



Federal Reserve  
Bank of Dallas

# Estimating Impulse Response Functions When the Shock Series Is Observed

---

Chi-Young Choi and Alexander Chudik

## Globalization Institute Working Paper 353

Research Department

<https://doi.org/10.24149/gwp353>

Working papers from the Federal Reserve Bank of Dallas are preliminary drafts circulated for professional comment. The views in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

# Estimating Impulse Response Functions When the Shock Series Is Observed\*

Chi-Young Choi<sup>†</sup> and Alexander Chudik<sup>‡</sup>

March 4, 2019

## Abstract

We compare the finite sample performance of a variety of consistent approaches to estimating Impulse Response Functions (IRFs) in a linear setup when the shock of interest is observed. Although there is no uniformly superior approach, iterated approaches turn out to perform well in terms of root mean-squared error (RMSE) in diverse environments and sample sizes. For smaller sample sizes, parsimonious specifications are preferred over full specifications with all ‘relevant’ variables.

**Keywords:** Observed shock, Impulse-response functions, Monte Carlo experiments, Finite sample performance

**JEL Classification:** C13, C50

---

\* We would like to thank Everett Grant, Oscar Jorda, Lutz Kilian, and Karel Mertens for helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas or the Federal Reserve System.

<sup>†</sup>Chi-Young Choi, Department of Economics, University of Texas at Arlington, 701 S. West Street, Arlington, Texas, [cy-choi@uta.edu](mailto:cy-choi@uta.edu).

<sup>‡</sup>Alexander Chudik, Federal Reserve Bank of Dallas, 2200 N. Pearl Street, Dallas, Texas 75201, [alexander.chudik@dal.frb.org](mailto:alexander.chudik@dal.frb.org).

# 1 Introduction

An important task in empirical economics is to track the effects of a shock on the variable(s) of interest (e.g., Kilian and Lütkepohl, 2017). Since shocks are rarely observed, much of the literature has been devoted to the issue of estimation and identification of shocks, usually within a vector autoregression (VAR) framework. If a shock can be measured, or is observed without error, however, then estimating IRFs is straightforward as noted by Kilian and Lütkepohl (2017, Chapter 7) and Stock and Watson (2017, p.63). In that case, many approaches are available to estimating IRFs, such as univariate distributed lag (DL) or iterated autoregressive distributed lag (ARDL) approaches, the local projection (LP) approach of Jordà (2005), and the asymptotically efficient full-system VARX. This paper focuses on the following three questions within the standard linear setup: (1) Do we need to care about which approach to use in practice?; (2) Is it desirable to include all the relevant variables<sup>1</sup> in estimating IRFs?; and (3) Is there any approach that always dominates the others? Our answers, respectively, are: yes, not really, and no. We arrive at these answers by investigating the root mean-squared error (RMSE) of estimating individual IRF coefficients in a series of Monte Carlo (MC) experiments.<sup>2</sup> We adopt a design with randomly generated coefficients as well as two additional designs calibrated to a quarterly international real output growth dataset and a U.S. macro dataset (relegated to the online Appendix).

We find a significant difference in the performance among various approaches under study, indicating that the choice of approach in empirical application can be consequential. Although there exists no uniformly superior approach, the iterated ARDL and vector ARDL (VARDL) approaches turn out to perform better in the simulation environments considered in our study. Moreover, parsimonious specifications outperform asymptotically efficient approaches when the time dimension is relatively small. These findings imply that one may not wish to include all ‘relevant’ variables for short samples even when the variables entering the true underlying model are identified.

---

<sup>1</sup>The term ‘relevant’ is used to denote being part of the true data generating process (DGP).

<sup>2</sup>While this paper focuses only on the RMSE of estimating individual IRF coefficients, other metrics are also of interest although they are not studied here. One important example is estimation of variance decompositions, which is considered in Gorodnichenko and Lee (2017). Besides the precision of estimates, another important aspect is inference. Small sample evidence on accuracy of inference of the (full-system) VAR and LP approaches is provided by Kilian and Kim (2011) in a setup where shock is identified by recursive ordering.

## 2 Approaches for Estimating IRFs When the Shock is Observed

Suppose that  $\mathbf{z}_t$  is an  $n \times 1$  vector of variables generated from a stationary VAR model.<sup>3</sup> A shock series of interest, denoted as  $\varepsilon_t$ , is observed, and the impact variable of interest (one element of  $\mathbf{z}_t$ ) is also observed and denoted by  $x_t$ . The remaining variables in  $\mathbf{z}_t$  are not necessarily observed. We refer the reader to the online Appendix for our linear VAR setup, its (standard) assumptions and further discussions. Our objective is to estimate the IRF coefficients,  $b_h = E(x_{t+h} | \varepsilon_t = 1, \mathcal{I}_{t-1}) - E(x_{t+h} | \mathcal{I}_{t-1})$ , where  $\mathcal{I}_{t-1} = \{\mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \dots\}$ .

We group individual approaches into ‘direct’ and ‘iterated,’ depending on whether the IRF coefficients are directly estimated from auxiliary regressions, or computed from estimated parameters in an iterative way. We abstract from deterministic terms, but they must be included in the regressions below when they are present.

### 2.1 Direct approaches

**The most parsimonious approach.** The simplest and most parsimonious approach is to regress  $x_t$  on  $\varepsilon_{t-h}$  (for given  $h \geq 0$ ) to obtain a consistent estimate of  $b_h$ .

**DL approach.** The DL approach is based on regressing  $x_t$  on the distributed lags of  $\varepsilon_t$ , namely

$$x_t = \sum_{h=0}^{h_{\max}} b_h \varepsilon_{t-h} + v_t, \quad (1)$$

where  $v_t$  is the regression error term, and  $h_{\max}$  is the maximum horizon considered. This approach has been employed by Kilian (2008a, 2009), Kimball et al. (2006), Romer and Romer (2010), and Baumeister and Kilian (2014), among others.

**Augmented DL (ADL) approach.** Augmenting the DL regression in (1) by any variables that are uncorrelated with the set of regressors,  $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-h_{\max}}$ , will also lead to a consistent estimator, possibly with improved small sample performance if the augmenting variables contribute meaningfully to the fit of the regression. Let  $\mathbf{y}_t$  be an  $s \times 1$  vector of covariates that are assumed to be uncorrelated with *future* values of  $\varepsilon_t$ , and also are considered to be good candidates for contributing to the fit.  $\mathbf{y}_t$  can include  $x_t$  and any of its lags, and any other variables that are not

---

<sup>3</sup>The most parsimonious and the DL approaches described below are not applicable to the case where the impact variable is subject to unit roots.

necessarily part of the true DGP. Then, the ADL regression is

$$x_t = \sum_{h=0}^{h_{\max}} b_h \varepsilon_{t-h} + \beta' \mathbf{y}_{t-h_{\max}-1} + v_t, \quad (2)$$

where the variance of the error term  $v_t$  will in general be smaller than in regression (1) when  $\beta \neq 0$ .

