

INFORMATION CRITERIA FOR IMPULSE RESPONSE  
FUNCTION MATCHING ESTIMATION OF DSGE MODELS

Alastair Hall	Atsushi Inoue	James M Nason	Barbara Rossi
University of Manchester	NCSU	Federal Reserve Bank of Philadelphia	Duke University

September 23, 2009

Abstract: We propose new information criteria for impulse response function matching estimators (IRFMEs). These estimators yield sampling distributions of the structural parameters of dynamic stochastic general equilibrium (DSGE) models by minimizing the distance between sample and theoretical impulse responses. First, we propose an information criterion to select only the responses that produce consistent estimates of the true but unknown structural parameters: the Valid Impulse Response Selection Criterion (VIRSC). The criterion is especially useful for mis-specified models. Second, we propose a criterion to select the impulse responses that are most informative about DSGE model parameters: the Relevant Impulse Response Selection Criterion (RIRSC). These criteria can be used in combination to select the subset of valid impulse response functions with minimal dimension that yields asymptotically efficient estimators. The criteria are general enough to apply to impulse responses estimated by VARs, local projections, and simulation methods. We show that the use of our criteria significantly affects estimates and inference about key parameters of two well-known new Keynesian DSGE models. Monte Carlo evidence indicates that the criteria yield gains in terms of finite sample bias as well as offering tests statistics whose behavior is better approximated by first order asymptotic theory. Thus, our criteria improve on existing methods used to implement IRFMEs.

J.E.L. Codes: C32, E47, C52, C53.

Acknowledgements. We thank the editors, two anonymous referees, J. Boivin, C. Burnside, Jeff Campbell, P. Hansen, Ò. Jordà, J. Lindé, S. Mavroedis, F. Schorfheide and seminar participants at the 2006 EC<sup>2</sup>, 2007 SED, EMSG at Duke U., the UBC Macro Lunch, the 2007 Conference in Waterloo, the 2007 ESEM, Kyoto U., Osaka U., Brown U., Tilburg U., U. of Amsterdam and U. of Cincinnati for comments and C. Burnside and L. Christiano for sharing their codes. A. Inoue acknowledges financial support by the SSHRC under project 410-2007-1975. The views in this paper represent those of the authors and are not those of either the Federal Reserve Bank of Atlanta, the Federal Reserve System, or any of its staff.

## 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 estimation (IRFME) 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, CEE hereafter), Altig, Christiano, Eichenbaum, and Lindé (2005, ACEL hereafter), Iacoviello (2005), Jordà and Kozicki (2007), DiCecio (2005), Boivin and Gianoni (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 methods have been used to choose which IRFs and how many lags to match.

In a recent paper, Dridi et al (2007) present a comprehensive statistical framework for analyzing econometric estimation of DSGE in which the parameters of a structural economic model are estimated by matching moments using a binding function obtained from an instrumental model. Using this terminology in our context, the DSGE is the structural model, the VAR is the instrumental model and the impulse response function is the binding function. Their framework allows the model to be mis-specified and distinguishes the parameters into three categories: parameters of interest, estimated nuisance parameters and calibrated nuisance parameters. Since the structural model may be mis-specified, an important issue is whether the structural model encompasses - or partially encompasses - the instrumental model in the sense that estimation based on the binding function nevertheless yields consistent estimators of the parameters of interest. As well as introducing this conceptual framework, Dridi et al (2007) derive first order asymptotic properties of (Partial) Indirect Inference estimators within this set-up, and propose a statistic for testing whether the structural model (potentially) encompasses the instrumental model.

Dridi et al's (2007) discussion of encompassing highlights the importance of the choice of binding function within this type of estimation. As noted by Dridi et al, the choice of the binding function is analogous to the choice of moment function in moment based estimation. In the literature on moment based estimation, it is recognized that two aspects of the choice of moment function are important: (i) valid moment conditions are needed to obtain consistent estimation: (ii) not all valid moment conditions are informative and that it may be desirable in terms of finite sample

properties to base the estimation on a subset of valid moment conditions. Exploiting the analogy between moment conditions and binding functions, it can be seen that Dridi et al's encompassing test addresses (i) above; however, their analysis does not address the sequential testing issues that arise if the statistic is used repeatedly as part of a specification search. To our knowledge, (ii) above has not been considered in the context of IRFME, where it is common to use IRF up to a relatively large, pre-specified number of lags.

In this paper, we address the issue of which impulse response functions to use in impulse response function matching estimation. Working within the framework provided by Dridi et al (2007), we propose two new criteria: the Valid Impulse Response Selection Criterion (VIRSC) and the Relevant Impulse Response Selection Criterion (RIRSC). The VIRSC is designed to determine which impulse response functions to include in order to obtain consistent estimators of the parameters of interest. The RIRSC is designed to select from the set of valid impulse responses those which are informative about the parameters of interest by excluding those that are redundant. We provide conditions under which the two criteria are weakly and strongly consistent, and report simulation evidence that shows the sensitivity of IRFME to the choice of impulse responses and also the efficacy of our criteria.

The RIRSC criterion is applied to the DSGE models of CEE and ACEL. We often obtain point estimates that are little changed from those CEE and ACEL report. Nonetheless, the RIRSC yields economically important changes in inference regarding several key parameters that lead to strikingly different conclusions than those of CEE and ACEL. We conjecture that the parameter estimates in CEE and ACEL may be subject to small sample biases, and we investigate this issue in Monte Carlo experiments. The Monte Carlo exercises indicate that, in general, the small sample bias of IRFMEs is mitigated by RIRSC compared to using a relatively large fixed lag length, and that the VIRSC works well in small samples. Thus, the criteria that we propose 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.

As mentioned above, the framework considered in this paper is inspired by the work of Dridi et al (2007), and we believe that our results both complement and extend their analysis in the following ways. We propose information criteria for the selection of both valid and relevant responses, whereas they focus more on hypothesis testing and model selection. We focus on commonly used

methodologies for IRFME, including but not limited to simulation-based estimators, whereas they focus on general simulation-based estimators. Our criteria can be used not only for IRFMEs but also for general classical minimum distance and Indirect Inference Estimators. Relative to this literature, and in particular relative to Dridi et al. (2007), we add useful information criteria to select valid as well as relevant restrictions, thus significantly extending the scope of their analysis. Our approach is also connected to the literature on moment selection, with VIRSC and RIRSC being extensions to Classical Minimum Distance estimators of criteria proposed by Andrews (1999) and Hall et al (2007). The criteria that we propose are also 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. Jordà and Kozicki (2007) show that our RIRSC 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.

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 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, so that the sample IRFs are informative for the DSGE model parameters. We are interested in the VAR:

$$Y_t = \Psi_0 + \Psi_1 Y_{t-1} + \Psi_2 Y_{t-2} + \dots + \Psi_{p_0} Y_{t-p_0} + \varepsilon_t, \quad (1)$$

where  $Y_t = [Y_{1,t}, Y_{2,t}, \dots, Y_{n_Y,t}]'$  is  $n_Y \times 1$ ,  $t = 1, 2, \dots, T$ , and  $\varepsilon_t = [\varepsilon_{1,t}, \varepsilon_{2,t}, \dots, \varepsilon_{n_Y,t}]'$  is white noise with zero mean and variance  $\Sigma_\varepsilon$ . The population VAR lag order is  $p_0$ . For (1) to have an infinite order Vector Moving Average (VMA) representation and IRFs, we make the following standard assumption:<sup>1</sup>

**Assumption (A0).** *In eq. 1,  $\Psi(L) = I_{n_Y} - \Psi_1 L - \Psi_2 L^2 - \dots - \Psi_{p_0} L^{p_0}$  is invertible, where  $L$  is the lag operator and  $I_{n_Y}$  is the  $(n_Y \times n_Y)$  identity matrix and  $p_0$  is finite.*

The model's parameters are  $(\theta, \alpha, \phi)$ :  $\theta$  is a  $p_\theta \times 1$  vector of structural parameters of interest,  $\alpha$  is a  $p_\alpha \times 1$  vector of estimated nuisance parameters and  $\phi$  is a  $p_\phi \times 1$  vector of calibrated nuisance parameters.<sup>2</sup>  $\theta$ ,  $\alpha$  and  $\phi$  belong to  $\Theta \subset \mathbb{R}^{p_\theta}$ ,  $A \subset \mathbb{R}^{p_\alpha}$  and  $\Xi \subset \mathbb{R}^{p_\phi}$ . Let  $\gamma$  denote a vector of population IRFs with maximum horizon  $H$  estimated from data  $\underline{Y}_T \equiv [Y_1', \dots, Y_T']'$ . In particular, we will use  $\gamma_{i,j,\tau}$  to denote IRFs of each variable  $Y_{i,t+\tau}$  to a structural shock  $\varepsilon_{j,t}$  at horizon  $\tau$ , where  $i, j = 1, \dots, n_Y$ ,  $\tau = 1, \dots, H$ .<sup>3</sup> Let  $\underline{\gamma}_\tau$  be a  $(n_Y^2 \times 1)$  vector that collects the population IRFs at a particular horizon  $\tau$ :

$$\underline{\gamma}_\tau = \left( \overbrace{\gamma_{1,1,\tau}, \gamma_{1,2,\tau}, \dots, \gamma_{1,n_Y,\tau}}^{i=1}, \overbrace{\gamma_{2,1,\tau}, \gamma_{2,2,\tau}, \dots, \gamma_{2,n_Y,\tau}}^{i=2}, \dots, \overbrace{\gamma_{n_Y,1,\tau}, \dots, \gamma_{n_Y,n_Y,\tau}}^{i=n_Y} \right)'$$

The population IRFs at horizons  $\tau = 1, 2, \dots, H$  can be further collected in the  $(n_Y^2 H \times 1)$  vector  $\gamma$  so that  $\gamma = \left( \underline{\gamma}'_1, \underline{\gamma}'_2, \dots, \underline{\gamma}'_H \right)'$ . We will use  $\gamma_0$  to denote the true value of  $\gamma$ .

<sup>1</sup>Note that even if the true theoretical model has a VAR( $\infty$ ) representation, our results still apply for the Indirect Inference estimator that we discuss in Section 4.

<sup>2</sup>Our  $\theta$ ,  $\alpha$  and  $\phi$  correspond to  $\theta_{11}$ ,  $\theta_{21}$  and  $\theta_{22}$  of Dridi et al. (2007).

<sup>3</sup>For simplicity we assume that the dimension of  $y_t$ ,  $n_y$ , and that of  $\varepsilon_t$ ,  $n_\varepsilon$ , are equal. However,  $n_y$  can be greater than  $n_\varepsilon$ . 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  $18 \times 18$  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  $18 \times 8$  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.

Finally, let  $g(\theta, \alpha, \phi)$  be a  $n_Y^2 H \times 1$  vector of impulse responses obtained by a DSGE model. These impulse responses are the binding function between the structural parameters  $\theta, \alpha, \phi$  and the impulse responses  $\gamma$ . The structural parameters  $\theta$  and  $\alpha$  are estimated from

$$\gamma = g(\theta, \alpha, \bar{\phi})$$

where  $\bar{\phi}$  is calibrated.

The question that we address in this paper is which subset of  $\gamma$  and  $g$  should be used. To address this question, we introduce selection vectors that choose subsets of impulse responses.

**Definition of  $c$ .** *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,  $\gamma(c)$  and  $g(\cdot; c)$ . If  $c_j = 1$  then the  $j^{\text{th}}$  element of  $\gamma$  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  $Y_t$  with respect to the first element of  $\varepsilon_t$  are used,  $c = 1_{H \times 1} \otimes [0_{1 \times n_Y} \ 1 \ 0_{1 \times (n_Y-1)} \ 0_{1 \times (n_Y-2)n_Y}]'$ . 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

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

The methodologies proposed in this paper are valid for general minimum distance estimators as well as indirect inference estimators. In particular, a special case on which we focus is the IRF Matching Estimator (IRFME). The IRFME is a classical minimum distance estimator such that:

$$\begin{pmatrix} \hat{\theta}_T(\bar{\phi}; c) \\ \hat{\alpha}_T(\bar{\phi}; c) \end{pmatrix} = \arg \min_{\theta \in \Theta, \alpha \in A} [\hat{\gamma}_T(c) - g(\theta, \alpha, \bar{\phi}; c)]' \hat{\Omega}_T(c) [\hat{\gamma}_T(c) - g(\theta, \alpha, \bar{\phi}; c)], \quad (2)$$

where  $\hat{\gamma}_T(c)$  is an estimate of  $\gamma(c)$  and  $\hat{\Omega}_T(c)$  is a weighting matrix.  $\hat{\Omega}_T(c)$  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, which displays the variances of the IRFs. In general,  $\hat{\Omega}_T(c)$  can be readily obtained from standard package procedures that compute IRF standard error bands.<sup>4</sup>

<sup>4</sup>In this paper, we focus on optimal weighting matrix estimators because the VIRSC is based on the overidentifying restrictions test statistic, although both methods can be extended to non-optimal weighting matrices.

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 are mis-specified and/or that contain only redundant information; at the same time, we enable the researcher 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:

We make the following Assumptions, which are slight modifications of the assumptions in Dridi et al. (2007):

**Assumption (A1).** (a)  $\Theta \times \alpha$  is non-empty and compact in  $\mathbb{R}^{p_\theta} \times \mathbb{R}^{p_\alpha}$ . (b) The parameter value  $[\theta'_0, \bar{\alpha}']'$  belongs to the interior of  $\Theta \times \alpha$ .  $\theta_0$  is the true parameter value whereas  $\bar{\alpha}$  is the pseudo-true parameter value. (c) There is a function  $Q_T(\underline{Y}_T, \gamma(c); c)$  that is twice continuously differentiable in  $\gamma$  such that

$$\hat{\gamma}_T(c) = \gamma_0(c) - \left[ \frac{\partial^2 Q_T(\underline{Y}_T, \gamma_0(c); c)}{\partial \gamma(c) \partial \gamma'(c)} \right]^{-1} \frac{\partial Q_T(\underline{Y}_T, \gamma_0(c); c)}{\partial \gamma(c)} + o_p(1).$$

**Assumption (A2).** The true parameter value  $\theta_0 \in \Theta$  is a unique solution to  $\max_{\theta \in \Theta} \max_{c_0, H \in C_{0, H}(\theta, \bar{\phi})} |c_{0, H}|$  where

$$C_{0, H}(\theta, \bar{\phi}) = \{c \in C_H : \gamma(c) = g(\theta, \alpha, \bar{\phi}) \text{ for some } \alpha \in A\}. \quad (3)$$

In addition, we define  $C_{0, H}(\bar{\phi}) \equiv C_{0, H}(\theta_0, \bar{\phi})$ .<sup>5</sup>

**Assumption (A3).** Let  $P_0$  denote the true probability distribution. (a)  $\frac{1}{\sqrt{T}} \frac{\partial Q_T}{\partial \gamma(c)}(\underline{Y}_T, \gamma_0(c); c)$  is asymptotically normally distributed with zero mean and a finite p.d. asymptotic covariance matrix:

$$I_0(c) = P_0 \lim_{T \rightarrow \infty} \text{Var} \left( T^{-1/2} \frac{\partial Q_T}{\partial \gamma(c)}(\underline{Y}_T, \gamma_0(c); c) \right) \quad (4)$$

(b) There is a  $|c| \times |c|$  finite matrix  $J_0(c)$  such that:

$$J_0(c) = P_0 \lim_{T \rightarrow \infty} T^{-1} \frac{\partial^2 Q_T}{\partial \gamma(c) \partial \gamma'(c)}(\underline{Y}_T, \gamma_0(c); c). \quad (5)$$

---

<sup>5</sup>This assumption ensures that the true parameter value  $\theta_0$  is the unique parameter value at which there are most valid IRFs. For just-identified cases, there may be other parameter values at which the restriction holds, but none of other parameter values give as many valid IRFs as the true parameter value does.

**Assumption (A4).**  $g(\theta, \alpha, \phi)$  is continuously differentiable in  $(\theta, \alpha)$  and  $\frac{\partial g(\theta_0, \bar{\alpha}(c), \bar{\phi}; c)}{\partial [\theta', \alpha']'}$  has full column rank  $(p_\theta + p_\alpha)$ .

Finally, let:

$$\begin{aligned}\Phi_0(c) &\equiv J_0^{-1}(c)I_0(c)J_0^{-1}(c) \\ \Omega(c) &\equiv \Phi_0^{-1}(c) \\ W(\bar{\phi}; c) &\equiv \left[ \frac{\partial g'(\theta_0, \bar{\alpha}(c), \bar{\phi}; c)}{\partial [\theta', \alpha']'} \Phi_0^{-1}(c) \frac{\partial g(\theta_0, \bar{\alpha}(c), \bar{\phi}; c)}{\partial [\theta', \alpha']'} \right]^{-1},\end{aligned}\quad (6)$$

and let  $W^{(1,1)}(\bar{\phi}; c)$  denote the  $(p_\theta \times p_\theta)$  upper-left diagonal sub-matrix of the  $(p_\theta + p_\alpha) \times (p_\theta + p_\alpha)$  matrix  $W(\bar{\phi}; c)$ . Using results from Dridi et al. (2007), Lemma 9 in the Appendix shows that the estimates of  $\theta$  and  $\alpha$  are asymptotically normal, centered around their true and pseudo-true parameter values, respectively, with joint asymptotic covariance matrix  $W(\bar{\phi}; c)$ .

