FEDERAL RESERVE BANK *of* ATLANTA

Information Criteria for Impulse Response
Function Matching Estimation of DSGE Models

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

WORKING PAPER SERIES

Information Criteria for Impulse Response Function Matching Estimation of DSGE Models

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

Abstract: We propose a new information criterion for impulse response function matching estimators of the structural parameters of macroeconomic models. The main advantage of our procedure is that it allows the researcher to select the impulse responses that are most informative about the deep parameters, therefore reducing the bias and improving the efficiency of the estimates of the model's parameters. We show that our method substantially changes key parameter estimates of representative dynamic stochastic general equilibrium models, thus reconciling their empirical results with the existing literature. Our criterion is general enough to apply to impulse responses estimated by vector autoregressions, local projections, and simulation methods.

JEL classification: C32, E47, C52, C53

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

1.   Introduction

Since the seminal work by Rotemberg and Woodford (1997), there has been an increasing interest in estimating macroeconomic models by using Impulse Response Function (IRF) matching estimators. The method is appealing because of its simplicity, and because it focuses on estimating the parameters on the basis of impulse responses, which directly capture the dynamics that are of primary interest to macroeconomists. Among the recent papers that have used IRF matching estimators we have: Christiano Eichenbaum and Evans (2005), Altig et al. (2004), Jordà and Kozicki (2005), Boivin and Giannoni (2006), Uribe and Yue (2006), DiCecio (2005) and DiCecio and Nelson (2006).

This paper proposes a new method to improve the efficiency of such estimators and to select the IRFs lag length based on statistical criteria. Our method is essentially an information criterion, and it is very easy to implement. It provides an econometrically sound and convenient procedure for implementing IRF matching estimators (IRFME) that improves their statistical inference.

We show that our method can substantially change the parameter estimates of existing representative Dynamic Stochastic General Equilibrium (DSGE) models. We focus on two recent important contributions by Altig et al. (2004) and Christiano et al. (2005). Overall, qualitatively our empirical results confirm the findings in these papers; however, we also find that some key parameter estimates are statistically significantly different from those in Altig et al. (2004). In particular, the average duration between firms' price re-optimization in the homogeneous capital model is shorter than that in Altig et al. (2004), thus going in the direction of reconciling their results with the previous literature. We argue that the difference in the estimates may be caused by small sample biases, and report Monte Carlo simulations which show that the use of our methodology provides substantially more precise estimates of the deep parameters of similar models.

From a theoretical point of view, our proposed estimator answers two questions. The first is: "Is there a method to improve the performance of IRFME?". Since standard macroeconomic models can be written in a state-space form, whose dynamics is constrained to be a function of a few parameters, this imposes some degree of linear dependence on the IRFs. Such restrictions are not imposed in the current IRFME, therefore causing a loss of efficiency in the estimates of the deep parameters in small samples, and possibly a bias. Our goal is to propose a criterion to select IRFs for the matching estimator that will perform well in small samples. Our criterion does so by keeping relevant impulse responses and discarding irrelevant impulse responses.

The second question is: "How many lags in the IRFs should be matched?". So far the literature has proceeded in ad-hoc ways. Christiano et al. (2005) and Altig et al. (2004) use a diagonal weighting matrix, and match a pre-specified number of IRF lags.[1] Jordà and Kozicki (2005) generalize the IRF matching procedure to efficient weighting matrices and select the number of IRF lags based on their significance at conventional levels. However, this procedure fails to recognize that: (i) even impulse responses that are zero may well be informative about the parameters of interest (e.g. restrictions imposed via long-run identification); (ii) confidence bands for each IRFs are obtained independently of each other. Instead, we propose a procedure that exploits the information content in the data to choose which lags to match. Our procedure also provides a framework to choose not only the IRF lag length, but also "which IRFs to match". Thus, we offer applied researchers a selection criterion that is easy to use and addresses an important and, so far, neglected issue.

Our estimator has several advantages over those proposed in the literature: (i) the estimator significantly improves inference; in fact, we show in the Monte Carlo simulations that our estimator has a significantly lower bias, and furthermore in both the Monte Carlo simulations and in the

---

[1]Christiano et al. (2005) and Altig et al. (2004) do not motivate their choice of a specific number of IRF lags.

empirical application we find that the standard errors of the deep parameter estimates significantly decrease; (ii) the estimator is very easy to implement; in fact, it only requires an estimate of the variance of the deep parameters given a particular choice of the number of lags in the IRFs; such an estimate is usually already available because it is computed for the purposes of inference on the deep parameters; (iii) the estimator can be used with or without an efficient weighting matrix; (iv) the estimator can be used in the presence of calibrated parameters (these will just be ignored in the criterion, since they won't affect estimation uncertainty in any way); (v) the estimator can be used when the IRFs are identified by using the most common identifying restrictions, including short-run and long-run restrictions;[2] (vi) our criterion provides both a way to select the appropriate lag length for a given subset of IRFs and, in case the researcher does not have a strong opinion on which IRFs to match, a way to select them (e.g. which, between an IRF to a monetary policy shock and an IRF to a technology shocks, to use).

The IRF matching estimator is a limited information approach to estimation of DSGE models. Limited information estimators do not rely on a full model specification. Thus, the IRF matching estimator can ignore the full set of predictions of which the DSGE model is capable. For example, Christiano, Eichenbaum, and Evans (2005) estimate DSGE model parameters by matching the empirical and theoretical responses of inflation and other macro variables only to an identified monetary policy shock. This contrasts with full information approaches in which the likelihood expresses the complete set of predictions inherent in the DSGE model. Although the solutions of the linearized DSGE models we study have well defined likelihoods, we adopt the limited information motivation of the IRF matching estimator to better understand its properties and behavior.[3]

---

[2] A type of restrictions not allowed in our current framework are sign restrictions, as they do not produce point estimates but only intervals.

[3] Besides IRF matching and maximum likelihood, simulation estimators in the frequentist domain are used to

Our method is related to several contributions that recently appeared in the literature. As in Rotemberg and Woodford (1996), Christiano et al. (2005), Altig et al. (2004), we estimate the model's deep parameters by minimizing the difference between estimated and theoretical IRFs. However, our method allows for (although it is not limited to) an efficient estimation of the parameters by optimally weighting the IRFs. Jordà and Kozicki (2005) propose a IRFME estimator based on local projections and an efficient weighting matrix. We instead propose a method to select the IRFs. In addition, our estimator does not necessarily rely on local projections: our method can be used with commonly used VAR-based methods as well as with local projections.[4]

Before getting into technical details, here below we outline the simple *step-by-step procedure* to implement our estimator in the VAR case (of finite or infinite order):

- I. Choose a particular horizon h between a chosen upper bound (H) and lower bound ($\underline{h}$);

- Estimate the impulse responses up to horizon h, and collect them in the vector $\widehat{\gamma}_h$;

- Obtain analytic or numerical expressions for the theoretical impulse responses up to horizon h as a function of the parameter of interest θ: $\gamma_h(\theta)$;

- Estimate the parameters θ via IRF matching (i.e. by minimizing the objective function (2) below) – denote the estimated parameter $\widehat{\theta}_h$;

- Estimate the variance of the parameters, $\hat{V}_h$;

- Calculate the Redundant Impulse Response Selection Criterion, "$RIRSC$", by using

formula (4) below, that is: $RIRSC(h) = \log(|\hat{V}_h|) + h\frac{\ln(\sqrt{T})}{\sqrt{T}}$ if the VAR has a finite order p, and $RIRSC(h) = \log(|\hat{V}_h|) + h\frac{\ln(\sqrt{T}/p)}{\sqrt{T}/p}$ if the VAR has an infinite order but is estimated with p lags, where T is the sample size.

- II. Repeat the previous steps for h equal to all possible values from $\underline{h}$ to H, and collect the values of $RIRSC(h)$ in a vector;

- III. Finally, choose the value of h associated to the minimum value of the RIRSC, call it $\hat{h}$. The parameter estimates for that particular horizon are our suggested parameter estimates.[5]

The paper is organized as follows. Section 2 presents our new estimator in the leading VAR case, and discusses the assumptions that guarantee its validity. Section 3 provides a clarifying example, Section 4 discusses the projection and the simulation-based estimators, Section 5 presents the main empirical results, and Section 6 reports robustness and Monte Carlo analyses. Section 7 concludes. All the technical proofs and assumptions are collected in the Appendix.

