

DISCUSSION PAPER SERIES

No. 7266

DO LOCAL PROJECTIONS SOLVE THE BIAS PROBLEM IN IMPULSE RESPONSE INFERENCE?

Lutz Kilian and Yun Jung Kim

INTERNATIONAL MACROECONOMICS



Centre for Economic Policy Research

www.cepr.org

Available online at:

www.cepr.org/pubs/dps/DP7266.asp

DO LOCAL PROJECTIONS SOLVE THE BIAS PROBLEM IN IMPULSE RESPONSE INFERENCE?

Lutz Kilian, University of Michigan and CEPR
Yun Jung Kim, University of Michigan

Discussion Paper No. 7266
April 2009

Centre for Economic Policy Research
53–56 Gt Sutton St, London EC1V 0DG, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INTERNATIONAL MACROECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Lutz Kilian and Yun Jung Kim

ABSTRACT

Do Local Projections Solve the Bias Problem in Impulse Response Inference?

It is well documented that the small-sample accuracy of asymptotic and bootstrap approximations to the pointwise distribution of VAR impulse response estimators is undermined by the estimator's bias. A natural conjecture is that impulse response estimators based on the local projection (LP) method of Jordà (2005, 2007) are less susceptible to this problem and hence potentially more reliable in small samples than VAR-based estimators. We show that - contrary to this conjecture - LP estimators tend to have both higher bias and higher variance, resulting in pointwise impulse response confidence intervals that are typically less accurate and wider on average than suitably constructed VAR-based intervals. Bootstrapping the LP estimator only worsens its finite-sample accuracy. We also evaluate recently proposed joint asymptotic intervals for VAR and LP impulse response functions. Our analysis suggests that the accuracy of joint intervals can be erratic in practice, and neither joint interval is uniformly preferred over the other.

JEL Classification: C32, C52 and C53

Keywords: bias, confidence interval, impulse response function, joint interval, local projection and vector autoregression

Lutz Kilian
Department of Economics
University of Michigan
238 Lorch Hall
611 Tappan Street
Ann Arbor, MI 48109-1220
USA

Email: lkilian@umich.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=132598

Yun Jung Kim
Department of Economics
University of Michigan
611 Tappan Street
Ann Arbor, MI 48109-1220
USA

Email: yunjungk@umich.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=169714

Submitted 06 April 2009

1 Introduction

Estimates of structural impulse response functions are of central interest in empirical macroeconomics. Conventional estimates of structural impulse responses are based on vector autoregressive (VAR) models. VAR impulse response coefficients are nonlinear functions of the estimates of the VAR slope parameters and of the VAR innovation variance-covariance matrix. It is well known that the finite-sample accuracy of conventional asymptotic and bootstrap approximations to the distribution of the impulse response estimator is undermined by the bias of the impulse response estimator. This bias arises from two distinct sources: the small-sample bias of the estimates of the VAR slope parameters and the additional bias induced by the non-linear transformations of the estimated parameters.

Over the last two decades, several econometric methods have been proposed to address this problem. For example, the bias-corrected bootstrap confidence interval of Kilian (1998a) explicitly accounts for the first source of bias by constructing bias-corrected slope coefficient estimates. This bias-adjusted bootstrap provides a significant improvement in coverage accuracy over the standard bootstrap of Runkle (1987) and the asymptotic delta method interval of Lütkepohl (1990), but it does not necessarily yield accurate coverage when the process is highly persistent or when the model includes deterministic time trends.¹ A large number of studies has investigated the small-sample and asymptotic accuracy of these methods (see, e.g., Griffiths and Lütkepohl (1993), Fachin and Bravetti (1996), Kilian (1998a,b,c; 1999), Berkowitz and Kilian (2000), Benkowitz, Lütkepohl and Neumann (2000), and Kilian (2001)).²

Kilian and Chang (2000) demonstrated that the coverage accuracy of all traditional asymptotic and bootstrap methods, including bias-adjusted methods, tends to deteriorate in large-dimensional VAR models at longer horizons, when the data are highly persistent. This finding motivated the use of nonstandard asymptotic approximations based on local-to-unity models. Since the bias of the impulse responses typically worsens, as the dominant autoregressive root approaches unity, that approach seemed well suited to dealing with the bias problem. For example, Wright (2000), building on Stock (1991), proposed conservative asymptotic impulse response confidence intervals based on local-to-unity approximations to the largest root of the autoregressive process. His method proved computationally intractable

¹An alternative bias-adjusted method for impulse responses in univariate autoregressions has been proposed by Andrews and Chen (1994). Related work also includes Rudebusch (1992).

²Related studies focusing on univariate autoregressions include Berkowitz, Birgean and Kilian (1999), Inoue and Kilian (2002a, 2003), and Pesavento and Rossi (2007).

in VAR models, however. In related work, Gospodinov (2004) proposed an asymptotic impulse response confidence interval based on the inversion of the likelihood ratio statistic. His method differs from Wright (2000) in that his interval is not conservative, but asymptotically exact. Gospodinov's approach, however, is limited to univariate autoregressions and requires knowledge of a point null which is unlikely to be available in practice. As yet another alternative, Hansen (1999) proposed a grid bootstrap method for the dominant root in univariate autoregressive processes. This grid bootstrap provides correct asymptotic coverage regardless of whether the autoregressive model is near-integrated or exactly integrated. Gospodinov (2004), however, reports that the coverage accuracy of the grid bootstrap method applied to impulse responses is too low at short horizons. Pesavento and Rossi (2006) were the first to propose operational confidence intervals for VAR impulse responses based on a local-to-unity approximation to the asymptotic distribution of that estimator. Their method is designed to achieve accurate coverage at long horizons when the data are highly persistent. At short horizons, its coverage accuracy may not be satisfactory. Pesavento and Rossi (2006) also proposed another method designed to yield accurate coverage at both short and long horizons, but that method may be conservative at medium horizons. Thus, even twenty years after the first confidence intervals were developed for VAR impulse responses, there is no single method that resolves the bias problem in all situations.

Our paper explores a new idea for dealing with this bias problem that is motivated by recent advances in estimating impulse response functions from local projections. Local (linear) projections (LPs) were proposed by Jordà (2005, 2007) on the grounds that such projections may be more robust to model misspecification.³ We instead focus on another potentially attractive feature of that method, namely its ability to ameliorate the bias problem that has undermined the accuracy of inference on impulse responses in practice. The basic idea of the LP method is that we directly estimate a sequence of linear projections of the future value of the dependent variable on the current information set. This approach will be asymptotically equivalent to the VAR-based approach, provided the data generating process is stationary and linear. Unlike VAR impulse response estimates, however, impulse responses based on local projections do not require any nonlinear transformations of the estimated slope parameters. Rather the slope parameters themselves are the reduced-form impulse response coefficients and - with the help of any estimate of the structural impact multiplier

³For another empirical application of LP methods to impulse response estimation see Haug and Smith (2007). Hall, Inoue, Nason and Rossi (2007) 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).

matrix - can easily be used to construct the structural impulse response function. Since local projections do not involve any nonlinear transformation, they are likely to be better approximated by Gaussian distributions. A very similar point has been made by Davidson (2000) in the context of bootstrap methods. Thus, local projection methods have the potential of greatly reducing the bias of impulse response point estimates and of increasing the coverage accuracy of impulse response confidence intervals.

The LP method is not without drawbacks, however. One of its potential disadvantages is that the LP estimator tends to have higher variance, when the data generating process is well approximated by a vector autoregression, since local projections impose less structure on the estimation problem. This increase in the variance would be expected to increase the mean-squared error (MSE) of impulse response point estimates and the average length of LP impulse response confidence intervals. Moreover, nothing is known about the extent of small-sample bias in the slope parameters of local projections compared with the well-known small-sample bias in the VAR slope parameters (see Pope 1990). Hence, the question of whether local projections should replace VAR models in constructing structural impulse response point and interval estimates is ultimately an empirical question.

In this paper, we address this question by comparing the finite-sample properties of impulse response confidence intervals estimated by VAR models and by local projections using simulation studies. Our objective 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 most reliable in practice, when the VAR framework is considered a good approximation to the data generating process. Throughout the paper we will maintain the assumption of stationarity. A summary of the alternative approaches considered in this paper is provided in Exhibit 1.

