FEDERAL RESERVE BANK *of* ATLANTA

Information Criteria for Impulse Response Function Matching Estimation of DSGE Models

Alastair Hall, Atsushi Inoue,
James M. Nason, and Barbara Rossi

Working Paper 2007-10a
October 2008

WORKING PAPER SERIES

FEDERAL RESERVE BANK *of* ATLANTA | WORKING PAPER SERIES

Information Criteria for Impulse Response Function Matching Estimation of DSGE Models

Alastair Hall, Atsushi Inoue, James M. Nason, and Barbara Rossi

Working Paper 2007-10a
October 2008

Abstract: We propose a new information criterion for impulse response function matching estimators (IRFMEs) of the structural parameters of dynamic stochastic general equilibrium (DSGE) macroeconomic models. An advantage of our procedure is that it allows researchers to select the impulse responses that are most informative about DSGE model parameters and ignore the rest. The idea of tossing out superfluous impulse responses motivates our Redundant Impulse Response Selection Criterion (RIRSC). The RIRSC is general enough to apply to impulse responses estimated by VARs, local projections, and simulation methods. We show that our criterion significantly affects estimates and inference about key parameters of two well-known New Keynesian DSGE models. Monte Carlo evidence indicates that the RIRSC yields gains in terms of finite sample bias as well as offering tests statistics whose behavior is better approximated by first order asymptotic theory. Thus, RIRSC improves on existing methods used to implement IRFMEs.

JEL classification: C32, E47, C52, C53

Key words: impulse response function, matching estimator, redundant selection criterion

---

The authors thank Craig Burnside for sharing his codes and for many useful suggestions along the way. They also thank L. Christiano for making the Altig, Christiano, Eichenbaum, and Linde (2005) codes available in his webpage. We gratefully acknowledge comments from Jean Boivin, Jeff Campbell, Peter R. Hansen, Oscar Jorda, Jesper Linde, Sophocles Mavroedis, Frank Schorfheide, and seminar participants at the 2006 EC^2 Conference, the EMSG at Duke University, the UBC Macro Lunch Group, the 2007 Conference on the Macroeconomics of Technology Shocks in Waterloo, the 2007 ESSM, the 2007 SED Conference, Kyoto University, Osaka University, Brown University, and University of Cincinnati. Atsushi Inoue acknowledges financial support by the SSHRC under project 410-2007-1975. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Alastair Hall, Manchester University, Economics Department, School of Social Sciences, Manchester, United Kingdom M13 9PL, alastair.hall@manchester.ac.uk; Atsushi Inoue, Department of Agricultural and Resource Economics, North Carolina State University, Campus Box 8109, Raleigh, NC 27695-8109, 919-515-5969, atsushi@unity.ncsu.edu; James M. Nason, Federal Reserve Bank of Atlanta, Research Department, 1000 Peachtree Street, N.E., Atlanta, GA 30309, jim.nason@atl.frb.org; or Barbara Rossi, Duke University, Department of Economics, 204 Social Science Building, Durham, NC 27708, brossi@econ.duke.edu.

1 Introduction

Since the seminal work of Rotemberg and Woodford (1997), there has been increasing use of impulse response function matching to estimate parameters of dynamic stochastic general equilibrium (DSGE) models. Impulse response function matching is a limited information approach that minimizes the distance between sample and DSGE model generated impulse responses. Those applying this estimator to DSGE models include, among others, Christiano, Eichenbaum and Evans (2005), Altig, Christiano, Eichenbaum, and Lindé (2005), Iacoviello (2005), Jordà and Kozicki (2007), DiCecio (2005), Boivin and Giannoni (2006), Uribe and Yue (2006) DiCecio and Nelson (2007), and Dupor, Han, and Tsai (2007). Despite the widespread use of impulse response function (IRF) matching, only ad hoc standards have been available to choose which IRFs and how many of their lags to match.

This paper presents a new method to improve impulse response function matching estimators (IRFMEs). We develop a criterion, the Redundant Impulse Response Selection Criterion (RIRSC), to select IRFs and their dimension for IRFME that has the flexibility to work in many different estimation environments. The RIRSC provides the means to select which IRFs to match and the number of lags of each IRF to include. Our method has solid econometric foundations, applies to several different classes of IRFMEs, is easy to implement, and offers improved statistical inference for IRFMEs.

Our criterion answers two questions. The first is: "Can the performance of the matching estimator be improved by selecting the IRF that are most informative about DSGE model parameters?". Since the linearized approximate solutions of many DSGE models take a state-space form with cross-equation restrictions tied to a relatively small number of parameters, some linear dependence is imposed on the IRFs. It seems reasonable to conjecture that not accounting for this linear dependence can lead to small sample bias in and incorrect inference about DSGE model parameters. Our proposed criterion selects IRFs for the matching estimator that appears to perform well in small samples. The key idea is that our criterion holds onto the relevant IRFs for estimation, while discarding the irrelevant IRFs.

The second question is: "How many lags in the relevant IRFs should be matched?". So far the literature has proceeded with ad-hoc rules for choosing which IRF elements to match. For example, Christiano, Eichenbaum and Evans (CEE) and Altig, Christiano, Eichenbaum, and Lindé

(ACEL) match a pre-specified number of IRF lags. However, this procedure fails to recognize that even impulse responses that are zero may well be informative about the parameters of interest (e.g. restrictions imposed via long-run identification) and that some impulse responses might contain more information on the parameters of interest than other IRFs. Our criterion exploits information in the data to choose the IRF lag length, as well as which lags to match. Thus, we provide a criterion to choose the IRF lag length, as well as which IRFs to match. This gives applied researchers a theoretically sound criterion for selecting the dimension of the relevant IRFs to match that is easy to use and addresses an important and, up to this point, neglected issue.

This paper develops a criterion to select IRFs and their dimension for IRFME that has the flexibility to work in many different estimation environments. Our criterion (*i*) provides the means to separate the relevant from the redundant IRFs and to choose the appropriate lag length for the relevant subset of IRFs; this can be especially useful for the researcher who does not have strong opinions on which IRFs to match (e.g. say, between an IRF tied to an identified monetary policy shock or to an IRF generated by a technology shock); (*ii*) can be used when IRFs are identified by short-run, long-run, or a combination of both types of restrictions;[1] (*iii*) does not depend on having access to an optimal weighting matrix; (*iv*) works in the presence of calibrated parameters; (*v*) improves inference; and (*vi*) is easy to implement; it only requires an estimate of the variances of the DSGE parameters given a particular choice of the number of lags in the IRFs, which is already available because it is computed to conduct inference.[2]

Our selection criterion is applied to the DSGE models of CEE and ACEL. We often obtain point

[1]Sign restrictions cannot be used for identification in our framework because only intervals of IRFs are produced, not point estimates.

[2]The IRF matching estimator is a limited information approach to estimation of DSGE models. Limited information estimators do not rely on a full model specification. Thus, the IRF matching estimator can ignore the full set of predictions of which the DSGE model is capable, and be more robust to misspecification. For example, Christiano, Eichenbaum, and Evans (2005) estimate DSGE model parameters by matching the sample and theoretical responses of inflation and other macro variables only to an identified monetary policy shock. This contrasts with full information approaches in which the likelihood expresses the complete set of predictions inherent in the DSGE model. Although the solutions of the linearized DSGE models we study have well defined likelihoods, we adopt the limited information motivation of the IRF matching estimator to better understand its properties and behavior. Besides IRF matching and maximum likelihood, simulation estimators in the frequentist domain are used to estimate DSGE models. A useful survey of simulation estimators is Gourieroux and Monfort (1997).

estimates that are little changed from those CEE and ACEL report. Nonetheless, the RIRSC yields economically important changes in several key parameters that leads to inference with strikingly different conclusions than those of CEE and ACEL. Monte Carlo exercises indicate that the small sample bias of IRFMEs is mitigated by RIRSC compared to using a fixed lag length. Thus, the RIRSC should be attractive to analysts at central banks and other institutions conducting policy evaluation with DSGE models, as well as academic researchers testing newly developed DSGE models.

The criterion that we propose is connected to several strands of the literature that estimate DSGE models. Rotemberg and Woodford (1996), CEE (2005), and ACEL (2005) employ IRFMEs that minimize the difference between sample and theoretical IRFs using a non-optimal weighting matrix to which our selection criterion can be applied. Jordà and Kozicki (2007) show that our criterion meshes with an IRFME estimator based on local projections and an optimal weighting matrix. Note that our criterion is applicable whether the weighting matrix is efficient or not. Finally, we show that our criterion can be an element of the Sims (1989) and Cogley and Nason (1995) simulation estimator.[3]

The paper is organized as follows. Section 2 presents our new criterion for the IRFME in the leading VAR case and discusses the assumptions that guarantee its validity. In section 3, we provide a clarifying example. The projection and simulation-based estimators are studied in section 4. Sections 5 and 6 present the empirical results and Monte Carlo analyses. Section 7 concludes. All technical proofs and assumptions are collected in the Appendix.

2 The VAR-based IRF matching estimator

In this section, we consider the leading case in which the researcher is interested in estimating the parameters of a DSGE model by using a VAR-based IRFME. This estimator is obtained by

[3]The method proposed in this paper is further related to Andersen and Sorensen (1996) and Hall, Inoue, Jana, and Shin (2007). The former paper shows (in a different context than ours) that minimum distance and GMM estimators do not work well in finite samples when the number of overidentifying restriction is large. The latter set of co-authors propose a "relevant moment selection criterion" based on entropy arguments that is useful for solving that problem. They show that the limiting distribution of a GMM estimator can be written in terms of long-run canonical correlations between the moment function and the true score vector. We adapt their concept, although our focus is on minimum distance estimators, which is appropriate for the IRF matching problem.

minimizing the distance between the sample IRFs obtained by fitting a VAR to the actual data and the theoretical IRFs generated by the DSGE model. The sample and the theoretical IRFs are identified by restrictions implied by the DSGE model. This requires we assume that the DSGE model admits a structural VAR representation, and so that, the sample IRFs are informative for the DSGE model parameters. We are interested in the VAR:

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + ... + \Phi_{p_0} y_{t-p_0} + \varepsilon_t, \tag{1}$$

where y_t is $n_y \times 1$, $t = 1, 2, ..., T$, and ε_t is white noise with zero mean and variance Σ_ε. The population VAR lag order, p_0, can be either finite or infinite. For (1) to have an infinite order Vector Moving Average (VMA) representation and IRFs, we make the following standard assumption:

Assumption 1 *In equation (1), $\Phi(L) = I_{n_y} - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_{p_0} L^{p_0}$ is invertible, where L is the lag operator and I_{n_y} is the $(n_y \times n_y)$ identity matrix.*

Let $\gamma_{i,j,\tau}$ denote IRFs of each variable $y_{i,t+\tau}$ to a structural shock $\varepsilon_{j,t}$ at horizon τ, where $i, j = 1, ..., n_y$, $\tau = 1, ..., H$, and H is the maximum horizon of the IRFs.[4] Let $\underline{\gamma}_\tau$ be a $\left(n_y^2 \times 1\right)$ vector that collects the population IRFs at a particular horizon τ :

$$\underline{\gamma}_\tau = (\overbrace{\gamma_{1,1,\tau},\ \gamma_{1,2,\tau},\ ...,\ \gamma_{1,n_y,\tau}}^{i=1},\ \overbrace{\gamma_{2,1,\tau},\ \gamma_{2,2,\tau},\ ...,\ \gamma_{2,n_y,\tau}}^{i=2},\ ...,\ \overbrace{\gamma_{n_y,1,\tau},\ ...,\ \gamma_{n_y,n_y,\tau}}^{i=n_y})'$$

The population IRFs at horizons $\tau = 1, 2, ..., H$ will be collected in a $\left(n_y^2 H \times 1\right)$ vector γ_H :

$$\gamma_H = \left(\underline{\gamma}_1',\ \underline{\gamma}_2',\ ...,\ \underline{\gamma}_H'\right)'.$$

[4]For simplicity we assume that the dimension of y_t, n_y, and that of ε_t, n_ε, are equal. However, n_y can be greater than n_ε. For example, suppose that a tri-variate VAR(2) with two shocks is fitted to the actual data in order to estimate eight DSGE model parameters using an optimal weighting matrix. When $H = 2$, suppose the 18×18 asymptotic covariance matrix of all possible IRFs is singular with rank of 12. Suppose the Moore-Penrose generalized inverse of the asymptotic covariance matrix is used as the weighting matrix and that the 18×8 Jacobian matrix of the theoretical IRF has rank of 8, which is implicit in assumption (1). In this case, the eight DSGE model parameters will be identified. If instead the tri-variate VAR(2) is driven only by one shock, the asymptotic covariance matrix has rank six. As a result, the inverse of the asymptotic covariance matrix of the IRFME is singular and the DSGE model parameters will not be identified. The dimension of shocks matters for identification but not necessarily relative to the dimension of y_t. Provided rank conditions are satisfied, adding a redundant vector of variables to the VAR system, while holding the number of shocks fixed, will not violate the identification condition. However, the finite-sample performance of the IRFME estimator can deteriorate.