2. THE VAR-BASED IRF MATCHING ESTIMATOR

We provide a framework for researchers interested in estimating the parameters of a structural model based on a series of variables of interest. Let y_t be the $(n_y \times 1)$ vector containing the values of the n_y variables at date t, $t = 1, 2, ..., T$, where T is the total sample size.

In this section, we consider the leading case in which the researcher is interested in estimating the deep parameters of the model by using a VAR-based IRFME, that is an estimator obtained by minimizing the distance between the empirical IRFs (obtained by fitting a VAR to the data and imposing some identification restrictions implied by the model) and the theoretical IRFs implied

---

[5]Note that the first 5 steps have to be programmed anyway by a researcher wishing to estimate and perform inference on the deep parameters of the structural model by using the IRF matching method – the only computational step that our procedure adds is step 6, which is very easy to calculate.

by the model itself. We therefore assume that the structural model admits a structural VAR

representation, so that, eventually, the IRFs estimated from a (reduced-form) VAR are informative

for the parameters contained in the structural model:

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

where ε_t is a white noise with mean 0 and variance Σ_ε. The VAR lag order, p, can be either finite

or infinite, and each of these cases will be treated separately.[6]

Let $\gamma_{i,j,\tau}$ denote IRFs of each variable $y_{i,t+\tau}$ to a structural shock $\varepsilon_{j,t}$ at horizon τ, where

$i = 1, ..., n_y$ and $j = 1, ..., n_\varepsilon$; $\tau = \underline{h}, \underline{h} + 1, ..., H$, where $\underline{h}$ denotes the lower bound on h such

that the deep parameters of the model are not identified for $h < \underline{h}$ and are identified for $h \geq \underline{h}$,[7]

and H is the maximum horizon of the IRFs. We assume that H is a finite number when dealing

with structural models that have a finite and known VAR order, and that instead H can be infinite

when dealing with structural models that have an infinite VAR representation.

Let $\underline{\gamma}_\tau$ be a $(n_\varepsilon n_y \times 1)$ vector that collects the IRFs at a particular horizon τ :

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

The IRFs at horizons $\tau = \underline{h}, \underline{h} + 1, ..., h$ will be collected in a $(n_\varepsilon n_y (h - \underline{h}) \times 1)$ vector γ_h :

$$\gamma_h = \left(\underline{\gamma}'_{\underline{h}}, \ \underline{\gamma}'_{\underline{h}+1}, \ ..., \ \underline{\gamma}'_h \right)'$$

---

[6] We make standard assumptions on the VAR lag operator defined as: $\Phi(L) \equiv I - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p$, where I is the identity matrix, L is the lag operator: we assume that $\Phi(L)$ is invertible, so that the MA representation (and therefore the IRFs) exists.

[7] Our criterion does not require the researcher to know $\underline{h}$.

Let the model's parameters (referred to as deep parameters) be collected in a $(q \times 1)$ vector θ, $\theta \in \Theta$, and the theoretical IRFs up to horizon h be denoted by $\gamma_h(\theta)$.[8]

The IRF Matching Estimator (IRFME) using horizons $\{\underline{h},\ \underline{h}+1,\ ...,\ h\}$ is defined as:

$$\hat{\theta}_h = \arg\min_{\theta \in \Theta} [\widehat{\gamma}_h - \gamma_h(\theta)]' \, \widehat{W}_h \, [\widehat{\gamma}_h - \gamma_h(\theta)] \tag{2}$$

where $\widehat{W}_h$ is an estimated weighting matrix. As explained in the Appendix, $\widehat{W}_h$ could be the covariance matrix of the IRFs, or, as often found in practice, a restricted version of this matrix that has zeros everywhere except along its diagonal. In general, $\widehat{W}_h$ can be readily obtained from standard package procedures that compute IRF standard error bands.

In order to implement the IRFME in practice, the researcher has to choose the horizon h in (2). Our contribution to the existing literature is to provide statistical criteria to choose h.

We derive our results under several technical assumptions, formally stated in the Appendix as Assumptions 1 and 1'. They differ depending on whether the VAR has a finite order $(p < \infty)$ or not $(p = \infty)$. To summarize, we require that: (i) the parameter estimate $\hat{\theta}_h$ be consistent and asymptotically normal for every horizon h; (ii) the asymptotic variance of the parameter estimate, denoted by V_h, be positive definite; (iii) V_h be consistently estimated by $\hat{V}_h$. Examples of such estimators are discussed in detail in the Appendix around equations (19) and (20).

The criterion that we propose allows the researcher to identify the "relevant horizon" of the IRFs (denoted by h_0) and to avoid using IRFs that contain only redundant information. Since IRFs that do not contain additional information only add noise to the estimation of the deep parameters,

---

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

these IRFs should be eliminated. The following definitions formalize these concepts.[9] Note that we do allow h_0 to be infinite (see the Appendix).

Definition 1 [Redundant IRFs]. *Suppose two horizons h_1 and h_2 satisfy $h_2 > h_1$. Then the IRFs at horizons $h_1 + 1, ..., h_2$ are redundant if $V_{h_2} = V_{h_1}$. The IRFs that correspond to horizons $h_1 + 1, ..., h_2$ are non-redundant if $V_{h_1} - V_{h_2}$ is positive semi-definite and distinct from the matrix of zeros.*

Definition 2 [Relevant IRFs]. *We say that h_0 is the horizon associated with the relevant IRFs if the following properties hold: (i) $h_0 \in \{\underline{h}, ...\}$; (ii) $V_{h_1} - V_{h_0}$ is positive semi-definite where $h_1 = h_0 - \Delta h$ and Δh is a positive number; (iii) $V_{h_0} = V_H$ if $h_0 \leq H$.*

Finally,[10] we need a penalty function to ensure that our criterion consistently selects the relevant horizon, h_0. One such penalty function that seems to work well in practice is:[11]

$$\kappa(h, T, p) = \begin{cases} h \frac{\ln\left(\sqrt{T}\right)}{\sqrt{T}} & \text{for } p < \infty \\ h \frac{\ln\left(\sqrt{T}/p\right)}{\sqrt{T}/p} & \text{for } p \to \infty \end{cases} \tag{3}$$

Our main result, namely the IRFME that we propose, is defined in the following theorem:[12]

---

[9]For notational simplicity, in what follows we assume that, if the researcher is matching multiple IRFs, he uses the same horizons for all of them. All the results in the paper do not change in the more general case in which the relevant horizons can differ across IRFs.

[10]Note that non-relevant IRFs can be non-redundant because relevance is defined relative to the use of all IRFs whereas redundancy is not. Even if some IRF is not relevant, it can be non-redundant relative to some other IRF.

[11]The Appendix derives results for more general penalty functions. Here we consider only the case in which $h = \underline{h}, .., H$ (H can be infinite), although in general the penalty function would depend on the number of constraints.

[12]Note that in the theorem we allow h to be unknown, and therefore let the horizon chosen by the researcher be s.t. $h \in \{1, ..., H\}$.

Theorem 3 [Consistent IRFs selection (VAR case)]. *Let the structural model have a VAR(p) representation (1), and the estimator of the deep parameters be defined as (2), where h is chosen s.t.*

$$
\begin{aligned}
\hat{h} &= \arg\min_{h\in\{1,\dots,H\}} RIRSC(h) \\
RIRSC(h) &= \log(|\hat{V}_h|) + \kappa(h, T, p),
\end{aligned}
\tag{4}
$$

and $\kappa(.)$ *is defined in (3) and* $|.|$ *denotes the determinant. Then:*

(a) if $p < \infty$, *under Assumption 1,* $\hat{h} \xrightarrow{p} h_0$;

(b) if $p \to \infty$, *under Assumption 1',* $\hat{h} \xrightarrow{p} h_0$.

3. INTERPRETATION OF THE RIRSC

The following example clarifies identification issues of the deep parameters as well as the definitions of redundant and relevant IRFs.