We make three distinct contributions. First, for each method, we compare the pointwise coverage accuracy and average length of the asymptotic and bootstrap confidence interval in stationary models. Our analysis allows for model specification uncertainty. Specifically, for the VAR model 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, in contrast, we investigate (3) the asymptotic interval proposed by Jordà (2005, 2007) and (4) we propose a bootstrap percentile interval based on a suitably designed block bootstrap method.⁴ We provide some intuition for the relative

⁴Although Jordà (2007) discusses the potential benefits from bootstrapping local projections, he does not explore bootstrap methods in his work. Our bootstrap percentile interval provides an alternative to

performance of these approaches by investigating the bias, standard deviation, and MSE of the impulse response point estimators. Our simulation analysis is based on a commonly used stylized bivariate VAR(1) data generating process as well as an additional high-dimensional VAR data generating process of the type used in studying responses to monetary policy shocks. Such realistic models are rarely analyzed in the literature, given the computational costs of simulating large-dimensional VAR systems.⁵

We find that, contrary to our conjecture, the LP point estimator tends to be both more biased and more variable than VAR based estimator, resulting in often excessively wide LP intervals with less than nominal coverage. Given its greater average width, the asymptotic LP interval tends to be more accurate than the VAR delta method interval. It also is more accurate than the bootstrap percentile LP interval, although again at the cost of added width. The reason is that standard applications of the bootstrap tend to reinforce the small-sample bias, much like in the Runkle (1987) method (see Kilian 1998a). Unlike in the VAR context, there is no obvious way of correcting for this bias in the LP estimator. The bias-adjusted VAR bootstrap interval, in contrast, tends to be more accurate than either LP interval in small samples and typically shorter on average. It also is more accurate than traditional asymptotic delta method intervals and only slightly wider on average, consistent with previous findings in the literature. Thus, for pointwise impulse response intervals, among the methods considered, no method is more accurate than the bias-adjusted bootstrap method for VAR models. For larger sample sizes, the LP asymptotic interval and the VAR bootstrap interval have similar coverage accuracy, but the LP interval tends to be much wider on average. Additional simulation evidence suggests that similar results hold even for VARMA data generating processes, when the finite lag order VAR model is only an approximation.

Second, in addition to evaluating the pointwise coverage accuracy of the intervals at each impulse response horizon, we also evaluate the coverage accuracy of the joint confidence interval proposed by Jordà (2007) whose small-sample properties have not been examined previously. These joint intervals can be computed from either local projections or VAR models. Joint confidence intervals are of particular interest to empirical researchers who want to evaluate the uncertainty about the entire path of the impulse response function. We focus on the asymptotic interval estimates (5) and (6). Our simulation results suggest that

asymptotic inference. We do not investigate the use of percentile- t intervals for the LP method since it is not clear how to estimate the variance of the bootstrap impulse response estimator.

⁵Kilian and Chang (2000) is a notable exception.

the coverage accuracy of the joint intervals can be erratic, regardless of the method used. Thus, joint intervals have to be used with caution. Our analysis also shows that neither interval is uniformly preferred over the other, although the joint LP interval is invariably much wider on average. Bootstrapping the standard error does not improve the accuracy of the asymptotic LP interval.

Third, we illustrate the practical differences that may arise from the choice of different methods of estimation and inference by re-examining the empirical findings for a standard monthly VAR model of monetary policy. Using data for 1970.1-2007.12, the VAR model provides evidence that monetary policy contractions cause a temporary reduction in output and a temporary drop in real commodity prices, but we find no evidence that inflation is significantly reduced. Despite the inclusion of commodity prices, the estimates exhibit the well known price puzzle. The use of joint VAR intervals does not change any of these conclusions, although it affects the degree of statistical significance especially at longer horizons. Since most statistically significant impulse response estimates are obtained at shorter horizons, when pointwise and joint intervals tend to be similar, we conclude that typically the use of joint intervals will not overturn the substantive findings of VAR studies based on pointwise intervals. Likewise, estimates from the LP and VAR method are qualitatively similar, but the LP estimates tend to be more erratic and less precisely estimated. This tendency is even more pronounced if we focus on subsamples such as the post-Volcker period. Our analysis provides no compelling reason to abandon traditional VAR methods of constructing impulse response estimates in favor of the LP method.

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) data-generating process used in the previous literature. The simulation results for a more realistic VAR(12) model are presented in section 4 along with a comparison of the empirical estimates. Section 5 contains some preliminary simulation results for infinite-order VAR processes. We conclude in Section 6.

2 Review of VARs and Local Projections

2.1 Data-Generating Process

Consider a K -dimensional linear vector autoregressive data-generating processes (DGP) of finite order p :⁶

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

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 Σ_e is non-singular and positive definite.⁷ All values of z satisfying $\det(I_K - B_1 z - \cdots - B_p z^p) = 0$ 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:

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

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

2.2 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 Σ_e . Estimates $\hat{\Theta}_h^{VAR(p)}$ are constructed by substituting the least-squares estimates of B and Σ_e obtained from regression (1).

⁶This 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.

⁷The 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).

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

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

where u_t may be serially correlated or heteroskedastic (see Jordà 2005, 2007). 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 :

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

$\Phi_0^{LP(q)} = I_K$. 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.⁸ $\theta_{ij,h}^{LP(q)}$ 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 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 $\theta_{ij,h}^{VAR(p)}$.

2.3 Confidence Intervals

2.3.1 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

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

⁸Jordà (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 $\widehat{\sigma}(\widehat{\theta}_{ij,h}^{VAR(p)})$ is the square root of element $(K(j-1)+i, K(j-1)+i)$ of $(\widehat{C}_h \widehat{\Sigma}_{\widehat{\beta}} \widehat{C}'_h + \widehat{C}_h \widehat{\Sigma}_{\widehat{\sigma}} \widehat{C}'_h)$, $z_{1-\alpha/2}$ 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

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

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

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

where $M_x = I - X(X'X)^{-1}X'$, $X = [1 \ y_{t-1} \ y_{t-2} \ \dots \ y_{t-q}]$, $\bar{G}_h = (I_K \otimes \Phi_h^{LP(q)}) \partial \text{vec}(A_0^{-1}) / \partial \sigma'$, and $\widehat{\Sigma}_u = E(u_{t+h} u_{t+h}')$ in equation (5). The first additive component of this variance-covariance matrix captures the variance of $\text{vec}(\widehat{\Phi}_h^{LP(q)} A_0^{-1})$ 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 2007).⁹ Following Jordà (2005), we employ the Newey-West estimator of $\widehat{\Sigma}_u$.¹⁰

2.3.2 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). The reader is referred to the relevant literature for details of that procedure.¹¹ For the LP impulse response estimator, no bootstrap methods

⁹Jordà (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.

¹⁰Jordà (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 Σ_u by least squares would seriously undermine the accuracy of the LP interval.

¹¹We 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).

have been considered to date. Although Jordà (2007) 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 $\hat{\Sigma}_e^{VAR}$.¹²

A nominal $(1 - \alpha)\%$ percentile confidence interval may be constructed, conditional on the data, as

$$P\left(\hat{\theta}_{ij,h,\alpha/2}^{LP(q)*} \leq \hat{\theta}_{ij,h}^{LP(q)} \leq \hat{\theta}_{ij,h,(1-\alpha/2)}^{LP(q)*}\right) = 1 - \alpha, \quad (12)$$

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). In principle, an alternative could have been to construct the symmetric percentile- t interval:

$$P\left(\hat{\theta}_{ij,h}^{LP(q)} - t_{1-\alpha}^* \frac{1}{\sqrt{T}} \hat{\sigma}\left(\hat{\theta}_{ij,h}^{LP(q)*}\right) \leq \hat{\theta}_{ij,h}^{LP(q)} \leq \hat{\theta}_{ij,h}^{LP(q)} + t_{1-\alpha}^* \frac{1}{\sqrt{T}} \hat{\sigma}\left(\hat{\theta}_{ij,h}^{LP(q)*}\right)\right) = 1 - \alpha, \quad (13)$$

where $t_{1-\alpha}^*$ denotes the $1 - \alpha$ quantile of the distribution of $\left|\hat{\theta}_{ij,h}^{LP(q)*} - \hat{\theta}_{ij,h}^{LP(q)}\right| / \hat{\sigma}\left(\hat{\theta}_{ij,h}^{LP(q)*}\right)$, and $\hat{\sigma}\left(\hat{\theta}_{ij,h}^{LP(q)*}\right)$ is an estimate of the standard deviation of $\hat{\theta}_{ij,h}^{LP(q)*}$. 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.

¹²Preliminary simulation experiments suggested that treating Σ_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.