**Local projection (LP) approaches.** The term local projection, introduced by Jordà (2005), has been used in the literature to denote any regressions of the following form

$$x_{t+h} = \sum_{\ell=0}^p b_{h,\ell} \varepsilon_{t-\ell} + \sum_{\ell=1}^q \beta'_{h,\ell} \mathbf{y}_{t-\ell} + v_{h,t+h}, \quad (3)$$

where  $\mathbf{y}_t$  denotes an  $s \times 1$  vector of regressors, which are uncorrelated with the future values of  $\varepsilon_t$ . Regressions in (3) are estimated separately for  $h = 0, 1, \dots, h_{\max}$ . This approach is used by Guajardo et al. (2014), Ramey and Zubairy (2018), Sekine and Tsuruga (2018), and Zeev (2019) among others. Let  $\hat{b}_{h,\ell}$  be the least squares (LS) estimate of  $b_{h,\ell}$  in (3) for a given  $h$ . The IRF coefficients  $b_h$  are then estimated by  $\hat{b}_{h,0}$ . Specification in (3) is very similar to the ADL specification in (2), with a couple of differences. First, for  $h > 0$  innovations  $\varepsilon_{t+1}, \dots, \varepsilon_{t+h-1}$  do not feature in (3), while they are included in (2). It can be argued that unless the first  $(h-1)$  IRF coefficients are zero, the inclusion of  $\varepsilon_{t+1}, \dots, \varepsilon_{t+h-1}$  is asymptotically beneficial (although not necessarily in small samples). Teulings and Zubanov (2014) advocate such inclusion to reduce a bias in a panel data context. The second difference is that separate regressions are estimated for each  $h$  in (3).

Note that regressions of the form in (3) are more flexibly specified than those in Jordà (2005) who assumed that  $\mathbf{z}_t$  and the vector of contemporaneous responses  $\mathbf{r}_e = E(\mathbf{z}_t | \varepsilon_t = 1, \mathcal{I}_{t-1}) - E(\mathbf{z}_t | \mathcal{I}_{t-1})$  (as opposed to  $\varepsilon_t$ ) are both observed. In particular, Jordà, (2005, equation (2)) considered the following regression, where we set the lag length of the DGP to one for the simplicity of exposition,

$$\mathbf{z}_{t+h} = \mathbf{B}_h \mathbf{z}_t + \mathbf{v}_{h,t+h}. \quad (4)$$

Regressions in (4) are estimated separately for  $h = 1, 2, \dots, h_{\max}$ . The estimates of the IRF coefficients  $b_h$  are then given by equation (3) in Jordà (2005):

$$\hat{b}_h = \mathbf{s}'_{n,1} \hat{\mathbf{B}}_h \mathbf{r}_1, \text{ for } h = 0, 1, 2, \dots, h_{\max}, \quad (5)$$

where  $\mathbf{B}_0 = \mathbf{I}_n$ . When  $\varepsilon_t$  is observed, a consistent estimate of  $\mathbf{r}_1$ , obtained from regressing  $\varepsilon_t$  onto the residuals of the standard VAR model for  $\mathbf{z}_t$ , can be used in (5) to obtain a consistent estimate of the IRF. We refer to this estimator as JLP to distinguish it from the set of LP regressions in (3).

## 2.2 Iterated approaches

**ARDL approach.** The ARDL approach is based on the following regression

$$x_t = \sum_{\ell=1}^p \psi_{\ell} x_{t-\ell} + \sum_{\ell=0}^q \beta_{\ell} \varepsilon_{t-\ell} + v_t. \quad (6)$$

Let  $\psi(L) = 1 - \sum_{\ell=1}^p \psi_{\ell} L^{\ell}$  and  $\beta(L) = \sum_{\ell=0}^{\infty} \beta_{\ell} L^{\ell}$ , and denote their LS estimates from (6) as  $\hat{\psi}(L)$  and  $\hat{\beta}(L)$ , respectively. Then the IRF coefficients in  $b(L) = \sum_{\ell=0}^{\infty} b_{\ell} L^{\ell}$  can, under some regularity conditions discussed in the online Appendix, be consistently estimated by  $\hat{b}(L) = \hat{\psi}^{-1}(L) \hat{\beta}(L)$  if the truncation lags  $p$  and  $q$  are chosen as an appropriately increasing function of the sample size. In general, choosing  $p$  and  $q$  to be of order  $T^{1/3}$  is sufficient for consistency. The ARDL approach was adopted by Anzuini et al. (2013), Bachmeier and Cha (2011), Coibion (2012), Kimball et al. (2006), Romer and Romer (2004, 2010) or Kilian (2008a, 2008b), among others.

**Vector ARDL (VARDL) approach.** Replacing  $x_t$  with a vector  $\mathbf{z}_{st} = (x_t, q_{1t}, q_{2t}, \dots, q_{s-1,t})'$  that is a subset of  $\mathbf{z}_t$  in (6), we obtain the VARDL approach:

$$\mathbf{\Psi}_s(L) \mathbf{z}_{st} = \boldsymbol{\beta}_s(L) \varepsilon_t + \mathbf{v}_t, \quad (7)$$

where  $\mathbf{\Psi}_s(L) = \mathbf{I}_s - \sum_{\ell=1}^p \mathbf{\Psi}_{s\ell} L^{\ell}$  and  $\boldsymbol{\beta}_s(L) = \sum_{\ell=0}^q \boldsymbol{\beta}_{s\ell} L^{\ell}$  are defined similarly to  $\psi(L)$  and  $\beta(L)$  in (6). Then, under some regularity conditions discussed in the online Appendix, a consistent estimator of the IRF coefficients is given by  $\hat{b}(L) = \mathbf{s}'_1 \hat{\mathbf{\Psi}}_s(L)^{-1} \hat{\boldsymbol{\beta}}_s(L)$  for appropriate truncation lags  $p$  and  $q$ , where  $\mathbf{s}_1$  represents the selection vector that selects the first element, and  $\hat{\mathbf{\Psi}}_s(L)$  and  $\hat{\boldsymbol{\beta}}_s(L)$  are the LS estimates of the polynomials in (7).

**Full-system VARX approach.** If the full set of variables in  $\mathbf{z}_t$  is known, then estimation of  $b_h$  is straightforward. Assuming that the lag order is one (without any loss of generality), we estimate the following full-system VAR augmented by  $\varepsilon_t$  (which we refer to as a full-system VARX),

$$\mathbf{z}_t = \mathbf{\Phi} \mathbf{z}_{t-1} + \mathbf{r}_{\varepsilon} \varepsilon_t + \mathbf{v}_t. \quad (8)$$

The full-system VARX estimates of  $b_h$  are given by  $\hat{b}_h = \mathbf{s}'_1 \hat{\Phi}^h \hat{\mathbf{r}}_\varepsilon$ , for  $h = 0, 1, \dots$ , where  $\hat{\Phi}$  and  $\hat{\mathbf{r}}_\varepsilon$  denote the estimates of  $\Phi$  and  $\mathbf{r}_\varepsilon$  in (8).<sup>4</sup>

### 3 Monte Carlo experiments

We conduct three sets of MC experiments. The first set is presented below and it features stochastically generated parameters. The remaining two sets are designed to match an international real output dataset and a U.S. macro dataset, which are relegated to the online Appendix. These experiments provide qualitatively similar results.

