Journal of Macroeconomics 33 (2011) 582–594

Contents lists available at ScienceDirect

Journal of Macroeconomics

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

Bootstrapping structural VARs: Avoiding a potential bias in confidence intervals for impulse response functions

Kerk L. Phillips *, David E. Spencer

Department of Economics, P.O. Box 22363, Brigham Young University, Provo, UT 84602, USA

A R T I C L E I N F O

Article history:
Received 3 February 2010
Accepted 22 February 2011
Available online 11 March 2011

JEL classification:
E32
E37
C32

Keywords:
Impulse response function
Structural VAR
Bias
Bootstrap

A B S T R A C T

Constructing bootstrap confidence intervals for impulse response functions (IRFs) from structural vector autoregression (SVAR) models has become standard practice in empirical macroeconomic research. The accuracy of such confidence intervals can deteriorate severely, however, if the bootstrap IRFs are biased. We document an apparently common source of bias in the estimation of the VAR error covariance matrix which can be easily reduced by a scale adjustment. This bias is generally unrecognized because it only affects the bootstrap estimates of the error variance, not the original OLS estimates. Nevertheless, as we illustrate here, analytically, with sampling experiments, and in an example from the literature, the bootstrap error variance bias can have significant distorting effects on bootstrap IRF confidence intervals. We also show that scale-adjusted bootstrap confidence intervals can be expected to exhibit improved coverage accuracy.

1. Introduction

Impulse response functions (IRFs) from structural vector autoregression (SVAR) models are widely employed to investigate the response of macroeconomic variables to identified structural shocks. Leading and influential examples of such studies include Blanchard and Quah (1989) examining the effects of aggregate demand and aggregate supply shocks on output and unemployment, Galí (1999) which investigates the effects of technology shocks, and Christiano et al. (1999) which assesses the effects of monetary policy shocks.

To assess uncertainty and draw inferences, these and other studies construct confidence intervals (CIs) around the estimated IRF. Increasingly, these intervals are constructed using bootstrap techniques.[1] In this paper we document a commonly occurring, but easily corrected, source of apparent bias in bootstrap estimates of IRFs from SVAR models.[2] Given the pervasiveness of the techniques that lead to this bias, it has important implications.[3] For example, it can lead to distorted CIs with such severe spurious asymmetry that the bootstrap CIs do not even include the estimated IRF. Sims and Zha (1999, p. 1125, fn 13)

* Corresponding author. Tel.: +1 801 422 5928.
E-mail address: kerk_phillips@byu.edu (K.L. Phillips).

[1] See, e.g., Runkle (1987) and Berkowitz and Kilian (2000).

[2] There are several potential source of bias in the estimation of bootstrap CIs for SVARs; see Kilian (1998). We focus on one source that has not been generally recognized. Though this source of bias may not be the most important in any particular application, it has the advantage of being easily resolved with a simple scale adjustment.

[3] This bias arises from the downward bias in the standard bootstrap estimate of the reduced form VAR error covariance matrix. Any object that depends on these estimates will be affected. This includes not only IRFs but bootstrap confidence intervals for error variance decompositions and bootstrap prediction intervals as well.

0164-0704/$ - see front matter
doi:10.1016/j.jmacro.2011.02.007

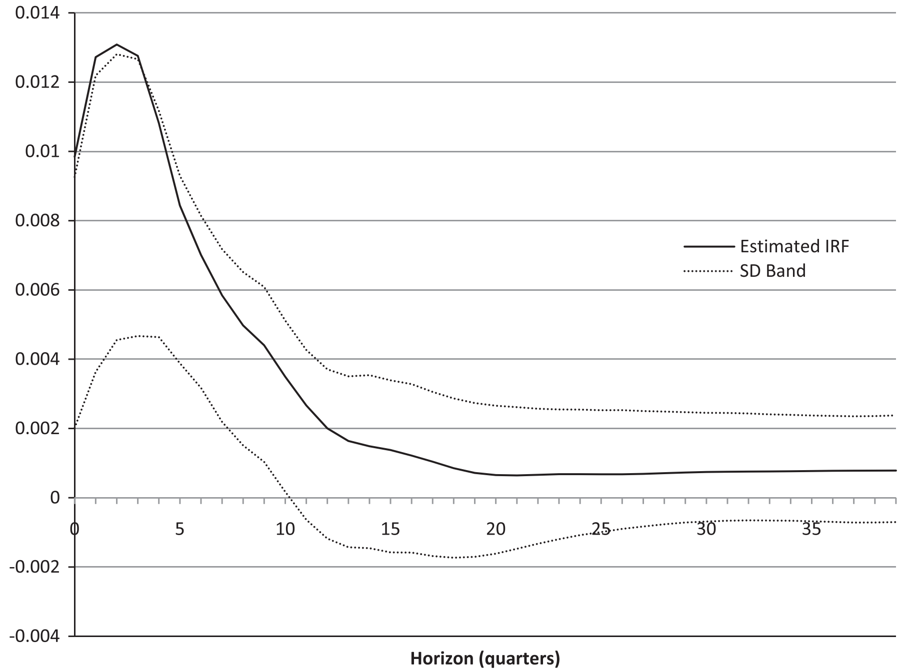

Fig. 1. A reestimated version of Fig. 3 in Blanchard and Quah (1989). It shows the response of output to aggregate demand shocks with asymmetric one standard deviation (SD) bands based on standard bootstrap estimates.

note that some SVAR studies have found it necessary to "use a modification of [the bootstrap confidence interval] that makes *ad hoc* adjustments to prevent the computed bands from failing to include the point estimates."

This bias-caused distortion can be seen in the results reported by Blanchard and Quah (1989); see especially their Figs. 3 and 5. Our Fig. 1 is a reestimated[4] version of their Fig. 3 with asymmetric one standard deviation bands.[5] Notice that the upper one standard deviation band actually lies below the original estimated IRF over the early horizon interval.[6] Anticipating our later discussion, Fig. 2 shows the same impulse response function with one standard deviation bands after implementing a degrees of freedom adjustment to reduce bias. The original asymmetry is greatly attenuated reflecting the fact that it is largely a spurious consequence of bias in the bootstrap estimates of the IRF.

If, as in the case of Galí (1999), researchers do not allow for asymmetric confidence intervals and simply plot error bands that are the estimated IRFs plus or minus one or two standard deviations, then the CIs are symmetric by construction, any bias is completely invisible, and the reported error bands are incorrect.[7]

Not all researchers attribute this seemingly odd behavior of IRFs completely to true skewness. Christiano et al. (2006), for example, note that, in their case, the mean value of the bootstrapped IRFs is quite different from the initial estimated IRF. They note that the "asymmetric percentile confidence intervals show that when data are generated by these [bootstrap] VARs, ... the impulse response functions have a downward bias."[8]

The finite-sample bias we examine arises from the fact that the *bootstrap* IRF for a SVAR depends on the *bootstrap* OLS estimate of the error covariance matrix in the reduced form vector autoregression (VAR), standard estimates of which are biased downward. This bias is apparently common[9] but, as we demonstrate below, it can be ameliorated by a degrees of

[4] We make the same data adjustments made by Blanchard and Quah and estimate the model over the same sample period. Our results differ slightly because we use revised data.

[5] We compute our asymmetric one standard deviation bands by obtaining 1000 bootstrap IRFs and then taking, in each direction, the square root of the mean squared deviation from the mean bootstrap IRF.

[6] The fact that the corresponding Blanchard-Quah IRF does not actually cross the bounds is due to the way they compute their one standard deviation bands. They obtain 1000 bootstrap IRFs which, for each horizon, they separate into those above and those below the original IRF. They then compute the standard deviation for each class to obtain the asymmetric one standard deviation bounds. This procedure assures that the IRF will not "cross" the bounds. A bound that is coincident with the original IRF indicates that, at that horizon, *none* of the bootstrap IRFs were above (or below) the original IRF.

[7] This practice of forced symmetry is followed in some econometric software packages like EViews.