2.3.3 Joint Confidence Intervals

In addition to the pointwise confidence intervals described so far, we consider the joint confidence regions proposed by Jordà (2007). Let the $(K(H+1) \times K)$ matrix Θ be a set of impulse responses for horizon 0 to H . Let θ denote $\text{vec}(\Theta)$ and $\hat{\theta}$ the corresponding estimator. Jordà observes that

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Omega_\theta),$$

where Ω_θ is the $(K^2(H+1) \times K^2(H+1))$ limiting variance-covariance matrix of all structural impulse response coefficients up to some horizon H . The confidence region that contains the entire impulse response path with probability $100(1 - \alpha)\%$ in repeated sampling is a multidimensional ellipsoid which cannot be displayed easily. Jordà suggests that this joint confidence region can be approximated by Scheffé's (1953) S-method. The resulting joint LP confidence interval is:

$$\left[\hat{\theta}_{ij} \pm \frac{\text{chol}(\hat{\Omega}_{\theta,ij})}{\sqrt{T}} \sqrt{\frac{c_{1-\alpha}^2}{H+1} i_{(H+1)}} \right] \quad (14)$$

where $\theta_{ij} = [\theta_{ij,0}, \theta_{ij,1}, \theta_{ij,2}, \dots, \theta_{ij,H}]'$, $c_{1-\alpha}^2$ is the critical value of the chi-square distribution with $H + 1$ degrees of freedom, i_{H+1} is a $(H + 1) \times 1$ vector of ones. Explicit expressions for $\hat{\Omega}_{\theta,ij}$ that are applicable to both VAR and LP estimators are presented in Jordà (2007). The joint variance-covariance matrix of the impulse response coefficients is evaluated using the closed-form solution of Jordà (2007):

$$\begin{aligned} \Omega_\theta &= (A_0^{-1} \otimes I_{K(H+1)})' \left((y_t' M_x y_t)^{-1} \otimes \Sigma_u \right) (A_0^{-1} \otimes I_{K(H+1)}) + \\ &\quad 2(I_K \otimes \Phi) C D_K^+ (\Sigma_e \otimes \Sigma_e) D_K^{++'} C' (I_K \otimes \Phi)', \end{aligned} \quad (15)$$

where Φ is the $(K(H+1) \times K)$ matrix of reduced-form impulse responses, $C = L_K' \{L_K (I_{K^2} + K_{KK}) (A_0^{-1} \otimes I_K) L_K'\}^{-1}$, and L_K is the elimination matrix, K_{KK} the commutation matrix, and D_K^+ is the duplication matrix defined in Lütkepohl (2005, Appendix 12.2). In practice, we substitute consistent estimators for the unknown expressions Φ , Σ_u , Σ_e and A_0^{-1} . We deal with possible serial correlation in the error term of the local projection by employing the Newey-West estimator of Σ_u with a truncation lag corresponding to the maximum impulse response horizon. Additional simulation evidence (not shown) suggests that somewhat tighter joint intervals may be obtained at the cost of more erratic coverage

accuracy by lowering the truncation lag. Alternatively, the joint variance-covariance matrix may be evaluated using the bootstrap method conditional on the original data.

In the applications below we adopt a suggestion by Jordà and Marcellino (2009) to replace expression (14) by:

$$\left[\hat{\theta}_{ij} \pm \frac{chol(\widehat{\Omega}_{\theta,ij})}{\sqrt{T}} \sqrt{\frac{c_{1-\alpha}^2(h)}{h}} \right]_{h=0}^H \quad (16)$$

Additional simulation evidence shows that this modification tends to improve the finite-sample accuracy of the joint interval.

2.4 Evaluation Criteria

We are interested in comparing the small-sample performance of impulse responses estimated by local projections and by conventional VAR methods. One set of criteria are the effective coverage accuracy and average length of pointwise impulse responses confidence intervals. Effective coverage is defined as the relative frequency with which the confidence interval covers the true, but in practice unknown value of the impulse response in repeated sampling. Average length is the average distance between the upper and lower bounds of the confidence interval in repeated trials. In addition, we report the bias, standard deviation and mean-squared error of the impulse response point estimates to help explain differences in coverage accuracy and average length. For the joint confidence interval, we calculate the probability that the interval estimator contains the entire path of the true impulse response function in repeated sampling. We also report the average interval length. Throughout the paper, we focus on nominal 95% confidence intervals. Qualitatively similar results are obtained for nominal 68% intervals.

3 Simulation Evidence: Bivariate VAR(1) Model

In this section, we perform a simulation study to evaluate the relative small-sample performance of VARs and LPs in the context of a stationary VAR(1)-DGP. The results will help build intuition before we turn to a more realistic DGP in the next section. The model is:

$$y_t = \begin{pmatrix} B_{11} & 0 \\ 0.5 & 0.5 \end{pmatrix} y_{t-1} + e_t, \quad e_t \stackrel{iid}{\sim} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix} \right) \quad (17)$$

where $B_{11} \in \{0.5, 0.9, 0.97\}$. The intercept has been normalized to zero in population. This DGP has been used widely in the literature as a benchmark (see, e.g., Griffiths and Lütkepohl

1993, Fachin and Bravetti 1996, Kilian 1998a,b; Berkowitz and Kilian 2000).

We draw 1,000 time series of length $T = 100$ from this process.¹³ For each trial, we fit the VAR model and a sequence of LP models, one each for each horizon. All regression models include an intercept.¹⁴ The lag-order for each local projection is chosen by the Akaike information criterion (AIC) with an upper bound of four lags.¹⁵ 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. All bootstrap confidence intervals are based on 2,000 bootstrap replications. 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.¹⁶ The maximum horizon H of the impulse response function is 16.

3.1 Pointwise Intervals

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} . This result illustrates that the small-sample bias in the impulse response estimates by the VAR

¹³Initially, we also considered a sample size of $T = 50$. In that case, the Σ^* 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.

¹⁴With the exception of the construction of joint confidence regions in section 3.2 and 4.1, 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.

¹⁵The 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.

¹⁶We 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.

is substantial and that the asymptotic normal approximation in the delta method interval may not be a good approximation to the impulse response distribution in small samples.

Unlike the coverage accuracy of the asymptotic VAR interval, that of the asymptotic LP interval does not deteriorate sharply 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. This finding suggests that indeed there are potential advantages to using local projections. It may be tempting to attribute these differences to the fact that the LP impulse response estimator does not require any non-linear transformations and hence is less biased. This is not the case. As Figure 2 shows, the bias of the VAR impulse response estimator is actually similar at low horizons and smaller than the bias of the LP estimator at long horizons. There is no evidence of reduced bias. Instead, the reason for the superior coverage accuracy of the LP interval is the higher variability of the LP estimator, shown in Figure 2, 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. This increase in variability is consistent with the less parametric nature of the LP estimator.

Returning to the first panel of Figure 1, we see that the bias-adjusted VAR bootstrap interval generally does fairly well at all horizons, with a tendency for the coverage accuracy to decline slightly as B_{11} increases. In contrast, the LP percentile bootstrap interval is even less accurate than the asymptotic LP interval in many cases, except at short horizons. The reason is that the bootstrap tends to amplify the bias in the initial point estimates, much like in the case of the Runkle (1987) VAR bootstrap method (see Kilian 1998a). To overcome that problem would require a bias adjustment of the LP regressions, but such an adjustment is not straightforward since there are no closed form solutions, nor does the block bootstrap method lend itself to bias adjustments.

To summarize, we find that, contrary to the conjecture in the introduction, there is no evidence that the LP impulse response estimator has lower bias than the VAR estimator, while the conjecture that the LP estimator has higher variance proved to be correct. Thus, local projections tend to deliver less accurate point estimates of impulse responses in terms of the MSE. As for inference, we found that the bias-corrected bootstrap interval for VAR models dominates all other pointwise confidence intervals in this simulation example. Moreover, there is no evidence that bootstrapping improves the accuracy of LP confidence intervals.

3.2 Joint Intervals

Table 1 investigates the coverage accuracy and average length of the corresponding asymptotic joint intervals. In computing the coverage accuracy of the nominal 95% joint interval, we compute the relative frequency with which the interval estimator includes the entire true response function within the interval bounds. Thus, there is one coverage rate for each of the four impulse response functions. We find that the joint VAR interval is typically more accurate than the joint LP interval. The coverage rate of the nominal 95% joint VAR interval ranges from 86% to 98%. The corresponding results for the joint LP interval range from 80% to 91%. At the same time, the average interval length of the joint LP interval is always greater than that of the joint VAR interval, often by a factor of more than two. Thus, neither type of joint interval necessarily comes close to nominal coverage, even in this simple model.

