NBER WORKING PAPER SERIES

A NOTE ON VARIANCE DECOMPOSITION WITH LOCAL PROJECTIONS

Yuriy Gorodnichenko
Byoungchan Lee

Working Paper 23998
http://www.nber.org/papers/w23998

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2017

We thank Oscar Jorda and Mikkel Plagborg-Møller for comments on an earlier version of the paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

A Note on Variance Decomposition with Local Projections
Yuriy Gorodnichenko and Byoungchan Lee
NBER Working Paper No. 23998
November 2017
JEL No. C53,E37,E47

ABSTRACT

We propose and study properties of several estimators of variance decomposition in the local-projections framework. We find for empirically relevant sample sizes that, after being bias corrected with bootstrap, our estimators perform well in simulations. We also illustrate the workings of our estimators empirically for monetary policy and productivity shocks.

Yuriy Gorodnichenko
Department of Economics
530 Evans Hall #3880
University of California, Berkeley
Berkeley, CA 94720-3880
and IZA
and also NBER
ygorodni@econ.berkeley.edu

Byoungchan Lee
University of California, Berkeley
Department of Economics
Berkeley, CA
bc1105@berkeley.edu

I. Introduction

Macroeconomists have been long interested in estimating dynamic responses of output, inflation and other aggregates to structural shocks. While many analyses use vector autoregressions (VARs) or dynamic stochastic general equilibrium (DSGE) models to construct estimated responses, an increasing number of researchers focus on a single structural shock and employ single-equation methods to study the dynamic responses. This approach allows concentrating on well-identified shocks and leaving other sources of variation unspecified. In addition, these approaches often impose no restrictions on the shape of the impulse response function. As a result, the local projections method (Jordà 2005, Stock and Watson 2007) has gained prominence in applied macroeconomic research.

The properties of impulse responses estimated with these methods are well studied (see e.g. Coibion 2012) but little is known about how one can estimate quantitative significance of shocks in the single-equation framework. Specifically, the vast majority of studies using single-equation approaches do not report variance decomposition for the variable of interest and hence one does not know if a given shock accounts for a large share of variation for the variable.[1] This practice contrasts sharply with the nearly universal convention to report variance decompositions in VARs and DSGE models. In this paper, we propose several methods to construct variance decomposition in the local projection framework.

We show that local projections lead to a simple and intuitive way to assess the contribution of identified shocks to variation at different horizons. However, there are several options to implement this insight. While the details of implementation do not matter in large samples, we observe heterogeneity in the performance of various options in small, empirically relevant samples. To illustrate the properties of various methods, we use several data generating processes which cover main profiles of variance decompositions documented in previous works. We show that estimated contributions to variation may be biased in small samples and one should use bootstrap to correct for possible biases in the local projections' estimates of variation decompositions. We also demonstrate how our method works in settings with multiple identified shocks. We illustrate the performance of our method using actual data and commonly used identified shocks as well as data simulated according to the Smets and Wouters (2007) DSGE model. Our work is concurrent and

[1] Coibion et al. (2017) is among the very few papers reporting variance decomposition in the local projection method. More precisely, if we use definitions of Plagborg-Møller and Wolf (2017), the object of our analysis is forecast variance ratio.

complementary to Plagborg-Møller and Wolf (2017) who provide set-identified variance decompositions in the local projections framework.

The rest of the paper is structured as follows. Section II lays out a basic setting to derive the estimators. Section III presents simulation results for bivariate and multivariate settings. Section IV provide an application of our estimators to estimate the contribution of monetary policy and productivity shocks to variation of output and inflation in the local projections framework. Section V concludes.

II. Basics of variance decomposition

Consider a generic setup encountered in studies using local projections. Let y_t be an endogenous variable of interest. We assume that variation in y_t has two components: an identified white-noise shocks series x_t with mean zero and variance σ_x^2 and the "rest" captured by series z_t so that

$$y_t = \sum_{i=0}^{\infty} \psi_{x,i} x_{t-i} + z_t = \psi_x(L) x_t + z_t. \quad (1)$$

We are interested in estimating coefficients in the lag polynomial $\psi_x(L)$ which provides us with the impulse response function of variable y_t to shock x_t. We make only a few assumptions about properties of x_t and z_t. Specifically, we assume that z_t admits an integrated $MA(\infty)$ representation,

$$\Delta z_t = g_y + \psi_e(L) e_t \quad (2)$$

where e_t is a zero-mean white noise series with variance σ_e^2. Following the conventions of local projection applications, we assume that x_t and e_t are uncorrelated and that $\sum_{i=0}^{\infty} \psi_{x,i}^2 < \infty$ and $\sum_{i=0}^{\infty} \psi_{e,i}^2 < \infty$. Without loss of generality we set $\psi_{e,0} = 1$. We assume that $\{(x_t, \Delta y_t)\colon t = 1, \ldots, T\}$ is observable.

Forecast error for h-period ahead value of the endogenous variable is given by

$$f_{t+h|t-1} \equiv y_{t+h} - y_{t+h|t-1} = (y_{t+h} - y_{t-1}) - E[y_{t+h} - y_{t-1} | \Omega_{t-1}]$$

where $y_{t+h|t-1} \equiv E[y_{t+h} | \Omega_{t-1}]$ is the prediction of y_{t+h} given information set $\Omega_{t-1} \equiv \{\Delta y_{t-1}, x_{t-1}, \Delta y_{t-2}, x_{t-2}, \ldots\}$. We can decompose forecast error due to innovations in x_t and other sources of variation as follows[2]

$$f_{t+h|t-1} = \psi_{x,0} x_{t+h} + \cdots + \psi_{x,h} x_t + v_{t+h|t-1}. \quad (3)$$

[2] If $\psi_e(L)$ is invertible, $v_{t+h|t-1}$ is equal to $\psi_{e,0} e_{t+h} + \cdots + (\psi_{e,0} + \cdots + \psi_{e,h}) e_t$. This representation in e_t's is obtained, because $e_t \in \Omega_t$. See Appendix A for details. Note that we do *not* need invertibility of $\psi_e(L)$ to construct the contribution of x_t to variability in y_t. Intuitively, we only need an estimate of either $Var(f_{t+h|t-1})$ in equations (4) and (4'), or $Var(v_{t+h|t-1})$ in equation (4'') which does not require us separating $\psi_e(L)$ and e_t.

Following Sims (1980), we can define the population share of variance explained by the future innovations in x_t to the total variations in $f_{t+h|t-1}$:

$$s_h = \frac{Var(\psi_{x,0}x_{t+h} + \cdots + \psi_{x,h}x_t)}{Var(f_{t+h|t-1})} \tag{4}$$

$$= \frac{\left(\sum_{i=0}^{h} \psi_{x,i}^2\right)\sigma_x^2}{Var(f_{t+h|t-1})} \tag{4'}$$

$$= \frac{\left(\sum_{i=0}^{h} \psi_{x,i}^2\right)\sigma_x^2}{\left(\sum_{i=0}^{h} \psi_{x,i}^2\right)\sigma_x^2 + Var(v_{t+h|t-1})}. \tag{4''}$$

Equation (4) demonstrates that we have several options for estimating s_h and these options vary in their reliance on imposing parametric structure. In what follows, we propose and evaluate several methods to estimate s_h.

A. R^2 method

Let $X_t^h = (x_{t+h}, \dots, x_t)'$. It can be shown with some algebra that equation (4) can be written as

$$s_h = \frac{Cov(f_{t+h|t-1}, X_t^h)\left[Var(X_t^h)\right]^{-1} Cov(X_t^h, f_{t+h|t-1})}{Var(f_{t+h|t-1})}. \tag{5}$$

This quantity can be understood as an R^2 of the population projection of $f_{t+h|t-1}$ on X_t^h, or probability limit of sample R^2. This observation suggests a natural estimator for s_h. First, the forecast errors for each horizon h are estimated using local projections. Second, forecast error for horizon h at time t $f_{t+h|t-1}$ is regressed on shocks x that happen between t and $t+h$. The R^2 in this regression is an estimate of s_h.

More precisely, the estimated forecast error $\hat{f}_{t+h|t-1}$ is the residual of the following regression:

$$y_{t+h} - y_{t-1} = c_h + \sum_{i=1}^{L_y} \gamma_i^h \, \Delta y_{t-i} + \sum_{i=1}^{L_x} \beta_i^h \, x_{t-i} + f_{t+h|t-1}, \tag{6}$$

which is an approximation to $y_{t+h} - y_{t-1} = c_h + \sum_{i=1}^{\infty} \gamma_i \, \Delta y_{t-i} + \sum_{i=1}^{\infty} \beta_i^h \, x_{t-i} + f_{t+h|t-1}$ in population. Then we run the following regression and calculate its R^2:

$$\hat{f}_{t+h|t-1} = \alpha_{x,0}x_{t+h} + \cdots + \alpha_{x,h}x_t + \tilde{v}_{t+h|t-1}. \tag{7}$$

Thus, our first estimator is $\hat{s}_h^{R2} = R^2$ which, by construction, is between 0 and 1. Note that $\alpha_{x,i}$ in equation (7) corresponds to $\psi_{x,i}$ in equation (1). Because $\hat{f}_{t+h|t-1}$ in equation (7) is a residual of an OLS regression with the intercept in equation (6) and x_t is assumed to be zero mean, an intercept term in equation (7) is not required. Moreover, the population mean of both $f_{t+h|t-1}$ and X_t^h are zeros, so both centered and non-centered R^2's are the same in population. We report results for the non-centered R^2, but properties are similar when we use the centered R^2. The following proposition derives the asymptotic distribution of the estimator.

Proposition 1. Suppose $f_h = \left(f_{T|T-h-1}, f_{T-1|T-h-2}, \dots, f_{L_{max}+h+1\,|L_{max}}\right)'$ and $X_h = \left(X_{T-h}^h, X_{T-1}^h, \dots, X_{L_{max}+1}^h\right)'$ for all $h \geq 0$ where $L_{max} = \max\{L_x, L_y\}$. Then the R^2 of the regression of $f_{t+h|t-1}$ on X_t^h, given by $\left(f_h' P_{X_h} f_h\right)/(f_h' f_h)$ where $P_{X_h} = X_h(X_h' X_h)^{-1} X_h'$, has the following asymptotic distribution for some V_{h,R^2}:

$$\sqrt{T}\left(\frac{f_h' P_{X_h} f_h}{f_h' f_h} - s_h\right) \xrightarrow{d} \mathcal{N}\left(0, \quad V_{h,R^2}\right).$$

Proof. See Appendix B1.

In practice, we may plug the estimated forecast errors from equation (6) in the place of f_h. Appendix B1 contains details of implementation. Note that, instead of using shocks $x_t, \dots, x_{t+h}$ in equation (7), one may want to use residuals from projecting $x_t, \dots, x_{t+h}$ on lags of x_t and Δy_t from equation (6) to guarantee that one does not use forecastable movements in $x_t, \dots, x_{t+h}$ to account for variation in $\hat{f}_{t+h|t-1}$. In practice, however, shocks x_t are constructed in ways to ensure that x_t is not predictable by lags of macroeconomic variables. As a result, we find in our simulations and applications that purifying structural shocks make little difference. Relatedly, one may implement this estimator by augmenting equation (6) with shocks $x_t, \dots, x_{t+h}$ and calculating partial R^2. This insight also justifies using $\hat{f}_{t+h|t-1}$ instead of $f_{t+h|t-1}$ in Proposition 1.

B. Local projection based methods

The R^2 approach requires estimation of two regressions for each horizon (first, construct forecast errors; second, compute the contribution of shocks x between t and $t+h$). However, one can estimate variance decomposition from the local projection directly. Following Jordà (2005), we can estimate $\psi_{x,h}$ from the following equation:

$$y_{t+h} - y_{t-1} = c_h^{\mathrm{LP}} + \sum_{i=1}^{L_y} \gamma_i^{h,LP}\, \Delta y_{t-i} + \sum_{i=0}^{L_x} \beta_i^{h,LP}\, x_{t-i} + r_{t+h|t-1} \qquad (8)$$

where $\hat{\beta}_0^{h,LP}$ is an estimate of $\psi_{x,h}$. Since we can estimate σ_x^2 directly from x_t, we can calculate $\left(\sum_{i=0}^{h} \psi_{x,i}^2\right)\sigma_x^2$ in the numerator of equation (4'). To compute the denominator in equation (4'), we note that the residual in equation (8) can be related to the forecast error $f_{t+h|t-1}$ in equation (6). For example, $\hat{f}_{t|t-1} = \hat{\beta}_0^{0,LP} x_t + \hat{r}_{t|t-1}$, that is, a part of forecast error $f_{t|t-1}$ is explained by shock x happening at time t which is now included as one of the regressors in equation (8). In a similar spirit, we can use equation (3) to compute $\hat{f}_{t+h|t-1} = \hat{\beta}_0^{h,LP} x_t + \hat{r}_{t+h|t-1}$. With these estimates of $\hat{f}_{t+h|t-1}$, we can compute $\widehat{Var}(\hat{f}_{t+h|t-1})$ where $\widehat{Var}(\cdot)$ denotes a sample variance. Using these insights, we define a local projection estimator of variance decomposition "LPA" as

$$s_h^{LPA} = \frac{\left(\sum_{i=0}^{h} \left\{\hat{\beta}_0^{i,LP}\right\}^2\right) \hat{\sigma}_x^2}{\widehat{Var}\left(\hat{\beta}_0^{h,LP} x_t + \hat{r}_{t+h|t-1}\right)} \tag{9}$$

where $\hat{\sigma}_x^2 \equiv \widehat{Var}(x_t)$.

Although simple, LPA estimator does not guarantee that in small samples the estimated s_h is between 0 and 1. A simple solution to this issue is to split the denominator into variation due to x and due to v so that $\left(\sum_{i=0}^{h} \psi_{x,i}^2\right)\sigma_x^2$ appears in both the numerator and denominator as in equation (4''). Note that

$$\begin{aligned} \hat{v}_{t+h|t-1} &= \hat{f}_{t+h|t-1} - \hat{\beta}_0^{h,LP} x_t - \hat{\beta}_0^{h-1,LP} x_{t+1} - \cdots - \hat{\beta}_0^{0,LP} x_{t+h} \\ &= \hat{r}_{t+h|t-1} - \hat{\beta}_0^{h-1,LP} x_{t+1} - \cdots - \hat{\beta}_0^{0,LP} x_{t+h} \end{aligned}$$

so that

$$\widehat{Var}\left(f_{t+h|t-1}\right) = \widehat{Var}\left(\hat{r}_{t+h|t-1} - \hat{\beta}_0^{h-1,LP} x_{t+1} - \cdots - \hat{\beta}_0^{0,LP} x_{t+h}\right) + \sum_{i=0}^{h} \left(\hat{\beta}_0^{i,LP}\right)^2 \hat{\sigma}_x^2$$

which we use to define another local projection estimator of variance decomposition "LPB":

$$s_h^{LPB} = \frac{\left(\sum_{i=0}^{h} \left\{\hat{\beta}_0^{i,LP}\right\}^2\right) \hat{\sigma}_x^2}{\sum_{i=0}^{h} \left(\hat{\beta}_0^{i,LP}\right)^2 \hat{\sigma}_x^2 + \widehat{Var}\left(\hat{r}_{t+h|t-1} - \hat{\beta}_0^{h-1,LP} x_{t+1} - \cdots - \hat{\beta}_0^{0,LP} x_{t+h}\right)}. \tag{9'}$$

Using tools from Proposition 1, we can derive the asymptotic distribution of the LPA and LPB estimators.

Proposition 2. The local projections based estimators when $f_{t+h|t-1}$ is observable have the following asymptotic distributions for some $V_{h,LPA}$ and $V_{h,LPB}$:

$$\sqrt{T}\left(\frac{\sum_{i=0}^{h}\left(\hat{\beta}_0^{i,LP}\right)^2\hat{\sigma}_x^2}{\widehat{Var}(f_{t+h|t-1})}-s_h\right)\xrightarrow{d}\mathcal{N}(0,\ \ V_{h,LPA}),\quad \text{and}$$

$$\sqrt{T}\left(\frac{\sum_{i=0}^{h}\left(\hat{\beta}_0^{i,LP}\right)^2\hat{\sigma}_x^2}{\sum_{i=0}^{h}\left(\hat{\beta}_0^{i,LP}\right)^2\hat{\sigma}_x^2+\widehat{Var}\left(r_{t+h|t-1}-\sum_{i=0}^{h-1}\hat{\beta}_0^{i,LP}\,x_{t+h-i}\right)}-s_h\right)\xrightarrow{d}\mathcal{N}(0,\ \ V_{h,LPB}).$$

Proof. See Appendix B2.

C. Small-sample refinements

To correct for potential small-sample biases in the estimates of s_h and to enhance coverage rates for confidence bands, we bootstrap $\hat{s}_h^{R2}$, $\hat{s}_h^{LPA}$, and $\hat{s}_h^{LPB}$ using an estimated VAR model which includes two variables $\{x_t, \Delta y_t\}$. While our implementation of bootstrap is aimed to remove potential biases, alternative implementations may also refine asymptotic inference. Details on how bootstrap is implemented are relegated to Appendix E.

D. Extension

While our analysis has focused on the bivariate case, the framework can be easily generalized to include more controls in equation (6):

$$y_{t+h}-y_{t-1}=\sum_{i=1}^{L_x}\beta_i^h\,x_{t-i}+\sum_{i=1}^{L_C}C_{t-i}'\Gamma_i^h+f_{t+h|t-1} \tag{10}$$

where C_t is the vector of control variables which may include structural shocks other than x_t. In the base case, C_t consists only of Δy_t. Note that for VAR-based bootstrap, one has to include x_t and *all* variables in C_t to simulate data.[3] Similar adjustments are also possible for LPA and LPB methods.

One should bear in mind that, although including or excluding C_t or changing the composition of variables in C_t should make little difference of impulse responses estimated with local projections (provided x is uncorrelated with other shocks), what goes in C_t is potentially important for variance decomposition. Intuitively, by including more controls in C_t, we (weakly) reduce the size of the forecast error (that is, information set Ω_t expands) and hence the amount of variation to be explained shrinks. In other words, the regressand in equation (7) and therefore s_h

[3] As the number of variables in C_t increases, the number of parameters in the VAR increases rapidly. When C_t is a large vector, or when a VAR is not a good representation of the DGP for control variables, VAR-based bootstrap might not be an appealing option. In such a case, one may correct for biases by simulating asymptotic distribution of primitive quantities in (4) such as $\hat{\psi}_{x,i}, \hat{\sigma}_x^2$, and $\widehat{Var}\left(\hat{v}_{t+h|t-1}\right)$. By considering s_h as a non-linear function of those parameters, such simulations would detect biases due to the non-linearity. See Appendix B for implementation and E and F for the results.

change with a change in the list of variables included in C_t. Thus, one should not be surprised to observe that the share of variation explained by x may be sensitive to changes in C_t.

III. Simulations

This section presents two sets of simulations. The first set shows results for the baseline bivariate case and studies the performances of the three estimators for various profiles of contribution of x to variance of y at different horizons. The second set uses the estimated Smets and Wouters (2007) model to investigate the performance in a setting with many control variables.

For each data generating process (DGP), we simulate data 2,000 times. When we employ bootstrap to correct for biases, the number of bootstrap replications is set to B=2,000. As a benchmark, we report results based on a corresponding VAR. This benchmark corresponds to the practice of including shocks into VARs directly (e.g., Basu et al. 2006, Ramey 2011, Barakchian and Crowe 2013, Romer and Romer 2004, 2010). We choose the Hannan-Quinn information criterion (HQIC) as our benchmark criterion to determine the number of lags in VAR. To make VAR and LP models comparable, we use HQIC number of lags in the VAR to set L_x and L_y. Results are similar when we use AIC instead of HQIC.

The sample size for simulated data is $T = 160$. Results for other sample sizes are reported in Appendices E and F. Standard errors are computed as the standard deviation of estimates across bootstrapped samples. The coverage rates are calculated as $\Pr\left(\left|\frac{\hat{s}_h - s_h}{s.e.(\hat{s}_h)}\right|\right) \leq 1.65$.

A. Bivariate Data Generating Processes

We study three DGPs to cover different shapes of s_h. The basic structure is as follows:

$$
\begin{aligned}
&y_t = \psi_x(L)x_t + z_t \\
&z_t = p_t + a_t, \\
&(\Delta p_t - g_y) = \rho_p(\Delta p_{t-1} - g_y) + e_t^p, \qquad e_t^p \sim iid\ N(0, \sigma_p^2), \\
&a_t = \rho_a a_{t-1} + e_t^a, \qquad e_t^a \sim iid\ N(0, \sigma_a^2), \\
&x_t \sim iid\ N(0, \sigma_x^2),
\end{aligned}
$$

where x_t, e_t^p and e_t^a are mutually independent, p_t and a_t are permanent and transitory components of z_t. Appendix C derives the population $MA(\infty)$ representation of Δz_t.

DGP1 is characterized by hump-shaped ψ_x and s_h. We assume that $\psi_x(L)x_t$ follows an $MA(100)$ process with the maximum response set to 3 after 8 periods.[4] DGP2 has a strong response of y to x only in the short-run and thus the shape of s_h is downward-sloping. Finally, DGP3 assumes that $\psi_x(L)$ has a unit root so that x has persistent effects on y and the shape of s_h is upward-sloping. Table 1 reports parameter values for each DGP. Figure 1 plots true impulse responses of y to x (Panel A) and the contribution of x to variation in y (Panel B).

