and different applications.

Unless the true share of technology shocks is very large, the RMSE of estimated impact coefficients are very large, often surpassing more than 100% of the true value. Interestingly, the RMSE do not differ much across the different methods, as can be seen in the right-most column of Figure 2. If anything, SF-NW is outperforming CEV-NW on bias, at the expense of a worse RMSE. This is likely due to an overfitting of the residual dynamics by SF-NW.

Turning to the estimated variance shares of output and hours shown in Figure 3, the relative performance of the various methods looks quite different. The panels in the top row of the figure show bias and RMSE for variance decompositions of output, the bottom row for variance decompositions of hours. For this figure, spectral densities have been estimated with the Andrews-Monahan estimator. Results are broadly similar when using the Newey-West estimator (see the separate web-appendix).

Strikingly, for technology shares in output, bias and RMSE are very similar when using either OLS or CEV. The mismatch in total variance discussed in Section 3, does not seem to distort the computations of *relative* variance measures in this case. But, the two methods differ when decomposing the variance of hours. Bias and RMSE in the variance decomposition of hours are an order of magnitude larger than for output, cautioning very strongly against neglecting small sample issues when comparing SVAR estimates against model predictions. Moreover, the variance decompositions of hours provide a useful counterexample against disregarding OLS methods altogether, since OLS dominates the spectral methods both in terms of simulated bias and RMSE for all calibrations considered here. All in all, these results underline how truncation and Hurwicz bias interact with the different methods in ways which are hard to anticipate for an empirical researcher who does not know the true dynamics of the data.

5. Conclusions

In finite sample, truncation bias and Hurwicz bias pose fundamental problems when identifying structural shocks from restrictions on the long-run behavior of the data. These issues are present in the time domain when working with a VAR, as well as in the frequency domain when working with spectral estimators. Basically, the same estimates of the data's autocovariances are employed for constructing non-parametric estimates of the spectrum as well as for computing OLS coefficients. In both cases, truncation bias arises since there are only as many sample autocovariances as there are data points. And due to the Hurwicz bias, variance estimates tend to be biased downwards the smaller the sample and the larger the persistence of the data—again affecting both OLS estimates of VAR coefficients as well as non-parametric estimates of the spectral density. Thus, spectral estimates offer no panacea against the small sample problems known from OLS.

The performance of different estimators appears to be very specific to the underlying model and its calibration, making it hard to predict, which procedure would do well in future applications using new data. Even for a given calibration, when a method performs better in terms of one model statistic, say impact coefficients, this does not necessarily translate into better performance for another statistic, like a variance share. All in all, these results strongly suggest to compare SVAR estimates (from any procedure), against the *small sample* predictions, not the true moments, of a specific model as in Cogley and Nason (1995), Kehoe (2006) or Dupaigne and Feve (2009). Alternatively, full-information methods based on maximum-likelihood estimation or Bayesian methods — see for example Ireland (2004) or Smets and Wouters (2007) — could be used.

Acknowledgements

This paper is based on the second chapter of my dissertation written at the University of Lausanne. 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. In particular, I would like to thank V.V. Chari, Patrick Kehoe and Ellen McGrattan for discussions and sharing their computer codes. In addition I am grateful for discussions with Lawrence Christiano, Bill

Dupor, Mark Watson as well as seminar participants at the University of Basel, the University of Zurich, the Young Swiss Economist Meetings, and the members of my dissertation committee, Jean-Pierre Danthine, Philippe Bacchetta, Robert G. King, and Peter Kugler. All remaining errors are of course mine.

Appendix A. Model Economy

This appendix expands on the description of the model economy used in the paper . The model is a common one-sector RBC economy driven by two shocks: First, a unit root shock to technology, z_t. This is the permanent shock to be estimated by the VAR. Second, a transitory non-technology shock, τ_{lt}, which drives a wedge between private household's labor-consumption decision.[21]

The representative household maximizes his lifetime utility over (per-capita) consumption, c_t, and labor services, l_t, $E_0 \sum_{t=0}^{\infty}(\beta(1+\gamma))^t u(c_t, l_t)$, 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 $\beta < 1$).