These estimates are based on the closed-form solution of the asymptotic variance provided in Jordà (2007). We also investigated the extent to which the accuracy of these intervals can be improved by estimating the joint variance-covariance matrix of the impulse response coefficient estimators by bootstrap methods. We found that substituting suitably constructed bootstrap estimates of this matrix based on 2,000 bootstrap draws did not improve the accuracy of either the VAR or LP interval. The relative ranking of the two methods remained unchanged. These additional results are not shown to conserve space.¹⁷

3.3 Larger Sample Sizes

The preceding analysis highlighted some practical limitations of the LP method in small samples. It is clear that, as the sample size increases, the performance of the LP point and interval estimators will improve. As expected, doubling the sample size from $T = 100$ to $T = 200$ greatly reduces the bias and variance of the LP estimator and improves the accuracy of the pointwise 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 interval for pointwise inference. For the joint intervals coverage improves for $T = 200$, but not dramatically so. The accuracy of

¹⁷We did not attempt to construct joint intervals based on the bootstrap. While it would not be difficult to simulate the joint distribution of the pointwise impulse response estimators, the resulting confidence region could not be displayed easily. It was this fact that prompted Jordà (2007) to propose the use of Scheffé's (1953) S-method resulting in the asymptotic joint interval defined in section 2.

the LP and VAR intervals becomes more similar. The joint VAR interval has coverage rates between 88% and 98%; the joint LP interval between 88% and 96%. For a given response function, the ranking by accuracy is in general ambiguous, but the joint LP interval always is considerably wider on average.

4 Four-Variable VAR(12) Model

The preceding results were confined to a stylized VAR(1) model. In this section we demonstrate that our main conclusions continue to hold in a realistic example with many lags and variables. Our example 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 (π_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 (π_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.

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.

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. Our measure of the real output gap is the CFNAI, a weighted average of a multitude of monthly indicators of U.S. real economic

activity, constructed by the Federal Reserve Bank of Chicago. This is a principal components index based on 85 real indicators including measures of production, income, employment, and consumption. It is constructed to have an average value of zero and a standard deviation of one and is stationary by construction. The CFNAI can be interpreted as a measure of the U.S. business cycle. The use of the CFNAI has several advantages relative to other output measures. First, real GDP data are not available at monthly frequency and industrial production data capture only a small and declining share of output. Second, it is well established that the Federal Reserve considers many measures of real output rather than one time series only (see, e.g., Evans 1999). The use of principal components allows us to capture a broader set of business cycle indicators. Third, conventional measures of output tend to imply implausibly large and persistent effects of monetary policy shocks on output, whereas our measure generates the expected temporary response. This result is in line with recent work stressing the importance of incorporating information from larger data sets in VAR models (see, e.g., Stock and Watson 2005; Banbura, Giannone and Reichlin 2008).

4.1 Simulation Evidence

Figure 3 shows effective coverage rates and average lengths of alternative pointwise confidence intervals up to a horizon of $H = 24$. 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 3 reveals severe coverage deficiencies for the LP bootstrap interval especially at long horizons. Its coverage rates may drop as low as 75%. The asymptotic LP method is fairly accurate at short horizons (except on impact), but its coverage accuracy also deteriorates at longer 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 more than twice as wide on average as the other intervals. Even the bootstrap LP interval, however, tends to be wider than the VAR intervals.

Similarly, when it comes to point estimates of the impulse responses, Figure 4 shows that the LP estimator has higher bias in most cases as well as higher variance, resulting in unambiguously higher MSEs. These results are very much consistent with the insights obtained from the stylized VAR(1) model.

Table 2 shows the corresponding results for the joint interval. Here the results differ somewhat from the VAR(1) data generating process in that the joint LP interval has typically more accurate coverage with rates between 88% and 94%. In contrast, the accuracy of the joint VAR interval is typically lower and more erratic with rates between 61% and 88%. As in the VAR(1) example, the joint LP interval is much wider on average often by a factor of more than three. This conclusion is further supported by results for subsamples, which generally produced much less accurate results for both methods. For example, if we restrict the sample to the pre-Greenspan period of 1970.1-1987.8, joint coverage rates drop as low as 33% for the VAR interval and as low as 36% for the LP interval. Moreover, the rankings of the two methods are mixed. We infer that joint intervals are unlikely to be reliable in practice, and that neither joint interval is clearly preferred over the other.

4.2 Empirical Application

We conclude this section with an illustration of how different methods of inference and choices of models may affect the consensus about the responses of macroeconomic aggregates to an unanticipated monetary policy tightening. The data and model are identical with the model used as a data generating process in this section. The first row of Figure 5 shows the responses estimated by the VAR model and the second row the responses estimated by local projections. The point estimates are plotted together with conventional pointwise nominal 95% confidence intervals. The VAR estimates suggest that the output gap temporarily turns negative with a full recovery within two years. The temporary decline is highly statistically significant. Real commodity price inflation also drops significantly. There is no evidence that CPI inflation is significantly reduced. Rather the inflation response exhibits the well known price puzzle with a statistically significant peak after two months. There is little to choose between bias-corrected bootstrap and delta method intervals in this application.

The LP estimates in the second row paint a very similar picture. The main difference is that the intervals tend to be substantially wider, lowering the statistical significance, and that the estimates are more erratic. Both findings are expected based on the earlier Monte Carlo simulation evidence. The pointwise bootstrap LP intervals are so wide at short horizons that not even the price puzzle remains significant. In contrast, based on the asymptotic LP

interval the price puzzle remains.

Figure 6 compares the pointwise and joint confidence intervals for the same example. The first row illustrates that the joint VAR intervals are similar to pointwise intervals at short horizons, but systematically wider at long horizons. A similar pattern applies to the LP intervals in the second row. At short horizons the joint and pointwise asymptotic intervals are quite similar. At longer horizons, discrepancies emerge, as the joint interval widens disproportionately. Compared with the joint VAR intervals, these intervals are substantially wider, as predicted by the simulation study.

These differences do not necessarily affect the substantive conclusions, however, because statistically significant point estimates tend to be concentrated at short horizons, when joint and pointwise intervals tend to be similar. For the VAR results, for none of the three response functions the use of joint intervals overturns the finding of a significant response function. For the LP results, both the price puzzle and the output contraction remain statistically significant, if barely so, even using the joint interval. Thus, at least in this empirical example, the use of LP models as opposed to VAR models and the use of joint intervals as opposed to pointwise intervals makes little difference for the main conclusions.

5 Approximate VAR Models

Based on the simulation results presented so far, there is no compelling reason to abandon traditional VAR methods of constructing impulse response estimates in favor of the LP method. As long as the finite lag order VAR model provides a good approximation to the stationary data generating process, the LP estimator suffers from greater small-sample bias and higher variance, resulting in very wide, yet often still inaccurate confidence intervals and erratic point estimates. An interesting avenue for future research is to investigate how poor the vector autoregressive approximation has to be for the LP method to become an attractive alternative to VAR approximations. For example, more general linear stationary processes can be represented as a VAR(∞) model and approximated by a sequence of finite-lag order vector autoregressions (see, e.g., Lütkepohl and Poskitt 1991; Inoue and Kilian 2002).

In this section, we provide some preliminary evidence based on a tri-variate invertible VARMA(1,1)-DGP used as an example in Braun and Mittnik (1993) and Inoue and Kilian (2002). The model includes quarterly investment growth, deflator inflation and the commercial paper rate in this order:

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

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

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

This model can be represented as a recursively identified VAR(∞) process which can be approximated using either local projections or a finite-lag order VAR model. We follow Inoue and Kilian (2002) in studying the response of the model variables to an innovation in the commercial paper rate. We focus on the asymptotic LP interval and the bias-adjusted bootstrap algorithm for VAR models.¹⁸ The sample size is $T = 200$. Figure 7 shows results for approximating lag orders of $p = 5$ and $q = 5$. The simulation results are remarkably robust to changes in these lag orders. Only for very low lag orders, the coverage accuracy deteriorates.

The bias-adjusted bootstrap method performs quite well, as illustrated in Figure 7. Based on the approximating VAR(5) model, its coverage rate is near the nominal coverage for all three responses. Interestingly, Figure 7 suggests that, for a suitably large choice of q , the asymptotic LP interval is just as accurate, 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 if the VAR model is merely an approximation to the data generating process.