[8] Christiano et al. (2006, p. 26).

[9] Of course, we have not documented this for all or even most SVAR studies. We have, however, examined programs that authors have posted on web sites. In *none* of the cases was the appropriate degrees of freedom adjusted bootstrap error covariance estimator used. Some programs (including those which use the standard VCV instruction in RATS) calculate the MLE of the bootstrap covariance matrix and thus make no degrees of freedom adjustment at all. We therefore conclude that this bias is likely to be common in practice.

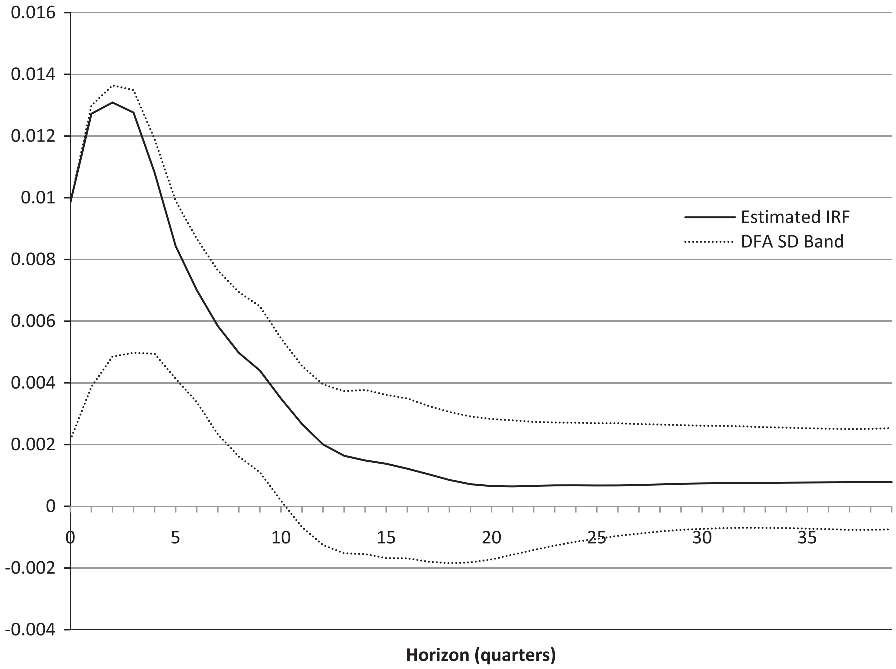

Fig. 2. A reestimated version of Fig. 3 in Blanchard and Quah (1989) with asymmetric one standard deviation bands based on *degrees of freedom adjusted* (DFA) bootstrap estimates.

freedom adjustment. Even though it can lead to substantially distorted bootstrap IRF CIs, this bias is generally unrecognized and not corrected in practice because it only affects the *bootstrap* estimates of the error variance, not the original OLS estimates.

Though the main insight of this paper is motivated by analogy to analytical results in a simple regression model and confirmed by Monte Carlo evidence in more general settings, it is important to indicate that the suggested degrees of freedom adjustment is ultimately inherently heuristic since exact analytical underpinnings are not available in the case of a VAR model.

In the next section, we examine the specific source of this bias in the bootstrap estimate of error variances in the context of a simple regression model and show how a degrees of freedom adjustment eliminates the bias. We then consider autoregressive models. Since exact finite-sample results are not available in this case, we proceed by analogy to suggest a similar degrees of freedom adjustment and confirm its usefulness with Monte Carlo evidence. In Section 3 we extend the analogy to show how the bias in bootstrap error variance estimates effects the bootstrap IRFs and thus the bootstrap confidence intervals for the original IRF and, again, suggest a degrees of freedom adjustment in this case of SVAR models. In Section 4 we illustrate how making the recommended degrees of freedom adjustment affects the IRF confidence intervals obtained in a widely-cited previous study. We also compare coverage rates for the alternative bootstrap CIs. The final section offers a brief conclusion.

2. A source of bias

2.1. Standard regression models

The simplest way to illustrate the bias under investigation is to examine a standard linear regression model with nonstochastic regressors. We first consider a univariate regression model represented by

$$y = X\beta + u \tag{1}$$

where y is a $T \times 1$ vector of observations on a dependent variable, X is a $T \times R$ matrix of observations on R nonstochastic regressors (perhaps including a constant), β is an $R \times 1$ vector of regression coefficients, and u is a $T \times 1$ vector of errors. We assume that $E(u) = 0$ and $E(uu') = \sigma^2 I_T$. Applying ordinary least squares (OLS), we obtain coefficient and error variance estimates: $\hat{\beta} = (X'X)^{-1}X'y$, $\hat{\sigma}^2 = \frac{1}{T-R}(\hat{u}'\hat{u})$, where $\hat{u} = y - X\hat{\beta}$. The indicated degrees of freedom correction makes $\hat{\sigma}^2$ an unbiased estimator for σ^2.

To help us understand the key argument to follow, it is useful to interpret the degrees of freedom adjustment from the perspective that it is necessary to compensate for the fact that the OLS residuals tend to be "smaller" than the error terms. Note that the expected value of the average squared error is σ^2; i.e., $E\left(\frac{u'u}{T}\right) = \sigma^2$. On the other hand, $E\left(\frac{\hat{u}'\hat{u}}{T}\right) = \left(\frac{T-R}{T}\right)E\left(\frac{u'u}{T}\right)$, which reflects that, on average, the squared residuals are $((T-R)/T)$ times as large as the squared errors.[10] Thus, to obtain an

[10] See Davidson and MacKinnon (1993, pp. 69–70).

unbiased estimate, we must rescale each residual by $(\frac{T}{T-R})^{\frac{1}{2}}$ and then compute the average squared *rescaled* residual giving the usual unbiased estimate for σ^2, $\hat{\sigma}^2 = \frac{1}{T-R}(\hat{u}'\hat{u})$.

Now consider obtaining a bootstrap variance estimate in this simple case. The bootstrap methodology relies on an analogy between the *unknown* population probability distribution of the "real world" and the *known* empirical distribution in the "bootstrap world."[11] The bootstrap analyst hopes to learn about the population distribution of $(\hat{\sigma}^2 - \sigma^2)$ by examining the distribution of $(\tilde{\sigma}^2 - \hat{\sigma}^2)$ where $\hat{\sigma}^2$ is the "pseudo-population" variance of the empirical distribution (not a random variable in the bootstrap world) and $\tilde{\sigma}^2$ is a candidate bootstrap variance estimate (which, of course, is a random variable in the bootstrap world).[12] Typically, this is done by drawing many samples from the pseudo-population given by the original sample. Because we can resample as many times as we want, we can estimate the mean of $(\tilde{\sigma}^2 - \hat{\sigma}^2)$ and, thus, the bias of $\tilde{\sigma}^2$ using Monte Carlo experiments.

Pursuing this bootstrap analogy and recalling the insight discussed above, we might expect an analogous degrees of freedom adjustment to be helpful for bootstrap variance estimates. This has been confirmed for the simple regression model by Freedman and Peters (1984, p. 99) and Peters and Freedman (1984, p. 408).

Suppose we obtain *bootstrap* estimates of the error variance as follows. For bootstrap replications $b = 1, \ldots, B$, generate

$$y_b^* = X\hat{\beta} + u_b^* \tag{2}$$

where the elements of u_b^* are drawn with replacement from the OLS residuals, $\hat{u}$. Then, apply OLS to Eq. (2) to get bootstrap estimates of $\hat{\beta}$ (not β), which we denote $\tilde{\beta}_b$, and bootstrap residuals, $\tilde{u}_b$. In the bootstrap, the variance estimate, $\tilde{\sigma}_b^2$, is an estimate of $\hat{\sigma}^2$ (not σ^2), the "population" error variance in the pseudo-population given by the error variance of the original OLS residuals, $\hat{u}$. The usual bootstrap variance estimate is given by $\tilde{\sigma}_{b,1}^2 = \frac{1}{T-R}(\tilde{u}_b'\tilde{u})$.