#### 3.1 Simulation design

We generate  $\mathbf{z}_t = (x_t, q_{1t}, q_{2t}, \dots, q_{n-1,t})'$  from

$$\mathbf{z}_t = \Phi \mathbf{z}_{t-1} + \mathbf{R} \boldsymbol{\xi}_t, \quad (9)$$

for  $t = -M + 1, -M + 2, \dots, 0, 1, 2, \dots, T$ , with starting values  $\mathbf{z}_{-M} = \mathbf{0}$ . The first  $M = 100$  generated observations are discarded to minimize the effects of initial values, which leaves us with the available sample size of  $T$ . Structural shocks are generated as  $\boldsymbol{\xi}_t = (\varepsilon_t, \zeta_{1t}, \zeta_{2t}, \dots, \zeta_{n-1,t})' \sim IIDN(\mathbf{0}, \mathbf{I}_n)$ . Individual coefficients in the reduced-form matrix  $\Phi$ , denoted as  $\phi_{ij}$  for  $i, j = 1, 2, \dots, n$ , are generated randomly from

$$\phi_{ij} \sim \begin{cases} IIDU[a, d], & \text{for } i = j \\ IIDU\left[\frac{\phi_{ii} - 0.98}{n-1}, \frac{0.98 - \phi_{ii}}{n-1}\right], & \text{for } i \neq j \end{cases}, \quad (10)$$

where we set  $a = 0.2$  and  $d = 0.98$ . This ensures that the support of  $|\lambda_1(\Phi)|$  is bounded by 0.98, regardless of  $n$ . The matrix  $\mathbf{R}$  is then generated as  $\mathbf{R} = (\mathbf{I}_n - \Phi) \Theta$ , where we set the first element of the long-run elasticity matrix  $\Theta$  to one, and the remaining elements of  $\Theta$  are generated from  $IIDU(0.2, 1)$ .<sup>5</sup> This ensures  $\sum_{h=0}^{\infty} b_h = 1$ , where  $b_h = E(x_{t+h} | \varepsilon_t = 1, \mathcal{I}_{t-1}) - E(x_{t+h} | \mathcal{I}_{t-1})$  is the IRF function defined in the previous section. We consider  $n \in \{6, 12\}$ ,  $T \in \{30, 150, 500\}$ , and conduct 10,000 replications for each simulation experiment.

<sup>4</sup>Unlike the VARDL specifications in (7), the specification in (8) does not require any lag terms of  $\varepsilon_t$  because it utilizes the information on all the variables entering DGP as well as the true lag length of the DGP VAR (set to one here for simplicity).

<sup>5</sup> $\Phi$  and  $\Theta$  are generated randomly for each  $n$  at the beginning of experiments and kept fixed across replications.

We implement all of the approaches outlined in Section 2 as summarized in Table 1. We include a constant term in each regression, and set  $h_{\max} = 8$ . The number of lags  $(p, q)$  in the ADL, LP, JLP, and the full-system VARX regressions are set to one.<sup>6</sup> Lag lengths in the iterated ARDL and VARDL approaches are selected from the BIC criterion.<sup>7</sup> For the vectors of additional covariates in the ADL, LP, and VARDL approaches, we consider selecting one and two additional variables from  $\mathbf{z}_t$ .<sup>8</sup> In addition, we also consider using all  $(n - 1)$  covariates,  $q_{1t}, q_{2t}, \dots, q_{n-1,t}$ .

### 3.2 Simulation results

Table 2 presents the results of our simulation exercise, with the RMSE of the IRF coefficients  $(b_h)$  for  $h = 1, 2, \dots, h_{\max}$  estimated from each competing approach *relative* to the full-system VARX approach used as a benchmark.<sup>9</sup> That is, the table displays a horse race of the relative performance of each approach and the benchmark model in estimating IRFs in a linear setup. We report average relative RMSE by averaging across horizons,  $h = 0, 1, 2$  (columns labeled ‘short’ horizon),  $h = 3, 4, 5$  (‘medium’ horizon), and  $h = 6, 7, 8$  (‘long’ horizon). Entries lower than one indicate a superior performance of the corresponding approach over the benchmark, whereas values higher than one indicates the dominance of the benchmark. Beware that our benchmark is infeasible in the sense that it utilizes the knowledge of the true lag length and the identity of all relevant variables entering the DGP. Table 2 consists of six blocks, with different combinations of  $n$  and  $T$ .

The results in Table 2 illustrate several points. First, the difference in RMSE across various approaches under study is substantial, suggesting that the selection of an appropriate approach can be consequential. In all sample sizes considered, the RMSE difference between the worst and the best approach is at least two-fold.

Second, no single approach uniformly dominates. We find that the iterated approaches like ARDL and VARDL generally work well, but their performance hinges on the sample size. While univariate ARDL or bivariate VARDL (using  $s = 2$  variables consisting of the target variable  $x_t$  and one additional covariate selected by the BIC rule) are the best performers for relatively

---

<sup>6</sup>This is asymptotically optimal when  $\mathbf{y}_t = \mathbf{z}_t$ . But it is unclear how to select an optimal lag length for ADL and LP specifications when  $\mathbf{y}_t \neq \mathbf{z}_t$ .

<sup>7</sup>The maximum lag lengths are set as  $p_{\max} = q_{\max} = \min \left\{ \left[ T^{1/3} \right], p_{80} \right\}$ , where  $\left[ T^{1/3} \right]$  is the integer part of  $T^{1/3}$ , and  $p_{80}$  is the largest lag length that ensures the degrees of freedom span at least 80% of the sample size.

<sup>8</sup>In the case of VARDL, the selection of variables and lag length is done jointly by the BIC. In the case of direct approaches, variables are selected based on the best fit.

<sup>9</sup>The chosen benchmark is asymptotically efficient.



small samples ( $T = 30$ ), they are outperformed by higher-dimensional VARDL in larger samples. This is intuitive because including all variables that are part of the true DGP is more efficient asymptotically (as  $T \rightarrow \infty$ ), but including all relevant variables in relatively small sample sizes can be counterproductive due to larger number of unknown parameters to estimate and thus possibly larger sampling uncertainty. When we compare the best performing direct approach with the best performing iterated approach, the latter outperforms by about 20-12,000 percent across sample sizes. The dominance of iterated approaches is stronger for longer horizons.<sup>10</sup> We also find that the ARDL approach dominates the most parsimonious LP approach ( $s = 1$ ) by about 22 to 6,600 percent depending on the sample size.

Next, when it comes to the direct approach, augmentation of the LP regressions in (3) by  $\varepsilon_{t+1}, \dots, \varepsilon_{t+h-1}$  can be beneficial, but only if the time dimension is sufficiently large and the regression is not augmented with too many unknown parameters (relative to  $T$ ). Improvement is noted in the range of 14-42 percent for  $T = 500$ , 11-41 percent for  $T = 150$ , but the results are mixed for  $T = 30$ . Compared to the ADL approach, LP and ALP approaches tend to achieve lower RMSE, implying that running separate regressions for different horizons is beneficial, in line with the asymptotic arguments.

These main conclusions also hold in the additional two sets of calibrated MC experiments presented in the online Appendix.

## 4 Conclusion

We find a substantial heterogeneity in the RMSE performance of various approaches to estimating IRFs in a linear setting when shocks are observed. Our simulation results suggest that iterated (ARDL/VARDL) approaches tend to outperform other competing approaches. In addition, the common practice of including all variables that are deemed to be part of the DGP can perform poorly in small samples. This paper does not address an equally important issue of inference about the impulse responses, which is left for future research.

---

<sup>10</sup>The same pattern is noticed in our second set of MC experiments (calibrated to international output growth dataset), but not in the third set of experiments (calibrated to a U.S. macro dataset). The DGP in the third experiment is very persistent, and in one case ( $T = 500$ , and longer horizon) the best performing direct approach (ALP with  $s = n$ ) outperforms the VARDL approach.