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: $\log Z_t = \mu_z + \log Z_{t-1} + \sigma_z \varepsilon_t^Z$ and $\log \tau_{l,t+1} = (1-\rho_l)\bar{\tau}_l + \rho_l \log \tau_{l,t-1} + \sigma_l \varepsilon_t^l$, 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 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%.[22] Following CEV, the transitory shock is calibrated as an AR(1) with persistence $\rho_l = 0.986$. This calibration is identical to the values used by CKM except for their values of $\phi = 1.6$ and $\rho_l = 0.95$.

The model economy is calibrated over different ratios in the variance of transitory to permanent shocks, σ_l^2/σ_z^2, which translate into different assumptions about the share of output fluctuations explained by technology shocks.[23] As a benchmark, maximum-likelihood estimates

[21]The non-technology shock 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 et al. (2007) show how this labor "wedge" can be understood more generally as the *reduced form* process for more elaborate distortions, such as sticky wages.

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

[23]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 in the limit, $\sigma_l/\sigma_z \to 0$, a finite order VAR (even a $p = 1$) in productivity growth and hours recovers the true responses—though the true system does not have a finite-order VAR representation. In this special case the model reduces to a standard, one-shock RBC model.

of CEV obtained from fitting the model to U.S. post-war data imply that around two-thirds of the bandpass-filtered variance in output can be attributed to technology shocks.[24] The bandpass filter employed throughout this paper considers only fluctuations with durations between two-and-a-half and eight years, which is consistent with the NBER definitions of Burns and Mitchell (1946).

Data is simulated for samples of length $T = 180$, corresponding to 45 years of quarterly data; identical to the simulations of CKM and CEV. Following CEV and CKM, bivariate VARs are estimated using simulated data of the (log) growth rate of labor productivity and hours worked; $X_t = \left[\Delta \log (y_t/l_t) \quad \log l_t\right]^T$. For each simulated sample, the lag length of the VAR(p) is chosen by minimizing the Schwartz Information Criterion (SIC), typically picking small values close to one.[25] When computing population moments, a VAR(1) is used. For each calibration, 1,000 samples are simulated.

As argued in Section 3.3 of the paper, CKM and CEV discuss a truncation bias which is hard to detect based on VAR lag-length selection procedures. In terms of the moving average $D(L)$, their results can be read as finding $D_i \approx 0$ but $D(1) \neq I$. For the model economy describe above, Figure A.4 plots the population values of the cumulated sums $\sum_{k=0}^{K} D_k$ when $p = 1$ for different calibrations of the share of fluctuations in output explained by technology shocks. (Results are similar for other values of p.) At each lag, the increments are small and close to zero, but summing over many lags leads to $D(1) \neq I$.

Appendix B. Spectral Factorization Method

Spectral factorization has a long tradition in the fields of linear quadratic control, robust estimation and control as surveyed for example by Whittle (1996).[26] Below is a version of the spectral factorization theorem, which has been adapted from Hannan (1970, p. 66), which has has been slightly strengthened here by excluding the case of zero power in the spectral density at zero-frequency, to ensure the invertibility of the MA(b).[27]

Theorem 1 (Spectral Factorization, Hannan (1970))**.** *Suppose a variable v_t has autocovariances $\Gamma_k \equiv E\left[v_t v_{t-k}^T\right] = (\Gamma_{-k})^T$ and a spectral density $S_v(\omega) \equiv \sum_{k=-q}^{q} \Gamma_k e^{-ik\omega}\ \forall\ \omega \in [-\pi, \pi]$ which is non-singular at each frequency ($|S_v(\omega)| \neq 0\ \forall\ \omega$), as well as being non-zero at the zero frequency, $S_v(0) \neq 0$. There is a unique factorization of $S_v(\omega)$ into $S_v(\omega) = D(e^{-ik\omega})\,\Omega\, D(e^{-ik\omega})^T$ where Ω is positive definite and $D(z)$ is a q'th order polynomial $D(z) = I + \sum_{k=1}^{q} D_k z^k$ which has all its roots outside the unit circle.*