## 2.1 The Valid IRF Selection Criterion

Our first goal is to identify the largest subset of IRFs that guarantee consistent estimation of the parameters of interest  $\theta$ , that is the set of valid IRFs. We define the Valid Impulse Responses Selection Criterion by

$$VIRSC_T(\bar{\phi}; c) = \widehat{\Gamma}_T(\bar{\phi}; c) - h(|c|)\kappa_T, \quad (7)$$

where

$$\begin{aligned}\widehat{\Gamma}_T(\bar{\phi}; c) &= T[\widehat{\gamma}_T(c) - g(\widehat{\theta}_T(\bar{\phi}; c), \widehat{\alpha}_T(\bar{\phi}; c), \bar{\phi}; c)]' \widehat{\Omega}_T(c) \\ &\quad \times [\widehat{\gamma}_T(c) - g(\widehat{\theta}_T(\bar{\phi}; c), \widehat{\alpha}_T(\bar{\phi}; c), \bar{\phi}; c)],\end{aligned}$$

$\widehat{\Omega}_T(c)$  is a consistent estimator of  $\Omega(c)$ ;  $h(|c|)\kappa_T$  is a deterministic penalty that is an increasing function of the number of impulse responses. For example, the SIC-type penalty term imposes  $h(|c|) = |c|$  and  $\kappa_T = \ln(T)$  and it is acceptable for the VIRSC under the restrictions of Assumption B3 listed and discussed below.

We select impulse response functions by minimizing the criterion (7):

$$\widehat{c}_{VIRSC,T} = \arg \min_{c \in C_H} VIRSC_T(\bar{\phi}; c). \quad (8)$$

In this section, our main result shows that  $\widehat{c}_{VIRSC,T}$  converges in probability to the (unique) selection vector,  $c_0$ , that chooses only valid restrictions. This selection vector provides consistent

estimates of the DSGE model parameters (which satisfy almost sure convergence in Section 5). We define the following sets of selection vectors for valid restrictions:

$$C_{\max,H}(\bar{\phi}) = \{c \in C_{0,H}(\theta, \bar{\phi}) : |c| \geq |c_{0,H}| \text{ for all } c_{0,H} \in C_{0,H}(\theta_0, \bar{\phi})\}, \quad (9)$$

$C_{0,H}(\theta, \bar{\phi})$  was defined in (3) and it is the set of selection vectors in which the remaining restrictions are valid.  $C_{\max,H}(\bar{\phi})$  is the set of selection vectors in  $C_{0,H}(\theta, \bar{\phi})$  that maximizes the number of selected vectors in  $C_{0,H}(\theta, \bar{\phi})$ . Also, in addition to Assumptions (A0)–(A4), we consider:

**Assumption (B1).**  $C_{\max,H}(\bar{\phi}) = \{c_0\}$ .

**Assumption (B2).**  $\hat{\Omega}_T(c) \xrightarrow{P} \Omega(c)$ , where  $\Omega(c)$  is positive definite.

**Assumption (B3).**  $h(\cdot)$  is strictly increasing and  $\kappa_T \rightarrow \infty$  as  $T \rightarrow \infty$  with  $\kappa_T = o(T)$ .

Note that Assumption B1 ensures that the set of valid IRFs is uniquely identified.<sup>6</sup> Assumption B2 ensures that there are consistent estimators for the asymptotic variances that enter our formulae. Such consistent estimators of  $\Phi_0(c)$  can be found in Lütkepohl (1990) and Lütkepohl and Poskitt (1991). Assumption B3 imposes appropriate assumptions on the penalty terms that guarantee the validity of the proposed information criterion. Also, note that the SIC-type penalty term ( $h(|c|) = |c|$  and  $\kappa_T = \ln(T)$ ) and the Hannan-Quinn-type penalty term ( $h(|c|) = |c|$  and  $\kappa_T = \ln[\ln(T)]$ ) satisfy Assumption B3, but the AIC-type penalty term, for which  $h(|c|) = |c|$  and  $\kappa_T = 2$ , does not.

**Theorem 1 (Valid IRF Selection Criterion (VAR case))** *Suppose that Assumptions A0, A1, A3–A4 hold for  $c \in C_H$ , that Assumption A2 holds for  $c \in C_{0,H}(\theta, \bar{\phi})$  and that Assumptions B1–B3 hold. Let  $\hat{c}_{VIRSC,T}$  be defined in (8). Then  $\hat{c}_{VIRSC,T} \xrightarrow{P} c_0$ .*

## 2.2 The Relevant IRF Selection Criterion

Our second goal is to identify the fewest number of IRFs that guarantee that, asymptotically, the covariance matrix of the IRFME is as small as possible.<sup>7</sup> We define the Relevant Impulse Responses

<sup>6</sup>Suppose that there are two elements,  $c_1$  and  $c_2$ , and that they are distinct. Then define  $c_3$  as the maximum of  $c_1$  and  $c_2$ . Then  $c_3$  includes more IRFs than  $c_1$  and  $c_2$  and  $c_3$  consists of valid IRFs, a contradiction. So it has to be unique.

<sup>7</sup>For two covariance matrices,  $A$  and  $B$ , we say that  $A$  is smaller than  $B$  if  $B - A$  is positive semi-definite.

Selection Criterion by

$$RIRSC_T(\bar{\phi}; c) = \ln(|\hat{W}_T^{(1,1)}(\bar{\phi}; c)|) + k(|c|)m_T, \quad (10)$$

where  $\hat{W}_T^{(1,1)}(\bar{\phi}; c)$  is a consistent estimator of  $W^{(1,1)}(\bar{\phi}; c)$ , and  $k(|c|)m_T$  is a deterministic penalty that is an increasing function of the number of impulse responses. For example, the SIC-based criterion selects  $k(|c|) = |c|$  and  $m_T = \frac{\ln(\sqrt{T})}{\sqrt{T}}$  (see Assumption C3 and the remarks thereafter for more discussion on the choice of these parameters).

We select impulse response functions by minimizing the criterion (10):

$$\hat{c}_{RIRSC,T} = \arg \min_{c \in C_{0,H}(\bar{\phi})} RIRSC_T(\bar{\phi}; c). \quad (11)$$

where  $C_{0,H}(\bar{\phi})$  is the set of selection vectors in which the restrictions are valid, defined below (3).

In this section, the main result shows that  $\hat{c}_{RIRSC,T}$  converges in probability to the (unique) selection vector  $c_r$ . The vector  $c_r$  chooses, among the valid IRFs, those: (i) with the smallest asymptotic variance and (ii) if a *relevant* IRF is dropped, the asymptotic variance is larger.<sup>8</sup> It is useful to define selection vectors that pick IRFs yielding efficient estimators:

$$\begin{aligned} C_{\mathcal{E},H}(\bar{\phi}) &= \{c \in C_{0,H}(\bar{\phi}) : W^{(1,1)}(\bar{\phi}; c) = W^{(1,1)}(\bar{\phi}; c_0)\}, \\ C_{\min,H}(\bar{\phi}) &= \{c \in C_{\mathcal{E},H}(\bar{\phi}) : |c| \leq |c_{\mathcal{E},H}| \text{ for all } c_{\mathcal{E},H} \in C_{\mathcal{E},H}(\bar{\phi})\}. \end{aligned}$$

$C_{\mathcal{E},H}(\bar{\phi})$  is the set of selection vectors in which selected restrictions yield efficient estimators.  $C_{\min,H}(\bar{\phi})$  is the set of selection vectors that pulls out non-redundant IRFs from the valid IRFs. We sequence the selection vectors first to find an element in  $C_{\mathcal{E},H}(\bar{\phi})$  and second to acquire an element in  $C_{\min,H}(\bar{\phi})$ . In addition to Assumptions (A0)–(A4), we consider:

**Assumption (C1).**  $C_{\min,H}(\bar{\phi}) = \{c_r\}$ .

**Assumption (C2).**

$$\hat{W}_T^{(1,1)}(\bar{\phi}; c) = W^{(1,1)}(\bar{\phi}; c) + O_p(T^{-1/2})$$

for  $c \in C_{I,H}(\bar{\phi})$ , where

$$C_{I,H}(\bar{\phi}) \equiv \{c \in C_{0,H}(\bar{\phi}) : \text{Assumptions } A(0) - A(4) \text{ hold}\},$$

---

<sup>8</sup>The reason why we require (ii) is that even if one adds redundant IRFs to the relevant IRFs one still obtains the same smallest asymptotic variance.

and

$$|\hat{W}_T(\bar{\phi}; c)| \xrightarrow{p} \infty.$$

for  $c \in C_{0,H}(\bar{\phi}) \cap C_{I,H}^c(\bar{\phi})$ .

**Assumption (C3).**  $k(\cdot)$  is strictly increasing and  $m_T$  satisfies  $m_T \rightarrow 0$  and  $T^{1/2}m_T \rightarrow \infty$  as  $T \rightarrow \infty$ .

*Remarks.*  $C_{I,H}(\bar{\phi})$  is the set of selection vectors that maximizes the number of selected valid responses. Assumption C1 ensures that the set of relevant IRFs is uniquely identified. Although the validity of the assumption depends on the model and has to be checked on a case by case basis, we can show that Assumption C1 is satisfied in the simplified practical method discussed in Section 2.3 below as well as in Example 1 and in the AR(1) case discussed respectively in Sections 3 and 7.1 below. Even when Assumption C1 is not satisfied, we expect our criterion will select elements in the set  $C_{\min,H}(\bar{\phi})$ , and therefore the estimator remains consistent.<sup>9</sup> Note that we allow for some selection vectors for which the parameters are not identified provided Assumption C1 holds. Assumption C2 ensures that there are consistent estimators for the asymptotic variances that enter our formulas. Assumption C3 imposes appropriate assumptions on the penalty terms that guarantee the validity of the proposed information criterion. Furthermore, note that the SIC-type penalty term,  $k(|c|) = |c|$  and  $m_T = \frac{\ln(\sqrt{T})}{\sqrt{T}}$  satisfies Assumption C3 whereas the AIC-type penalty term, for which  $k(|c|) = |c|$  and  $m_T = \frac{2}{\sqrt{T}}$  does not.

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

**Theorem 2 (Relevant IRFs Selection Criterion (VAR case))** *Suppose that Assumptions (A0), (A1)–(A4) and (C1)–(C3) hold. Let  $\hat{c}_{RIRSC,T} = \arg \min_{c \in C_{0,H}(\bar{\phi})} RIRSC_T(\bar{\phi}; c)$ . Then  $\hat{c}_{RIRSC,T} \xrightarrow{p} c_r$ .*

### 2.3 Suggestions for Practical Implementation

Theorems 1 and 2 describe the asymptotic behavior of the two criteria, the VIRSC and the RIRSC, that consider all possible combinations of IRFs. However, implementing these criteria in practice

---

<sup>9</sup>Suppose  $C_{\min,H}(\bar{\phi}) = \{c_{1r}, c_{2r}\}$ . We expect that  $\hat{c} \in \{c_{1r}, c_{2r}\}$  with probability approaching one. Then  $\hat{\theta}_T = 1_{\pi,r} \hat{\theta}_T(\bar{\phi}; c_{1r}) + (1 - 1_{\pi,r}) \hat{\theta}_T(\bar{\phi}; c_{2r}) + o_p(1)$ , where  $1_{\pi,r}$  equals 1 if  $c_{1r}$  is selected and equals zero if  $c_{2r}$  is selected. Since  $\hat{\theta}_T(\bar{\phi}; c_{1r})$  and  $\hat{\theta}_T(\bar{\phi}; c_{2r})$  converge in probability to  $\theta_0$ , then  $\hat{\theta}_T \xrightarrow{p} \theta_0$ .

might become very computationally intensive, given the large number of variables included in typical VARs and the fact that  $H$  can be large (for example, CEE chose  $H = 12$ ). Applied researchers, however, often impose an ad hoc maximum lag length to all the IRFs a DSGE model is asked to match.

For practical reasons, we suggest the following. First, select the valid IRFs among the  $(n_Y \times n_Y)$  set of all IRFs given a pre-specified choice for  $H$ . Second, among the valid IRFs, select the relevant IRFs across horizons  $h$ , for  $h \in \{1, 2, \dots, H\}$ . Our VIRSC and RIRSC can be easily tailored to this special case.

This ad hoc approach selects the valid IRFs using a set of selection vectors that consider all IRFs up to the maximum horizon  $H$ . Let

$$\widehat{c}_{VIRSC,T} = \arg \min_{c \in \overline{C}_H} VIRSC_T(\overline{\phi}; c). \quad (12)$$

where

$$\overline{C}_H = \{c \in C_H : c = 1_{H \times 1} \otimes d \text{ for some } d = [d_1, d_2, \dots, d_{n_Y^2}]', d_j \in \{0, 1\}, j = 1, 2, \dots, n_Y^2\}.$$

Define  $\overline{C}_{0,H}(\overline{\phi})$  by  $C_{0,H}(\overline{\phi})$  as in (3), with  $C_H$  replaced by  $\overline{C}_H$ . Also define  $\overline{C}_{\max,H}(\overline{\phi})$  by  $C_{\max,H}(\overline{\phi})$  as in (9) with  $C_{0,H}(\overline{\phi})$  replaced by  $\overline{C}_{0,H}(\overline{\phi})$ . We have:

**Corollary 3 (Simplified VIRSC Criterion (VAR case))** *Let the structural model have a VAR representation (1), and the estimator of the parameters be defined as (2), where  $c$  is chosen by (12). Suppose that Assumptions A1-A4, B1-B3 hold with  $\overline{C}_H$ ,  $\overline{C}_{\min,H}$  and  $\overline{c}_0$  replacing  $C_H$ ,  $C_{\min,H}$  and  $c_0$ , respectively. Then  $\widehat{c}_{VIRSC,T} \xrightarrow{p} \overline{c}_0$ .*

Relevant IRFs are selected among the set of possible IRFs  $\gamma_1, \gamma_2, \dots, \gamma_H$  that satisfy the validity criterion. Let  $c_Y$  denote the  $n_Y^2 \times 1$  vector that selects the valid IRFs for  $h = H$ , that is the last  $n_Y^2 \times 1$  component of  $\widehat{c}_{VIRSC,T}$ . Let  $c_{Y^h} = \underbrace{[c'_Y, c'_Y, \dots, c'_Y]}'_{1 \times n_Y^2 \cdot h} 0_{1 \times n_Y^2 \cdot (H-h)}$  and

$$\overline{\overline{C}}_H = \left\{ c_{Y^h} = [1_{1 \times n_Y^2 \cdot h} \ 0_{1 \times n_Y^2 \cdot (H-h)}]', \text{ for } h = 1, 2, \dots, H \right\}.$$

Define  $\overline{\overline{C}}_{\mathcal{E},H}(\overline{\phi})$ ,  $\overline{\overline{C}}_{\min,H}(\overline{\phi})$  and  $\overline{\overline{c}}_r$  by  $C_{\mathcal{E},H}(\overline{\phi})$ ,  $C_{\min,H}(\overline{\phi})$  and  $c_r$ , respectively, with  $C_H$  replaced by  $\overline{\overline{C}}_H$ . Note that, in this case,  $\overline{\overline{C}}_{\min,H}(\overline{\phi})$  directly satisfies Assumption C1.<sup>10</sup>

<sup>10</sup>In fact, if the order of the VAR is finite, then elements of  $\overline{\overline{C}}_{\min,H}$  are all finite. Suppose that there are two

Using these definitions, we implement the RIRSC by selecting the maximum horizon of IRFs that minimizes the RIRSC across horizons for the given choice of IRF shocks and variables obtained by the VIRSC:

$$\hat{h}_T = \arg \min_{h \in \{1, 2, \dots, H\}} RIRSC_T(\bar{\phi}; c_{Y^h}). \quad (13)$$

Let  $h_r$  denote the corresponding IRF lag length implied by  $\bar{c}_r : \bar{c}_r = [1_{1 \times n_Y^2 h_r} \quad 0_{1 \times n_Y^2 (H-h_r)}]'$ . It follows immediately from Theorems 1 and 2 that  $\hat{h}_T$  is a consistent estimate of  $h_r$ :

**Corollary 4 (Simplified RIRSC Criterion (VAR case))** *Let the structural model have a VAR representation (1), and the estimator of the parameters be defined as (2), where  $c$  is chosen by (12). Suppose that Assumptions A1-A4, C1-C3 hold with  $C_H(\bar{\phi})$ ,  $C_{\min, H}(\bar{\phi})$  and  $c_r$  replaced by  $\bar{C}_H(\bar{\phi})$ ,  $\bar{C}_{\min, H}(\bar{\phi})$  and  $\bar{c}_r$ , respectively. Then  $\hat{h}_T \xrightarrow{p} h_r$ .*

### 3 Interpretation of the Criteria

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, as well as the usefulness of the VIRSC in the presence of model mis-specification.