Let the model's parameters (referred to as deep parameters) be collected in a $(q \times 1)$ vector θ, $\theta \in \Theta$, and the theoretical IRFs up to horizon H be denoted by $g_H(\theta)$.[5] The validity of IRFMEs requires that theoretical IRFs equal population IRFs:

Assumption 2 $g_\infty(\theta_0) = \gamma_\infty$ *for some* $\theta_0 \in \Theta$.

Let c be a $n_y^2 H \times 1$ selection vector that indicates which elements of the candidate impulse responses are included in estimation. We use c to index functions of impulse responses, that is, $\gamma_H(c)$ and $g_H(\cdot; c)$, and, for notational simplicity, we will drop the subscript H. If $c_j = 1$ then the j^{th} element of γ_H is included in $\gamma(c)$, and $c_j = 0$ implies this element is excluded. For example, if only impulse responses up to horizons $h < H$ are selected, $c = [1_{1\times n_y^2 h} \;\; 0_{1\times n_y^2(H-h)}]'$ where $1_{m\times n}$ and $0_{m\times n}$ denote $m \times n$ matrices of ones and zeros, respectively. If the impulse responses of the second element of y_t to the first element of ε_t are used, $c = I_{n_y \times 1} \otimes [0_{1\times n_y} \;\; 0 \;\; 1 \;\; 0_{1\times(n_y-2)} \;\; 0_{1\times(n_y-2)n_y}]'$. Note that $|c| = c'c$ equals the number of elements in $\gamma(c)$. The set of all possible selection vectors is denoted by C_H, that is

$$C_H = \left\{ c \in \Re^{n_y^2 H};\; c_j = 0, 1, \text{ for } j = 1, 2, \ldots n_y^2 H, \text{ and } c = (c_1, \ldots c_{n_y^2 H})',\; |c| \geq 1 \right\}.$$

The IRF Matching Estimator (IRFME) is a classical minimum distance estimator based on Assumption 2:

$$\hat{\theta}(c) = \arg\min_{\theta\in\Theta} [\widehat{\gamma}_T(c) - g(\theta; c)]' \widehat{W}_T(c) [\widehat{\gamma}_T(c) - g(\theta; c)], \tag{2}$$

where $\hat{\gamma}_T(c)$ is an estimate of $\gamma(c)$ and $\widehat{W}_T(c)$ is a weighting matrix. $\widehat{W}_T(c)$ could be the inverse of the covariance matrix of the IRFs $\hat{\gamma}_T(c)$ or, as often found in practice, a restricted version of this matrix that has zeros everywhere except along its diagonal. In general, $\widehat{W}_T(c)$ can be readily obtained from standard package procedures that compute IRF standard error bands.

In order to implement the IRFME in practice, the researcher has to choose which impulse responses to use in (2). Our contribution to the existing literature is to provide statistical criteria to choose c. The criterion that we propose allows the researcher to avoid using the IRFs that contain

[5]Theoretical models may also contain additional parameters whose values are not estimated but calibrated. We denote such parameters by ϕ. Let $g_H(\theta, \phi)$ denote the mapping between the parameters of the DSGE model and its theoretical IRFs. Since the calibrated parameters do not play any role in our analysis, in order to simplify notation we ignore ϕ and write $g_H(\theta)$ rather than $g_H(\theta, \phi)$.

only redundant information and to identify the "relevant horizon" of the IRFs. Since IRFs that do not contain additional information only add noise to the estimation of the deep parameters, these IRFs should be eliminated. The following definitions formalize these concepts: Let $V_\theta(c)$ denote the asymptotic covariance matrix of the IRMFE based on selection vector c. We denote the set of selection vectors that are asymptotically efficient relative to the candidate set by

$$C_{\mathcal{E},H} \;\equiv\; \{c;\; c \in C_H;\; V_\theta(\iota_{n_y^2 H}) = V_\theta(c)\},$$

where ι_n is an $n \times 1$ vector of ones, and the subset of $C_{\mathcal{E},H}$ containing vectors of minimum length by

$$C_{min,H} \;\equiv\; \{c;\, c \in C_{\mathcal{E},H},\, |c| \leq |\bar{c}| \text{ for all } \bar{c} \in C_{\mathcal{E},H}\}\,.$$

If $c \in C_{min,H}$ we call $\gamma(c)$ a relevant IRF. Let's denote the selection vector of the relevant IRFs with c_r. We assume that there is a unique relevant IRF in $C_{min,H}$.

Assumption 3 $C_{min,H} = \{c_r\}$,

Our goal is to identify the fewest number of IRFs that guarantee that, asymptotically, the covariance matrix of the IRFME estimator is as small as possible.

We define the redundant impulse responses selection criterion by

$$RIRSC(c) \;=\; \ln(|\hat{V}_T(c)|) + \kappa(|c|, p_T, T), \tag{3}$$

where p_T is the lag order of the VAR fitted to (1) that can depend on the sample size, and $\kappa(|c|, p_T, T)$ is a deterministic penalty that is an increasing function of the number of impulse responses.[6]

We select impulse response functions by minimizing the criterion (3):

$$\hat{c}_T \;=\; \arg\min_{c \in C_H} RIRSC(c). \tag{4}$$

Note that we allow for some selection vectors for which the parameters are not identified provided Assumption 3 holds. We will show that $\hat{c}_T \xrightarrow{p} c_r$ and $\hat{c}_T \xrightarrow{a.s.} c_r$. Let $C_{I,H}$ denote the set of selection vectors for which the deep parameter θ is identified and $C_{NI,H}$ denote the set of c for which the deep parameter θ is not identified. For weak consistency, we assume

[6] It can be shown that our criterion can be given a canonical correlations interpretation along the lines of Hall, Inoue, Jana and Shin (2007).

Assumption 4 *(a)* $g : \Theta \times C_H \to \Re^{|c|}$ *is twice continuously differentiable in* θ *in a neighborhood of* θ_0 *for every* $c \in C_H$ *and* H*.*

(b) $C_H = C_{I,H} \cup C_{NI,H}$*. For all sufficiently large* H*,* $C_{I,H}$ *is non-empty and satisfies*

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V_\theta(c)) \tag{5}$$

for every $c \in C_{I,H}$*.*

(c) For sufficiently large H *so that* $C_{I,H} \neq \phi$*, there is an estimator* $\hat{V}_{\theta,T}(c)$ *such that* $|\hat{V}_{\theta,T}(c)| = |V_\theta(c)| + O_p(p_T^2 T^{-1/2})$ *uniformly in* $c \in C_{I,H}$*. For any* $c \in C_{NI,H}$*,* $|\hat{V}_{\theta,T}(c)| \xrightarrow{p} \infty$*.*

(d) For every $c \in C_H$*,* $\hat{W}_T(c)$ *is a sequence of positive semi-definite matrices and satisfies* $\hat{W}_T(c) = W(c) + O_p(T^{-1/2})$*, where* $W(c)$ *is positive definite.*

(e) $H = kp_T$ *where* $k \geq 1$ *is some constant.*

Remark 1. Assumption 4(b) requires that if θ is identified then $\hat{\theta}_T(c)$ is asymptotically normal, which follows from more primitive assumptions for classical minimum distance estimators (e.g., Newey and McFadden, 1994, Theorem 3.2). $C_{NI,H}$ is the set of c for which the deep parameter θ is not identified. If it is unidentified or weakly identified, the asymptotic covariance matrix $V_\theta(c)$ is not well-defined and Assumption 4(c) assumes that the asymptotic covariance matrix estimator $\hat{V}_\theta(c)$ diverges, which is expected because the Jacobian of $g(\theta_0; c)$ is rank-deficient.

Remark 2. Assumption 4(e) requires that the maximum horizon H diverges at the same rate as the lag order p_T in case $p_T \to \infty$.

We must also impose certain "identification" restrictions involving the deterministic penalty term.

Assumption 5 $\underset{T\to+\infty}{lim} (T^{\frac{1}{2}}/p_T)(\kappa(|c_1|, p_T, T) - \kappa(|c_2|, p_T, T)) = +\infty$ *for any sequences of* $c_{1,H}, c_{2,H} \in C_H$ *such that* $|c_{1,H}| > |c_{2,H}|$*, and* $\lim_{T\to+\infty} \kappa(|c|, p_T, T) = 0$ *for any sequence of* $c_H \in C_H$*.*

Remark 3. The SIC-type penalty term,

$$\kappa(|c|, p_T, T) = \begin{cases} |c|\frac{\ln(\sqrt{T})}{\sqrt{T}} & \text{for } p_T = p_0 < \infty \\ |c|\frac{\ln(\sqrt{T}/p_T)}{\sqrt{T}/p_T} & \text{for } p_T \to \infty \end{cases} \tag{6}$$

satisfies Assumption 5 whereas the AIC-type penalty term,

$$\kappa(|c|, p_T, T) = \begin{cases} \frac{2|c|}{\sqrt{T}} & \text{for } p_T = p_0 < \infty \\ \frac{2|c|}{\sqrt{T}/p_T} & \text{for } p_T \to \infty \end{cases} \tag{7}$$

does not.

We show that our criterion is weakly consistent in the following theorem:

Theorem 1 (Weakly consistent IRFs selection (VAR case)) *Let the structural model have a VAR(p_0) representation (1), and the estimator of the deep parameters be defined as (2), where c is chosen by (4). Suppose that Assumptions 1, 2, 3, 4 and 5 hold. Then* $\hat{c}_T \xrightarrow{p} c_r$.

By focusing on the case in which $p_0 < \infty$, our results can be strengthened to strong consistency results. The strong consistency result may shed light on explaining the difference between the SIC-type and Hannan-Quinn-type information criteria, as discussed below.

Assumption 6 $\kappa(|c|, p_T, T) = k(|c|)m_T$ *where* $k(\cdot)$ *is strictly increasing,* $m_T \to \infty$ *as* $T \to \infty$ *with* $m_T = o(T^{1/2})$ *and either: (i)* $\liminf_{T\to\infty} T^{1/2}m_T/(lnlnT)^{1/2} = \mu$ *where* $z < \mu < \infty$ *and z is a positive constant that is defined in Appendix B; or (ii)* $\liminf_{T\to\infty} T^{1/2}m_T/(lnlnT)^{1/2} = +\infty$.

Theorem 2 (Strongly consistent IRFs selection (VAR case)) *Suppose that Assumptions 1, 2, 3, 4, 6, 7 and 8 hold where Assumptions 7 and 8 are presented in Appendix A. Let* $\hat{W}_T(c) = \hat{V}_T^{-1}(c)$ *and* $p_T = p_0 < \infty$. *Then* $\hat{c}_T \xrightarrow{a.s.} c_r$.

Remark 4. Theorem 2 establishes conditions under which $\hat{c}_T$ is strongly consistent for c_r. It can be seen that the conditions on the penalty term are necessarily satisfied if $\kappa(|c|, p, T) = (|c| - q)ln[T^{1/2}]/T^{1/2}$, which is the penalty term associated with the Schwarz information criterion. However, the conditions are not necessarily satisfied if $\kappa(|c|, p, T) = (|c|-q)ln[ln\left(T^{1/2}\right)]/T^{1/2}$, which is the penalty term associated with the Hannan and Quinn information criterion. In the latter case, if selection is over all possibilities then strong consistency requires that $z = 2^{1/2}(\omega_\xi(c_r) + \omega_\xi(c))$ for all $c \in C_{\mathcal{E},H}$ where $\omega_\xi(c)$ is defined in Appendix B. Notice that if this condition fails for some $\bar{c} \in C_{\mathcal{E},H}$ then one of two scenarios unfolds: if $\mu = z$ then the assignment is random between c_r and $\bar{c}$; if $\mu < z$ then $\hat{c}_T \rightarrow_{a.s.} \bar{c}$ and so more moments are included than is necessary to achieve the minimum variance. The data dependence of the condition governing these outcomes makes this choice of penalty term unattractive.

Remark 5. It is interesting to contrast the conditions on the penalty term for the case considered here in which the order of convergence of $\hat{V}_{\theta,T}(c)$ to $V_\theta(c)$is $T^{-1/2}$. For weak consistency, it is only necessary that $m_T \rightarrow \infty$ and $m_T = o(T^{1/2})$. Given Remark 4, the strong consistency results suggest the use of a penalty term for which $\liminf_{T\rightarrow\infty} T^{1/2} m_T/(lnlnT)^{1/2}$diverges. Theorem 2 therefore provides more guidance on the choice of penalty term than the corresponding weak consistency result.

