

# How Reliable are Local Projection Estimators of Impulse Responses?\*

Lutz Kilian                      Yun Jung Kim

University of Michigan    University of Michigan

CEPR

October 31, 2009

## Abstract

We compare the finite-sample performance of impulse response confidence intervals based on local projections (LPs) and vector autoregressive (VAR) models in linear stationary settings. We find that in small samples the asymptotic LP interval often is less accurate than the bias-adjusted bootstrap VAR interval, notwithstanding its excessive average length. Although the asymptotic LP interval has adequate coverage in sufficiently large samples, its average length still far exceeds that of bias-adjusted bootstrap VAR intervals with comparable accuracy. Bootstrap LP intervals (with or without bias correction) and asymptotic VAR intervals are shorter on average, but often lack coverage accuracy in finite samples.

**JEL Classification Number:** C32, C52, C53

**Keywords:** impulse response function, vector autoregression, local projection, confidence interval, bootstrap.

---

\*Acknowledgements: This project was stimulated by discussions with Jim Stock. We thank Matias Cattaneo, Oscar Jordà, Barbara Rossi, the editor, and two anonymous referees for helpful comments on an earlier draft of the paper.

## 1 Introduction

Local (linear) projections (LPs) were proposed by Jordà (2005, 2009) as an alternative to traditional VAR based estimators of impulse responses.<sup>1</sup> Our objective in this paper is to provide some practical guidance about which of the combinations of estimation method (VAR versus LP) and method of inference (asymptotic versus bootstrap) is likely to be the most reliable method of impulse response analysis in practice when the data generating process is linear and stationary. Specifically, we focus on (1) the asymptotic delta method interval of Lütkepohl (1990) and (2) percentile intervals based on the bias-corrected bootstrap method proposed by Kilian (1998a,b; 1999). For local projections we investigate (3) an improved version of the asymptotic interval proposed in Jordà (2005) and (4) we propose a bias-adjusted bootstrap percentile interval based on the block bootstrap method. For each method, we compare the pointwise coverage accuracy and average length of the asymptotic and bootstrap confidence intervals. Our analysis is based on a set of stylized bivariate VAR(1) data generating processes (DGPs) as well as a high-dimensional VAR(12) DGP of the type used in studying responses to monetary policy shocks. We also investigate VARMA(1,1) and ARMA(1,1) DGPs based on the asymptotic results for VAR( $\infty$ ) models in Lütkepohl and Poskitt (1991) and the corresponding bootstrap results in Inoue and Kilian (2002).

The remainder of the paper is organized as follows. Section 2 briefly establishes the notation and contrasts the construction of impulse response estimates from VAR models and from local projections. In section 3, we build intuition based on results from a Monte Carlo study that employs a stylized bivariate VAR(1) DGP used in the previous literature. The simulation results for a more realistic VAR(12) DGP are presented in section 4. Section 5 focuses on VARMA DGPs. We conclude in Section 6.

## 2 Review of VARs and Local Projections

Consider a  $K$ -dimensional linear vector autoregressive data-generating processes (DGP) of finite order  $p$ :<sup>2</sup>

$$y_t = B_1 y_{t-1} + \dots + B_p y_{t-p} + e_t, \quad (1)$$

---

<sup>1</sup>For an empirical application of LP methods to impulse response estimation see Haug and Smith (2007). Hall, Inoue, Nason and Rossi (2009) use LP impulse responses in estimating DSGE models. For closely related methodologies also see Chang and Sakata (2007), Jordà and Kozicki (2007), and Jordà and Marcellino (2009).

<sup>2</sup>This approach is standard in the literature. Alternatively, one could interpret a vector autoregression as an approximation to general stationary linear DGP (see, e.g., Lütkepohl and Poskitt 1991; Inoue and Kilian 2002). This case will be addressed in section 5.

where  $t = p + 1, \dots, T$ ,  $y_t = (y_{1t}, \dots, y_{Kt})'$  is a  $(K \times 1)$  random vector,  $B_i, i = 1, \dots, p$ , are  $(K \times K)$  coefficient matrices and  $e_t = (e_{1t}, \dots, e_{Kt})'$  is  $K$ -dimensional i.i.d. white noise, i.e.,  $E(e_t) = 0$ ,  $E(e_t e_s') = 0$  for  $s \neq t$  and  $E(e_t e_t') = \Sigma_e$  where  $\Sigma_e$  is non-singular and positive definite.<sup>3</sup> All values of  $z$  satisfying  $\det(I_K - B_1 z - \dots - B_p z^p) = 0$  lie outside the unit circle. For expository purposes, we abstract from deterministic regressors, although we will allow for an intercept in estimation throughout this paper. This VAR process can be written in structural form as:

$$A_0 y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t, \quad (2)$$

where  $\Sigma_\varepsilon = I_K$  without loss of generality.

## 2.1 Impulse Responses

Impulse responses to VAR reduced-form disturbances are obtained recursively as

$$\Phi_h^{VAR(p)} = \sum_{l=1}^h \Phi_{h-l}^{VAR(p)} B_l, \quad h = 1, 2, \dots, H, \quad (3)$$

where  $\Phi_0^{VAR(p)} = I_K$  and  $B_l = 0$  for  $l > p$ . The corresponding responses to structural shocks are given by:

$$\Theta_h^{VAR(p)} = \Phi_h^{VAR(p)} A_0^{-1}, \quad h = 0, 1, \dots, H, \quad (4)$$

where  $A_0^{-1}$  satisfies  $A_0^{-1} (A_0^{-1})' = \Sigma_e$ . For the purpose of the analysis below, we postulate that  $A_0^{-1}$  is a lower triangular matrix. Element  $(i, j)$  of  $\Theta_h^{VAR(p)}$  is  $\theta_{ij,h}^{VAR(p)}$  and represents the response of variable  $i$  to a one-time structural shock  $j$ ,  $h$  periods ago. By construction,  $\theta_{ij,h}^{VAR(p)}$  is a nonlinear function of  $B$  and  $\Sigma_e$ . Estimates  $\hat{\Theta}_h^{VAR(p)}$  are constructed by substituting the least-squares estimates of  $B$  and  $\Sigma_e$  obtained from regression (1).