## References

- Anzuini, A., M. J. Lombardi, and P. Pagano (2013). The Impact of Monetary Policy Shocks on Commodity Prices. *International Journal of Central Banking* 9(3), 125–150.
- Bachmeier, L. J. and I. Cha (2011). Why Don't Oil Shocks Cause Inflation? Evidence from Disaggregate Inflation Data. *Journal of Money, Credit and Banking* 43(6), 1165–1183. <https://doi.org/10.1111/j.1538-4616.2011.00421.x>.
- Baumeister, C. and L. Kilian (2014). Do oil price increases cause higher food prices? *Economic Policy* 29(80), 691–747. <https://doi.org/10.1111/1468-0327.12039>.
- Coibion, O. (2012). Are the Effects of Monetary Policy Shocks Big or Small? *American Economic Journal: Macroeconomics* 4(2), 1–32. <https://doi.org/10.1257/mac.4.2.1>.
- Evans, C. L. and D. A. Marshall (1998). Monetary policy and the term structure of nominal interest rates: Evidence and theory. *Carnegie-Rochester Conference Series on Public Policy* 49, 53–111. [https://doi.org/10.1016/S0167-2231\(99\)00004-4](https://doi.org/10.1016/S0167-2231(99)00004-4).
- Gorodnichenko, Y. and B. Lee (2017, November). A Note on Variance Decomposition with Local Projections. NBER Working Papers 23998, National Bureau of Economic Research, Inc.
- Guaajardo, J., D. Leigh, and A. Pescatori (2014). Expansionary Austerity? International Evidence. *Journal of the European Economic Association* 12(4), 949–968. <https://doi.org/10.1111/jeea.12083>.
- Jordà, Ò. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review* 95(1), 161–182. <https://doi.org/10.1257/0002828053828518>.
- Kilian, L. (2008a). A Comparison of the Effects of Exogenous Oil Supply Shocks on Output and Inflation in the G7 Countries. *Journal of the European Economic Association* 6(1), 78–121. <https://doi.org/10.1162/JEEA.2008.6.1.78>.
- Kilian, L. (2008b). Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy? *The Review of Economics and Statistics* 90(2), 216–240. <https://doi.org/10.1162/rest.90.2.216>.
- Kilian, L. (2009). Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market. *American Economic Review* 99(3), 1053–1069. <https://doi.org/10.1257/aer.99.3.1053>.
- Kilian, L. and Y. J. Kim (2011). "how reliable are local projection estimators of impulse responses?". *The Review of Economics and Statistics* 93, 1460–1466. [https://doi.org/10.1162/REST\\_a\\_00143](https://doi.org/10.1162/REST_a_00143).
- Kilian, L. and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press.
- Kimball, M. S., J. G. Fernald, and S. Basu (2006). Are Technology Improvements Contractionary? *American Economic Review* 96(5), 1418–1448. <https://doi.org/10.1257/aer.96.5.1418>.

- Ramey, V. A. and S. Zubairy (2018). Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data. *Journal of Political Economy* 126(2), 850–901. <https://doi.org/10.1086/696277>.
- Romer, C. D. and D. H. Romer (2004). A New Measure of Monetary Shocks: Derivation and Implications. *American Economic Review* 94(4), 1055–1084. <https://doi.org/10.1257/0002828042002651>.
- Romer, C. D. and D. 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. <https://doi.org/10.1257/aer.100.3.763>.
- Sekine, A. and T. Tsuruga (2018). Effects of commodity price shocks on inflation: A cross-country analysis. *Oxford Economic Papers* 70(4), 1108–1135. <https://doi.org/10.1093/oep/gpy015>.
- Stock, J. H. and M. W. Watson (2017). Twenty Years of Time Series Econometrics in Ten Pictures. *Journal of Economic Perspectives* 31(2), 59–86. <https://doi.org/10.1257/jep.31.2.59>.
- Teulings, C. N. and N. Zubanov (2014). Is Economic Recovery A Myth? Robust Estimation Of Impulse Responses. *Journal of Applied Econometrics* 29(3), 497–514. <https://doi.org/10.1002/jae.2333>.
- Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Uppsala: Almqvist and Wiksel.
- Zeev, N. B. (2019). Global credit supply shocks and exchange rate regimes. *Journal of International Economics* 116, 1–32. <https://doi.org/10.1016/j.jinteco.2018.10.002>.

**Table 1: Overview of approaches**

Approach	Regression
<b>A. Direct approaches</b>	
The most parsimonious approach	$x_t = b_h \varepsilon_{t-h} + v$ , for $h = 0, 1, \dots, h_{\max}$
Distributed lag (DL)	$x_t = \sum_{h=0}^{h_{\max}} b_h \varepsilon_{t-h} + v_t$
Augmented distributed lag (ADL)	$x_t = \sum_{h=0}^{h_{\max}} b_h \varepsilon_{t-h} + \beta' \mathbf{y}_{t-h_{\max}-1} + v_t$
Local projections (LP)	$x_{t+h} = \sum_{\ell=0}^p b_{h,\ell} \varepsilon_{t-\ell} + \sum_{\ell=1}^q \beta'_{h,\ell} \mathbf{y}_{t-\ell} + v_{h,t+h}$ , or (JLP) $\mathbf{z}_{t+h} = \mathbf{B}_h \mathbf{z}_t + \mathbf{v}_{h,t+h}$ , for $h = 0, 1, \dots, h_{\max}$
<b>B. Iterated approaches</b>	
Autoregressive distributed lag (ARDL)	$x_t = \sum_{\ell=1}^p \psi_\ell x_{t-\ell} + \sum_{\ell=0}^q \beta_\ell \varepsilon_{t-\ell} + v_t$
Vector autoregressive distributed lag (VARDL)	$\Psi_s(L) \mathbf{z}_{st} = \beta_s(L) \varepsilon_t + \mathbf{v}_t$
Full-system augmented vector autoregression (VARX)	$\mathbf{z}_t = \Phi \mathbf{z}_{t-1} + \mathbf{r}_\varepsilon \varepsilon_t + \mathbf{v}_t$ .

Notes:  $x_t$  is the target variable, and  $\varepsilon_t$  is the shock of interest (not necessarily structural). We assume  $\mathbf{z}_t$  is an  $n \times 1$  vector of variables generated from a stationary VAR model, see the online Appendix for assumptions.  $x_t$  is included in  $\mathbf{z}_t$ .  $\mathbf{z}_{st}$  is a subset of  $\mathbf{z}_t$  and it includes  $x_t$ .  $\mathbf{y}_t$  is an  $s \times 1$  vector of covariates that are assumed to be uncorrelated with *future* values of  $\varepsilon_t$ .