Remark 6. Theorem 2 relies crucially on Assumptions 7 and 8, which are high level assumptions that guarantee approximations by the law of iterated logarithms. For simplicity, Theorem 2 also relies on the use of the optimal weighting matrix, which however is not crucial. Theorem 2 would still hold with any positive definite weighting matrix.

Theorems 1 and 2 describe the asymptotic behavior of a criterion, the RIRSC, that considers all possible combinations of IRFs. Applied researchers, however, often impose an ad hoc maximum lag length to all the IRFs a DSGE model is asked to match. In this ad hoc approach, the set of possible IRFs consists of $\gamma_1, \gamma_2, ..., \gamma_H$ only:

$$\bar{C}_H = \left\{ c_h = [1_{1\times n_y^2 h} \;\; 0_{1\times n_y^2(H-h)}]', \text{ for } h = 1, 2, ..., H \right\}.$$

where $1_{m\times n}$ and $0_{m\times n}$ denote $m \times n$ matrices of ones and zeros, respectively. Our RIRSC can be easily tailored to this special case. Define $\bar{C}_{\mathcal{E},H}$, $\bar{C}_{min,H}$ and $\bar{c}_r$ by $C_{\mathcal{E},H}$, $C_{min,H}$ and c_r, respectively, with C_H replaced by $\bar{C}_H$. Using these definitions, we implement the RIRSC by selecting the maximum of lag of IRFs to minimize:

$$\hat{h}_T = \arg\min_{h \in \{1,2,\ldots,H\}} RIRSC(c_h). \quad (8)$$

Let h_r denote the corresponding IRF lag length implied by $\bar{c}_r$. It follows immediately from Theorems 1 and 2 that h_r is consistent both weakly and strongly:

Corollary 3 (Weakly consistent IRFs selection (VAR case)) *Let the structural model have a VAR(p_0) representation (1), and the estimator of the deep parameters be defined as (2), where c is chosen by (8). Suppose that Assumptions 1, 2, 3, 4 and 5 hold with $\bar{C}_H$, $\bar{C}_{min,H}$ and $\bar{c}_r$ replaced by C_H, $C_{min,H}$ and c_r, respectively. Then $\hat{h}_T \xrightarrow{p} h_r$.*

Corollary 4 (Strongly consistent IRFs selection (VAR case)) *Suppose that Assumptions 1, 2, 3, 4, 6, 7 and 8 hold with $\bar{C}_H$, $\bar{C}_{min,H}$ and $\bar{c}_r$ replaced by C_H, $C_{min,H}$ and c_r, respectively, where Assumptions 7 and 8 are presented in Appendix A. Let $\hat{W}_T(c_h) = \hat{V}_T^{-1}(c_h)$ and $p_T = p_0 < \infty$. Then $\hat{h}_T \xrightarrow{a.s.} h_r$.*

3 Interpretation of the RIRSC

This section provides examples that clarify the identification problem for DSGE model parameters estimated by IRF matching. The examples also make concrete the definitions of redundant and relevant IRFs under the RIRSC.

Example 1 (Labor productivity and hours) *Christiano, Eichenbaum and Vigfusson (2006, Section 2) discuss an RBC model in which a technology shock is the only disturbance that affects labor productivity in the long run. By imposing the parameterization $\rho_l = \phi = 1$ and the short-run restriction $a_z = 0$, the model implies a VAR(2) of labor productivity (y_t/l_t) and employment (l_t):*

$$\begin{aligned}
\begin{bmatrix} \Delta ln\left(\frac{y_t}{l_t}\right) \\ \Delta ln(l_t) \end{bmatrix} &= \begin{bmatrix} \beta_0\delta & -\alpha\beta_0\delta \\ \delta & -\alpha\delta \end{bmatrix} \begin{bmatrix} \Delta ln\left(\frac{y_{t-1}}{l_{t-1}}\right) \\ \Delta ln(l_{t-1}) \end{bmatrix} \\
&+ \begin{bmatrix} \beta_0\delta & -\alpha\beta_0\delta \\ \delta & -\alpha\delta \end{bmatrix} \begin{bmatrix} \Delta ln\left(\frac{y_{t-2}}{l_{t-2}}\right) \\ \Delta ln(l_{t-2}) \end{bmatrix} + \begin{bmatrix} 1 & \beta_0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t \\ \nu_t \end{bmatrix}
\end{aligned}$$

where $\beta_0 = -\beta_1$, $\delta = -\tilde{a}_z/(1-\alpha)$, $\eta_t = (1-\alpha)\sigma_z\varepsilon_t^z$, $v_t = a_l\sigma_l\varepsilon_t^l$, *and* ε_t^l, ε_t^z *have zero mean, unit variance and are uncorrelated. The structural parameters of interest are* α, σ_l^2, σ_z^2, a_l *and* $\tilde{a}_z$ *(cfr. Watson, 2006, eqs. 3,4). Let*

$$\Gamma_j = \begin{bmatrix} \gamma_{j,11} & \gamma_{j,12} \\ \gamma_{j,21} & \gamma_{j,22} \end{bmatrix}$$

denote the j*th-step structural IRFs. The restrictions on the first three-steps IRFs are*

$$\gamma_{0,11} = (1-\alpha)^2\sigma_z^2 \tag{9}$$

$$\gamma_{0,12} = -\alpha\,(a_l\sigma_l)^2 \tag{10}$$

$$\gamma_{0,21} = 0 \tag{11}$$

$$\gamma_{0,22} = (a_l\sigma_l)^2 \tag{12}$$

$$\gamma_{1,11} = -\alpha(1-\alpha)\tilde{a}_z\sigma_z^2 \tag{13}$$

$$\gamma_{1,12} = 2\alpha^2\tilde{a}_z\,(a_l\sigma_l)^2/(1-\alpha) \tag{14}$$

$$\gamma_{1,21} = (1-\alpha)\tilde{a}_z\sigma_z^2 \tag{15}$$

$$\gamma_{1,22} = -2\alpha\tilde{a}_z\,(a_l\sigma_l)^2/(1-\alpha) \tag{16}$$

$$\gamma_{2,11} = (\alpha(1-\alpha)\tilde{a}_z + 2\alpha^2\tilde{a}_z^2)\sigma_z^2 \tag{17}$$

$$\gamma_{2,12} = -2\alpha^2\,(a_l\sigma_l)^2\,\tilde{a}_z/(1-\alpha) \tag{18}$$

$$\gamma_{2,21} = -((1-\alpha)\tilde{a}_z + 2\alpha\tilde{a}_z^2)\sigma_z^2 \tag{19}$$

$$\gamma_{2,22} = 2\alpha\,(a_l\sigma_l)^2\,\tilde{a}_z/(1-\alpha). \tag{20}$$

Since a_l and σ_l cannot be separately identified (only their product is identified), four out of the five deep parameters are identified. For example, α is identified from restrictions (10) and (12), while $a_l\sigma_l$, $\tilde{a}_z$ and σ_z^2 are identified from restrictions (12), (13) and (9), respectively.

There are two trivial examples of redundant impulse responses. One is restriction (11). Another is restrictions on Γ_j, for $j > 2$: since the model is a VAR(2) model, restrictions for $j > 2$ are non-linear transformation of (9)–(20), and thus are first-order equivalent to some linear combinations of the above restrictions. Therefore, adding these restrictions will not reduce the asymptotic variance. However, even if an impulse response depends on the parameters of interest and its horizon is less than or equal to p, the impulse response may be redundant.[7]

When the order of a VAR is finite, it is unlikely that there is efficiency gain from using additional IRFs once sufficiently many IRFs are included. Even when the underlying processes are of infinite order, it is not always necessary to use all IRFs that are available. Consider a simple growth model in which total factor productivity growth follows a first-order moving average process with parameter θ and the Cobb-Douglas production has parameter α. In this model consumption growth follows an ARMA(1,1) process:

$$\Delta c_t = \theta \Delta c_{t-1} + \varepsilon_t - (\theta - \alpha)\varepsilon_{t-1}, \tag{21}$$

and thus it has an infinite-order moving average representation. We report below the asymptotic efficiency loss from estimating α and θ from a finite number of impulse responses by computing the ratio of determinants of asymptotic covariance matrices of the IRFME estimator of α and θ.[8] A number greater than one denotes the existence of efficiency losses deriving from not using infinitely many impulse responses.

horizon	1	2	3	4	5	6	7	8
efficiency loss	∞	1.857	1.177	1.044	1.011	1.003	1.001	1.000

While the two parameters are identified by the IRFs at lags one and two, the table shows there are some efficiency gain from using IRFs at lags higher than two. This efficiency gain disappers quicky after seven lags, however. This simple example shows that even when the data generating processes are of infinite-order, efficiency can be effectively achieved by using only a finite number of impulse responses.

[7]For example, suppose $\alpha = 0.5$, $\sigma_z = a_l = \sigma_l = \widetilde{a}_z = 1$, where α is to be estimated and the latter parameters are instead known. Let the covariance matrix of the impulse responses be the identity matrix. When α is estimated by using the optimal weighting matrix, using (9), (10), (14)-(20) produces the same asymptotic variance as using (9),(10),(13)-(20). Thus (13) is a redundant impulse response (it does not help to identify α, as it is used to identify other parameters that are assumed to be known in the example in this note). Although the redundant impulse response does not change the asymptotic variance, it can inflate the variance in finite samples. The other IRF are all relevant. Omitting any of these IRFs increases the asymptotic variance and will likely increase the finite sample variance.

[8]In the calculations, we set $\alpha = 0.5$, $\theta = 0.25$ and $var(\varepsilon_t) = 1$. The asymptotic covariance matrix of the IRFME estimator for the infinite-order case is approximated by the one that uses the first thousand impulse responses. The results do not change when a larger number of impulse responses is used.

4 Alternative IRF matching estimators

Although the VAR-based IRFME is the most widely used IRFME, alternative IRFME have been proposed in the literature. Jordà and Kozicki (2007) proposed IRFME based on local projections. In addition, researchers have been interested in simulation-based methods to approximate theoretical impulse responses. This section extends the RIRSC to these IRF matching estimators, and describes how our criterion is implemented in these contexts.

4.1 The IRF Matching Projection Estimator

Consider first the local projections method advocated by Jordà (2005) and used in Jordà and Kozicki (2007). The simplest version of his estimator for the $\tau-$th step impulse response is $\hat{B}_{1,\tau}D$, where $\hat{B}_{1,\tau}$ is directly estimated from

$$y_{t+\tau} \;=\; B_{0,\tau} + B_{1,\tau}y_{t-1} + B_{2,\tau}y_{t-2} + \cdots + B_{p,\tau}y_{t-p} + u_{t+\tau}$$

for $\tau = \underline{h}, ..., H$, and D is a matrix derived from the identification procedure.

Jordà's local projection impulse responses estimator is:

$$\hat{\theta}_J(c) \;=\; \arg\min_{\theta\in\Theta}(\widehat{\gamma}_T(c) - \gamma(\theta;c))'\hat{W}_T(c)(\widehat{\gamma}_T(c) - \gamma(\theta;c)) \tag{22}$$

where $\widehat{\gamma}_T(c)$ is a vector of structural impulse responses estimated by local projections, $g(\theta;c)$ is the vector of the model's theoretical impulse responses given structural parameter θ, and $\hat{W}_T(c)$ is a weighting matrix.

Theorem 5 (Consistent IRF selection (Local projections case)) *Suppose that Assumptions 1, 2, 3, 4 and 5 hold with $\hat{V}_T(c)$ replaced by the asymptotic covariance matrix of $\hat{\theta}_{J,T}(c)$, $\hat{V}_{J,T}(c)$. Let the estimator of θ be (22), where c is chosen such that:*

$$\begin{aligned} \hat{c}_T &= \arg\min_{h\in C_H} RIRSC_J(c), \;\; \textit{and} \\ RIRSC_J(h) &= log(|\hat{V}_J(c)|) + \kappa(h, p_T, T). \end{aligned}$$

Then $\hat{c}_T \xrightarrow{p} c_r$.

4.2 The IRF Matching Simulation Estimator

The third estimator that we consider is the simulation-based estimator, which we will refer to as the Sims-Cogley-Nason (SCN) estimator. In this case, we assume that the DSGE model implies an infinite-order VAR process; the reason is that when the VAR is of finite order there is no advantage from using simulation-based estimators because there is no lag-truncation problem.