For DGP1, we find (Table 2) that local projections capture the hump-shaped impulse response correctly but s^{R2}, s^{LPA} and s^{LPB} fail to match the hump-share dynamics of s_h. s^{R2}, s^{LPA} and s^{LPB} tend to monotonically increase with the horizon. The VAR misses the hump both in the impulse response and variance decomposition as HQIC selects too few lags (on average the number of lags is 1.27). Confidence bands yield poor coverage rates. This performance reflects the fact that, by construction, shock x contributes zero variation in y for this DGP at short horizons. Since s_h is between zero and one, we effectively have estimates close to the boundary and, therefore, standard methods are likely to fail. While bootstrap appears to provide some improvement (e.g., the bias at long horizons when x accounts for a larger share of variance in y is corrected)[5], it does not perform consistently better because the parameter is at the boundary. When we allow x to explain 5 percent or more of the variation in y at short horizons, bootstrap brings coverage rates close to nominal (results are available upon request). Note that, although VAR is strongly biased, the VAR estimates tend to have smaller variance so that the root mean squared error (RMSE) is similar in magnitude to RMSE of the s^{R2}, s^{LPA} and s^{LPB} estimators. Finally, we observe that the s^{R2}, s^{LPA} and s^{LPB} estimators have similar performance.

Because DGP2 permits an exact, finite-order VAR representation,[6] VAR has good properties in terms of bias, RMSE and coverage rates (Table 3). The local projections recover the share of the impulse response correct, but the estimates of contribution of x to variance of y again overstate the contribution at long horizons. Bootstrap can correct this bias. Given that VAR nests

[4] This value and pattern is motivated by a 3 percent response of real GDP to a 100bp monetary policy shock estimated in Coibion (2012).

[5] The bias can be further reduced by using higher values of L_x and L_y by reducing errors in $\hat{f}_{t+h|t-1}$ due to the truncation.

[6] Given the parameter values in Table 1, $\Delta y_t = g_y + (1-L)(1-0.9L)^{-1}x_t + (1-0.9L)^{-1}e_t^p$. By pre-multiplying $(1-0.9L)$, we have $\Delta y_t = 0.1g_y + 0.9\Delta y_{t-1} - x_{t-1} + x_t + e_t^p$.

the DGP and that VAR is more parsimonious than local projections, VAR has a better performance than the s^{R2}, s^{LPA} and s^{LPB} estimators.

Because x has long-lasting effects on y in DGP3, the VAR underestimates the responses at long horizons in small samples. Impulse responses estimated with local projections perform better but also exhibit a downward bias at long horizons. In a similar spirit, $\hat{s}_h$ shows a strong downward bias for VAR and a smaller, but still considerable bias for the s^{R2}, s^{LPA} and s^{LPB} estimators (this is the case even after we use bootstrap to correct for possible biases). This performance reflects the fact that HQIC chooses a low number of lags (1.34 lags on average across simulations). As a result, VARs used to simulate bootstrap samples fail to capture the degree of persistence in the data. To demonstrate the importance of the lag order, we report results (Table 4) when we use VAR(5) and VAR(10) for bootstrap. As the number of lags increases, we observe some improvement but these enhancements are achieved at the price of higher variance in the estimates. These results suggest that one may want to overfit VAR for persistent processes at the bootstrap stage.

In summary, we find for small samples that the s^{R2}, s^{LPA} and s^{LPB} estimators perform reasonably well across the DGPs and that bootstrap helps to improve the estimators' properties. In addition, there is relatively little difference between the s^{R2}, s^{LPA} and s^{LPB} estimators. In contrast, VARs that include structural shock x tend to perform poorly when a DGP is not nested in a small-order VAR.

B. Smets-Wouters model

While the bivariate DGPs provide important insights about how the R^2, LPA and LPB estimators perform, researchers face potentially more complex DGPs and often have more information in practice. In this section, we use the Smets and Wouters (2007) model to study performance of our estimators in an environment with multiple shocks and many control variables.

As discussed above, different information sets determine different population s_h. In the simulations, we assume that the researcher is interested in explaining variation in output and that the researcher observes output growth rate, inflation, federal funds rate, and monetary policy shocks.[7] This choice of variables is motivated by the popularity of small VARs which include

[7] For this information set, we construct the true variance decomposition using a stationary Kalman filter similar to the method in Appendix C. We also tried various combinations of shocks and endogenous variables and found similar results. Figures for inflation and results with large samples are in Appendix F. Note that monetary policy shocks are *nearly* invertible in the Smets-Wouters model (see Wolf (2017) for more details). While this may be a problem if we

output, inflation and a policy rate to study effects of monetary policy on the economy. In this exercise, the shock is ordered first because the Smets-Wouters model allows contemporaneous responses of macroeconomic variables to policy shocks. When estimating impulse responses using local projections, we augment equation (8) with inflation and federal funds rate as controls.

We find (Figure 2) that local projections correctly recover the response of output to monetary policy shocks, while a low order VAR (lag length is chosen with HQIC) fails to capture the transitory effect of monetary shocks on output. Consistent with our bivariate analysis, $\hat{s}_h^{R2}$, $\hat{s}_h^{LPA}$ and $\hat{s}_h^{LPB}$ increase with the horizons while the true s_h exhibits hump-shaped dynamics. s_h estimated with a VAR also fails to capture the true dynamics as $\hat{s}_h$ flattens out after about $h = 5$. Similar to our results in the previous section, we find that bias correction helps $\hat{s}_h^{R2}$, $\hat{s}_h^{LPA}$ and $\hat{s}_h^{LPB}$ to recover the true hump-shaped profile of s_h. Coverage rates (after bias correction) are 10 percentage points lower their nominal values at short horizons ($h \leq 5$) but the coverage rates are close to nominal at longer horizons. Again, although VAR estimates of s_h are strongly biased, the variance of estimates is low so that RMSE is broadly similar cross methods. We conclude that our proposed methods to estimate variance decomposition work well in more complex settings.

IV. Application

To illustrate the properties our estimators, we use two structural shocks identified in the literature. The first shock is the monetary policy innovation identified as in Romer and Romer (2004) and extended in Coibion et al. (2017). The second shock is the total factor productivity (TFP) shock identified as in Fernald (2014).[8] The correlation between the shocks is -0.059. Our objective is to quantify the contribution of these shocks to variation of output and inflation. The sample covers 1969Q1-2007Q4 which excludes the period of binding zero lower bound. The set of variables for local projections includes inflation (annualized growth rate of GDP deflator, i.e. $400\Delta\ln(P_t)$), annual GDP growth rate ($400\Delta\ln(Y_t)$), federal funds rate, and the two-shock series. We set $L_C = L_x = 4$ in equation (10) and add control variables similarly when estimating impulse responses. In the benchmark VAR, we have all five variables and allow four lags.[9]

use shocks identified and recovered from a DSGE model, the spirit of our exercise is to assume that we have access to other information (as in e.g. Romer and Romer (2004)) so that we can observe monetary policy shocks directly.

[8] When we use empirically identified shocks, measurement errors might be an issue. Given measurement errors, we show that asymptotic biases of our estimators are negative in Appendix D. Therefore, results here can be understood as conservative estimates. In addition, shocks are often estimated and thus are generated regressors, but if the researcher is interested in testing the null of no response then there is no need to adjust inference (Pagan 1984).

[9] The ordering of shocks in the VAR is TFP shock, output growth rate, inflation, monetary policy shock, fed funds rate.

Consistent with previous studies, we find (Figures 3 and 4) that a contractionary monetary policy shocks lowers output and prices, and that a positive TFP shock raises output and lowers prices. Impulse responses estimated with VAR and local projections are similar. VAR estimates for variance decomposition suggest that each of the shocks accounts for approximately 10 percent of variation in output. According to the VAR estimates, monetary policy shocks account for approximately 25 percent of variation in inflation at long horizons and little variation at short horizons while the contribution of TFP shocks is generally small. Bias correction makes no material difference for the variance decomposition estimates for all cases but one: the bias-corrected estimate of the contribution of monetary policy shocks to variation of inflation at long horizons is reduced to about 10 percent.

Local projections estimate that the contribution of the two shocks to variation of output is approximately twice as large as the contribution in VAR estimates. Consistent with simulations, bias correction tends to generate lower contributions but generally the magnitudes are similar. Specifically, when we use the s^{R2} estimator, monetary policy shocks account for approximately 20 percent of variation in output according to local projection estimates (25 percent without bias correction) and approximately 10 percent according to VAR estimates. While the LPB estimator yields similar results, the LPA estimator assigns a much larger role to the monetary policy shocks. This pattern reflects the fact that $\hat{s}_h^{LPA}$ may be greater than 1 in finite samples. Also, note that, in contrast to the profile of s_h estimated with VAR for output (which is generally flat after $h = 5$), $\hat{s}_h^{R2}$, $\hat{s}_h^{LPA}$ and $\hat{s}_h^{LPB}$ have richer dynamics.

In a similar spirit, the contribution of TFP and monetary policy shocks to variation in inflation is much greater according to our local-projections estimates. The difference is particularly large for monetary shocks: $\hat{s}_h^{R2}$ and $\hat{s}_h^{LPB}$ are close to 40 percent (after bias correction) and $\hat{s}_h^{VAR}$ is about 10 percent at long horizons. Again, $\hat{s}_h^{LPA}$ estimates an even greater contribution of monetary shocks and confidence intervals are much wider for $\hat{s}^{LPA}$ than for $\hat{s}^{LPB}$ or $\hat{s}^{R2}$. Again, this stems from the fact that $\hat{s}_h^{LPA}$ may be greater than 1 in finite samples.

V. Concluding remarks

Single-equation methods can offer flexibility and parsimony that many economists seek. The increasing popularity of these methods, specifically the local projections, calls for further development of these tools. An important limitation for practitioners using this framework has been a lack of simple tools to assess quantitative significance of a given set of shocks, that is, the contribution of the shocks to variance of the variable of interest. We propose several methods to provide such a metric. In a series of simulation exercises, we document that these methods have good small-sample properties. We also show that conventional approaches to assess the quantitative significance of two popular structural shocks (monetary policy shocks and total factor productivity shocks) could have been understated the importance of these two shocks.

References

Barakchian, S. Mahdi, and Christopher Crowe, 2013. "Monetary policy matters: Evidence from new shocks data," *Journal of Monetary Economics*, 60(8): 950-966.

Basu, Susanto, John G. Fernald, and Miles S. Kimball, 2006. "Are Technology Improvements Contractionary?" *American Economic Review,* 96(5): 1418–1448.

Coibion, Olivier, 2012. "Are the Effects of Monetary Policy Shocks Big or Small?" *American Economic Journal: Macroeconomics*, 4(2): 1–32.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia, 2017. "Innocent Bystanders? Monetary policy and inequality," *Journal of Monetary Economics*, 88(C): 70–89.

Fernald, John, 2014. "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," Working Paper 2012-19.

Jordà, Oscar, 2005. "Estimation and Inference of Impulse Responses by Local Projections," A*merican Economic Review*, 95(1): 161–182.

Pagan, Adrian, 1984. "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25(1): 221-247.

Plagborg-Møller, Mikkel, and Christian K. Wolf, 2017. "Instrumental Variable Identification of Dynamic Variance Decompositions," manuscript.

Ramey, Valerie A., 2011. "Identifying Government Spending Shocks: It's all in the Timing," *Quarterly Journal of Economics*, 126(1): 1–50.

Romer, Christina D., and David H. Romer, 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review*, 100(3): 763-801.

Romer, Christina, D., and David H. Romer, 2004. "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94(4): 1055-1084.

Sims, Christopher A., 1980. "Macroeconomics and reality." *Econometrica*, 48(1): 1-48.

Smets, Frank, and Rafael Wouters, 2007. "Shocks and frictions in US business cycles: A Bayesian DSGE approach," *American Economic Review*, 97(3): 586-606.

Stock, James, and Mark Watson, 2007. "Why Has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Banking and Credit*, 39(1): 3–33.

Wolf, Christian K., 2017. "Masquerading Shocks in Sign-Restricted VARs," manuscript.

Table 1. Parameter values for data generating processes (DGPs) used in simulations.

	$\psi_x(L)$	σ_x	g_y	ρ_p	σ_p	ρ_a	σ_a
DGP1	Hump-shaped	1	0.5	0.9	0.5	0.9	3
DGP2	$(1-0.9\mathrm{L})^{-1}$	3	0.5	0.9	1.5	-	-
DGP3	$(1-\mathrm{L})^{-1}(1-0.9\mathrm{L})^{-1}$	1	0.5	0.5	2	0.9	3

Table 2. Simulation results for DGP 1.

	Horizon h					
	0	4	8	12	16	20
Impulse response						
True	0.00	1.39	3.00	2.06	0.88	0.29
Local projections	0.00	1.36	2.99	2.03	0.85	0.29
VAR(HQIC)	0.00	0.18	0.26	0.27	0.27	0.27
Variance decomposition						
True	0.00	0.04	0.19	0.21	0.18	0.14
Average estimate						
R2	0.01	0.06	0.20	0.25	0.26	0.27
LP A	0.01	0.04	0.18	0.23	0.23	0.23
LP B	0.01	0.04	0.17	0.22	0.21	0.21
VAR(HQIC)	0.01	0.02	0.02	0.03	0.03	0.03
Root mean squared error						
R2	0.01	0.05	0.11	0.16	0.19	0.22
LP A	0.01	0.04	0.11	0.15	0.18	0.20
LP B	0.01	0.04	0.10	0.14	0.15	0.15
VAR(HQIC)	0.01	0.03	0.17	0.20	0.16	0.14
Coverage (90 % level) (asymptotic)						
R2	1.00	0.94	0.74	0.71	0.75	0.73
LP A	1.00	0.94	0.53	0.57	0.74	0.80
LP B	1.00	0.93	0.53	0.55	0.70	0.78
VAR(HQIC)	1.00	0.57	0.06	0.05	0.07	0.09
Variance decomposition (bias corrected, VAR(HQIC))						
True	0.00	0.04	0.19	0.21	0.18	0.14
Average estimate						
R2	0.00	0.02	0.13	0.16	0.13	0.11
LP A	0.00	0.02	0.14	0.17	0.16	0.14
LP B	0.00	0.02	0.14	0.17	0.15	0.13
VAR(HQIC)	0.00	0.00	0.01	0.02	0.02	0.02
Root mean squared error						
R2	0.01	0.05	0.13	0.16	0.17	0.18
LP A	0.01	0.04	0.12	0.16	0.17	0.18
LP B	0.01	0.04	0.12	0.14	0.15	0.15
VAR(HQIC)	0.01	0.04	0.19	0.21	0.18	0.15
Coverage (90 % level)						
R2	1.00	0.93	0.59	0.61	0.67	0.79
LP A	1.00	0.82	0.45	0.49	0.59	0.79
LP B	1.00	0.81	0.46	0.47	0.57	0.73
VAR(HQIC)	1.00	0.41	0.05	0.05	0.06	0.08

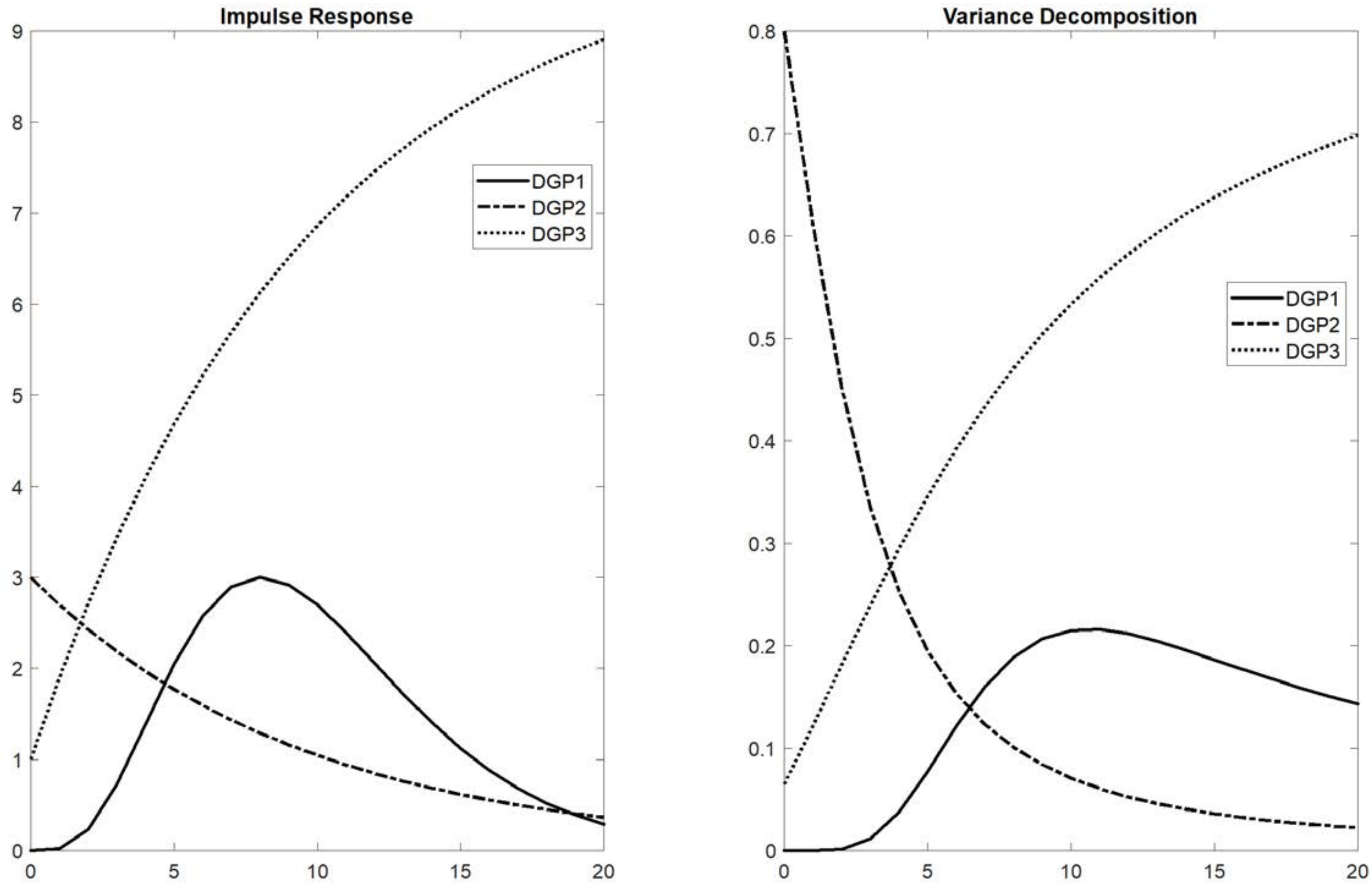
Table 3. Simulation results for DGP 2

	Horizon h					
	0	4	8	12	16	20
Impulse response						
True	3.00	1.97	1.29	0.85	0.56	0.36
Local projections	2.99	1.88	1.18	0.71	0.43	0.22
VAR(HQIC)	2.96	1.98	1.40	1.04	0.82	0.67
Variance decomposition						
True	0.80	0.25	0.10	0.05	0.03	0.02
Average estimate						
R2	0.79	0.27	0.15	0.14	0.15	0.18
LP A	0.80	0.27	0.13	0.10	0.09	0.09
LP B	0.79	0.26	0.13	0.09	0.09	0.09
VAR(HQIC)	0.80	0.27	0.13	0.08	0.06	0.05
Root mean squared error						
R2	0.03	0.11	0.12	0.14	0.17	0.21
LP A	0.03	0.09	0.08	0.08	0.09	0.11
LP B	0.03	0.08	0.07	0.08	0.09	0.10
VAR(HQIC)	0.03	0.08	0.07	0.06	0.05	0.05
Coverage (90 % level) (asymptotic)						
R2	0.92	0.87	0.92	0.89	0.85	0.80
LP A	0.93	0.90	0.94	0.95	0.93	0.93
LP B	0.90	0.88	0.93	0.94	0.92	0.90
VAR(HQIC)	0.90	0.88	0.91	0.97	0.99	0.99
Variance decomposition (bias corrected, VAR(HQIC))						
True	0.80	0.25	0.10	0.05	0.03	0.02
Average estimate						
R2	0.81	0.25	0.09	0.04	0.01	0.00
LP A	0.79	0.25	0.10	0.05	0.03	0.02
LP B	0.81	0.25	0.10	0.05	0.03	0.02
VAR(HQIC)	0.80	0.25	0.10	0.05	0.03	0.02
Root mean squared error						
R2	0.03	0.10	0.09	0.09	0.11	0.13
LP A	0.03	0.08	0.07	0.07	0.07	0.08
LP B	0.03	0.08	0.07	0.07	0.07	0.08
VAR(HQIC)	0.03	0.08	0.06	0.05	0.04	0.04
Coverage (90 % level)						
R2	0.91	0.90	0.97	0.98	0.98	0.97
LP A	0.93	0.88	0.91	0.98	0.97	0.97
LP B	0.89	0.83	0.89	0.97	0.97	0.96
VAR(HQIC)	0.89	0.88	0.89	0.92	0.99	1.00

Table 4. Simulation results for DGP 3 with alternative lag orders in VARs.