**Table 2: Monte Carlo findings for the relative RMSE of individual approaches**

Approach \ horizon	$n = 6, T = 30$			$n = 6, T = 150$			$n = 6, T = 500$		
	short	medium	long	short	medium	long	short	medium	long
The most parsimonious	1.55	2.09	3.44	2.76	3.70	7.29	3.20	4.25	9.21
DL	2.33	3.05	4.04	1.85	2.34	4.65	1.90	2.43	5.34
ADL, $s = 1$	2.87	4.13	5.60	1.90	2.43	4.83	1.91	2.44	5.38
ADL, $s = 2$	3.35	5.40	6.54	1.94	2.53	5.02	1.93	2.47	5.45
ADL, $s = n$	6.37	12.65	18.77	2.03	2.73	5.54	1.95	2.52	5.59
LP, $s = 1$	1.24	2.29	3.74	1.77	3.51	7.30	1.97	3.95	9.17
ALP, $s = 1$	0.92	1.91	4.38	1.22	2.18	4.71	1.33	2.35	5.34
LP, $s = 2$	1.42	2.44	3.85	1.84	3.67	7.52	1.98	4.01	9.26
ALP, $s = 2$	0.97	1.97	4.65	1.21	2.23	4.84	1.31	2.36	5.39
LP, $s = n$	2.31	3.45	4.88	2.05	4.16	8.22	2.06	4.19	9.56
ALP, $s = n$	1.43	3.63	12.92	1.29	2.45	5.40	1.33	2.42	5.54
ARDL	0.87	0.69	<b>0.31</b>	1.10	1.12	0.69	1.21	1.23	0.81
VARDL, $s = 2$	<b>0.58</b>	<b>0.37</b>	0.33	1.15	1.25	<b>0.65</b>	1.39	1.60	<b>0.78</b>
VARDL, $s = 3$	0.73	0.66	0.45	1.08	1.20	1.19	1.20	1.29	1.23
VARDL, $s = n$	1.01	1.01	1.01	<b>1.00</b>	<b>1.00</b>	1.00	<b>1.00</b>	<b>1.00</b>	1.00
JLP	2.11	1.90	2.22	1.94	4.13	8.27	1.68	4.30	10.34

  

Approach \ horizon	$n = 12, T = 30$			$n = 12, T = 150$			$n = 12, T = 500$		
	short	medium	long	short	medium	long	short	medium	long
The most parsimonious	0.75	2.04	2.39	1.55	8.01	22.23	1.67	9.31	27.07
DL	1.50	3.81	3.56	1.45	7.08	19.07	1.43	7.60	22.03
ADL, $s = 1$	1.82	4.77	4.43	1.48	7.23	19.40	1.44	7.63	22.15
ADL, $s = 2$	2.46	6.63	5.94	1.51	7.48	20.05	1.44	7.70	22.31
ADL, $s = n$	>100	>100	>100	1.65	8.30	22.44	1.48	7.92	22.99
LP, $s = 1$	0.77	2.21	2.62	1.44	8.06	22.44	1.53	9.27	27.11
ALP, $s = 1$	0.71	2.29	3.52	1.24	6.70	19.02	1.31	7.46	22.02
LP, $s = 2$	0.84	2.39	2.82	1.46	8.18	22.91	1.53	9.31	27.22
ALP, $s = 2$	0.79	2.64	4.33	1.25	6.87	19.58	1.30	7.52	22.16
LP, $s = n$	1.63	4.64	6.66	1.61	8.92	24.85	1.59	9.53	27.81
ALP, $s = n$	1.59	9.47	>100	1.36	7.66	22.00	1.33	7.73	22.86
ARDL	0.63	0.30	0.04	0.96	0.50	<b>0.16</b>	1.02	0.90	<b>0.54</b>
VARDL, $s = 2$	<b>0.37</b>	<b>0.10</b>	<b>0.04</b>	0.63	0.39	0.45	0.80	0.67	0.92
VARDL, $s = 3$	0.40	0.12	0.05	<b>0.59</b>	<b>0.35</b>	0.43	<b>0.63</b>	<b>0.43</b>	0.60
VARDL, $s = n$	1.10	1.31	2.34	1.00	1.00	1.00	1.00	1.00	1.00
JLP	0.79	0.93	1.20	1.45	7.28	20.32	1.50	9.51	27.82

Notes: Entries represent the ratio of the average RMSE of estimating  $b_h$  from each approach to that of the benchmark full-system VARX approach. The values smaller than one indicate the cases where the RMSE of the corresponding approach is smaller (and hence better) than that of the benchmark approach. Numbers in bold face indicate the cases with the lowest (best) values. IRF horizon ‘short’ stands for  $h = 0, 1, 2$ , ‘medium’ for  $h = 3, 4, 5$  and ‘long’ for  $h = 6, 7, 8$ . The DGP is a VAR(1) model shown in (9), where the reduced form coefficient matrix ( $\Phi$ ) is generated randomly from (10) for each  $n$  and kept fixed across replications. Descriptions of each approach are provided in Section 2. The benchmark approach and the JLP approach assume that the identity of variables entering the DGP and the true lag length are known. The other approaches do not necessarily utilize all variables. Section 3 provides a full description of MC experiments.

# A Online Appendix for "Estimating Impulse Response Functions When the Shock Series is Observed" by C-Y Choi and A. Chudik

This online appendix is organized as follows. Section A.1 describes the modeling setup and outlines the main assumptions. Section A.2 discusses the consistency of the most parsimonious, DL, ADL, ARDL, and VARDL approaches. Section A.3 presents Monte Carlo experiments using a design calibrated to a multicountry output growth VAR. Section A.4 presents Monte Carlo experiments using a design calibrated to a US macro dataset.

## A.1 Modeling Setup

Suppose an  $n \times 1$  dimensional vector of variables  $\mathbf{z}_t$  is generated by the following stationary VAR model,

$$\mathbf{z}_t = \mathbf{\Phi} \mathbf{z}_{t-1} + \mathbf{u}_t, \tag{A.1}$$

for  $t = \dots, 0, 1, 2, \dots, T$ , where  $\mathbf{\Phi}$  is an  $n \times n$  matrix of coefficients and  $\mathbf{u}_t$  is an  $n \times 1$  vector of reduced form shocks, which is partitioned as

$$\mathbf{u}_t = \mathbf{r} \varepsilon_t + \boldsymbol{\zeta}_t, \tag{A.2}$$

where  $\varepsilon_t$  denotes the observed shock of interest which is uncorrelated with  $\boldsymbol{\zeta}_t$ . The observed shock ( $\varepsilon_t$ ) might or might not have a structural interpretation. We assume one lag and no deterministic terms in (A.1) for the simplicity of the exposition. Augmenting additional lags and/or deterministic terms in (A.1) is conceptually straightforward. Not all variables in  $\mathbf{z}_t$  are necessarily observed as it might be the case in empirical work where the choice of variables is often a thorny issue. It is assumed that  $\mathbf{z}_t$  can be partitioned as  $\mathbf{z}_t = (x_t, \mathbf{q}_t)'$ , where  $x_t$  is the observed variable of interest and  $\mathbf{q}_t$  is an  $(n - 1) \times 1$  vector of the remaining variables (also referred to as covariates) whose components can be partially observed.

The following assumptions are postulated to hold for the discussions below.

**ASSUMPTION 1**  $|\lambda_1(\mathbf{\Phi})| \leq \rho < 1$ , where  $\lambda_1(\mathbf{\Phi})$  is the largest eigenvalue of  $\mathbf{\Phi}$ .

**ASSUMPTION 2** Let  $\mathbf{u}_t = \mathbf{R} \boldsymbol{\xi}_t$ , where the elements of  $\mathbf{R}$  are bounded, and  $\boldsymbol{\xi}_t \sim IID(\mathbf{0}_{k \times k}, \mathbf{I}_k)$ . The first element of  $\boldsymbol{\xi}_t$  is denoted as  $\varepsilon_t$  so that  $\boldsymbol{\xi}_t = (\varepsilon_t, \boldsymbol{\xi}'_{-1,t})'$ .

These standard assumptions ensure stationarity. Assumption 1 is the standard stationarity condition for the coefficient matrix  $\Phi$ . Assumption 2 is also standard in the literature which implies the decomposition in (A.2) where  $\mathbf{r}$  is the first element of  $\mathbf{R} = (\mathbf{r}, \mathbf{R}_{-1})$ .  $\zeta_t = \mathbf{R}_{-1}\xi_{-1,t}$  is by assumption uncorrelated with the observed shock  $\varepsilon_t$ , which is necessary for the consistency of the estimation approaches below.