In this case, we can get some analytical insight for the properties of $\tilde{\sigma}_{b,1}^2$ by conditioning on the unknown population distribution. Proceeding as above, we note that though $E(\frac{\hat{u}'\hat{u}}{T-R}) = \sigma^2$,

$$E\left(\frac{\tilde{u}_b'\tilde{u}_b}{T-R}\right) = \left(\frac{T-R}{T}\right)E\left(\frac{u_b^{*\prime}u_b^*}{T-R}\right) = \left(\frac{T-R}{T}\right)\sigma^2$$

since the elements of u_b^* are drawn randomly from $\hat{u}$. This reflects that, on average, the squared bootstrap residuals are $((T-R)/T)$ times as large as the squared OLS residuals which are the pseudo-population errors. Consequently, we suggest that a better bootstrap estimate might be given by $\tilde{\sigma}_{b,2}^2 = \frac{T}{(T-R)^2}(\tilde{u}_b'\tilde{u}) = \frac{T}{T-R}\tilde{\sigma}_{b,1}^2$. This is the same rescaling suggested by Freedman and Peters (1984) and Peters and Freedman (1984).

If this analogy holds exactly, we would expect the size of the (proportional) bias for the natural estimator to be $-R/T$.[13] While this vanishes asymptotically, it can be important in small samples when R is large relative to T. To illustrate, we conduct a Monte Carlo experiment in which we simulate obtaining bootstrap estimates of the error variance in a univariate regression model like (1). We estimate models with nine regressors including a constant term, $R = 9$, for three sample sizes: $T = 30, 50, 100$.[14] Consequently, the expected bias for $\tilde{\sigma}_{b,1}^2$ is -30%, -18% and -9% respectively. For each sample size, we draw 1000 samples of size T from a normal distribution with mean zero and variance 0.81. For each of these Monte Carlo draws we generate observations for y, estimate (1) by OLS, and compute the usual population-unbiased estimate of the error variance, $\hat{\sigma}^2$. The average estimate is given in Table 1. To examine the bias of the two bootstrap error variance estimates, $\tilde{\sigma}_{b,1}^2$ and $\tilde{\sigma}_{b,2}^2$,[15] we take each of the 1000 Monte Carlo samples and obtain 200 bootstrap estimates in each case. The average values are reported in Table 1 for our three sample sizes. We call this the estimated bootstrap bias.

The results in Table 1 confirm our expectation very nicely. The "natural" bootstrap estimator, $\tilde{\sigma}_{b,1}^2$, has bias approximately equal to $-R/T$ while $\tilde{\sigma}_{b,2}^2$ is approximately unbiased.

This bias in the standard bootstrap "error" variance carries over exactly to the case of a multivariate seemingly unrelated regression model with nonstochastic regressors. To confirm the theory, we have conducted simple Monte Carlo experiments similar to those undertaken for the univariate regression model discussed above. To save space, we do not report the results here but simply indicate that the conclusions are the same.[16]

2.2. Autoregressive models

Consider a univariate AR(p) with a constant term, ν, so that $R = p + 1$:

$$y_t = \nu + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + u_t; \quad t = -p+1, \ldots, 0, 1, \ldots, T \tag{3}$$

[11] See Efron and Tibshirani (1993), especially Chapter 8, for discussion of this analogy.

[12] So, in the bootstrap world, $\tilde{\sigma}^2$ is an estimate of $\hat{\sigma}^2$.

[13] It should be noted that bias arising from maximum likelihood estimation (MLE) of the error variance will be even larger. As is well known, the MLE of σ^2, $\breve{\sigma}^2 = \frac{1}{T}(\hat{u}'\hat{u})$, is biased; i.e., $E(\breve{\sigma}^2) = (\frac{T-R}{T})\sigma^2$. Thus, the proportional bias is $-R/T$. Now, when we bootstrap and obtain the MLE of $\breve{\sigma}^2$, $\widetilde{\breve{\sigma}}_b^2 = \frac{1}{T}(\tilde{u}_b'\tilde{u}_b)$, the bias is magnified since we have a biased estimate of a biased estimate. $\widetilde{\breve{\sigma}}_b^2 = \frac{(T-R)^2}{T^2}\tilde{\sigma}_{b(2)}^2$, so $E\left(\widetilde{\breve{\sigma}}_b^2\right) = \frac{(T-R)^2}{T^2}\sigma^2$ and the expected proportional bias is $\left(\frac{(T-R)^2}{T^2}\right) - 1 = \frac{R^2-2TR}{T^2}$ which is negative and larger (in absolute value) than $-R/T$.

[14] The values of the regressors are 1.1, 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3 with the first element being the constant term; $\sigma^2 = 0.81$.

[15] Note that Table 1reports the bias relative to, $\hat{\sigma}^2$, the pseudo-population variance.

[16] The results are reported in an unpublished appendix available on request.

Table 1
Bootstrap error variance estimates in standard univariate linear regression model with $R = 9$; number of Monte Carlo trials = 1000, number of bootstrap draws = 200. True value of variance = 0.81.

Estimator	Sample size	Mean estimate	Bias ($-R/T$)	Estimated bias[a] (%)
$\hat{\sigma}^2$	30	0.8051		
$\tilde{\sigma}^2_{b(1)}$	30	0.5633	−30.0%	−30.03
$\tilde{\sigma}^2_{b(2)}$	30	0.8047	0	−0.05
$\hat{\sigma}^2$	50	0.8053		
$\tilde{\sigma}^2_{b(1)}$	50	0.6603	−18.0%	−18.01
$\tilde{\sigma}^2_{b(2)}$	50	0.8053	0	0.00
$\hat{\sigma}^2$	100	0.8054		
$\tilde{\sigma}^2_{b(1)}$	100	0.7327	−9.0%	−9.03
$\tilde{\sigma}^2_{b(2)}$	100	0.8052	0	−0.02

[a] The estimated bootstrap bias is the average difference between the relevant bootstrap error variance estimate and $\hat{\sigma}^2$, the pseudo-population variance across the 1000 Monte Carlo trials.

Table 2
Bootstrap error variance estimates in an AR(8) model with a constant term ($R = 9$); number of Monte Carlo trials = 1000, number of bootstrap draws = 200. True value of variance = 0.81.

Estimator	Sample size	Mean estimate	Proportional difference ($-R/T$)	Estimated bias[a] (%)
$\hat{\sigma}^2$	30	0.8476		
$\tilde{\sigma}^2_{b(1)}$	30	0.6178	−30.0%	−27.11
$\tilde{\sigma}^2_{b(2)}$	30	0.8826	0	4.13
$\hat{\sigma}^2$	50	0.8129		
$\tilde{\sigma}^2_{b(1)}$	50	0.6766	−18.0%	−16.77
$\tilde{\sigma}^2_{b(2)}$	50	0.8252	0	1.51
$\hat{\sigma}^2$	100	0.8120		
$\tilde{\sigma}^2_{b(1)}$	100	0.7418	−9.0%	−8.65
$\tilde{\sigma}^2_{b(2)}$	100	0.8152	0	0.39

[a] The estimated bootstrap bias is the average difference between the relevant bootstrap error variance estimate and $\hat{\sigma}^2$, the pseudo-population variance across the 1000 Monte Carlo trials.

where u_t is white noise with variance σ^2 and T is the number of *usable* observations. Because the regressors are stochastic, the finite sample theory of the previous section does not apply. However, following Stine (1987, p. 1074) and Berkowitz and Kilian (2000, p. 5), we might speculate (correctly) that similar bias problems exist for bootstrap estimators of the error variance in this case.