	Horizon h					
	0	4	8	12	16	20
Impulse response						
True	1.00	4.10	6.13	7.46	8.33	8.91
Local projections	0.99	3.96	5.78	6.86	7.44	7.66
VAR(5)	0.93	3.74	4.76	5.01	5.10	5.14
VAR(10)	0.92	3.65	5.34	6.05	6.17	6.23
Variance decomposition (bias corrected, VAR(5))						
True	0.06	0.29	0.47	0.58	0.65	0.70
Average estimate						
R2	0.06	0.26	0.41	0.50	0.55	0.58
LP A	0.05	0.24	0.38	0.48	0.54	0.58
LP B	0.06	0.25	0.40	0.49	0.54	0.57
VAR(5)	0.06	0.24	0.33	0.36	0.38	0.39
Root mean squared error						
R2	0.04	0.11	0.16	0.19	0.21	0.23
LP A	0.04	0.11	0.17	0.21	0.25	0.28
LP B	0.04	0.11	0.16	0.19	0.20	0.22
VAR(5)	0.04	0.11	0.19	0.26	0.31	0.34
Coverage (90 % level) (asymptotic)						
R2	0.77	0.80	0.80	0.81	0.82	0.83
LP A	0.85	0.83	0.82	0.83	0.84	0.85
LP B	0.84	0.79	0.78	0.78	0.79	0.80
VAR(5)	0.83	0.78	0.66	0.51	0.40	0.33
Variance decomposition (bias corrected, VAR(10))						
True	0.06	0.29	0.47	0.58	0.65	0.70
Average estimate						
R2	0.07	0.30	0.47	0.57	0.63	0.66
LP A	0.05	0.23	0.37	0.47	0.52	0.55
LP B	0.05	0.27	0.44	0.54	0.60	0.63
VAR(10)	0.06	0.27	0.42	0.50	0.54	0.56
Root mean squared error						
R2	0.05	0.12	0.16	0.19	0.21	0.22
LP A	0.04	0.12	0.17	0.21	0.25	0.29
LP B	0.04	0.12	0.16	0.18	0.20	0.21
VAR(10)	0.04	0.11	0.15	0.19	0.21	0.23
Coverage (90 % level)						
R2	0.72	0.76	0.78	0.80	0.82	0.83
LP A	0.87	0.87	0.89	0.90	0.91	0.92
LP B	0.85	0.77	0.75	0.76	0.78	0.79
VAR(10)	0.83	0.83	0.81	0.79	0.77	0.74

Figure 1. Population impulse responses and variance decomposition for each DGP



Notes: the left panel shows the impulse response functions for three bivariate data generating processes (DGPs). The right panel shows the contribution of the identified shock to variation of an outcome variable for the DGPs.

Figure 2: Smets and Wouters (2007) model, real GDP and monetary policy shock, T = 160.

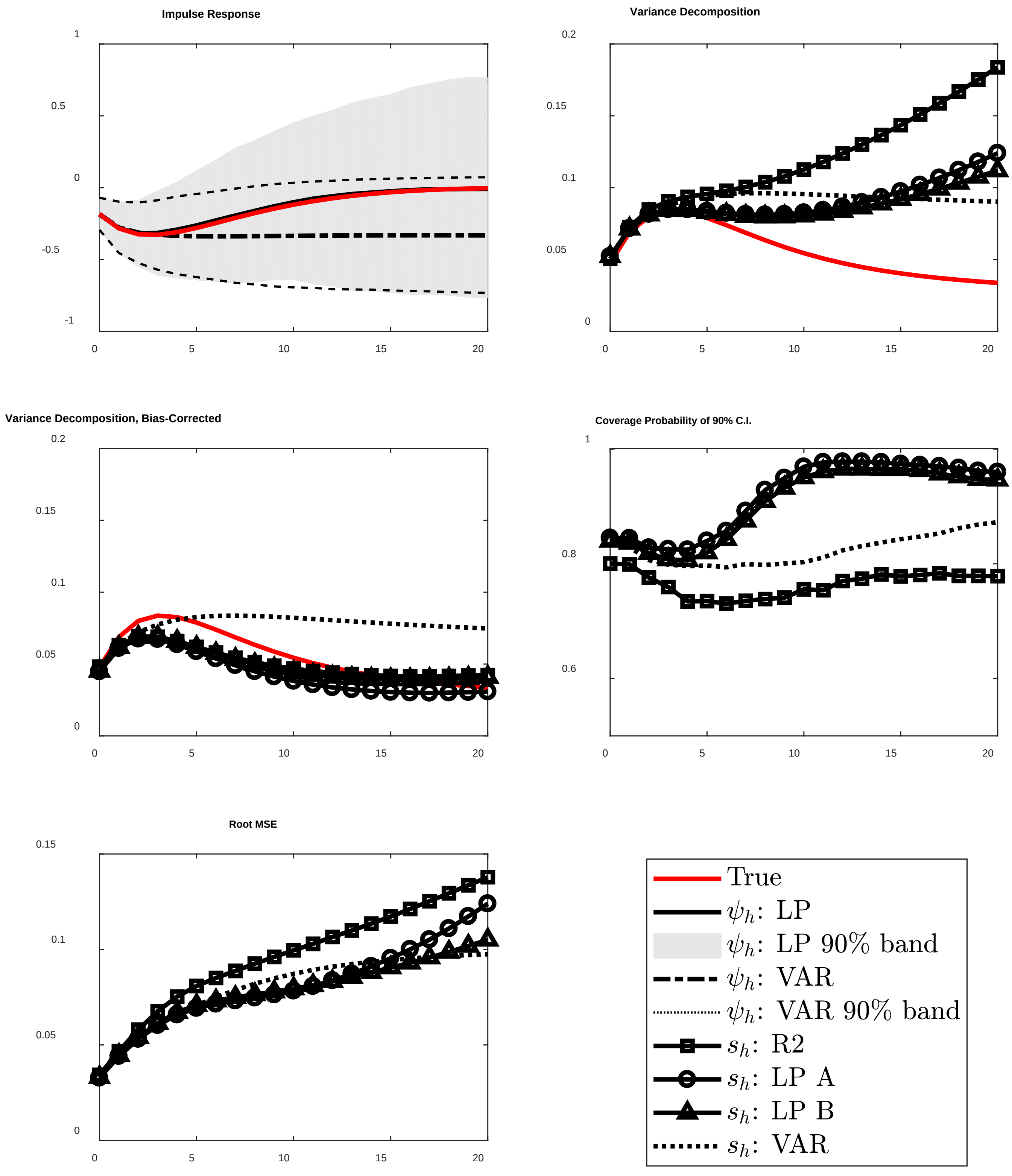


Figure 3 Real GDP. 1969:Q1-2007:Q4.

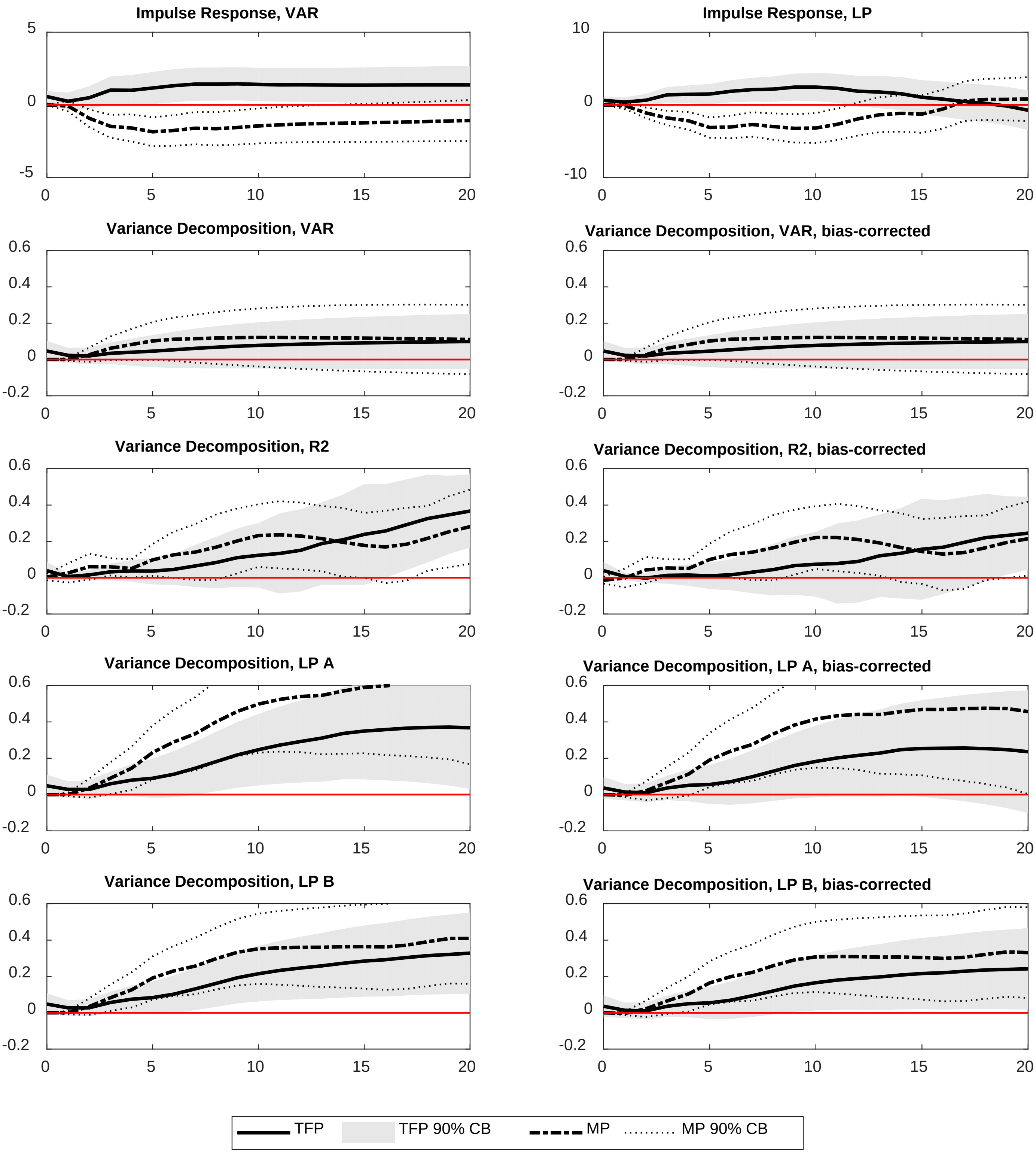
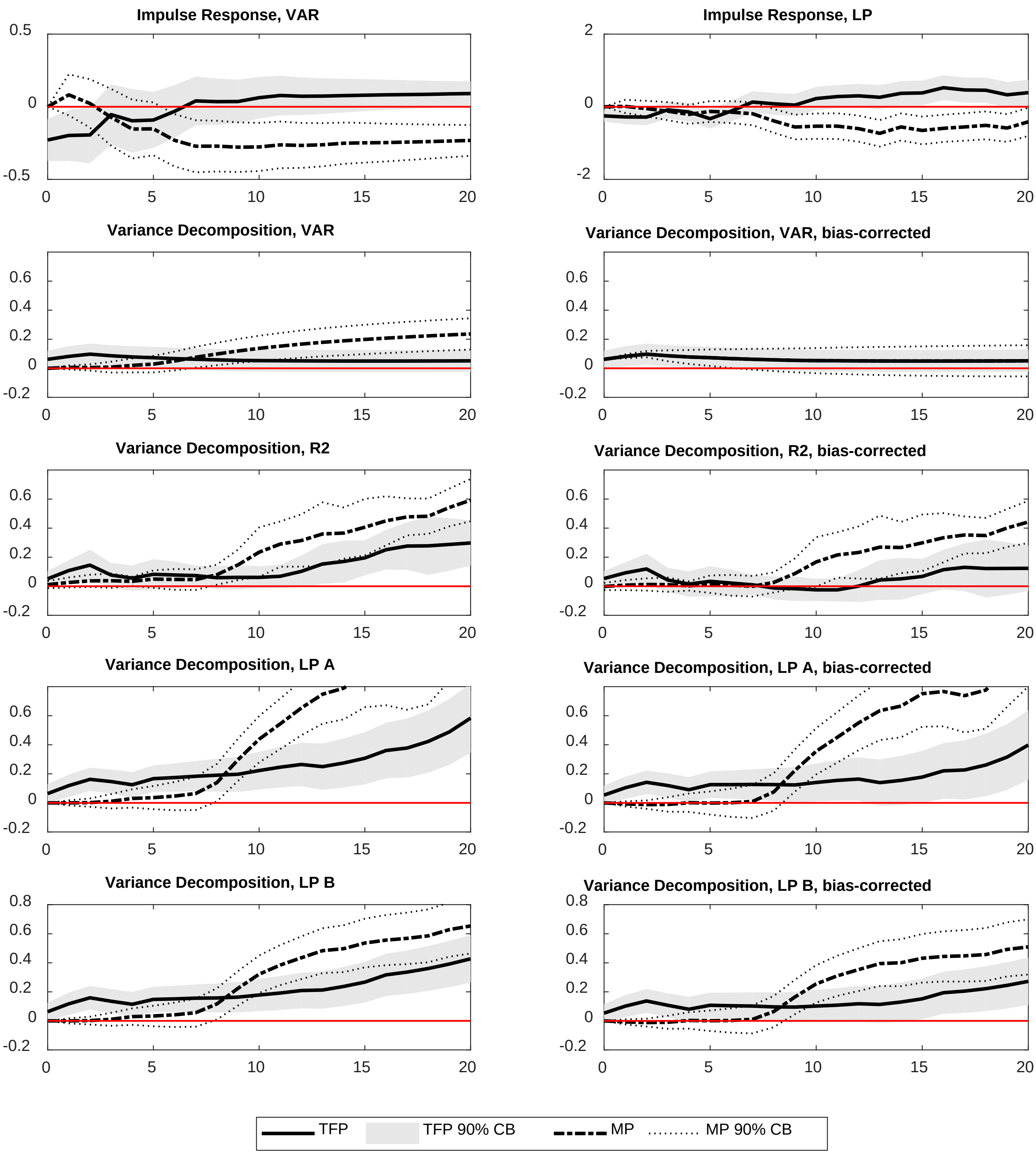


Figure 4. Inflation. 1969:Q1-2007:Q4.



APPENDIX FOR

A NOTE ON VARIANCE DECOMPOSITION WITH LOCAL PROJECTIONS

Yuriy Gorodnichenko
University of California – Berkeley
and NBER

Byoungchan Lee
University of California – Berkeley

Appendix A. Identification of e_t

In Section II, we derive the h-period ahead forecast error as following:

$$f_{t+h|t-1} = y_{t+h} - y_{t+h|t-1} = (y_{t+h} - y_{t-1}) - E[y_{t+h} - y_{t-1}|\Omega_{t-1}]$$
$$= \psi_{x,o}x_{t+h} + \cdots + \psi_{x,h}x_t + v_{t+h|t-1}$$

where $y_{t+h|t-1} \equiv E[y_{t+h}|\Omega_{t-1}]$ and $\Omega_{t-1} = \{\Delta y_{t-1}, x_{t-1}, \Delta y_{t-2}, x_{t-2}, \ldots\}$. In the footnote 2, we argue that given invertibility of $\psi_e(L)$, $v_{t+h|t-1} = \psi_{e,0}e_{t+h} + \cdots + (\psi_{e,0} + \cdots + \psi_{e,h})e_t$.

There is a technically subtle issue that the above forecast error seems to be based on the information set $\Omega_{t-1} \cup \{e_{t-1}, \ldots\}$, not the set of observables Ω_{t-1}. Thus, we need to prove that knowing e_t is redundant, once we have Ω_t. In other words, $\{e_{t-1}, e_{t-2}, \ldots\} \subset closure(span(\Omega_{t-1}))$.

Let's assume that we have only Ω_t. Following the idea of Jordà (2005), $\psi_{x,h}$ is identified as $\frac{Cov(y_t - y_{t-h-1},\ x_{t-h})}{Var(x_t)}$ for all h. This implies that Δz_t is identified, because

$$\Delta z_t = \Delta y_t - (1 - L)\psi_x(L)x_t.$$

The drift term of z_t is also easily identified because $g_y = E[\Delta z_t] = E[\Delta y_t]$, where $E[\cdot]$ is the unconditional expectation operator. Therefore,

$$w_t \equiv \psi_e(L)e_t = \Delta z_t - g_y \in \text{closure}(\text{span}(\Omega_t)).$$

Finally, it follows from the uniqueness of the Wold decomposition[1] that

$$e_t = w_t - Projection(w_t|w_{t-1}, w_{t-2}, \ldots)$$

and $\psi_{e,h} = \frac{Cov(w_t - w_{t-h-1}, e_{t-h})}{Var(e_t)}$ for all h, where $Projection(a|A)$ is defined by the orthogonal projection of a vector a in a Hibert space to a closed subspace generated by a set of vectors A, $closure(\text{span}(A))$.[2] Therefore, $\{e_t, e_{t-1}, \ldots\} \subset closure(span(\Omega_t))$, and specifically, $E[y_{t+h}|\Omega_{t-1}] = \text{E}[y_{t+h}|\Omega_{t-1} \cup \{e_{t-1}, e_{t-2}, \ldots\}]$.

This result illustrates how we can back out $f_{t+h|t-1}$ in practice. First, we consider $y_{t+h} - y_{t-1}$:

$$y_{t+h} - y_{t-1} = \psi_{x,0}x_{t+h} + \psi_{x,1}x_{t+h-1} + \cdots + \psi_{x,h}x_t + [(S^*)^{h+1} - I]\psi_x(L)x_{t-1} + z_{t+h} - z_{t-1},$$

[1] See Brockwell and Davis (1991) for details.
[2] See Conway (1990) for details on projections.

where S^* is the adjoint operator of the unilateral shift on $l^2(N)$ and I is the identity operator. In other words,

$$S(\psi_0, \psi_1, \dots) = (0, \psi_0, \psi_1, \dots), \quad and \quad S^*(\psi_0, \psi_1, \dots) = (\psi_1, \psi_2, \dots).$$

For simple notations, we additionally assume that $\sum_{j=0}^{\infty} j \cdot |\psi_{e,j}| < \infty$. This condition holds for any stationary ARMA processes. By applying the Beveridge-Nelson decomposition, we obtain the followings:

$$z_t = \Delta z_t + \cdots + \Delta z_1 + z_0 = g_y \cdot t + \psi_e(1) \cdot (e_t + \cdots + e_1) + \zeta_t - \zeta_0 + z_0,$$

where $\psi_e(1) = \sum_{j=0}^{\infty} \psi_{e,j}$, $\zeta_t = \sum_{j=0}^{\infty} \bar{\psi}_j e_{t-j}$, $\overline{\psi_j} = -(\psi_{j+1} + \psi_{j+2} + \cdots)$, and $\sum_{j=0}^{\infty} |\overline{\psi_j}| < \infty$. Thus, we can rewrite $z_{t+h} - z_{t-1}$ by $\psi_e(1) \cdot (e_{t+h} + \cdots + e_t) + \zeta_{t+h} - \zeta_{t-1}$.[3] Finally,

$$E(y_{t+h} - y_{t-1} | \Omega_{t-1}) = (\psi_{x,h+1} - \psi_{x,0})x_{t-1} + (\psi_{x,h+2} - \psi_{x,1})x_{t-2} + \cdots +$$
$$E(\zeta_{t+h} - \zeta_{t-1} | \Omega_{t-1}).$$

This illustrates what we actually do when we try to estimate the forecast errors by taking residuals after regressing $y_{t+h} - y_{t-1}$ on $\Delta y_{t-1}, x_{t-1}$ and their lagged values. In the regression, x_{t-1} and its lagged values control for two things. First thing to be captured is the component directly related to $\{x_{t-1}\}$ through $\psi_x(L)$, which is $[(S^*)^{h+1} - I]\psi_x(L)x_{t-1} = (\psi_{x,h+1} - \psi_{x,0})x_{t-1} + (\psi_{x,h+2} - \psi_{x,1})x_{t-2} + \cdots$ in the above expression. Moreover, $(1 - L)\psi_x(L)x_{t-1}$ in Δy_{t-1} is also controlled, generating $w_{t-1} = \psi_e(L)e_{t-1}$. A closed subspace generated by w_{t-1} and its lagged values will be the same as that by $e_{t-1}, e_{t-2}, \dots$. Finally, this part of the projection will control for $E(\zeta_{t+h} - \zeta_{t-1} | \Omega_{t-1})$ because ζ_{t-1} is a limit of linear combinations of $\{e_{t-1}, e_{t-2}, \dots\}$. This completes purification of the $y_{t+h} - y_{t-1}$ to $f_{t+h|t-1}$.

---

[3] Of course, we can proceed without the additional assumption. In that case, notations become messy because everything should be written in terms of e_t's instead of $\zeta_{t+h} - \zeta_{t-1}$.

Appendix B1. Proof of Proposition 1 and implementation detail

Proposition 1. Suppose $f_h = \left(f_{T|T-h-1}, f_{T-1|T-h-2}, \dots, f_{L_{max}+h+1\,|L_{max}}\right)'$ and $X_h = \left(X_{T-h}^h, X_{T-1}^h, \dots, X_{L_{max}+1}^h\right)'$ for all $h \geq 0$ where $L_{max} = \max\{L_x, L_y\}$. Then the R^2 of the regression of $f_{t+h|t-1}$ on X_t^h, given by $(f_h' P_{X_h} f_h)/(f_h' f_h)$ where $P_{X_h} = X_h(X_h'X_h)^{-1}X_h'$, has the following asymptotic distribution for some V_{h,R^2}:

$$\sqrt{T}\left(\frac{f_h' P_{X_h} f_h}{f_h' f_h} - s_h\right) \xrightarrow{d} \mathcal{N}(0, \quad V_{h,R^2}).$$

Proof. Although $f_{t+h|t-1}$ is a time $t+h$ variable, not time t, we can proceed without loss of validity of the results below by considering the moment conditions below as time $t+h$ conditions, not t. We use this notation instead of $f_{t|t-h-1}$ for consistency of the presentation.