An alternative approach to estimating reduced form impulse responses is to fit the linear projection

$$y_{t+h} = \mu + F_1 y_t + F_2 y_{t-1} + \dots + F_q y_{t-q+1} + u_{t+h} \quad \text{for } h = 1, \dots, H, \quad (5)$$

where  $u_t$  may be serially correlated or heteroskedastic (see Jordà 2005, 2009). The lag length  $q$  needs not be common across different horizons. By construction, the slope  $F_1$  can be interpreted as the response of  $y_{t+h}$  to a reduced-form disturbance in period  $t$ :

$$\Phi_h^{LP(q)} = F_1 = E(y_{t+h} | e_t = 1; y_t, \dots, y_{t-q}) - E(y_{t+h} | e_t = 0; y_t, \dots, y_{t-q}), \quad h = 1, \dots, H. \quad (6)$$

---

<sup>3</sup>The assumption of i.i.d. innovations is common in applied work and provides a useful benchmark for our purposes. It could be relaxed with suitable changes in the theory and implementation of the asymptotic and bootstrap approach (see Goncalves and Kilian 2004, 2007).

$\Phi_0^{LP(q)} = I_K$ . The corresponding structural impulse responses are

$$\Theta_h^{LP(q)} = \Phi_h^{LP(q)} A_0^{-1}, \quad h = 0, 1, \dots, H, \quad (7)$$

where  $A_0^{-1}$  is obtained based on the VAR model as described earlier.<sup>4</sup>  $\theta_{ij,h}^{LP(q)}$  denotes the response of variable  $i$  to a one-time structural shock  $j$ ,  $h$  periods ago. Estimates  $\hat{\Theta}_h^{LP(q)}$  are constructed from the the VAR( $p$ ) estimate  $\hat{A}_0^{-1}$  and the  $\hat{\Phi}_h^{LP(q)}$  estimates obtained from a sequence of least-squares regressions (5) for each horizon  $h$ . Under the maintained assumption of the DGP in equation (1), both  $\hat{\theta}_{ij,h}^{VAR(p)}$  and  $\hat{\theta}_{ij,h}^{LP(q)}$  will be consistent for  $\theta_{ij,h}^{VAR(p)}$ .

## 2.2 Asymptotic Confidence Intervals

Let  $\beta = \text{vec}(B_1, B_2, \dots, B_p)$  and  $\sigma = \text{vech}(\Sigma_e)$ . Under suitable moment restrictions, the asymptotic distribution of the VAR impulse response estimator can be derived by the delta method:

$$\sqrt{T} \text{vec} \left( \hat{\Theta}_h^{VAR(p)} - \Theta_h^{VAR(p)} \right) \xrightarrow{d} N \left( 0, C_h \Sigma_{\hat{\beta}} C_h' + \bar{C}_h \Sigma_{\hat{\sigma}} \bar{C}_h' \right) \quad (8)$$

where  $C_0 = 0$ ,  $C_h = (A_0^{-1} \otimes I_K) G_h$  with  $G_h = \partial \text{vec}(\Phi_h^{VAR(p)}) / \partial \beta'$ , and  $\bar{C}_h = (I_K \otimes \Phi_h^{VAR(p)}) \partial \text{vec}(A_0^{-1}) / \partial \sigma'$ . Explicit expressions for the asymptotic variance of the impulse response estimator can be found in Lütkepohl (1990). The nominal  $(1 - \alpha)\%$  confidence interval satisfies

$$P \left( \hat{\theta}_{ij,h}^{VAR(p)} - z_{1-\alpha/2} \frac{1}{\sqrt{T}} \hat{\sigma} \left( \hat{\theta}_{ij,h}^{VAR(p)} \right) \leq \theta_{ij,h}^{VAR(p)} \leq \hat{\theta}_{ij,h}^{VAR(p)} + z_{1-\alpha/2} \frac{1}{\sqrt{T}} \hat{\sigma} \left( \hat{\theta}_{ij,h}^{VAR(p)} \right) \right) = 1 - \alpha, \quad (9)$$

where  $\hat{\sigma} \left( \hat{\theta}_{ij,h}^{VAR(p)} \right)$  is the square root of element  $(K(j-1) + i, K(j-1) + i)$  of  $\left( \hat{C}_h \hat{\Sigma}_{\hat{\beta}} \hat{C}_h' + \bar{C}_h \hat{\Sigma}_{\hat{\sigma}} \bar{C}_h' \right)$ ,  $z_{1-\alpha/2}$  denotes the  $(1 - \alpha/2)$ -quantile of the  $N(0, 1)$  distribution.

The asymptotic confidence interval of the corresponding LP estimator proposed by Jordà (2005) is

$$P \left( \hat{\theta}_{ij,h}^{LP(q)} - z_{1-\alpha/2} \frac{1}{\sqrt{T}} \hat{\sigma} \left( \hat{\theta}_{ij,h}^{LP(q)} \right) \leq \theta_{ij,h}^{LP(q)} \leq \hat{\theta}_{ij,h}^{LP(q)} + z_{1-\alpha/2} \frac{1}{\sqrt{T}} \hat{\sigma} \left( \hat{\theta}_{ij,h}^{LP(q)} \right) \right) = 1 - \alpha. \quad (10)$$

Here  $\hat{\sigma} \left( \hat{\theta}_{ij,h}^{LP(q)} \right)$  is the square root of element  $(K(j-1) + i, K(j-1) + i)$  of

$$(A_0^{-1'} \otimes I_K) \left( (y_t' M_x y_t)^{-1} \otimes \hat{\Sigma}_u \right) (A_0^{-1} \otimes I_K) + \bar{G}_h \Sigma_{\hat{\sigma}} \bar{G}_h', \quad (11)$$

---

<sup>4</sup>Jordà (2005) does not explicitly discuss the distinction between the structural and reduced-form impulse responses. The Gauss code provided by Jordà, however, shows that his structural impulse responses are constructed using the VAR estimate of  $A_0^{-1}$ .

where  $M_x = I - X(X'X)^{-1}X'$ ,  $X = \begin{bmatrix} 1 & y_{t-1} & y_{t-2} & \dots & y_{t-q} \end{bmatrix}$ ,  $\bar{G}_h = (I_K \otimes \Phi_h^{LP(q)})\partial vec(A_0^{-1})/\partial \sigma'$ , and  $\hat{\Sigma}_u = E(u_{t+h}u'_{t+h})$  in equation (5). The first additive component of this variance-covariance matrix captures the variance of  $vec(\hat{\Phi}_h^{LP(q)}A_0^{-1})$  and reflects the uncertainty associated with the slope parameter estimates. The second additive component incorporates the estimation uncertainty associated with the estimate of  $A_0^{-1}$  (see Jordà 2009).<sup>5</sup> Following Jordà (2005), we employ the Newey-West estimator of  $\hat{\Sigma}_u$ .<sup>6</sup>