6 Conclusion

Local projections methods are a promising recent development in the literature on impulse response analysis, but little is known about their finite sample performance and their merits relative to more conventional VAR-based methods. In this paper, we compared the small-sample performance of impulse response confidence intervals and point estimates based on local (linear) projections and VAR models. We explored and compared alternative approaches to implementing the LP method in stationary environments and developed suitable boot-

¹⁸Inoue and Kilian (2002) discuss the validity of the bootstrap for stationary VAR(∞) processes and show that this specific bootstrap approach performs better than the delta method interval proposed in Lütkepohl and Poskitt (1991).

strap methods of inference.

Our main objective was to investigate the conjecture that LP intervals may help resolve the long-standing problem of bias driven by the nonlinearity of the VAR impulse-response estimator. This bias tends to undermine Gaussian approximations to the finite-sample distribution of the impulse response in vector autoregressions, resulting in confidence intervals with poor coverage accuracy. Since LP impulse responses can be represented as slope coefficients in a linear model, our conjecture was that confidence intervals based on the LP model might be more accurate in practice than VAR based intervals.

We showed that this conjecture is not correct. In particular, we found that the bias of the LP estimator is greater than the bias of the VAR impulse response estimator, notwithstanding the linearity of the LP estimator in the slope parameters. Combined with its excessive variability, the LP point estimator proved to be very unreliable. The accuracy of the LP interval estimator tended to be erratic in small samples. Although the asymptotic LP interval was more accurate than some VAR-based alternatives such as the delta method interval, neither the asymptotic nor the bootstrap LP interval proved more accurate than bias-adjusted bootstrap intervals for VAR models. We concluded that there is no compelling reason to abandon traditional VAR methods of constructing impulse response estimates in favor of the LP method, as long as the finite lag order VAR model provides a good approximation to the data generating process. While the accuracy of LP estimators quickly improves with increasing sample size, the average width of the asymptotic LP interval far exceeds that of bias-adjusted VAR bootstrap intervals with similar accuracy. Thus, even for large samples there are no apparent advantages to the LP method. This result tends to hold even when the data are generated by a stationary $\text{VAR}(\infty)$ process.

We also investigated the reliability of the joint impulse response confidence intervals proposed by Jordà (2007) for both VAR and LP models. Such intervals are of great potential interest for applied work, given the dependence of impulse response estimates across horizons. We concluded that these joint intervals were not accurate enough to be recommended for realistic applications. This conclusion held whether these intervals were constructed based on the VAR or the LP approach. An empirical application illustrated that joint VAR intervals tend to be similar to pointwise intervals at short horizons. Only at longer horizons they become substantially wider than pointwise intervals. Since most statistically significant results in applied work are obtained at short horizons, there is reason to believe that in most cases the use of joint intervals is unlikely to overturn the substantive conclusions of studies based on more conventional pointwise intervals.

References

- [1] Andrews, D.W.K., and H.-Y. Chen (1994), “Approximately Median-Unbiased Estimation of Autoregressive Models,” *Journal of Business and Economic Statistics*, 18, 139-165.
- [2] Banbura, M., D. Giannone and L. Reichlin (2008), “Large Bayesian VARs,” Working Paper No. 966, European Central Bank.
- [3] Benkowitz, A., H. Lütkepohl , and M.H. Neumann. (2000), “Problems Related to Confidence Intervals for Impulse Responses of Autoregressive Processes,” *Econometric Reviews*, 19, 69-103.
- [4] Berkowitz, J., I. Birgean, and L. Kilian (1999), “On the Finite-Sample Accuracy of Nonparametric 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.
- [5] Berkowitz, J., and L. Kilian (2000), “Recent Developments in Bootstrapping Time Series,” *Econometric Reviews*, 19, 1-48.
- [6] 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
- [7] 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.
- [8] 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.
- [9] Chang, P., and S. Sakata (2007), “Estimation of Impulse Response Functions Using Long Autoregressions,” *Econometrics Journal*, 10, 453-469.
- [10] Davidson, R. (2000), “Comment on ‘Recent Developments in Bootstrapping Time Series’,” *Econometric Reviews*, 19, 49-54.

- [11] Evans, C.L. (1999), “If You Were the Central Banker, How Many Data Series Would You Watch? An empirical analysis,” unpublished manuscript, Federal Reserve Bank of Chicago.
- [12] Fachin, S. and L. Bravetti (1996), “Asymptotic Normal and Bootstrap Inference in Structural VAR Analysis,” *Journal of Forecasting*, 15, 329-341.
- [13] Forni, M. and L. Reichlin (1998), “Let’s Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics,” *Review of Economic Studies*, 65, 453-473.
- [14] Goncalves, S., and L. Kilian (2004), “Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form,” *Journal of Econometrics*, 123, 89-120.
- [15] Goncalves, S., and L. Kilian (2007), “Asymptotic and Bootstrap Inference for AR(∞) Processes with Conditional Heteroskedasticity,” *Econometric Reviews*, 26, 609-641.
- [16] Gospodinov, N. (2004), “Asymptotic Confidence Intervals for Impulse Responses of Near-integrated Processes”, *Econometrics Journal*, 7, 505-527.
- [17] Griffiths, W., and H. Lütkepohl (1993), “Confidence Intervals for Impulse Responses from VAR Models: A Comparison of Asymptotic Theory and Simulation Approaches,” manuscript, Institut für Statistik und Ökonometrie, Humboldt-Universität, Berlin.
- [18] Hall, A., A. Inoue, J.M. Nason, and B. Rossi (2007), “Information Criteria for Impulse Response Function Matching Estimation of DSGE Models”, mimeo, Duke University.
- [19] Hansen, B. (1999), “Bootstrapping the Autoregressive Model”, *Review of Economics and Statistics*, 81, 594-607.
- [20] Haug, A.A., and C. Smith (2007), “Local Linear Impulse Responses for a Small Open Economy”, Economics Discussion Paper No. 0707, University of Otago.
- [21] Inoue, A., and L. Kilian (2002a), “Bootstrapping Autoregressive Processes with Possible Unit Roots,” *Econometrica*, 70, 377-391.
- [22] Inoue, A., and L. Kilian (2002b), “Bootstrapping Smooth Functions of Slope Parameters and Innovation Variances in VAR(∞) Models,” *International Economic Review*, 43, 309-332.