The SCN estimator is implemented as follows. First, fit a VAR(p) to the actual data to obtain sample impulse responses $\widehat{\gamma}_T(c)$.[9] Next, draw a vector of shock innovations of length T and simulate synthetic data from the DSGE model with parameter vector θ and apply the VAR(p) procedure to obtain a vector of simulated theoretical IRFs. Let $\widetilde{g}_T^{(s)}(\theta;c)$ denote the vector of simulated impulse responses from the s-th synthetic sample, $s = 1,..,S$, where S is the total number of simulation replications. Finally, define $\widetilde{g}_T(\theta;c)$ to be the average across the ensemble of simulated IRFs, which we refer to as the approximate theoretical impulse responses: $\widetilde{g}_T(\theta;c) = (1/S)\sum_{s=1}^{S}\widetilde{g}_T^{(s)}(\theta;c)$. Note that the vector of shock innovations is drawn only once and held fixed as θ is adjusted to move $\widetilde{g}_T(\theta;c)$ closer to $\widehat{\gamma}_T(c)$.

The SCN estimator of θ minimizes the distance between the average simulated theoretical impulse responses and the sample impulse responses:

$$\hat{\theta}_{SCN,T}(c) \;=\; \arg\min_{\theta\in\Theta}(\widehat{\gamma}_T(c) - \widetilde{g}_T(\theta;c))'\hat{W}_T(c)(\widehat{\gamma}_T(c) - \widetilde{g}_T(\theta;c)), \tag{23}$$

where $\hat{W}_T(c)$ is a weighting matrix.[10] Let $\hat{V}_{SCN}(c)$ denote a consistent estimate of the asymptotic variance of $\hat{\theta}_{SCN}(c)$, which is computed in the Appendix (see eq. (64)).

Next, consider the problem of selecting the impulse responses for the IRFME. Theorem (6) describes the IRF selection criterion we propose for the SCN estimator:

Theorem 6 (Consistent IRF selection (Simulation-based estimators case)) *The estimator of $\hat{\theta}_{SCN}(c)$ is (23), where h is chosen s.t.:*

[9] All the subsequent estimated parameters should also be function of p, the estimated VAR lag length. However, in order to simplify notation, we drop this dependence in the notation.

[10] In other words, the SCN estimator can be viewed as an indirect-inference estimator with a sequence of finite-order VAR models used as an auxiliary model (see Smith (1993) and Gourieroux, Monfort and Renault (1993) for examples of indirect inference applied to DSGE models and financial models, respectively). The Appendix shows that, under quite mild conditions, $\hat{\theta}_{SCN}(c)$ is consistent and asymptotically normal.

$$\begin{aligned} \hat{c}_T &= \arg\min_{c \in C} RIRSC_{SCN}(c), \\ RIRSC_{SCN}(c) &= ln(|\hat{V}_{SCN,T}(c)|) + \kappa(|c|, p_T, T), \end{aligned}$$

where $\hat{V}_{SCN,T}(c)$ *is an estimate of the asymptotic covariance matrix of* $\hat{\theta}_{SCN,T}(c)$*. Let Assumptions 1, 2, 3, 4 and 5 hold, with* $\hat{\theta}_T(c)$ *and* $\hat{V}_{\theta,T}(c)$ *replaced by* $\hat{\theta}_{SCN,T}(c)$ *and* $\hat{V}_{SCN,T}(c)$*; then* $\hat{c}_T \xrightarrow{p} c_r$*.*

5 Empirical analysis of two representative DSGE models

This section applies a VAR-based IRFME and the RIRSC to the new Keynesian DSGE models of CEE and ACEL. The goal is to assess the impact of the RIRSC on the estimated parameters of these DSGE models. Thus, we estimate the CEE and ACEL models fixing the maximum number of impulse response lags at 20 (excluding those that are zero by assumption) and employing the RIRSC. In either case, the IRFME is implemented with a diagonal weighting matrix.[11]

The CEE and ACEL DSGE models use different schemes to identify IRFs. The former model is estimated by matching the responses of nine aggregate variables only to an identified monetary policy shock. The identification relies on an impact restriction that orthogonalizes the monetary policy shock with respect to the nine aggregate series. We use this identification to estimate nine parameters of the CEE DSGE model.

We follow ACEL by identifying the sample and theoretical IRFs with long-run neutrality restrictions. The ACEL DSGE model is constructed to satisfy three restrictions: (i) neutral and capital embodied shocks are the only shocks that affect productivity in the long run; (ii) the capital embodied shock is the only shock that affects the price of investment goods; and (iii) monetary policy shocks do not contemporaneously affect aggregate quantities and prices. These restrictions identify IRFs for ten aggregate variables with respect to neutral technology, capital embodied and monetary policy shocks. The ACEL DSGE model presents 18 parameters to estimate.

Table 1(a) reports the results for the ACEL DSGE model. From the left to right of the table, the columns list parameters, parameter estimates and standard errors under RIRSC, and parameter estimates and standard errors given a fixed IRF lag length of 20. We implement the RIRSC by

[11] ACEL remark that the diagonal weighting matrix ensures that the estimated DSGE model parameters are such that theoretical IRFs lie as much as possible within confidence bands of estimated IRFs.

matching the IRFs with respect to the three shocks and progressively reduce the lags in all three IRFs one by one. Next, the RIRSC criterion (3) is applied as the number of lags in each IRF ranges from two to 20, which gives a total of number of IRF points (h) ranging between 6 and 60. The RIRSC selects $h = 3$ for the three IRFs, which makes it possible for the 18 ACEL DSGE model parameters to be identified.

The RIRSC has one important effect on ACEL DSGE model parameter estimates. Across the RIRSC and fixed lag length IRFMEs, there are six ACEL DSGE model parameters with t-ratios greater than two, with qualitatively similar point estimates. The fixed lag length IRFME yields an additional parameter, ρ_{μ_z}, which is the AR(1) coefficient on the growth rate of the labor neutral productivity shock, whose point estimate is 0.89 with a standard error of 0.16. This implies a persistent growth rate of the labor neutral productivity shock (e.g., its half-life to an own shock is over six quarters) that contrasts with the RIRSC-based estimate $\rho_{\mu_z} = 0.24$ and a standard error of 0.70. Since this standard error is nearly three times larger than its point estimate, under RIRSC, inference points to a random walk labor neutral productivity shock for the ACEL DSGE model. Although the remaining 11 ACEL DSGE model parameters have t-ratios less than two, note the distance across the RIRSC and fixed lag estimates (which are close to those reported by ACEL). For example, the RIRSC and fixed lag IRFMEs produce an estimate of the coefficient on marginal cost in the new Keynesian Phillips curve (NKPC), γ, of 0.21 and 0.04, respectively. The latter estimate produces a steeply sloped NKPC, while the latter suggests monetary policymakers face a weaker trade-off. Nonetheless, these estimates of γ are smaller than the associated standard errors. Another appealing feature of the RIRSC-IRFME appears from the standard errors reported in parentheses below the estimates reported in Table 1(a). Note that the RIRSC-IRFME is overall more efficient, which tends to be in smaller standard errors.

INSERT TABLES 1,2 HERE

A crucial aspect of the ACEL DSGE model is the implied average time between firms' price re-optimization, which is a function of γ. Since the RIRSC-IRFME estimate of γ is larger than the fixed lag IRFME estimate, according to Table 1(b) the former estimate implies that on average monopolistically competitive firms change their prices at most about every three quarters in the homogeneous capital model. This contrasts with the fixed lag IRFME, which estimates price changes every five quarters on average. From the standard errors reported in parentheses below the estimates reported in Table 1(a): note that the differences are statistically significant at conventional levels.

Table 2 presents estimates of the CEE DSGE model, where the monetary policy shock is the only shock of interest. In this case, the RIRSC chooses 6 lags for the impulse response. We see that RIRSC and the fixed lag length IRFMEs generate nearly identical results for the five CEE DSGE model parameters with t-ratios greater than two. The remaining four parameter estimates differ across the RIRSC and the fixed lag length IRFMEs. However, the fixed lag length IRFME delivers estimates that are often close to those reported by CEE.[12]

As a robustness analysis, we investigate whether the insensitivity of our point estimates in Tables 1 and 2 to a different IRF lag length is robust to different choices for the initial parameter values and to the step size for the numerical derivatives. Unreported results show that a slight perturbation of the initial parameter values does not substantially change the main results, although the estimates might change considerably when the magnitude of the perturbation is large.[13] The results are considerably less sensitive to the choice of the step size; in that case, the estimates and standard errors change only very slightly.

6 Monte Carlo robustness analysis

The striking difference in the estimates of some key parameters in the previous section deserves an additional careful investigation into the causes of why this happens. In this section, we argue that the difference in the estimates is likely caused by small sample biases, and report Monte Carlo simulations to show that the use of our methodology provides substantially more precise estimation of the deep parameters of the structural models. Unfortunately, a careful Monte Carlo analysis of

[12] We attribute any disparities between the fixed lag estimates of Table 2 and those of CEE to modifications to the computational procedure used to implement the IRFME. For example, we make it more robust to changes in the initial parameter values. Further, we aim to obtain more precise results by (i) using a Newton-Raphson type algorithm rather than a simplex algorithm; (ii) increasing the maximum iterations to 1000 rather than 10; and (iii) changing the grid sizes for numerical derivatives. The latter two are responsible for most of the differences in the numerical parameter values.

[13] In particular, results were robust to adding a Normal(0, σ) shock to the initial parameter values with $\sigma \in [1, 10]$, but were not robust to ad-hoc initial parameter values (e.g. the origin).

ACEL and CEE is currently computationally infeasible, so we consider a simple univariate AR(1) process; and the structural VAR(2) example discussed in (1).

6.1 The AR(1)

To start, first consider the following simple univariate AR(1):

$$y_t = \rho y_{t-1} + \varepsilon_t \ , \ t = 1, 2, ...T$$

where ε_t are random draws from a normal distribution with mean zero and variance one, $\rho = 0.4$ and $T = 100$. We estimate the deep parameter ρ by the IRFME that minimizes the distance between the vector of IRFs estimated by fitting an AR(2) to the data and the theoretical IRF derived from the AR(1). The weighting matrix W is the inverse of the covariance matrix of the estimated IRFs calculated by using Monte Carlo simulation. In this section we let H denote either the number of IRFs matched by the IRFME with a fixed number of IRF lags (when we refer to the usual IRFME) or the maximum number of IRFs considered when criterion (3) is used to select the relevant IRF lag length.

Table 3 reports, for various values of H, both the estimated average bias ("bias") and the empirical rejection rates ("rej. rate") of nominal 5% significance level tests for the following estimators: the IRF matching estimator with H IRF lags, labeled "IRFME"; the IRF matching estimator using only the IRFs selected by (3), labeled "IRFME$_{RIRSC}$"; and the usual AR(1) estimator, labeled "AR(1)". Note that the IRFME with $H = 1$ is the maximum likelihood estimator. We performed 1,000 Monte Carlo replications, discarding replications in which the estimator did not converge numerically.

The table shows that the bias of IRFME tends to increase (in absolute value) with the number of IRFs used (H) and its rejection rates are well above the nominal level of 0.05 for $H \geq 5$, and tend to go to one as H increases. The table also shows that the RIRSC method that we propose does not suffer from over-rejections, and that it substantially reduces the bias of the traditional IRFME.

INSERT TABLE 3 HERE

6.2 The structural VAR(2) discussed in example (1)

We consider estimation of α and γ_l in example 1 by IRFME. We set $\alpha = 0.35$, $\gamma_l = 1$, $\rho = 1$, $\sigma_l = 1$ and $\sigma_z = 1$. [14] The sample sizes considered are $T = 100, 200, 400$ and the number of Monte Carlo replications is set to 1000. We focus on the choice of horizons and the minimum and maximum horizons are 1 and 12, respectively.

Table 4 reports the median of absolute bias, variance and coverage probabilities of the 95% confidence interval baed on the t test when the number of impulse responses is fixed. As expected, both the bias and the variance become smaller and the coverage becomes more accurate as the sample size increases. In this DGP, the coverage probability is most affected by the number of impulse responses. The best coverage probability is obtained when $h = 1$ or $h = 2$ and it deteriorates as more impulse responses are included.

Table 5 shows that the performance of the IRFME using all the impulse responses and the IRFME using the only the IRFs selected by the RIRSC. The table shows that the RIRSC significantly improves the coverage probability. It also reduces bias and variance for γ_1. Although the AIC-type penalty term does not satisfy Assumption 5, it reduces the number of impulse responses which results in the improved performance of the IRFME.

Table 6 presents summary statistics of the selected numbers of impulse responses. The RIRSC with the SIC-type penalty term tends to choose $h = 2$ as the sample size grows in the sense that the variance becomes small. The RIRSC with the AIC-type penalty term tends to choose larger numbers of impulse responses and the variance is also larger than the other types of the penalty term.

INSERT TABLES 4,5,6 HERE

7 Conclusions

This paper's objective is to contribute to the literature on the estimation of dynamic stochastic general equilibrium (DSGE) models by using impulse response function matching estimators

[14] We have looked at all the cases in which $\alpha \in \{0.275, 0.35, 0.425\}$, $\rho_l = \{0.75, 0.85, 0.9, 0.95, 0.975, 1\}$, $\sigma_l = \{0.5, 0.75, 1, 1.25, 1.5\}$. They are qualitatively similar to the reported results and are available upon request to the authors.