### 2.3 Bootstrap Confidence Intervals

Confidence intervals can also be obtained by bootstrap approximations. For the VAR impulse response estimator, we consider the well-established bias-corrected bootstrap confidence interval proposed by Kilian (1998a, 1999) based on 2,000 bootstrap replications. The reader is referred to the relevant literature for details of that procedure.<sup>7</sup> For the LP impulse response estimator, no bootstrap methods have been considered to date. Although Jordà (2009) discusses the potential benefits from bootstrapping the LP estimator, he does not explore any bootstrap methods in his work. In this paper, we propose a block bootstrap approach since the error term in LP regressions is serially correlated. By construction, the LP impulse response estimate for horizon  $h$  depends on the  $(1+q)$  tuple  $(y_{t+h}, y_t, y_{t-1}, \dots, y_{t-q+1})$ . To preserve the correlation in the data, we first construct the set of all possible  $(1+q)$  tuples. Then blocks of  $l$  consecutive  $(1+q)$  tuples are drawn (see, e.g., Berkowitz, Birgean and Kilian (1999) for a review of this bootstrap method) and used in the construction of  $\hat{\Phi}_h^{LP*}$ . In constructing  $\hat{\Theta}_h^{LP*}$ , for each bootstrap replication, we construct  $A_0^{-1*}$  based on a draw  $\hat{\Sigma}_e^*$  from the asymptotic distribution of  $\hat{\Sigma}_e^{VAR}$ .<sup>8</sup> A nominal  $(1-\alpha)\%$  percentile confidence interval may

---

<sup>5</sup>Jordà (2005) abstracts from this second component. This causes the asymptotic interval for LP too narrow at short horizons. Additional simulation results show that adding the second term significantly improves coverage accuracy of the asymptotic confidence interval at short horizons, with a marginal increase in average length.

<sup>6</sup>Jordà (2005) shows that the disturbance terms in a local projection has a moving average component of order  $h$  under our assumptions:  $u_{t+h} = e_{t+h} + \Phi_1^{VAR}e_{t+h-1} + \Phi_2^{VAR}e_{t+h-2} + \dots + \Phi_{h-1}^{VAR}e_{t+1}$ . This suggests that we set the truncation lag for the Newey-West estimator to be  $h$  for each local projection horizon  $h$ . While the results are not overly sensitive to the choice of the truncation lag, estimating  $\Sigma_u$  by least squares would seriously undermine the accuracy of the LP interval.

<sup>7</sup>We implement this method as discussed in Kilian (1998b,c, 1999) using the full double loop rather than using the computational short-cut proposed in Kilian (1998a). The first-order bias is estimated using the asymptotic closed-form solutions proposed by Pope (1990) rather than the bootstrap method. For a detailed description see, e.g., Kilian (1998b).

<sup>8</sup>Preliminary simulation experiments suggested that treating  $\Sigma_e^*$  as random improved the coverage accuracy of intervals compared with intervals based on the initial point estimate  $\hat{\Sigma}_e^{VAR}$  for all bootstrap replications.

be constructed, conditional on the data, as  $P\left(\hat{\theta}_{ij,h,\alpha/2}^{LP(q)*} \leq \theta_{ij,h}^{LP(q)} \leq \hat{\theta}_{ij,h,(1-\alpha/2)}^{LP(q)*}\right) = 1 - \alpha$ , where  $\hat{\theta}_{ij,h,\alpha/2}^{LP(q)*}$  and  $\hat{\theta}_{ij,h,(1-\alpha/2)}^{LP(q)*}$  are the  $\alpha/2$  and  $1 - \alpha/2$  quantiles of the distribution of  $\hat{\theta}_{ij,h}^{LP(q)*}$ . Under asymptotic normality, this interval provides a valid first-order approximation (see Efron and Tibshirani 1993).<sup>9</sup> The LP bootstrap results shown in this paper further incorporate a bias correction of the slopes analogous to the bias correction in Kilian (1998). The bootstrap bias estimate is constructed from blocks of the tuples of the original data and applied to all bootstrap slope estimates as in Kilian (1998a).<sup>10</sup> We do not report bootstrap estimates without bias corrections, but note that - unlike in the VAR model - bias corrections improve the performance of the LP bootstrap only slightly. We use 1,000 bootstrap draws for bias estimation and 2,000 bootstrap draws to construct the empirical distribution of the estimator.

### 3 Simulation Evidence: VAR(1) Model

In this section, we perform a simulation study based on a set of stationary VAR(1)-DGPs. The results will help build intuition before we turn to a more realistic DGP in the next section. The population model is:

$$y_t = \begin{pmatrix} B_{11} & 0 \\ 0.5 & 0.5 \end{pmatrix} y_{t-1} + e_t, \quad e_t \stackrel{iid}{\sim} N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix} \right) \quad (12)$$

where  $B_{11} \in \{0.5, 0.9, 0.97\}$ . The intercept has been normalized to zero in population. This DGP has been widely used in the literature as a benchmark (see, e.g., Fachin and Bravetti 1996, Kilian 1998a,b; Berkowitz and Kilian 2000). We draw 1,000 time series of length  $T = 100$  from this process.<sup>11</sup> For each trial, we fit the VAR model and a sequence of LP

---

<sup>9</sup>In principle, an alternative could have been to construct a symmetric percentile- $t$  interval. The problem with this proposal is that it is not clear how to estimate the variance of the bootstrap LP impulse response estimator in constructing the studentized statistic  $t^*$ . The Newey-West estimator of Jordà (2007) cannot be used since we rely on the block bootstrap for generating bootstrap draws. Nor can the variance of the estimator be simulated by the bootstrap method, conditional on a given realization of the block-bootstrap estimator. For that reason we do not consider the percentile- $t$  interval.

<sup>10</sup>We also experimented with reestimating the bias in each bootstrap loop, as proposed in Kilian (1998b, 1999), using a resample of the blocks used in generating the initial bootstrap slope estimate. This procedure produced very similar results at considerable added computational cost, so only results for the simpler procedure are reported.