Under Assumptions 1-2,  $x_t$  has the following moving average representation,

$$x_t = \sum_{h=0}^{\infty} \mathbf{s}'_{n,1} \Phi^h \mathbf{u}_{t-h} = \sum_{h=0}^{\infty} \mathbf{s}'_{n,1} \Phi^h \mathbf{r} \varepsilon_{t-h} + \sum_{h=0}^{\infty} \mathbf{s}'_{n,1} \Phi^h \zeta_{t-h} = \sum_{h=0}^{\infty} b_h \varepsilon_{t-h} + e_t, \quad (\text{A.3})$$

where  $\mathbf{s}_{n,1} = (1, 0, \dots, 0)'$  is an  $n \times 1$  selection vector that selects the first element,

$$b_h = \mathbf{s}'_{n,1} \Phi^h \mathbf{r}, \text{ for } h = 0, 1, \dots, \quad (\text{A.4})$$

and  $e_t = \sum_{h=0}^{\infty} \mathbf{s}'_{n,1} \Phi^h \zeta_{t-h}$ . The sequence  $\{b_h\}_{h=0}^{\infty}$  is the impulse response function of a unit shock to  $\varepsilon_t$  on the variable of interest,  $x_t$ . Each approach for a consistent estimation of  $\{b_h\}$  is discussed in Section 2 of the paper, with a more detailed discussion provided below on the consistency of the most parsimonious, DL, ADL, ARDL, and VARDL approaches.

## A.2 Discussion of competing approaches

### A.2.1 The most parsimonious approach

The simplest, most parsimonious approach is based on:

$$x_t = b_h \varepsilon_{t-h} + v_t, \text{ for } h = 0, 1, 2, \dots, h_{\max}, \quad (\text{A.5})$$

where  $h_{\max}$  can be a non-decreasing function of the available sample size,  $T$ , and  $v_t$  is a generic regression error term, which clearly depends on the regression specification (suppressed in terms of notations), and it takes different forms throughout the paper and in this Appendix. In the context of (A.5),  $v_t$  is given by (using the moving-average representation in (A.3)),

$$v_t = e_t + \sum_{j=0, j \neq h}^{\infty} b_j \varepsilon_{t-j}. \quad (\text{A.6})$$

Clearly,  $v_t$  is serially correlated but is uncorrelated with  $\varepsilon_{t-h}$ , and hence it is not surprising that the LS estimate of  $b_h$  using the auxiliary regression (A.5), denoted as  $\hat{b}_h$ , is consistent for any given  $h$ .

### A.2.2 DL

The DL approach is based on (1) where the corresponding regression error term is given by

$$v_t = e_t + \sum_{h=h_{\max}+1}^{\infty} b_h \varepsilon_{t-h}. \quad (\text{A.7})$$

Clearly,  $v_t$  in (A.7) is uncorrelated with regressors,  $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-p_T}$ , in (1), and therefore the corresponding LS estimates of the IRF coefficient vector  $\mathbf{b}_{h_{\max}} = (b_0, b_1, \dots, b_{h_{\max}})'$ , denoted as  $\hat{\mathbf{b}}_{h_{\max}}^{DL} = (\hat{b}_0^{DL}, \hat{b}_1^{DL}, \dots, \hat{b}_{h_{\max}}^{DL})'$ , will be consistent for any fixed  $h_{\max}$ .<sup>11</sup>

### A.2.3 ADL

So long as the regressors  $\mathbf{y}_{t-h_{\max}-1}$  are uncorrelated with  $\varepsilon_{t-h}$  for  $h = 0, 1, \dots, h_{\max}$ , it is not surprising that, as in the case of the DL regressions, the same conclusion on the consistency continues to apply. When augmenting the DL regression in (1), it should be noted that it is not necessarily advisable to use many regressors in  $\mathbf{y}_t$  to avoid overfitting.

### A.2.4 ARDL

Autoregressive distributed lag approach involves univariate regressions featuring the current and lagged values of  $\varepsilon_t$  as well as the lagged terms of the dependent variable,  $x_t$ . Using (A.3),  $x_t$  can be decomposed into two orthogonal components, one that depends on  $\{\varepsilon_{t-h}\}_{h=0}^{\infty}$  and the other that depends on the remaining shocks  $\{\zeta_{t-h}\}_{h=0}^{\infty}$ :

$$x_t = b(L) \varepsilon_t + e_t, \quad (\text{A.8})$$

where  $b(L) = \sum_{h=0}^{\infty} b_h L^h$  with  $b_h$ , for  $h = 0, 1, \dots$ , defined in (A.4). The error term  $e_t$  is covariance stationary, and from the Wold decomposition theorem (Wold, 1938), it has an MA( $\infty$ )

<sup>11</sup>Similarly to the LP approach, regression (1) can also be run for  $h_{\max} = 0, 1, 2, \dots, H$ , which gives us  $H - h + 1$  different estimates of  $b_h$ . Any of these estimates is consistent. These estimators are not pursued in the current paper.



representation

$$e_t = \alpha(L) \nu_t, \quad (\text{A.9})$$

where  $\alpha(L) = 1 - \sum_{h=1}^{\infty} \alpha_h L^h$ ,  $\nu_t = e_t - \hat{E}(e_t | e_{t-1}, e_{t-2}, \dots)$ , and  $\hat{E}$  denotes the linear prediction operator. We assume that  $\psi(L) = \alpha^{-1}(L) = 1 - \sum_{h=0}^{\infty} \psi_h L^h$  exists and its coefficients decay at an exponential rate. Then, multiplying both sides of (A.8) by  $\psi(L)$  yields

$$\psi(L) x_t = \psi(L) b(L) \varepsilon_t + \psi(L) e_t = \beta(L) \varepsilon_t + \nu_t,$$

or

$$x_t = \sum_{h=1}^{\infty} \psi_h x_{t-h} + \sum_{h=0}^{\infty} \beta_h \varepsilon_{t-h} + \nu_t, \quad (\text{A.10})$$

where  $\beta(L) = \sum_{h=0}^{\infty} \beta_h L^h = \psi(L) b(L)$  and  $\psi(L) e_t = \nu_t$ . Since  $\psi(L)$  and  $b(L)$  feature exponentially declining coefficients, their product term,  $\beta(L)$ , has exponentially declining coefficients as well. The ARDL estimation of the IRF coefficients is based on the truncated version of (A.10) as shown in (6). Since all the coefficients decay at an exponential rate, the truncation errors are negligible when the truncation lags increase with the sample size at an appropriate rate.