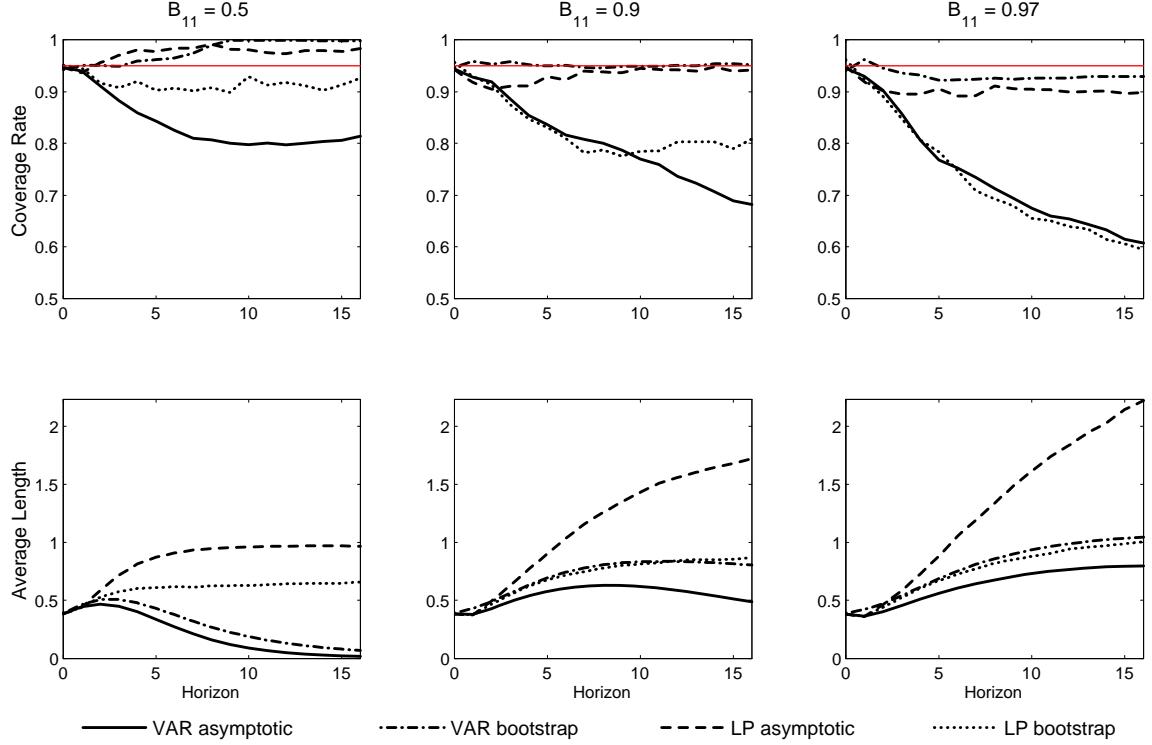
- [23] Inoue, A., and L. Kilian (2003), “The Continuity of the Limit Distribution in the Parameter of Interest is not Essential for the Validity of the Bootstrap,” *Econometric Theory*, 19, 944–961.
- [24] Jordà, Ò. (2005), “Estimation and Inference of Impulse Responses by Local Projections,” *American Economic Review*, 95, 161–182.
- [25] Jordà, Ò. (2007), “Simultaneous Confidence Regions for Impulse Responses,” *Review of Economics and Statistics*, forthcoming.
- [26] Jordà, Ò., and S. Kozicki (2007), “Estimation and Inference by the Method of Projection Minimum Distance,” mimeo, Bank of Canada.
- [27] Jordà, Ò., and M. Marcellino (2009), “Path Forecast Evaluation,” *Journal of Applied Econometrics*, forthcoming.
- [28] Kilian, L. (1998a), “Small-sample confidence intervals for impulse response functions,” *Review of Economics and Statistics*, 80, 218–230.
- [29] Kilian, L. (1998b), “Confidence intervals for impulse responses under departures from normality,” *Econometric Reviews*, 17, 1–29.
- [30] Kilian, L. (1998c), “Accounting for Lag Order Uncertainty in Autoregressions: The Endogenous Lag Order Bootstrap Algorithm,” *Journal of Time Series Analysis*, 19, 531–548.
- [31] 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.
- [32] Kilian, L. (2001), “Impulse Response Analysis in Vector Autoregressions with Unknown Lag Order,” *Journal of Forecasting*, 20, 161–179.
- [33] Kilian, L., and P.-L. Chang (2000), “How Accurate are Confidence Intervals for Impulse Responses in Large VAR Models?” *Economics Letters*, 69, 299–307.
- [34] 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.

- [35] Lütkepohl, H., and D.S. Poskitt (1991), “Estimating orthogonal impulse responses via vector autoregressive models,” *Econometric Theory*, 7, 487–496.
- [36] Pope, A.L. (1990), “Biases of Estimators in Multivariate Non-Gaussian Autoregressions,” *Journal of Time Series Analysis*, 1, 249-258.
- [37] Pesavento, E. and B. Rossi (2006), “Small-Sample Confidence Intervals for Multivariate Impulse Response Functions at Long Horizons,” *Journal of Applied Econometrics*, 21, 1135-1155.
- [38] Pesavento, E. and B. Rossi (2007), “Impulse Response Confidence Intervals for Persistent Data: What Have We Learned?” *Journal of Economic Dynamics and Control*, 31, 2398-2412.
- [39] Rudebusch, G.D. (1992), “Trends and Random Walks in Macroeconomic Time Series: A Re-Examination,” *International Economic Review*, 33, 661-680.
- [40] Runkle, D.E. (1987), “Vector Autoregression and Reality,” *Journal of Business and Economic Statistics*, 5, 437–442.
- [41] Scheffé, H. (1953), “A Method for Judging All Contrasts in the Analysis of Variance,” *Biometrika*, 40, 87-104.
- [42] Stine, R.A.(1987), “Estimating Properties of Autoregressive Forecasts,” *Journal of the American Statistical Association*, 82, 1072-1078.
- [43] Stock, J. H. (1991), “Confidence Intervals for the Largest Autoregressive Root in U.S. Economic Time Series,” *Journal of Monetary Economics*, 28, 435-460.
- [44] Stock, J.H. and M.W. Watson (2005), “Implications of Dynamic Factor Models for VAR Analysis,” mimeo, Department of Economics, Harvard University.
- [45] Wright, J.H.(2000), “Confidence Intervals for Univariate Impulse Responses with a Near Unit Root”, *Journal of Business and Economic Statistics*, 18, 368-373.

Exhibit 1: Models for Estimating Impulse Responses and Methods of Inference

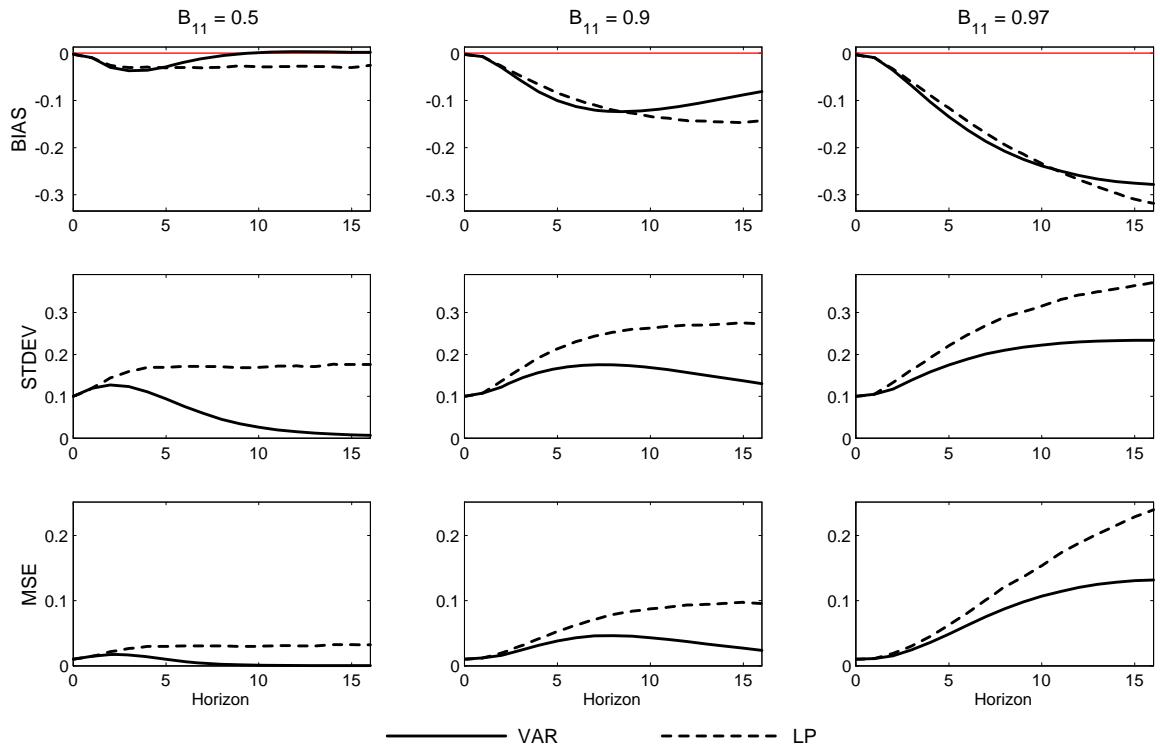
Model	Inference		
	Pointwise Interval	Bootstrap	Joint Interval
	Asymptotic	Asymptotic	
VAR	(1)	(2)	(5)
LP	(3)	(4)	(6)

Figure 1: Coverage Rates and Average Lengths of
95% Pointwise Confidence Intervals for $\theta_{21,h}$