(IRFMEs). We propose a simple and econometrically sound method for doing so. We show by Monte Carlo simulations that our method can substantially improve the precision of the parameter estimates and decrease their small sample biases. We also show that our method can substantially change key parameter estimates of existing representative DSGE models. We hope that the simplicity and the usefulness of the criterion that we propose will increase the applicability of impulse response function matching estimators in practice.

Our framework assumes, as in most of the literature on IRF matching, that the IRFs to be used in the estimation are correctly specified. Although it could be interesting to identify correctly specified IRFs and those that are not, it is outside the scope of this paper.

Our paper provides an information criterion to improve upon commonly used IRFMEs. We do not provide a systematic analysis of the relative merits of using IRFMEs versus alternative estimators (e.g. MLE or Bayesian methods). The latter use the entire likelihood of the model whereas the IRF matching focuses only on selected aspects of the model (e.g. limited information methods), therefore giving rise to the usual trade-off between efficiency and robustness. We leave these issues to future work.

8 References

Altig, David, Lawrence Christiano, Martin Eichenbaum and Jesper Linde (2005): "Firm-Specific Capital, Nominal Rigidities and the Business Cycle", *NBER* Working Paper 11034, Cambridge, MA.

Andersen, T.G. and B.E. Sorensen (1996), "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study", *Journal of Business and Economic Statistics*, 14, 328–352.

Blinder, Alan S., Elie Canetti, David Lebow and Jeremy Rudd (1998): *Asking about Prices: A New Approach to Understanding Price Stickiness*, Russell Sage Foundation.

Boivin, Jean and Marc Giannoni (2006): "Has Monetary Policy Become More Effective?", *Review of Economics and Statistics* 88(3), 445-462.

Christiano, Lawrence and Martin Eichenbaum (1992): "Current Real Business Cycle Theories and Aggregate Labor Market Fluctuations", *American Economic Review* 82(3), 430-450.

Christiano, Lawrence, Martin Eichenbaum and Charles Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy", *Journal of Political Economy*, 113(1), 1-45.

Christiano, Lawrence, Martin Eichenbaum and Robert Vigfusson (2006): "Assessing Structural VARs", *Macroeconomics Annual 2006*, by Daron Acemoglu, Kenneth Rogoff and Michael Woodford, MIT Press, Boston, MA.

Cogley, Timothy and James M. Nason (1995): "Output Dynamics in Real-Business-Cycle Models", *American Economic Review* 85(3), 492-511.

Davidson, James (1994): *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford University Press: Oxford, UK.

Dhrymes, Phoebus J. (1984), *Mathematics for Econometrics,* Second Edition, Springer Verlag: New York, NY.

DiCecio, Riccardo (2005): "Comovement: It's Not a Puzzle", Working Paper 2005-035, *Federal Reserve Bank of St. Louis.*

DiCecio, Riccardo, and Edward Nelson (2007): "An Estimated DSGE Model for the United Kingdom", *Review*, Federal Reserve Bank of St. Louis, 89(4), 215-231.

Dupor, Bill, Jing Han, and Yi Chan Tsai (2007): "What Do Technology Shocks Tell Us about the New Keynesian Paradigm?", *mimeo*, Department of Economics, Ohio State University.

Gourieroux, Christian S., Alain Monfort and Eric Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S118.

Gourieroux, Christian S., and Alain Monfort (1997): *Simulation Based Econometric Methods.* Oxford University Press, Oxford, UK.

Hall, Alastair, Atsushi Inoue, Kalidas Jana and Changmock Shin (2007): "Information in Generalized Method of Moments Estimation and Entropy Based Moment Selection", *Journal of Econometrics* 138, 488-512.

Iacoviello, Matteo (2005), "House Prices, Borrowing Constraints and Monetary Policy in the Business Cycle", *American Economic Review*, 95(3), 739-764.

Jordà, Òscar (2005): "Estimation and Inference for Impulse Responses by Local Projections," *American Economic Review* 95(1), 161-182.

Jordà, Òscar and Sharon Kozicki (2007): "Estimation and Inference by the Method of Projection Minimum Distance", *Bank of Canada* Working Paper 2007-56.

Jordà, Òscar (2006): "Inference for Impulse Responses," *mimeo*, Department of Economics, University of California.

Lewis, R. and G. C. Reinsel (1985): "Prediction of Multivariate Time Series by Autoregressive Model Fitting", *Journal of Multivariate Analysis*, 16, 393–411.

Lütkepohl, Helmut and D.S. Poskitt (1991): "Estimating Orthogonal Impulse Responses Via Vector Autoregressive Models," *Econometric Theory* 7, 487–496.

Magnus, Jan R. and Heinz Neudecker (1999): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised Edition, Chicester, England: John Wiley & Sons.

Newey, Whitney and Daniel McFadden (1994): "Large Sample Estimation and Hypothesis Testing", in: R.F. Engle and D.L. McFadden, *Handbook of Econometrics*, Vol. 4, Elsevier.

Phillips, Peter C.B. and Werner Ploberger (2003), "Empirical Limits for Time Series Econometric Models," *Econometrica,* 71, 627–673.

Rotemberg, Julio and Michael Woodford (1997): "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy", in Bernanke, B. and J. Rotemberg, eds., *NBER Macroeconomics Annual*, Cambridge, MA: MIT Press.

Sims, Christopher A. (1989): "Models and Their Uses", Discussion Paper 11, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis.

Smith, Anthony A. (1993): "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions", *Journal of Applied Econometrics* 8, S63-S84.

Uribe, Martin and Vivien Z. Yue (2006): "Country Spreads and Emerging Markets: Who Drives Whom?", *Journal of International Economics* 69, 6-36.

Appendix A: Notation and Additional Assumptions

Notation. In what follows, $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{d}$ denotes convergence in distribution, $dim(x)$ denotes the length of vector x, and for a matrix A: $\|A\|^2 \equiv tr(A'A)$, $\widehat{A}$ denotes an estimate of A, "p.s.d." denotes positive-semidefinite, "p.d." denotes positive-definite, and $E\,(.)$ denotes the expectation operator. Finally, B^c denotes the complement of a set B.

Assumption 7

$$\begin{aligned} T^{1/2}(\hat{\theta}_T(c)-\theta_0) &= -\left\{G(c)'W(c)G(c)\right\}^{-1} G(c)'W(c)T^{1/2}[\hat{\gamma}_T(c)-f(\theta_0;c)] \\ &\quad + o(1) \quad a.s. \end{aligned} \tag{24}$$

$$T^{1/2}vec\{\hat{G}_T(c)-G(c)\} = \bar{G}(c)T^{-1/2}(\hat{\theta}_T(c)-\theta_0) + o(1) \quad a.s. \tag{25}$$

$$T^{1/2}vech\{\hat{V}_T(c)-V(c)\} = T^{1/2}\sum_{t=1+m}^{T} h_v(y_t,\theta_0;c) + o(1) \quad a.s. \tag{26}$$

$$T^{1/2}(\hat{\gamma}_T(c)-g(\theta_0;c)) = T^{-1/2}\sum_{t=1+m}^{T} h_\gamma(y_t,\theta_0;c) + o(1) \quad a.s. \tag{27}$$

for some $0<m<\infty$*, and where* $\bar{G}(c) = (\partial/\partial\theta')vec\left\{\partial g(\theta;c)/\partial\theta'\right\}\big|_{\theta=\theta_0}$.

Assumption 8 *Let* $h(y_t,\theta_0;c) = [h_\gamma(y_t,\theta_0;c)'\ h_v(y_t,\theta_0;c)']'$.
Define $\Omega_h(c) = lim_{T\to\infty}Var[T^{-1/2}\sum_{t=1}^{T} h(v_t,\theta_0;c)]$. $\sum_{t=1}^{T} h(v_t,\theta_0;c)$ *satisfies the Law of Iterated Logarithms (LIL) in the sense that for all* $b\in\Re^{dim(h)}$ *with* $\|b\|=1$,

$$lim\ sup_{T\to\infty}\left\{\frac{1}{(2TlnlnT)^{1/2}}\left|b'\Omega_h(c)^{-1/2}\sum_{t=1}^{T} h(v_t,\theta_0;c)\right|\right\} = 1,\ a.s.$$

for all $c\in\mathcal{C}$.

Assumption 9 (Asymptotic Normality of Simulation-Based Estimators) *In model (1): (a) As* $p_T, T\to\infty$, $p_T^4/T\to 0$. *(b) The parameter space* Θ *is compact. (c) Let* $g(\theta;c,p_T)$ *denote a vector of population impulse responses implied by a VAR(*p_T*) model fitted to the data simulated with* θ *and selected by c; to simplify notation, we let* $g_T(\theta;c)\equiv g(\theta;c,p_T)$. *There is a sequence of covariance matrix* $\{\Sigma_{g_T(\theta;c)}\}$ *such that, for any sequence of vectors* $\{\ell_{p_T}\}$ *satisfying* $0<L_1\le\|\ell_{p_T}\|\le L_2<\infty$ *for all* p_T, $\sqrt{T}\ell_{p_T}'(\widehat{\gamma}_T(c)-g_T(\theta_0;c)) \xrightarrow{d} N(0,\lim_{p_T\to\infty}\ell_{p_T}'\Sigma_{g_T(\theta_0;c)}\ell_{p_T})$ *and* $\sqrt{T}\ell_{p_T}'(\widetilde{g}_T^{(s)}(\theta;c)-g_T(\theta;c)) \xrightarrow{d} N(0,\lim_{p_T\to\infty}\ell_{p_T}'\Sigma_{g_T(\theta;c)}\ell_{p_T})$ *jointly and independently for every* $\theta\in\Theta$ *and* $s=1,2,...,S$. *(d)* $\lim_{p_T\to\infty}\|g_T(\theta;c)-g_T(\theta_0;c)\|=0$ *if and only if* $\theta=\theta_0$. *(e)*

$\{g_T(\theta;c)\}$ is continuously differentiable and the rank of $G_T(\theta_0;c) \equiv (\partial/\partial\theta')vec\{\partial g_T(\theta;c)/\partial\theta'\}\big|_{\theta=\theta_0}$ is $dim(\theta)$ for $c \in C_{I,H}$ and sufficiently large H. (f) $g_T(\theta;c)$ and $\widetilde{g}_T(\theta;c)$ satisfy Lipschitz conditions, $\|g_T(\theta_1;c) - g_T(\theta_2;c)\| < \mathcal{L}\|\theta_1 - \theta_2\|$ and $\|\widetilde{g}_T(\theta_1;c) - \widetilde{g}_T(\theta_2;c)\| < \tilde{\mathcal{L}}\|\theta_1 - \theta_2\|$ where $\mathcal{L}$ and $\hat{\mathcal{L}}$ do not depend on θ and θ' and are $O(1)$ and $O_p(1)$, respectively, uniformly in T. (g) Each row of $\{G_T(\theta_0;c)\}$ is summable. (h) There is a sequence of matrices $\{W_T(c)\}$ such that, for any absolutely summable sequence of vectors $\{\ell_{p_T}\}$, $\ell'_{p_T}(\hat{W}_T(c) - W_T(c))\ell_{p_T} = O_p\left(p_T^2/\sqrt{T}\right)$. (k) The eigenvalues of $\{W_T(c)\}$ are all positive and bounded away from zero and bounded above by some finite constant. (l) For any absolutely summable sequence of vectors, $\{v_{p_T}\}$, $\lim_{p_T\to\infty} v'_{p_T} W_T v_{p_T}$ is well-defined. (m) There are consistent estimators of $\Sigma_{g_T(\theta_0;c)}$ and $\Sigma_{g_T(\theta;c)}$, $\hat{\Sigma}_{g_T(\theta_0,c)}$ and $\tilde{\Sigma}^{(s)}_{g_T(\theta;c)}$ respectively, such that, for any absolutely summable sequence of vectors $\{\ell_{p_T}\}$, $\ell'_{p_T}\hat{\Sigma}_{g_T(\theta_0,c)}\ell_{p_T} - \ell'_{p_T}\Sigma_{g_T(\theta_0;c)}\ell_{p_T} = O_p(p_T^2/\sqrt{T})$ and $\ell'_{p_T}\tilde{\Sigma}^{(s)}_{g_T(\theta;c)}\ell_{p_T} - \ell'_{p_T}\Sigma_{g_T(\theta;c)}\ell_{p_T} = O_p(p_T^2/\sqrt{T})$

Remark 7. The asymptotic normality of structural impulse responses estimators in Assumption 2 is a high-level assumption, and follows from arguments similar to those in Lewis and Reinsel (1985) and Lütkepohl and Poskitt (1991).