Example 4 [**Labor productivity and hours**]. *Consider a simplified Real Business Cycle model of labor productivity and employment (cfr. Watson, 2006, eqs. 3,4). By imposing the parameterization* $\rho_l = \phi = 1$ *and the short-run restriction* $a_z = 0$, *the model can be written as*

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

where $\beta_0 = -\beta_1$, $\delta = -\tilde{a}_z/(1-\alpha)$, $\eta_t = (1-\alpha)\sigma_z\varepsilon_t^z$, $v_t = a_l\sigma_l\varepsilon_t^l$, *and* ε_t^l, ε_t^z *have zero mean, unit*

variance and are uncorrelated. The structural parameters of interest are α, σ_l^2, σ_z^2, a_l and $\tilde{a}_z$. Let

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

denote the jth-step structural IRFs. The restrictions on the first three-steps IRFs are

$$\gamma_{0,11} = (1-\alpha)^2 \sigma_z^2 \tag{5}$$

$$\gamma_{0,12} = -\alpha (a_l \sigma_l)^2 \tag{6}$$

$$\gamma_{0,21} = 0 \tag{7}$$

$$\gamma_{0,22} = (a_l \sigma_l)^2 \tag{8}$$

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

$$\gamma_{1,12} = 2\alpha^2 \tilde{a}_z (a_l \sigma_l)^2 / (1-\alpha) \tag{10}$$

$$\gamma_{1,21} = (1-\alpha)\tilde{a}_z \sigma_z^2 \tag{11}$$

$$\gamma_{1,22} = -2\alpha \tilde{a}_z (a_l \sigma_l)^2 / (1-\alpha) \tag{12}$$

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

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

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

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

Since a_l and σ_l cannot be separately identified (only their product is identified), four out of the five deep parameters are identified. For example, α is identified from restrictions (6) and (8), while $a_l \sigma_l$, $\tilde{a}_z$ and σ_z^2 are identified from restrictions (8), (9) and (5), respectively.

There are two trivial examples of redundant impulse responses. One is restriction (7). Another is restrictions on Γ_j, for $j > 2$: since the model is a VAR(2) model, restrictions for $j > 2$ are nonlinear transformation of (5)–(16), and thus are first-order equivalent to some linear combinations of the above restrictions. Therefore, adding these restrictions will not reduce the asymptotic variance.[13] 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.[14]

4. Alternative IRF matching estimators

Jordà and Kozichi (2006) have proposed IRF matching estimators 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). The simplest version of his estimator for the $\tau-$th step impulse response is $\hat{B}_{1,\tau} D$, where $\hat{B}_{1,\tau}$ is directly estimated from

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

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

---

[13] The Appendix shows that our criterion can be given a Canonical Correlations interpretation along the lines of Hall et al. (2006).

[14] For example, suppose $\alpha = 0.5$, $\sigma_z = a_l = \sigma_l = \tilde{a}_z = 1$, where α is to be estimated and the latter parameters are instead known. Let the covariance matrix of the impulse responses be the identity matrix. When α is estimated by using the optimal weighting matrix, using (5), (6), (10)-(16) produces the same asymptotic variance as using (5),(6),(9)-(16). Thus (9) is a redundant IRF (it does not help to identify α, as it is used to identify other parameters that are assumed to be known in the example in this note). While the redundant IRF does not change the asymptotic variance, it can inflate the variance in finite samples. The other IRF are all relevant. Omitting any of these IRFs increases the asymptotic variance and will likely increase the finite sample variance.

Jordà's local projection impulse responses estimator is:

$$\hat{\theta}_{J,h} \;=\; \arg\min_{\theta \in \Theta}(\widehat{\gamma}_h - \gamma_h(\theta))'\hat{W}_h(\widehat{\gamma}_h - \gamma_h(\theta)) \tag{17}$$

where $\widehat{\gamma}_h$ is a vector of structural impulse responses estimated by local projections, $\gamma_h(\theta)$ is the vector of the model's theoretical impulse responses up to horizon h given structural parameter θ, and $\hat{W}_h$ is a weighting matrix.

Theorem 5 [Consistent IRF selection (Local projections case)]. *Let Assumption 1' hold and the estimator of θ be (17), where h is chosen such that:*

$$\hat{h} \;=\; \arg\min_{h \in \{1,...,H\}} RIRSC_J(h),$$
$$RIRSC_J(h) \;=\; log(|\hat{V}_{J,h}|) + \kappa(h,p,T)$$

where $\hat{V}_{J,h}$ is an estimate of the asymptotic covariance matrix of $\hat{\theta}_{J,h}$ and $\kappa(h,p,T) = h\ln\left(\sqrt{T}/p\right)\big/\left(\sqrt{T}/p\right)$. Then $\hat{h} \xrightarrow{p} h_0$.

4.2. The IRF Matching Simulation Estimator. The second estimator that we consider is the simulation-based estimator, which we will refer to as the Sims-Cogley-Nason (SCN) estimator. In this case, we assume that the DSGE model implies an infinite-order VAR process; the reason is that when the VAR is of finite order there is no advantage from using simulation-based estimators because there is no lag-truncation problem. The SCN estimator is implemented in practice as follows. First, VAR(p) models are fitted actual data to obtain empirical impulse responses $\widehat{\gamma}_h$.[15]

---

[15] All the subsequent estimated parameters should also be function of p, the estimated number of lags in the VAR. However, in order to simplify notation, we will not explicitly do so in the notation.

Next, simulate data of length T from the model with parameter θ and apply the VAR(p) procedure to obtain a vector of simulated impulse responses. Let $\widetilde{\gamma}_h^{(s)}(\theta)$ denote the vector of simulated impulse responses from the $s-$th simulated data, $s = 1, .., S$, where S is the total number of simulation replications. Finally, let $\widetilde{\gamma}_h(\theta)$ denote the average across the ensemble of simulated IRFs, which we refer to as the approximate theoretical impulse responses: $\widetilde{\gamma}_h(\theta) = (1/S) \sum_{s=1}^{S} \widetilde{\gamma}_h^{(s)}(\theta)$.

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

$$\hat{\theta}_{SCN,h} = \arg \min_{\theta \in \Theta} (\widehat{\gamma}_h - \widetilde{\gamma}_h(\theta))' \hat{W}_h (\widehat{\gamma}_h - \widetilde{\gamma}_h(\theta)) \tag{18}$$

where $\hat{W}_h$ is a weighting matrix.[16] Let $\hat{V}_{SCN,h}$ denote a consistent estimate of the asymptotic variance of $\hat{\theta}_{SCN,h}$. One such estimator is provided in the Appendix (see eq. (25)).

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

Theorem 6 [Consistent IRF selection (Simulation-based estimators case)]. *The estimator of* $\hat{\theta}_{SCN,h}$ *is (18), where h is chosen s.t.:*

$$\hat{h} = \arg \min_{h \in \{1,...,H\}} RIRSC_{SCN}(h),$$

$$RIRSC_{SCN}(h) = \log(|\hat{V}_{SCN,h}|) + \kappa(p, h, T),$$

---

[16] In other words, the Sims-Cogley-Nason estimator can be viewed an indirect-inference estimator with a sequence of finite-order VAR models used as an auxiliary model (see Gourieux, Monfort and Renault (1993) for indirect inference). The Appendix shows that, under quite mild conditions, $\hat{\theta}_{SCN,h}$ is consistent and asymptotically normal.

where$\hat{V}_{SCN,h}$ is an estimate of the asymptotic covariance matrix of $\hat{\theta}_{SCN,h}$ and $\kappa(p, h, T) = h \frac{\ln(\sqrt{T}/p)}{T/p}$.

Under Assumption 1", $\hat{h} \xrightarrow{p} h_0$.

5. EMPIRICAL ANALYSIS OF TWO REPRESENTATIVE DSGE MODELS

We apply the methods developed in this paper to two important representative macroeconomic models recently developed by Altig et al. (2004) and Christiano et al. (2005) and estimated by IRFME. Altig et al. (2004) account for the dynamic response of ten key postwar macroeconomic variables to monetary policy, neutral technology and capital embodied shocks. Christiano et al. (2005) focus instead exclusively on the monetary policy shock. One of the key parameters of these models is the average time between price re-optimization. Both papers estimate the model's parameter values as well as the implied average time between re-optimization that minimize the distance between the dynamic responses to shocks in the model and the impulse responses estimated by using a VAR. The number of lags in the IRFs is fixed and equal to 20 (excluding those that are zero by assumption), and the weighting matrix is diagonal.[17]

The goal of this section is to efficiently select the number of lags and compare our results to theirs. We follow Altig et al. (2004) in estimating the empirical IRFs by using the following structural identification assumptions: (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; (iii) monetary policy shocks do not contemporaneously affect aggregate quantities and prices. The number of deep parameters to be estimated is 18. We focus on the case of a diagonal weighting matrix, as in Altig et al. (2004) and Christiano et al. (2005).