Since analytical results are not available, we examine the finite-sample bias issue for the AR(p) model using a Monte Carlo exercise similar to the one described above. We generate data for, and estimate, a model like (3) in which $p = 8$ so $R = 9$.[17] For each of three sample sizes, $T = 30, 50, 100$, we draw 1000 samples for u_t of size $T + p$ from a normal distribution with mean zero and variance 0.81. For each of these Monte Carlo draws we generate observations for y, estimate (3) by OLS, and compute the usual estimate of the error variance, $\hat{\sigma}^2$. The average estimate is given in Table 2. To examine the bias (relative to $\hat{\sigma}^2$) of the two bootstrap error variance estimates, $\tilde{\sigma}^2_{b,1}$ and $\tilde{\sigma}^2_{b,2}$, we obtain 200 bootstrap estimates for each of the Monte Carlo samples.[18] The average values are reported in Table 2 for each of our three sample sizes.

The results are quite informative. The heuristically expected bias for the corresponding standard linear regression is a rather good guide for the bias in the AR(p) model. We confirm that the bootstrap estimator of the error variance given by $\tilde{\sigma}^2_{b,1}$ is biased and thus likely to result in significant distortion when the number of slope coefficients is large relative to the sample size.

Proceeding by analogy, we expect these bias results to carry over to the case of a (non-structural) VAR(p) with K variables. In that case, our interest is the $K \times K$ error (innovation) covariance matrix Σ. Assuming a constant term, the usual degrees-of-freedom-corrected OLS estimator for Σ is $\widehat{\Sigma} = (\frac{1}{T-R})\widehat{U}'\widehat{U}$ where $\widehat{U}$ is the $T \times K$ matrix of OLS residuals and $R = Kp + 1$. The "natural" but perhaps biased *bootstrap* estimator of $\widehat{\Sigma}$ is $\widetilde{\Sigma}_{b,1} = (\frac{1}{T-R})\widetilde{U}'_b\widetilde{U}_b$ where $\widetilde{U}_b$ is the $T \times K$ matrix of bootstrap residuals from the bth bootstrap iteration. The degrees of freedom adjusted (DFA) bootstrap estimator of Σ is $\widetilde{\Sigma}_{b,2} = \left(\frac{T}{(T-R)^2}\right)\widetilde{U}'_b\widetilde{U}_b$.

We have investigated the bootstrap error variance bias for a two-equation VAR(8) model with a constant term using Monte Carlo methods similar to those described above and find the bias to be quite close to the bias expected from the above

[17] The model coefficients are 0.008, 0.25, 0.11, −0.03, −0.004, −0.12, 0.03, −0.02, −0.08 with the first element being the constant term; $\sigma^2 = 0.81$.

[18] For each bootstrap iteration, we obtain the initial p observations $\{y_{-p+1}, \ldots, y_o\}$ by drawing (with replacement) from the original generated sample $\{y_t\}^T_{-p+1}$.

heuristic analysis above. To conserve space, we do not report the results here since they are quite similar to those reported for the AR(8) model above.[19] In particular, the bias for $\widetilde{\Sigma}_{b,1}$ is approximately $-(\frac{Kp+1}{T})$ where K is the number of equations (variables) in the VAR(p). For a two-equation VAR(8) model, this implies an approximate bias of -17% for each element of Σ when $T = 100$.[20]

3. Bootstrapping IRFs for SVARs

The downward bias of the standard bootstrap estimator of the VAR error covariance matrix is of particular concern when we are interested in drawing inferences about IRFs from a SVAR model since the IRFs are nonlinear functions of both VAR slope parameters *and* the elements of the error covariance matrix.[21] In this section we show how bias in the bootstrap estimate of the VAR error covariance matrix affects bootstrap IRFs and, thus, bootstrap CIs.

To estimate a SVAR model of the behavior of a $K \times 1$ vector of variables, y_t, and the corresponding IRFs, we begin by specifying a finite-order *reduced form* VAR model which can always be estimated:

$$B(L)y_t = u_t \tag{4}$$

where $B(L)$ is a matrix of p-order polynomials in the lag operator, L, u_t is a vector of K reduced form errors, and $E(u_t u_t') = \Sigma$. In general, consistent OLS estimates of $B(L)$ and Σ can be obtained: $\widehat{B}(L)$ and $\widehat{\Sigma}$, where $\widehat{\Sigma} = (\frac{1}{T-R})\widehat{U}'\widehat{U}$, $\widehat{U}$ is the $T \times K$ matrix of OLS residuals, and $R = Kp + 1$ since we assume a constant term.

The reduced form moving average representation is obtained by inverting (4):

$$y_t = B(L)^{-1}u_t = C(L)u_t \tag{5}$$

where $C_0 = I$, the identity matrix. The shocks in this representation are one-step-ahead forecast errors and do not, in general, correspond to structural economic shocks for which we want to obtain impulse response functions.

We assume that there exists a structural moving average representation of the model from which the IRFs can be obtained:

$$y_t = A(L)\varepsilon_t \tag{6}$$

where ε_t is a vector of K *structural* shocks and we make the standard assumption that $E(\varepsilon_t\varepsilon_t') = I_K$. This assumption provides a normalization as well as a set of identifying restrictions. The elements of the matrix polynomial $A(L)$ give the impulse response functions: $a_{ij,l}(i, j = 1, \ldots, K;\ l = 0, 1, \ldots)$ indicates the response of variable i in l periods to a one unit (standard deviation) movement in the jth structural shock today.

Though the IRFs are frequently the objects of interest in macroeconomic analysis, they cannot generally be estimated directly from time series data since the SVAR model (6) is not identified without further restrictions. However, equating terms in (4) and (6) allows us to conclude the following:

$$u_t = A_0\varepsilon_t \tag{7}$$

$$A_l = C_l A_0 \quad l = 1, \ldots \tag{8}$$

Thus, it is clear that knowledge of the K^2 elements of A_0 is sufficient to obtain the IRF.

From (7) and the assumption that $E(\varepsilon_t\varepsilon_t') = I_K$, we infer the key relationship between the covariance matrices of the structural and reduced form errors:

$$\Sigma = A_0 A_0' \tag{9}$$

Symmetry of Σ provides $\left(\frac{K(K+1)}{2}\right)$ restrictions on A_0. With $\left(\frac{K(K-1)}{2}\right)$ additional restrictions, A_0 can be identified and IRFs computed. Eqs. (8) and (9) assure us that the estimated IRFs depend on estimates of $B(L)$ and Σ. I.e., $\hat{a}_{ij,l} = g(\hat{\beta}\hat{\sigma})\ (i,j = 1, \ldots, K; l = 1, \ldots)$ where $\hat{\beta} = vec(\widehat{B})$, a $K(Kp + 1) \times 1$ vector, and $\hat{\sigma} = vech(\widehat{\Sigma})$, a $\left(\frac{K(K+1)}{2}\right) \times 1$ vector and the form of the nonlinear function g depends on the identification strategy. Consequently, the properties of the IRFs depend on the properties of $\hat{\beta}$ and $\hat{\sigma}$. Similarly, the properties of the bootstrap IRFs depend in the same way on the properties of the bootstrap estimates of β and σ: $\tilde{a}_{ij,l} = g(\tilde{\beta}, \tilde{\sigma})\ (i,j = 1, \ldots, K; l = 0, 1, \ldots)$.

We can see from this that there are several potential sources of bias for the bootstrapped IRFs and, thus, bootstrap confidence intervals for the original IRFs. The source we focus on here arises when the *bootstrap* estimate of σ, $\tilde{\sigma}$, is biased for $\hat{\sigma}$, the elements of the pseudo-population covariance matrix. How much difference does the appropriate degrees or freedom adjusted bootstrap estimation of the error covariance matrix make for bootstrap estimates of the IRF? We provide some intuitive analytics to address this question.

[19] The results are reported in an unpublished appendix available on request.

[20] Note that if, for this VAR model, we had computed MLE rather than OLS estimates of Σ in both the initial and bootstrap stages, the approximate bias for the elements of the bootstrap estimate of Σ would have been magnified to -31%. See Footnote 12.

[21] Other objects of frequent interest that are also nonlinear functions of VAR slope parameters and elements of the error covariance matrix are forecast error variance decompositions and measures of predictability. Thus, related bootstrap confidence or prediction intervals would also suffer from the bias we discuss here. See Inoue and Kilian (2002).