<sup>11</sup>Initially, we also considered a sample size of  $T = 50$ . In that case, the  $\Sigma^*$  draws required for the LP bootstrap method sometimes are not positive definite, making it impossible to implement the LP method. Hence, we focus on the larger sample size.

models, one each for each horizon. All regression models include an intercept.<sup>12</sup> The lag-order for each local projection is chosen by the Akaike information criterion (AIC) with an upper bound of four lags.<sup>13</sup> The same criterion and upper bound is used in selecting the lag order of the fitted VAR model. For each of the impulse response estimates  $\hat{\theta}_{ij,h}^{VAR(p)}$  and  $\hat{\theta}_{ij,h}^{LP(q)}$  we construct the pointwise confidence intervals discussed earlier. For the block bootstrap method, the size of the block is set to four at all horizons. This produces the most accurate results for the LP intervals.<sup>14</sup>

Figure 1 shows some representative results for alternative values of  $B_{11}$ . The larger  $B_{11}$ , the higher the persistence of the process. The upper panel of Figure 1 plots the effective coverage rates of nominal 95% confidence interval for  $\theta_{21,h}$ , where  $\theta_{21,h}$  stands for the response of variable 2 to structural shock 1 in period 0 at horizon  $h = 0, \dots, 16$ . The difference in performance between the four methods is substantial. The effective coverage rates of the standard VAR delta method interval drops quickly with increasing horizon, consistent with earlier results in the literature. At horizon 16, it is 81% for  $B_{11} = 0.5$ , 69% for  $B_{11} = 0.9$  and 60% for  $B_{11} = 0.97$ . On the other hand, the effective coverage of the bias-corrected bootstrap for the VAR remains fairly close to 95% at all time horizons and for all values of  $B_{11}$ .

Unlike the coverage accuracy of the asymptotic VAR interval, that of the asymptotic LP interval does not deteriorate steadily as the horizon increases, but its overall coverage accuracy declines somewhat with increasing  $B_{11}$ . Whereas the asymptotic LP interval tends to attain close to nominal coverage for  $B_{11} = 0.5$ , its accuracy may drop as low as 91% for  $B_{11} = 0.9$  and 89% for  $B_{11} = 0.97$ . While far from perfect, these results are clearly superior to the VAR delta method.<sup>15</sup> Even higher accuracy is achieved with the bias-adjusted VAR

---

<sup>12</sup>We follow Jordà's (2005) proposal of fitting individual projections rather than doing one projection for all horizons jointly. In additional sensitivity analysis we determined that the joint linear projection is less accurate than the individual linear projections in small samples.

<sup>13</sup>The local projections with the lag order selected for each horizon have better small sample properties than the local projections using the same lag order for all horizons. The AIC with the maximum lag order of four performs better than the AIC with the upper bound of eight lags or the SIC with either a maximum lag order of four or eight. A natural conjecture is that the performance of the LP method may be improved by enforcing greater parsimony. For example, one could set  $q = 1$  in all LP regressions. Further analysis (not shown) suggests that this modification may greatly reduce the coverage accuracy of the LP interval in practice.

<sup>14</sup>We also investigated whether the LP bootstrap interval performed better when allowing for different block sizes by horizon, and found that a fixed block size produces more accurate intervals.

<sup>15</sup>It may be tempting to attribute this result to the fact that the LP impulse response estimator does not require any non-linear transformations and hence may be less biased. This is not the case. It can be shown that the bias of the LP impulse response estimator is actually similar at short horizons and larger than the bias of the VAR estimator at long horizons. Instead, the reason for the superior coverage accuracy of

bootstrap interval which does fairly well at all horizons with a tendency for its coverage accuracy to decline slightly as  $B_{11}$  increases. In contrast, the bias-adjusted LP bootstrap interval tends to be more accurate than the asymptotic VAR interval, yet less accurate than either the asymptotic LP interval or the bias-adjusted VAR bootstrap interval.

As the sample size increases, the performance of the LP point and interval estimators improve. Doubling the sample size from  $T = 100$  to  $T = 200$  can be shown to reduce greatly the bias and variance of the LP estimator and to improve the accuracy of the asymptotic LP interval to near nominal coverage even for  $B_{11} = 0.97$ , while reducing its average length. Although the asymptotic LP intervals for  $T = 200$  are about as accurate as the VAR bootstrap intervals, they remain about three times as wide on average, however. Thus, one would still prefer the VAR-based bootstrap interval for inference.

#### 4 Simulation Evidence: VAR(12) Model

Our main conclusions based on the stylized VAR(1) DGP continue to hold in a more realistic example with many lags and variables. This DGP is a prototypical partially identified four-variable VAR model of the type commonly employed in the analysis of monetary policy shocks (see, e.g., Christiano, Eichenbaum and Evans 1999). We postulate a VAR(12) model with intercept for  $y_t = [gap_t, \pi_t, \pi_t^{RPCOM}, i_t]'$ . Underlying this model is the notion that the Federal Reserve sets the interest rate ( $i_t$ ), conditional on all past data, as a function of the current inflation rate ( $\pi_t$ ) and output gap ( $gap_t$ ). We follow the literature in augmenting the model with the growth rate in real industrial commodity prices, as a leading indicator of inflationary pressures ( $\pi_t^{RPCOM}$ ). The presumption is that this additional variable helps alleviate the well-known price puzzle. The model is semi-structural in that only the monetary policy shock is identified. As is standard in this literature, the identifying assumption is that there is no contemporaneous feedback from policy decisions to the output gap, to commodity prices, or to the inflation rate. We specify the model at monthly frequency since the identifying assumptions are more credible at monthly than at quarterly frequency. Our measure of inflation is based on the seasonally adjusted monthly CPI for all urban consumers. Real commodity price inflation is constructed as the change in the Commodity Research Board's price index for raw industrials adjusted for CPI inflation. The Federal Funds rates serves as our proxy for the interest rate. The measure of the real output gap

---

the LP interval is the higher variability of the LP estimator, especially at long horizons. This difference is reflected in a substantial increase in the average interval length, as shown in the lower panel of Figure 1. The increase in variability is consistent with the less parametric nature of the LP estimator.