Table 1(a) reports the results for the Altig et al. (2004) model. The table reports the estimated

---

[17] Altig et al. (2005) remark that the diagonal weighting matrix ensures that the estimated value of the parameter is such that the model's IRFs lie as much as possible within the confidence bands based on the estimated IRFs.

parameters along with their standard errors[18] based on (4) in the columns labeled "RIRSC" (see Altig et al. (2004) for a detailed description of the parameters). We match all three shocks' IRFs, and progressively reduce the lags in all three IRFs by one. We apply the RIRSC criterion (4) for a number of lags in each IRF between 2 and 20, for a total of number of IRF points (h) ranging between 6 and 60. The criterion selects 3 as the relevant number of lags for each IRF. Table 1(a) also reports the estimated parameter values and standard errors for the case of a fixed number of IRFs lags (namely 20, which is the case considered in the Altig et al., 2004) in the columns labeled "Fixed lags (20)". The table shows that the point estimates obtained by using our method are overall qualitatively comparable to those in Altig et al. (2004), thus confirming their empirical findings, although some parameter estimates are different. In particular, note that the estimate of σ_a falls from 2 to 0.487, therefore going in the direction of reconciling the estimate with that in Christiano et al. (2005).

A crucial aspect of the model is the implied average time between firms' price re-optimization, which is a function of the key parameter γ. Interestingly,note that our estimate of γ is substantially (and statistically significantly) higher than that with $h = 20$. As a consequence, our estimate of the implied average time between price re-optimization, reported in Table 1(b), implies that such average time is about 3 quarters even in the homogeneous capital model (rather than the 5 quarters that one would obtain in the fixed IRF lag scenario), which is more in line with the empirical evidence discussed in Blinder et al. (1998). From the standard errors reported in parentheses below such estimates in Table 1(b), note that the difference is statistically significant at conventional levels. Furthermore, our estimator is more efficient, and this shows up in significantly lower standard errors

---

[18]Our standard errors are smaller than their counterparts in Altig et al. (2004). This is because the latter report the square root of the components on the main diagonal of the asymptotic covariance matrix, whereas we report those values divided by the square root of the sample size, namely the standard errors.

of the parameters, including the implied average time between price re-optimization.

Table 2 shows the results for the Christiano et al. (2005) model, where the monetary policy shock is the only shock of interest. In this case, the RIRSC chooses 6 lags for the impulse response. Although the RIRSC changes some parameter estimates, the differences in this case are smaller than in Table 1. This is likely due to the fact that Christiano et al. (2005) match only the IRFs to one shock, which drastically reduces the dimensionality problem.

Note that our point estimates and standard errors are slightly different from those in the original papers by Christiano et al. (2005) and Altig et al. (2004). The reason is that we modified the numerical estimation procedure to make it more robust to changes in the initial parameter values and to obtain more precise results. Specifically: (i) we use a Newton-Raphson type algorithm rather than a simplex algorithm; (ii) our number of maximum iterations is 1000 rather than 10; (iii) the grid sizes for numerical derivatives are different. We found that the latter two are responsible for most of the differences in the numerical parameter values.[19]

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 slightly 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.[20] 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.

---

[19]In particular note that the estimate of σ_a in the Christiano et al. (2005) model was around 2 whereas in our case is much smaller.

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

6. Monte Carlo robustness analysis

The striking difference in the estimates of some key parameters in the previous section deserves an additional careful investigation into the causes of why this happens. In this section, we argue that the difference in the estimates is likely caused by small sample biases, and report Monte Carlo simulations to show that the use of our methodology provides substantially more precise estimation of the deep parameters of the structural models. Unfortunately, a careful Monte Carlo analysis of Altig et al. (2004) and Christiano et al. (2005) is currently computationally unfeasible, so we investigate the properties of smaller scale models that nevertheless contain many features of the more complicated models considered in the previous section. We consider a variety of examples, some of which are purely statistical and others which are influential business cycle macroeconomic models: a simple univariate AR(1) process; the structural VAR(2) example discussed in (4); the Christiano and Eichenbaum (1992) and the Burnside et al. (1993) models, which have a VAR(∞) representation. In the Christiano and Eichenbaum (1992) example, the impulse responses are identified by using long-run identifying restrictions, whereas in the Burnside et al. (1993) example, the impulse responses are identified by using short-run restrictions.

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

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

where ε_t are random draws from a normal distribution with mean zero and variance one, and $T = 100$. We estimate the deep parameter ρ by the IRFME that minimizes the distance between the vector of IRFs estimated by fitting an AR(2) to the data and the theoretical IRF derived from the AR(1). The weighting matrix W is the inverse of the covariance matrix of the estimated IRFs

calculated by using Monte Carlo simulation. In this section we let H denote either the number of IRFs matched by the IRFME with a fixed number of IRF lags (when we refer to the usual IRFME) or the maximum number of IRFs considered when criterion (4) is used to select the relevant IRF lag length.

Table 3(a) reports, for various values of ρ and H, both the estimated average bias ("bias") and the empirical rejection rates ("rej. rate") of nominal 5% significance level tests for the following estimators: the IRF matching estimator with H IRF lags, labeled "IRFME"; the IRF matching estimator using only the IRFs selected by (4), labeled "IRFME$_{RIRSC}$"; and the usual AR(1) estimator, labeled "AR(1)". Note that the IRFME with $H = 1$ is the maximum likelihood estimator. We performed 1,000 Monte Carlo replications, discarding replications in which the estimator did not converge numerically.

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

6.2. The structural VAR(2) discussed in example (4). In example (4), the structural model has a finite order structural VAR representation. We set $T = 1,000$ to assess the performance of the estimator in large samples. Table 3(b) reports bias (columns labeled "bias") and empirical rejection frequencies of t-tests with a nominal 5% rejection level (columns labeled "rej. rates") for both the IRF matching estimator with a fixed number of lags equal to H (labeled "IRFME") and the IRF matching estimator using only the IRFs selected by (4) (labeled "IRFME$_{RIRSC}$"). The

table shows that the bias afflicts mostly β_0, and that the RIRSC helps to keep it under control. Note that tests based on the traditional IRFME become unreliable when H increases, with rejection rates well above the nominal level, a problem that does not arise if one uses the RIRSC.

6.3. Burnside, Eichenbaum and Rebelo (1993). In this case, the structural model has an infinite VAR representation for endogenous variables, such as output. Government spending and productivity are exogenous and follow independent univariate autoregressive processes. We estimate the empirical IRFs by using a bivariate VAR with output and government spending. The Burnside et al. (1993) theoretical model implies that the former has an infinite VAR representation, the latter an AR(1) representation.

The simulations are performed as follows. We generate a sample of artificial data of output and government spending from the model (treating the parameter estimates of Burnside et al. (1993) as the true values – such values are reported in the row labeled "true param."). The empirical IRFs are estimated by using a VAR(p), where p is selected by the BIC. The identification strategy assumes that the productivity shock does not contemporaneously affect government spending, whereas the government spending shock does. We perform 100 Monte Carlo replications, and evaluate only the bias of the parameter estimates.[21]

Results are reported in Table 3(c). Parameter notation is the same as in Burnside et al. (1993, cfr. Table 1). The table shows that in this environment the RIRSC performs quite well. It reduces the average absolute bias across parameter estimates by 15% or more (see the last column, labeled "*Average Bias*"). Not all the parameters in the Burnside et al. (1993) model are sensitive to

---

[21] For the Burnside et al. (1993) and the Christiano and Eichenbaum (1992) cases, we do not consider coverage rates because they would require a very large number of Monte Carlo replications. Although the current number of Monte Carlo replications is enough to give a flavor of the biases incurred in these examples, it is not sufficient to provide a precise analysis of coverage properties.

changes in the horizon of the IRF used for estimation: for example, parameters like θ and ρ_A do not seem to be influenced by that. From comparing the size of the biases with the magnitude of the true parameter values, it is clear that the biases can be substantial, and that therefore our RIRSC can be extremely valuable.