In the context of this paper, $S_v(\omega)$ will be the spectral density of $v_t^{\text{OLS}} = D(L)e_t$ where $E e_t e_t^T = \Omega = A_0 A_0^T$. Non-parametric estimates of $S_v(\omega)$ can be constrcuted from weighted sums of the *sample* autocovariance function as described in Section 3.2 of the main paper.[28]

---

[24] To be precise, CEV estimate $\sigma_l = 0.0056^2$, $\sigma_z = 0.00953$ corresponding to a technology share of 67.5%.

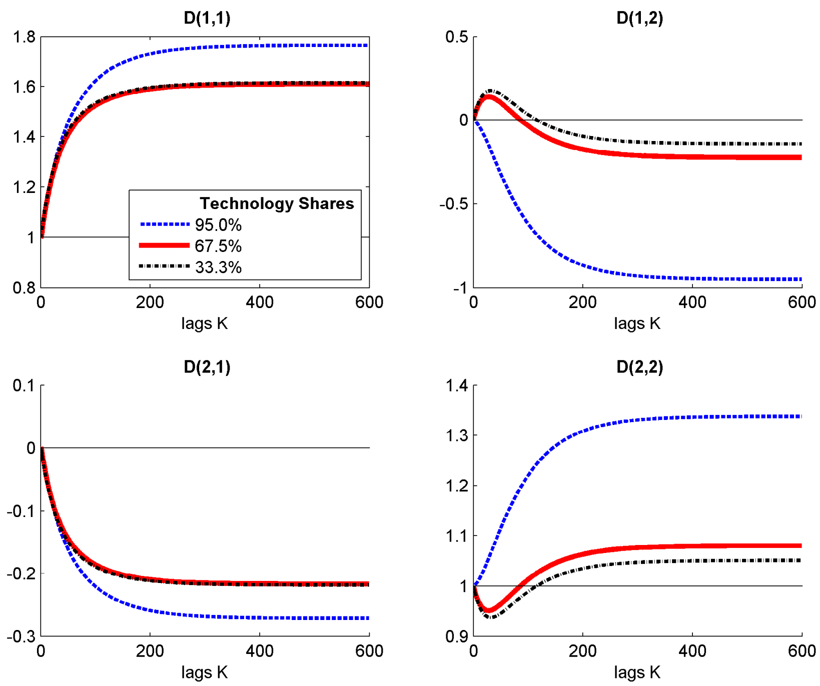
[25] Results are insensitive to using other information criteria, such as the Akaike criterion (AIC). In general, AIC is known for picking higher values of p compared to SIC. For this lab economy, AIC has been found to pick lag lengths of up to $p = 6$ with an average of $p = 2$.

[26] For a reference in the context of economics see Hansen and Sargent 2007; 2005.

[27] 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(b).

[28] The Γ_k from the Spectral Factorization Theorem are a smoothed version of the sample autocovariance since they are the coefficients of an inverse Fourier transform of the Newey-West estimate of the spectral density.

Figure A.4: $D(1) \neq I$



Note: Each panel shows an element of $\sum_{k=0}^{K} D_k$ for different lags K. $D(L)$ is computed from population moments of a VAR(1) applied to the model economy of Appendix A for three different percentages of the bandpass-filtered variability in output explained by technology shocks. A technology share of 67.5% (solid line) corresponds to the maximum likelihood estimates of CEV.