Let $\theta_0 = \left(\theta_{1,0}', \theta_{2,0}', \theta_{3,0}'\right)'$ where

$$X_t^h = (x_{t+h}, \dots, x_t)',$$

$$\theta_{1,0} = \left(E[X_t^h X_t^{h\prime}]\right)^{-1}\left(E\left[X_t^h f_{t+h|t-1}\right]\right) = \left(\psi_{x,0}, \psi_{x,1}, \dots, \psi_{x,h}\right)',$$

$$\theta_{2,0} = E\left[X_t^h f_{t+h|t-1}\right] = \theta_{1,0}\sigma_x^2,$$

$$\theta_{3,0} = E\left[f_{t+h|t-1}^2\right] \equiv \sigma_{f,h}^2.$$

We define the method of moments estimator $\hat{\theta} = \left(\hat{\theta}_1', \hat{\theta}_2', \hat{\theta}_3'\right)'$ as following:

$$\hat{\theta}_1 = (X_h'X_h)^{-1}(X_h'f_h), \qquad \hat{\theta}_2 = \frac{X_h' f_h}{T_h}, \qquad \hat{\theta}_3 = \frac{f_h' f_h}{T_h}$$

where $T_h = T - (L_{max} + 1)$. It follows that $s_h = \xi(\theta_0)$ and $\frac{f_h' P_{X_h} f_h}{f_h' f_h} = \xi(\hat{\theta})$ where $\xi(\theta) = \xi(\theta_1, \theta_2, \theta_3) = \frac{\theta_2' \theta_1}{\theta_3}$. Therefore, we first derive the asymptotic distribution of $\sqrt{T}(\hat{\theta} - \theta_0)$ and then apply the delta method.

To begin, we consider the moment conditions that $E\left[g\left(f_{t+h|t-1}, X_t^h, \theta\right)\right] = 0$ where

$$g_{t+h}(\theta) \equiv g\left(f_{t+h|t-1}, X_t^h, \theta\right) = \begin{pmatrix} X_t^h\left(f_{t+h|t-1} - \left(X_t^h\right)'\theta_1\right) \\ X_t^h f_{t+h|t-1} - \theta_2 \\ f_{t+h|t-1}^2 - \theta_3 \end{pmatrix}.$$

It is clear that the conditions are satisfied only when $\theta = \theta_0$ and the system is just-identified. As shown by Hansen (1982), we know that

$$\sqrt{T}\left(\hat{\theta}-\theta_0\right) \overset{d}{\to} \mathcal{N}(0, \quad G^{-1}\Omega(G')^{-1})$$

where $G = E[\nabla_\theta g_{t+h}(\theta_0)]$ and $\Omega = \sum_{l=-\infty}^{\infty}\Gamma(l)$ and $\Gamma(l)$ is the autocovariance of $g_{t+h}(\theta_0)$ at lag l. With some algebra, we can show that $G = -diag\left(E\left[X_t^h\left(X_t^h\right)'\right], I_{h+2}\right)$ where $diag(A, B)$ is the block diagonal matrix whose diagonal components are A and B in order.

Regarding the delta method, we define $\Delta \equiv \frac{\partial\xi(\theta_0)}{\partial\theta'} = \frac{1}{\theta_{3,0}}\left(\theta_{2,0}', \theta_{1,0}', -s_h\right)$. Combining the above derivations, and being explicit about the fact that the moment conditions $g_{t+h}(\cdot)$ are for the R^2 approach at horizon h, we have the desired result.

$$\sqrt{T}\left(\frac{f_h' P_{X_h} f_h}{f_h' f_h} - s_h\right) \overset{d}{\to} \mathcal{N}\left(0, \quad V_{h,R^2}\right)$$

where $V_{h,R^2} = \Delta_{h,R^2}\left(G_{h,R^2}\right)^{-1}\Omega_{h,R^2}\left(G_{h,R^2}'\right)^{-1}\Delta_{h,R^2}'$. □

Implementation. We discuss how to implement Proposition 1. First of all, we use $\hat{f}_{t+h|t-1}$ obtained from Equation (6) instead of $f_{t+h|t-1}$ in practice, because $f_{t+h|t-1}$ is not observable. Then $\hat{f}_h = \left(\hat{f}_{T|T-h-1}, \dots, \hat{f}_{L_{max}+h+1\,|L_{max}}\right)'$, and $\hat{s}_h^{R2}$ is given by $\left(\hat{f}_h' P_{X_h}\hat{f}_h\right)/(\hat{f}_h'\hat{f}_h)$.

We also need to estimate V_{h,R^2} because it depends on the population parameters. Let's begin with Δ_{h,R^2}. A practically feasible estimator of θ we use is $\tilde{\theta} = \left(\tilde{\theta}_1', \tilde{\theta}_2', \tilde{\theta}_3'\right)'$ where

$$\tilde{\theta}_1 = (X_h'X_h)^{-1}\left(X_h'\hat{f}_h\right), \qquad \tilde{\theta}_2 = \frac{X_h'\hat{f}_h}{T_h}, \qquad \tilde{\theta}_3 = \frac{\hat{f}_h'\hat{f}_h}{T_h}.$$

A natural estimator of Δ_{h,R^2} is $\widehat{\Delta}_{h,R^2} \equiv \frac{\partial\xi(\tilde{\theta})}{\partial\theta'} = \frac{1}{\tilde{\theta}_3}\left(\tilde{\theta}_2', \tilde{\theta}_1', -\hat{s}_h^{R2}\right)$. The last element is based on a bias-corrected estimates instead of $\xi\left(\tilde{\theta}\right)$ because we find that this specification provides better performances in simulations.[4] How to obtain the bias-corrected estimator $\hat{s}_h^{R2}$ in this set-up will be discussed later.

We next turn to $G = -diag\left(E\left[X_t^h\left(X_t^h\right)'\right], I_{h+2}\right)$. It can be easily estimated by $\hat{G}_{h,R^2} = -diag(X_h'X_h/T_h, I_{h+2}) = -diag\left(\sum_{t=L_{max}+1}^{T-h} X_t^h\left(X_t^h\right)'/T_h, I_{h+2}\right)$.

It remains to estimate $\Omega_{h,R^2} = \sum_{l=-\infty}^{\infty}\Gamma(l)$ where $\Gamma(l)$ is the autocovariance of $g_t(\theta_0)$ at lag l. We use the pre-whitening procedure following Andrews and Monahan (1992) to avoid

[4] Results are available upon requests.

underestimation problem of the long-run variance of $g_{t+h}(\theta_0)$. To that end, we define a $2h+3$ dimensional vector Z_{t+h} as following:

$$Z_{t+h} \equiv \begin{pmatrix} X_t^h\left(\hat{f}_{t+h|t-1} - (X_t^h)'\tilde{\theta}_1\right) \\ X_t^h\hat{f}_{t+h|t-1} - \tilde{\theta}_2 \\ \hat{f}_{t+h|t-1}^2 - \tilde{\theta}_3 \end{pmatrix}.$$

It is worth noting that the sample average of Z_t is a zero vector, *i.e.* $\frac{1}{T_h}\sum_{t=L_{max}+1}^{T-h} Z_t = 0$ given the definition of $\tilde{\theta}$. To whiten the series, we use a VAR(1) that $Z_t = AZ_{t-1} + U_t$. The estimated autoregressive matrix and the residual are denoted by $\hat{A}$ and $\widehat{U}_t$. Then we estimate the long-run variance of $\widehat{U}_t$ by applying the method suggested by Newey and West (1987) with the Bartlett kernel to the residuals $\{\widehat{U}_{L_{max}+2}, \ldots, \widehat{U}_{T-h}\}$. Specifically, the estimated long-run variance is given by

$$\widehat{LRV}(\widehat{U}_t) = \hat{\Gamma}_{U,0} + \frac{1}{L_{NW}+1}\left(\hat{\Gamma}_{U,1} + \hat{\Gamma}_{U,-1}\right) + \cdots + \frac{L_{NW}}{L_{NW}+1}\left(\hat{\Gamma}_{U,L_{NW}} + \hat{\Gamma}_{U,-L_{NW}}\right)$$

where $\hat{\Gamma}_{U,l}$ is the estimated autocovariance matrix of U_t at lag l. We use a simple rule suggested by Stock and Watson (2011) to select the number of autocovariance matrices to be included in the estimation. Following the rule, $L_{NW}+1$ is the closest natural number of $0.75T_h^{1/3}$. Finally, $\widehat{\Omega}_{h,R^2}$ is obtained by $\left(I_{2h+3} - \hat{A}\right)^{-1}\widehat{LRV}(\widehat{U}_t)\left(I_{2h+3} - \hat{A}'\right)^{-1}$.

In sum, the asymptotic standard error of $\hat{s}_h^{R2}$ is given by

$$[s.e.(\hat{s}_h^{R2})]^2 = \frac{1}{T_h}\widehat{\Delta}_{h,R^2}\left(\hat{G}_{h,R^2}\right)^{-1}\widehat{\Omega}_{h,R^2}\left(\hat{G}'_{h,R^2}\right)^{-1}\widehat{\Delta}'_{h,R^2}$$

$$\text{where } \widehat{\Delta}_{h,R^2} = \frac{1}{\tilde{\theta}_3}\left(\tilde{\theta}'_2, \tilde{\theta}'_1, -\hat{s}_h^{R2}\right),$$

$$\hat{G}_{h,R^2} = -diag\left(\frac{X'_hX_h}{T_h}, I_{h+1}\right),$$

$$\widehat{\Omega}_{h,R^2} = \left(I_{2h+3} - \hat{A}\right)^{-1}\widehat{LRV}(\widehat{U}_t)\left(I_{2h+3} - \hat{A}'\right)^{-1}.$$

Bias-correction. We conjecture that most of the finite sample bias is due to the non-linear transformation $\xi(\cdot)$, not estimation of θ. Note that θ consists of projection coefficients of $f_{t+h|t-1}$ on X_t^h, covariance between $f_{t+h|t-1}$ and X_t^h, and variance of $f_{t+h|t-1}$. Estimation of all three quantities are rather standard, and significant biases regarding the method of moments estimator have not been reported. Below we suggest a method to capture biases originating from $\xi(\cdot)$ in small samples.

Because $\sqrt{T}\left(\hat{\theta}-\theta_0\right) \overset{d}{\to} \mathcal{N}(0, \quad G^{-1}\Omega(G')^{-1})$, we can approximate the asymptotic variance of the feasible estimator $\tilde{\theta}$ by $\frac{1}{T_h}\left(\hat{G}_{h,R^2}\right)^{-1}\widehat{\Omega}_{h,R^2}\left(\hat{G}'_{h,R^2}\right)^{-1}$. We simulate θ^b for B times from the following normal distribution:

$$\theta^b \sim \mathcal{N}\left(\tilde{\theta}, \qquad \frac{1}{T_h}\left(\hat{G}_{h,R^2}\right)^{-1}\widehat{\Omega}_{h,R^2}\left(\hat{G}'_{h,R^2}\right)^{-1}\right).$$

We discard cases with $\theta_3^b \leq 0$ while drawing θ^b's, because $\theta_3 = Var\left(f_{t+h|t-1}\right)$. Then the bias is estimated by

$$bias_h^{R2} \equiv \overline{\xi(\theta^b)} - \xi\left(\tilde{\theta}\right), \quad \text{where} \quad \overline{\xi(\theta^b)} \equiv \frac{1}{B}\sum_{b=1}^{B}\xi(\theta^b).$$

Finally, the bias-corrected estimator is given by

$$\hat{s}_h^{R2} = \xi\left(\tilde{\theta}\right) - bias_h^{R2} = 2\xi\left(\tilde{\theta}\right) - \overline{\xi(\theta^b)}.$$

Appendix B2. Proof of Proposition 2 and implementation details

Proposition 2. The local projections based estimators when $f_{t+h|t-1}$ is observable have the following asymptotic distributions for some $V_{h,LPA}$ and $V_{h,LPB}$:

$$\sqrt{T}\left(\frac{\sum_{i=0}^{h}\left(\hat{\beta}_0^{i,LP}\right)^2 \hat{\sigma}_x^2}{\widehat{Var}(f_{t+h|t-1})} - s_h\right) \xrightarrow{d} \mathcal{N}\left(0,\ \ V_{h,LPA}\right), \quad \text{and}$$

$$\sqrt{T}\left(\frac{\sum_{i=0}^{h}\left(\hat{\beta}_0^{i,LP}\right)^2 \hat{\sigma}_x^2}{\sum_{i=0}^{h}\left(\hat{\beta}_0^{i,LP}\right)^2 \hat{\sigma}_x^2 + \widehat{Var}\left(f_{t+h|t-1} - \sum_{i=0}^{h}\hat{\beta}_0^{i,LP}\, x_{t+h-i}\right)} - s_h\right) \xrightarrow{d} \mathcal{N}\left(0,\ \ V_{h,LPB}\right)$$

Proof. Similar to Proposition 1, the moment conditions below should be understood as time $t+h$ conditions, not time t.

(i) LP-A estimator

We first derive the joint distribution of $\hat{\psi}_{x,i}$'s, $\hat{\sigma}_x^2$, and $\hat{\sigma}_{f,h}^2 \equiv \widehat{Var}\left(f_{t+h|t-1}\right)$. Then we will use the delta method to find $V_{h,LPA}$.

To begin, we describe the moment conditions for the local projections for $\hat{\psi}_{x,i}$'s. We run the following OLS regression and take the coefficient on x_t:

$$y_{t+i} - y_{t-1} = \beta_0^i x_t + \cdots + \beta_{J^{LP}}^i x_{t-J^{LP}} + \gamma_1^i \Delta y_{t-1} + \cdots + \gamma_{I^{LP}}^i \Delta y_{t-I^{LP}} + c_i + r_{t+i|t-1}$$

for all $i = 0,1,\dots,h$. In the above representation, $\beta_0^i = \psi_{x,i}$. For a simple notation, we rewrite the above equation by

$$p_{i,t} = q_t' \mathrm{B}_i + r_{t+i|t-1}$$

$$\text{where} \quad p_{i,t} = y_{t+i} - y_{t-1},$$

$$q_t = \left(1, x_t, \dots, x_{t-J^{LP}}, \Delta y_{t-1}, \dots, \Delta y_{t-I^{LP}}\right)',$$

$$\mathrm{B}_i = \left(c_i, \beta_0^i, \dots, \beta_{J^{LP}}^i, \gamma_1^i, \dots, \gamma_{I^{LP}}^i\right)'.$$

Then the OLS estimator $\widehat{\mathrm{B}}_i$ becomes the method of moments estimator of the following moment conditions that

$$E\left[q_t\left(p_{i,t} - q_t'\mathrm{B}_i\right)\right] = 0.$$

Also, $\hat{\psi}_{x,i}$ is given by $\iota_1'\widehat{\mathrm{B}}_i$ where ι_1 is a $I^{LP} + J^{LP} + 2$ dimensional vector whose first element is one and the others are zero.

To study all parameters required simultaneously, we let $\theta_0 = \left(\mathrm{B}_0', \ldots, \mathrm{B}_h', \sigma_x^2, \sigma_{f,h}^2\right)'$ where $\sigma_{f,h}^2 \equiv Var\left(f_{t+h|t-1}\right)$. We use the moment conditions such that $E[g_{t+h}(\theta_0)] = 0$ where $g_{t+h}(\theta_0)$ is given as following:

$$g_{t+h}(\theta_0) = \begin{pmatrix} q_t\left(p_{0,t} - q_t'\mathrm{B}_0\right) \\ \vdots \\ q_t\left(p_{h,t} - q_t'\mathrm{B}_h\right) \\ x_t^2 - \sigma_x^2 \\ f_{t+h|t-1}^2 - \sigma_{f,h}^2 \end{pmatrix}.$$

We define $g_{t+h}(\theta)$ similarly. It is clear that it is a just-identified system. Similar to the proof of Proposition 1, we know that

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, \quad G^{-1}\Omega(G')^{-1})$$

where $G = E[\nabla_\theta g_{t+h}(\theta_0)]$ and $\Omega = \sum_{l=-\infty}^{\infty} \Gamma(l)$ and $\Gamma(l)$ is the autocovariance of $g_{t+h}(\theta_0)$ at lag l. With some algebra, we can show that

$$G = -E\begin{pmatrix} I_{h+1} \otimes q_t q_t' & 0 \\ 0 & I_2 \end{pmatrix}$$

where $\otimes$ is the Kronecker's product.

A transformation ξ is required to connect θ with s_h. We define

$$\xi(\theta_0) = s_h = \frac{\sum_{i=0}^{h} (\iota_1'\mathrm{B}_i)^2 \ \sigma_x^2}{\sigma_{f,h}^2}.$$

$\xi(\theta)$ is also defined similarly.

Regarding the delta method, we need $\Delta \equiv \frac{\partial \xi(\theta_0)}{\partial \theta'}$. With some algebra, we show that

$$\Delta = \frac{1}{\sigma_{f,h}^2} \begin{pmatrix} 2\psi_{x,0}\sigma_x^2 \iota_1 \\ \vdots \\ 2\psi_{x,h}\sigma_x^2 \iota_1 \\ \sum_{i=0}^{h} \psi_{x,i}^2 \\ -s_h \end{pmatrix}'.$$

Combining the above derivations, and being explicit about the fact that the moment conditions $g_{t+h}(\cdot)$ are for the LP-A approach at horizon h, we have the desired result.

$$\sqrt{T}\left(\frac{\sum_{i=0}^{h}\left(\hat{\psi}_{x,i}\right)^2 \hat{\sigma}_x^2}{\widehat{Var}\left(\hat{f}_{t+h|t-1}\right)} - s_h\right) \xrightarrow{d} \mathcal{N}\left(0, \quad V_{h,LPA}\right).$$

where $V_{h,LPA} = \Delta_{h,LPA}\left(G_{h,LPA}\right)^{-1}\Omega_{h,LPA}\left(G'_{h,LPA}\right)^{-1}\Delta'_{h,LPA}$. □

(ii) LP-B estimator

The joint distribution of $\hat{\psi}_{x,i}$'s is obtained similarly. To study all parameters required simultaneously, we let $\theta_0 = \left(\mathrm{B}'_0, \dots, \mathrm{B}'_h, \sigma_x^2, \sigma_{v,h}^2\right)'$ where $\sigma_{v,h}^2 \equiv Var\left(f_{t+h|t-1} - \sum_{i=0}^{h} \psi_{x,i}\, x_{t+h-i}\right)$. We use the moment conditions such that $E[g_{t+h}(\theta_0)] = 0$ where $g_{t+h}(\theta_0)$ is given as following:

$$g_{t+h}(\theta_0) = \begin{pmatrix} q_t\left(p_{0,t} - q'_t \mathrm{B}_0\right) \\ \vdots \\ q_t\left(p_{h,t} - q'_t \mathrm{B}_h\right) \\ x_t^2 - \sigma_x^2 \\ \left(f_{t+h|t-1} - \sum_{i=0}^{h} (\iota'_1 \mathrm{B}_i)\, x_{t+h-i}\right)^2 - \sigma_{v,h}^2 \end{pmatrix}.$$

We define $g_{t+h}(\theta)$ similarly. It is clear that it is a just-identified system. In such a case, the method of moments estimator $\hat{\theta}$ can be understood as a two-step estimator. It first find $\widehat{\mathrm{B}}_i$'s using the OLS moment conditions and then plug the estimates into the remaining conditions. Then $\hat{\sigma}_x^2$ and $\hat{\sigma}_{v,h}^2$ are derived given $\widehat{\mathrm{B}}_i$'s. It is worth noting that this is the same procedure we follow when we define $\hat{s}_h^{LPB}$. The only difference is that we are using here $f_{t+h|t-1}$ instead of its estimate.

Similar to the proof of Proposition 1, we know that

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, \quad G^{-1}\Omega(G')^{-1})$$

where $G = E[\nabla_\theta g_{t+h}(\theta_0)]$ and $\Omega = \sum_{l=-\infty}^{\infty} \Gamma(l)$ and $\Gamma(l)$ is the autocovariance of $g_{t+h}(\theta_0)$ at lag l. With some algebra, we can show that

$$G = -E\left(\begin{array}{ccc|c} & I_{h+1} \otimes q_t q'_t & & 0 \\ \hline 0 & \cdots & 0 & \\ 2v_{t+h|t-1}x_{t+h}\iota'_1 & \cdots & 2v_{t+h|t-1}x_{t+1}\iota'_1 & I_2 \end{array}\right)$$

where $\otimes$ is the Kronecker's product. For the bottom left part, we use the fact that $\iota'_1 \mathrm{B}_i = \psi_{x,i}$ and $v_{t+h|t-1} = f_{t+h|t-1} - \sum_{i=0}^{h} \psi_{x,i} x_{t+h-i}$. Because $v_{t+h|t-1} = \psi_{e,0} e_{t+h} + \cdots + \left(\psi_{e,0} + \cdots + \psi_{e,h}\right) e_t$ is orthogonal to $\{x_t\}$, the bottom left block of G becomes a zero matrix.

A transformation ξ is required to connect θ with s_h. We define

$$\xi(\theta_0) = s_h = \frac{\sum_{i=0}^{h} (\iota_1' \mathrm{B}_i)^2 \ \sigma_x^2}{\sum_{i=0}^{h} (\iota_1' \mathrm{B}_i)^2 \ \sigma_x^2 + \sigma_{v,h}^2}.$$

$\xi(\theta)$ is also defined similarly.