6.4.   Christiano and Eichenbaum (1992).   We select two variables, employment and output; the theoretical model implies that they jointly have an infinite VAR representation. We simulate data of employment and output from the theoretical model by using the Christiano and Eichenbaum's (1992) estimated parameter values, reported in the table as "True param." to provide a comparison for the bias estimates. We then estimate a VAR(p) with a fixed lag length ($p = 2$) to control for the effect caused by uncertainty on the lag length on the IRFME performance.[22] Parameter notation is the same as in Christiano and Eichenbaum (1992, Table 1, the indivisible labor case with $\alpha = 0$).

According to the model, the government spending shock has no long run effect on output nor employment, whereas the productivity shock does. We therefore estimate the empirical IRFs by using such long run identification procedure. The deep parameter of the structural model by matching the IRF estimated from the data with the theoretical IRF generated by the model. Since the estimation procedure is computationally burdensome, the number of Monte Carlo simulations is limited to 100.

Table 3(d) shows the results. The rows labeled "IRFME" report the bias of the IRFME implemented with H lags, whereas the rows labeled "IRFME$_{RIRSC}$" report the bias of the IRFME implemented with a number of lags chosen by the $RIRSC$. The table shows that increasing the

---

[22] One lag was insufficient to the VAR to replicate the shape of the theoretical IRFs, and sometimes a BIC criterion selected that and the estimator did not converge; in general, two lags did a pretty good job. Cogley and Nason (1995) also impose a lag length equal to two.

number of IRF lags used in the IRFME usually increases the mean absolute bias of the parameter estimates. On average across the various parameters that we consider, the mean absolute bias increases from 0.0938 when $H = 3$ to 0.2119 when $H = 10$. The rows labeled "IRFME$_{RIRSC}$" show instead that our criterion is very successful in attenuating this bias.

7. CONCLUSIONS

This paper's objective is to contribute to the literature on the estimation of DSGE models by using IRF matching estimators. We propose a simple and econometrically sound method for doing so. We show by Monte Carlo simulations that our method can substantially improve the precision of the parameter estimates and decrease their small sample biases. We also show that our method can substantially change key parameter estimates of existing representative DSGE models. We hope that the simplicity and the usefulness of the criterion that we propose will increase the applicability of impulse response function matching estimators in practice.

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

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

8. REFERENCES

Altig, David, Lawrence Christiano, Martin Eichenbaum and Jesper Linde (2004): "Firm-Specific Capital, Nominal Rigidities and the Business Cycle", *mimeo*, Northwestern University.

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

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

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

Burnside, Craig, Martin Eichenbaum and Sergio Rebelo (1993): "Labor Hoarding and the Business Cycle", *Journal of Political Economy* 101(2), 245-273.

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.*

Cogley, Timothy and James 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.

Del Negro, Marco and Frank Schorfheide (2004): "Priors from General Equilibrium Models for VARs", *International Economic Review* 45, 643–673.

Del Negro, Marco, Frank Schorfheide, Frank Smets and Rafael Wouters (2007): "On the Fit and Forecasting Performance of New Keynesian Models, *Journal of Business and Economic Statistics.*

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

DiCecio, Riccardo and Edward Nelson (2006): "An Estimated DSGE Model for the United Kingdom", *mimeo*, Federal Reserve Bank of St. Louis.

Fernandez-Villaverde, Jesus and Juan Rubio-Ramirez (2005): "Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood", *Journal of Applied Econometrics* 20, 891-910.

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

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

Hall, Alastair, Atsushi Inoue, Kalidas Jana and Changmock Shin (2005): "Information in Generalized Method of Moments Estimation and Entropy Based Moment Selection", *mimeo*.

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

Jordà, Oscar and Sharon Kozichi (2005): "An Efficient IRF Matching Estimator for Rational Expectations Models", *mimeo*, University of California at Davis.

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

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

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

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

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

Rabanal, Pau and Juan Rubio-Ramirez (2005): "Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach", *Journal of Monetary Economics* 52, 1151-1166.

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.

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

9. Appendix

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

Assumption 1 (VAR(p)). *(a) For every $h \in \{\underline{h}, \underline{h}+1, ..., H\}$, $\sqrt{T}(\hat{\theta}_h - \theta_0) \overset{d}{\rightarrow} N(0, V_h)$, where V_h is p.d. (b) There is a consistent estimator of V_h, $\hat{V}_h = V_h + O_p(T^{-1/2})$ for every $h \in \{\underline{h}, ..., H\}$. (c) $\hat{W}_h$ is a sequence of p.s.d. matrices and satisfies $\hat{W}_h = W_h + O_p(T^{-1/2})$ where W_h is p.d. for every $h \in \{\underline{h}, ..., H\}$. (d) There is a unique $h_0 \in \{\underline{h}, ..., H\}$ such that: (i) $V_h - V_{h_0}$ is p.s.d. for all $h \in \{\underline{h}, ..., H\}$; (ii) If there is another $h_1 \neq h_0$ such that $V_h - V_{h_1}$ is p.s.d. for all $h \in \{\underline{h}, ..., H\}$, then $h_0 < h_1$. (e) For any $h_1, h_2 \in \{\underline{h}, ..., H\}$ such that $h_1 < h_2$, $\sqrt{T}[\kappa(h_2, T) - \kappa(h_1, T)] \rightarrow \infty$ as $T \rightarrow \infty$ and $\kappa(h, T) = o(1)$ for every $h \in \{\underline{h}, ..., H\}$.*

Assumption 1' (VAR(∞)). *Assumption 1 holds with 1(b,d,e,f) replaced by: (b') $\hat{V}_h = V_h + O_p(p^2 T^{-1/2})$ uniformly in h [for the local projection estimator, $\hat{V}_{J,h} = V_{J,h} + O_p(p^2 T^{-1/2})$]; (d') There is a unique $h_0 \in \{\underline{h}, ...\}$ such that: (i) $V_h - V_{h_0}$ is p.s.d. for all $h \in \{\underline{h}, ...\}$; (ii) If there is another $h_1 \neq h_0$ such that $V_h - V_{h_1}$ is p.s.d. for all $h \in \{\underline{h}, ...\}$, then $h_0 < h_1$. (e') for any $h_1, h_2 \geq \underline{h}$ such that $h_1 < h_2$, $(\sqrt{T}/p^2)[\kappa(h_2, p, T) - \kappa(h_1, p, T)] \overset{p}{\rightarrow} \infty$ as $T \rightarrow \infty$ and $\kappa(p, p, T) = o(1)$ for every $h \geq \underline{h}$; (f') $H = cp$, where $c > 0$.*

Assumption 1" (SCN). *(a") For every $h \in \{\underline{h}, ...\}$, $\sqrt{T}(\hat{\theta}_{SCN,h} - \theta_0) \overset{d}{\rightarrow} N(0, V_{SCN,h})$, where $V_{SCN,h}$ is p.d. (b") $\hat{V}_{SCN,h} = V_{SCN,h} + O_p(p^2 T^{-1/2})$ uniformly in h. (c") There is a unique $h_0 \in \{\underline{h}, ...\}$ such that: (i) $V_{SCN,h} - V_{SCN,h_0}$ is p.s.d. for all $h \in \{\underline{h}, ...\}$; (ii) If there is another $h_1 \neq h_0$ such that $V_h - V_{h_1}$ is p.s.d. for all $h \in \{\underline{h}, ...\}$, then $h_0 < h_1$. (d") For any $h_1, h_2 \in \{\underline{h}, ...\}$*

such that $h_1 < h_2$, $\frac{\sqrt{T}}{p^2}[\kappa(h_2, p, T) - \kappa(h_1, p, T)] \to \infty$ as $T \to \infty$ and $\kappa(p, p, T) = o(1)$ for every

$h \in \{\underline{h}, ...\}$; (e") $H = cp$, where $c > 0$.

Remarks on Assumptions 1 and 1'. The IRFME (2) is a Classical Minimum Distance Estimator (CMD). I. 1(a) is a high-level assumption and follows from more primitive assumptions for CMD estimators (e.g., Newey and McFadden, 1994, Theorem 3.2). For example, under standard assumptions, the asymptotic covariance matrix of the estimator $\hat{\theta}_h$ takes the form of

$$V_h = [\Gamma_h(\theta_0)'W_h\Gamma_h(\theta_0)]^{-1}\Gamma_h(\theta_0)'W_h\Sigma_{\gamma_h}W_h\Gamma_h(\theta_0)[\Gamma_h(\theta_0)'W_h\Gamma_h(\theta_0)]^{-1},$$

where $\Gamma_h(\theta) \equiv \partial\gamma_h(\theta)/\partial\theta$, Σ_{γ_h} is the covariance matrix of the IRFs, and W_h is the weighting matrix.