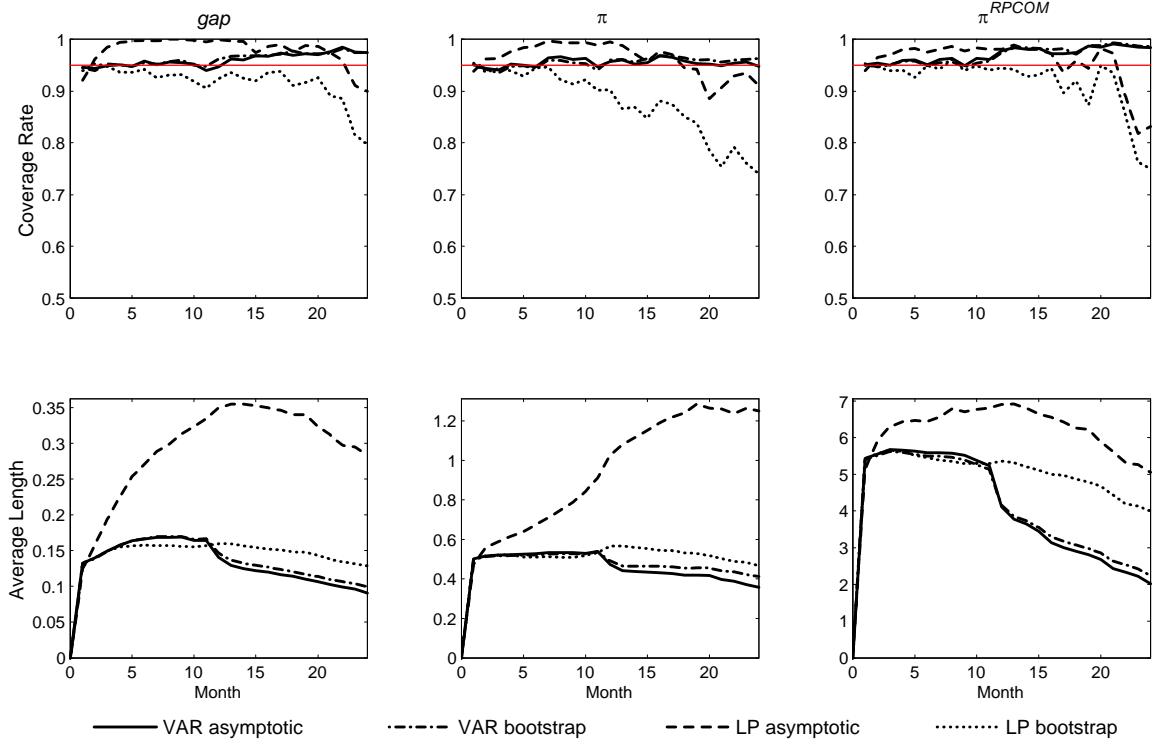
Notes: Simulation results based on 1,000 trials of length $T=100$ from the VAR(1) DGP described in text. VAR asymptotic denotes the asymptotic delta method for VAR impulse responses. VAR bootstrap refers to the bias-corrected bootstrap method for VAR impulse responses. LP asymptotic is the asymptotic interval for LPs. LP bootstrap refers to a 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: Bias, Standard Deviation, and MSE of $\hat{\theta}_{21,h}$



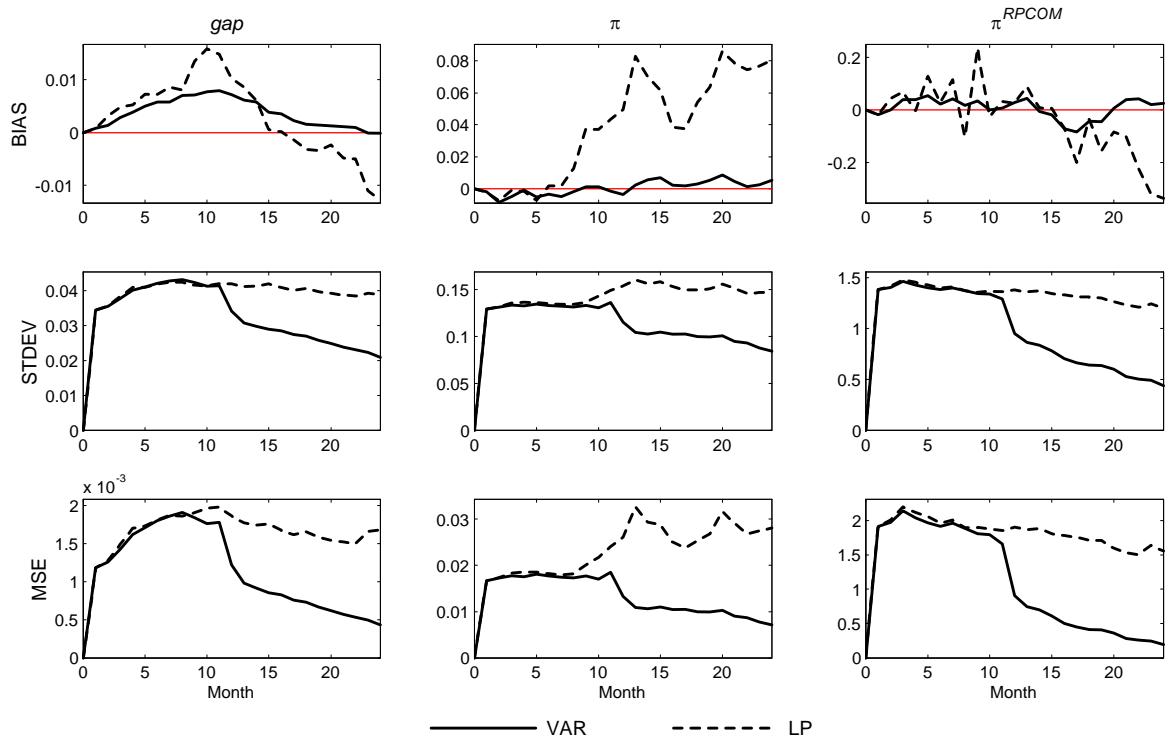
Notes: Based on 5,000 trials.

Figure 3: Coverage Rates and Average Lengths of
95% Pointwise Confidence Intervals for Responses to a Monetary Tightening



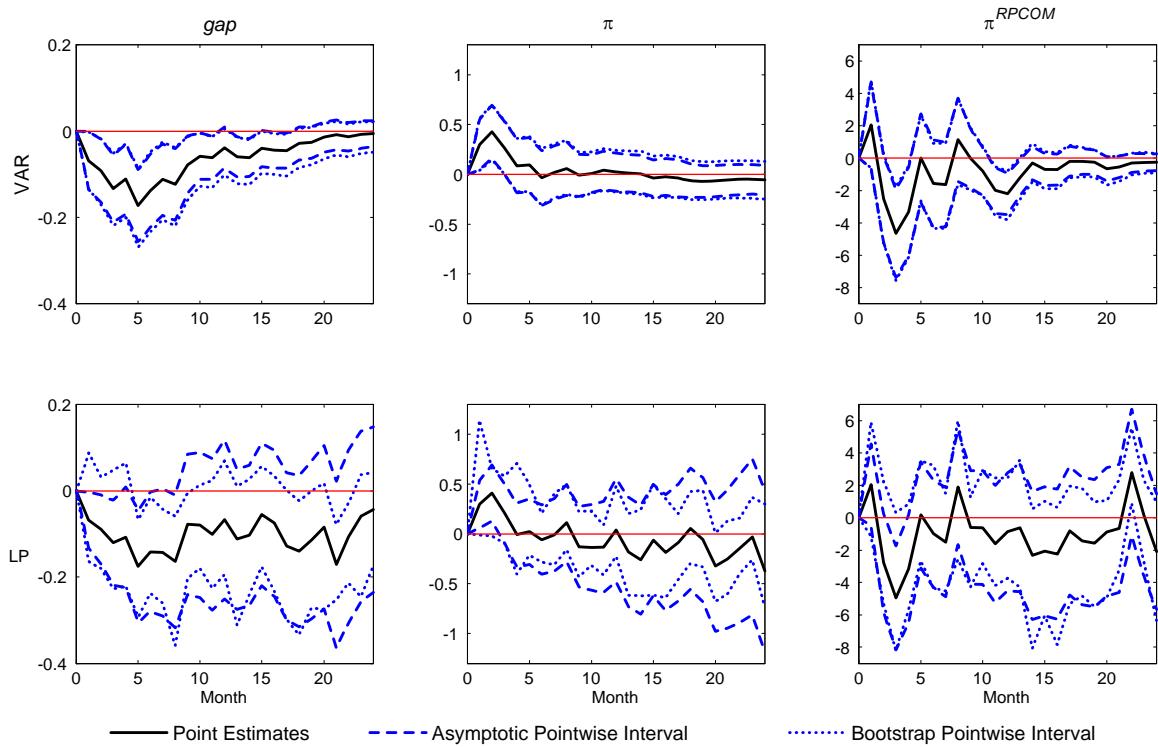
Notes: Simulation results are based on 1,000 trials of length 456 from the VAR(12) model described in the text. Gap denotes the CFNAI, π is CPI inflation, and π^{RPCOM} is real commodity price inflation. *VAR asymptotic* refers to the delta method interval for VAR impulse responses. *VAR bootstrap* refers to the bias-adjusted bootstrap method for VAR models. *LP asymptotic* refers to asymptotic interval for LP models. *LP bootstrap* refers to the block bootstrap interval for LP models. Lag orders are selected by the AIC with an upper bound of 12 lags in all cases. Since there is no uncertainty about the impact response of these variables, we do not construct a coverage rate for horizon 0.