Regarding the delta method, we need $\Delta \equiv \frac{\partial \xi(\theta_0)}{\partial \theta'}$. For a simple notation, we write $\sigma_{f,h}^2 \equiv Var\left(f_{t+h|t-1}\right) = \sum_{i=0}^{h} \psi_{x,i}^2 \sigma_x^2 + \sigma_{v,h}^2$. With some algebra, we show that

$$\Delta = \frac{1 - s_h}{\sigma_{f,h}^2} \begin{pmatrix} 2\psi_{x,0}\sigma_x^2 \iota_1 \\ \vdots \\ 2\psi_{x,h}\sigma_x^2 \iota_1 \\ \sum_{i=0}^{h} \psi_{x,i}^2 \\ -s_h/(1 - s_h) \end{pmatrix}'.$$

Combining the above derivations, and being explicit about the fact that the moment conditions $g_{t+h}(\cdot)$ are for the LPB approach at horizon h, we have the desired result.

$$\sqrt{T}\left(\frac{\sum_{i=0}^{h} \left(\hat{\beta}_0^{i,LP}\right)^2 \hat{\sigma}_x^2}{\sum_{i=0}^{h} \left(\hat{\beta}_0^{i,LP}\right)^2 \hat{\sigma}_x^2 + \widehat{Var}\left(f_{t+h|t-1} - \sum_{i=0}^{h} \hat{\beta}_0^{i,LP} x_{t+h-i}\right)} - s_h \right) \xrightarrow{d} \mathcal{N}\left(0, \quad V_{h,LPB}\right).$$

where $V_{h,LPB} = \Delta_{h,LPB}\left(G_{h,LPB}\right)^{-1} \Omega_{h,LPB} \left(G_{h,LPB}'\right)^{-1} \Delta_{h,LPB}'$. □

Joint inference. In the below, we explain how to obtain joint distribution of the LP-B estimator $(\hat{s}_0^{LPB}, \hat{s}_1^{LPB}, \dots, \hat{s}_H^{LPB})'$. Results for the LP-A estimator can be obtained similarly.

We consider augmented moment conditions that $E\left[g_{t+H}^{Joint}(\theta_0)\right] = 0$ where $\theta_0 = (\mathrm{B}_0', \dots, \mathrm{B}_H', \sigma_x^2, \sigma_{v,0}^2, \dots, \sigma_{v,H}^2)'$ is a $(H+1) * (I^{LP} + J^{LP} + 3) + 1$ dimensional vector, and

$$g_{t+H}^{Joint}(\theta_0) = \begin{pmatrix} q_t\left(p_{0,t} - q_t' \mathrm{B}_0\right) \\ \vdots \\ q_t\left(p_{H,t} - q_t' \mathrm{B}_H\right) \\ x_t^2 - \sigma_x^2 \\ \left(f_{t|t-1} - \sum_{i=0}^{0} (\iota_1' \mathrm{B}_i)\, x_{t-i}\right)^2 - \sigma_{v,0}^2 \\ \vdots \\ \left(f_{t+H|t-1} - \sum_{i=0}^{H} (\iota_1' \mathrm{B}_i)\, x_{t+H-i}\right)^2 - \sigma_{v,H}^2 \end{pmatrix}.$$

Then it is straightforward to extend Proposition 2 to the joint distribution of $(\hat{s}_0^{LP}, \hat{s}_1^{LP}, \ldots, \hat{s}_H^{LP})'$. In practice, both I^{LP} and J^{LP} should be small not to make $(H+1)*(I^{LP}+J^{LP}+3)+1$ too large relative to available sample sizes.

Implementation. We discuss how to implement Proposition 2. In the below, we focus on the LP-B estimator. Again, the LP-A estimator can be implemented in a similar way.

First of all, we use $\hat{f}_{t+h|t-1}$ from Equation (6) instead of $f_{t+h|t-1}$ in practice.

We also need to estimate $V_{h,LPB}$ because it depends on population parameters. Let's begin with $\mathrm{G}_{h,LPB} = -diag(I_{h+1} \otimes E[q_t q_t'], I_2)$. It is natural to have

$$\hat{G}_{h,LPB} = -diag\left(I_{h+1} \otimes \frac{1}{T_h} \sum_{t=L_{max}+1}^{T-h} q_t {q_t}', I_2\right).$$

The feasible estimator of θ is denoted by $\tilde{\theta} \equiv \left(\widehat{\mathrm{B}}_0', \ldots, \widehat{\mathrm{B}}_h', \hat{\sigma}_x^2, \hat{\sigma}_{v,h}^2\right)'$ where $\hat{\sigma}_x^2 = \frac{1}{T}\sum_{t=1}^{T} x_t^2$, and $\hat{\sigma}_{v,h}^2 = \frac{1}{T_h}\sum_{t=L_{max+1}}^{T-h}\left(\hat{f}_{t+h|t-1} - \sum_{i=0}^{h} \iota_1' \widehat{\mathrm{B}}_i\, x_{t+h-i}\right)^2$. [5] We define $(H+1)*(I^{LP}+J^{LP}+2)+2$ dimensional vector Z_{t+h} as following:

$$Z_{t+h} \equiv \begin{pmatrix} q_t\left(p_{0,t} - q_t'\widehat{\mathrm{B}}_0\right) \\ \vdots \\ q_t\left(p_{h,t} - q_t'\widehat{\mathrm{B}}_h\right) \\ x_t^2 - \hat{\sigma}_x^2 \\ \left(\hat{f}_{t+h|t-1} - \sum_{i=0}^{h}\left(\iota_1'\widehat{\mathrm{B}}_i\right) x_{t+h-i}\right)^2 - \hat{\sigma}_{v,h}^2 \end{pmatrix}.$$

Then $\widehat{\Omega}_{h,LPB}$ is obtained by applying the Newey-West estimator to Z_{t+h} with pre-whitening similar to Proposition 1.

It remains to estimate $\Delta_{h,LPB}$. It is straightforward to define

$$\widehat{\Delta}_{h.LPB} = \frac{1-\hat{s}_h^{LPB}}{\hat{\sigma}_{f,h}^2}\begin{pmatrix} 2\hat{\psi}_{x,0}\hat{\sigma}_x^2\iota_1 \\ \vdots \\ 2\hat{\psi}_{x,h}\hat{\sigma}_x^2\iota_1 \\ \sum_{i=0}^{h}\hat{\psi}_{x,i}^2 \\ -\hat{s}_h^{LPB}/(1-\hat{s}_h^{LPB}) \end{pmatrix}'$$

[5] The denominator T_h ight be adjusted according to the degrees of freedom without affecting the asymptotics.

where $\hat{\sigma}^2_{f,h} = \frac{1}{T_h}\sum_{t=L_{max}+1}^{T-h} \hat{f}^2_{t+h|t-1}$. We plug the bias-corrected $\hat{s}_h^{LPB}$ in the place of s_h. How to obtain a bias-corrected estimator $\hat{s}_h^{LPB}$ in this set-up will be discussed later.

Given $\widehat{\Delta}_{h.LPB}$, the standard error of $\hat{s}_h^{LPB}$ is given as following:

$$[s.e.(\hat{s}_h^{LPB})]^2 = \frac{1}{T_h}\widehat{\Delta}_{h,LPB}\left(\hat{G}_{h,LPB}\right)^{-1}\widehat{\Omega}_{h,LPB}\left(\hat{G}'_{h,LPB}\right)^{-1}\widehat{\Delta}'_{h,LPB}.$$

Bias-correction. Similar to discussion regarding Proposition 1, we conjecture that most of the finite sample bias is due to the non-linear transformation $\xi(\cdot)$.

We approximate the asymptotic variance of the feasible estimator $\tilde{\theta}$ by $\frac{1}{T_h}\left(\hat{G}_{h,LPB}\right)^{-1}\widehat{\Omega}_{h,LPB}\left(\hat{G}'_{h,LPB}\right)^{-1}$. Then we simulate θ^b for B times from the following normal distribution:

$$\theta^b \sim \mathcal{N}\left(\tilde{\theta}, \quad \frac{1}{T_h}\left(\hat{G}_{h,LPB}\right)^{-1}\widehat{\Omega}_{h,LPB}\left(\hat{G}'_{h,LPB}\right)^{-1}\right).$$

We drop cases when simulated $\hat{\sigma}^2_x$ and $\hat{\sigma}^2_{v,h}$ are negative. The bias is estimated by

$$bias_h^{LP} \equiv \overline{\xi(\theta^b)} - \xi\left(\tilde{\theta}\right), \qquad \text{where} \qquad \overline{\xi(\theta^b)} \equiv \frac{1}{B}\sum_{b=1}^{B}\xi(\theta^b).$$

Finally, the bias-corrected estimator is given by

$$\hat{s}_h^{LP} = \xi\left(\tilde{\theta}\right) - bias_h^{LP} = 2\xi\left(\tilde{\theta}\right) - \overline{\xi(\theta^b)}.$$

Appendix C. Finding a MA(∞) representation for a process driven by multiple underlying shocks

Suppose the following data generating process as in Section II. In this section, we explain how an infinite-order MA representation driven by a single white noise process is obtained for the residual process $\Delta p_t + \Delta a_t$.

$$y_t = \psi_x(L)x_t + z_t \qquad \text{where} \qquad z_t = p_t + a_t,$$

$$\left(\Delta p_t - g_y\right) = \rho_p\left(\Delta p_{t-1} - g_y\right) + \sigma_p e_t^p, \qquad e_t^p \sim iid\ N(0,1),$$

$$a_t = \rho_a a_{t-1} + \sigma_a e_t^a, \qquad e_t^a \sim iid\ N(0,1),$$

$$x_t \sim iid\ N(0, \sigma_x^2), \qquad \text{and } \{x_t\}, \{e_t^p\} \text{ and } \{e_t^a\} \text{ are mutually independent.}$$

We first show why having a representation $g_y + \psi_e(L)e_t$ of $\Delta p_t + \Delta a_t$ is needed. When all three shocks are in the information set, then the corresponding forecast error with $\widetilde{\Omega}_t = \{x_t, \Delta y_t, e_t^p, e_t^a, \dots\}$ is

$$\tilde{f}_{t+h|t-1} = y_{t+h} - \mathrm{E}\left[y_{t+h}|\widetilde{\Omega}_{t-1}\right] = \psi_{x,0}x_{t+h} + \cdots + \psi_{x,h}x_t$$
$$+\sigma_p e_{t+h}^p + \left(1+\rho_p\right)\sigma_p e_{t+h-1}^p + \cdots + \left(1 + \rho_p + \cdots + \rho_p^h\right)\sigma_p e_t^p$$
$$+\sigma_a e_{t+h}^a + \rho_a\sigma_a e_{t+h-1}^a + \cdots + \rho_a^h \sigma_a e_t^a.$$

Then

$$\tilde{s}_h = \frac{\left(\sum_{i=0}^h \psi_{x,i}^2\right)\sigma_x^2}{\left(\sum_{i=0}^h \psi_{x,i}^2\right)\sigma_x^2 + \sum_{i=0}^h\left(\sum_{j=0}^i \rho_p^j\right)^2 \sigma_p^2 + \sum_{i=0}^h \rho_a^{2i}\sigma_a^2}.$$

However, what we estimate in the simulations using Δy_t and x_t is s_h, not $\tilde{s}_h$. It is because our information set is Ω_t, not the augmented one $\widetilde{\Omega}_t$. Because Ω_t is coarser than $\widetilde{\Omega}_t$, $s_h \leq \tilde{s}_h$. To obtain the true value of s_h, we need $\psi_e(L)$ and σ_e.

We use a stationary Kalman filter (Hamilton (1994), pp.391-394) to that end. We cast the above process in a form of state-space representation.

State equation:

$$s_t = Fs_{t-1} + Bu_t,$$

$$\text{where} \qquad s_t = (\Delta p_t - g_y, \Delta a_t, e_t^a)',$$

$$u_t = \left(e_t^p, e_t^a\right)' \sim (0, I),$$

$$\mathrm{F} = \begin{pmatrix} \rho_p & 0 & 0 \\ 0 & \rho_a & -\sigma_a \\ 0 & 0 & 0 \end{pmatrix},$$

$$B = \begin{pmatrix} \sigma_p & 0 \\ 0 & \sigma_a \\ 0 & 1 \end{pmatrix}.$$

Measurement equation: $\Delta z_t = g_y + H' s_t$ where $H = (1,1,0)'$.

By defining $Q = BIB' = BB'$ and $R = 0$, the stationary P and K are obtained from the matrix equation (13.5.3) and (13.5.4) on Hamilton (1994).

$$P = F[P - PH(H'PH + R)^{-1}H'P]F' + Q,$$

$$K = FPH(H'PH + R)^{-1}.$$

This is a discrete time algebraic Riccati equation for P which can be solved numerically. Then deriving K is straightforward from the second equation. Given K, it is known that

$$\Delta z_t = g_y + (I + H'(I - FL)^{-1}KL)e_t, \qquad e_t \sim WN(\sigma_e^2), \qquad \text{and} \qquad \sigma_e = \sqrt{H'PH + R}.$$

To convert $(I + H'(I - FL)^{-1}KL)$ into $\psi_e(L)$, we use the identity that $(I - FL)^{-1} = I + FL + F^2L^2 + \cdots$. Note all three eigenvalues of F, ρ_p, ρ_a and 0, are less than one in absolute values.

Given the MA representation of Δz_t, we can find s_h accordingly.

In Section III.B, the model of Smets and Wouters (2007) is analyzed. We find the s_h under the assumed information set in a similar way.

Appendix D. Unobservable Shocks and Measurement Errors

In some cases, an estimated structural shock is only a part of the true shock that can be identified with high confidences. For example, unexplained innovations in the Federal Funds Rates from Romer and Romer (2004) may be a part of the entire change in the monetary policy including changes in members on the board of governors, change in institutional details, or regime shifts as in the Volcker periods. Similarly, legislative tax changes identified from narratives by Romer and Romer (2010) would be understood as a part of the whole fiscal policy shocks affecting the U.S. economy. The measurement error is yet another potential issue. It is unavoidable in practical studies, especially when shocks are generated from narratives like Ramey (2011) and Romer and Romer (2010). In this section, we show that our approach can still provide interesting and meaningful quantities, because the estimates can be understood as a *conservative* estimate of the 'true' estimates available only when all hidden confounding factors are observable.

We decompose the true shock into two components $x_t = x_t^o + x_t^u$. The superscript o means observable, and u unobservable. We understand the vector process of two components has a representation as

$$\begin{pmatrix} x_t^o \\ x_t^u \end{pmatrix} = \begin{pmatrix} \sigma_o & 0 \\ \rho_{o,u}\sigma_u & \sqrt{1-\rho_{o,u}^2}\,\sigma_u \end{pmatrix} \delta_t,$$

where $\delta_t \sim wn(I_2),\ \delta_t \perp e_t,\ \sigma_o = \left(Var(x_t^o)\right)^{1/2},\ \sigma_u = \left(Var(x_t^u)\right)^{1/2}$, and $\rho_{o,u} = corr(x_t^o, x_t^u)$.

Three different situations are possible for the sign of correlation between two components, (1) $corr(x_t^o, x_t^u) = 0$: they are orthogonal, (2) $corr(x_t^o, x_t^u) > 0$: this might be the case when we are able to observe only some parts of the true shock, and (3) $corr(x_t^o, x_t^u) < 0$: the presence of measurement error m_t might impose such a correlation structure, since $x_t^o = x_t + m_t$, and $x_t^u = -m_t$. Narrative approaches might be exposed to such a concern.

Our claim is that the suggested estimators have a negative asymptotic bias, regardless of the sign of correlation between two components.[6] The population variance share can be written as a fraction of the amount explained by the innovations in $\{x_t\}$ to the variance of forecast error, $s_h =$

[6] All of our estimators have the same probability limit. So, the discussion in this section applies to R2, LP-A, and LP-B estimators.

$\frac{\sum_{i=0}^{h}\psi_{x,i}^2\sigma_x^2}{Var(f_{t+h|t-1})}$. We argue that there are (1) a positive asymptotic bias for the denominator, and (2) a negative asymptotic bias for the numerator when we apply our method to $\{x_t^o\}$ only while ignoring the existence of $\{x_t^u\}$. Therefore, the estimated share can be understood as a conservative estimate for the population quantity with the full information about $\{x_t^o\}$ and $\{x_t^u\}$.

Let's start with the denominator. The first problem we encounter is about recovering the forecast errors. It will be proven that the estimated forecast error without information on $\{x_t^u\}$ will have larger variances than the 'true' forecast error $f_{t+h|t-1}$. With the full information, we can back out the forecast error as a difference between $y_{t+h} - y_{t-1}$ and its projected values on the information set at time $t-1$ where $\Omega_{t-1} = \{\Delta y_{t-1}, x_{t-1}^o, x_{t-1}^u, \Delta y_{t-2}, x_{t-2}^o, x_{t-2}^u, \ldots\}$ as

$$f_{t+h|t-1} = y_{t+h} - y_{t-1} - E(y_{t+h} - y_{t-1}|\Omega_{t-1})$$

where the latter conditional expectation is the projection of $y_{t+h} - y_{t-1}$ on the closed subspace spanned by Ω_{t-1}. However, an econometrician has only $\Omega_{t-1}^e = \{x_{t-1}^o, \Delta y_{t-1}, x_{t-2}^o, \Delta y_{t-2}, \ldots\}$. It is evident that $\Omega_t^e \subset \Omega_t$. We define the closed subspaces spanned by variables in each information set as

$$V_t = closure(span(\Omega_t)),$$
$$V_t^e = closure\big(span(\Omega_t^e)\big).$$

We now compare the true forecast errors and identifiable ones to econometricians having Ω_{t-1}^e. By using a notation of $Projection(s|S)$ to denote the population least square projection of the element s on the closed subspace S, we can rewrite $f_{t+h|t-1}$ as

$$f_{t+h|t-1} = y_{t+h} - y_{t-1} - Projection(y_{t+h} - y_{t-1}|V_{t-1}) = Projection(y_{t+h} - y_{t-1}|(V_{t-1})^{\perp})$$

where $V_t^{\perp}$ is the orthogonal complement of V_t.

On the other hand, the econometrician's forecast error $f_{t+h|t-1}^e$ is given by

$$f_{t+h|t-1}^e = y_{t+h} - y_{t-1} - Projection(y_{t+h} - y_{t-1}|V_{t-1}^e) = f_{t+h|t-1} + r_{t+h|t-1}^e$$

where $r_{t+h|t-1}^e \equiv Projection(y_{t+h} - y_{t-1}|V_{t-1}) - Projection(y_{t+h} - y_{t-1}|V_{t-1}^e)$.

It is worth to mention that $f_{t+h|t-1}$ and $r_{t+h|t-1}^e$ are orthogonal, because the first term is an element of $(V_{t-1})^{\perp}$, and the second of V_{t-1}.[7] Therefore, $Var\big(f_{t+h|t-1}^e\big) = Var\big(f_{t+h|t-1}\big) +$

[7] This result is in fact due to a decomposition of the entire vector space, V, into a direct sum of three mutually orthogonal closed subspaces as $V = V_{t-1}^e \oplus (V_{t-1} \cap (V_{t-1}^e)^{\perp}) \oplus (V \cap (V_{t-1})^{\perp})$, where the symbol '$\oplus$' means a direct sum. From the representation, it directly follows that $y_{t+h} - y_{t-1} = Projection(y_{t+h} - y_{t-1}|V) = Projection(y_{t+h} - y_{t-1}|V_{t-1}^e) + Projection(y_{t+h} - y_{t-1}|V_{t-1} \cap (V_{t-1}^e)^{\perp}) +$

$Var(r^e_{t+h|t-1}) \geq Var(f_{t+h|t-1})$. Also, the equality holds only when x^u_t and its lagged values have no additional power in explaining y_{t+h} given V^e_{t-1} implying $Projection(y_{t+h} - y_{t-1}|V_{t-1}) = Projection(y_{t+h} - y_{t-1}|V^e_{t-1})$. This is not true except for some uninteresting special situations such as $\psi_x(L) = 0$ or $corr(x^o_t, x^u_t) = \pm 1$.

The second step is to show the econometrician's numerator converges to a values less than the true numerator in probability. The true numerator is $\sum_{i=0}^{h} \psi^2_{x,i}\, \sigma^2_x$ as before. Defining $X^h_t = (x_{t+h}, \dots, x_t)'$, we can write it as following:

$$E\left(f_{t+h|t-1} \cdot X^{h'}_t\right) E\left(X^h_t\, X^{h'}_t\right)^{-1} E\left(X^h_t \cdot f_{t+h|t-1}\right)$$

$$= \left[E\left(X^h_t X^{h'}_t\right)^{-1} E\left(X^h_t \cdot f_{t+h|t-1}\right)\right]' \left[E\left(X^h_t\, X^{h'}_t\right)\right] \left[E\left(X^h_t X^{h'}_t\right)^{-1} E\left(X^h_t \cdot f_{t+h|t-1}\right)\right].$$

The term inside the last square bracket is a vector of population regression coefficients of $f_{t+h|t-1}$ on X^h_t. By the specification, we know that it is equal to $\Psi^h = \left(\psi_{x,0}, \dots, \psi_{x,h}\right)'$.

Now we investigate the econometrician's numerator $E\left(f^e_{t+h|t-1} \cdot X^{h,e'}_t\right) E\left(X^{h,e}_t\, X^{h,e'}_t\right)^{-1} E\left(X^{h,e}_t \cdot f^e_{t+h|t-1}\right)$ where $X^{h,e}_t = (x^o_{t+h}, \dots, x^o_t)'$. Because $r^e_{t+h|t-1} = Projection(\, Projection(y_{t+h} - y_{t-1}|V_{t-1}) \mid (V^e_{t-1})^{\perp}\,) \in (V^e_{t-1})^{\perp}$, $E\left(X^{h,e}_t \cdot f^e_{t+t|t-1}\right) = E\left(X^{h,e}_t \cdot f_{t+h|t-1}\right)$. The corresponding projection coefficient is

$$E\left(X^{h,e}_t\, X^{h,e'}_t\right)^{-1} E\left(X^{h,e}_t \cdot f_{t+h|t-1}\right) = \left(1 + \frac{Cov(x^o_t, x^u_t)}{Var(x^o_t)}\right)\Psi^h = \frac{\left(\sigma_o + \rho_{o,u} \cdot \sigma_u\right)}{\sigma_o}\Psi^h.$$

We used the fact that $f_{t+h|t-1} = \sum_{i=0}^{h} \psi_{x,i} x_{t+h-i} + \sum_{i=0}^{h} \sum_{j=0}^{i} (\psi_{e.j})\; e_{t+h-i}$.