II. Examples of consistent estimators of V_h are as follows. For a general weighting matrix (for example, $\hat{W}_h = I$) the following estimator satisfies Assumptions 1(b,c) under standard assumptions:

$$\hat{V}_h = [\Gamma_h(\hat{\theta}_h)'\hat{W}_h\Gamma_h(\hat{\theta}_h)]^{-1}\Gamma_h(\hat{\theta}_h)'\hat{W}_h\hat{\Sigma}_{\gamma_h}\hat{W}_h\Gamma_h(\hat{\theta}_h)[\Gamma_h(\hat{\theta}_h)'\hat{W}_h\Gamma_h(\hat{\theta}_h)]^{-1}, \tag{19}$$

where $\hat{\Sigma}_{\gamma_h}$ is a consistent estimator of the covariance matrix of impulse responses, Σ_{γ_h}, $\hat{W}_h$ is a consistent estimator of W_h by Assumption 1(c). In Assumption 1(b), H can be bigger than p, and therefore the covariance of the IRFs may be singular, in which case we assume that a generalized inverse is used to invert such a matrix. When $\hat{W}_h$ is the optimal weighting matrix, $\hat{W}_h = \hat{\Sigma}_{\gamma_h}^{-1}$, then (19) simplifies to:

$$\hat{V}_h = [\Gamma_h(\hat{\theta}_h)'\hat{\Sigma}_{\gamma_h}^{-1}\Gamma_h(\hat{\theta}_h)]^{-1}. \tag{20}$$

III. Assumption 1(d) is the identification condition for the number of relevant impulse responses,

h_0. IV. Assumption 1(e) is a condition on the penalty term for unnecessary redundant impulse responses. For example, Hall et al. (2006) recommend a BIC-type penalty, $h \ln\left(\sqrt{T}\right)/\sqrt{T}$, which is the penalty function that has been used throughout the main paper. V. Note that assumption 1c(i) does not require $n_y \leq n_\varepsilon$, nor that $n_y = n_\varepsilon$. Nor it requires non-singularity of W_h.[23] VI. Note that $\kappa(p, h, T) = h\frac{\ln(\sqrt{T}/p)}{T/p}$ satisfies Assumption 1'(e') and 1"(d").

Assumption 2.[24] *In model (1): (a) As $p, T \to \infty$, $p^4/T \to 0$. (b) The parameter space Θ is compact. (c) Let $\gamma_h(\theta)$ denote a vector of population impulse responses implied by a VAR(p) model fitted to the data simulated with θ. There is a sequence of covariance matrix $\{\Sigma_{\gamma_h(\theta)}\}$ such that, for any sequence of vectors $\{\ell_p\}$ satisfying $0 < L_1 \leq \|\ell_p\| \leq L_2 < \infty$ for all p, $\sqrt{T}\ell'_p(\widehat{\gamma}_h - \gamma_h(\theta_0))$ $\overset{d}{\to} N(0, \lim_{p\to\infty} \ell'_p \Sigma_{\gamma_h(\theta_0)}\ell_p)$ and $\sqrt{T}\ell'_p(\widetilde{\gamma}_h^{(s)}(\theta) - \gamma_h(\theta)) \overset{d}{\to} N(0, \lim_{p\to\infty} \ell'_p \Sigma_{\gamma_h(\theta)}\ell_p)$ jointly and independently for every $\theta \in \Theta$ and $s = 1, 2, ..., S$. (d) $\lim_{p\to\infty} \|\gamma_h(\theta) - \gamma_h(\theta_0)\| = 0$ if and only if $\theta = \theta_0$. (e) $\{\gamma_h\}$ is continuously differentiable and the rank of $\Gamma_h(\theta_0)$ is $dim(\theta)$ for $p = \underline{h}, \underline{h} + 1, ...$ for some $\underline{h}$. (f) $\gamma_h(\theta)$ and $\widetilde{\gamma}_h(\theta)$ satisfy Lipschitz conditions, $\|\gamma_h(\theta') - \gamma_h(\theta)\| < L\|\theta' - \theta\|$ and $\|\widetilde{\gamma}_h(\theta') - \widetilde{\gamma}_h(\theta)\| < \tilde{L}\|\theta' - \theta\|$ where L and $\hat{L}$ do not depend on θ and θ' and are $O(1)$ and $O_p(1)$, respectively. (g) Each row of $\{\Gamma_h(\theta_0)\}$ is summable. (h) There is a sequence of matrices $\{W_h\}$ such that, for any absolutely summable sequence of vectors $\{\ell_p\}$, $\ell'_p(\hat{W}_h - W_h)\ell_p = O_p\left(p^2/\sqrt{T}\right)$. (k) The eigenvalues of $\{W_h\}$ are all positive and bounded away from zero and above. (l) For*

---

[23] In fact, for example, suppose that you estimate a tri-variate VAR(2) with two shocks and estimate eight deep parameters using the optimal weighting matrix. When $H = 2$, the 18 by 18 asymptotic covariance matrix of all the possible IRFs is singular and has rank 12, say. If you use the Moore-Penrose generalized inverse of the asymptotic covariance matrix as the weighting matrix and if the 18 by 8 Jacobian matrix $F_h(\theta_h)$ has rank 8, which is implicit in assumption (a), the result will still carry through even though $n_y > n_\varepsilon$. Suppose now that the tri-variate VAR(2) is driven by one shock instead of two. Then the asymptotic covariance matrix has rank 6. As a result the matrix $[F_h(\theta_h)'\hat{W}_h F_h(\theta_h)]$ is singular and the deep parameters will not be identified. In conclusion, the dimension of shocks matters but not necessarily relative to the dimension of y_t. Provided the rank conditions are satisfied, one could keep adding a redundant vector of variables to the VAR system in which the number of shocks is fixed and not violate the identification condition. Of course the finite-sample performance of the CMD estimator will deteriorate.

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

any absolutely summable sequence of vectors, $\{v_p\}$, $\lim_{p\to\infty} v'_p W_h v_p$ is well-defined. (m) There

are consistent estimators of $\Sigma_{\gamma_h(\theta_0)}$ and $\Sigma_{\gamma_h(\theta)}$, $\hat{\Sigma}_{\gamma_h}$ and $\tilde{\Sigma}^{(s)}_{\gamma_h(\theta)}$ respectively, such that, for any

absolutely summable sequence of vectors $\{\ell_p\}$, $\ell'_p \hat{\Sigma}_{\gamma_h} \ell_p - \ell'_p \Sigma_{\gamma_h(\theta_0)} \ell_p = O_p(p^2/\sqrt{T})$ and $\ell'_p \tilde{\Sigma}^{(s)}_{\gamma_h(\theta)} \ell_p -$

$\ell'_p \Sigma_{\gamma_h(\theta)} \ell_p = O_p(p^2/\sqrt{T})$.

Proof. [Proof of Theorem 3] *Recall that $\underline{h}$ denotes the lowest horizon of the IRF for which the*

parameters are identified. (a) Suppose that $h > \underline{h}$ and $h \neq h_0$. First consider the case in which

$V_h = V_{h_0}$. Because Assumption 1(d) implies that $h > h_0$, it follows from Assumptions 1(b,e) that

$$T^{1/2}(RIRSC(h) - RIRSC(h_0)) = T^{1/2}(\log(|\hat{V}_h|) - \log(|\hat{V}_{h_0}|)) + T^{1/2}(\kappa(h,T) - \kappa(h_0,T))$$

$$\to \infty.$$

Thus $T^{1/2}(RIRSC(h) - RIRSC(h_0))$ is positive with probability approaching one as $T \to \infty$. Next

consider the case in which $V_h \neq V_{h_0}$. By Theorem 22 of Magnus and Neudecker (1999, p.21), it

follows from Assumption 1(d) that $\log(|V_h|) - \log(|V_{h_0}|) > 0$, where V_h is defined in Assumption

1(a). Thus it follows from Assumptions 1(b,e) that

$$RIRSC(h) - RIRSC(h_0) = \log(|\hat{V}_h|) - \log(|\hat{V}_{h_0}|) + \kappa(h,T) - \kappa(h_0,T)$$

$$= \log(|V_h|) - \log(|V_{h_0}|) + o_p(1)$$

$$> 0$$

with probability approaching one. When $h < \underline{h}$, it follows from the definition of $\underline{h}$ that the rank of

$\Gamma_h(\widehat{\theta}_h)$ will be less than the dimension of θ with probability approaching one. Thus $|\Gamma_h(\widehat{\theta}_h)' W_h \Gamma_h(\widehat{\theta}_h)| \xrightarrow{p}$

0 and $RIRSC(h) \xrightarrow{p} \infty$ for all $h < \underline{h}$. Since $RIRSC(h) = O_p(1)$ for $h \geq \underline{h}$, $\hat{h} \geq \underline{h}$ with probabil-

ity one. Because RIRSC(h) is uniquely minimized at h_0 with probability approaching one and $RIRSC(\hat{h}) \leq RIRSC(h)$ for all h, $\hat{h} \xrightarrow{p} h_0$.