Appendix B: Proofs

Proof of Theorem 1 when $p_T = p_0$: First suppose that $c \in C_{\mathcal{E},H}$ and $c \neq c_r$. It follows from Assumptions 3, 4(c) and 5 that

$$\begin{aligned} T^{1/2}(RIRSC(c) - RIRSC(c_r)) &= T^{1/2}(ln(|\hat{V}_{\theta,T}(c)|) - ln(|\hat{V}_{\theta,T}(c_r)|)) \\ &\quad + T^{1/2}(\kappa(|c|, p_T, T) - \kappa(|c_r|, p_T, T)) \\ &\to +\infty \end{aligned} \tag{28}$$

as the first term is $O_p(1)$ by Assumption 4(c) and the second term diverges to infinity by Assumptions 3 and 5. Thus $T^{1/2}(RIRSC(c) - RIRSC(c_0))$ is positive with probability approaching one as $T \to \infty$. Next consider the case in which $c \in C_{I,H} \cap C^c_{H,\mathcal{E}}$. By Theorem 22 of Magnus and Neudecker (1999, p.21), it follows from Assumption 3 that $ln(|V_\theta(c)|) - ln(|V_\theta(c_r)|) > 0$. Thus it follows from Assumptions 4(c) and 5

$$\begin{aligned} RIRSC(c) - RIRSC(c_r) &= ln(|\hat{V}_{\theta,T}(c)|) - ln(|\hat{V}_{\theta,T}(c_r)|) + \kappa(|c|, p_T, T) - \kappa(|c_r|, p_T, T) \\ &= ln(|V_\theta(c)|) - ln(|V_\theta(c_r)|) + o_p(1) \\ &> 0 \end{aligned} \tag{29}$$

with probability approaching one. Third, when $c \in C_{NI,H} \cap C^c_{\mathcal{E},H}$, it follows from Assumption 4(c) that

$$RIRSC(c) - RIRSC(c_r) \xrightarrow{p} +\infty. \tag{30}$$

Because $C_{\mathcal{E},H} \cup (C_{I,H} \cap C^c_{\mathcal{E},H}) \cup (C_{NI,H} \cap C^c_{\mathcal{E},H}) = C_{I,H} \cup C_{NI,H} = C_H$,

$$RIRSC(c_r) < RIRSC(c) \tag{31}$$

for all $c \in C_H$ such that $c \neq c_r$ with probability approaching one asymptotically. Since

$$RIRSC(\hat{c}_T) \leq RIRSC(c)$$

for all $c \in C_H$,

$$RIRSC(\hat{c}_T) \leq RIRSC(c_r) \tag{32}$$

Therefore it follows from (31), (32) and Assumption 3 that $\hat{c}_T \xrightarrow{p} c_r$.

Proof of Theorem 1 when $p_T \to \infty$ as $T \to \infty$: First consider the case in which $|c_r| < \infty$. By Assumption 4(e), $c_r \in C_H$ for sufficiently large H. By using the proof for the case $p_T = p_0$ with $T^{1/2}$ replaced by $T^{1/2}/p_T$, we obtain

$$\begin{aligned} RIRSC(c) - RIRSC(c_r) &= ln(|\hat{V}_\theta(c)|) - ln(|\hat{V}_\theta(c_r)|) + \kappa(|c|, p_T, T) - \kappa(|c_r|, p_T, T) \\ &= ln(|V_\theta(c)|) - log(|V_\theta(c_r)|) + o_p(1) \\ &> 0 \end{aligned} \tag{33}$$

with probability approaching one. Second, we consider the case $|c_r| = \infty$. For all c such that $|c| = \infty$, either $V_\theta(c) = V_\theta(c_r)$ or $V_\theta(c) > V_\theta(c_r)$ holds because c_r is an efficient selection vector by Assumption 3. If $V_\theta(c) > V_\theta(c_r)$ then (33) holds for sufficiently large T by Assumption 5. If

$V_\theta(c) = V_\theta(c_r)$

$$\begin{aligned} \frac{T^{1/2}}{p_T}(RIRSC(c) - RIRSC(c_r)) &= \frac{T^{1/2}}{p_T}(ln(|\hat{V}_{\theta,T}(c)|) - ln(|\hat{V}_{\theta,T}(c_r)|)) \\ &\quad + \frac{T^{1/2}}{p_T}(\kappa(|c|, p_T, T) - \kappa(|c_r|, p_T, T)) \\ &\to +\infty, \end{aligned} \tag{34}$$

where the first term is $O_p(1)$ by Assumption 4(c) and the second term diverges to infinity by Assumptions 3 and 5. Thus, for any $c \in C_H$

$$RIRSC(c) - RIRSC(c_r) \ > \ 0 \tag{35}$$

with probability approaching one. Therefore $\hat{c}_T - c_r \xrightarrow{p} 0$ by (32), (35) and Assumption 3.

Proof of Theorem 2: Let

$$\hat{M}_T(c) = \hat{V}_{\theta,T}(c)^{-1} = \hat{G}_T(c)'\widehat{W}_T(c)\hat{G}_T(c),$$

which implies

$$RIRSC(c) = -ln[|\hat{M}_T(c)|] + \kappa(|c|, T). \tag{36}$$

Also define $M(c) = G(c)'W(c)G(c)$.

We have the following expression for $\hat{M}_T(c) - M(c)$:

$$\hat{M}_T(c) - M(c) \ = \ \hat{G}_T(c)'\widehat{W}_T(c)\hat{G}_T(c) \ - \ G(c)'W(c)G(c) \tag{37}$$

$$\begin{aligned} &= \ \hat{G}_T(c)'\widehat{W}_T(c)\{\hat{G}_T(c) \ - \ G(c)\} \ + \ \hat{G}_T(c)'\{\widehat{W}_T(c) - W(c)\}G(c) \\ &\quad + \{\hat{G}_T(c) \ - \ G(c)\}'W(c)G(c) \end{aligned} \tag{38}$$

We also have the following representation:

$$\widehat{W}_T(c) - W(c) \ = \ W(c)\left\{V(c) \ - \ \hat{V}_T(c)\right\}\widehat{W}_T(c) \tag{39}$$

and

$$vec\left\{V_T(c) \ - \ \hat{V}_T(c)\right\} \ = \ Bvech\left\{V_T(c) \ - \ \hat{V}_T(c)\right\} \tag{40}$$

where (39) follows since $\hat{W}_T(c) = (\hat{V}_T(c))^{-1}$ and B is the matrix such that $vec(.) = Bvech(.)$.

The above equations are the foundations for the analysis. The proof rests on equations derived in the following three steps.

- *Step 1:* From (38)–(40) and Assumptions 7-8, it follows that

$$\begin{aligned}\hat{M}_T(c) - M(c) &= G(c)'\widehat{W}_T(c)\{\hat{G}_T(c) - G(c)\} + G(c)'\{\widehat{W}_T(c) - W(c)\}G(c) \\ &\quad + \{\hat{G}_T(c) - G(c)\}'W(c)G(c) + o(\{T/lnlnT\}^{-1/2}) \quad a.s. \\ &\equiv a_T(c) + o(\{T/lnlnT\}^{-1/2}) \qquad (41)\end{aligned}$$

- *Step 2:* Using Dhrymes (1984)[Proposition 89, p.105], we have

$$tr\left\{M(c)^{-1}(\hat{M}_T(c) - M(c))\right\} = vec\{M(c)^{-1}\}'vec\{\hat{M}_T(c) - M(c)\} \qquad (42)$$

From (41), it follows that $a_T(c)$ can be written as

$$\begin{aligned}a_T(c) &= vec\{G(c)'\widehat{W}_T(c)[\hat{G}_T(c) - G(c)]\} + vec\{G(c)'[\widehat{W}_T(c) - W(c)]G(c)\} \\ &\quad + vec\{[\hat{G}_T(c) - G(c)]'W(c)G(c)\} \qquad (43) \\ &= a_{1,T}(c) + a_{2,T}(c) + a_{3,T}(c). \qquad (44)\end{aligned}$$

Taking the terms of the right hand side of (44) in turn, we have

  - $a_{1,T}(c)$:

    From Dhrymes (1984)[Corollary 25, p.103], it follows that

$$\begin{aligned}a_{1,T}(c) &= vec\{G(c)'\widehat{W}_T(c)[\hat{G}_T(c) - G(c)]\} \\ &= \left[I_p \otimes G(c)'\widehat{W}_T(c)\right] vec\{\hat{G}_T(c) - G(c)\} \qquad (45)\end{aligned}$$

    From Assumptions 7-8 and equation (45), it follows that

$$\begin{aligned}a_{1,T}(c) &= -\left[I_p \otimes G(c)'W(c)\right]\left\{\bar{G}(c)\left\{G(c)'W(c)G(c)\right\}^{-1} G(c)'W(c)(\hat{\gamma}_T(c) - g(\theta_0; c))\right\} \\ &\quad + o(\{T/lnlnT\}^{-1/2}) \quad a.s. \qquad (46)\end{aligned}$$

    where the rate follows from the law of iterated logarithms in Assumption 8.

  - $a_{2,T}(c)$:

    From $A^{-1} - B^{-1} = B^{-1}(B - A)A^{-1}$ and Dhrymes (1984)[Corollary 25, p.103], it follows that

$$\begin{aligned}a_{2,T}(c) &= vec\{G(c)'[\widehat{W}_T(c) - W(c)]G(c)\} \\ &= \left[G(c)'\widehat{W}_T(c) \otimes G(c)'W(c)\right] vec\{S(c) - \hat{S}_T(c)\} \qquad (47)\end{aligned}$$

Using (40), (47), and Assumptions 7-8, it follows that

$$\begin{aligned} a_{2,T}(c) &= -\left[G(c)'W(c) \otimes G(c)'W(c)\right] B \left\{\hat{\gamma}(c) - g(\theta_0;c)\right\} \\ &\qquad + o(\{T/lnlnT\}^{-1/2}) \quad a.s. \end{aligned} \tag{48}$$

- $a_{3,T}(c)$:

  From Dhrymes (1984)[Corollary 25, p.103], it follows that

$$\begin{aligned} a_{3,T}(c) &= vec\{[\hat{G}_T(c) - G(c)]'W(c)G(c)\} \\ &= \left[G(c)'W(c) \otimes I_q\right] vec\{\hat{G}_T(c)' - G(c)'\} \\ &= \left[G(c)'W(c) \otimes I_q\right] N\, vec\{\hat{G}_T(c) - G(c)\} \end{aligned} \tag{49}$$

  where N is the permutation matrix such that $vec(A') = Nvec(A)$. It follows from (49) and Assumptions 7-8 that

$$\begin{aligned} a_{3,T}(c) &= \left[G(c)'W(c) \otimes I_q\right] N\overline{G}(c)(\hat{\theta}_T(c) - \theta_0) \\ &\qquad + o((T/lnlnT)^{-1/2}) \quad a.s. \end{aligned} \tag{50}$$

  From Assumptions 7-8 and equation (50), it follows that

$$\begin{aligned} a_{3,T}(c) &= -\left[G(c)'W(c) \otimes I_q\right] N \\ &\qquad \{\, \overline{G}_T(c) \left\{G(c)'W(c)G(c)\right\}^{-1} G(c)'W(c)(\hat{\gamma}(c) - g(\theta_0;c))\} \\ &\qquad + o(\{T/lnlnT\}^{-1/2}) \quad a.s. \end{aligned} \tag{51}$$

- *Step 3:* From Phillips and Ploberger's (2003, p.665) Proposition A8, we have the following Taylor series expansion of $ln[|M|]$ around $M = M_0$ for non-negative definite M, M_0 such that $\|M - M_0\|\|M_0^{-1}\| < 1$,

$$\begin{aligned} ln[|M|] &= ln[|M_0|] + tr\left\{M_0^{-1}(M - M_0)\right\} - tr\left\{(M - M_0)M_0^{-1}(M - M_0)M_0^{-1}\right\} \\ &\qquad + o\left(\frac{\|M^{-1}\|^3\|(M - M_0)\|^3}{1 - \|M^{-1}\|\|M - M_0\|}\right) \end{aligned} \tag{52}$$

  Setting $M = \hat{M}_T(c)$ and $M_0 = M(c)$ and using (41), (42), (44), (46), (48) and (51) we obtain

$$ln[|\hat{M}_T(c)|] = ln[|M(c)|] + tr\left\{M(c)^{-1}(\hat{M}_T(c) - M(c))\right\} + o(\nu_T^{-1}) \quad a.s. \tag{53}$$

where $\{T/lnlnT\}^{1/2}/\nu_T \to 0$ and $\|M - M_0\|\|M_0^{-1}\| < 1$ as $T \to \infty$.

