

# 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 the Monte Carlo experiments using a design calibrated to a multicountry output growth VAR. Section A.4 presents the Monte Carlo experiments using a design calibrated to a U.S. macro dataset. Section A.4 presents all the Monte Carlo experiment results re-computed using the AIC in lieu of the BIC rule.

## A.1 Modeling setup

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

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

**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 (2) where  $\mathbf{r}$  is the first element of  $\mathbf{R} = (\mathbf{r}, \mathbf{R}_{-1})$ .  $\boldsymbol{\zeta}_t = \mathbf{R}_{-1}\boldsymbol{\xi}_{-1,t}$  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,

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

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

$$b_h = \mathbf{s}'_{n,1} \Phi^h \mathbf{r}, \quad \text{for } h = 0, 1, \dots, \quad (\text{A.2})$$

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:

$$x_t = b_h \varepsilon_{t-h} + v_t, \quad \text{for } h = 0, 1, 2, \dots, h_{\max}, \quad (\text{A.3})$$

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 the notations), and it takes different forms throughout the paper including this Appendix. In the context of (A.3),  $v_t$  is given by (using the moving-average representation in (A.1)),

$$v_t = e_t + \sum_{j=0, j \neq h}^{\infty} b_j \varepsilon_{t-j}. \quad (\text{A.4})$$

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.3), denoted as  $\hat{b}_h$ , is consistent for any given  $h$ .

### A.2.2 DL

The DL approach is based on (5) where the corresponding regression error term is given by

$$v_t = e_t + \sum_{h=h_{\max}+1}^{\infty} b_h \varepsilon_{t-h}. \quad (\text{A.5})$$

$v_t$  given by (A.5) is uncorrelated with regressors,  $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-p_T}$ , in (5), 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>A.1</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 holds. When augmenting the DL regression in (5), 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.1),  $x_t$  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  $\{\zeta_{t-h}\}_{h=0}^{\infty}$ :

$$x_t = b(L)\varepsilon_t + e_t, \quad (\text{A.6})$$

where  $b(L) = \sum_{h=0}^{\infty} b_h L^h$  with  $b_h$ , for  $h = 0, 1, \dots$ , defined in (A.2). The error term  $e_t$  is covariance stationary, and from the Wold decomposition theorem (Wold, 1938), it has an MA( $\infty$ ) representation,

$$e_t = \alpha(L)\nu_t, \quad (\text{A.7})$$

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.6) 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

$$x_t = \sum_{h=1}^{\infty} \psi_h x_{t-h} + \sum_{h=0}^{\infty} \beta_h \varepsilon_{t-h} + \nu_t, \quad (\text{A.8})$$

---

<sup>A.1</sup>Similar to the LP approach, regression in (5) can also be run for  $h_{\max} = 0, 1, 2, \dots, H$ , which gives us  $H - h + 1$  different estimates of  $b_h$ . All of these estimates are consistent, but these estimators are not pursued in the current paper.

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.8) as shown in (10). 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 (1), we have the following decomposition similar to (A.6),

$$\begin{aligned}
\mathbf{z}_{st} &= \mathbf{S}'_s \mathbf{z}_t = \mathbf{S}'_s \sum_{h=0}^{\infty} \Phi^h \mathbf{R} \xi_{t-h} \\
&= \sum_{h=0}^{\infty} \mathbf{S}'_s \Phi^h \mathbf{r}_1 \varepsilon_{t-h} + \sum_{h=0}^{\infty} \mathbf{S}'_s \Phi^h \mathbf{R}_{-1} \zeta_{t-h} \\
&= \sum_{h=0}^{\infty} \mathbf{b}_{sh} \varepsilon_{t-h} + \mathbf{e}_{st}, \\
&= \mathbf{b}_s(L) \varepsilon_t + \mathbf{e}_{st}
\end{aligned} \tag{A.9}$$

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

$$\mathbf{e}_{st} = \sum_{h=0}^{\infty} \mathbf{S}'_s \Phi^h \mathbf{R}_{-1} \zeta_{t-h}, \quad \text{for } s = 2, 3, \dots, n-1 \text{ and } h = 0, 1, \dots \tag{A.10}$$