#### 3.1 Interpretation of the VIRSC

**Example 1 (A Simple New Keynesian Model)** *Consider the following simplified New Keynesian model (cfr. Canova and Sala, 2009):*

$$y_t = k_1 + a_1 E_t y_{t+1} + a_2 (i_t - E_t \pi_{t+1}) + e_{1t} \quad (14)$$

$$\pi_t = k_2 + a_3 E_t \pi_{t+1} + a_4 y_t + e_{2t} \quad (15)$$

$$i_t = k_3 + a_5 E_t \pi_{t+1} + e_{3t} \quad (16)$$

where  $y_t$  is the output gap,  $\pi_t$  is the inflation rate,  $i_t$  is the nominal interest rate, and  $e_{1t}$ ,  $e_{2t}$ ,  $e_{3t}$  are *i.i.d.*(0,1) contemporaneously uncorrelated shocks. The first equation is the IS curve, the second

---

elements in the set  $\bar{C}_{\min, H}$ , and denote these elements by  $h_{r,1}$  and  $h_{r,2}$ , and let  $h_{r,1}$  and  $h_{r,2}$  be different, with  $h_{r,1} < h_{r,2}$ . By definition of  $\bar{C}_{\min, H}$ , they achieve the same asymptotic variance but since  $h_{r,1} < h_{r,2}$  then  $h_{r,2}$  cannot be element of  $\bar{C}_{\min, H}$ , thus inducing a contradiction. Therefore, the set  $\bar{C}_{\min, H}$  must be unique.

is a Phillips curve and the third characterizes monetary policy.

To consider the issue of mis-specification, assume that the researcher imposes  $a_4 = \bar{a}_4$  in his/her model, whereas  $a_4$  is different from zero in the true data generating process (15). In addition, suppose the true values of the parameters  $a_1, a_3$  are known and equal to 0.5 and  $k_1 = 0$ . Then the solution of the model is:

$$\begin{pmatrix} i_t \\ y_t \\ \pi_t \end{pmatrix} = \begin{pmatrix} i \\ y \\ \pi \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ a_2 & 1 & 0 \\ a_2 a_4 & a_4 & 1 \end{pmatrix} \begin{pmatrix} e_{1,t} \\ e_{2,t} \\ e_{3,t} \end{pmatrix} \quad (17)$$

$$\text{where } \begin{pmatrix} i \\ y \\ \pi \end{pmatrix} = \mu \begin{pmatrix} -a_5 a_4 & -0.25 a_5 & -0.25 + a_2 a_4 \\ -0.5 & (1 - a_5) a_2 & -1.5 a_2 \\ -a_4 & -0.5 & -a_2 a_4 \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} \text{ and } \mu = \frac{1}{-0.25 + (a_5 - 1) a_2 a_4}.$$

Suppose the researcher is mainly interested in estimating the slope of the IS equation ( $a_2$ ). Using the notation of the previous sections,  $\phi = \{k_1, a_1, a_3, a\}$ ;  $\theta = \{a_2\}$ ; and  $\alpha = \{k_2, k_3, a_5\}$  is the empty set. He has two options to estimate the parameter of interest: MLE or IRFME. Recall that the researcher works with a mis-specified model that assumes that  $a_4 = \bar{a}_4$ , where we let  $\bar{a}_4 = 1$ . MLE will recover consistent estimates of the mean parameters ( $y$ ,  $\pi$  and  $i$ ). Let the researcher recover the parameter of interest from (17) assuming  $a_4 = 1$  using information on the variances and covariances:

$$\begin{pmatrix} i_t \\ y_t \\ \pi_t \end{pmatrix} \sim N \left( \begin{pmatrix} i \\ y \\ \pi \end{pmatrix}, \begin{pmatrix} 1 & a_2 & a_2 \\ a_2 & a_2^2 + 1 & a_2^2 + 1 \\ a_2 & a_2^2 + 1 & a_2^2 + 2 \end{pmatrix} \right) \quad (18)$$

For example, the researcher may estimate  $\hat{a}_2 = \widehat{cov}(\pi_t, i_t)$ , while in reality the distribution implied by (17) is

$$\begin{pmatrix} i_t \\ y_t \\ \pi_t \end{pmatrix} \sim N \left( \begin{pmatrix} i \\ y \\ \pi \end{pmatrix}, \begin{pmatrix} 1 & a_2 & a_2 a_4 \\ a_2 & a_2^2 + 1 & a_4 a_2^2 + a_4 \\ a_2 a_4 & a_4 a_2^2 + a_4 & a_2^2 a_4^2 + a_4^2 + 1 \end{pmatrix} \right), \quad (19)$$

then  $\hat{a}_2 \rightarrow \bar{a}_2 \equiv cov(\pi_t, i_t)$ . Note that  $\bar{a}_2 = a_2 a_4 \neq a_2$ , which in general will not recover the true value of  $a_2$  unless the true parameter value is  $a_4 = 1$ . It is important to note that in this example

the response of  $y_t$  to  $e_{1,t}$  is the valid and relevant response among the three IRFs at impact, whereas the responses of  $\pi_t$  to  $e_{1,t}, e_{2,t}$  are not valid.

Consider now the IRFME based on the response of  $y_t$  to  $e_{1,t}$  from (17). Since that IRF is not affected by assuming  $a_4 = 1$  or not, then the researcher will correctly estimate  $a_2$  whether or not the model is correctly specified or not.

### 3.2 Interpretation of the RIRSC

**Example 2 (Labor Productivity and Hours)** *This example follows Watson (2006). Watson (2006) derives the VAR representation for a simple RBC model outlined in Christiano, Eichenbaum and Vigfusson (2006, Section 2), in which a technology shock is the only disturbance that affects labor productivity in the long run. Watson (2006) shows that the RBC model can be written as a bivariate VAR(3) model in labor productivity ( $y_t/l_t$ ) and employment ( $l_t$ ):*

$$\begin{aligned} \begin{bmatrix} 1 & \alpha_y \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta \ln \left( \frac{y_t}{l_t} \right) \\ \ln(l_t) \end{bmatrix} &= \begin{bmatrix} 0 & \alpha_y \\ \gamma_1 & \rho_l - \alpha_y \gamma_1 \end{bmatrix} \begin{bmatrix} \Delta \ln \left( \frac{y_{t-1}}{l_{t-1}} \right) \\ \ln(l_{t-1}) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -\rho_l \gamma_1 & \alpha_y (1 + \rho_l) \gamma_1 \end{bmatrix} \begin{bmatrix} \Delta \ln \left( \frac{y_{t-2}}{l_{t-2}} \right) \\ \ln(l_{t-2}) \end{bmatrix} \\ &+ \begin{bmatrix} 0 & 0 \\ 0 & -\alpha_y \rho_l \gamma_1 \end{bmatrix} \begin{bmatrix} \Delta \ln \left( \frac{y_{t-3}}{l_{t-3}} \right) \\ \ln(l_{t-3}) \end{bmatrix} + \begin{bmatrix} \eta_t \\ \nu_t \end{bmatrix} \end{aligned} \quad (20)$$

where  $\alpha$  is the capital share in production;  $\rho_l$  is the serial correlation coefficient in the tax rate on labor income process;  $\tilde{a}_z$  and  $a_z$  are, respectively, the parameters associated with the lagged state of technology in the policy rule for labor in the standard and recursive versions of the model;  $\gamma_1 = (\tilde{a}_z - a_z \rho_l)/(1 - \alpha_y)$ ;  $\gamma_2 = -\tilde{a}_z \rho_l/(1 - \alpha_y)$ ;  $\eta_t = (1 - \alpha_y) \sigma_z \varepsilon_t^z$ ,  $\nu_t = a_l \sigma_l \varepsilon_t^l$ ; and  $\varepsilon_t^l$ ,  $\varepsilon_t^z$  are *i.i.d.* zero mean and unit variance shocks (cfr. Watson, 2006, eqs. 3 and 4). The structural parameters of interest in this example are  $\alpha_y$ ,  $\gamma_1$  and  $\rho_l$ . For identification purposes, we impose the short-run restriction that  $a_z = 0$ , which yields the short-run restriction,  $\gamma_1 = \tilde{a}_z/(1 - \alpha_y)$ . Thus  $\gamma_1$  is informative for the structural parameter  $\tilde{a}_z$ , given the estimate of  $\alpha_y$ .

In the notation of the previous sections, the structural parameters of interest are  $\theta = \{\alpha_y, \rho_l\}$ , the estimated nuisance parameter is  $\alpha = \{\gamma_1\}$ , and the calibrated nuisance parameters are  $\phi = \{a_z, a_l, \sigma_l, \sigma_z\}$ . There are two trivial examples of redundant impulse responses. One is the response of  $\Delta \ln \left( \frac{y_t}{l_t} \right)$  to  $\varepsilon_t^z$ , which is always zero – see (20). Another is restrictions on the impulse responses  $\gamma_j$  for  $j > 3$ : since the model is a VAR(3) model, restrictions for  $j > 3$  are nonlinear transformation of

the first three impulse responses, 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.

## 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

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

for  $\tau = 1, \dots, H$ , and  $D$  is a matrix derived from the identification procedure.

Let the vector of structural impulse responses estimated by local projections be denoted by  $\hat{\gamma}_{J,T}(c)$ . Jordà's local projection impulse response estimator is:

$$\begin{pmatrix} \hat{\theta}_{J,T}(\bar{\phi}; c) \\ \hat{\alpha}_{J,T}(\bar{\phi}; c) \end{pmatrix} = \arg \min_{\theta \in \Theta, \alpha \in A} (\hat{\gamma}_{J,T}(c) - g(\theta, \alpha, \bar{\phi}; c))' \hat{\Omega}_T(c) (\hat{\gamma}_{J,T}(c) - g(\theta, \alpha, \bar{\phi}; c)) \quad (21)$$

where  $g(\theta, \alpha, \bar{\phi}; c)$  is the vector of the model's theoretical impulse responses given structural parameter  $\theta$ , and  $\hat{\Omega}_{J,T}(c)$  is the inverse of a consistent estimator of the asymptotic covariance matrix of  $\hat{\gamma}_{J,T}(c)$ .

Our main result for the local projection estimator is:

**Theorem 5 (Consistent IRF Selection (Local Projections case))** *Suppose that Assumptions A1-A4, B1-B3, and C1-C3 hold with  $\widehat{W}_T(\bar{\phi}; c)$  replaced by  $\widehat{W}_{J,T}(\bar{\phi}; c)$  which is a consistent estimator of (6) constructed using  $\widehat{\gamma}_{J,T}(c)$  and  $\widehat{\Omega}_{J,T}(c)$ , and  $\widehat{W}_{J,T}^{(1,1)}(\bar{\phi}; c)$  the  $(p_\theta \times p_\theta)$  upper-left diagonal sub-matrix of  $\widehat{W}_{J,T}(\bar{\phi}; c)$ . Let the estimator of  $\theta$  be (21), where  $c$  is chosen such that:*

$$\hat{c}_{VIRSC,T} = \arg \min_{c \in C_H} VIRSC_{J,T}(\bar{\phi}; c), \text{ where}$$

$$VIRSC_{J,T}(\bar{\phi}; c) = \widehat{\Gamma}_{J,T}(\bar{\phi}; c) - h(|c|)\kappa_T,$$

$$\widehat{\Gamma}_{J,T}(\bar{\phi}; c) = T(\widehat{\gamma}_{J,T}(c) - g(\widehat{\theta}_{J,T}(\bar{\phi}; c), \widehat{\alpha}_{J,T}(\bar{\phi}; c), \bar{\phi}; c))' \widehat{\Omega}_{J,T}(c) (\widehat{\gamma}_{J,T}(c) - g(\widehat{\theta}_{J,T}(\bar{\phi}; c), \widehat{\alpha}_{J,T}(\bar{\phi}; c), \bar{\phi}; c))$$

and:

$$\hat{c}_{RIRSC,T} = \arg \min_{c \in C_{0,H}(\bar{\phi})} RIRSC_{J,T}(\bar{\phi}; c), \text{ and}$$

$$RIRSC_{J,T}(\bar{\phi}; c) = \log(|\widehat{W}_{J,T}^{(1,1)}(\bar{\phi}; c)|) + k(|c|)m_T.$$

Then  $\hat{c}_{VIRSC,T} \xrightarrow{P} c_0$  and  $\hat{c}_{RIRSC,T} \xrightarrow{P} c_r$ .

## 4.2 Indirect Inference Estimators

The third estimator that we consider is the simulation-based estimator. The simulation-based estimator is an indirect-inference (II) estimator with a sequence of finite-order VAR models used as an auxiliary model (see Smith (1993) and Dridi, Guay and Renault (2007), and Gouriou, Monfort and Renault (1993) for examples of indirect inference applied to DSGE models and financial models, respectively). In the macroeconomics literature, the application of simulation-based estimators to IRFMEs is referred to as the Sims-Cogley-Nason estimator.<sup>11</sup>

The II estimator is implemented as follows. First, fit a VAR( $p$ ) to the actual data to obtain sample impulse responses  $\widehat{\gamma}_T(c)$ .<sup>12</sup> Note that all the results in this section still hold if the VAR is of infinite order provided that  $H$  is finite. Next, simulate synthetic data of length  $T$  from the DSGE model with parameter vector  $(\theta, \alpha, \bar{\phi})$ . Let the  $s$ -th simulated synthetic data obtained using initial condition  $Y_0^s$  be denoted by  $\underline{Y}_T^s(\theta, \alpha, \bar{\phi}, Y_0^s)$ , and repeat this process for  $s = 1, \dots, S$ , where  $S$  is the total number of simulation replications. Estimate the VAR( $p$ ) on the synthetic

<sup>11</sup>The Sims-Cogley-Nason estimator was popularized by Kehoe (2006).

<sup>12</sup>All the subsequent estimated parameters should also be function of  $p$ , the estimated VAR lag length. However, in order to simplify notation, we suppress this dependence in the notation.

samples  $s = 1, \dots, S$  to obtain a matrix of simulated theoretical IRFs. Let  $\tilde{g}_T^{(s)}(\theta, \alpha, \bar{\phi})$  denote the vector of simulated impulse responses from the  $s$ -th synthetic sample. Finally, define  $\tilde{g}_T^S(\theta, \alpha, \bar{\phi})$  to be the average across the ensemble of simulated IRFs, which we refer to as the (approximate) theoretical impulse responses,  $\tilde{g}_T^S(\theta, \alpha, \bar{\phi}) = (1/S) \sum_{s=1}^S \tilde{g}_T^{(s)}(\theta, \alpha, \bar{\phi})$ . As the paper did earlier,  $c$  is used to index subsets of IRFs,  $\tilde{g}_T^S(\theta, \alpha, \bar{\phi}; c)$ . Note that the matrix of shock innovations is drawn only once and held fixed as  $\theta$  is adjusted to move  $\tilde{g}_T^S(\theta, \alpha, \bar{\phi}; c)$  closer to  $\hat{\gamma}_T(c)$ .

The II estimator of  $\theta$  minimizes the distance between the average simulated theoretical impulse responses and the sample impulse responses:

$$\begin{pmatrix} \hat{\theta}_{II,T}(S, \bar{\phi}; c) \\ \hat{\alpha}_{II,T}(S, \bar{\phi}; c) \end{pmatrix} = \arg \min_{\theta \in \Theta} (\hat{\gamma}_T(c) - \tilde{g}_T^S(\theta, \alpha, \bar{\phi}; c))' \hat{\Omega}_T(c) (\hat{\gamma}_T(c) - \tilde{g}_T^S(\theta, \alpha, \bar{\phi}; c)), \quad (22)$$

where  $\hat{\Omega}_T(c)$  is a weighting matrix.<sup>13</sup>

Next, consider the problem of selecting the impulse responses for the IRFME. We impose additional assumptions:

**Assumption (A1').** Let  $P_*$  denote the probability measure of the simulated data. For the simulated theoretical impulse responses  $\tilde{g}_T^S(\theta, \alpha, \bar{\phi}; c)$ ,  $Q_T(\underline{Y}_T, \gamma(c); c)$  satisfies

$$\begin{aligned} \tilde{g}_T^S(\theta, \alpha, \bar{\phi}; c) &= \frac{1}{S} \sum_{s=1}^S \tilde{g}_T^{(s)}(\theta, \alpha, \bar{\phi}; c), \\ \tilde{g}_T^{(s)}(\theta, \alpha, \bar{\phi}; c) &= \gamma_0(c) - \left[ \frac{\partial^2 Q_T(\underline{Y}_T^s(\theta, \alpha, \bar{\phi}, Y_0^s), \gamma_0(c); c)}{\partial \gamma(c) \partial \gamma'(c)} \right]^{-1} \\ &\quad \times \frac{\partial Q_T(\underline{Y}_T^s(\theta, \alpha, \bar{\phi}, Y_0^s), \gamma_0(c); c)}{\partial \gamma(c)} + o_p(1). \end{aligned}$$

**Assumption (A3').** (a)  $\frac{1}{\sqrt{T}} \frac{\partial Q_T}{\partial \gamma(c)}(\underline{Y}_T^s(\theta_0, \bar{\alpha}(c), \bar{\phi}, Y_0^s), \gamma_0(c); c)$  is asymptotically normally distributed with zero mean and asymptotic covariance matrix

$$I_0^*(\bar{\phi}; c) = P_* \lim_{T \rightarrow \infty} \text{Var} \left( T^{-1/2} \frac{\partial Q_T}{\partial \gamma(c)}(\underline{Y}_T^s(\theta_0, \bar{\alpha}(c), \bar{\phi}, Y_0^s), \gamma_0(c); c) \right)$$

and independent of the initial values  $Y_0^s$ ,  $s = 1, 2, \dots, S$ . (b) There is a  $|c| \times |c|$  matrix  $J_0^*(\bar{\phi}; c)$  such that

$$J_0^*(\bar{\phi}; c) = P_* \lim_{T \rightarrow \infty} T^{-1} \frac{\partial^2 Q_T}{\partial \gamma(c) \partial \gamma'(c)}(\underline{Y}_T^s(\theta_0, \bar{\alpha}(c), \bar{\phi}, Y_0^s), \gamma_0(c); c).$$

<sup>13</sup>The Appendix shows that, under quite mild conditions,  $\hat{\theta}_{II,T}(S, \bar{\phi}; c)$  is consistent and asymptotically normal.