The theorem 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.[29]

Algorithms for implementing the factorization go back to Whittle (1963) and have recently been surveyed by Sayed and Kailath (2001). The simulations reported here use the algorithm of Li (2005), which is based on a state space representation of v_t and performed very reliably.[30] The remainder of this appendix describes the algorithm in more detail.

Suppose v_t follows an MA(q) as above. To represent it in a state space system, define the state vector $s_t = E\left\{\begin{bmatrix} v_t & v_{t+1} & \dots & v_{t+q-1}\end{bmatrix}^T \middle| v^{t-1}\right\}$ where v^{t-1} is the entire history of realizations

[29]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 $|\gamma_1/\gamma_0| < 0.5$.

[30]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 inverting larger matrices in the Riccati iterations which proved to be numerically less stable in the simulations computed for this paper.

of v_t up to time $t-1$. Li then constructs the following state space system

$$s_{t+1} = As_t + De_t$$
$$v_t = Cs_t + e_t$$

$$A = \begin{bmatrix} 0_m & I_m & 0_m & \dots & & 0_m \\ 0_m & 0_m & I_m & 0_m & \dots & 0_m \\ \vdots & & & \ddots & & \vdots \\ 0_m & \dots & & & 0_m & I_m \\ 0_m & \dots & & & 0_m & 0_m \end{bmatrix} \qquad D = \begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ D_q \end{bmatrix} \qquad C = \begin{bmatrix} I_m & 0_m & \dots & 0_m \end{bmatrix}$$

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

What is needed is a mapping from the autocovariances of v_t, Γ_k, to the state space objects. The objects of interest are the matrix D containing the stacked MA coefficients D_i as well as the variance $\Omega = Ee_te_t^T$ of the innovations process. To obtain this mapping, it is useful to stack the autocovariances into a matrix $M = \begin{bmatrix} \Gamma_1^T & \Gamma_2^T & \vdots & \Gamma_q^t \end{bmatrix}^T$

Li (2005, Theorem 2) shows that the variance-covariance matrix of the states $\Psi \equiv Es_ts_t^T$ solves the Riccati equation $\Psi = A\Psi A^T + (M - A\Psi C^T)(\Gamma_0 - C\Psi C^T)^{-1}(M - A\Psi C^T)^T$ and that the MA(q) coefficients can be recovered as $D = (M - A\Psi C^T)(\Gamma_0 - C\Psi C^T)^{-1}$ and $\Omega = \Gamma_0 - C\Psi C^T$. As shown by Li (2005), the above Riccati equation can be solved recursively, starting from $\Psi^{(0)} = 0$ and iterating over $\Psi^{(n+1)} = A\Psi^{(n)}A^T + (M - A\Psi^{(n)}C^T)(\Gamma_0 - C\Psi^{(n)}C^T)^{-1}(M - A\Psi^{(n)}C^T)^T$ since $\Psi = \lim_{n\to\infty} \Psi^{(n)}$ and $\Psi^{(n+1)} \geq \Psi^{(n)}$.

At the end of each factorization computed for this paper, it has been verified that the factorization produces an invertible MA(q) polynomial, which matches the original spectral density. In all simulations, this held up to machine accuracy.

Appendix C. VARs Implied by Lab Economy

This section outlines how to derive the following: First, values from the lab economy for true VAR objects like A_0, $A(1)$, $B(1)$, and the autocovariances of X_t. Second, population coefficients of finite-order VARs implied by the lab economy.[31]

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

$$X_t = \begin{bmatrix} \Delta \log (y_t/l_t) \\ \log l_t \end{bmatrix} = CZ_t \qquad \text{with} \quad Z_t = \mathcal{A}Z_{t-1} + \mathcal{B}\varepsilon_t \tag{C.1}$$

State vector and shock vector are:

$$Z_t = \begin{bmatrix} \hat{k}_t & \varepsilon_t^z & \tau_{l,t}^z & \hat{k}_{t-1} & \varepsilon_{t-1}^z & \tau_{l,t-1}^z \end{bmatrix}^T \qquad \varepsilon_t = \begin{bmatrix} \varepsilon_t^z & \varepsilon_t^l \end{bmatrix}^T \tag{C.2}$$

[31] For this specific two-shock economy, details can also be found in McGrattan (2005). For general state space models details can be found in Fernandez-Villaverde et al. (2005). To simplify the VAR notation, X_t has been demeaned prior to the analysis.

where $\hat{k}_t$ is the log-deviation of detrended capital from its steady 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 C is straightforward, and a detailed presentation can be found in CKM.