Note that  $\mathbf{e}_{st}$  is an  $s$ -dimensional vector of covariance stationary variables. Similar to (A.7),  $\mathbf{e}_{st}$  has an MA( $\infty$ ) representation,

$$\mathbf{e}_{st} = \mathbf{A}_s(L) \boldsymbol{\nu}_{st}, \tag{A.11}$$

where  $\boldsymbol{\nu}_{st}$  has a zero mean with a constant variance,  $\boldsymbol{\Omega}_{uk}$ . It is serially uncorrelated and independent with  $\varepsilon_{t'}$  for all  $t, t'$ , and the coefficients of  $\mathbf{A}_s(L) = \mathbf{I}_s - \sum_{h=1}^{\infty} \mathbf{A}_{sh} L^h$  decay at an exponential rate. As in the ARDL approach,  $\mathbf{e}_{st}$  is required to have an invertible vector MA( $\infty$ ) representation so that it 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.9) by  $\Psi_s(L)$  from the left and making use of the representation in (A.11), we obtain similar to (A.8),

$$\Psi_s(L) \mathbf{z}_{st} = \beta_s(L) \varepsilon_t + \nu_{st}, \quad (\text{A.12})$$

where  $\nu_{st} = \Psi_s(L) \mathbf{e}_{st}$ ,  $\beta_s(L) = \mathbf{A}_s(L) \mathbf{b}_s(L)$ , and the coefficients of  $\beta_s(L)$  decay at an exponential rate. The truncated version of (A.12) is given by (11), 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 log first difference 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,

$$\mathbf{z}_t = \mathbf{c} + \Phi \mathbf{z}_{t-1} + \mathbf{u}_t, \quad (\text{A.13})$$

using the sample 1980:Q3 - 2015:Q2 (140 quarterly observations). Let  $\hat{\mathbf{c}}$  and  $\hat{\Phi}$  respectively denote the LS estimates of  $\mathbf{c}$  and  $\Phi$  in (A.13), and let  $\hat{\Sigma}$  be the estimate of the variance-covariance matrix of  $\mathbf{u}_t$  in (A.13). We then generate simulated data  $\mathbf{z}_t^{(r)}$  based on

$$\mathbf{z}_t^{(r)} = \hat{\mathbf{c}} + \hat{\Phi} \mathbf{z}_{t-1}^{(r)} + \mathbf{u}_t^{(r)}, \quad (\text{A.14})$$

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

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

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 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	1.43	3.53	9.92	2.33	7.00	23.06	2.50	7.55	24.04
DL	2.29	5.48	12.68	1.83	5.06	16.88	1.78	5.00	16.16
ADL, $s = 1$	2.74	6.97	16.98	1.86	5.17	17.34	1.79	5.02	16.26
ADL, $s = 2$	3.56	9.05	21.28	1.91	5.33	17.82	1.80	5.04	16.35
ADL, $s = N$	>100	>100	>100	2.01	5.62	18.85	1.82	5.11	16.55
LP, $s = 1$	1.39	3.87	10.88	2.03	6.86	23.28	2.11	7.25	23.93
ALP, $s = 1$	1.19	3.43	12.85	1.66	4.77	16.97	1.71	4.90	16.14
LP, $s = 2$	1.36	4.19	11.80	1.78	6.88	23.61	1.85	7.15	23.99
ALP, $s = 2$	1.18	3.94	16.32	1.46	4.79	17.39	1.47	4.81	16.23
LP, $s = N$	1.75	6.61	20.85	1.66	7.18	24.78	1.64	7.17	24.27
ALP, $s = N$	1.52	7.57	>100	1.36	4.97	18.37	1.34	4.81	16.42
ARDL	1.12	0.81	0.33	1.66	1.95	1.08	2.19	3.96	2.78
VARDL, $s = 2$	0.82	<b>0.37</b>	<b>0.26</b>	1.24	1.03	1.06	1.37	2.10	2.25
VARDL, $s = 3$	<b>0.82</b>	0.46	0.32	1.15	<b>0.85</b>	<b>0.70</b>	1.43	1.22	1.06
VARDL, $s = N$	1.03	1.05	1.05	<b>1.00</b>	1.00	1.00	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
JLP	1.35	2.08	5.46	1.53	5.47	18.77	1.48	6.21	21.25

Notes: The entries represent the RMSE of each approach relative to the VARX benchmark. The DGP is given by VAR(1) model in (A.14), which features coefficients in  $\hat{\mathbf{c}}$  and  $\hat{\mathbf{\Phi}}$  estimated using real GDP growth data on  $n = 10$  economies over the sample period 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 U.S. 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>A.2</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

<sup>A.2</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).