(b) First consider the case $h_0 < \infty$. When h_0 and $h < \underline{h}$ are compared, or when h_0 and h that is greater than or equal to $\underline{h}$ and does not depend on p are compared, h_0 will be selected with probability approaching one by using arguments analogous to the proof of Part (a). When h_0 and h that is increasing in p are compared, the value of the penalty term for such h will be larger than that for h_0. Because the first term of the criterion will converge to the same limit, h_0 will be chosen with probability approaching one.

Second, consider the case $h_0 = \infty$ and $p \rightarrow \infty$ under Assumption 1'. Let h_T be a sequence such that $h_T \rightarrow \infty$ as $T \rightarrow \infty$, $h_T = O(p)$, and $\kappa(p, p, T) = o(1)$. Then for every finite h,

$$
\begin{aligned}
RIRSC(h) - RIRSC(h_T) &= log(|\hat{V}_h|) - log(|\hat{V}_{h_T}|) + \kappa(h, p, T) - \kappa(h_T, p, T) \\
&= log(|V_h|) - log(|V_{h_T}|) + o_p(1) \\
&> 0
\end{aligned}
\tag{21}
$$

with probability approaching one.[25] Since $RIRSC(\hat{h}) \leq RIRSC(h_T)$, it follows that for any finite h,

$$
RIRSC(h) - RIRSC(\hat{h}) > 0
\tag{22}
$$

with probability approaching one. Thus $\hat{h} \xrightarrow{p} \infty$. Therefore, summarizing the results in Part (b), our criterion will choose the right horizon whether h_0 is finite or not (at least asymptotically). ∎

Proof.   [Canonical Correlation interpretation of RIRSC] *The fact that our criterion can be given*

---

[25] *This follows because $h_0 = \infty$. No matter how large h is, there is always an $h' > h$ such that the IRF at horizon h' is relevant. As a result, the difference is positive with probability one.*

a Canonical Correlations interpretation along the lines of Hall et al. (2006) can be shown as follows.

Let $\widehat{\Psi}_h$ denote a vector of slope coefficients and covariance matrix estimates and Ψ_0 be the vector of their true parameter values. In general, estimators of the slope coefficients and the covariance matrix are $\sqrt{T}$-consistent and can be written as $\sqrt{T}(\widehat{\Psi}_h - \Psi_0) = GT^{-\frac{1}{2}}\sum_{t=1}^{T} s_t(\theta_0) + o_p(1)$, for some non-singular matrix G and random vectors $\{s_t(\theta)\}$ such that it is serially uncorrelated and $E[s_t(\theta_0)] = 0$. Because impulse responses are smooth functions of slope coefficients and the covariance matrix of a VAR model, it follows that

$$
\begin{aligned}
\sqrt{T}(\widehat{\gamma}_h - \gamma_h(\theta_0)) &= \sqrt{T}\left(f(\widehat{\Psi}_h) - f(\Psi_0)\right) \\
&= \sqrt{T}F_h(\Psi_0)(\widehat{\Psi}_h - \Psi_0) + o_p(1) \\
&= F_h(\Psi_0)GT^{-\frac{1}{2}}\sum_{t=1}^{T} s_t(\theta_0) + o_p(1),
\end{aligned}
$$

where $F_h(\Psi)$ is the Jacobian matrix of the smooth functions. Thus, an impulse function matching estimator is first-order equivalent to the GMM estimator based on $F_h(\Psi_0)GE(s_t(\theta_0)) = 0$, provided that asymptotically equivalent weighting matrices are used. Thus, the results in Hall et al. (2006) apply: the RIRSC criterion selects the set of moment conditions (or the impulse responses that correspond to these moment conditions) whose canonical correlations with the true score is highest among the set of candidate moment conditions (impulse responses). ■

Proof. *[Proof of Theorem (5).] The proof is as in Theorem 1(b).* ■

Theorem 7 [Asymptotic normality of simulation-based estimators]. *Under Assumption 2(a)–(l), $\hat{\theta}_{SCN,h}$ is consistent and is asymptotically normally distributed:*

$$
\sqrt{T}(\hat{\theta}_{SCN,h} - \theta_0) \xrightarrow{d} N(0, V_{SCN,h}) \tag{23}
$$

where

$$V_{SCN,h} = \left(1 + \frac{1}{S}\right) \lim_{p\to\infty} (\Gamma_{h,p}(\theta_0)'W_h\Gamma_{h,p}(\theta_0))^{-1}\Gamma_{h,p}(\theta_0)'W_h\Sigma_{\gamma_h(\theta_0)}W_h\Gamma_{h,p}(\theta_0)$$

$$\times(\Gamma_{h,p}(\theta_0)'W_h\Gamma_{h,p}(\theta_0))^{-1}.$$

Proof. [Proof of Theorem (7)] *By Theorem 2.1 of Newey and McFadden (1994), for a given horizon h, $\hat{\theta}_{SCN,h}$ is consistent if: (i) $Q_h^0(\theta) \equiv \lim_{p\to\infty} Q_h(\theta)$, where $Q_h(\theta) \equiv (\gamma_h(\theta_0) - \gamma_h(\theta))'W_h(\gamma_h(\theta_0) - \gamma_h(\theta))$ is uniquely minimized at θ_0, where $\gamma_h(\theta)$ is defined in Assumption 2(c); (ii) Θ is compact; (iii) $Q_h^0(\theta)$ is continuous; and (iv) $\hat{Q}_h(\theta)$ (the objective function evaluated at the estimated parameters and at the estimated weighting function) converges uniformly in probability to $Q_h^0(\theta)$. By Assumptions 2(d,k), $Q_h^0(\theta)$ is uniquely minimized at θ_0. By Assumption SCN1(d), Θ is compact. By Assumptions 2(b,f,k),*

$$
\begin{aligned}
|Q_h(\theta') - Q_h(\theta)| &= |(\gamma_h(\theta') - \gamma_h(\theta))'W_h\gamma_h(\theta') + \gamma_h(\theta)'W_h(\gamma_h(\theta') - \gamma_h(\theta))| \\
&\leq \left[(\gamma_h(\theta') - \gamma_h(\theta))'W_h(\gamma_h(\theta') - \gamma_h(\theta))\right]^{\frac{1}{2}} (Q_h(\theta)^{\frac{1}{2}} + Q_h(\theta')^{\frac{1}{2}}) \\
&\leq C\|\gamma_h(\theta') - \gamma_h(\theta)\|,
\end{aligned}
\tag{24}
$$

where C is a constant that does not depend on p. Since

$$|Q_h^0(\theta') - Q_h^0(\theta)| \leq C \lim_{p\to\infty} \|\gamma_h(\theta') - \gamma_h(\theta)\|,$$

it follows from Assumption 2(f) that $Q_h^0(\theta)$ is continuous in θ. To show that $\hat{Q}_h(\theta)$ uniformly converges in probability to $Q_h^0(\theta)$, we need to show pointwise convergence and stochastic equicontinuity

of $\hat{Q}_h(\theta)$. The pointwise convergence of $\hat{Q}_h(\theta)$ to $Q_h^0(\theta)$ follows from Assumptions 2(c,g). The stochastic equicontinuity of $\hat{Q}_h(\theta)$ follows from the Lipschitz condition in Assumption 2(f). By the uniform law of large number (e.g., Theorem 21.9 of Davidson, 1994, p.337), $\hat{Q}_h(\cdot)$ converges uniformly in probability to $Q_h^0(\cdot)$. Therefore, $\hat{\theta}_{SCN,h}$ converges in probability to θ_0.