From (53), (42), (46), (48) and (51), it follows that

$$\begin{aligned} ln[|\hat{M}_T(c)|] &= ln[|M(c)|] + vec\{M(c)^{-1}\}'D(c)T^{-1}\sum_{t=1}^{T} h(v_t, \theta_0; c) \\ &\quad + o(\{T/lnlnT\}^{-1/2}) \quad a.s. \end{aligned} \tag{54}$$

where $D(c)' = [D_1(c), D_2(c)]$ and

$$\begin{aligned} D_1(c) &= -\left\{\{[I_p \otimes G(c)'W(c)] + [G(c)'W(c) \otimes I_q]N\}\overline{G}_T(c)\right. \\ &\quad \times \left\{G(c)'W(c)G(c)\right\}^{-1} G(c)'W(c) \end{aligned} \tag{55}$$

$$D_2(c) = -\left[G(c)'W(c) \otimes G(c)'W(c)\right] B \tag{56}$$

Now define $\xi_t(c) = vec\{M(c)^{-1}\}'D(c)h(v_t, \theta_0; c)$ and $\omega_\xi^2(c) = \lim_{T\to\infty} Var[T^{-1/2}\sum_{t=1}^{T} \xi_t(c)]$. Then

$$\begin{aligned} \left(\frac{T}{lnlnT}\right)^{1/2} RIRSC(c) &= -\left(\frac{T}{lnlnT}\right)^{1/2} ln[|M(c)|] - \left(\frac{T}{lnlnT}\right)^{1/2} T^{-1}\sum_{t=1}^{T} \xi_t(c) \\ &\quad + \left(\frac{T}{lnlnT}\right)^{1/2} \kappa(|c|, T) + o(1) \quad a.s. \end{aligned} \tag{57}$$

We now use the above results to establish Theorem 2. The proof proceeds by considering two cases.

Part (i): Consider c_1 and c_2 such that $V_\theta(c_1) - V_\theta(c_2)$ is $p.s.d.$ and hence $ln[|M(c_2)|] > ln[|M(c_1)|]$. Since $\kappa(|c|, T) = o(T^{1/2})$ from Assumption 6, it follows from (57) and Assumption 8 that

$$\left(\frac{T}{lnlnT}\right)^{\frac{1}{2}} [RIRSC(c_1) - RIRSC(c_2)] = \left(\frac{T}{lnlnT}\right)^{\frac{1}{2}} (ln[|M(c_2)|] - ln[|M(c_1)|]) + o(1) \quad a.s. \tag{58}$$

Since $V_\theta(c) - V_\theta(c_r)$ is p.s.d. for all $c \in C_H$, it follows from (58) that $RIRSC(c) \geq RIRSC(c_\mathcal{E})$ a.s. for any $c \in C_H$ and $c_\mathcal{E} \in C_{\mathcal{E},H}$. Because $RIRSC(c) \geq RIRSC(\hat{c}_T)$ holds for any $c \in C_H$ by definition of $\hat{c}_T$, it has to be the case that $\hat{c}_T \in C_{\mathcal{E},H}$ a.s. for T sufficiently large.

Part (ii): Consider $c_a \in \mathcal{C}_{\mathcal{E},H}$ such that $c_r \neq c_a$. From Assumption 8, it follows that for $c = c_r, c_a$ we have

$$\limsup_{T\to\infty} \left(\frac{T}{lnlnT}\right)^{1/2} |T^{-1}\sum_{t=1}^{T}\xi_t(c)| \leq 2^{1/2}\omega_\xi(c),\ a.s. \tag{59}$$

Set $r(c_r, c_a) = RIRSC(c_r) - RIRSC(c_a)$. Since $M(c_r) = M(c_a)$ by definition in this case, it follows from (57) and (59) that

$$\begin{aligned}\limsup_{T\to\infty} \left(\frac{T}{lnlnT}\right)^{1/2} r(c_r, c_a) &\leq \limsup_{T\to\infty} \left(\frac{T}{lnlnT}\right)^{1/2} |T^{-1}\sum_{t=1}^{T}\xi_t(c_r)| + \limsup_{T\to\infty} \left(\frac{T}{lnlnT}\right)^{1/2} |T^{-1}\sum_{t=1}^{T}\xi_t(c_a)| \\ &\quad - \liminf_{T\to\infty} \left(-\kappa(|c_r|, T) + \kappa(|c_a|, T)\right\} \\ &\leq 2^{1/2}\left(\omega_\xi(c_r) + \omega_\xi(c_a)\right) + \\ &\quad - \liminf_{T\to\infty} \left\{ k(|c_a|) - k(|c_r|)(T^{1/2}m_T/(lnlnT)^{1/2}) \right\}\end{aligned} \tag{60}$$

Using Assumptions 8 and 3, it follows from (60) that $\hat{c}_T = c_r$ for T sufficiently large *a.s.* if Assumption 6 holds with (i) and $z = 2^{1/2}(\omega_\xi(c_r) + \omega_\xi(c_a))/\bar{k}$ where $\bar{k} = min_{c\in\mathcal{C}_{\mathcal{E},H}} k(|c|) - k(|c_r|)$ or (ii).

Proof of Theorems 5 and 6: The proofs are as in Theorem 1.

Appendix C: Auxiliary Results for Simulation-Based Estimators

Theorem 7 (Asymptotic normality of simulation-based estimators) *Under Assumption 9(a)–(m), $\hat{\theta}_{SCN,T}(c)$ is consistent and is asymptotically normally distributed:*

$$\sqrt{T}(\hat{\theta}_{SCN,T}(c) - \theta_0) \ \xrightarrow{d}\ N(0, V_{\theta,SCN}(c)) \tag{61}$$

where

$$\begin{aligned}V_{\theta,SCN}(c) &= \left(1 + \frac{1}{S}\right) \lim_{p_T\to\infty} (\Gamma_T(\theta_0;c)'W_T(c)\Gamma_T(\theta_0;c))^{-1}\Gamma_T(\theta_0;c)'W_T(c)\Sigma_{g(\theta_0;c,p_T)}W_T(c)\Gamma_T(\theta_0;c) \\ &\quad \times(\Gamma_T(\theta_0;c)'W_T(c)\Gamma_T(\theta_0;c))^{-1}.\end{aligned}$$

and $\Gamma_T(\theta;c) = \partial g_T(\theta;c)/\partial\theta$ and $W_T(c)$ is defined in Assumption (9).

Proof of Theorem 7: By Theorem 2.1 of Newey and McFadden (1994), for a given horizon h, $\hat{\theta}_{SCN,T}(c)$ is consistent if: (i) $Q(\theta;c) \equiv \lim_{p_T\to\infty} Q_T(\theta;c)$, where

$$Q_T(\theta;c) \equiv (g_T(\theta_0;c) - g_T(\theta;c))'W_T(c)(g_T(\theta_0;c) - g_T(\theta;c)),$$

is uniquely minimized at θ_0, and where $g_T(\theta; c)$ is defined in Assumption 9(c); (ii) Θ is compact; (iii) $Q(\theta; c)$ is continuous; and (iv) $\hat{Q}_T(\theta; c)$ (the objective function evaluated at the estimated parameters and at the estimated weighting function) converges uniformly in probability to $Q(\theta; c)$. By Assumptions 9(d,k), $Q(\theta; c)$ is uniquely minimized at θ_0. By Assumption 9(b), Θ is compact. By Assumptions 9(b,f,k),

$$\begin{aligned}
& |Q_T(\theta_1; c) - Q_T(\theta_2; c)| \\
= \ & |(g_T(\theta_1; c) - g_T(\theta_2; c))' W_T(c) g_T(\theta_1; c) + g_T(\theta_2; c)' W_T(c) g_T(\theta_1; c) - g_T(\theta_2; c))| \\
\leq \ & \left[(g_T(\theta_1; c) - g_T(\theta_2; c))' W_T(c)(g_T(\theta_1; c) - g_T(\theta_2; c))\right]^{\frac{1}{2}} (Q_T(\theta_1; c)^{\frac{1}{2}} + Q_T(\theta_2; c)^{\frac{1}{2}}) \\
\leq \ & K \| g_T(\theta_1; c) - g_T(\theta_2; c) \|, \qquad (62)
\end{aligned}$$

where K is a constant that does not depend on T, the first inequality follows from Cauchy-Schwartz and the last inequality follows by the continuity of the objective function, the compactness of the parameter space and the fact that the eigenvalues of $W_T(c)$ are bounded. Since

$$|Q(\theta_1; c) - Q(\theta_2; c)| \ \leq \ K \lim_{T \to \infty} \| g_T(\theta_1; c) - g_T(\theta_2; c) \|,$$

it follows from Assumption 9(f) that $Q(\theta; c)$ is continuous in θ. To show that

$$\hat{Q}_T(\theta; c) \equiv (\hat{\gamma}_T(c) - \tilde{g}_T(\theta; c))' \hat{W}_T(c)(\hat{\gamma}_T(c) - \tilde{g}_T(\theta; c)),$$

uniformly converges in probability to $Q(\theta; c)$, we need to show pointwise convergence and stochastic equicontinuity of $\hat{Q}_T(\theta; c)$. The pointwise convergence of $\hat{Q}_T(\theta; c)$ to $Q(\theta; c)$ follows from Assumptions 9(c,h,k). The stochastic equicontinuity of $\hat{Q}_T(\theta; c)$ follows from the Lipschitz condition in Assumption 9(f). By the uniform law of large number (e.g., Theorem 21.9 of Davidson, 1994, p.337), $\hat{Q}_T(\cdot; c)$ converges uniformly in probability to $Q(\cdot; c)$. Therefore, $\hat{\theta}_{SCN,T}(c)$ converges in probability to θ_0.

Since $\hat{\theta}_{SCN,T}(c) \xrightarrow{p} \theta_0$, it follows from the first-order condition and the mean-value theorem that

$$\begin{aligned}
\sqrt{T}(\hat{\theta}_{SCN,T}(c) - \theta_0) \ = \ & (\Gamma(\hat{\theta}_{SCN,T}(c); c)' \hat{W}_T(c) \Gamma(\bar{\theta}_{SCN,T}(c); c))^{-1} \\
& \times \Gamma(\hat{\theta}_{SCN,T}(c); c)' \hat{W}_T(c) [\sqrt{T}(\widehat{\gamma}_T(c) \\
& - g_T(\theta_0; c) - \sqrt{T}(\widetilde{g}_T(\theta_0; c) - g_T(\theta_0; c))],
\end{aligned}$$

where $\bar{\theta}_{SCN,T}(c)$ is a point between $\hat{\theta}_{SCN,T}(c)$ and θ_0. Then, by Assumptions 9(c,f,g,h,k,l,m) and the central limit theorem, we have

$$\sqrt{T}(\hat{\theta}_{SCN,T}(c) - \theta_0) \xrightarrow{d} N(0, V_{\theta,SCN}(c)),$$

where

$$\begin{aligned} V_{\theta,SCN}(c) &= \lim_{T\to\infty} (\Gamma_T(\theta_0;c)'W_T(c)\Gamma_T(\theta_0;c))^{-1}\Gamma_T(\theta_0;c)'W_T(c)\,(1+1/S) \\ &\quad \times \Sigma_{g_T(\theta_0;c)} W_T(c)\Gamma_T(\theta_0;c)(\Gamma_T(\theta_0;c)'W_T(c)\Gamma_T(\theta_0;c))^{-1}. \end{aligned} \tag{63}$$