is the CFNAI, a principal components index U.S. real economic activity, constructed by the Federal Reserve Bank of Chicago.<sup>16</sup>

Figure 2 shows effective coverage rates and average lengths of the four confidence intervals of interest. These are obtained by generating 1,000 trials of the same length as the original data from the VAR(12) model fitted on the actual data. In that simulation exercise, the VAR error term is assumed to be Gaussian. The lag orders of the fitted regression models are obtained using the AIC with an upper bound of twelve lags. We focus on the responses of the output gap, of CPI inflation and of real commodity price inflation to an unanticipated monetary policy tightening. Figure 2 reveals some coverage deficiencies of the LP bootstrap interval especially at long horizons. Its coverage rates may drop as low as 76%. The asymptotic LP interval is fairly accurate at short horizons (in the sense that its coverage probability if anything tends to exceed the nominal coverage rate), but its coverage accuracy also deteriorates at long horizons. For example, for the output gap its coverage rate may drop as low as 90%, for the inflation rate as low as 88% and for real commodity price inflation as low as 82%. In contrast, both the asymptotic VAR interval and the bias-adjusted bootstrap VAR interval are consistently quite accurate. If anything, their coverage is excessive. Moreover, there is little to choose between the two VAR intervals in terms of their average length. The VAR intervals not only tend to be more accurate, but also systematically shorter than the LP intervals. The asymptotic LP interval is sometimes three times as wide on average as the other intervals. Even the bootstrap LP interval, however, tends to be wider than the VAR intervals.<sup>17</sup>

## 5 Simulation Evidence: VARMA(1,1) Models

An interesting question is whether the relative performance of the LP and VAR approaches changes when the data are generated from linear stationary processes that cannot be represented as finite-order VAR models. This class of models includes invertible VARMA

---

<sup>16</sup>This system is similar to models discussed in Christiano et al. (1999) and closely resembles the VAR model of Bernanke and Gertler (1995). One difference between their models and this model is that our measure of output is broader and clearly stationary and that in our model the price level is specified in log-differences. This transformation ensures that the model is stationary, if still persistent. That fact is important since the maintained assumption in this paper is stationarity. The sample period is January 1970 through December 2007. We also repeated our analysis for subsamples. The simulation results were qualitatively similar and are not reported.

<sup>17</sup>As in the case of the VAR(1) model, the excess width of the LP intervals reflects the higher variance of the LP estimator. It can be shown that the LP estimator has not only higher variance in the VAR(12) case, but typically higher bias as well, resulting in unambiguously higher MSEs. These results are very much consistent with the insights obtained from the stylized VAR(1) model, but the LP bias is more pronounced.

processes. In this section, we provide additional evidence based on a tri-variate VARMA(1,1)-DGP used as an example in Braun and Mittnik (1993) and Inoue and Kilian (2002). The implied VAR( $\infty$ ) processes may be approximated by a sequence of finite-lag order vector autoregressions (see, e.g., Lütkepohl and Poskitt 1991; Inoue and Kilian 2002). This affects the construction of the VAR delta method interval, but not that of VAR bootstrap intervals or LP intervals. The DGP includes quarterly investment growth, deflator inflation and the commercial paper rate in this order. We postulate that

$$y_t = A_1 y_{t-1} + \varepsilon_t + M_1 \varepsilon_{t-1},$$

$$\text{where } A_1 = \begin{bmatrix} 0.5417 & -0.1971 & -0.9395 \\ 0.0400 & 0.9677 & 0.0323 \\ -0.0015 & 0.0829 & 0.8080 \end{bmatrix}, M_1 = \begin{bmatrix} -0.1428 & -1.5133 & -0.7053 \\ -0.0202 & 0.0309 & 0.1561 \\ 0.0227 & 0.1178 & -0.0153 \end{bmatrix},$$

$$\text{and } \varepsilon_t \sim NID(0, PP') \text{ with } P = \begin{bmatrix} 9.2352 & 0 & 0 \\ -1.4343 & 3.6070 & 0 \\ -0.7756 & 1.2296 & 2.7555 \end{bmatrix}.$$

We focus on the response of the model variables to an innovation in the commercial paper rate. We draw 1,000 time series of length  $T = 200$  from this process. Figure 3 shows results for approximating lag orders of  $p = 5$  and  $q = 5$ .<sup>18</sup> The bias-adjusted VAR bootstrap method performs quite well. Based on the approximating VAR(5) model, its coverage rate is near the nominal coverage for all three sets of responses. Figure 3 suggests that, for a suitably large choice of  $q$ , the asymptotic LP interval attains the same coverage accuracy, but its average width is systematically wider by a factor of about three. Thus, the VARMA-DGP results are quite similar to the results we obtained earlier for finite-lag order VAR models for large  $T$ . This tentative evidence suggests that it is not clear that there are advantages to the LP approach, even in the VARMA setting.

One might conjecture that the magnitude of the MA component in the VARMA DGP could affect the relative performance of the LP and (sieve) VAR approaches in finite samples. Given the difficulty of parameterizing the magnitude of the MA component in the vector setting, we address this issue with the help of a ARMA(1,1) simulation example in which we can systematically compare the LP and VAR methods, as we vary the magnitude of the MA

---

<sup>18</sup>The simulation results are robust to changes in these lag orders. Only for very low lag orders, the coverage accuracy deteriorates.

coefficient, for a given sample size. We set  $T = 200$ , corresponding to 50 years of quarterly data (the approximate length of postwar macroeconomic data sets) or somewhat more than 16 years of monthly data. The DGP is  $y_t = 0.9y_{t-1} + \varepsilon_t + M_1\varepsilon_{t-1}$  where  $\varepsilon_t \sim NID(0, 1)$  and  $M_1 \in \{0, 0.25, 0.5, 0.75\}$ . For the AR approximation we choose lag orders that have been found to work well for a wide range of data-based ARMA DGPs in Berkowitz et al. (1999) and Inoue and Kilian (2002) – just as one would have done in applied work. Again the results are not sensitive to slight changes in the lag order. Figure 4 confirms that the relative performance of the four methods is remarkably robust even for very large MA coefficients.