**Assumption (A4').**  $\tilde{g}_T^{(s)}(\theta, \alpha, \phi; c)$  is continuously differentiable in  $(\theta, \alpha, \phi)$  and  $P_* \lim_{T \rightarrow \infty} \frac{\partial \tilde{g}_T^{(s)}(\theta_0, \bar{\alpha}(c), \bar{\phi}; c)}{\partial [\theta', \alpha']'} = \frac{\partial g(\theta_0, \bar{\alpha}(c), \bar{\phi}; c)}{\partial [\theta', \alpha']'}$ .

**Assumption (A5).** Let  $Cov_*$  denote covariance under  $P_*$ . There are  $|c| \times |c|$  matrices  $K_0(\bar{\phi}; c)$  and  $K_0^*(\bar{\phi}; c)$  such that

$$\lim_{T \rightarrow \infty} Cov_* \left\{ \frac{1}{\sqrt{T}} \frac{\partial Q_T}{\partial \gamma(c)}(\underline{Y}_T, \gamma_0(c); c), \frac{1}{\sqrt{T}} \frac{\partial Q_T}{\partial \gamma(c)}(\underline{Y}_T^s(\theta_0, \bar{\alpha}(c), \bar{\phi}, Y_0^s), \tilde{\gamma}^0(\theta_0, \bar{\alpha}(c), \bar{\phi}; c); c) \right\} = K_0(\bar{\phi}; c)$$

$$\lim_{T \rightarrow \infty} Cov_* \left\{ \frac{1}{\sqrt{T}} \frac{\partial Q_T}{\partial \gamma}(\underline{Y}_T^s(\theta_0, \bar{\alpha}(c), \bar{\phi}, Y_0^s), \gamma_0(c); c), \frac{1}{\sqrt{T}} \frac{\partial Q_T}{\partial \gamma}(\underline{Y}_T^l(\theta_0, \bar{\alpha}(c), \bar{\phi}, Y_0^l), \tilde{\gamma}^0(\theta_0, \bar{\alpha}(c), \bar{\phi}; c); c) \right\} = K_0^*(\bar{\phi}; c)$$

independent of the initial values  $Y_0^s$  and  $Y_0^l$ ,  $s \neq l$ ,  $s, l = 1, 2, \dots, S$ .

Let the following definitions hold:

$$\begin{aligned} \Phi_0(S, \bar{\phi}; c) &= J_0^{-1}(c)I_0(c)J_0^{-1}(c) + \frac{1}{S}J_0^{*-1}(\bar{\phi}; c)I_0^*(\bar{\phi}; c)J_0^*(\bar{\phi}; c)^{-1} \\ &\quad + \left(1 - \frac{1}{S}\right)J_0^{*-1}(\bar{\phi}; c)K_0^*(\bar{\phi}; c)J_0^{*-1}(\bar{\phi}; c) \\ &\quad - J_0^{-1}(c)K_0(\bar{\phi}; c)J_0^{*-1}(\bar{\phi}; c) - J_0^{*-1}(\bar{\phi}; c)K_0'(\bar{\phi}; c)J_0^{-1}(c) \\ \Omega(S, \bar{\phi}; c) &= \Phi_0^{-1}(S, \bar{\phi}; c) \\ W(S, \bar{\phi}; c) &= \left[ \frac{\partial g'(\theta_0, \bar{\alpha}(c); c)}{\partial [\theta', \alpha']'}(\Phi_0(S, \bar{\phi}; c))^{-1} \frac{\partial g(\theta_0, \bar{\alpha}(c); c)}{\partial [\theta', \alpha']'} \right]^{-1} \end{aligned} \quad (23)$$

and  $W^{(1,1)}(S, \bar{\phi}; c)$  denotes the  $(p_\theta \times p_\theta)$  upper-left diagonal sub-matrix of the  $(p_\theta + p_\alpha) \times (p_\theta + p_\alpha)$  matrices  $W(S, \bar{\phi}; c)$ . Lemma 10 in the Appendix shows that the simulation-based estimators  $\hat{\theta}_{II,T}(\bar{\phi}; c), \hat{\alpha}_{II,T}(\bar{\phi}; c)$  are asymptotically normal, centered around their true and pseudo-true parameter values, respectively, with an asymptotic covariance matrix equal to  $W(S, \bar{\phi}; c)$ . Finally, let  $\widehat{W}_{II,T}(S, \bar{\phi}; c)$  be a consistent estimator of  $W(S, \bar{\phi}; c)$  and  $\widehat{\Gamma}_{II,T}(S, \bar{\phi}; c)$  be defined as

$$\begin{aligned} \widehat{\Gamma}_{II,T}(S, \bar{\phi}; c) &= T(\hat{\gamma}_T(c) - \tilde{g}_T^S(\hat{\theta}_{II,T}(S, \bar{\phi}; c), \hat{\alpha}_{II,T}(S, \bar{\phi}; c), \bar{\phi}; c))' \\ &\quad \times \widehat{\Omega}_T(c)(\hat{\gamma}_T(c) - \tilde{g}_T^S(\hat{\theta}_{II,T}(S, \bar{\phi}; c), \hat{\alpha}_{II,T}(S, \bar{\phi}; c), \bar{\phi}; c)). \end{aligned}$$

Finally, let  $\widehat{W}_{II,T}^{(1,1)}(S, \bar{\phi}; c)$  be the  $(p_\theta \times p_\theta)$  upper-left diagonal sub-matrix of the  $(n_Y^2 H \times n_Y^2 H)$  matrix  $\widehat{W}_{II,T}(S, \bar{\phi}; c)$ . Note that Gouriéroux, Monfort and Renault (1993, S112–S113) propose consistent estimators of  $\Phi_0(S, \bar{\phi}; c)$  for some special cases which can be used to construct  $\widehat{W}_{II,T}(S, \bar{\phi}; c)$  and  $\widehat{W}_{II,T}^{(1,1)}(S, \bar{\phi}; c)$  can be computed using (71) in the Appendix.

Theorem 6 describes the IRF selection criteria we propose for the II estimator:

**Theorem 6 (Consistent IRF Selection (Simulation-based case))** *Let Assumptions A1, A1', A2, A3', A4', A5 hold. Let  $c$  be chosen such that:*

$$\begin{aligned}\hat{c}_{VIRSC,T} &= \arg \min_{c \in C_H} VIRSC_{II,T}(\bar{\phi}; c), \text{ where} \\ VIRSC_{II,T}(\bar{\phi}; c) &= \hat{\Gamma}_{II,T}(S, \bar{\phi}; c) - h(|c|)\kappa_T,\end{aligned}$$

and

$$\begin{aligned}\hat{c}_{RIRSC,T} &= \arg \min_{c \in C_{0,H}(\bar{\phi})} RIRSC_{II,T}(\bar{\phi}; c), \text{ and} \\ RIRSC_{II,T}(\bar{\phi}; c) &= \log(|\widehat{W}_{II,T}^{(1,1)}(S, \bar{\phi}; c)|) + k(|c|)m_T.\end{aligned}$$

Then, under Assumptions B1-B3,  $\hat{c}_{VIRSC,T} \xrightarrow{P} c_0$  and, under the additional Assumptions C1-C3,  $\hat{c}_{RIRSC,T} \xrightarrow{P} c_T$ .

## 5 Strongly Consistent IRFs Selection

This section derives additional results that guarantee almost sure convergence of the VIRSC and RIRSC. This analysis will provide more guidance on the choice of the penalty term than the weak consistency results of Sections 2.1 and 2.2 (in fact, the weak consistency results only require Assumption C3 which is valid for many choices for  $\kappa(|c|)$  and  $m_T$ ), although at the cost of imposing additional assumptions on the data. For simplicity, we will do so only in the simulation-based estimator: the results for the VAR-based and the Projection-based estimators follow directly as special cases. We impose additional assumptions:

**Assumption (D).** (a) *There is a unique  $\theta_0(\bar{\phi}; c)$  and  $\alpha_0(\bar{\phi}; c)$  such that*

$$\begin{aligned}[\theta_0(\bar{\phi}; c)', \alpha_0(\bar{\phi}; c)']' &= \arg \min_{\theta \in \Theta, \alpha \in A} [\gamma_0(c) - g(\theta(\bar{\phi}; c), \alpha(\bar{\phi}; c), \bar{\phi}; c)]' \Omega(S, \bar{\phi}; c) \\ &\quad [\gamma_0(c) - g(\theta(\bar{\phi}; c), \alpha(\bar{\phi}; c), \bar{\phi}; c)],\end{aligned}$$

and  $[\hat{\theta}_{II,T}(S, \bar{\phi}; c)', \hat{\alpha}_{II,T}(S, \bar{\phi}; c)']' \xrightarrow{a.s.} [\theta_0(\bar{\phi}; c)', \alpha_0(\bar{\phi}; c)']'$  for all  $c \in C_H$ . (b) *There is a sequence of positive semi-definite matrices  $\{\hat{\Omega}_T(S, \bar{\phi}; c)\}$  such that  $\hat{\Omega}_T(S, \bar{\phi}; c) \xrightarrow{a.s.} \Omega(S, \bar{\phi}; c)$  where  $\Omega(S, \bar{\phi}; c)$  is positive definite for all  $c \in C_{0,H}(\bar{\phi})$ . (c) *The Law of the Iterated Logarithm (LIL) holds:**

$$\limsup_{T \rightarrow \infty} \frac{1}{(2T \ln \ln T)^{1/2}} \left| b' \Omega^{1/2}(S, \bar{\phi}; c) (\hat{\gamma}_T(c) - g(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c); c)) \right| = 1, \text{ a.s.}$$

for all  $b \in \mathfrak{R}^{|c|}$  such that  $|b|=1$  (where  $b$  is a generic vector) and  $c \in C_{0,H}(\bar{\phi})$  where  $\ln \ln T$  denotes  $\ln(\ln(T))$ . (d)  $\sup_{\theta \in \Theta, \alpha \in A} \|D\tilde{g}_T^S(\theta, \alpha, \bar{\phi}; c) - Dg(\theta, \alpha, \bar{\phi}; c)\| = o_{as}(1)$ .

The following results holds:

**Theorem 7 (Strong Consistency of VIRSC )** Suppose that Assumptions A4', B1, B3 and D hold. Then  $\hat{c}_{VIRSC,T} \xrightarrow{a.s.} c_0$ .

Theorem 7 shows that the results of Theorem 1 can be strengthened to almost sure convergence.

We can derive similar results for the RIRSC under the following Assumption. Let  $Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s)$  denote the time  $t$  component of the vector  $\underline{Y}_T^s(\theta, \alpha, \bar{\phi}, Y_0^s)$ .

**Assumption (E).** (a) Let  $G(c) \equiv \partial g(\theta, \alpha, \bar{\phi}; c) / \partial [\theta', \alpha']' |_{\theta=\theta_0, \alpha=\bar{\alpha}(c)}$ ,

$\bar{G}(c) = (\partial / \partial \theta') \text{vec} \{ \partial g(\theta, \alpha, \bar{\phi}; c) / \partial [\theta', \alpha']' \} |_{\theta=\theta_0, \alpha=\bar{\alpha}(c)}$ ,  $\hat{G}(c)$  be a consistent estimator of  $G(c)$ , and let the following approximations hold:

$$T^{1/2}(\hat{\theta}_T(\bar{\phi}; c) - \theta_0(c)) = -\{G(c)' \Omega(S, \bar{\phi}; c) G(c)\}^{-1} G(c)' \Omega(S, \bar{\phi}; c) \times T^{1/2}[\hat{\gamma}_T(c) - \tilde{g}_T^S(\theta_0, \bar{\alpha}, \bar{\phi}; c)] + o(1) \quad a.s. \quad (24)$$

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

Also,  $\omega_v(Y_t, Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s), \theta_0; c)$  and  $\omega_\gamma(Y_t, Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s), \theta_0; c)$  exist such that:

$$T^{1/2} \text{vech}\{\hat{\Phi}_T(S, \bar{\phi}; c) - \Phi_0(S, \bar{\phi}; c)\} = T^{-1/2} \sum_{t=1+m}^T \omega_v(Y_t, Y_t^s(\theta_0, \bar{\alpha}(c), \bar{\phi}, Y_0^s), \theta_0; c) + o(1) \quad a.s. \quad (26)$$

$$T^{1/2}(\hat{\gamma}_T(c) - \tilde{g}_T^S(\theta_0, \bar{\alpha}(c), \bar{\phi}; c)) = T^{-1/2} \sum_{t=1+m}^T \omega_\gamma(Y_t, Y_t^s(\theta_0, \bar{\alpha}(c), \bar{\phi}, Y_0^s), \theta_0; c) + o(1) \quad a.s.$$

for some  $0 < m < \infty$ .

(b) Let  $\omega(Y_t, Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s), \theta_0; c) = [\omega_\gamma(Y_t, Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s), \theta_0; c)' \omega_v(Y_t, Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s), \theta_0; c)]'$ .

Define  $\Omega_\omega(S, \bar{\phi}; c) = \lim_{T \rightarrow \infty} \text{Var}[T^{-1/2} \sum_{t=1}^T \omega(Y_t, Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s), \theta_0; c)]$ .  $\sum_{t=1}^T \omega(Y_t, Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s), \theta_0; c)$  satisfies the LIL in the sense that for all  $b \in \mathfrak{R}^{\dim(\omega)}$  with  $\|b\| = 1$ ,

$$\lim_{T \rightarrow \infty} \sup \left\{ \frac{1}{(2T \ln \ln T)^{1/2}} \left| b' \Omega_\omega(S, \bar{\phi}; c)^{-1/2} \sum_{t=1}^T \omega(Y_t, Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s), \theta_0; c) \right| \right\} = 1, \quad a.s.$$

for all  $c \in C$ .

**Assumption (F).** Let the penalty function be  $k(|c|)m_T$  where  $k(\cdot)$  is strictly increasing and  $m_T = o(1)$  and either: (i)  $\liminf_{T \rightarrow \infty} T^{1/2}m_T/(\ln \ln T)^{1/2} = \mu$  where  $z < \mu < \infty$  and  $z$  is a positive constant that is defined in Appendix A; or (ii)  $\liminf_{T \rightarrow \infty} T^{1/2}m_T/(\ln \ln T)^{1/2} = +\infty$ .

The following results holds:

**Theorem 8 (Strong Consistency of RIRSC )** Suppose that Assumptions  $A4'$ ,  $C1$ ,  $C3$ ,  $E$  and  $F$  hold. Then  $\hat{c}_{RIRSC,T} \xrightarrow{a.s.} c_r$ .

*Remarks.* Theorem 8 establishes conditions under which  $\hat{c}_{RIRSC,T}$  is strongly consistent for  $c_r$ . It can be seen that the conditions on the penalty term are necessarily satisfied if  $\kappa(|c|, p, T) = (|c| - p)\ln[T^{1/2}]/T^{1/2}$ , which is the penalty term associated with the Schwarz information criterion. However, the conditions are not necessarily satisfied if  $\kappa(|c|, p, T) = (|c| - p)\ln[\ln(T^{1/2})]/T^{1/2}$ , 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 A. 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  $\hat{c}_{RIRSC,T} \xrightarrow{a.s.} \bar{c}$  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.

It is interesting to contrast the conditions on the penalty term for the case considered here in which the order of convergence of  $\widehat{W}_T^{(1,1)}(S, \bar{\phi}; c)$  to  $W^{(1,1)}(S, \bar{\phi}; c)$  is  $T^{-1/2}$ . For weak consistency, it is only necessary that  $m_T \rightarrow \infty$  and  $m_T = o(T^{1/2})$ . Given the discussion above, the strong consistency results suggest the use of a penalty term for which  $\liminf_{T \rightarrow \infty} T^{1/2}m_T/(\ln \ln T)^{1/2}$  diverges. Theorem 8 therefore provides more guidance on the choice of penalty term than the corresponding weak consistency result.

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

## 6 Empirical Analysis of Two Representative DSGE Models

This section applies a VAR-based IRFME and the information criteria to the new Keynesian DSGE models of CEE and ACEL. The goal is to assess the impact of the VIRSC and the RIRSC on the estimated parameters of these DSGE models. Since the VIRSC confirms that all IRFs are valid, we focus on the RIRSC. 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.<sup>14</sup>

The CEE and ACEL DSGE models use different schemes to identify IRFs. The CEE 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.

Long-run neutrality restrictions identify the IRFs engaged to estimate the parameters of the ACEL DSGE model. The ACEL DSGE model is constructed such that: (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 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 (10) 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

---

<sup>14</sup>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.

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,  $\rho_{\mu_z}$ , 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),  $\gamma$ , 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  $\gamma$  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 has smaller standard errors.

INSERT TABLES 1 AND 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  $\gamma$ . Since the RIRSC-IRFME estimate of  $\gamma$  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.<sup>15</sup>

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.<sup>16</sup> 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.