True VAR objects

The decomposition in Section 4 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 since true impulse responses and spectrum are given by $A(L) = C(I - \mathcal{A}L)^{-1}\mathcal{B}$ and $S_X(\omega) = A(e^{-i\omega})A(e^{-i\omega})^T$. The impact coefficients $A_0 = C\mathcal{B}$ are apparent from (C.1). The covariance matrix of the forecast errors equals $\Omega = C\mathcal{B}\mathcal{B}C^T$.

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

The non-structural moving average representation of X_t is $X_t = A(L)A_0^{-1}e_t = C(L)e_t$. The coefficients of the non-structural VAR(∞) representation of the model can be obtained by inverting this moving average, yielding $B(L)L = I - C(L)^{-1}$.

The autocovariances $EX_tX_{t-k}^T$ can be directly computed from the state space model. The covariance matrix of the states $EZ_tZ_t^T \equiv \Omega$ is obtained as the solution to a discrete Lyapunov equation: $\Omega = \mathcal{A}\Omega\,\mathcal{A}^T + \mathcal{B}\mathcal{B}^T$ and the autocovariances of X_t are $EX_tX_{t-k}^T = C\mathcal{A}^k\Omega\,C^T$.

VAR(p) coefficients in population

Finite-order VAR(p) can be computed as projections of X_t on a finite number of its past values, $X_{t-1}\dots X_{t-p}$. In line with the notation of the main text, population coefficients of a VAR(p) are denoted with a superscript "*OLS*", $X_t = B(L)^{\text{OLS}}X_{t-1} + v_t^{\text{OLS}}$. The coefficients of the lag polynomial $B(L)^{\text{OLS}} = \sum_{i=0}^{p-1} B_i^{\text{OLS}}L^i$ solve the OLS normal equations

$$E\left(X_t - \sum_{i=0}^{p-1} B_i^{\text{OLS}}X_{t-1-i}\right)X_{t-j}^T = 0 \quad \forall\, j = 1\dots p$$

which are evaluated using the autocovariance matrices of X_t whose computations are described in the preceding paragraph. For instance if $p = 1$, $B_1^{\text{OLS}} = \left(EX_tX_{t-1}^T\right)\left(EX_tX_t^T\right)^{-1}$. Detailed formulas for higher VARs can be found in Fernandez-Villaverde et al. (2005).

Chari et al. (2005, Proposition 1) show that the VAR representation of X_t in the model is of infinite order and residuals from a VAR(p) will not be martingales. By construction, the projection residuals v_t^{OLS} are orthogonal to $X_{t-1}, \dots, X_{t-p}$, but they are not orthogonal to the complete history of X_t. The moving average representation of the forecast errors $v_t^{\text{OLS}} = D(L)e_t$ is easily constructed from $D(L) = \left(I - B(L)^{\text{OLS}}L\right)(I - B(L)L)^{-1}$.