## 6 Conclusion

Local projections methods are a promising recent development in the literature on impulse response analysis, but little is known about their merits relative to more conventional VAR-based methods. In this paper, we compared the finite-sample performance of impulse response confidence intervals based on local (linear) projections and VAR models in linear stationary settings. When the data are generated by finite-dimensional VAR processes, we found that the asymptotic LP interval often is less accurate than the bias-adjusted VAR bootstrap interval, notwithstanding its excessive average length. Although the accuracy of asymptotic LP interval quickly improves with increasing sample size, its average length far exceeds that of bias-adjusted VAR bootstrap intervals with comparable accuracy. Thus, even for large samples there are no apparent advantages to the LP method. Bootstrap LP intervals (with or without bias correction) and asymptotic VAR intervals are shorter on average than asymptotic LP intervals, but often lack coverage accuracy in finite samples. Similar result hold even when the data are generated by a stationary VARMA process.

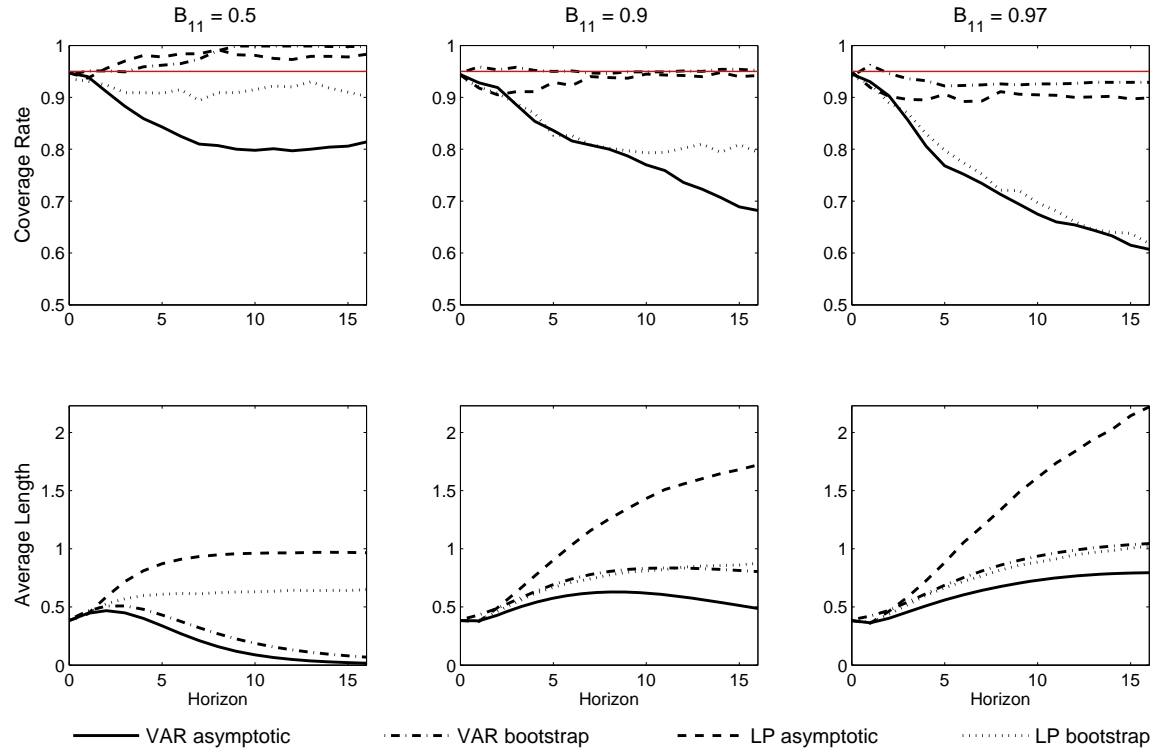
## References

- [1] Berkowitz, J., I. Birgean, and L. Kilian (1999), “On the Finite-Sample Accuracy of Non-parametric Resampling Algorithms for Economic Time Series,” in T.B. Fomby and C. Hill (eds.): *Advances in Econometrics: Applying Kernel and Nonparametric Estimation to Economic Topics*, vol. 14, 1999, JAI Press, Connecticut, 77-107.
- [2] Berkowitz, J., and L. Kilian (2000), “Recent Developments in Bootstrapping Time Series,” *Econometric Reviews*, 19, 1-48.

- [3] Bernanke, B.S. and M. Gertler (1995), “Inside the Black Box: The Credit Channel of Monetary Policy Transmission,” *Journal of Economic Perspectives*, 9, 27-48
- [4] Braun, P. A., and S. Mittnik (1993), “Misspecifications in Vector Autoregressions and Their Effects on Impulse Responses and Variance Decompositions,” *Journal of Econometrics*, 59, 319-41.
- [5] Chang, P., and S. Sakata (2007), “Estimation of Impulse Response Functions Using Long Autoregressions,” *Econometrics Journal*, 10, 453-469.
- [6] Christiano, L.J., M. Eichenbaum and C.L. Evans (1999), “Monetary Policy Shocks: What Have We Learned and to What End?” in: *Handbook of Macroeconomics*, 1, Edited by J.B. Taylor and M. Woodford.
- [7] Fachin, S. and L. Bravetti (1996), “Asymptotic Normal and Bootstrap Inference in Structural VAR Analysis,” *Journal of Forecasting*, 15, 329-341.
- [8] Goncalves, S., and L. Kilian (2004), “Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form,” *Journal of Econometrics*, 123, 89-120.
- [9] Goncalves, S., and L. Kilian (2007), “Asymptotic and Bootstrap Inference for  $AR(\infty)$  Processes with Conditional Heteroskedasticity,” *Econometric Reviews*, 26, 609-641.
- [10] Haug, A.A., and C. Smith (2007), “Local Linear Impulse Responses for a Small Open Economy”, Economics Discussion Paper No. 0707, University of Otago.
- [11] Hall, A., A. Inoue, J.M. Nason, and B. Rossi (2009), “Information Criteria for Impulse Response Function Matching Estimation of DSGE Models”, mimeo, Duke University.
- [12] Inoue, A., and L. Kilian (2002), “Bootstrapping Smooth Functions of Slope Parameters and Innovation Variances in  $VAR(\infty)$  Models,” *International Economic Review*, 43, 309-332.
- [13] Jordà, Ò. (2005), “Estimation and Inference of Impulse Responses by Local Projections,” *American Economic Review*, 95, 161-182.
- [14] Jordà, Ò. (2009), “Simultaneous Confidence Regions for Impulse Responses,” *Review of Economics and Statistics*, 91, 629-647.