From Eq. (8) we infer that the bootstrap estimates of the IRF are given by

$$\widetilde{A}_l = \widetilde{C}_l \widetilde{A}_0 \quad l = 1, \ldots \tag{10}$$

where $\widetilde{A}_l$, $\widetilde{C}_l$, and $\widetilde{A}_0$ are bootstrap estimates. Any bias in $\widetilde{A}_0$ will, thus, likely carry over to all the $\widetilde{A}_l$. We see from Eq. (9) that $\widetilde{A}_0$ depends only on the bootstrap estimate of Σ, $\widetilde{\Sigma}$, which, based on results reported above, we expect to be systematically biased. Consequently, it will be instructive to consider how potential bias in the bootstrapped estimates of the VAR covariance matrix can affect the bootstrapped IRF. The original sample estimate of the VAR error covariance matrix is $\widehat{\Sigma} = \left(\frac{1}{T-R}\right)\widehat{U}'\widehat{U}$. As in the previous section, we consider two alternative bootstrap estimates of $\widehat{\Sigma}$ based on the pseudo-population. The standard bootstrap estimate is given by $\widetilde{\Sigma}_{b,1} = \left(\frac{1}{T-R}\right)\widetilde{U}_b'\widetilde{U}_b$ where $\widetilde{U}_b$ is the $T \times K$ matrix of bootstrap residuals from the bth bootstrap iteration. The DF-adjusted bootstrap estimate is given by $\widetilde{\Sigma}_{b,2} = \left(\frac{T}{T-R}\right)\left(\frac{1}{T-R}\right)\widetilde{U}_b'\widetilde{U}_b = \left(\frac{T}{T-R}\right)\widetilde{\Sigma}_{b,1}$. So,

$$\widetilde{\Sigma}_{b,1} = (1+b)\widetilde{\Sigma}_{b,2} \tag{11}$$

where $b = -\frac{R}{T}$ is the proportional difference between the standard and DF-adjusted bootstrap estimates of $\widehat{\Sigma}$. Based on the results of the previous section (including the Monte Carlo evidence alluded to), we might expect b to approximate the bias in $\widetilde{\Sigma}_{b,1}$.

We can use Eq. (11) to derive the implied proportional difference between the corresponding bootstrapped IRFs. Eq. (9) implies that $\widetilde{\Sigma}_{b,1} = \widetilde{A}_{0,1}\widetilde{A}_{0,1}'$ and $\widetilde{\Sigma}_{b,2} = \widetilde{A}_{0,2}\widetilde{A}_{0,2}'$. So, from (11) we have $\widetilde{A}_{0,1}\widetilde{A}_{0,1}' = (1+b)\widetilde{A}_{0,2}\widetilde{A}_{0,2}'$ which, in turn, implies that

$$\widetilde{A}_{0,1} = (1+b)^{1/2}\widetilde{A}_{0,2} = (1+a)\widetilde{A}_{0,2} \tag{12}$$

Equating $(1 + b)^{1/2}$ and $(1 + a)$ in (12) implies that b and a are related by

$$a = (1+b)^{1/2} - 1 \tag{13}$$

Since $-1 < b = -\frac{R}{T} < 0$, we see that $b < a < 0$. As T increases without bound, b and a approach zero so the standard and DF-adjusted IRFs are asymptotically equivalent.

Now, consider how this proportional difference in the bootstrap estimate of $\widehat{A}_0$ affects the alternative bootstrap IRFs. First, it is important to recognize that $\widetilde{C}_l$ does not depend on which bootstrap estimate of $\widehat{\Sigma}$ we choose. Thus, as implied by Eq. (10), the two IRFs are given by

$$\widetilde{A}_{l,j} = \widetilde{C}_l\widetilde{A}_{0,j} \quad j = 1, 2;\ l = 1, \ldots \tag{14}$$

In particular,

$$\widetilde{A}_{l,2} = \widetilde{C}_l\widetilde{A}_{0,2} = \widetilde{C}_l\left(\frac{1}{1+a}\right)\widetilde{A}_{0,1} = \left(\frac{1}{1+a}\right)\widetilde{A}_{l,1} \tag{15}$$

where the second equality follows from (12) and the final equality follows from (14). From Eq. (15), it follows that the proportional difference between the two IRFs is the same for all values of l:

$$\frac{\widetilde{A}_{l,1} - \widetilde{A}_{l,2}}{\widetilde{A}_{l,2}} = a, \quad l = 1, \ldots \tag{16}$$

Thus, the bootstrap IRF proportional difference is constant and equal to a for the entire IRF horizon. So, for example, if we have a SVAR model with $K = 2$, $p = 8$, $T = 100$ and a constant term, the elements of the standard bootstrap covariance matrix estimate is 17% less than the DF-adjusted estimate and the corresponding IRFs differ by 9% of the DF-adjusted IRF.[22]

4. An example

As indicated earlier,[23] the procedures that generate this bias seem to be quite common in the empirical SVAR literature. In this section we illustrate its effect in practice by replicating the biased results obtained in a single influential paper by Christiano, Eichenbaum, and Evans (CEE) (1999). We then compute the corresponding DF-adjusted IRF and associated bootstrap confidence intervals to draw our comparison. Finally, we examine coverage accuracy by comparing the coverage rates for standard bootstrap CIs with those for the DF-adjusted CI.

In their paper, CEE examine the effects of monetary policy shocks on several economic variables of interest using models imposing a recursive structure to identify the relevant shocks.[24] Their first benchmark model includes a constant term and four lags ($p = 4$) of seven variables ($K = 7$) with the federal funds rate as the chosen monetary policy instrument. They estimate their models using quarterly data over the period 1965:3–1995:2. Given the loss of observations due to the four lags in the VAR,

[22] These are the implied values of b and a in percentage terms. As indicated above, $b = -R/T$, and a can be computed from (13).

[23] See Footnote 8 above.

[24] The seven variables they include, in recursive order, are the log of real GDP, the log of the implicit GDP deflator, a series of smoothed changes in a commodity price index, the federal funds rate, the log of nonborrowed reserves plus extended credit, the log of total reserves, and the log of $M1$.

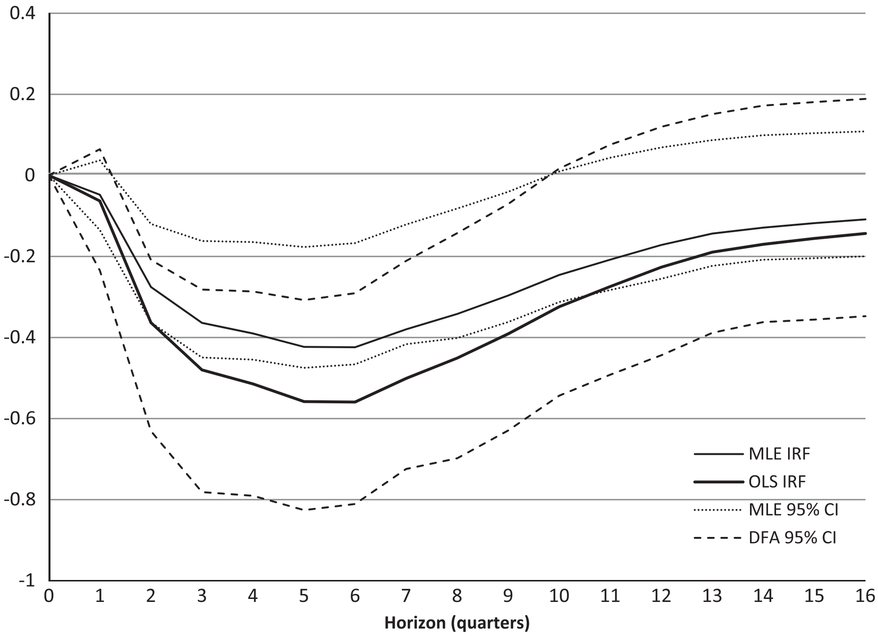

Fig. 3. Impulse response functions showing the effect of a contractionary monetary policy on real GDP with 95% confidence intervals. The solid line gives the original MLE IRF and the long-dashed bold line gives the OLS IRF; CEE use MLE. The dotted lines give the MLE bootstrap 95% confidence intervals and the dashed lines give the DF-adjusted 95% confidence intervals.