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 the case of a VAR(1), $X_t = B_1^{\text{OLS}}X_{t-1} + v_t^{\text{OLS}}$. The normal equations imply

$$\begin{aligned}\operatorname{Var} X_t &= B_1^{\text{OLS}}\ (\operatorname{Var} X_t)\ (B_1^{\text{OLS}})^T + \Omega_v^{\text{OLS}} \\ &= \sum_{k=0}^{\infty} (B_1^{\text{OLS}})^k\ \Omega_v^{\text{OLS}}\ ((B_1^{\text{OLS}})^k)^T \\ &= \sum_{k=0}^{\infty} C_k^{\text{OLS}}\ \Omega_v^{\text{OLS}}\ (C_k^{\text{OLS}})^T\end{aligned} \tag{C.3}$$

The second line is obtained by recursive substitution of $\operatorname{Var} X_t$ and the third line follows from the construction of moving-average coefficients of a VAR(1), $C_k^{\text{OLS}} = (B_1^{\text{OLS}})^k$. The argument is easily extended to VARs with higher lag orders by using their companion form.

Appendix D. Bandpass-Filtered Variance Share from SVARs

This appendix describes how to compute the share of bandpass-filtered fluctuations attributed to technology shocks from a set of SVAR parameters, $\hat{B}(L)$ and $\hat{A}_0$. The bandpass filter employed here considers only cycles with a duration between two-and-a-half and eight years. Denoting the bandpass-filtered level of output $\tilde{y}_t$, its variance can be easily computed from the transfer function

$$T_y(\omega) = \begin{bmatrix}(1 - e^{-i\omega})^{-1} & 1\end{bmatrix} \hat{C}(e^{-i\omega})\hat{A}_0$$

Depending on the identification scheme, $\hat{A}_0$ corresponds to A_0^{OLS}, $A_0^{\text{CEV-AM}}$, $A_0^{\text{CEV-NW}}$, $A_0^{\text{SF-AM}}$ or $A_0^{\text{SF-NW}}$ and $\hat{C}(L)$ corresponds to VMA implied by each method.

Using $\underline{\omega} = \frac{2\pi}{8 \cdot 12}$ and $\overline{\omega} = \frac{2\pi}{2.5 \cdot 12}$ the bandpass-filtered variance is

$$\operatorname{Var} \tilde{y}_t = \int_{\underline{\omega}}^{\overline{\omega}} T_y(\omega) T_y(\omega)^T d\omega$$

and the share of fluctuations attributed to technology shocks is the ratio $(\operatorname{Var} \tilde{y}_t | \varepsilon_t^z)/(\operatorname{Var} \tilde{y}_t)$, where $\operatorname{Var} \tilde{y}_t | \varepsilon_t^z$ conditions only on fluctuations attributed to technology shocks.

$$\operatorname{Var} \tilde{y}_t | \varepsilon_t^z = \int_{\underline{\omega}}^{\overline{\omega}} T_y(\omega) \begin{bmatrix}1 & 0 \\ 0 & 0\end{bmatrix} T_y(\omega)^T d\omega$$

Similar computations yield the variance shares for hours, when using the transfer function

$$T_l(\omega) = \begin{bmatrix}0 & 1\end{bmatrix} \hat{C}(e^{-i\omega})\, \hat{A}_0\ .$$

References

Acemoglu, D., Rogoff, K., Woodford, M. (Eds.), 2006. NBER Macroeconomics Annual 2006. MIT Press, Cambridge, MA.

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–58.

Andrews, D.W.K., Monahan, J.C., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60, 953–66.

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

Burns, A.F., Mitchell, W.C., 1946. Measuring Business Cycles. NBER.

Chari, V., Kehoe, P.J., McGrattan, E.R., 2008. Are structural VARs with long-run restrictions useful in developing business cycle theory? Journal of Monetary Economics 55, 1337–1352.

Chari, V.V., Kehoe, P.J., McGrattan, E.R., 2005. A critique of structural VARs using business cycle theory. Staff Report 364. Federal Reserve Bank of Minneapolis.

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

Christiano, L.J., Eichenbaum, M., Vigfusson, R., 2004. The response of hours to a technology shock: Evidence based on a direct measure of technology. Journal of the European Economic Association 2, 381–395.