Finally, the econometrician's numerator becomes

$$\left[E\left(X^{h,e}_t\, X^{h,e'}_t\right)^{-1} E\left(X^{h,e}_t \cdot f_{t+h|t-1}\right)\right]' \left[E\left(X^{h,e}_t\, X^{h,e'}_t\right)\right] \left[E\left(X^{h,e}_t\, X^{h,e'}_t\right)^{-1} E\left(X^{h,e}_t \cdot f_{t+h|t-1}\right)\right]$$

$$= \frac{\left(\sigma_o + \rho_{o,u} \cdot \sigma_u\right)}{\sigma_o}\Psi^{h'} \cdot\; \sigma^2_o I \;\cdot \frac{\left(\sigma_o + \rho_{o,u} \cdot \sigma_u\right)}{\sigma_o}\Psi^h = \sum_{i=0}^{h} \psi^2_{x,i}\left(\sigma_o + \rho_{o,u} \cdot \sigma_u\right)^2.$$

Thus, any asymptotic bias in the numerators are from the differences between σ^2_x and $\left(\sigma_o + \rho_{o,u} \cdot \sigma_u\right)^2$. Because $\sigma^2_x - \left(\sigma_o + \rho_{o,u} \cdot \sigma_u\right)^2 = \left(1 - \rho^2_{o,u}\right)\sigma^2_u$,

$Projection(y_{t+h} - y_{t-1}|V \cap (V_{t-1})^{\perp}) = Projection(y_{t+h} - y_{t-1}|V^e_{t-1}) + r^e_{t+h|t-1} + f_{t+h|t-1}$, and the last three terms are mutually orthogonal.

$$E\left(f_{t+h|t-1} \cdot X_t^{h\prime}\right) E\left(X_t^h X_t^{h\prime}\right)^{-1} E\left(X_t^h \cdot f_{t+h|t-1}\right)$$
$$= E\left(f_{t+h|t-1}^e \cdot X_t^{h,e\prime}\right) E\left(X_t^{h,e} X_t^{h,e\prime}\right)^{-1} E\left(X_t^{h,e} \cdot f_{t+h|t-1}^e\right) + \sum_{i=0}^{h} \psi_{x,i}^2 \left(1-\rho_{o,u}^2\right)\sigma_u^2.$$

As claimed, the econometrician's numerator is asymptotically less and denominator is asymptotically greater than their full information counterparts. Thus, we have a negative asymptotic bias. So, we can understand our method as a conservative estimator for the true s_h. Moreover, the size of bias becomes small when the observed and unobserved parts are highly correlated, or variance of the unobserved parts is small. In such a case, both biases for the denominator $Var\left(r_{t+h|t-1}^e\right)$, and the numerator $\sum_{i=0}^{h} \psi_{x,i}^2 \left(1-\rho_{o,u}^2\right)\sigma_u^2$ are small.[8]

[8] In the above, we assume that x_t^o and x_t^u have the sample impulse response polynomial $\psi_x(L)$ for simplicity. Instead, we may consider $\psi_x^o(L)x_t^o + \psi_x^u(L)x_t^u$. This does not change our result, and the above derivations admit a straight-forward extension to the general case. In such a case, the difference between two numerators becomes $\sum_{i=0}^{h}\left(\psi_{x,i}^u\right)^2 \left(1-\rho_{o,u}^2\right)\sigma_u^2$. Therefore, we can conclude that if the unobservable component has a less contribution to the endogenous variable than the observable component, the bias in the numerator would be small.

Appendix E. Supplementary Figures for Univariate Simulations

Details on VAR-based Bootstrap

We first choose L_{VAR} using the HQIC. VAR models for $(x_t, \Delta y_t)'$ are estimated for different lag lengths between 1 and 10. The information criterion is given by

$$\log(det(V)) + \frac{2k \log(\log(T))}{T}$$

where V is the estimated variance matrix of the reduced-form residual process, k is the number of parameters, in this bi-variate case, 4 times lag length, and T is the sample size. For a fair comparison, we adjust the initial observation across L_{VAR} and make the effective sample sizes same. Once it is selected as a minimizer of the HQIC, both L_x and L_y are set to L_{VAR}.

We use the estimated $VAR(L_{VAR})$ model to bootstrap. First of all, we randomly choose t between 1 and $T - L_{VAR}$. Then $(x_t, \Delta y_t)', \ldots, \left(x_{t+L_{VAR}}, \Delta y_{t+L_{VAR}}\right)'$ are used as an initial condition when simulating the bootstrapped time series.

Second, we randomly draw the reduced form residuals with replacement. Using the estimated model with the above initial conditions and shuffled residuals, artificial data points are generated. The first T_{BunIn} number of observations are discarded as burn-in. $T_{BurnIn} = 100$ in all simulations.

We apply our estimators to the bootstrapped time series obtaining $\hat{s}_h^{R2,b}$, $\hat{s}_h^{LPA,b}$, and $\hat{s}_h^{LPB,b}$ for $b = 1, \ldots, 2000$. Given $\hat{s}_h^{VAR}$ obtained from the estimated $VAR(L_{VAR})$ model, the biases for local projection-based estimators are obtained by

$$\frac{1}{B}\sum_{b=1}^{B} \hat{s}_h^{m,b} - \hat{s}_h^{VAR}$$

for $m = R2, LPA,$ and LPB.

For other details regarding simulations, see Section III.

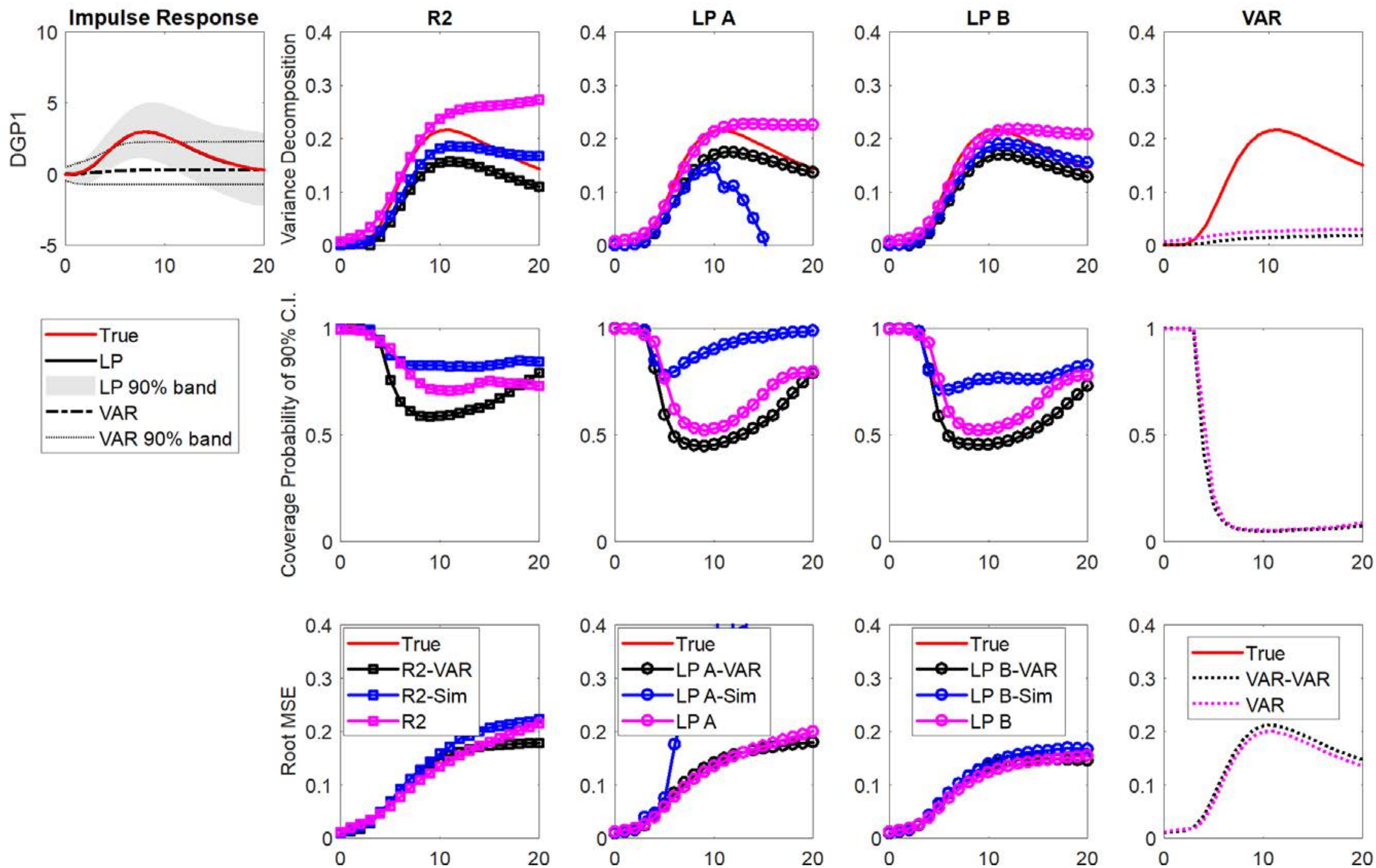
How to read the legend:

1. Impulse response
   - The 90% bands are based on 5% and 95% quantiles of estimates across 2,000 replications.

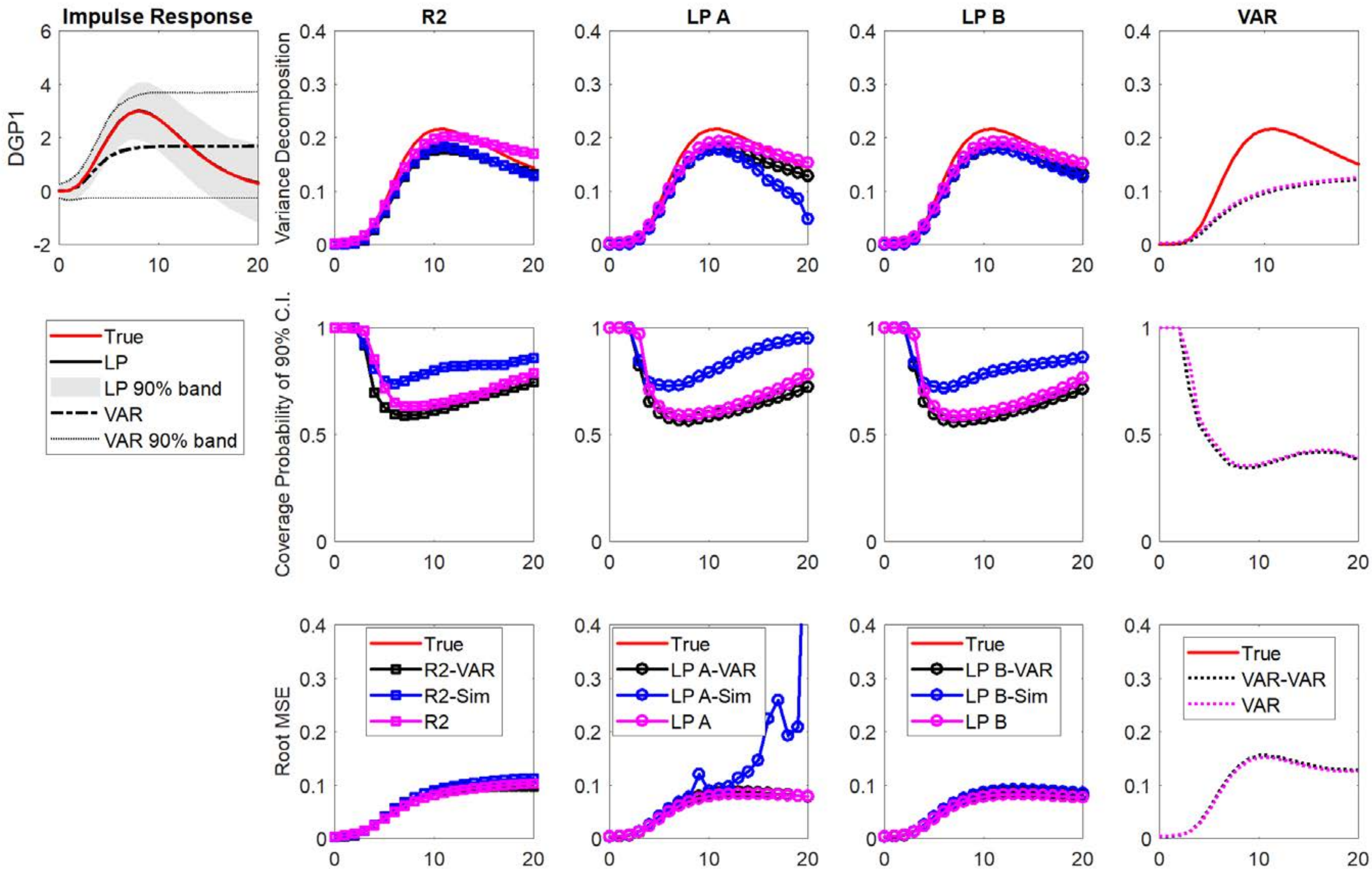
2. Variance decomposition, Coverage probability, and Root MSE

- ‘R2-VAR’ means the bias-corrected R2 estimator by bootstrapping an estimated VAR model. Its standard error is the standard deviation across bootstrap estimates.
- ‘R2-Sim’ uses the method discussed in Appendix B. The coverage probability is based on the asymptotic standard error with pre-whitening as discussed in Appendix B.
- ‘R2’ denotes for the estimator without any finite sample correction. It uses the same standard error as ‘R2-VAR.’
- ‘LP A/B-VAR’, ‘LP A/B-Sim’, ‘LP A/B’, ‘VAR-VAR’, and ‘VAR’ are defined similarly.

DGP 1, T = 160.

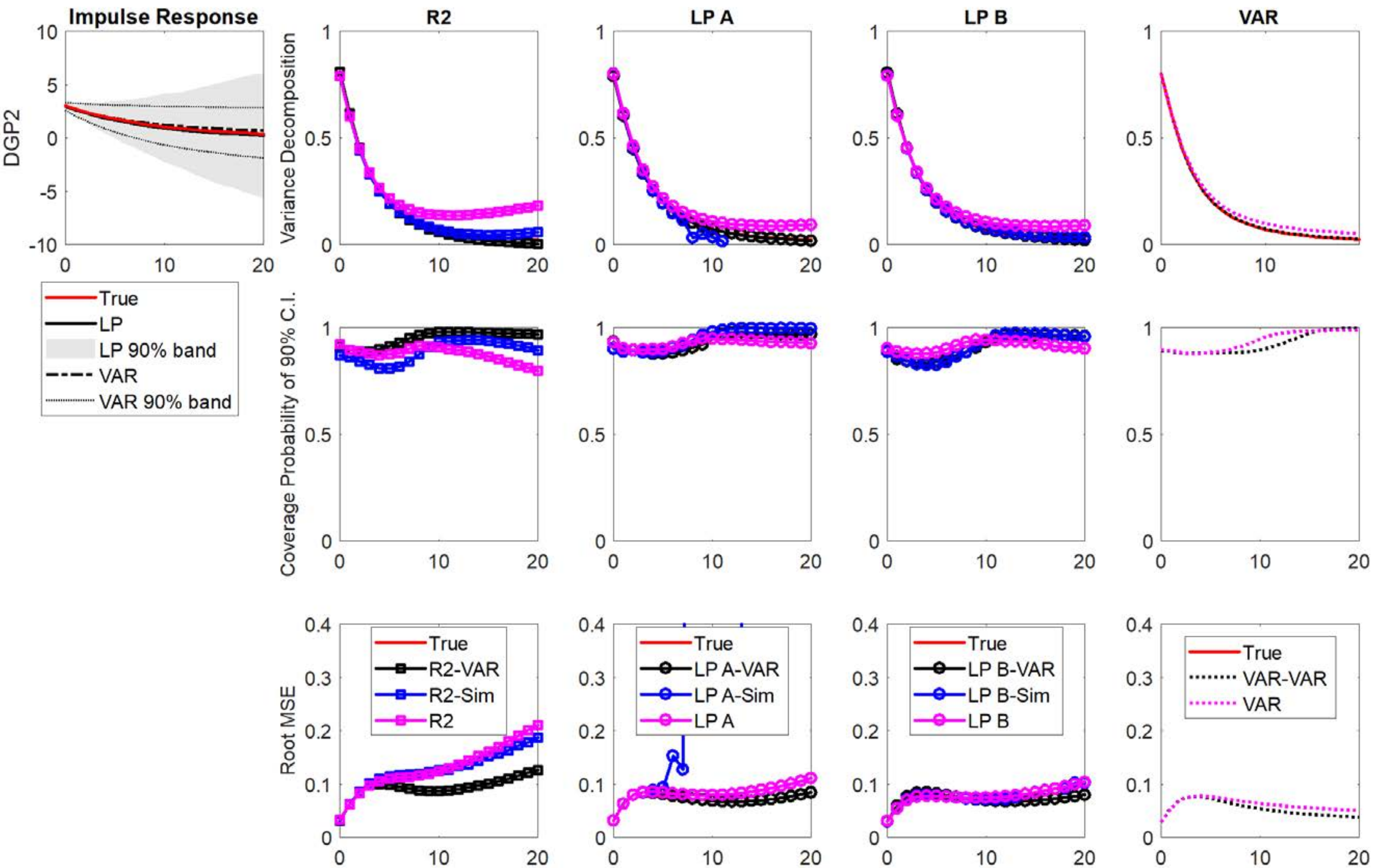

Impulse Response
R2
LP A
LP B
VAR
DGP1
Variance Decomposition
Coverage Probability of 90% C.I.
Root MSE
True
LP
LP 90% band
VAR
VAR 90% band
True
R2-VAR
R2-Sim
R2
True
LP A-VAR
LP A-Sim
LP A
True
LP B-VAR
LP B-Sim
LP B
True
VAR-VAR
VAR

DGP1, T = 500.



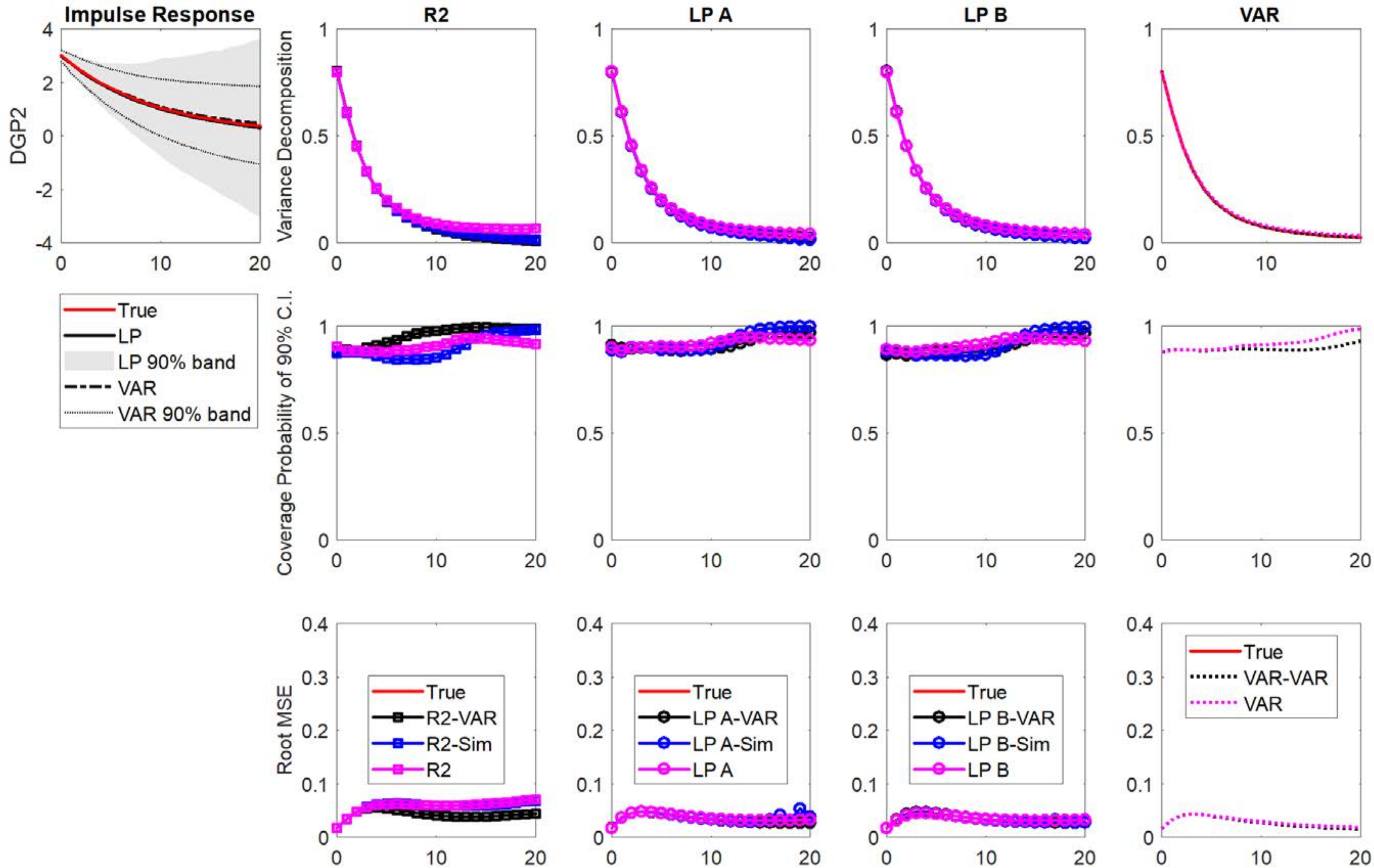
Impulse Response
DGP1
R2
LP A
LP B
VAR
Variance Decomposition
Coverage Probability of 90% C.I.
Root MSE
True
LP
LP 90% band
VAR
VAR 90% band
True
R2-VAR
R2-Sim
R2
True
LP A-VAR
LP A-Sim
LP A
True
LP B-VAR
LP B-Sim
LP B
True
VAR-VAR
VAR

DGP2, T = 160.