$T = 116$ in our notation. We replicate their results by estimating their model over the same sample period.[25] For illustrative purposes, we report only the IRF indicating the effects of a negative monetary policy shock on output. While this is an IRF of particular interest, the same bias will be present in all the other 48 IRFs as well.[26] As seen in Fig. 3 here and Fig. 2 of CEE (1999, p. 86), given a positive federal funds rate shock, "after a delay of 2 quarters, there is a sustained decline in real GDP " (p. 87). We note that CEE use MLE to estimate the VAR error covariance estimate so the estimated IRF will be biased. Furthermore, we see that the bootstrap confidence intervals reflect considerable asymmetry which, we shall see momentarily, is partially due to bias in the confidence intervals arising from biased bootstrap IRF estimates.

To illustrate the effect of bias due to MLE and the further bias due to the CEE bootstrap IRFs, we estimate the CEE model once again but this time including the degrees of freedom correction we suggest in this paper. These results for the first-stage IRF and the bootstrap confidence intervals are also reported in Fig. 3. We first notice that the fundamental conclusion regarding the IRF is unchanged: a contractionary federal funds rate shock will, after a lag, have a sustained negative effect on real GDP.[27] We also notice that using the OLS rather than the MLE estimate of original error covariance matrix causes the corresponding OLS IRF to lie entirely below the MLE IRF obtained by CEE.

In addition, we see that the confidence intervals also shift significantly when we adjust the degrees of freedom in the bootstrap estimates of the error covariance matrix. We note three consequences. First, we see that for much of the time horizon, the DF-adjusted OLS IRF actually lies below the CEE 95% confidence intervals. Second, we see that adjusting the degrees of freedom has greatly reduced the asymmetry in the confidence intervals.[28] Third, we notice that between 2 and 11 quarters, the upper 95% confidence bounds are farther away from zero after degrees of freedom adjustment. This provides stronger evidence supporting the conclusion that a contractionary monetary policy has a significant negative effect on output over that horizon.

Since part of the distortion in the CEE results is a consequence of their choice to use MLE estimates of the error covariance matrix, we also illustrate how much distortion remains when we use OLS estimates. The results are reported in Fig. 4. In the

[25] Indeed, we have estimated the CEE model using their data which Larry Christiano has generously made available on his website.

[26] This is because, as Eq. (12) shows, the proportional difference between the bootstrap estimates of the A_0 matrix is a multiplicative scalar that affects all elements of the matrix the same way. Eq. (14) shows that this same proportional difference will carry over to every IRF. The complete results are reported in an unpublished appendix available on request.

[27] Indeed, we will always draw the same conclusion about statistical significance when our interest is in whether or not the IRF is significantly different from zero. This is a consequence of the fact, illustrated in the previous section Eq. (15), that the DF-adjusted bootstrap IRF is *proportional* to the standard IRF at all horizons with the constant of proportionality positive but less than one. Accordingly, both confidence interval bounds will cross the horizontal axis (zero line) at exactly the same horizons. This implies that the range over which the IRF is significantly greater or less than zero will be the same whether or not a degrees of freedom adjustment is applied. Adjusting the degrees of freedom can lead to a reversal of conclusion, however, if the null hypothesis takes on a value other than zero.

[28] This leads us to conjecture that the often puzzling asymmetry in IRF CIs found in the literature is largely due to the bias documented in this paper.

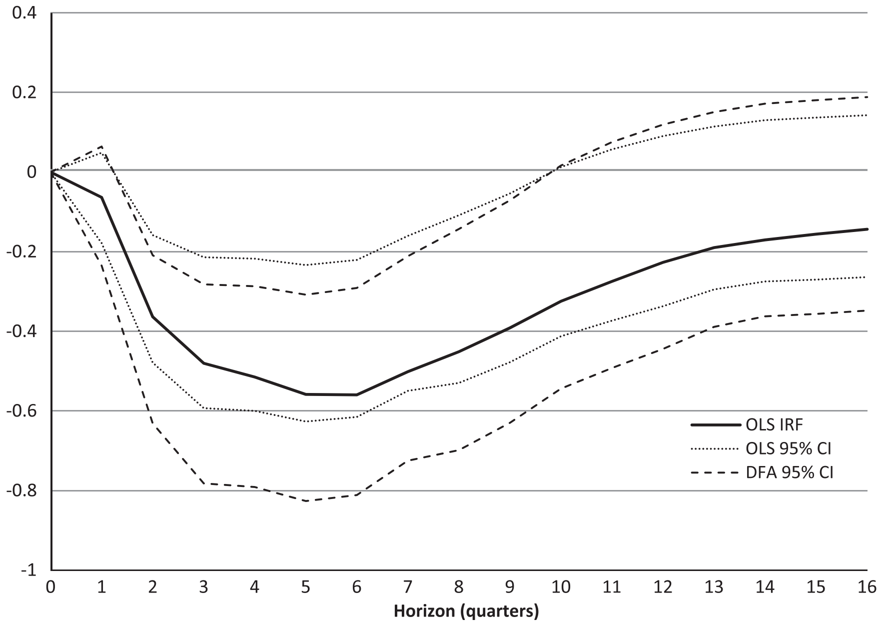

Fig. 4. Impulse response function showing the effect of a contractionary monetary policy on real GDP with 95% confidence intervals. The solid line gives the original OLS IRF. The dotted lines give the typical bootstrap 95% confidence intervals not adjusted for degrees of freedom and the dashed lines give the DF-adjusted 95% confidence intervals.

typical approach incorporating the natural OLS degrees of freedom correction, the original IRF is already DF-adjusted so we only have a single IRF estimate. However, the typical procedure does result in biased bootstrap confidence intervals. As in Fig. 3, we again see that the typical biased procedure results in quite asymmetric confidence intervals which are, in part, a consequence of the bias; the DF-adjusted confidence intervals exhibit much less asymmetry. Also, as noted in the discussion of Fig. 3, over a range of intermediate horizons, the upper bound of the DF-adjusted confidence intervals lie below their biased counterparts giving us greater confidence in our conclusion that a monetary contraction has a significant negative effect on output.

These examples illustrate that adjusting the degrees of freedom in both the original IRF and especially in the bootstrap confidence interval estimates can remove distortions that change the quantitative (if not qualitative) conclusions when SVAR models are used.

Of course, for the degrees of freedom adjustment we recommend to be of practical value, we must have confidence that it will result in greater coverage accuracy for the resulting CIs. Accordingly, we conclude this section by reporting the results of a series of Monte Carlo experiments that investigate the coverage rates of alternative bootstrap CIs. To avoid the potential arbitrariness of an *ad hoc* data generating process (DGP), we treat the benchmark CEE model as our initial DGP from which we obtain the "true" IRF.[29] Using that model and assuming jointly normal errors with the CEE estimated covariance matrix, we generate 1000 Monte Carlo trials of the same length as the CEE sample. Once again, to keep the analysis focused, we look only at the IRF representing the effect of a negative monetary policy shock on output.[30] For each Monte Carlo trial, we then take 200 bootstrap replications and construct three sets of 95% bootstrap IRF confidence intervals: MLE (following CEE), standard OLS, and DF-adjusted. We then report the coverage rates[31] for each of these respective confidence intervals across the 1000 trials.