Christiano, L.J., Eichenbaum, M., Vigfusson, R., 2006a. Alternative procedures for estimating vector autoregressions identified with long-run restrictions. Journal of the European Economic Association 2, 475–483.

Christiano, L.J., Eichenbaum, M., Vigfusson, R., 2006b. Assessing structural VARs, in: Acemoglu et al. (2006). chapter 1. pp. 1–72.

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

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

Dupaigne, M., Feve, P., 2009. Technology shocks around the world. Review of Economic Dynamics 12, 592–607.

Dupor, B., Kiefer, L., 2007. Exectuing long run restrictions. *mimeo*.

Erceg, C.J., Guerrieri, L., Gust, C., 2005. Can long-run restrictions identify technology shocks? Journal of the European Economic Association 3, 1237–1278.

Fernandez-Villaverde, J., Rubio-Ramirez, J., Sargent, T.J., 2005. A, B, C's (and D)'s for Understanding VARs. Technical Working Paper 0308. NBER.

Fernàndez-Villaverde, J., Rubio-Ramirez, J.F., Sargent, T.J., Watson, M.W., 2007. ABCs (and Ds) of understanding VARs. American Economic Review 97, 1021–1026.

Galì, J., 1999. Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations? The American Economic Review 89, 249–271.

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

Hansen, L.P., Sargent, T.J., 2005. Recursive models of dynamic linear economies. Unpublished Manuscript.

Hansen, L.P., Sargent, T.J., 2007. Robustness. Princeton University Press.

Hurwicz, L., 1950. Least-squares bias in time series, in: Koopmans, T.C. (Ed.), Statistical Inference in Dynamic Economic Models. John Wiley & Sons, Inc., New York. volume 10 of *Cowles Commission for Resarch in Economics*. chapter XV, pp. 365–383.

Ireland, P.N., 2004. Technology shocks in the New Keynesian model. The Review of Economics and Statistics 86, 923–936.

Jordá, Ó., 2005. Estimation and inference of impulse responses by local projections. American Economic Review 95, 161–182.

Kascha, C., Mertens, K., 2009. Business cycle analysis and VARMA models. Journal of Economic Dynamics and Control 33, 267–282.

Kehoe, P.J., 2006. Comment on Christiano, Eichenbaum, and Vigfusson's 'Assessing structural VARs', in: Acemoglu et al. (2006). chapter 1. pp. 73–96.

Li, L.M., 2005. Factorization of moving-average spectral densities by state-space representations and stacking. Journal of Multivariate Analysis 96, 425–438.

McGrattan, E., 2005. Appendix: A critique of structural VARs using real business cycle theory. Federal Reserve Bank of Minneapolis, *mimeo*.

Mertens, E., 2010. Are spectral estimatiors useful for implementing long-run restrictions in SVARs?. FEDS Working Paper 2010-09. Federal Reserve Board.

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

Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. Review of Economic Studies 61, 631–53.

Phillips, P.C.B., Sun, Y., Jin, S., 2006. Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. International Economic Review 47, 837–894.

Ravenna, F., 2007. Vector autoregressions and reduced form representations of DSGE models. Journal of Monetary Economics 54, 2048–2064.

Sayed, A., Kailath, T., 2001. A survey of spectral factorization methods. Numerical Linear Algebra with Applications 8, 467–496.

Sims, C.A., 1972. The role of approximate prior restrictions in distributed lag estimation. Journal of the American Statistical Association 67, 169–175.

Smets, F., Wouters, R., 2007. Shocks and frictions in U.S. business cycles: A Bayesian DSGE approach. The American

Economic Review 97, 586–606.
Sun, Y., Phillips, P.C.B., Jin, S., 2008. Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. Econometrica 76, 175–194.
Whittle, P., 1963. On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. Biometrika 50, 129–134.
Whittle, P., 1996. Optimal Control, Basics and Beyond. Wiley-Interscience Series in Systems and Optimization, John Wiley & Sons, Chichester.