## 7 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 ACEL and CEE is computationally infeasible at the moment. Thus, we consider a simple univariate AR(1) process; the structural VAR(3) discussed in example 2; and the simplified New Keynesian model discussed in example 1.

---

<sup>15</sup>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.

<sup>16</sup>In particular, results were robust to adding a Normal(0,  $\sigma$ ) 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).

## 7.1 The AR(1)

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

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

where  $\varepsilon_t$  are random draws from a normal distribution with mean zero and variance one,  $\rho = 0.4$  and  $T = 100$ . We estimate the deep parameter  $\rho$  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  $\widehat{\Omega}_T$  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 (10) is used to select the relevant IRF lag length. In this example, all IRFs are valid; also note that Assumption C1 holds, and the unique relevant IRF is the first.

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  $H$  lags, labeled “IRFME”; and the IRF matching estimator using only the IRFs selected by (10), labeled “IRFME<sub>RIRSC</sub>”. 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

## 7.2 The Structural VAR(3) in Example 2

We consider estimation of  $\alpha$  and  $\gamma_l$  in example 2 by IRFME. We set  $\alpha_y = 0.35$ ,  $\sigma_z = 1$ ,  $\sigma_l = 1$ ,  $\rho_l = 0.95$ ,  $\gamma_l = 1$ ,  $\alpha_l = 0$  (because of the short run identification restriction) and  $\tilde{\alpha}_z = 0.325$ , so

that  $\gamma_1 = 0.5$ .<sup>17</sup> 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 based 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 data generating process, 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 only the IRFs selected by the VIRSC (labeled "Valid IRF Selection Only"), only those selected by the RIRSC (labeled "Relevant IRF Selection Only") or in a sequential procedure where the VIRSC is used first, and then the RIRSC is applied to the valid IRFs only (labeled "Valid and Relevant IRF Selection"). The criteria were implemented using the following choices. For the VIRSC: (AIC)  $h(|c|) = 2|c|$ ,  $\kappa_T = 1$ ; (SIC)  $h(|c|) = |c|$ ,  $\kappa_T = \ln(T)$ ; (HQC)  $h(|c|) = 2|c|$ ,  $\kappa_T = \ln(\ln(T))$ . For the RIRSC: (AIC)  $k(|c|) = 2|c|$ ,  $m_T = 1/\sqrt{T}$ ; (SIC)  $k(|c|) = |c|$ ,  $m_T = \ln(T)/\sqrt{T}$ ; (HQC)  $k(|c|) = 2|c|$ ,  $m_T = \ln(\ln(T))/\sqrt{T}$ . The table shows that the RIRSC significantly improves the coverage probability of the IRFME based on using all IRFs and also on using the valid IRFs. Although the AIC-type penalty term does not satisfy Assumption C3, 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  $h = 1$  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. Overall, the Monte Carlo simulations show that using jointly Valid and Relevant IRF Selection Criteria results in significant improvements in the performance of the estimators.

INSERT TABLES 4, 5 AND 6 HERE

---

<sup>17</sup>We have looked at all the cases in which  $\alpha \in \{0.275, 0.35, 0.425\}$ ,  $\rho_i = \{0.75, 0.85, 0.9, 0.95, 0.975, 1\}$ ,  $\sigma_i = \{0.5, 0.75, 1, 1.25, 1.5\}$ . They are qualitatively similar to the reported results and are available upon request.

### 7.3 The Simple New Keynesian Model in Example 1

In this example, the researcher does not have to choose across horizons but to choose across impulse responses only. Note that a Cholesky decomposition will directly recover the IRFs at horizon zero from (18).

Assume the researcher believes that  $a_4 \equiv \bar{a}_4$ , where for example  $\bar{a}_4 = 0.2$ . Then he will estimate:

$$\begin{pmatrix} i_t \\ y_t \\ \pi \end{pmatrix} = \begin{pmatrix} i \\ y \\ \pi \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ a_2 & 1 & 0 \\ a_2\bar{a}_4 & \bar{a}_4 & 1 \end{pmatrix} \begin{pmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{pmatrix} \quad (27)$$

where it is known that the variance of the shocks is normalized to unity. We thus focus on extracting information from:  $[\frac{\partial y_t}{\partial e_{1,t}}, \frac{\partial \pi_t}{\partial e_{1,t}}, \frac{\partial \pi_t}{\partial e_{2,t}}]'$   $= [a_2, a_2\bar{a}_4, \bar{a}_4]'$ , where the selection vector  $c$  is defined accordingly as  $c = [c_1, c_2, c_3]'$ . We thus have, for each shock, three impulse responses of the three macroeconomic variables to the shock, which could be taken individually, in combinations of two, or all three, for a total of 7 combinations. The data generating process is (17) with  $a_4 = 0.8$  and  $a_2 = 0.5$ , and we let  $T = 100$ . The empirical IRFs are derived from (19) and the theoretical model used by the researcher to recover the responses and the parameters is (27), where  $a_4$  is assumed to be equal to  $\bar{a}_4$ . In this example,  $c_1$  is the valid and relevant impulse response to estimate  $a_2$  when  $\bar{a}_4$  is far from the true value  $a_4$  (as the other IRFs impose the constraint  $a_4 = \bar{a}_4$  which can be mis-specified unless  $\bar{a}_4 = 0.8$ );  $c_2$  is a valid response when  $\bar{a}_4 = a_4 = 0.8$ .

Table 7 reports which IRFs are selected by the proposed criteria. Table 7(a) reports results for the VIRSC and Table 7(b) reports results for the RIRFSC, where the RIRFSC is applied to the IRFs that satisfy the VIRSC. The results are based on 10,000 Monte Carlo replications. The criteria were implemented using SIC-type criteria: for the VIRSC:  $h(|c|) = |c|$ ,  $\kappa_T = \ln(T)$ ; for the RIRSC:  $k(|c|) = |c|$ ,  $m_T = \ln(T)/\sqrt{T}$ . It is clear that the VIRSC tends to select IRFs that include the first one ( $c_1 = 1$ ). However, as  $\bar{a}_4$  gets closer to 0.8, the VIRSC will sometimes select the second and/or the third responses, as the parameter value imposed by the researcher becomes closer and closer to its true value. In no case does the VIRSC discards the first IRF. Table 7(b) shows that the additional redundant responses are easily wiped out by the RIRSC, which almost always selects only the first IRF.

Table 8(a,b,c) show that the median bias, variance and mean coverage probabilities are sig-

nificantly improved by using the VIRSC and the RIRSC. For example, if the researcher used all three IRFs to estimate the parameter of interest, he/she would incur in mis-specification biases that are sometimes three times as big as those of the parameter estimate based on the VIRSC and RIRSC. The researcher would also incur in much worse coverage probabilities if he/she focused on the wrong responses (and the empirical coverage can be as low as zero for those cases), and obtain standard errors that are significantly larger if some IRFs are erroneously included.

INSERT TABLES 7 AND 8 HERE

## 8 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 (IRFMEs). We propose simple and econometrically sound methods for doing so, that consist of information criteria. We show by Monte Carlo simulations that our methods can substantially improve the precision of the parameter estimates and decrease their biases, both in small samples (when the IRFs are correctly specified) as well as asymptotically (when the IRFs are mis-specified). We also show that our methods can substantially change inferences regarding key parameters of existing representative DSGE models. We hope that the simplicity and the usefulness of the criteria that we propose will increase the applicability of impulse response function matching estimators in practice.

Our criteria can be used not only for IRFMEs but also for general classical minimum distance and Indirect Inference Estimators. Relative to this literature, and in particular relative to Dridi et al. (2007), we add useful information criteria to select valid as well as relevant restrictions, thus significantly extending the scope of their analysis.

Finally, we do not provide a systematic analysis of the relative merits of using IRFMEs versus alternative estimators such as classical full information MLE or Bayesian methods. The latter estimators employ the entire likelihood of the model rather than the limited information approach of the IRFME with its focus on selected aspects of a DSGE model. The decision to pursue the IRFME over a full information approach gives rise to the usual trade-off between efficiency and robustness. We leave these issues for future research.

## 9 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.

Andrews, Donald W.K. (1999): "Consistent Moment Selection Procedures for Generalized Method of Moments Estimators," *Econometrica* 67(3), 543-564.

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.

Dridi, Ramdan, Alain Guay, and Eric Renault (2007), “Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models”, *Journal of Econometrics* 136, 397-430.

Dupor, Bill, Jing Han, and Yi Chan Tsai (2007): “What Do Technology Shocks Tell Us about the New Keynesian Paradigm?”, *Journal of Monetary Economics* 56, 560-569.

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 Changmook Shin (2007): “Information in Generalized Method of Moments Estimation and Entropy Based Moment Selection”, *Journal of Econometrics* 138, 488-512.

Hall, Alastair, Atsushi Inoue, James Nason and Barbara Rossi (2007): “Information Criteria for Impulse Response Function Matching Estimation of DSGE Models”, *Duke University Working Paper* 2007-04.

Hamilton, James (1994), *Time Series Analysis*, Princeton University Press.

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.

Kehoe, Patrick J. (2006): “How to Advance Theory with Structural VARs: Use the Sims-Cogley-Nason Approach”, *NBER Working Paper* 12575, Cambridge MA.

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 (1990), “Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models,” *Review of Economic Statistics* 72, 116– 125.

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, Chichester, 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”, *Institute for Empirical Macroeconomics Discussion Paper 11*, 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.

Watson, Mark W. (2006): “Comment on “Assessing Structural VARs” by L. Christiano, M. Eichenbaum, and R. Vigfusson”, *NBER Macroeconomics Annual*.

## Appendix A: Proofs

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

**Proposition 9 (Asymptotic Normality of Parameter Estimates – VAR case)** *Suppose that Assumptions (A1)–(A4) are satisfied. Then:*

$$\sqrt{T} \begin{pmatrix} \hat{\theta}_T(\bar{\phi}; c) - \theta_0(c) \\ \hat{\alpha}_T(\bar{\phi}; c) - \bar{\alpha}(c) \end{pmatrix} \xrightarrow{d} N(0, W(\bar{\phi}; c)).$$

*The proof follows from Proposition 3.5 of Dridi et al. (2007).*

**Proposition 10 (Asymptotic Normality of Parameter Estimates – Simulation-based case)**

*Suppose that Assumptions (A1), (A2), (A1'), (A3'), (A4'), (A5) are satisfied. Then:*

$$\sqrt{T} \begin{pmatrix} \hat{\theta}_{II,T}(\bar{\phi}; c) - \theta_0(c) \\ \hat{\alpha}_{II,T}(\bar{\phi}; c) - \bar{\alpha}(c) \end{pmatrix} \xrightarrow{d} N(0, W(S, \bar{\phi}; c)),$$

where

$$\begin{aligned} W(S, \bar{\phi}; c) &= \left[ \frac{\partial g'(\theta_0, \bar{\alpha}(c); c)}{\partial [\theta', \alpha']'} (\Phi_0(S, \bar{\phi}; c))^{-1} \frac{\partial g(\theta_0, \bar{\alpha}(c); c)}{\partial [\theta', \alpha']'} \right]^{-1} \\ \Omega(S, \bar{\phi}; c) &= \Phi_0^{-1}(S, \bar{\phi}; c) \\ \Phi_0(S, \bar{\phi}; c) &= J_0^{-1}(c) I_0(c) J_0^{-1}(c) + \frac{1}{S} J_0^{*-1}(\bar{\phi}; c) I_0^*(\bar{\phi}; c) J_0^{*-1}(\bar{\phi}; c) \\ &\quad + \left(1 - \frac{1}{S}\right) J_0^{*-1}(\bar{\phi}; c) K_0^*(\bar{\phi}; c) J_0^{*-1}(\bar{\phi}; c) \\ &\quad - J_0^{-1}(c) K_0(c) J_0^{*-1}(\bar{\phi}; c) - J_0^{*-1}(\bar{\phi}; c) K_0'(c) J_0^{-1}(c). \end{aligned}$$

*The proof follows from Proposition 3.5 of Dridi et al. (2007).*

**Proof of Theorem (1).** Consider two cases: (1) the case in which  $c_0$  and  $c \in C_H \cap C_{0,H}(\bar{\phi})^c$  are compared; and (2) the case in which  $c_0$  and  $c \in C_H \cap C_{0,H}(\bar{\phi})$  are compared. First, when  $c \in$

$C_H \cap C_{0,H}(\bar{\phi})^c$ ,  $(1/T)\hat{\Gamma}_T(\bar{\phi}; c) \xrightarrow{p} j(\bar{\phi}; c)$  for some constant  $j(\bar{\phi}; c) > 0$  whereas  $\hat{\Gamma}_T(\bar{\phi}; c_0) = O_p(1)$ .

Thus, it follows from Assumption B3 that

$$T^{-1}(VIRSC_T(\bar{\phi}; c_0) - VIRSC_T(\bar{\phi}; c)) = -j(\bar{\phi}; c) + o_p(1).$$

That is,  $(1/T)(VIRSC_T(\bar{\phi}; c_0) - VIRSC_T(\bar{\phi}; c))$  is negative with probability approaching one.

Next, when  $c \in C_H \cap C_{0,H}(\bar{\phi})$ ,  $\hat{\Gamma}_T(\bar{\phi}; c_0)$  and  $\hat{\Gamma}_T(\bar{\phi}; c)$  are both  $O_p(1)$ . Note that  $\hat{\Gamma}_T(\bar{\phi}; c)$  is  $O_p(1)$

whether or not  $\theta_0$  and  $\bar{\alpha}(c)$  are identified by the IRFs selected by  $c$ . By definition  $|c_0| > |c|$  and

Assumption B3,  $-h(|c_0|)\kappa_T + h(|c|)\kappa_T \rightarrow -\infty$ . Thus  $VIRSC_T(\bar{\phi}; c_0) - VIRSC_T(\bar{\phi}; c) \xrightarrow{p} -\infty$ .

Combining these two results,  $VIRSC_T(\bar{\phi}; c_0) < VIRSC_T(\bar{\phi}; c)$  for all  $c \neq c_0$  with probability approaching one.

**Proof of Theorem (2).** First suppose that  $c \in C_{\mathcal{E},H}(\bar{\phi})$  and  $c \neq c_r$ . It follows from Proposition 9 and Assumptions C2 and C3 that

$$\begin{aligned} T^{1/2}(RIRSC(\bar{\phi}; c) - RIRSC(\bar{\phi}; c_r)) &= T^{1/2}(\ln(|\hat{W}_T(\bar{\phi}; c)|) - \ln(|\hat{W}_T(\bar{\phi}; c_r)|)) \\ &\quad + T^{1/2}(k(|c|) - k(|c_r|))m_T \\ &\rightarrow +\infty \end{aligned} \tag{28}$$

as the first term is  $O_p(1)$  by Assumption C2 and the second term diverges to infinity by Assumption

C3. Thus  $T^{1/2}(RIRSC(\bar{\phi}; c) - RIRSC(\bar{\phi}; c_0))$  is positive with probability approaching one as

$T \rightarrow \infty$ . Next consider the case in which  $c \in C_{I,H}(\bar{\phi}) \cap [C_{\mathcal{E},H}(\bar{\phi})]^c$ . By Theorem 22 of Magnus and Neudecker (1999, p.21), it follows from Assumption C1 that  $\ln(|W_T(\bar{\phi}; c)|) - \ln(|W_T(\bar{\phi}; c_r)|) > 0$ .

Thus it follows from Assumptions C2 and C3 that

$$\begin{aligned} &RIRSC(\bar{\phi}; c) - RIRSC(\bar{\phi}; c_r) \\ &= \ln(|\hat{W}_T(\bar{\phi}; c)|) - \ln(|\hat{W}_T(\bar{\phi}; c_r)|) + k(|c|)m_T - k(|c_r|)m_T \\ &= \ln(|W_T(\bar{\phi}; c)|) - \ln(|W_T(\bar{\phi}; c_r)|) + o_p(1) \\ &> 0 \end{aligned} \tag{29}$$

with probability approaching one. Third, when  $c \in C_{NI,H}(\bar{\phi}) \cap [C_{\mathcal{E},H}(\bar{\phi})]^c$  where

$$C_{NI,H}(\bar{\phi}) = \{c \in C_{0,H}(\bar{\phi}) : c \notin C_{I,H}(\bar{\phi})\}, \tag{30}$$

it follows from Assumption C2 that

$$RIRSC(\bar{\phi}; c) - RIRSC(\bar{\phi}; c_r) \xrightarrow{p} +\infty. \tag{31}$$

Because  $C_{\mathcal{E},H}(\bar{\phi}) \cup (C_{I,H}(\bar{\phi}) \cap C_{\mathcal{E},H}^c(\bar{\phi})) \cup (C_{NI,H}(\bar{\phi}) \cap C_{\mathcal{E},H}^c(\bar{\phi})) = C_{I,H}(\bar{\phi}) \cup C_{NI,H}(\bar{\phi}) = C_{0,H}(\bar{\phi})$ ,

$$RIRSC(\bar{\phi}; c_r) < RIRSC(\bar{\phi}; c) \quad (32)$$

for all  $c \in C_H(\bar{\phi})$  such that  $c \neq c_r$  with probability approaching one asymptotically. Since, by definition,  $\hat{c}_T$  minimizes  $RIRSC(\bar{\phi}; c)$ :

$$RIRSC(\bar{\phi}; \hat{c}_T) \leq RIRSC(\bar{\phi}; c)$$

for all  $c \in C_H(\bar{\phi})$ , then,

$$RIRSC(\bar{\phi}; \hat{c}_T) \leq RIRSC(\bar{\phi}; c_r) \quad (33)$$

Therefore it follows from (32), (33) and Assumption C3 that  $\hat{c}_{RIRSC,T} \xrightarrow{p} c_r$ .