Fig. 5 reports the results for the benchmark CEE model DGP along with a reference line at 0.95 reflecting the 95% nominal value of the confidence intervals. We label the methods: MLE, OLS, and DFA. We notice that none of the methods yields coverage rates that are consistently near the ideal value of 0.95 in this baseline case but the DFA method we recommend is uniformly superior to the traditionally-used alternatives. Coverage rates for the DFA method fall to about 0.65 but are generally above 0.75. The MLE has particularly poor coverage rates for intermediate horizons, falling as low as 0.2 while coverage rates fall in between for the OLS method. These poor coverage rates may reflect the fact that the bias we account for here is not the only bias affecting the results. Furthermore, it would not be appropriate to generalize that coverage rates for all IRFs are likely to be as poor as those reported in Fig. 5. The evidence reported by Kilian and Chang (2000) for the CEE model suggests

[29] Kilian and Chang (2000) argue that the results of studies that focus on simple *ad hoc* (e.g., bivariate) VAR models may not generalize to higher dimensional models that are typical of actual applied work. In their study investigating coverage rates, they use three leading models in the literature, including the CEE model, as data generating processes.

[30] For comparison, see the upper left graph in Fig. 3 of Kilian and Chang (2000).

[31] By coverage rate we mean the fraction of Monte Carlo trials for which the respective confidence interval includes the true IRF. We evaluate the coverage rate at each point of the IRF horizon.

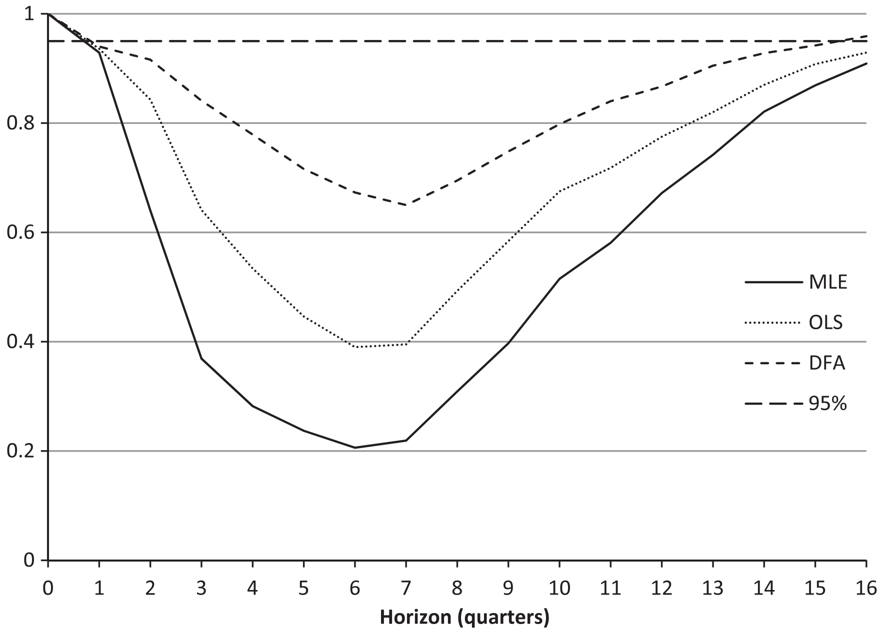

Fig. 5. A comparison of coverage rates for three alternative 95% bootstrap confidence intervals: MLE, OLS, and DFA. Applied to the impulse response function showing the effect of monetary policy on output in the baseline CEE model as originally parameterized.

that the IRF representing the effects of a monetary policy shock on output exhibited lower coverage rates than other IRFs. We have confirmed this for our case.[32] It turns out that the IRF we are most interested in has the poorest coverage rates.

To investigate the robustness of the finding that the DFA method gives greater coverage accuracy, we consider alternative variations on the benchmark CEE model. We first consider alternative parameterizations of the error covariance matrix and then examine variation in key slope coefficients. We also consider the effect of increasing the sample size and the consequences for DGPs with non-normal errors.

The first three alternative DGPs retains the slope parameter of the baseline CEE model but change the values of the elements of the error covariance matrix. The first of these doubles all those values, the second halves them, and the third sets all the off-diagonal elements to zero. To conserve space, and since the coverage rates were only marginally affected in each of these three cases, the results are not reported here but are available from the authors in an unpublished appendix. As in the baseline model, coverage rates were generally poor but the DFA method improved considerably on the OLS method and especially the MLE method.

Because of the impracticality of varying the very large number of slope parameters for the seven-equation, four-lag CEE model in a broadly systematic way, we focus only on the parameters that are most likely to affect the IRF of principal interest. In particular, we vary the first order autoregressive and cross-autoregressive coefficients relating to output and the federal funds rate. We first try a parameterization which is the same as the baseline model except that it halves the benchmark values of these four first order parameters. We also examine two further parameterizations which halve only the first order autoregressive slope coefficients for output and the federal funds rate respectively. The result for the third of these is reported in Fig. 6.[33] We see that halving the size of the coefficient on the first lagged value of the federal funds rate results in considerable improvement in coverage rates for all three methods with the performance of the recommended DFA method being quite good. For all horizons, the coverage rate is above 0.85 and often above 0.9. The coverage rates for the other parameterizations also improve relative to the baseline model with the halving of all four of the relevant first order slope coefficients producing slightly better coverage rates and the model that halves only the first order autoregressive coefficient on output doing a little worse than the results reported in Fig. 6.

Since the bias we are investigating shrinks as sample size increases, we examine the effect on coverage rates of increasing sample size in the baseline CEE model. We first double the usable CEE sample size (T) from 116 to 232 and report the results in Fig. 7. Not surprisingly, when compared with Fig. 5, we see that the coverage rates improve considerably, with rates for the DFA method exceeding 0.8 for most of the horizon. The DFA method also continues to be greatly superior to the OLS and MLE methods for estimating the bootstrap error covariance matrix. It is interesting to note that this larger sample size is typical of

[32] In the appendix available from the authors, we report coverage rate results for all the IRFs corresponding to the effects of a monetary policy (federal funds rate) shock. For the other IRFs as well, the DFA method yields higher coverage rates.
[33] The others are included in the unpublished appendix available on request.

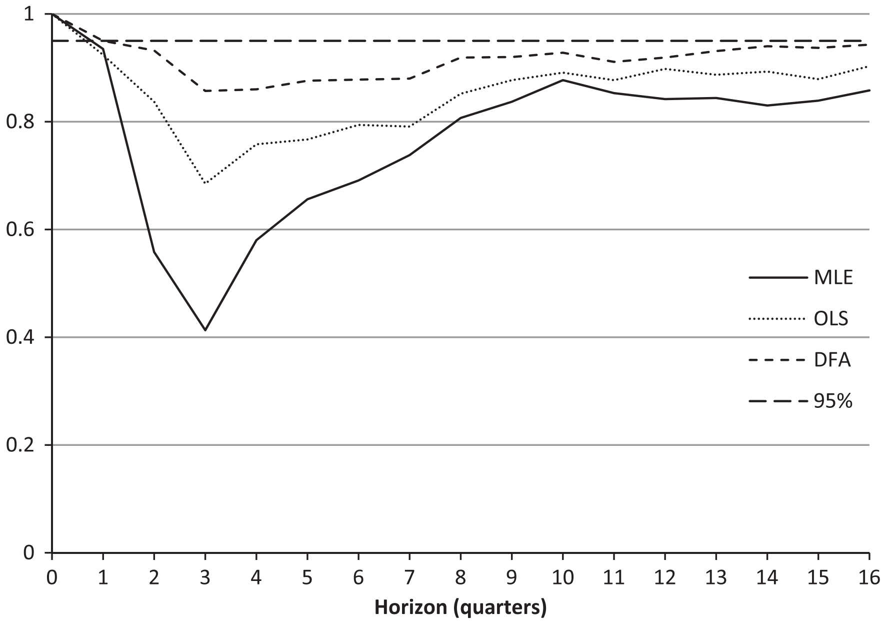

Fig. 6. A comparison of coverage rates for three alternative 95% bootstrap confidence intervals: MLE, OLS, and DFA. Applied to the impulse response function showing the effect of monetary policy on output in the CEE model setting the first order autoregressive slope coefficient for the federal funds rate at half its size in the baseline model.