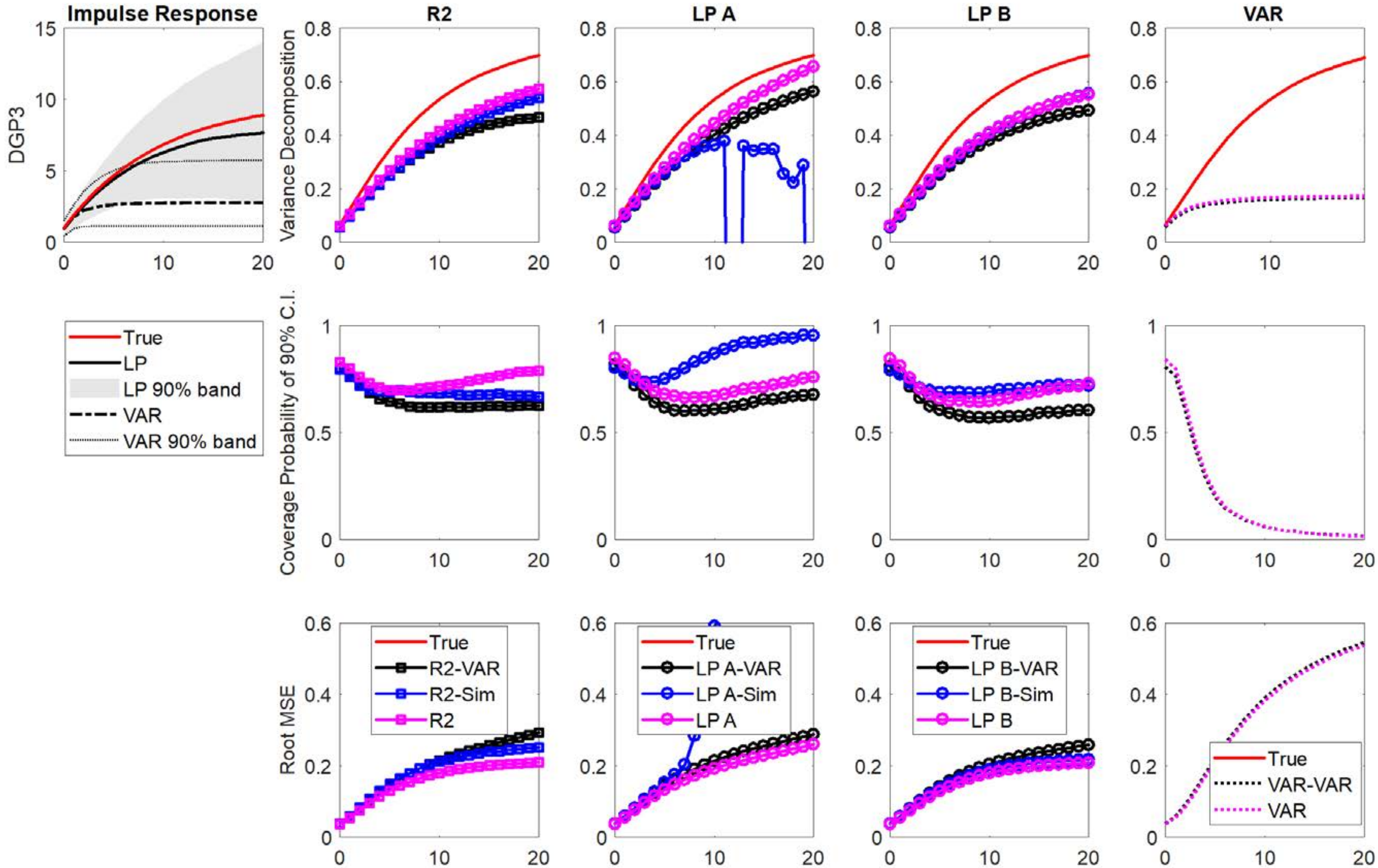
Impulse Response
R2
LP A
LP B
VAR
DGP2
Variance Decomposition
Coverage Probability of 90% C.I.
Root MSE
True
LP
LP 90% band
VAR
VAR 90% band
True
R2-VAR
R2-Sim
R2
True
LP A-VAR
LP A-Sim
LP A
True
LP B-VAR
LP B-Sim
LP B
True
VAR-VAR
VAR

DGP2, T= 500.

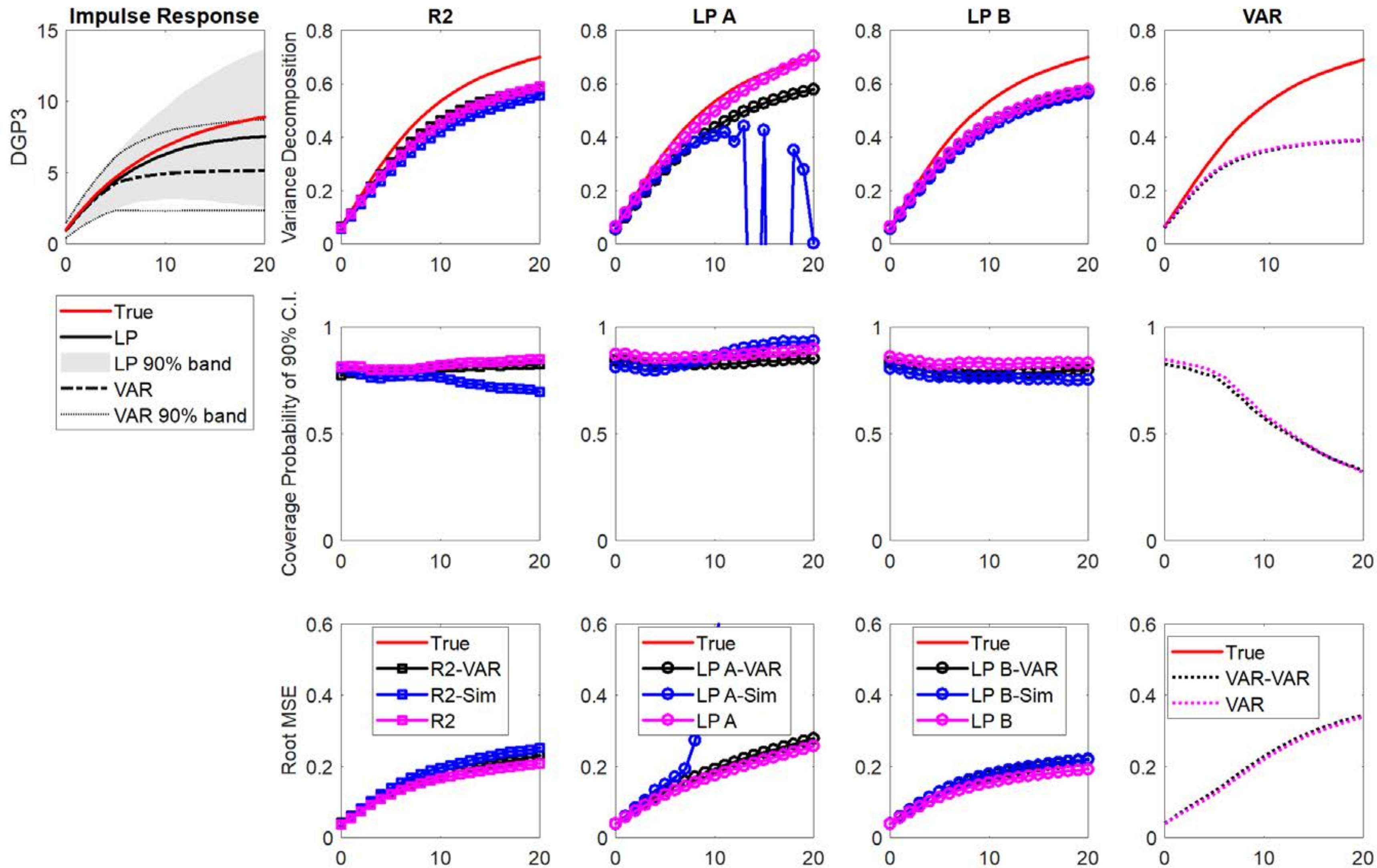


Impulse Response
R2
LP A
LP B
VAR
DGP2
Variance Decomposition
Coverage Probability of 90% C.I.
Root MSE
True
LP
LP 90% band
VAR
VAR 90% band
R2-VAR
R2-Sim
R2
LP A-VAR
LP A-Sim
LP A
LP B-VAR
LP B-Sim
LP B
VAR-VAR

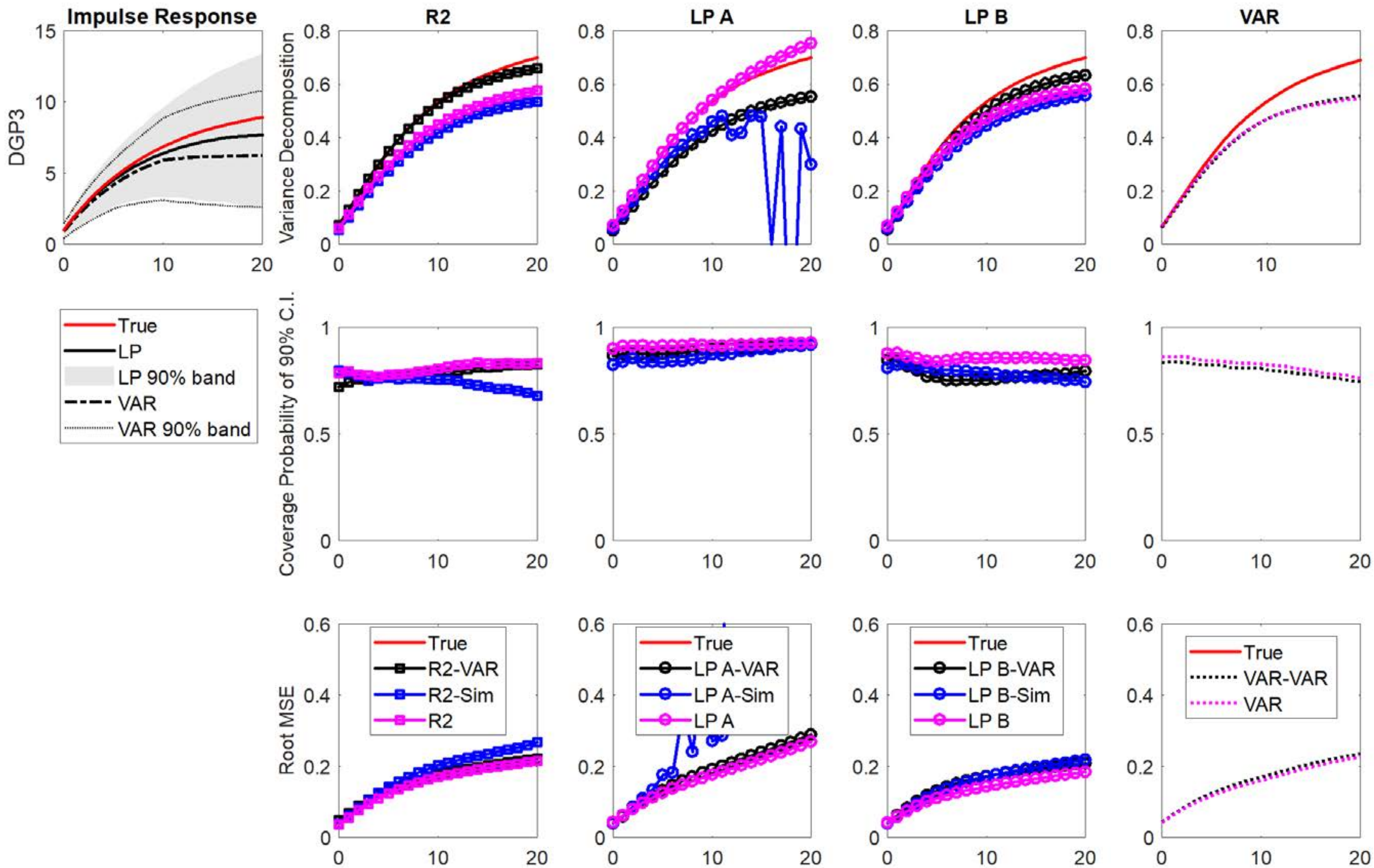
DGP3, T = 160, VAR(HQIC).



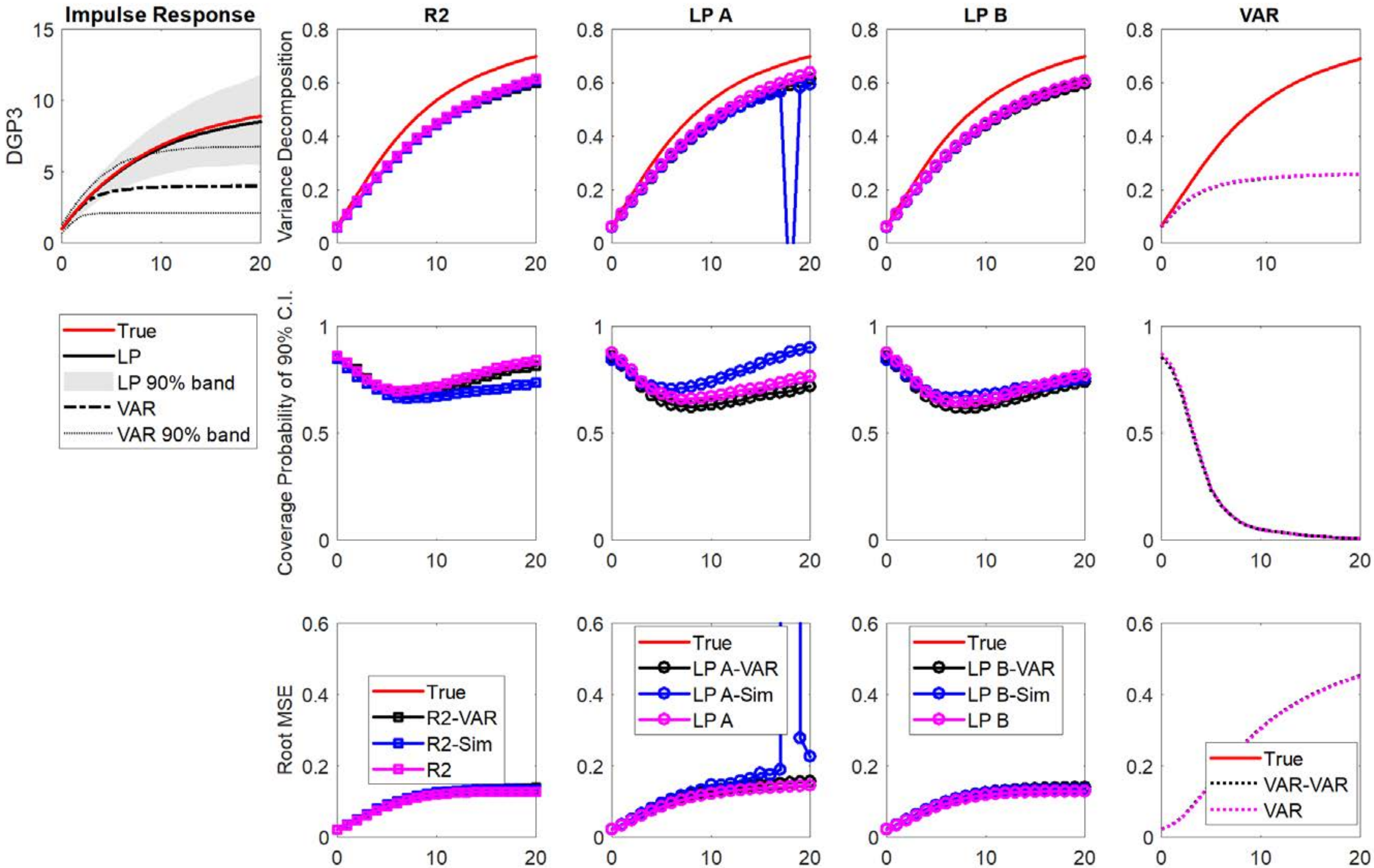
DGP3, T = 160, VAR(5).



DGP3, T = 160, VAR(10).



DGP3, T = 500, VAR(HQIC).



Appendix F. Supplementary Figures for Multivariate Simulations

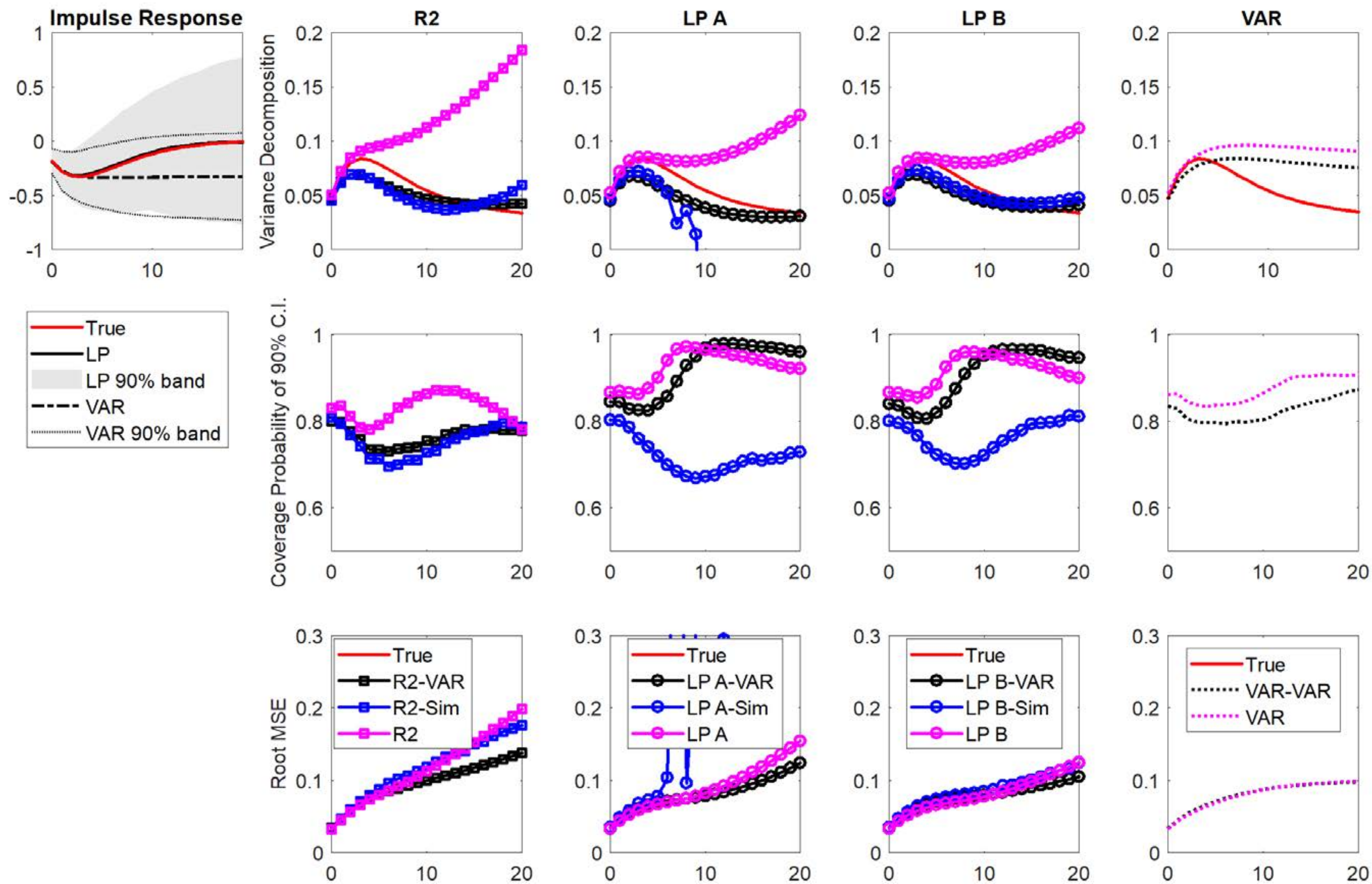
For details regarding simulations, see Section III.