Theorem 8 (Estimation of asymptotic variance of simulation-based estimators) *Let* $\Sigma^{(s)}_{g_T(\theta;c)}$ *denote the estimated asymptotic covariance matrix of the simulated impulse responses* $\widetilde{g}^s_T(\theta;c)$*, and* $\hat{\Sigma}_{\widehat{\gamma}_T(c)}$ *denote the estimate of the asymptotic covariance matrix of the sample impulse responses. Let*

$$\hat{W}_T(c) = \left(\hat{\Sigma}_{\widehat{\gamma}_T(c)} + \frac{1}{S^2} \sum_{s=1}^{S} \tilde{\Sigma}^{(s)}_{g(\tilde{\theta}_{SCN,T}(c);c)} \right)^{-1},$$

where $\tilde{\theta}_{SCN,T}(c)$ *is an estimator of* θ *(e.g., the SCN estimator with* $\hat{W}_T(c)$ *equal to the identity matrix). Then,*

$$\begin{aligned} \hat{V}_{SCN,T}(c) &= (\widehat{\Gamma}_T(\tilde{\theta}_{SCN,T}(c);c)'\hat{W}_T(c)\widehat{\Gamma}_T(\tilde{\theta}_{SCN,T}(c);c))^{-1} \\ &\quad \times \widehat{\Gamma}_T(\tilde{\theta}_{SCN,T}(c);c)'\hat{W}_T(c) \left(\hat{\Sigma}_{\widehat{\gamma}_T(c)} + \frac{1}{S^2}\sum_{s=1}^{S} \tilde{\Sigma}^{(s)}_{g(\tilde{\theta}_{SCN,T}(c);c)} \right) \hat{W}_T(c)\widehat{\Gamma}_T(\tilde{\theta}_{SCN,T}(c);c) \\ &\quad \times (g_T(\tilde{\theta}_{SCN,T}(c),c)'\hat{W}_T(c)\widehat{\Gamma}_T(\tilde{\theta}_{SCN,T}(c);c))^{-1} \end{aligned} \tag{64}$$

where $\widehat{\Gamma}_T(\theta;c) \equiv \frac{1}{S}\sum_{s=1}^{S} \Gamma^{(s)}_T(\theta;c)$*, for* $\Gamma^{(s)}_T(\theta;c) \equiv \partial\widetilde{g}^{(s)}_T(\theta;c)/\partial\theta'$*. Under Assumption 9,*

a. If $\tilde{\theta}_{SCN}(c)$ *is* $\sqrt{T}$*-consistent, with variance* $\hat{V}_{SCN,T}(c) = V_{\theta,SCN}(c) + O_p(p_T^2/T)$*, where* $V_{\theta,SCN}(c)$ *is defined in Theorem 7.*

b. If $c \in C_{NI,H}$*,* $|\hat{V}^{-1}(c)| \xrightarrow{p} \infty$*.*

When the weighting matrix $\hat{W}_T(c) = \left(\hat{\Sigma}_{\widehat{\gamma}_T(c)} + \frac{1}{S^2}\sum_{s=1}^{S} \widetilde{\Sigma}^{(s)}_{g(\widehat{\theta}_{SCN,T}(c);c)} \right)^{-1}$ *is used, the variance of* $\hat{\theta}_{SCN,T}(c)$ *can be estimated by:*

$$\hat{V}_{SCN,T}(c) = (\widehat{\Gamma}_T(\hat{\theta}_{SCN,T}(c);c)'\hat{W}_T(c)\widehat{\Gamma}_T(\hat{\theta}_{SCN,T}(c);c))^{-1} \tag{65}$$

Proof of Theorem 8: (a). The consistency of the covariance matrix estimator follows from Assumptions 9(c)(e)(f)(g)(k) and Theorem (7). The convergence rate follows from Assumptions 9(a,h).

(b). Since the rank of $\Gamma_T(\hat{\theta}_{SCN,T}(c);c)$ is less than full for $c \in C_{NI,H}$, $|\hat{V}^{-1}_{SCN,T}(c)| \xrightarrow{p} 0$, where is defined in (64) or (65).

9 Tables

Table 1(a). Empirical results (ACEL, 2005))

	RIRSC ($\hat{h}_T = 3$)		Fixed lags (h=20)	
Parameters	*Parameter Estimates*	*Standard Errors*	*Parameter Estimates*	*Standard Errors*
ρ_{xM}	-0.097	0.247	-0.040	0.292
ρ_{xz}	0.588	1.257	0.329	0.948
c_z	0.655	0.664	2.952	3.096
ρ_{μ_z}	0.237	0.703	0.894	0.159
$\rho_{x\Upsilon}$	0.997	0.107	0.822	0.345
c_Υ	0.307	0.435	0.247	0.440
$\rho_{\mu\Upsilon}$	0.344	0.240	0.239	0.425
σ_M	0.334	0.113	0.333	0.110
σ_{μ_z}	0.203	0.168	0.069	0.068
$\sigma_{\mu\Upsilon}$	0.287	0.084	0.304	0.093
ε	0.831	0.284	0.809	0.256
S''	6.907	9.842	3.350	3.477
ξ_w	0.832	0.225	0.713	0.261
b	0.779	0.124	0.706	0.135
σ_a	0.413	0.777	2.029	4.251
c_z^p	0.144	1.414	1.379	3.732
c_Υ^p	0.073	0.580	0.137	0.499
γ	0.207	0.434	0.039	0.069

Table 1(b). Implied Average Time Between Re-Optimization (ACEL, 2005)

	RIRSC ($\hat{h}_T = 3$)	Fixed lags (h=20)
Firm-Specific Capital Model	1.294	1.515
	(0.037)	(0.007)
Homogeneous Capital Model	2.769	5.655
	(0.167)	(0.046)

Table 2. Empirical results (CEE, 2005)

	RIRSC ($\hat{h}_T = 6$)		*Fixed Lags (h=20)*	
	Parameter	*Standard*	*Parameter*	*Standard*
Parameters	*Estimates*	*Errors*	*Estimates*	*Errors*
ρ_M	*-0.020*	*0.300*	*-0.114*	*0.272*
σ_M	*0.348*	*0.108*	*0.352*	*0.108*
ϵ	*0.897*	*0.275*	*0.836*	*0.255*
S''	*3.732*	*3.695*	*4.324*	*4.566*
ξ_w	*0.624*	*0.194*	*0.645*	*0.261*
b	*0.762*	*0.127*	*0.717*	*0.144*
λ_f	*1.002*	*0.231*	*1.097*	*0.277*
σ_a	*0.001*	*0.152*	*0.041*	*0.557*
γ	*0.106*	*0.243*	*0.208*	*0.546*

Note to Tables 1-2. The tables report parameter estimates and their standard errors for the IRFME with 20 *lags for each IRF, and the IRFME with* h *chosen according to the RIRSC (3), which selects h=3 for ACEL and h=6 for CEE. The notation is the same as that in Tables 2 and 3 in ACEL, and* λ_f *is calibrated to be 1.01. See ACEL for a complete description. The CEE model is a special case of ACEL when only monetary shocks are considered; for consistency, we maintain the same notation as ACEL, Tables 2 and 3.*

Table 3. Monte Carlo results for the AR(1) case.

H	IRFME		$IRFME_{RIRSC}$	
	bias	*rej. rate*	*bias*	*rej. rate*
1	0.0010	0.0531	0.0010	0.0511
5	-0.0243	0.2265	-0.0045	0.0521
10	-0.0135	0.4090	-0.0036	0.0442
20	0.0026	0.6194	-0.0072	0.0473
50	-0.0768	0.6815	-0.0480	0.0506
100	-0.0819	0.6236	-0.0451	0.0577

Note to Tables 3. The tables reports bias (i.e. true parameter value minus estimated value) and rejection rates of 95% nominal confidence intervals for examples AR(1).

***Table 4a. Bias, Variance, Coverage Probability** ($T = 100$)*						
	α			γ_1		
h	*bias*	*var*	*prob*	*bias*	*var*	*prob*
1	*0.058*	*0.007*	*0.925*	*0.049*	*0.006*	*0.928*
2	*0.056*	*0.008*	*0.887*	*0.042*	*0.005*	*0.904*
3	*0.040*	*0.004*	*0.781*	*0.051*	*0.006*	*0.815*
4	*0.033*	*0.003*	*0.655*	*0.044*	*0.005*	*0.709*
5	*0.044*	*0.005*	*0.607*	*0.052*	*0.007*	*0.677*
6	*0.035*	*0.003*	*0.586*	*0.045*	*0.005*	*0.626*
7	*0.048*	*0.005*	*0.539*	*0.051*	*0.007*	*0.584*
8	*0.041*	*0.004*	*0.496*	*0.045*	*0.005*	*0.535*
9	*0.048*	*0.005*	*0.455*	*0.053*	*0.007*	*0.510*
10	*0.040*	*0.004*	*0.429*	*0.047*	*0.005*	*0.481*
11	*0.048*	*0.006*	*0.376*	*0.054*	*0.008*	*0.436*
12	*0.041*	*0.004*	*0.354*	*0.046*	*0.006*	*0.408*

***Table 4b. Bias, Variance, Coverage Probability** ($T = 200$)*						
	α			γ_1		
h	*bias*	*var*	*prob*	*bias*	*var*	*prob*
1	*0.038*	*0.003*	*0.943*	*0.029*	*0.002*	*0.948*
2	*0.039*	*0.003*	*0.942*	*0.025*	*0.002*	*0.946*
3	*0.026*	*0.001*	*0.851*	*0.030*	*0.002*	*0.857*
4	*0.021*	*0.001*	*0.737*	*0.026*	*0.002*	*0.783*
5	*0.030*	*0.002*	*0.711*	*0.031*	*0.003*	*0.749*
6	*0.024*	*0.001*	*0.685*	*0.025*	*0.003*	*0.731*
7	*0.030*	*0.002*	*0.649*	*0.031*	*0.003*	*0.712*
8	*0.024*	*0.001*	*0.602*	*0.026*	*0.003*	*0.677*
9	*0.029*	*0.002*	*0.568*	*0.032*	*0.003*	*0.649*
10	*0.025*	*0.002*	*0.566*	*0.026*	*0.003*	*0.620*
11	*0.030*	*0.002*	*0.533*	*0.032*	*0.003*	*0.613*
12	*0.025*	*0.002*	*0.503*	*0.026*	*0.003*	*0.592*

***Table 4c. Bias, Variance, Coverage Probability** ($T = 400$)*						
	α			γ_1		
h	*bias*	*var*	*prob*	*bias*	*var*	*prob*
1	*0.026*	*0.001*	*0.952*	*0.021*	*0.001*	*0.951*
2	*0.025*	*0.002*	*0.951*	*0.017*	*0.001*	*0.950*
3	*0.018*	*0.001*	*0.871*	*0.021*	*0.001*	*0.886*
4	*0.015*	*0.001*	*0.784*	*0.017*	*0.001*	*0.829*
5	*0.020*	*0.001*	*0.761*	*0.021*	*0.001*	*0.804*
6	*0.015*	*0.001*	*0.740*	*0.017*	*0.001*	*0.778*
7	*0.020*	*0.001*	*0.709*	*0.022*	*0.001*	*0.771*
8	*0.016*	*0.001*	*0.682*	*0.017*	*0.001*	*0.745*
9	*0.020*	*0.001*	*0.658*	*0.022*	*0.001*	*0.723*
10	*0.016*	*0.001*	*0.634*	*0.017*	*0.001*	*0.720*
11	*0.020*	*0.001*	*0.607*	*0.022*	*0.001*	*0.705*
12	*0.016*	*0.001*	*0.571*	*0.017*	*0.001*	*0.691*

Table 5a. Median Absolute Bias, Variance, Coverage Probability (α)													
		All			*AIC*			*SIC*			*HQC*		
T	*H*	*bias*	*var*	*prob*	*bias*	*var*	*prob*	*bias*	*var*	*prob*	*bias*	*var*	*prob*
100	*12*	*0.041*	*0.004*	*0.354*	*0.047*	*0.006*	*0.679*	*0.046*	*0.005*	*0.714*	*0.044*	*0.005*	*0.738*
200	*12*	*0.025*	*0.002*	*0.503*	*0.031*	*0.002*	*0.728*	*0.030*	*0.002*	*0.778*	*0.030*	*0.002*	*0.775*
400	*12*	*0.016*	*0.001*	*0.571*	*0.020*	*0.001*	*0.759*	*0.020*	*0.001*	*0.804*	*0.020*	*0.001*	*0.795*

Table 5b. Median Absolute Bias, Variance, Coverage Probability (γ_1)													
		All			*AIC*			*SIC*			*HQC*		
T	*H*	*bias*	*var*	*prob*	*bias*	*var*	*prob*	*bias*	*var*	*prob*	*bias*	*var*	*prob*
100	*12*	*0.046*	*0.006*	*0.408*	*0.037*	*0.004*	*0.722*	*0.036*	*0.003*	*0.755*	*0.035*	*0.003*	*0.776*
200	*12*	*0.026*	*0.003*	*0.592*	*0.024*	*0.002*	*0.785*	*0.024*	*0.002*	*0.818*	*0.024*	*0.002*	*0.816*
400	*12*	*0.017*	*0.001*	*0.691*	*0.016*	*0.001*	*0.813*	*0.016*	*0.001*	*0.843*	*0.016*	*0.001*	*0.833*

Notes to Tables 4 and 5. The tables report the median of absolute bias, variance and the coverage probability of 95% nominal confidence intervals for (1).

Table 6. Selected Number of Impulse Responses										
		AIC			*SIC*			*HQC*		
T	*H*	*Mean*	*Mode*	*Var*	*Mean*	*Mode*	*Var*	*Mean*	*Mode*	*Var*
100	*12*	*3.477*	*3.000*	*4.514*	*3.033*	*2.000*	*2.833*	*2.838*	*2.000*	*2.186*
200	*12*	*3.618*	*4.000*	*3.363*	*2.982*	*2.000*	*1.803*	*3.012*	*2.000*	*1.888*
400	*12*	*3.972*	*4.000*	*3.537*	*3.069*	*2.000*	*1.706*	*3.278*	*3.000*	*2.091*

Notes to Table 6. The table reports the mean, median and variance of selected horizons of impulse responses for (1).