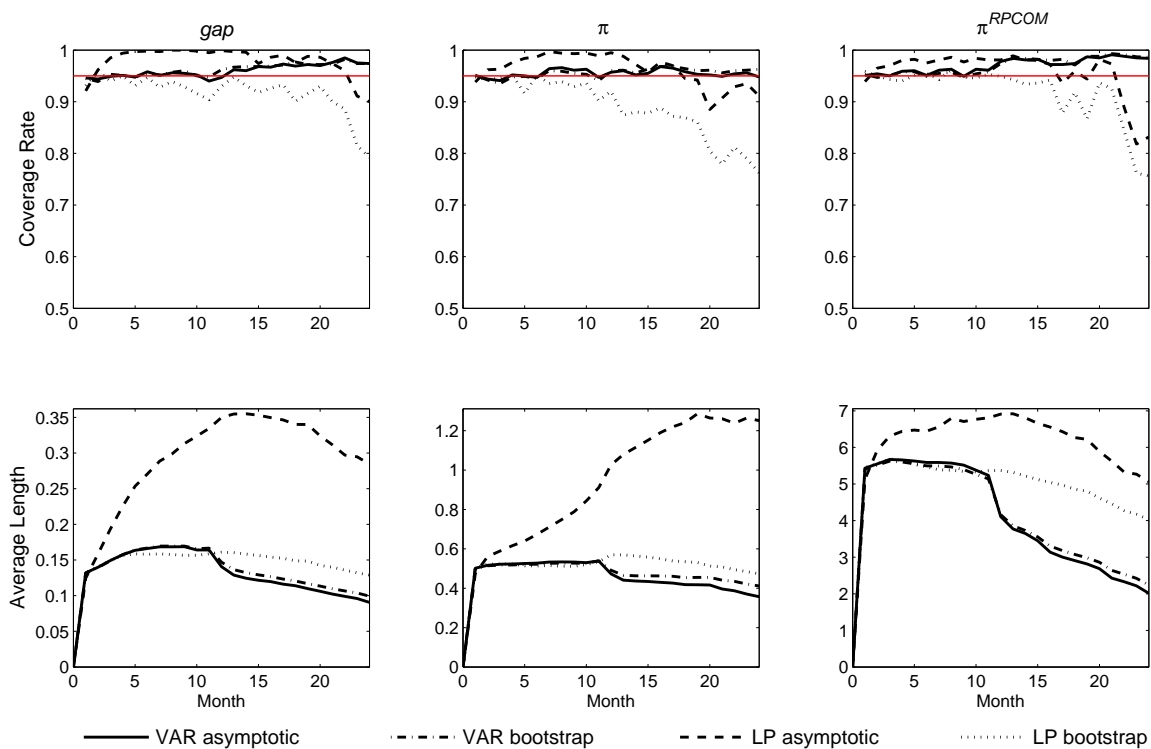
- [15] Jordà, Ò., and S. Kozicki (2007), “Estimation and Inference by the Method of Projection Minimum Distance,” mimeo, Bank of Canada.
- [16] Jordà, Ò., and M. Marcellino (2009), “Path Forecast Evaluation,” *Journal of Applied Econometrics*, forthcoming.
- [17] Kilian, L. (1998a), “Small-sample confidence intervals for impulse response functions,” *Review of Economics and Statistics*, 80, 218–230.
- [18] Kilian, L. (1998b), “Confidence intervals for impulse responses under departures from normality,” *Econometric Reviews*, 17, 1–29.
- [19] Kilian, L. (1998c), “Accounting for Lag Order Uncertainty in Autoregressions: The Endogenous Lag Order Bootstrap Algorithm,” *Journal of Time Series Analysis*, 19, 531–548.
- [20] Kilian, L. (1999), “Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses,” *Review of Economics and Statistics*, 81, 652–660.
- [21] Lütkepohl, H. (1990), “Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models,” *Review of Economics and Statistics*, 72, 116–125.
- [22] Lütkepohl, H., and D.S. Poskitt (1991), “Estimating orthogonal impulse responses via vector autoregressive models,” *Econometric Theory*, 7, 487–496.
- [23] Pope, A.L. (1990), “Biases of Estimators in Multivariate Non-Gaussian Autoregressions,” *Journal of Time Series Analysis*, 1, 249–258.

Figure 1: Coverage Rates and Average Lengths of 95% Confidence Intervals for  $\theta_{21,h}$  in the VAR(1) Model



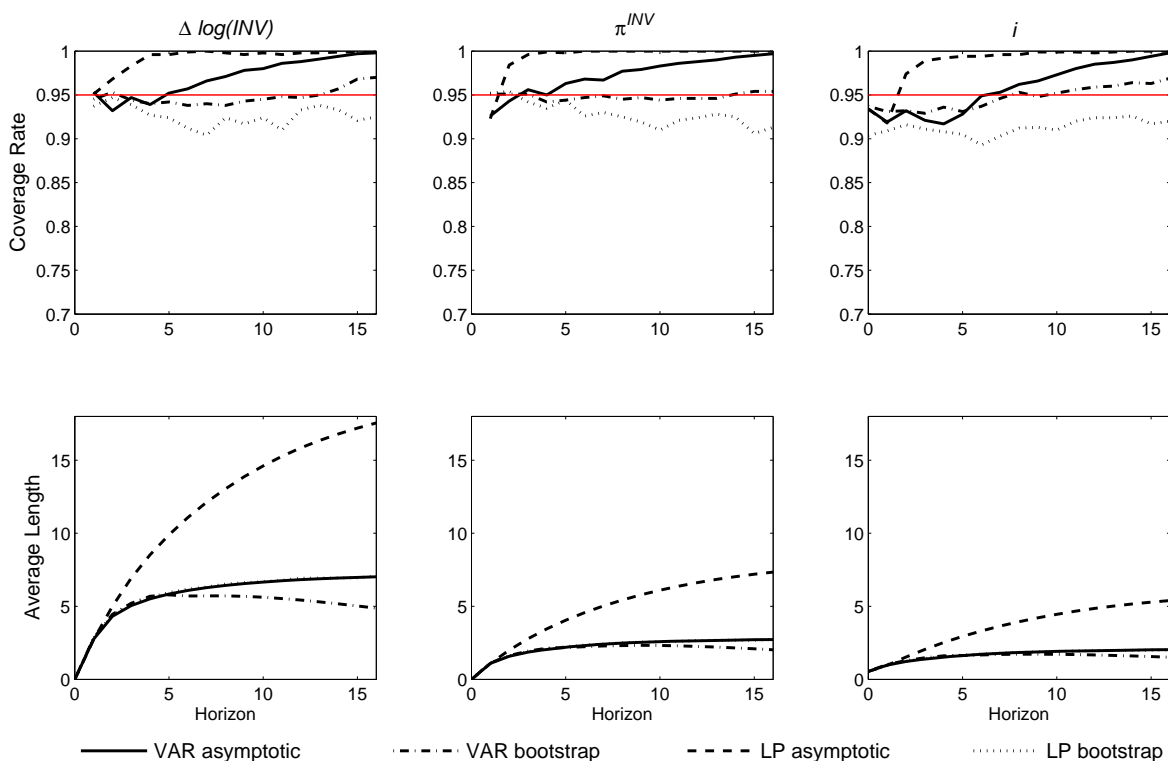
Notes: Simulation results based on 1,000 trials of length  $T = 100$  from the VAR(1) DGP described in the text. *VAR asymptotic* denotes the asymptotic delta method for VAR impulse responses. *VAR bootstrap* refers to the bias-corrected bootstrap method for VAR impulse responses discussed in Kilian (1998b, 1999). *VAR asymptotic* refers to the asymptotic delta method interval in Lütkepohl (1990). *LP asymptotic* denotes the asymptotic interval for LPs. *LP bootstrap* refers to the bias-corrected block bootstrap interval for LPs. All lag orders are selected by the AIC with an upper bound of four lags for all methods.