How to read the legend:

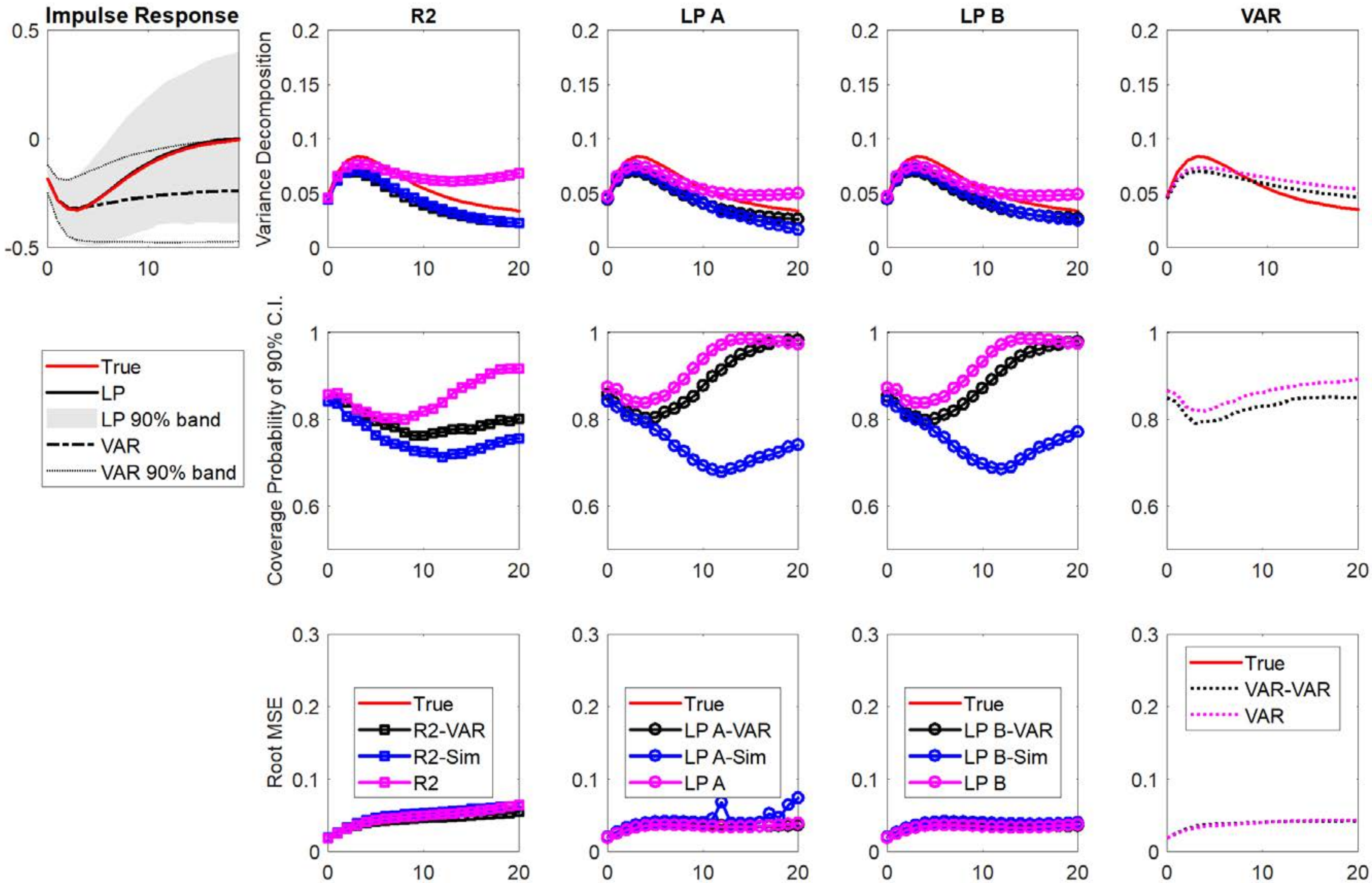
1. Impulse response
   - The 90% bands are based on 5% and 95% quantiles of estimates across 2,000 replications.

2. Variance decomposition, Coverage probability, and Root MSE
   - 'R2-VAR' means the bias-corrected R2 estimator by bootstrapping an estimated VAR model. Its standard error is the standard deviation across bootstrap estimates.
   - 'R2-Sim' uses the method discussed in Appendix B. The coverage probability is based on the asymptotic standard error with pre-whitening as discussed in Appendix B.
   - 'R2' denotes for the estimator without any finite sample correction. It uses the same standard error as 'R2-VAR.'
   - 'LP A/B-VAR', 'LP A/B-Sim', 'LP A/B', 'VAR-VAR', and 'VAR' are defined similarly.

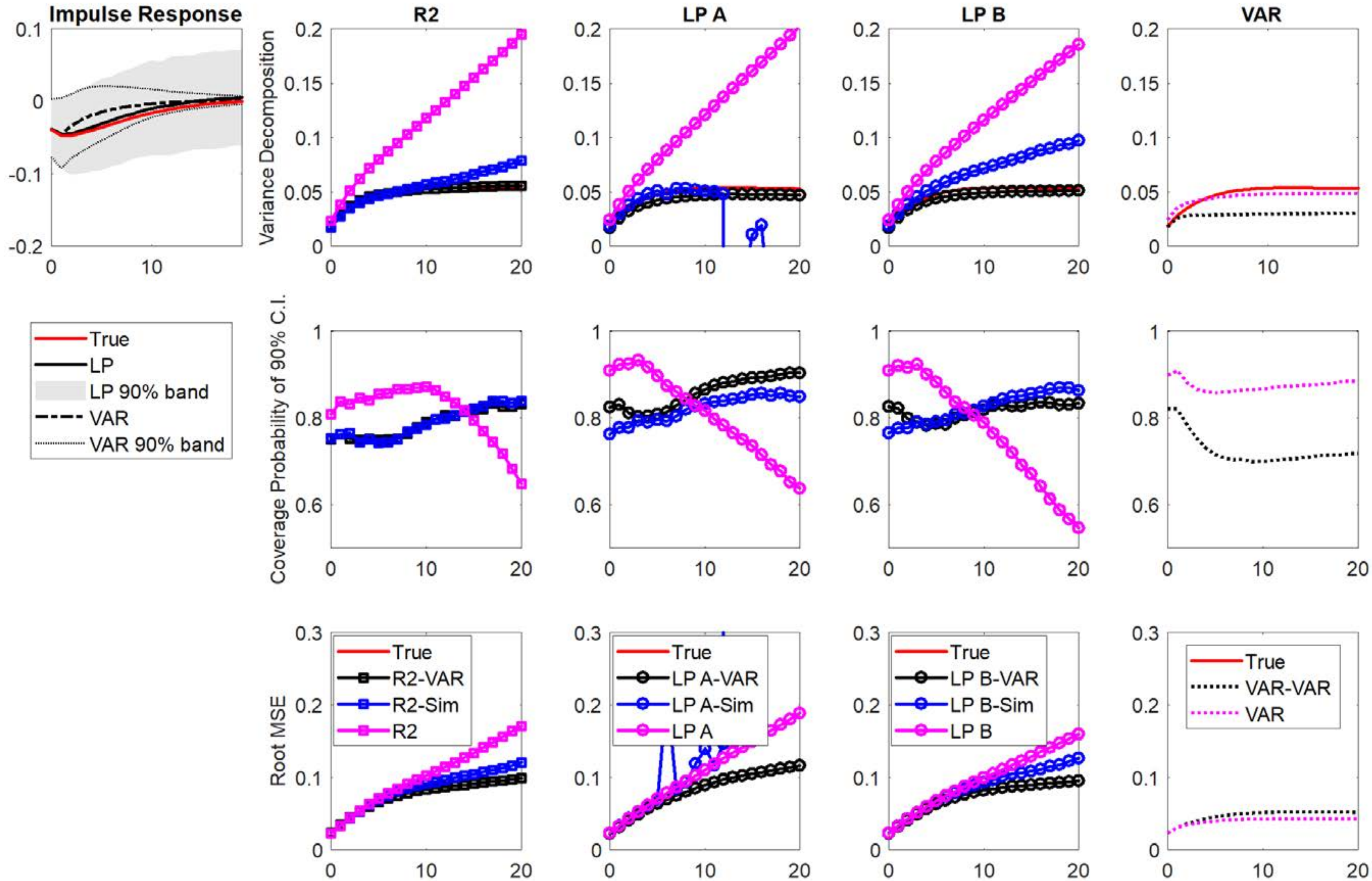
Real GDP and monetary policy shock, T = 160



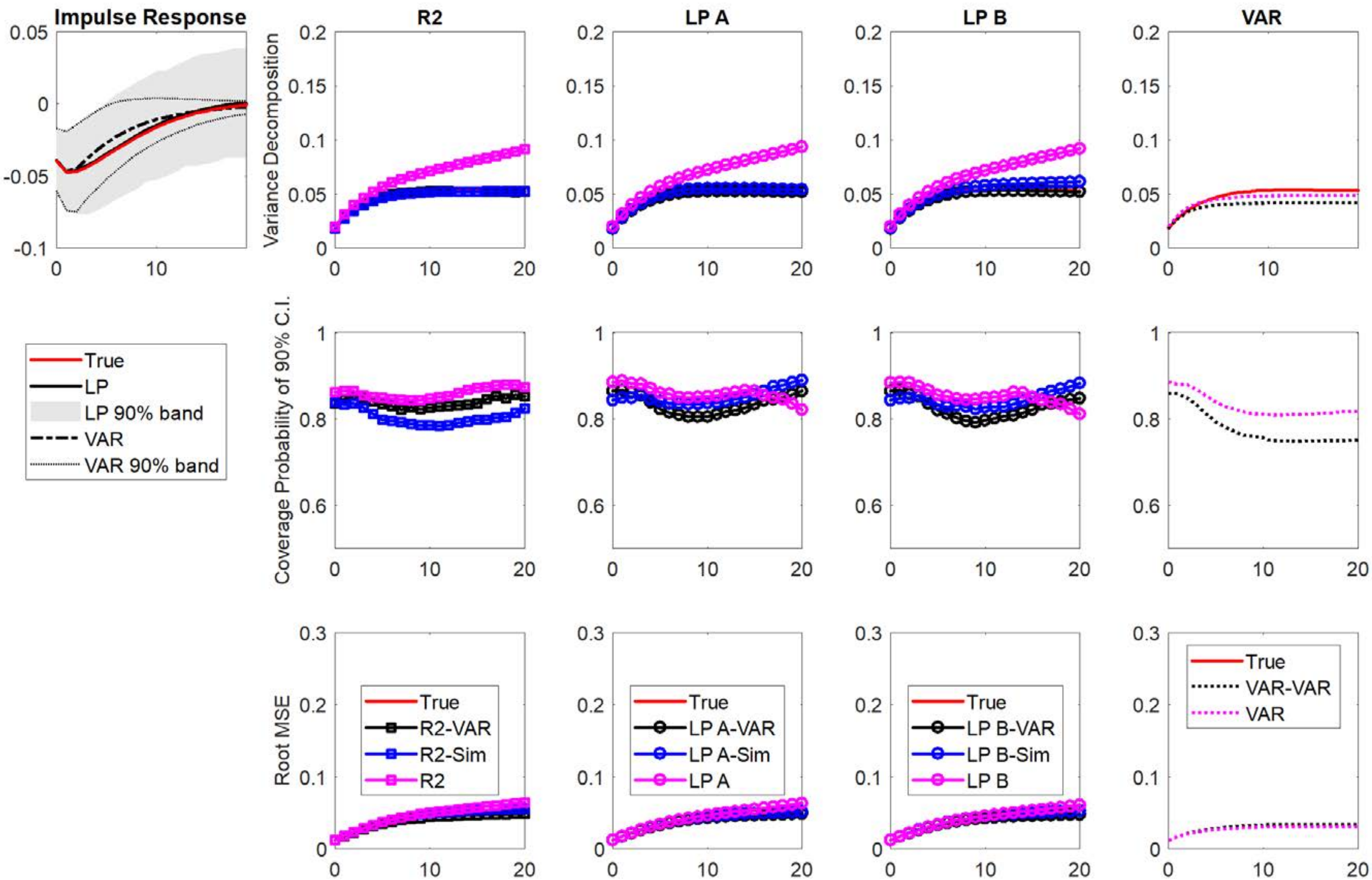
Real GDP and monetary policy shock, T = 500.



Price inflation and monetary policy shock, T = 160



Price inflation and monetary policy shock, T = 500.



Appendix G. Applications to Real GDP and Inflation

For the LP method, we use different local projections for bias-correction depending on whether the estimated VAR or simulations are used. For example, suppose that x_t is the monetary policy shock, and y_t is the real GDP. When we bootstrap the estimated VAR, the following local projection is estimated to have bootstrap impulse responses.

$$y_{t+h}^{\mathrm{b}} - y_{t-1}^{\mathrm{b}} = constant + \sum_{i=0}^{3} \beta_i^h \, x_{t-i}^{\mathrm{b}} + C_{0,1} \cdot TFP_t^b + C_{0,2} \cdot output\ growth_t^b$$
$$+ C_{0,3} \cdot inflation_t^{\mathrm{b}} + \sum_{i=1}^{3} \left(C_{t-i}^b\right)' \Gamma_i^h + r_{t+h|t-1}^b,$$

where C_t includes $TFP_t, output\ growth_t, inflation_t$ and $federal\ funds\ rate_t$. It corresponds to VAR(4) with the vector of TFP, output growth, inflation, monetary policy shock, and federal funds rate.

On the other hand, we cannot include many variables when we need asymptotic (joint) variance estimated. In this case, the following regression is estimated.

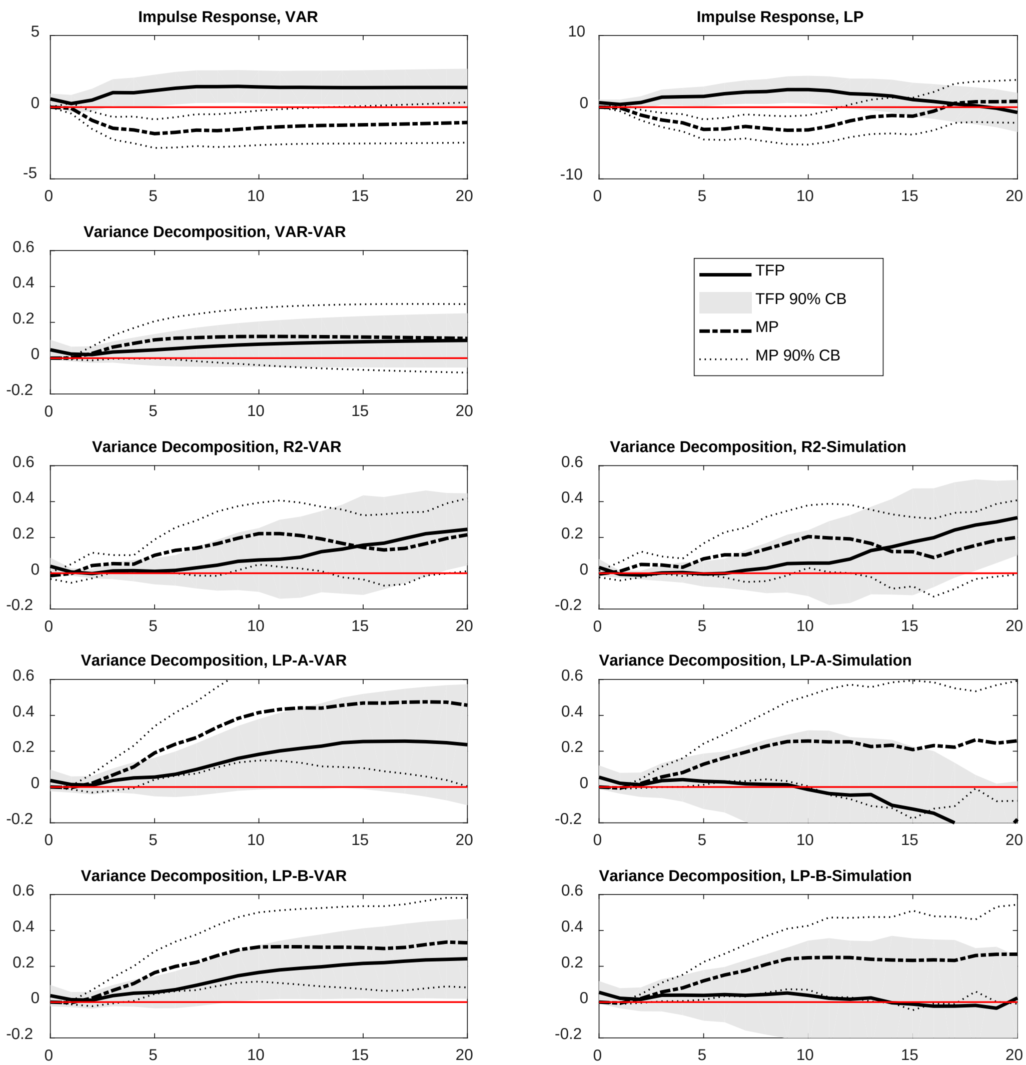
$$y_{t+h}^{\mathrm{b}} - y_{t-1}^{\mathrm{b}} = constant + \beta_0^h x_t^{\mathrm{b}} + C_{0,1} \cdot TFP_t^b + C_{0,2} \cdot output\ growth_t^b + C_{0,3} \cdot inflation_t^{\mathrm{b}}$$
$$+ r_{t+h|t-1}^b.$$

In a similar logic, it corresponds to VAR(1) with the same ordering. Therefore, we preserve the assumed ordering of variables in both cases.

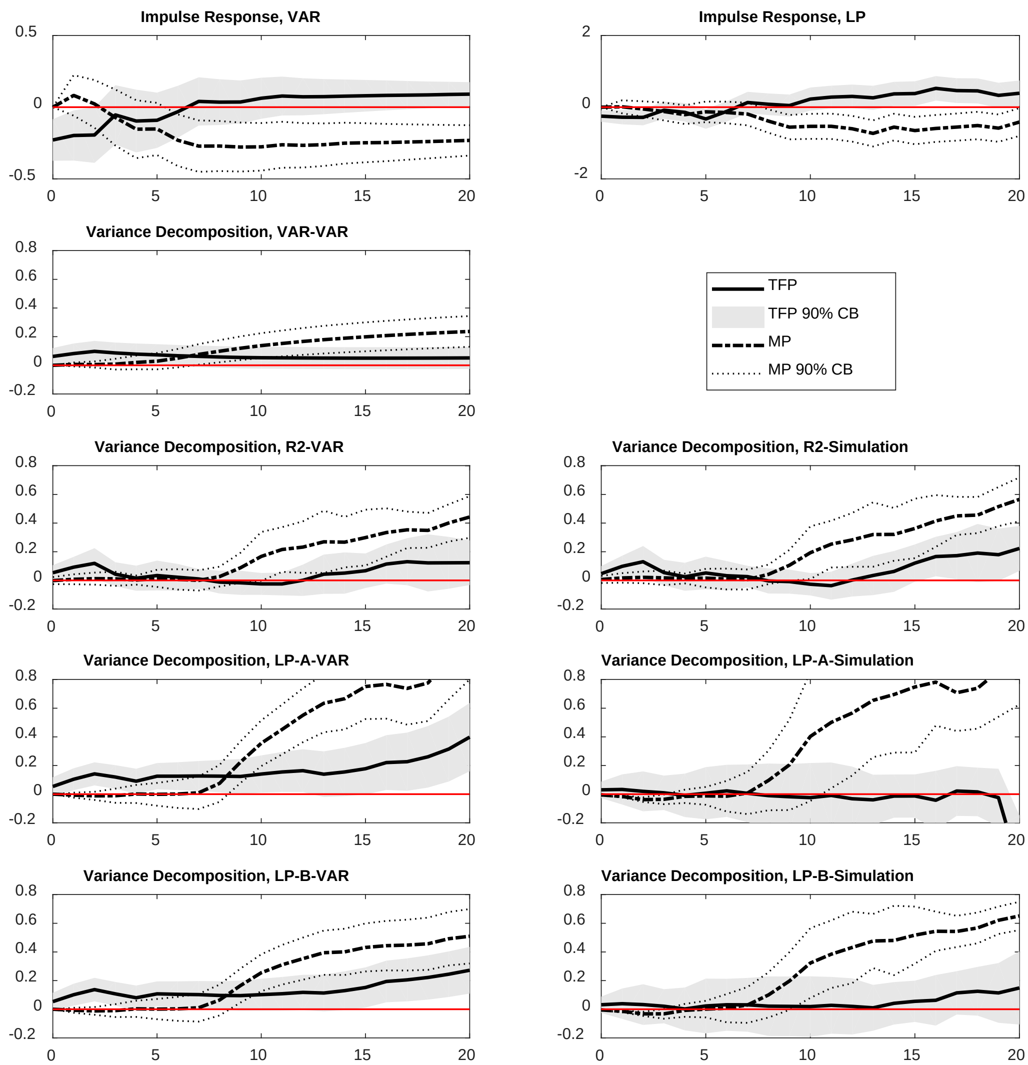
When x_t is TFP_t, x_t is the only time t variables on the right-hand side making it consistent to the ordering.

Figures below show the results for simulation based bias-corrections. Results are similar to VAR-based ones.

1969:Q1-2007:Q4. Real GDP.



1969:Q1-2007:Q4, Inflation.



References

Andrews, Donald W.K., and J. Christopher Monahan, 1992. "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator." *Econometrica*, 60(4): 953-966.

Brockwell, Peter J., and Richard A. Davis, 1991, *Time series: theory and methods*. Springer Science & Business Media.

Conway, John B., 1990. *A course in functional analysis*. Vol. 96. Springer Science & Business Media.

Hamilton, James D., 1994. *Time series analysis*. Princeton: Princeton university press.

Hansen, Lars Peter, 1982. "Large sample properties of generalized method of moments estimators." *Econometrica*, 50(4): 1029-1054.

Jordà, Oscar, 2005. "Estimation and Inference of Impulse Responses by Local Projections," A*merican Economic Review*, 95(1): 161–182.

Ramey, Valerie A., 2011. "Identifying Government Spending Shocks: It's all in the Timing," *Quarterly Journal of Economics*, 126(1): 1–50.

Romer, Christina, D., and David H. Romer, 2004. "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94(4): 1055-1084.

Romer, Christina D., and David H. Romer, 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review*, 100(3): 763-801.

Stock, James H., and Mark W. Watson, 2011. *Introduction to Econometrics*, 3/E. Boston: Addison Wesley.

Whitney K. Newey and Kenneth D. West, 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55(3): 703-708.