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 above ordering), we estimate a VAR( $p_0$ ) model

$$\mathbf{z}_t = \mathbf{c} + \sum_{\ell=0}^{p_0} \hat{\Phi}_\ell \mathbf{z}_{t-\ell} + \mathbf{u}_t, \quad (\text{A.15})$$

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  $\hat{\mathbf{c}}$ ,  $\hat{\Phi}_\ell$ ,  $\hat{\mathbf{u}}_t$  denote the LS reduced-form estimates, and let  $\hat{\Sigma}$  denote the estimated variance-covariance matrix of error term. Note that federal funds rate is ordered as the fourth after the variables (i)-(iii). After obtaining the lower triangular Cholesky factorization of  $\hat{\Sigma} = \hat{\mathbf{C}}\hat{\mathbf{C}}'$ , 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

$$\mathbf{z}_t^{(r)} = \hat{\mathbf{c}} + \sum_{\ell=0}^{p_0} \hat{\Phi}_\ell \mathbf{z}_{t-\ell}^{(r)} + \hat{\mathbf{C}}\boldsymbol{\varepsilon}_t^{(r)}, \quad (\text{A.16})$$

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 VAR( $p_0$ ) in  $\mathbf{z}_t$  and JLP infeasible for the relatively short time span 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 VARX model. 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 in (5) and (A.3) 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 largely confirm our main conclusions drawn from the previous two MC results.

**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, $s = 1$	1.95	1.08	0.63	3.57	1.86	1.24	8.24	4.49	3.15
ADL, $s = 2$	2.00	1.35	0.84	2.67	1.48	1.02	2.92	1.57	1.10
ADL, $s = N$	2.38	1.90	1.06	2.00	1.22	0.87	2.14	1.17	0.80
LP, $s = 1$ (benchmark)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
ALP, $s = 1$	0.71	0.88	1.44	0.75	0.99	1.07	0.75	0.96	1.01
LP, $s = 2$	1.47	1.55	1.03	1.02	1.22	1.17	0.97	1.00	1.01
ALP, $s = 2$	0.79	0.65	0.59	0.69	0.86	0.90	0.73	0.96	0.99
LP, $s = N$	2.78	2.35	0.84	1.24	1.98	1.70	0.91	1.02	0.96
ALP, $s = N$	1.18	0.97	0.67	0.63	0.72	0.74	0.63	0.76	<b>0.72</b>
ARDL	0.67	0.76	1.04	0.65	1.43	3.27	0.58	0.78	0.90
VARDL, $s = 2$	<b>0.66</b>	0.58	0.99	1.03	3.60	5.26	0.49	<b>0.74</b>	0.91
VARDL, $s = 3$	0.73	<b>0.44</b>	<b>0.54</b>	0.86	2.71	3.19	<b>0.47</b>	0.76	1.29
VARDL, $s = N$	1.17	0.85	0.68	<b>0.51</b>	<b>0.68</b>	<b>0.70</b>	0.55	2.67	4.89

Notes: The entries represent the RMSE of each approach relative to the benchmark of the LP approach with  $s = 1$ . The DGP is given by VAR(12) model in (A.16), 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. See Section A.4 and the notes in Table 2 for further details.



## A.5 Comparison of Iterative methods using AIC and BIC

**Table A3: Monte Carlo findings for the relative RMSE of estimating IRF coefficients with iterative methods using AIC and BIC in DGP1**