Figure 2: Coverage Rates and Average Lengths of 95% Confidence Intervals for Responses to an Interest Rate Innovation in the VAR(12) Model



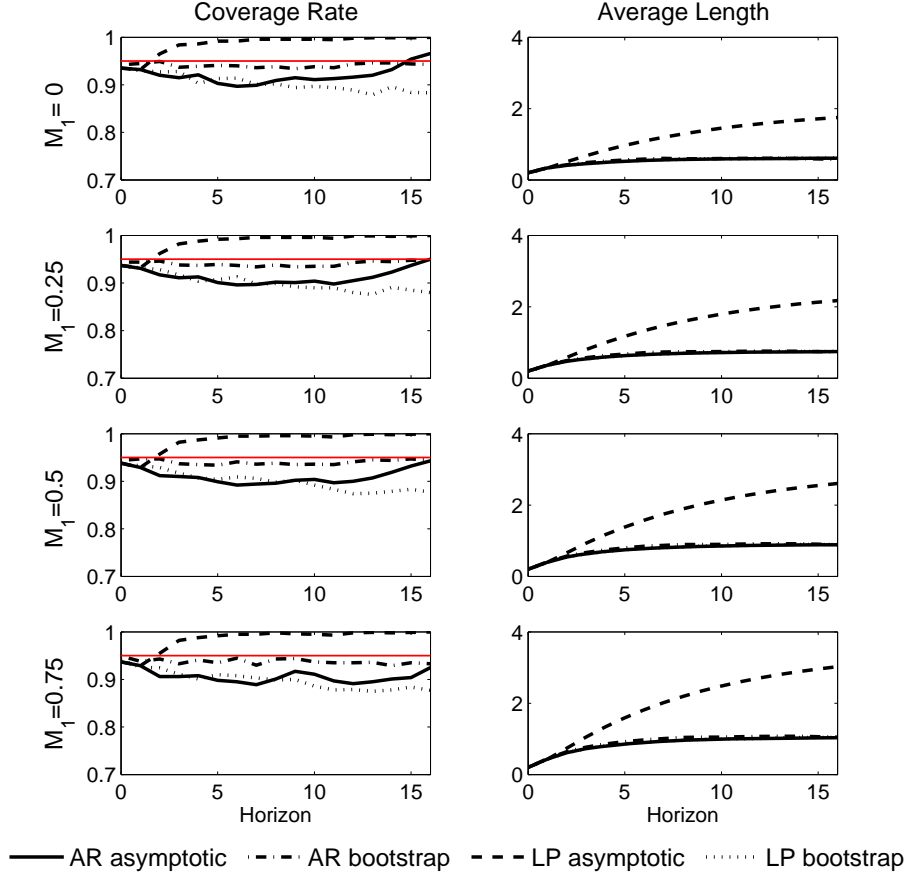
Notes: See Figure 1. The results are based on 1,000 trials of length  $T = 456$  from the VAR(12) DGP described in text.  $Gap$  denotes the CFNAI index of U.S. real activity,  $\pi$  is U.S. CPI inflation, and  $\pi^{RPCOM}$  is real commodity price inflation. Lag orders are selected by the AIC with an upper bound of 12 lags. Since there is no uncertainty about the impact response of these variables, we do not construct a coverage rate for horizon 0.

Figure 3: Coverage Rates and Average Lengths of 95% Confidence Intervals for Responses to an Interest Rate Innovation in the VARMA(1,1) Model



Notes: Simulation results for  $T = 200$  based on 1,000 trials from the VARMA(1,1)-DGP described in text. *VAR asymptotic* denotes the modified asymptotic delta method for VAR impulse responses described in Lütkepohl and Poskitt (1991). *VAR bootstrap* refers to the bias-corrected bootstrap method for VAR impulse responses, which accommodates departures from finite-dimensional DGPs without change (see Inoue and Kilian 2002). *LP asymptotic* and *LP bootstrap* are defined as in Figure 1. The approximating lag orders are  $p = 5$  and  $q = 5$ . The results are robust to reasonable changes in the lag order, as long as the lag orders are not too small.

Figure 4: Coverage Rates and Average Lengths of 95% Impulse Response Confidence Intervals in the ARMA(1,1) Model



Notes: See Figure 3. Simulation results based on 1,000 trials of length  $T=200$  from  $y_t = 0.9y_{t-1} + \varepsilon_t + M_1\varepsilon_{t-1}$  where  $\varepsilon_t \sim NID(0, 1)$ .  $M_1 \in \{0, 0.25, 0.5, 0.75\}$ . The approximating lag orders are  $p = 8$  and  $q = 8$ . The results are robust to reasonable changes in the lag order, as long as the lag orders are not too small.