**Proof of Theorem (5) and (6).** The proofs are as in Theorem 1 and 2.

**Proof of Theorem (7).** Let  $D\tilde{g}_T^S(\theta, \alpha, \phi)$  denote the Jacobian of  $\tilde{g}_T^S(\theta, \alpha, \phi)$  with respect to  $\theta$  and  $\alpha$ . Note that the first order conditions are written as:

$$D\tilde{g}_T^S(\hat{\theta}_{II,T}(S, \bar{\phi}; c), \hat{\alpha}_{II,T}(S, \bar{\phi}; c), \bar{\phi}; c)' \hat{\Omega}_T(S, \bar{\phi}; c) (\hat{\gamma}_T(c) - \tilde{g}_T^S(\hat{\theta}_{II,T}(S, \bar{\phi}; c), \hat{\alpha}_{II,T}(S, \bar{\phi}; c), \bar{\phi}; c)) = 0, \quad (34)$$

It follows from (34) that, for  $\theta_0(\bar{\phi}; c)$  and  $\alpha_0(\bar{\phi}; c)$  defined in Assumption D(a):

$$\begin{aligned} & \left( \frac{T}{2l \ln T} \right)^{\frac{1}{2}} \begin{bmatrix} \hat{\theta}_{II,T}(S, \bar{\phi}; c) - \theta_0(\bar{\phi}; c) \\ \hat{\alpha}_{II,T}(S, \bar{\phi}; c) - \alpha_0(\bar{\phi}; c) \end{bmatrix} \\ &= - \left[ D\tilde{g}_T^S(\hat{\theta}_{II,T}(S, \bar{\phi}; c), \hat{\alpha}_{II,T}(S, \bar{\phi}; c), \bar{\phi}; c)' \hat{\Omega}_T(S, \bar{\phi}; c) D\tilde{g}_T^S(\bar{\theta}_T(S, \bar{\phi}; c), \bar{\alpha}_T(S, \bar{\phi}; c), \bar{\phi}; c) \right]^{-1} \\ & \quad D\tilde{g}_T^S(\hat{\theta}_{II,T}(S, \bar{\phi}; c), \hat{\alpha}_{II,T}(S, \bar{\phi}; c), \bar{\phi}; c)' \hat{\Omega}_T(S, \bar{\phi}; c) \left( \frac{T}{2l \ln T} \right)^{\frac{1}{2}} (\hat{\gamma}_T(c) - \tilde{g}_T^S(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c)) \\ &= - \left[ Dg(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi})' \Omega(S, \bar{\phi}; c) Dg(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c) \right]^{-1} \\ & \quad Dg(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c)' \Omega(S, \bar{\phi}; c) \left( \frac{T}{2l \ln T} \right)^{\frac{1}{2}} (\hat{\gamma}_T(c) - \tilde{g}_T^S(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c)) + o(1) \quad \mathbf{a(35)} \end{aligned}$$

where  $\bar{\theta}_T(\bar{\phi}; c)$  [ $\bar{\alpha}_T(\bar{\phi}; c)$ ] is a point between  $\theta_0(\bar{\phi}; c)$  and  $\hat{\theta}_{II,T}(S, \bar{\phi}; c)$  [resp.  $\alpha_0(\bar{\phi}; c)$  and  $\hat{\alpha}_{II,T}(S, \bar{\phi}; c)$ ], the first equality follows from the mean-value theorem, and the second from Assumptions D(a)–D(c). Suppose that  $c \in C_{0,H}(\bar{\phi})$  so that the IRFs selected by  $c$  are valid. Then it follows from

Assumption D that

$$\begin{aligned}
& \left( \frac{1}{2l\ln lT} \right) \widehat{\Gamma}_T(\bar{\phi}; c) \\
= & \left( \frac{T}{2l\ln lT} \right) [\hat{\gamma}_T(c) - \bar{g}_T^S(\hat{\theta}_{II,T}(S, \bar{\phi}; c), \hat{\alpha}_{II,T}(S, \bar{\phi}; c), \bar{\phi}; c)]' \hat{\Omega}_T(S, \bar{\phi}; c) [\hat{\gamma}_T(c) - \bar{g}_T^S(\hat{\theta}_{II,T}(S, \bar{\phi}; c), \hat{\alpha}_{II,T}(S, \bar{\phi}; c), \bar{\phi}; c)] \\
= & \left( \frac{T}{2l\ln lT} \right) \left[ \hat{\gamma}_T(c) - g(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c) - Dg(\bar{\theta}_T(\bar{\phi}; c), \bar{\alpha}_T(\bar{\phi}; c), \bar{\phi}; c) \begin{pmatrix} \hat{\theta}_{II,T}(S, \bar{\phi}; c) - \theta_0(\bar{\phi}; c) \\ \hat{\alpha}_{II,T}(S, \bar{\phi}; c) - \alpha_0(\bar{\phi}; c) \end{pmatrix} \right]' \\
& \hat{\Omega}_T(S, \bar{\phi}; c) \left[ \hat{\gamma}_T(c) - g(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c) - Dg(\bar{\theta}_T(\bar{\phi}; c), \bar{\alpha}_T(\bar{\phi}; c), \bar{\phi}; c) \begin{pmatrix} \hat{\theta}_{II,T}(S, \bar{\phi}; c) - \theta_0(\bar{\phi}; c) \\ \hat{\alpha}_{II,T}(S, \bar{\phi}; c) - \alpha_0(\bar{\phi}; c) \end{pmatrix} \right] \\
= & \left( \frac{T}{2l\ln lT} \right) (\hat{\gamma}_T(c) - g(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c))' M_1' \Omega(S, \bar{\phi}; c) M_1 (\hat{\gamma}_T(c) - g(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c)) \\
& + o(1) \text{ a.s.} \\
= & \left( \frac{T}{2l\ln lT} \right) (\hat{\gamma}_T(c) - g(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c))' \Omega^{\frac{1}{2}}(S, \bar{\phi}; c) \\
& M_2 \Omega^{\frac{1}{2}}(S, \bar{\phi}; c) (\hat{\gamma}_T(c) - g(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c)) + o(1) \text{ a.s.} \tag{36}
\end{aligned}$$

where

$$\begin{aligned}
M_1 &= I_{|c|} - Dg(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}) [Dg(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c)]' \Omega(S, \bar{\phi}; c) Dg(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c)]^{-1} Dg(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi})' \Omega(S, \bar{\phi}; c) \\
M_2 &= I_{|c|} - \Omega^{\frac{1}{2}}(S, \bar{\phi}; c) Dg(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c) [Dg(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c)]' \Omega(S, \bar{\phi}; c) Dg(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c)]^{-1} \\
& Dg(\theta_0(\bar{\phi}; c), \alpha_0(\bar{\phi}; c), \bar{\phi}; c)' \Omega^{\frac{1}{2}}(S, \bar{\phi}; c).
\end{aligned}$$

Since  $M_2$  is symmetric and idempotent with rank  $|c| - p_\theta - p_\alpha$ , it follows from the Schur decomposition theorem that there is a  $(|c| \times (|c| - p_\theta - p_\alpha))$  matrix  $S$  such that  $M_2 = SS'$  and  $S'S = I_{|c| - p_\theta - p_\alpha}$ .

Thus it follows from (36) and Assumption D(c) that

$$\limsup_{T \rightarrow \infty} \left( \frac{1}{2l\ln lT} \right) \widehat{\Gamma}_T(\bar{\phi}; c) = |c| - p_\theta - p_\alpha. \text{ a.s.} \tag{37}$$

Suppose that  $c_1$  and  $c_2$  both select valid IRF with  $|c_1| > |c_2|$ . It follows from (37) that

$$\limsup_{T \rightarrow \infty} \frac{1}{2l\ln lT} \widehat{\Gamma}_T(\bar{\phi}; c_1) = |c_1| - p_\theta - p_\alpha, \text{ a.s.}, \tag{38}$$

$$\limsup_{T \rightarrow \infty} \frac{1}{2l\ln lT} \widehat{\Gamma}_T(\bar{\phi}; c_2) = |c_2| - p_\theta - p_\alpha, \text{ a.s.}, \tag{39}$$

whereas

$$\frac{T}{l\ln lT} (-h(|c_1|)\kappa_T + h(|c_2|)\kappa_T) = \frac{T}{l\ln lT} (-h(|c_1|) + h(|c_2|))\kappa_T \rightarrow -\infty \tag{40}$$

Combining (34), (35) and (36) yields

$$VIRSC_T(\bar{\phi}; c_1) < VIRSC_T(\bar{\phi}; c_2) + o_{as}(1) \tag{41}$$

whenever  $c_1, c_2 \in C_{0,H}(\bar{\phi})$  with  $|c_1| < |c_2|$ .

Consider the case in which  $c_1 \in C_{0,H}(\bar{\phi})$  and  $c_2 \in C_H(\bar{\phi}) \setminus C_{0,H}(\bar{\phi})$ . Then

$$\frac{1}{T}\hat{\Gamma}_T(\bar{\phi}; c_1) = o(1) \text{ a.s.}, \quad (42)$$

$$\frac{1}{T}\hat{\Gamma}_T(\bar{\phi}; c_2) > 0, \text{ a.s.}, \quad (43)$$

while it follows from Assumption B3 that

$$\frac{1}{T}h(|c_1|)\kappa_T \rightarrow 0, \quad (44)$$

$$\frac{1}{T}h(|c_2|)\kappa_T \rightarrow 0, \quad (45)$$

It follows from (44)–(45) that

$$\frac{1}{T}VIRSC_T(\bar{\phi}; c_1) < \frac{1}{T}VIRSC_T(\bar{\phi}; c_2) + o(1) \text{ a.s.} \quad (46)$$

whenever  $c_1 \in C_{0,H}(\bar{\phi})$  and  $c_2 \in C_H(\bar{\phi}) \setminus C_{0,H}(\bar{\phi})$ . The desired result follows from (41), (46) and Assumption B1.

**Proof of Theorem (8).** Let

$$\hat{M}_T(c) = \hat{W}_T^{(1,1)}(S, \bar{\phi}; c)^{-1} = \hat{G}_T(c)' \hat{\Omega}_T(S, \bar{\phi}; c) \hat{G}_T(c),$$

where  $\hat{G}_T(c) = \partial \hat{g}_T^S(\hat{\theta}_{II,T}(S, \bar{\phi}; c), \hat{\alpha}_{II,T}(S, \bar{\phi}; c), \bar{\phi}; c) / \partial([\theta', \alpha']')$ . It implies

$$RIRSC_{II,T}(\bar{\phi}; c) = -\ln[|\hat{M}_T(c)|] + k(|c|)m_T. \quad (47)$$

Also define  $M(c) = G(c)' \Omega(S, \bar{\phi}; c) G(c)$ . We have the following expression for  $\hat{M}_T(c) - M(c)$ :

$$\begin{aligned} \hat{M}_T(c) - M(c) &= \hat{G}_T(c)' \hat{\Omega}_T(S, \bar{\phi}; c) \hat{G}_T(c) - G(c)' \Omega(S, \bar{\phi}; c) G(c) \\ &= \hat{G}_T(c)' \hat{\Omega}_T(S, \bar{\phi}; c) \{ \hat{G}_T(c) - G(c) \} + \hat{G}_T(c)' \{ \hat{\Omega}_T(S, \bar{\phi}; c) - \Omega(S, \bar{\phi}; c) \} G(c) \\ &\quad + \{ \hat{G}_T(c) - G(c) \}' \Omega(S, \bar{\phi}; c) G(c) \end{aligned} \quad (48)$$

We also have the following representation:

$$\hat{\Omega}_T(S, \bar{\phi}; c) - \Omega(S, \bar{\phi}; c) = \Omega(S, \bar{\phi}; c) \left\{ \Phi_0(S, \bar{\phi}; c) - \hat{\Phi}_T(S, \bar{\phi}; c) \right\} \hat{\Omega}_T(S, \bar{\phi}; c) \quad (49)$$

and

$$\text{vec} \left\{ \Phi_0(S, \bar{\phi}; c) - \hat{\Phi}_T(S, \bar{\phi}; c) \right\} = \Delta \text{vech} \left\{ \Phi_0(S, \bar{\phi}; c) - \hat{\Phi}_T(S, \bar{\phi}; c) \right\} \quad (50)$$

where (49) follows since  $\hat{\Omega}_T(S, \bar{\phi}; c) = (\hat{\Phi}_T(S, \bar{\phi}; c))^{-1}$  and  $\Delta$  is the matrix such that  $vec(\cdot) = \Delta vec(\cdot)$ . The above equations are the foundations for the analysis. The proof rests on equations derived in the following three steps.

*Step 1:* From (48)–(50) and Assumptions E(a)–E(b), it follows that

$$\begin{aligned} \hat{M}_T(c) - M(c) &= G(c)' \hat{\Omega}_T(S, \bar{\phi}; c) \{\hat{G}_T(c) - G(c)\} + G(c)' \{\hat{\Omega}_T(S, \bar{\phi}; c) - \Omega(S, \bar{\phi}; c)\} G(c) \\ &\quad + \{\hat{G}_T(c) - G(c)\}' W(c) G(c) + o(\{T/\ln \ln T\}^{-1/2}) \quad a.s. \\ &\equiv a_T(c) + o(\{T/\ln \ln T\}^{-1/2}) \end{aligned} \quad (51)$$

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

$$tr \left\{ M(c)^{-1} (\hat{M}_T(c) - M(c)) \right\} = vec\{M(c)^{-1}\}' vec\{\hat{M}_T(c) - M(c)\} \quad (52)$$

>From (51), it follows that  $a_T(c)$  can be written as

$$\begin{aligned} a_T(c) &= vec\{G(c)' \hat{\Omega}_T(S, \bar{\phi}; c) [\hat{G}_T(c) - G(c)]\} + vec\{G(c)' [\hat{\Omega}_T(S, \bar{\phi}; c) - \Omega(S, \bar{\phi}; c)] G(c)\} \\ &\quad + vec\{[\hat{G}_T(c) - G(c)]' \Omega(S, \bar{\phi}; c) G(c)\} \end{aligned} \quad (53)$$

$$= a_{1,T}(c) + a_{2,T}(c) + a_{3,T}(c). \quad (54)$$

Taking the terms of the right hand side of (54) in turn, we have

For the first term,  $a_{1,T}(c)$ :

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

$$\begin{aligned} a_{1,T}(c) &= vec\{G(c)' \hat{\Omega}_T(S, \bar{\phi}; c) [\hat{G}_T(c) - G(c)]\} \\ &= \left[ I_p \otimes G(c)' \hat{\Omega}_T(S, \bar{\phi}; c) \right] vec\{\hat{G}_T(c) - G(c)\} \end{aligned} \quad (55)$$

Using Assumptions E(a) and E(b) and eq. (55), it follows that

$$\begin{aligned} a_{1,T}(c) &= - \left[ I_{p_\theta + p_\alpha} \otimes G(c)' \Omega(S, \bar{\theta}; c) \right] \\ &\quad \times \left\{ G(c) \left\{ G(c)' \Omega(S, \bar{\theta}; c) G(c) \right\}^{-1} G(c)' \Omega(S, \bar{\theta}; c) (\hat{\gamma}_T(c) - \tilde{g}_T^S(S, \theta_0, \bar{\alpha}(c), \bar{\phi}; c)) \right\} \\ &\quad + o(\{T/\ln \ln T\}^{-1/2}) \quad a.s. \end{aligned} \quad (56)$$

where the rate follows from the Law of the Iterated Logarithm in Assumption E(b).