### A.2.5 VARDL

Let  $\mathbf{z}_{st} = (x_t, q_{1t}, q_{2t}, \dots, q_{s-1,t})'$  denote an  $(s \times 1)$  vector of observed variables. Using the moving average representation of the data generating process in (A.1), we have the following decomposition similar to (A.8),

$$\begin{aligned} \mathbf{z}_{st} &= \mathbf{S}'_s \mathbf{z}_t = \mathbf{S}'_s \sum_{h=0}^{\infty} \mathbf{\Phi}^h \mathbf{R} \boldsymbol{\xi}_{t-h} \\ &= \sum_{h=0}^{\infty} \mathbf{S}'_s \mathbf{\Phi}^h \mathbf{r}_1 \varepsilon_{t-h} + \sum_{h=0}^{\infty} \mathbf{S}'_s \mathbf{\Phi}^h \mathbf{R}_{-1} \zeta_{t-h} \\ &= \sum_{h=0}^{\infty} \mathbf{b}_{sh} \varepsilon_{t-h} + \mathbf{e}_{st}, \\ &= \mathbf{b}_s(L) \varepsilon_t + \mathbf{e}_{st} \end{aligned} \quad (\text{A.11})$$

where  $\mathbf{S}_s$  is an  $(n \times s)$  selection matrix that selects the first  $s$  elements of  $\mathbf{z}_t$ ,  $\mathbf{b}_{sh} = \mathbf{S}'_s \Phi^h \mathbf{r}_1$ , and

$$\mathbf{e}_{st} = \sum_{h=0}^{\infty} \mathbf{S}'_s \Phi^h \mathbf{R}_{-1} \zeta_{t-h}, \quad (\text{A.12})$$

for  $s = 2, 3, \dots, n-1$ , and  $h = 0, 1, \dots$ . Note that  $\mathbf{e}_{st}$  is an  $s$ -dimensional vector of covariance stationary variables. Similarly to (A.9),  $\mathbf{e}_{st}$  has an MA( $\infty$ ) representation,

$$\mathbf{e}_{st} = \mathbf{A}_s(L) \boldsymbol{\nu}_{st}, \quad (\text{A.13})$$

where  $\boldsymbol{\nu}_{st}$  has a zero mean with constant variance  $\boldsymbol{\Omega}_{uk}$ . It is serially uncorrelated and independent with  $\varepsilon_{t'}$  for all  $t, t'$ , and coefficients of  $\mathbf{A}_s(L) = \mathbf{I}_s - \sum_{h=1}^{\infty} \mathbf{A}_{sh} L^h$  decay at an exponential rate. Similarly to the ARDL approach, we require that  $\mathbf{e}_{st}$  has an invertible vector MA( $\infty$ ) representation so that  $\mathbf{e}_{st}$  can be expressed as a VAR( $\infty$ ) process with exponentially decaying coefficients. To this end, we assume that  $\boldsymbol{\Psi}_s(L) = \mathbf{A}_s^{-1}(L) = \mathbf{I}_s - \sum_{h=0}^{\infty} \boldsymbol{\Psi}_{sh} L^h$  exists and its coefficients decay at an exponential rate.

Multiplying both sides of (A.11) by  $\boldsymbol{\Psi}_s(L)$  from the left, and making use of the representation in (A.13), we obtain similar to (A.10)

$$\boldsymbol{\Psi}_s(L) \mathbf{z}_{st} = \boldsymbol{\beta}_s(L) \varepsilon_t + \boldsymbol{\nu}_{st}, \quad (\text{A.14})$$

where  $\mathbf{u}_{st} = \boldsymbol{\Psi}_s(L) \mathbf{e}_{st}$ ,  $\boldsymbol{\beta}_s(L) = \mathbf{A}_s(L) \mathbf{b}_s(L)$ , and the coefficients of  $\boldsymbol{\beta}_s(L)$  decay at an exponential rate. The truncated version of (A.14) is given by (7), and the truncation lag errors become negligible when the truncation lag increases with the sample size at an appropriate rate.

### A.3 Experiments using design calibrated to a global output growth VAR

Our second simulation design is based on a quarterly international GDP dataset. Let  $\mathbf{z}_t = (z_{1,t}, z_{2,t}, \dots, z_{n,t})'$  be the  $(n \times 1)$  vector of observations for  $n$  countries in period  $t$  where  $z_{it}$  denotes the first difference of the logarithm of real GDP in country  $i$  at quarter  $t$ . We consider  $n = 10$  large economies: Canada, China, France, Germany, Italy, Japan, Korea, Mexico, the United Kingdom, and the United States, that account for a slightly more than a half of the global output in the

Purchasing Power Parity terms (in 2015). We estimate the following reduced-form VAR model,

$$\mathbf{z}_t = \mathbf{c} + \mathbf{\Phi}\mathbf{z}_{t-1} + \mathbf{u}_t, \quad (\text{A.15})$$

using the sample 1980:Q3 - 2015:Q2 (140 quarterly observations). Let  $\hat{\mathbf{c}}$  and  $\hat{\mathbf{\Phi}}$  in (A.15) respectively denote the LS estimates of  $\mathbf{c}$  and  $\mathbf{\Phi}$ , and let  $\hat{\mathbf{\Sigma}}$  be the estimate of the variance-covariance matrix of  $\mathbf{u}_t$  in (A.15). We then generate simulated data  $\mathbf{z}_t^{(r)}$  based on

$$\mathbf{z}_t^{(r)} = \hat{\mathbf{c}} + \hat{\mathbf{\Phi}}\mathbf{z}_{t-1}^{(r)} + \mathbf{u}_t^{(r)}, \quad (\text{A.16})$$

for  $t = -M + 1, -M + 2, \dots, 0, 1, 2, \dots, T$ , with starting values  $\mathbf{z}_{-M}^{(r)} = \mathbf{0}$ , where

$$\mathbf{u}_t^{(r)} \sim IIDN(\mathbf{0}, \hat{\mathbf{\Sigma}}).$$

We use superscript  $(r)$  to denote the individual MC replications,  $r = 1, 2, \dots, R$  where  $R = 10,000$ . The first  $M = 100$  generated observations are discarded to minimize the effects of initial values, which leaves us with the available sample size of  $T$ . The U.S. is ordered the first in  $\mathbf{z}_t$  so as to estimate the generalized IRF function for the shock to U.S. output growth on its neighbor Canada. We consider the same set of approaches and sample sizes discussed in the paper. Table A1 reports the results of this simulation exercise by comparing the RMSE of each competing approach relative to the benchmark of the full-system VARX approach. The results appear to be qualitatively very similar to that of Table 2.

**Table A1: MC findings for the relative RMSE of estimating IRF coefficients in design calibrated to international output dataset**

Approach \ horizon	$T = 30$			$T = 150$			$T = 500$		
	short	medium	long	short	medium	long	short	medium	long
The most parsimonious	1.39	3.36	7.25	2.34	7.02	23.42	2.47	7.50	23.50
DL	2.24	5.33	9.31	1.81	5.09	17.21	1.75	5.00	15.92
ADL, $s = 1$	2.66	6.64	12.53	1.84	5.21	17.73	1.76	5.03	16.03
ADL, $s = 2$	3.48	8.67	15.77	1.91	5.37	18.22	1.77	5.07	16.16
ADL, $s = N$	>100	>100	>100	1.99	5.65	19.25	1.79	5.15	16.35
LP, $s = 1$	1.34	3.68	7.99	2.02	6.90	23.60	2.08	7.21	23.43
ALP, $s = 1$	1.15	3.27	9.53	1.64	4.81	17.29	1.69	4.91	15.95
LP, $s = 2$	1.32	4.02	8.79	1.79	6.88	24.03	1.83	7.12	23.46
ALP, $s = 2$	1.14	3.85	12.09	1.44	4.81	17.76	1.46	4.82	16.05
LP, $s = N$	1.74	6.23	15.29	1.67	7.20	25.29	1.62	7.14	23.74
ALP, $s = N$	1.53	7.26	>100	1.35	4.97	18.86	1.33	4.81	16.24
ARDL	1.06	0.72	0.21	1.65	1.98	1.09	2.17	3.92	2.73
VARDL, $s = 2$	<b>0.78</b>	<b>0.33</b>	<b>0.18</b>	1.22	1.05	1.11	1.36	2.11	2.29
VARDL, $s = 3$	0.78	0.42	0.21	1.15	<b>0.86</b>	<b>0.72</b>	1.43	1.20	1.05
VARDL, $s = N$	1.03	1.05	1.07	<b>1.00</b>	1.00	1.00	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
JLP	1.35	1.97	3.95	1.54	5.46	19.12	1.48	6.10	20.77

Notes: The entries represent the RMSE of each approach relative to the full-system VARX benchmark. The DGP is given by VAR(1) model in (A.16), which features coefficients in  $\hat{\mathbf{c}}$  and  $\hat{\mathbf{\Phi}}$  estimated using real output growth data on  $n = 10$  economies over the sample 1980Q3 - 2015Q2. The lowest (best) entries are highlighted by bold fonts. See Section A.3 and the notes in Table 2 for further details.