Approach \ horizon	$n = 6, T = 30$			$n = 6, T = 150$			$n = 6, T = 500$		
	short	medium	long	short	medium	long	short	medium	long
ARDL, BIC	0.87	0.69	0.31	1.10	1.12	0.69	1.21	1.23	0.81
ARDL, AIC	0.99	0.91	0.42	1.19	1.44	0.99	1.27	1.53	1.66
VARDL, $s = 2$ , BIC	0.58	0.37	0.33	1.15	1.25	0.65	1.39	1.60	0.78
VARDL, $s = 2$ , AIC	0.62	0.38	0.33	1.21	1.35	0.77	1.27	1.43	1.07
VARDL, $s = 3$ , BIC	0.73	0.66	0.45	1.08	1.20	1.19	1.20	1.29	1.23
VARDL, $s = 3$ , AIC	0.77	0.68	0.46	1.27	1.69	1.58	1.35	2.13	2.42
VARDL, $s = n$ , BIC	1.01	1.01	1.01	1.00	1.00	1.00	1.00	1.00	1.00
VARDL, $s = n$ , AIC	1.05	1.07	1.07	1.02	1.02	1.02	1.03	1.01	1.01
Approach \ horizon	$n = 12, T = 30$			$n = 12, T = 150$			$n = 12, T = 500$		
	short	medium	long	short	medium	long	short	medium	long
ARDL, BIC	0.63	0.30	0.04	0.96	0.50	0.16	1.02	0.90	0.54
ARDL, AIC	0.72	0.55	0.07	1.17	2.78	0.61	1.23	3.48	3.19
VARDL, $s = 2$ , BIC	0.37	0.10	0.04	0.63	0.39	0.45	0.80	0.67	0.92
VARDL, $s = 2$ , AIC	0.42	0.10	0.04	0.95	1.76	1.05	1.10	2.63	2.18
VARDL, $s = 3$ , BIC	0.40	0.12	0.05	0.59	0.35	0.43	0.63	0.43	0.60
VARDL, $s = 3$ , AIC	0.45	0.13	0.06	0.88	1.09	0.72	1.00	1.38	1.06
VARDL, $s = n$ , BIC	1.10	1.31	2.34	1.00	1.00	1.00	1.00	1.00	1.00
VARDL, $s = n$ , AIC	1.15	1.46	2.63	1.02	1.05	1.01	1.01	1.03	1.01

Notes: Entries represent the ratio of the average RMSE of estimating  $b_h$  from each approach to that of the benchmark 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 (13), where the reduced form coefficient matrix ( $\Phi$ ) is generated randomly from (14) for each  $n$  and kept fixed across replications. Descriptions of each approach are provided in Section 2. Section 3 provides a full description of MC experiments.

**Table A4: MC findings for the relative RMSE of estimating IRF coefficients with iterative methods using AIC and BIC in design calibrated to international output dataset**

Approach \ horizon	$T = 30$			$T = 150$			$T = 500$		
	short	medium	long	short	medium	long	short	medium	long
ARDL, BIC	1.12	0.81	0.33	1.66	1.95	1.08	2.19	3.96	2.78
ARDL, AIC	1.21	1.16	0.46	1.66	2.84	1.75	1.81	3.43	4.46
VARDL, $s = 2$ , BIC	0.82	0.37	0.26	1.24	1.03	1.06	1.37	2.10	2.25
VARDL, $s = 2$ , AIC	0.84	0.39	0.30	1.41	2.10	1.51	1.48	2.79	2.47
VARDL, $s = 3$ , BIC	0.82	0.46	0.32	1.15	0.85	0.70	1.43	1.22	1.06
VARDL, $s = 3$ , AIC	0.85	0.48	0.36	1.33	1.74	1.41	1.34	1.92	1.89
VARDL, $s = n$ , BIC	1.03	1.05	1.05	1.00	1.00	1.00	1.00	1.00	1.00
VARDL, $s = n$ , AIC	1.08	1.15	1.17	1.01	1.01	1.01	1.01	1.01	1.00

Notes: The entries represent the RMSE of each approach relative to the VARX benchmark. The DGP is given by VAR(1) model in (A.14), 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. See Section A.3 and the notes in Table 2 for further details.

**Table A5: MC findings for the relative RMSE of estimating IRF coefficients with iterative methods using AIC and BIC 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
ARDL, BIC	0.67	0.76	1.04	0.65	1.43	3.27	0.58	0.78	0.90
ARDL, AIC	0.69	0.76	0.98	0.64	1.03	1.99	0.58	0.79	0.90
VARDL, $s = 2$ , BIC	0.66	0.58	0.99	1.03	3.60	5.26	0.49	0.74	0.91
VARDL, $s = 2$ , AIC	0.60	0.62	1.05	0.61	0.94	1.69	0.49	0.73	0.84
VARDL, $s = 3$ , BIC	0.73	0.44	0.54	0.86	2.71	3.19	0.47	0.76	1.29
VARDL, $s = 3$ , AIC	0.61	0.45	0.57	0.59	0.84	1.19	0.47	0.69	0.83
VARDL, $s = n$ , BIC	1.17	0.85	0.68	0.51	0.68	0.70	0.55	2.67	4.89
VARDL, $s = n$ , AIC	0.88	0.80	0.68	0.51	0.67	0.76	0.32	0.48	0.57

Notes: The entries represent the RMSE of each approach relative to the benchmark of the LP approach with  $s = 1$ . The DGP is given by VAR(12) model (A.16), 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). See Section A.4 and the notes in Table 2 for further details.