For the second term,  $a_{2,T}(c)$ :

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

$$\begin{aligned} a_{2,T}(c) &= vec\{G(c)' [\hat{\Omega}_T(S, \bar{\phi}; c) - \Omega(S, \bar{\phi}; c)] G(c)\} \\ &= \left[ G(c)' \hat{\Omega}_T(S, \bar{\phi}; c) \otimes G(c)' \Omega(S, \bar{\phi}; c) \right] vec\{\Phi_0(S, \bar{\phi}; c) - \hat{\Phi}_T(S, \bar{\phi}; c)\} \end{aligned} \quad (57)$$

Using (50), (57), and Assumptions E(a) and E(b), it follows that

$$\begin{aligned}
a_{2,T}(c) &= - [G(c)' \Omega(S, \bar{\phi}; c) \otimes G(c)' \Omega(S, \bar{\phi}; c)] T^{-1} \sum_{t=1+m}^T \omega_v(Y_t, Y_t^s(\theta_0, \bar{\alpha}(c), \bar{\phi}, Y_0^s), \theta_0; c) \\
&\quad + o(\{T/\ln \ln T\}^{-1/2}) \quad a.s.
\end{aligned} \tag{58}$$

For the third term,  $a_{3,T}(c)$ :

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

$$\begin{aligned}
a_{3,T}(c) &= \text{vec}\{[\hat{G}_T(c) - G(c)]' \Omega(S, \bar{\phi}; c) G(c)\} \\
&= [G(c)' \Omega(S, \bar{\phi}; c) \otimes I_{|c|}] \text{vec}\{\hat{G}_T(c)' - G(c)'\} \\
&= [G(c)' \Omega(S, \bar{\phi}; c) \otimes I_{|c|}] N \text{vec}\{\hat{G}_T(c) - G(c)\}
\end{aligned} \tag{59}$$

where  $N$  is the permutation matrix such that  $\text{vec}(A') = N \text{vec}(A)$ . It follows from (59) and Assumptions E(a) and E(b) that

$$\begin{aligned}
a_{3,T}(c) &= [G(c)' \Omega(S, \bar{\phi}; c) \otimes I_{|c|}] N \bar{G}(c) (\hat{\theta}_T(c) - \theta_0) \\
&\quad + o((T/\ln \ln T)^{-1/2}) \quad a.s.
\end{aligned} \tag{60}$$

Using Assumptions E(a) and E(b) and eq. (60), it follows that

$$\begin{aligned}
a_{3,T}(c) &= - [G(c)' \Omega(S, \bar{\phi}; c) \otimes I_{|c|}] N \\
&\quad \{ \bar{G}_T(c) \{G(c)' \Omega(S, \bar{\phi}; c) G(c)\}^{-1} G(c)' \Omega(S, \bar{\phi}; c) (\hat{\gamma}(c) - \hat{g}_T^S(\theta_0; c)) \} \\
&\quad + o(\{T/\ln \ln T\}^{-1/2}) \quad a.s.
\end{aligned} \tag{61}$$

*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  $\|M - M_0\| \|M_0^{-1}\| < 1$ ,

$$\begin{aligned}
\ln[|M|] &= \ln[|M_0|] + \text{tr} \{M_0^{-1}(M - M_0)\} - \text{tr} \{(M - M_0)M_0^{-1}(M - M_0)M_0^{-1}\} \\
&\quad + o\left(\frac{\|M^{-1}\|^3 \|(M - M_0)\|^3}{1 - \|M^{-1}\| \|M - M_0\|}\right)
\end{aligned} \tag{62}$$

Setting  $M = \hat{M}_T(c)$  and  $M_0 = M(c)$  and using (51), (52), (54), (56), (58) and (61) we obtain

$$\ln[|\hat{M}_T(c)|] = \ln[|M(c)|] + \text{tr} \left\{ M(c)^{-1} (\hat{M}_T(c) - M(c)) \right\} + o(\nu_T^{-1}) \quad a.s. \tag{63}$$

where  $-\text{tr}\{(M - M_0)M_0^{-1}(M - M_0)M_0^{-1}\} + o\left(\frac{\|M^{-1}\|^3 \|(M - M_0)\|^3}{1 - \|M^{-1}\| \|M - M_0\|}\right) = o(\nu_T^{-1})$ ,  $\{T/\ln \ln T\}^{1/2}/\nu_T \rightarrow 0$  and  $\|M - M_0\| \|M_0^{-1}\| < 1$  as  $T \rightarrow \infty$ .

From (63), (52), (56), (58) and (61), it follows that

$$\begin{aligned} \ln[|\hat{M}_T(c)|] &= \ln[|M(c)|] + \text{vec}\{M(c)^{-1}\}' D(c) T^{-1} \sum_{t=1}^T \omega(Y_t, Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s), \theta_0; c) \\ &\quad + o(\{T/\ln\ln T\}^{-1/2}) \quad a.s. \end{aligned} \quad (64)$$

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

$$\begin{aligned} D_1(c) &= -\{[I_{p_\theta+p_\alpha} \otimes G(c)' \Omega(S, \bar{\phi}; c)] + [G(c)' \Omega(S, \bar{\phi}; c) \otimes I_{|c|} N] \bar{G}_T(c) \\ &\quad \times \{G(c)' \Omega(S, \bar{\phi}; c) G(c)\}^{-1} G(c)' \Omega(S, \bar{\phi}; c) \end{aligned} \quad (65)$$

$$D_2(c) = -[G(c)' \Omega(S, \bar{\phi}; c) \otimes G(c)' \Omega(S, \bar{\phi}; c)] \Delta \quad (66)$$

where  $\Delta$  is as in eq. (50) and  $N$  is defined after (59). Now define

$$\xi_t(c) = \text{vec}\{M(c)^{-1}\}' D(c) \omega(Y_t, Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s), \theta_0; c)$$

and  $\omega_\xi^2(c) = \lim_{T \rightarrow \infty} \text{Var}[T^{-1/2} \sum_{t=1}^T \xi_t(c)]$ . Then

$$\begin{aligned} \left(\frac{T}{\ln\ln T}\right)^{1/2} RIRSC(\bar{\phi}; c) &= -\left(\frac{T}{\ln\ln T}\right)^{1/2} \ln[|M(c)|] - \left(\frac{T}{\ln\ln T}\right)^{1/2} T^{-1} \sum_{t=1}^T \xi_t(c) \\ &\quad + \left(\frac{T}{\ln\ln T}\right)^{1/2} \kappa(|c|) m_T + o(1) \quad a.s. \end{aligned} \quad (67)$$

We now use the above results to establish Theorem 8. The proof proceeds by considering two cases.

Part (i): Consider  $c_1$  and  $c_2$  such that  $W^{(1,1)}(S, \bar{\phi}; c_1) - W^{(1,1)}(S, \bar{\phi}; c_2)$  is *p.s.d.* and hence  $\ln[|M(c_2)|] > \ln[|M(c_1)|]$ . Since  $\kappa(|c|) m_T = o(1)$  from Assumption F, it follows from (67) and Assumption E(b) that

$$RIRSC(\bar{\phi}; c_1) - RIRSC(\bar{\phi}; c_2) = \ln[|M(c_2)|] - \ln[|M(c_1)|] + o(1) \quad a.s. \quad (68)$$

Since  $W(S, \bar{\phi}; c) - W(S, \bar{\phi}; c_r)$  is *p.s.d.* for all  $c \in C_H$ , it follows from (68) that  $RIRSC(\bar{\phi}; c) \geq RIRSC(\bar{\phi}; c_\mathcal{E})$  *a.s.* for any  $c \in C_H$  and  $c_\mathcal{E} \in C_{\mathcal{E}, H}$ . Because  $RIRSC(\bar{\phi}; c) \geq RIRSC(\bar{\phi}; \hat{c}_T)$  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  $c_a \in C_{\mathcal{E}, H}$  such that  $c_r \neq c_a$ . From Assumption E(b), it follows that for  $c = c_r, c_a$  we have

$$\limsup_{T \rightarrow \infty} \left(\frac{T}{\ln\ln T}\right)^{1/2} |T^{-1} \sum_{t=1}^T \xi_t(c)| \leq 2^{1/2} \omega_\xi(c), \quad a.s. \quad (69)$$

Set  $r(\bar{\phi}; c_r, c_a) = RIRSC(\bar{\phi}; c_r) - RIRSC(\bar{\phi}; c_a)$ . Since  $M(c_r) = M(c_a)$  by definition in this case, it follows from (67) and (69) that

$$\begin{aligned}
\limsup_{T \rightarrow \infty} \left( \frac{T}{\ln \ln T} \right)^{1/2} r(\bar{\phi}; c_r, c_a) &\leq \limsup_{T \rightarrow \infty} \left( \frac{T}{\ln \ln T} \right)^{1/2} \left| T^{-1} \sum_{t=1}^T \xi_t(c_r) \right| \\
&\quad + \limsup_{T \rightarrow \infty} \left( \frac{T}{\ln \ln T} \right)^{1/2} \left| T^{-1} \sum_{t=1}^T \xi_t(c_a) \right| \\
&\quad - \liminf_{T \rightarrow \infty} \left( -k(|c_r|, T) + k(|c_a|, T) \right) \\
&\leq 2^{1/2} (\omega_\xi(c_r) + \omega_\xi(c_a)) + \\
&\quad - \liminf_{T \rightarrow \infty} \left\{ k(|c_a|) - k(|c_r|) (T^{1/2} m_T / (\ln \ln T)^{1/2}) \right\} \quad (70)
\end{aligned}$$

Using Assumptions E(b) and C1, it follows from (70) that  $\hat{c}_T = c_r$  for  $T$  sufficiently large *a.s.* if Assumption F 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).

## Appendix B: Asymptotic Covariance Estimation Formulas

A common choice for the estimate of the weighting matrix  $\hat{\Omega}_T(c)$  is the inverse of the estimated asymptotic covariance of the IRFs. Let  $\hat{\Sigma}_{\hat{\gamma}_T(c)}$  denote the estimate of the asymptotic covariance matrix of the sample IRFs (see Hamilton, 1994, Section 11.7 for formulas).

**Definition 11 (Consistent Estimation of VAR-based Estimators.)** *Consistent estimates of the matrices of interest can be obtained by standard HAC estimators (e.g., Newey and West, 1987).*

*Let  $b(T)$  be a bandwidth that grows with  $T$  and suppose that there are  $q(Y_t, \gamma_0(c); c)$  such that*

$$Q_T(\underline{Y}_T, \gamma_0(c); c) = \sum_{t=1}^T q(Y_t, \gamma_0(c); c) + o_p\left(T^{1/2}\right) \text{ and}$$

$$\hat{\gamma}_T(c) = \gamma_0(c) - \left[ \frac{\partial^2 Q_T(\underline{Y}_T, \gamma_0(c); c)}{\partial \gamma(c) \partial \gamma'(c)} \right]^{-1} \sum_{t=1}^T \frac{\partial q(Y_t, \gamma_0(c); c)}{\partial \gamma(c)} + o_p(1).$$

*Then,*

$$\hat{I}_T(c) \equiv \sum_{i=-b(T)+1}^{b(T)-1} (1 - |i/b(T)|) \frac{1}{T} \sum_{t=1}^T \left[ \frac{\partial q(Y_t, \gamma_0(c); c)}{\partial \gamma(c)} \right] \left[ \frac{\partial q(Y_{t-i}, \gamma_0(c); c)}{\partial \gamma(c)} \right]'$$

$$\hat{J}_T(c) \equiv \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 Q_t}{\partial \gamma(c) \partial \gamma'(c)}(\underline{Y}_T, \gamma_0(c); c)$$

$$\begin{aligned}\widehat{\Phi}_T(c) &\equiv \widehat{J}_T(c)^{-1} \widehat{I}_T(c) \widehat{J}_T(c)^{-1} \\ \widehat{G}_T(\bar{\phi}; c) &\equiv \frac{\partial g(\widehat{\theta}_T(\bar{\phi}; c), \widehat{\alpha}_T(\bar{\phi}; c), \bar{\phi}; c)}{\partial[\theta', \alpha']'} \\ \widehat{W}_T(\bar{\phi}; c) &\equiv \left[ \widehat{G}_T(\bar{\phi}; c)' \widehat{\Phi}_T(c)^{-1} \widehat{G}_T(\bar{\phi}; c) \right]^{-1}\end{aligned}$$

are consistent estimators of  $I_0(c)$ ,  $J_0(c)$ ,  $\Phi_0(c)$ ,  $\frac{\partial g(\theta_0, \bar{\alpha}(c), \bar{\phi}; c)}{\partial[\theta', \alpha']'}$ ,  $W(\bar{\phi}; c)$ .

When eq. 2 holds, and for the optimal choice of the weighting matrix  $\widehat{\Omega}_T(c) = \widehat{\Sigma}_{\widehat{\gamma}_T(c)}^{-1}$ , the formulas simplify to:  $\widehat{I}_T(c) = \widehat{\Omega}_T(c) \widehat{\Sigma}_{\widehat{\gamma}_T(c)} \widehat{\Omega}_T(c)'$ ,  $\widehat{J}_T(c) = \widehat{\Omega}_T(c)$ ,  $\widehat{\Phi}_T(c) = \widehat{\Sigma}_{\widehat{\gamma}_T(c)}$ .

**Definition 12 (Estimation of Asymptotic Variance of Simulation-based Estimators)** Let

$Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s)$  denote the time  $t$  simulated data from the simulated sample  $\underline{Y}_T^s(\theta, \alpha, \bar{\phi}, Y_0^s)$ . Suppose that there are  $q(Y_t, \gamma_0(c); c)$  and  $q(Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s), \gamma_0(c); c)$  such that

$$\begin{aligned}Q_T(\underline{Y}_T, \gamma_0(c); c) &= \sum_{t=1}^T q(Y_t, \gamma_0(c); c), \\ Q_T(\underline{Y}_T^s(\theta, \alpha, \bar{\phi}, Y_0^s), \gamma_0(c); c) &= \sum_{t=1}^T q(Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s), \gamma_0(c); c), \\ \widehat{\gamma}_T(c) &= \gamma_0(c) - \left[ \frac{\partial^2 Q_T(\underline{Y}_T, \gamma_0(c); c)}{\partial \gamma(c) \partial \gamma'(c)} \right]^{-1} \sum_{t=1}^T \frac{\partial q(Y_t, \gamma_0(c); c)}{\partial \gamma(c)} + o_p(1).\end{aligned}$$

and

$$\begin{aligned}\widetilde{\gamma}_T^{(s)}(\theta, \alpha, \bar{\phi}; c) &= \gamma_0(c) - \left[ \frac{\partial^2 Q_T(\underline{Y}_T^s(\theta, \alpha, \bar{\phi}, Y_0^s), \gamma_0(c); c)}{\partial \gamma(c) \partial \gamma'(c)} \right]^{-1} \\ &\quad \times \sum_{t=1}^T \frac{\partial q(Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s), \gamma_0(c); c)}{\partial \gamma(c)} + o_p(1).\end{aligned}$$

Let

$$\begin{aligned}r_t(\underline{Y}_T, \underline{Y}_T^s(\theta, \alpha, \bar{\phi}, Y_0^s); c) &= - \left[ \frac{\partial^2 Q_T(\underline{Y}_T, \gamma_0(c); c)}{\partial \gamma(c) \partial \gamma'(c)} \right]^{-1} \frac{\partial q(Y_t, \gamma_0(c); c)}{\partial \gamma(c)} \\ &\quad + \left[ \frac{\partial^2 Q_T(\underline{Y}_T^s(\theta, \alpha, \bar{\phi}, Y_0^s), \gamma_0(c); c)}{\partial \gamma(c) \partial \gamma'(c)} \right]^{-1} \frac{\partial q(Y_t^s(\theta, \alpha, \bar{\phi}, Y_0^s), \gamma_0(c); c)}{\partial \gamma(c)}\end{aligned}$$

Let  $b(T)$  be a bandwidth that grows with  $T$ . Then

$$\begin{aligned}\widehat{\Phi}_T(S, \bar{\phi}; c) &\equiv \sum_{i=-b(T)+1}^{b(T)-1} (1 - |i/b(T)|) \frac{1}{T} \sum_{t=i+1}^T r_t(\underline{Y}_T, \underline{Y}_T^s(\widehat{\theta}_{II,T}(S, \bar{\phi}; c), \widehat{\alpha}_{II,T}(S, \bar{\phi}; c), \bar{\phi}, Y_0^s); c) \times \\ &\quad r_{t-i}(\underline{Y}_T, \underline{Y}_T^s(\widehat{\theta}_{II,T}(S, \bar{\phi}; c), \widehat{\alpha}_{II,T}(S, \bar{\phi}; c), \bar{\phi}, Y_0^s); c)',\end{aligned}$$

$$\widehat{G}_T(S, \bar{\phi}; c) \equiv \frac{\partial \widetilde{g}_T^S(\widehat{\theta}_{II,T}(S, \bar{\phi}; c), \widehat{\alpha}_{II,T}(S, \bar{\phi}; c), \bar{\phi}; c)}{\partial [\theta', \alpha']'}$$

and

$$\widehat{W}_{II,T}(S, \bar{\phi}; c) \equiv \left[ \widehat{G}_T(S, \bar{\phi}; c)' \widehat{\Phi}_T(S, \bar{\phi}; c)^{-1} \widehat{G}_T(S, \bar{\phi}; c) \right]^{-1} \quad (71)$$

are consistent for  $\Phi_0(c)$ ,  $\frac{\partial g(\theta_0, \bar{\alpha}(c), \bar{\phi}; c)}{\partial [\theta', \alpha']'}$  and  $W_0(\bar{\phi}; c)$ , respectively, under suitable assumptions.

## 10 Tables

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

Parameters	RIRSC ( $\hat{h}_T = 3$ )		Fixed lags (h=20)	
	Parameter	Standard	Parameter	Standard
	Estimates	Errors	Estimates	Errors
$\rho_{xM}$	-0.097	0.247	-0.040	0.292
$\rho_{xz}$	0.588	1.257	0.329	0.948
$c_z$	0.655	0.664	2.952	3.096
$\rho_{\mu_z}$	0.237	0.703	0.894	0.159
$\rho_{x\Upsilon}$	0.997	0.107	0.822	0.345
$c_\Upsilon$	0.307	0.435	0.247	0.440
$\rho_{\mu\Upsilon}$	0.344	0.240	0.239	0.425
$\sigma_M$	0.334	0.113	0.333	0.110
$\sigma_{\mu_z}$	0.203	0.168	0.069	0.068
$\sigma_{\mu\Upsilon}$	0.287	0.084	0.304	0.093
$\varepsilon$	0.831	0.284	0.809	0.256
$S''$	6.907	9.842	3.350	3.477
$\xi_w$	0.832	0.225	0.713	0.261
$b$	0.779	0.124	0.706	0.135
$\sigma_a$	0.413	0.777	2.029	4.251
$c_z^p$	0.144	1.414	1.379	3.732
$c_\Upsilon^p$	0.073	0.580	0.137	0.499
$\gamma$	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 (0.037)	1.515 (0.007)
Homogeneous Capital Model	2.769 (0.167)	5.655 (0.046)

Note to Table 1. The table reports 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 (10), which selects  $h = 6$  for CEE. 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. See ACEL for a complete description.