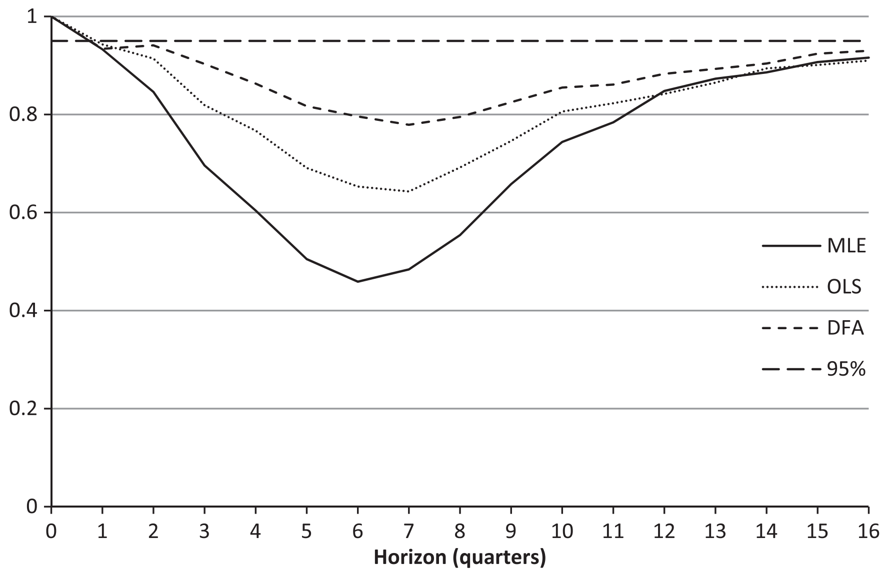

Fig. 7. A comparison of coverage rates for three alternative 95% bootstrap confidence intervals: MLE, OLS, and DFA. Applied to the impulse response function showing the effect of monetary policy on output in the CEE model with the sample size doubled ($T = 232$).

most recent empirical work using post-war US quarterly data. When we quadruple the sample size ($T = 464$), the coverage rate for the DFA method generally exceeds 0.90.[34]

As a final consideration, we consider DGPs with non-normal errors. In particular, we are interested in what happens if the errors come from a distribution with fatter tails than the normal. We investigate errors generated by two t-distributions, one with 10 degrees of freedom and one with five degrees of freedom. The results for the latter are given in Fig. 8. We see that coverage rates improve, relative to Fig. 5, for all three methods with the DFA method maintaining its superiority. It exhibits

[34] Results are included in the unpublished appendix.

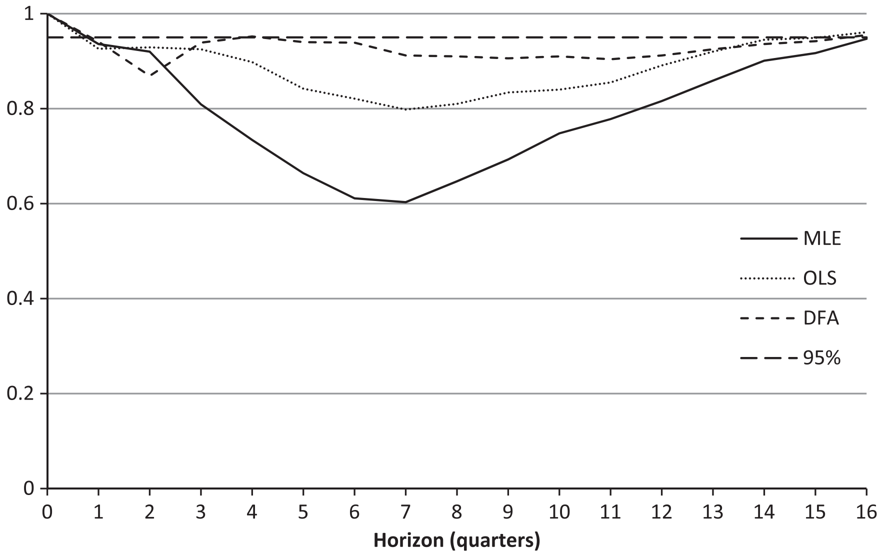

Fig. 8. A comparison of coverage rates for three alternative 95% bootstrap confidence intervals: MLE, OLS, and DFA. Applied to the impulse response function showing the effect of monetary policy on output in the CEE model with errors generated from a *t*-distribution with five degrees of freedom.

coverage rates generally quite close to the nominal level of 0.95. The results for the *t*-distribution with 10 degrees of freedom (reported in the unpublished appendix) also represent an improvement over the baseline case but not as much as we see in Fig. 8.

In summary, we conclude that the straightforward DFA method for obtaining bootstrap estimates results in considerably improved coverage accuracy. Since the bias we discuss here reflects only one of several potential sources of bias, we are not surprised to see that even the DFA method often results in poor coverage accuracy. However, the evidence reported here suggests that, as sample size approaches that of most modern macroeconomic research, and/or if the distribution generating the errors has fatter tails than a normal distribution, coverage rates for the DFA method may not only be improved but become reasonable. Furthermore, these potential gains are available with a simple degrees of freedom adjustment.

5. Conclusion

This paper has discussed a commonly occurring source of bias in bootstrap estimates of confidence intervals for IRFs in SVARs arising from the downward bias in the traditional bootstrap estimate of the VAR covariance matrix. Since the bootstrap IRFs depend on these biased estimates, they are systematically distorted along with the implied bootstrap IRF percentile confidence intervals. This distortion is potentially large but, fortunately, can be readily ameliorated by an additional degrees of freedom adjustment when estimating the VAR covariance matrix. Furthermore, the results of a series of Monte Carlo experiments suggest that we can expect the degrees of freedom adjusted confidence intervals to exhibit improved coverage accuracy relative to traditionally-used confidence intervals.

Acknowledgements

We thank Richard W. Evans, Lutz Kilian, and James B. McDonald for their comments on previous versions of this paper. We are grateful to Jason Blankenagel, Ryan Decker, Mark Hendricks and Bryan Perry for excellent research assistance. We also thank an anonymous referee whose comments led to significant improvement in the exposition and content of the paper.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmacro.2011.02.007.

References

Berkowitz, J., Kilian, L., 2000. Recent developments in bootstrapping time series. Econometric Reviews 19, 1–48.
Blanchard, O., Quah, D., 1989. The dynamic effects of aggregate demand and supply disturbances. American Economic Review 79, 655–673.
Christiano, L.J., Eichenbaum, M., Evans, C.L., 1999. Monetary policy shocks: what have we learned and to what end? In: Taylor, J.B., Woodford, M. (Eds.), The Handbook of Macroeconomics, vol. 1A. Amsterdam, North Holland, pp. 65–148.
Christiano, L.J., Eichenbaum, M., Vigfusson, R., 2006. Assessing structural VARs. NBER Macroeconomics Annual 21, 1–72.

Davidson, R., MacKinnon, J.G., 1993. Estimation and Inference in Econometrics. Oxford University Press, New York.
Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.
Freedman, D.A., Peters, S.C., 1984. Bootstrapping a regression equation: some empirical results. Journal of the American Statistical Association 79, 97–206.
Galí, J., 1999. Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations? American Economic Review 89, 249–271.
Inoue, A., Kilian, L., 2002. Bootstrapping smooth functions of slope parameters and innovation variances in VAR(∞) models. International Economic Review 43, 309–331.
Kilian, L., 1998. Small-sample confidence intervals for impulse response functions. Review of Economics and Statistics 80, 218–230.
Kilian, L., Chang, P., 2000. How accurate are confidence intervals for impulse responses in large VAR models? Economics Letters 69, 299–307.
Peters, S.C., Freedman, D.A., 1984. Some notes on the bootstrap in regression problems. Journal of Business and Economics Statistics 2, 406–409.
Runkle, D.E., 1987. Vector autoregression and reality. Journal of Business and Economics Statistics 5, 437–442.
Sims, C.A., Zha, T., 1999. Error bands for impulse responses. Econometrica 67, 1113–1155.
Stine, R.A., 1987. Estimating properties of autoregressive forecasts. Journal of the American Statistical Association 82, 1072–1078.