Figure 4: Bias, Standard Deviation, and MSE of Impulse Responses to a Monetary Tightening



Notes: Based on 5,000 trials. See Table 3.

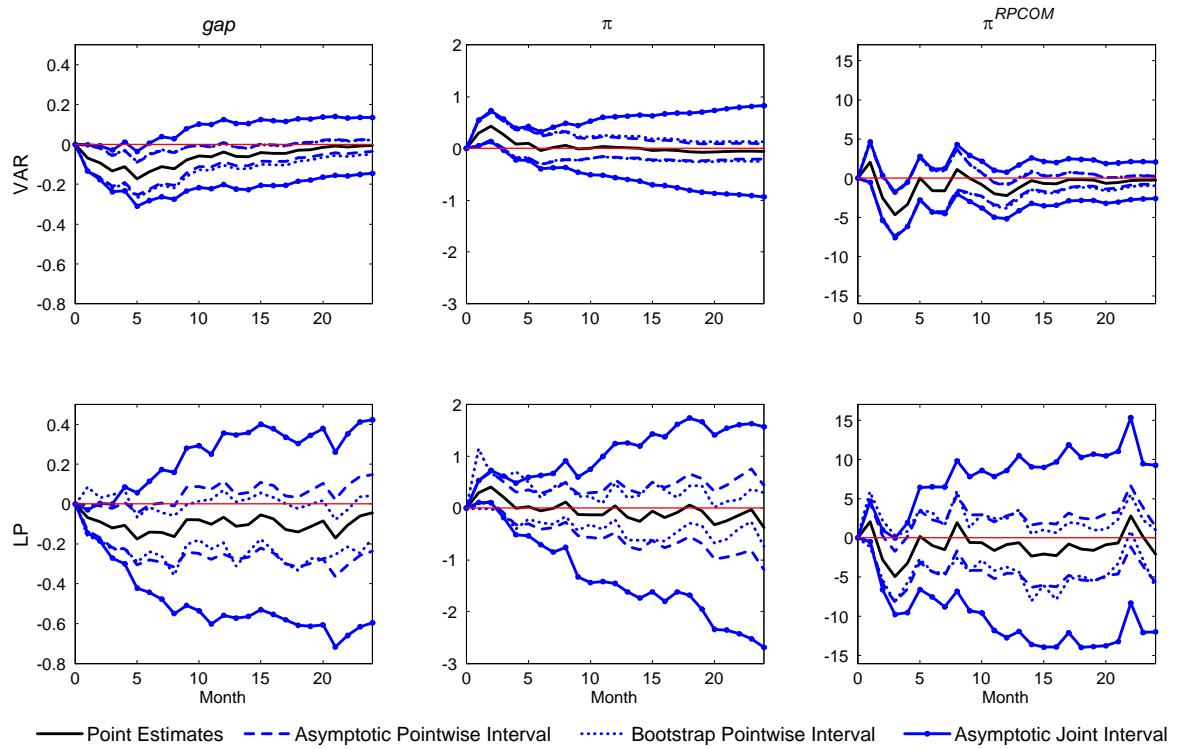
Figure 5: Responses to a Monetary Tightening with
95% Pointwise Confidence Intervals
1970.1-2007.12



Notes: The lag orders were chosen by the AIC with an upper bound of 12 lags

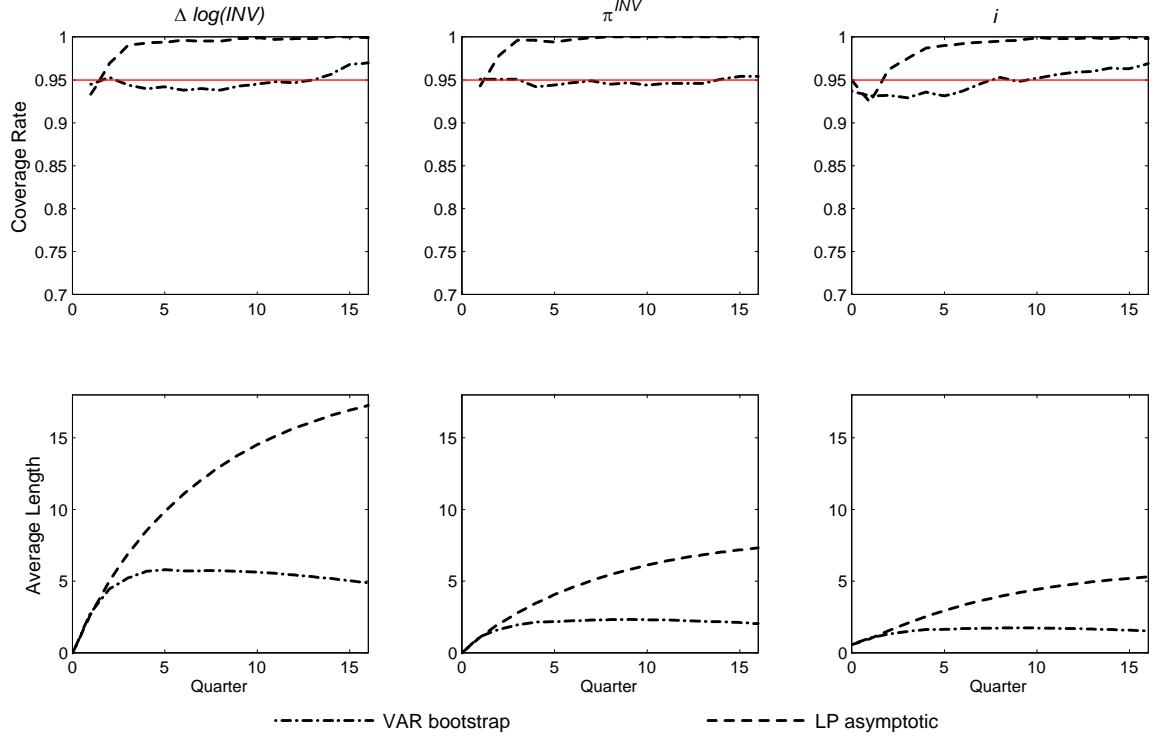
Figure 6: Responses to a Monetary Tightening with
95% Pointwise and Joint Confidence Intervals

1970.1-2007.12



Notes: See Figure 5.

Figure 7: Coverage Rates and Average Lengths of
95% Pointwise Confidence Intervals for Responses to an Interest Rate Innovation



Notes: Simulation results for $T = 200$ based on 1,000 trials from the VARMA(1,1)-DGP described in Inoue and Kilian (2002). *VAR bootstrap* refers to the bias-adjusted bootstrap method for VAR models. *LP asymptotic* refers to asymptotic interval for LP models. 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.

Table 1: Coverage Rates and Average Lengths of
Asymptotic 95% Joint Confidence Intervals for θ_{ij} in the VAR(1) DGP

	VAR			LP		
	B ₁₁ =0.5	B ₁₁ =0.9	B ₁₁ =0.97	B ₁₁ =0.5	B ₁₁ =0.9	B ₁₁ =0.97
Coverage Rate						
θ_{11}	0.868	0.882	0.859	0.816	0.874	0.849
θ_{12}	0.977	0.979	0.980	0.822	0.905	0.904
θ_{21}	0.899	0.907	0.896	0.867	0.872	0.843
θ_{22}	0.892	0.897	0.884	0.798	0.800	0.798
Average Length						
θ_{11}	0.513	1.395	1.829	1.239	3.323	4.195
θ_{12}	0.494	1.351	1.767	1.203	3.202	4.025
θ_{21}	0.782	1.428	1.751	1.886	3.310	3.898
θ_{22}	0.764	1.369	1.664	1.853	3.249	3.809

Notes: Simulation results based on 1,000 trials from the VAR(1) DGP described in text.

Table 2: Coverage Rates and Average Lengths of
 Asymptotic 95% Joint Confidence Intervals for
 Responses to a Monetary Tightening in the VAR(12)-DGP

	VAR	LP
Coverage Rate		
gap	0.877	0.920
π	0.764	0.914
π^{RPCOM}	0.614	0.936
i	0.874	0.881
Average Length		
gap	0.265	0.853
π	1.020	2.981
π^{RPCOM}	5.082	16.589
i	0.832	2.403

Notes: Simulation results based on 1,000 trials from the VAR(12) model described in text.