#### A.4 Experiments using design calibrated to a US macro dataset

Our third set of MC experiments is calibrated to a U.S. macro dataset. We obtain the dataset used in the MC study by Jordà (2005) from the AEA website.<sup>12</sup> This dataset contains monthly observations from January 1960 to February 2001 for the following six variables: (i) log of non-agricultural payroll employment; (ii) log of personal expenditure deflator; (iii) annual growth rate of the index of sensitive material prices by the Conference Board; (iv) federal funds rate; (v) ratio of nonborrowed reserves plus extended credit to total reserves; and (vi) annual growth rate of M2. Collecting these variables in the vector  $\mathbf{z}_t$  (using the same ordering), we estimate a VAR( $p_0$ ) model

$$\mathbf{z}_t = \mathbf{c} + \sum_{\ell=0}^{p_0} \mathbf{\Phi}_{\ell} \mathbf{z}_{t-\ell} + \mathbf{u}_t, \quad (\text{A.17})$$

<sup>12</sup>[http://www.aeaweb.org/aer/data/mar05\\_data\\_jorda.zip](http://www.aeaweb.org/aer/data/mar05_data_jorda.zip), (the unpacked file evnew.csv).

and identify the monetary policy shock using the Christiano-Eichenbaum-Evans recursive identification strategy as described in Section 2.2 of Evans and Marshall (1998). In particular, let  $\hat{\mathbf{c}}$ ,  $\hat{\mathbf{\Phi}}_\ell$ ,  $\hat{\mathbf{u}}_t$  denote the LS reduced-form estimates, and let  $\hat{\mathbf{\Sigma}}$  denote the estimated variance-covariance matrix of error term. Note the federal funds rate is ordered as the fourth after the variables (i)-(iii). After obtaining the lower triangular Cholesky factorization of  $\hat{\mathbf{\Sigma}} = \hat{\mathbf{C}}\hat{\mathbf{C}}'$ , we compute  $\hat{\boldsymbol{\varepsilon}}_t = \hat{\mathbf{C}}^{-1}\hat{\mathbf{u}}_t$ , of which the fourth element is the normalized identified monetary policy shock with a unit variance.

The DGP is then given by

$$\mathbf{z}_t^{(r)} = \hat{\mathbf{c}} + \sum_{\ell=0}^{p_0} \hat{\mathbf{\Phi}}_\ell \mathbf{z}_{t-\ell}^{(r)} + \hat{\mathbf{C}} \boldsymbol{\varepsilon}_t^{(r)}, \quad (\text{A.18})$$

for  $t = p + 1, p + 2, \dots, T$ , with initial values  $\mathbf{z}_\ell^{(r)} = \mathbf{z}_\ell$  for  $\ell = 1, 2, \dots, p_0$ . As in Evans and Marshall (1998), we consider the lag length of  $p_0 = 12$ . Such a large lag length makes the full system VAR and JLP infeasible for the relatively small choice of  $T = 30$ , due to a large number of unknown parameters to estimate. These two approaches are therefore no longer considered here, and hence we choose the parsimonious LP approach with  $s = 1$  as a benchmark instead of the full-system VARX. In addition to being proliferated with parameters, this design differs from the previous two designs in that it is calibrated to data series that are not mean reverting, which violates the stationarity assumption. This renders the most parsimonious and DL approaches not applicable, because the regressions (1) and (A.5) are no longer well balanced. These two approaches are therefore not considered in this design. We take the non-normalized monetary policy shock  $\varepsilon_{MP,t}^{(r)} = \hat{c}_{44}\varepsilon_{4,t}^{(r)}$  as observed, where  $\hat{c}_{44}$  is the element (4,4) of the matrix  $\hat{\mathbf{C}}$ . We use the applicable competing approaches under study to estimate the impact of the monetary policy shock on federal funds rate. Each estimation approach is implemented as in the previous two sets of experiments. Table A2 reports the relative RMSE results. As can be seen from Table A2, the simulation results are in line with our main conclusions drawn from the previous two MC results, with the single exception discussed in Footnote 10.

**Table A2: MC findings for the relative RMSE of estimating IRF coefficients in design calibrated to U.S. macro dataset.**

Approach \ horizon	$T = 30$			$T = 150$			$T = 500$		
	short	medium	long	short	medium	long	short	medium	long
ADL, $s = 1$	1.94	1.10	0.63	3.50	1.84	1.23	8.56	4.68	3.16
ADL, $s = 2$	1.93	1.31	0.83	2.66	1.49	1.01	2.95	1.61	1.07
ADL, $s = N$	2.37	1.89	1.07	1.99	1.23	0.88	2.19	1.19	0.78
LP, $s = 1$ (benchmark)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
ALP, $s = 1$	0.72	0.91	1.50	0.75	0.98	1.05	0.74	0.97	0.98
LP, $s = 2$	1.45	1.55	1.08	1.01	1.22	1.16	0.97	1.00	1.01
ALP, $s = 2$	0.78	0.66	0.60	0.68	0.84	0.89	0.73	0.96	0.96
LP, $s = N$	2.78	2.39	0.86	1.21	1.98	1.69	0.91	1.02	0.97
ALP, $s = N$	1.18	0.98	0.68	0.62	0.72	0.75	0.63	0.76	<b>0.70</b>
ARDL	0.67	0.75	1.00	0.65	1.50	3.32	0.58	0.79	0.88
VARDL, $s = 2$	<b>0.65</b>	0.58	0.99	1.05	3.70	5.29	0.49	<b>0.73</b>	0.88
VARDL, $s = 3$	0.72	<b>0.44</b>	<b>0.56</b>	0.88	2.78	3.19	<b>0.48</b>	0.76	1.26
VARDL, $s = N$	1.16	0.86	0.66	<b>0.49</b>	<b>0.67</b>	<b>0.69</b>	0.54	2.65	4.67

Notes: The entries represent the RMSE of each approach relative to the benchmark of the parsimonious LP approach with  $s = 1$ . The DGP is given by VAR(12) model (A.18), which features coefficients  $\hat{\mathbf{c}}, \hat{\mathbf{\Phi}}_\ell$ , for  $\ell = 1, 2, \dots, 12$ , and  $\hat{\mathbf{C}}$  estimated based on U.S. macro dataset taken from Evans and Marshall (1998) and Jordà (2005). The lowest (best) entries are highlighted by bold fonts. The benchmark method is LP with  $s = 1$ . See Section A.4 and the notes in Table 2 for further details.