**Table 2. Empirical results (CEE, 2005)**

Parameters	RIRSC ( $\hat{h}_T = 6$ )		Fixed Lags (h=20)	
	Parameter	Standard	Parameter	Standard
	Estimates	Errors	Estimates	Errors
$\rho_M$	-0.020	0.300	-0.114	0.272
$\sigma_M$	0.348	0.108	0.352	0.108
$\epsilon$	0.897	0.275	0.836	0.255
$S''$	3.732	3.695	4.324	4.566
$\xi_w$	0.624	0.194	0.645	0.261
$b$	0.762	0.127	0.717	0.144
$\lambda_f$	1.002	0.231	1.097	0.277
$\sigma_a$	0.001	0.152	0.041	0.557
$\gamma$	0.106	0.243	0.208	0.546

Note to Table 2. The table reports 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 (10), which selects h=3 for ACEL. The notation is the same as that in Tables 2 and 3 in ACEL, and  $\lambda_f$  is calibrated to be 1.01. See ACEL for a complete description.

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

$H$	IRFME		IRFME <sub>RIRSC</sub>	
	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 Table 3. The table reports bias (i.e. true parameter value minus estimated value) and rejection rates of 95% nominal confidence intervals for the AR(1) example.

**Table 4a. Bias, Variance, Coverage Probability ( $T = 100$ )**

h	$\alpha_y$			$\gamma_1$			$\rho_l$		
	bias	var	prob	bias	var	prob	bias	var	prob
1	0.013	0.008	0.938	0.002	0.012	0.938	-0.003	0.019	0.943
2	-0.002	0.005	0.910	0.001	0.003	0.926	-0.014	0.006	0.922
3	-0.000	0.004	0.871	0.000	0.003	0.900	-0.011	0.002	0.918
4	-0.006	0.005	0.757	0.004	0.004	0.803	-0.013	0.003	0.857
5	-0.010	0.005	0.679	0.001	0.004	0.749	-0.012	0.004	0.795
6	-0.008	0.005	0.638	0.002	0.004	0.704	-0.013	0.004	0.748
7	-0.004	0.005	0.614	-0.002	0.004	0.682	-0.014	0.003	0.659
8	-0.004	0.005	0.579	0.000	0.004	0.643	-0.017	0.009	0.609
9	-0.002	0.006	0.543	-0.000	0.004	0.620	-0.017	0.011	0.564
10	-0.004	0.006	0.482	0.001	0.005	0.583	-0.022	0.018	0.520
11	0.001	0.006	0.450	-0.002	0.005	0.545	-0.022	0.020	0.494
12	-0.002	0.006	0.430	-0.004	0.005	0.515	-0.022	0.025	0.458

**Table 4b. Bias, Variance, Coverage Probability ( $T = 200$ )**

h	$\alpha_y$			$\gamma_1$			$\rho_l$		
	bias	var	prob	bias	var	prob	bias	var	prob
1	-0.003	0.004	0.956	0.004	0.005	0.950	0.004	0.009	0.952
2	-0.006	0.002	0.947	-0.002	0.001	0.951	-0.004	0.003	0.926
3	-0.002	0.002	0.925	-0.002	0.001	0.937	-0.004	0.001	0.945
4	-0.005	0.002	0.824	-0.000	0.001	0.864	-0.004	0.001	0.910
5	-0.006	0.002	0.735	0.002	0.001	0.831	-0.005	0.001	0.837
6	-0.006	0.002	0.730	0.002	0.001	0.803	-0.005	0.002	0.791
7	-0.004	0.002	0.712	0.000	0.001	0.794	-0.005	0.001	0.741
8	-0.005	0.002	0.691	-0.001	0.001	0.774	-0.007	0.003	0.686
9	-0.005	0.002	0.663	-0.002	0.001	0.766	-0.009	0.007	0.660
10	-0.005	0.002	0.622	-0.001	0.002	0.730	-0.011	0.014	0.618
11	-0.005	0.002	0.602	-0.001	0.002	0.724	-0.011	0.012	0.605
12	-0.004	0.002	0.578	-0.002	0.002	0.692	-0.012	0.025	0.562

**Table 4c. Bias, Variance, Coverage Probability ( $T = 400$ )**

h	$\alpha_y$			$\gamma_1$			$\rho_l$		
	bias	var	prob	bias	var	prob	bias	var	prob
1	-0.000	0.002	0.949	-0.000	0.002	0.953	0.000	0.004	0.944
2	-0.004	0.001	0.947	0.001	0.001	0.956	-0.003	0.001	0.942
3	-0.003	0.001	0.944	0.000	0.001	0.941	-0.002	0.000	0.948
4	-0.003	0.001	0.842	0.001	0.001	0.886	-0.002	0.000	0.915
5	-0.005	0.001	0.775	0.001	0.001	0.841	-0.002	0.001	0.879
6	-0.004	0.001	0.781	0.001	0.001	0.829	-0.002	0.000	0.837
7	-0.003	0.001	0.775	0.001	0.001	0.824	-0.003	0.000	0.769
8	-0.003	0.001	0.757	0.000	0.001	0.813	-0.004	0.000	0.720
9	-0.003	0.001	0.716	0.000	0.001	0.803	-0.004	0.001	0.698
10	-0.003	0.001	0.677	0.000	0.001	0.793	-0.005	0.008	0.673
11	-0.003	0.001	0.665	0.000	0.001	0.779	-0.005	0.002	0.667
12	-0.003	0.001	0.648	-0.000	0.001	0.760	-0.006	0.019	0.628

Note to Table 4. The table reports the median bias ("bias"), variance ("var") and the coverage probability ("prob") of 95% nominal confidence intervals for Example 2. H=12.

Table 5a. Median Bias, Variance, Coverage Probability ( $\alpha_y$ )

$T$	All			AIC			SIC			HQC		
	bias	var	prob	bias	var	prob	bias	var	prob	bias	var	prob
<b>Valid IRF Selection Only</b>												
100	-0.002	0.006	0.430	-0.000	0.006	0.527	-0.002	0.006	0.456	-0.002	0.006	0.479
200	-0.004	0.002	0.578	-0.003	0.002	0.636	-0.004	0.002	0.604	-0.003	0.002	0.605
400	-0.003	0.001	0.648	-0.002	0.001	0.688	-0.002	0.001	0.668	-0.002	0.001	0.672
<b>Relevant IRF Selection Only</b>												
100	-0.002	0.006	0.430	0.013	0.008	0.938	0.013	0.008	0.938	0.013	0.008	0.938
200	-0.004	0.002	0.578	-0.003	0.004	0.956	-0.003	0.004	0.956	-0.003	0.004	0.956
400	-0.003	0.001	0.648	-0.000	0.002	0.949	-0.000	0.002	0.949	-0.000	0.002	0.949
<b>Valid and Relevant IRF Selection</b>												
100	-0.002	0.006	0.430	0.013	0.008	0.938	0.013	0.008	0.938	0.013	0.008	0.938
200	-0.004	0.002	0.578	-0.003	0.004	0.956	-0.003	0.004	0.956	-0.003	0.004	0.956
400	-0.003	0.001	0.648	-0.000	0.002	0.949	-0.000	0.002	0.949	-0.000	0.002	0.949

Table 5b. Median Bias, Variance, Coverage Probability ( $\gamma_1$ )

$T$	All			AIC			SIC			HQC		
	bias	var	prob	bias	var	prob	bias	var	prob	bias	var	prob
<b>Valid IRF Selection Only</b>												
100	-0.004	0.005	0.515	-0.002	0.004	0.628	-0.003	0.004	0.549	-0.003	0.004	0.574
200	-0.002	0.002	0.692	-0.001	0.002	0.755	-0.001	0.001	0.720	-0.001	0.001	0.728
400	-0.000	0.001	0.760	0.000	0.001	0.806	-0.001	0.001	0.776	-0.000	0.001	0.788
<b>Relevant IRF Selection Only</b>												
100	-0.004	0.005	0.515	0.002	0.012	0.938	0.002	0.012	0.938	0.002	0.012	0.938
200	-0.002	0.002	0.692	0.004	0.005	0.950	0.004	0.005	0.950	0.004	0.005	0.950
400	-0.000	0.001	0.760	-0.000	0.002	0.953	-0.000	0.002	0.953	-0.000	0.002	0.953
<b>Valid and Relevant IRF Selection</b>												
100	-0.004	0.005	0.515	0.002	0.012	0.938	0.002	0.012	0.938	0.002	0.012	0.938
200	-0.002	0.002	0.692	0.004	0.005	0.950	0.004	0.005	0.950	0.004	0.005	0.950
400	-0.000	0.001	0.760	-0.000	0.002	0.953	-0.000	0.002	0.953	-0.000	0.002	0.953

Table 5c. Median Bias, Variance, Coverage Probability ( $\rho_l$ )

$T$	All			AIC			SIC			HQC		
	bias	var	prob	bias	var	prob	bias	var	prob	bias	var	prob
<b>Valid IRF Selection Only</b>												
100	-0.022	0.025	0.458	-0.017	0.004	0.573	-0.018	0.003	0.504	-0.018	0.003	0.519
200	-0.012	0.025	0.562	-0.009	0.001	0.644	-0.010	0.001	0.607	-0.009	0.001	0.616
400	-0.006	0.019	0.628	-0.005	0.001	0.684	-0.005	0.000	0.663	-0.005	0.000	0.669
<b>Relevant IRF Selection Only</b>												
100	-0.022	0.025	0.458	-0.003	0.019	0.943	-0.003	0.019	0.943	-0.003	0.019	0.943
200	-0.012	0.025	0.562	0.004	0.009	0.952	0.004	0.009	0.952	0.004	0.009	0.952
400	-0.006	0.019	0.628	0.000	0.004	0.944	0.000	0.004	0.944	0.000	0.004	0.944
<b>Valid and Relevant IRF Selection</b>												
100	-0.022	0.025	0.458	-0.003	0.019	0.943	-0.003	0.019	0.943	-0.003	0.019	0.943
200	-0.012	0.025	0.562	0.004	0.009	0.952	0.004	0.009	0.952	0.004	0.009	0.952
400	-0.006	0.019	0.628	0.000	0.004	0.944	0.000	0.004	0.944	0.000	0.004	0.944

Note to Table 5. The table reports the median bias, variance and the coverage probability of 95% nominal confidence intervals for Example 2.  $H=12$ . Note that all IRFs in this example are correctly specified so the results using RIRSC and those using both VIRSC and RIRSC are the same.

**Table 6. Selected Number of Impulse Responses**

<b>Valid IRF Selection Only</b>										
<i>T</i>	<i>H</i>	AIC			SIC			HQC		
		Mean	Mode	Var	Mean	Mode	Var	Mean	Mode	Var
100	12	9.804	12.000	11.251	11.306	12.000	3.180	10.819	12.000	5.722
200	12	10.635	12.000	8.933	11.623	12.000	1.711	11.294	12.000	4.034
400	12	11.109	12.000	6.125	11.806	12.000	0.759	11.583	12.000	2.327
<b>Relevant IRF Selection Only</b>										
<i>T</i>	<i>H</i>	AIC			SIC			HQC		
		Mean	Mode	Var	Mean	Mode	Var	Mean	Mode	Var
100	12	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000
200	12	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000
400	12	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000
<b>Both Valid and Relevant IRF Selection</b>										
<i>T</i>	<i>H</i>	AIC			SIC			HQC		
		Mean	Mode	Var	Mean	Mode	Var	Mean	Mode	Var
100	12	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000
200	12	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000
400	12	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000

Note to Table 6. The table reports the mean, median and variance of selected horizons of impulse responses for Example 2.

**Table 7(a). Empirical Selection Frequency for VIRFSC**

$c_1$	$c_2$	$c_3$	$\bar{a}_4$									
			0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0	0	0.77	0.70	0.57	0.40	0.33	0.09	0.02	0.001	0.001	0.001
0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0
1	1	0	0.22	0.29	0.39	0.49	0.48	0.35	0.18	0.08	0.04	0.11
0	1	1	0	0	0	0	0	0	0	0	0	0
1	0	1	0.001	0.002	0.01	0.04	0.09	0.10	0.08	0.05	0.04	0.05
1	1	1	0.001	0.001	0.01	0.05	0.18	0.44	0.69	0.86	0.91	0.82

**Table 7(b). Empirical Selection Frequency for RIRFSC**

$c_1$	$c_2$	$c_3$	$\bar{a}_4$									
			0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0	0	1	1	0.999	0.999	0.998	0.998	0.998	0.998	0.998	0.998
0	1	0	0	0	0	0.001	0.001	0.001	0.001	0.001	0.001	0.001
0	0	1	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0.001	0	0.001	0.001	0.001	0.001	0.001	0.001
0	1	1	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	0	0	0	0

Note to Table 7. The table reports empirical selection frequency for Example 1.

Table 8(a) Median Bias

c1	c2	c3	$\bar{a}_4$									
			0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1.000	0.000	0.000	-0.007	-0.003	-0.003	-0.006	-0.009	-0.007	-0.008	-0.005	-0.004	-0.006
0.000	1.000	0.000	-0.500	-0.437	-0.373	-0.311	-0.248	-0.185	-0.121	-0.054	0.003	0.067
0.000	0.000	1.000	--	--	--	--	--	--	--	--	--	--
1.000	1.000	0.000	-0.021	-0.019	-0.009	-0.016	-0.019	-0.017	-0.020	-0.017	-0.017	-0.018
0.000	1.000	1.000	-0.500	-0.437	-0.381	-0.302	-0.232	-0.214	-0.121	-0.032	0.049	0.081
1.000	0.000	1.000	-0.007	-0.003	-0.004	-0.001	-0.000	-0.009	-0.014	-0.002	0.017	-0.000
1.000	1.000	1.000	-0.021	-0.019	-0.011	-0.011	-0.011	-0.019	-0.025	-0.014	0.004	-0.013
	VIRFSC		-0.007	-0.005	-0.003	-0.009	-0.013	-0.015	-0.023	-0.014	0.004	-0.013
	RIRFSC		-0.007	-0.003	-0.003	-0.006	-0.009	-0.007	-0.009	-0.005	-0.005	-0.006

Table 8(b) Median Variance

c1	c2	c3	$\bar{a}_4$									
			0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1.000	0.000	0.000	0.492	0.491	0.494	0.494	0.491	0.492	0.496	0.492	0.496	0.491
0.000	1.000	0.000	0	0.062	0.125	0.189	0.250	0.313	0.380	0.438	0.507	0.563
0.000	0.000	1.000	--	--	--	--	--	--	--	--	--	--
1.000	1.000	0.000	0.482	0.485	0.477	0.479	0.480	0.479	0.489	0.481	0.484	0.477
0.000	1.000	1.000	0	0.066	0.131	0.187	0.251	0.307	0.392	0.437	0.473	0.543
1.000	0.000	1.000	0.492	0.491	0.496	0.492	0.492	0.481	0.502	0.491	0.484	0.486
1.000	1.000	1.000	0.482	0.484	0.480	0.477	0.482	0.469	0.495	0.480	0.473	0.473
	VIRFSC		0.492	0.491	0.490	0.487	0.483	0.477	0.494	0.481	0.474	0.475
	RIRFSC		0.492	0.491	0.494	0.494	0.491	0.492	0.496	0.492	0.496	0.490

Table 8(c) Mean Coverage Probability

c1	c2	c3	$\bar{a}_4$									
			0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1.000	0.000	0.000	0.950	0.950	0.950	0.950	0.949	0.953	0.951	0.954	0.951	0.950
0.000	1.000	0.000	0.218	0.412	0.701	0.863	0.930	0.971	0.986	0.988	0.989	0.970
0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	1.000	0.000	0.925	0.928	0.933	0.942	0.939	0.951	0.950	0.952	0.950	0.948
0.000	1.000	1.000	0.218	0.412	0.674	0.878	0.944	0.950	0.986	0.990	0.985	0.965
1.000	0.000	1.000	0.950	0.950	0.950	0.950	0.950	0.953	0.951	0.954	0.949	0.951
1.000	1.000	1.000	0.925	0.928	0.933	0.943	0.942	0.950	0.948	0.954	0.952	0.949
	VIRFSC		0.942	0.941	0.942	0.945	0.941	0.950	0.948	0.953	0.951	0.949
	RIRFSC		0.945	0.944	0.946	0.948	0.945	0.950	0.946	0.953	0.956	0.949

Note to Table 8. The table reports the median bias, variance and mean coverage probability for Example 1. In particular, in Table 8(b). "--" means that the estimate is not available (this happens because the third IRF is not informative about the parameter of interest and thus the gradient with respect to that parameter is zero). In some cases, when  $\bar{a}_4 = 0$ , some variances are zero, and these correspond to cases in which we are interested in the IRF that picks restricted estimators, which are imposed to be zero by construction.