Since $\hat{\theta}_{SCN,h} \overset{p}{\to} \theta_0$, it follows from the first-order condition and the mean-value theorem that

$$
\begin{aligned}
\sqrt{T}(\hat{\theta}_{SCN,h} - \theta_0) \quad = \quad & (\Gamma(\hat{\theta}_{SCN,h})'\hat{W}_h\Gamma(\bar{\theta}_{SCN,h}))^{-1} \\
& \times \Gamma(\hat{\theta}_{SCN,h})'\hat{W}_h[\sqrt{T}(\hat{\gamma}_h - \gamma_h(\theta_0)) - \sqrt{T}(\tilde{\gamma}_h(\theta_0) - \gamma_h(\theta_0)],
\end{aligned}
$$

where $\bar{\theta}_{SCN,h}$ is a point between $\hat{\theta}_{SCN,h}$ and θ_0. Then

$$
\sqrt{T}(\hat{\theta}_{SCN,h} - \theta_0) \quad \overset{d}{\to} \quad N(0, V_{SCN,h}),
$$

where

$$
V_h \quad = \quad \lim_{p\to\infty} (\Gamma_h(\theta_0)'W_h\Gamma_h(\theta_0))^{-1}\Gamma_h(\theta_0)'W_h\,(1+1/S)\,\Sigma_{\gamma_h(\theta_0)}W_h\Gamma_h(\theta_0)(\Gamma_h(\theta_0)'W_h\Gamma_h(\theta_0))^{-1}.
$$

∎

Theorem 8 [Estimation of asymptotic variance of simulation-based estimators]. *Let* $\Sigma_{\gamma_h(\theta)}^{(s)}$ *denote the estimated asymptotic covariance matrix of the simulated impulse responses* $\tilde{\gamma}_h^s(\theta)$, *and* $\hat{\Sigma}_{\gamma_h}$ *denote the estimate of the asymptotic covariance matrix of the empirical impulse responses. Let*

$$
\hat{W}_h \quad = \quad \left(\hat{\Sigma}_{\gamma_h} + \frac{1}{S^2}\sum_{s=1}^{S} \tilde{\Sigma}_{\gamma_h(\tilde{\theta}_{SCN,h})}^{(s)} \right)^{-1},
$$

where $\tilde{\theta}_{SCN,h}$ is an estimator of θ (e.g., the SCN estimator with $W_h = I$). Then

$$
\begin{aligned}
\hat{V}_{SCN,h} \;=\; & (\Gamma_h(\tilde{\theta}_{SCN,h})'\hat{W}_h\Gamma_h(\tilde{\theta}_{SCN,h}))^{-1} \\
& \times \Gamma_h(\tilde{\theta}_{SCN,h})'\hat{W}_h \left(\hat{\Sigma}_{\gamma_h} + \frac{1}{S^2}\sum_{s=1}^{S} \tilde{\Sigma}^{(s)}_{\gamma_h(\tilde{\theta}_{SCN,h})} \right) \hat{W}_h\Gamma_h(\tilde{\theta}_{SCN,h}) \\
& \times (\Gamma_h(\tilde{\theta}_{SCN,h})'\hat{W}_h\Gamma_h(\tilde{\theta}_{SCN,h}))^{-1}
\end{aligned}
\tag{25}
$$

where $\widehat{\Gamma}_h(\theta) \equiv \frac{1}{S}\sum_{s=1}^{S} \Gamma_h^{(s)}(\theta)$, for $\Gamma_h^{(s)}(\theta) \equiv \partial\tilde{\gamma}_h^{(s)}(\theta)/\partial\theta$. Under Assumption 2,

a. If $\tilde{\theta}_{SCN,h}$ is $\sqrt{T}$-consistent, $\hat{V}_h = V_h + O_p(p^2/T)$.

b. If $h < \underline{h}$, $|\hat{V}_h^{-1}| \overset{p}{\to} 0$.

When the weighting matrix $\hat{W}_h = \left(\hat{\Sigma}_{\gamma_h} + \frac{1}{S^2}\sum_{s=1}^{S}\Sigma^{(s)}_{\gamma_h(\hat{\theta}_{SCN,h})} \right)^{-1}$ is used, the variance of $\hat{\theta}_{SCN,h}$

can be estimated by:

$$
\hat{V}_{SCN,h} = (\widehat{\Gamma}_h(\hat{\theta}_{SCN,h})'\hat{W}_h\widehat{\Gamma}_h(\hat{\theta}_{SCN,h}))^{-1}
\tag{26}
$$

Proof. [Proof of Theorem (8)] *(a). The consistency of the covariance matrix estimator follows from Assumptions 2(c)(e)(f)(g)(k) and Theorem (7). The convergence rate follows from Assumptions 2(a,h). (b). Since the rank of $\Gamma_h(\hat{\theta}_{SCN,h})$ is less than full for $h < \underline{h}$,*
$|\hat{V}_{SCN,h}^{-1}| = |\Gamma_h(\hat{\theta}_{SCN,h})'\left(\hat{\Sigma}_{\gamma_h(\hat{\theta}_{SCN,h})} + \frac{1}{S}\tilde{\Sigma}^{(s)}_{\gamma_h(\hat{\theta}_{SCN,h})} \right)\Gamma_h(\hat{\theta}_{SCN,h})| \overset{p}{\to} 0.$ When h_0 and $h \leq \underline{h}$ are compared, h_0 will be selected with probability approaching one by arguments similar to those in Theorem 1(b). When h_0 and h that is greater than or equal to $\underline{h}$ and does not depend on p are compared, h_0 will be selected with probability approaching one by using arguments analogous to those in Theorem 1. When h_0 and h that is increasing in p are compared, the value of penalty term for such h will be larger than the for h_0. Because the first term of the criterion will converge to the same limit, h_0 will be chosen with probability approaching one. ∎

10. Tables

Table 1(a). Empirical results (Altig et al. (2004))

Parameters	RIRSC		Fixed lags (20)	
	Parameter Estimates	Standard Errors	Parameter Estimates	Standard Estimates
ρ_{xM}	-0.097	0.019	-0.040	0.023
ρ_{xz}	0.588	0.097	0.329	0.073
c_z	0.655	0.051	2.952	0.240
ρ_{μ_z}	0.237	0.054	0.894	0.012
$\rho_{x\Upsilon}$	0.997	0.008	0.822	0.027
c_Υ	0.307	0.034	0.247	0.034
$\rho_{\mu\Upsilon}$	0.344	0.019	0.239	0.033
σ_M	0.334	0.009	0.333	0.008
σ_{μ_z}	0.203	0.013	0.069	0.005
$\sigma_{\mu\Upsilon}$	0.287	0.007	0.304	0.007
ε	0.831	0.022	0.809	0.020
κ	6.907	0.762	3.350	0.269
ξ_w	0.832	0.017	0.713	0.020
b	0.779	0.010	0.706	0.010
σ_a	0.413	0.060	2.029	0.329
c_z^p	0.144	0.109	1.379	0.289
c_Υ^p	0.073	0.045	0.137	0.039
γ	0.207	0.034	0.039	0.005

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

(Altig et al. (2004))

	RIRSC	*Fixed lags (20)*
Firm-Specific Capital Model	1.294	1.515
	(0.010)	(0.019)
Homogeneous Capital Model	2.770	5.655
	(0.047)	(0.127)

Table 2. Empirical results (Christiano et al. (2005))

	RIRSC		*Fixed lags (20)*	
	Parameter	Standard	Parameter	Standard
Parameters	Estimates	Errors	Estimates	Errors
ρ_M	-0.020	0.023	-0.114	0.021
σ_M	0.348	0.008	0.352	0.008
ϵ	0.897	0.021	0.836	0.020
κ	3.732	0.286	4.324	0.353
ξ_w	0.624	0.015	0.645	0.020
b	0.762	0.010	0.717	0.011
λ_f	1.002	0.018	1.097	0.021
σ_a	0.001	0.012	0.041	0.043
γ	0.106	0.019	0.208	0.042

Note to Tables 1-2. The tables report parameter estimates and their standard errors for the IRFME with 20 lags for each IRF, and the IRFME with lags chosen according to the RIRSC (4).

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

ρ	H	IRFME		IRFME_{RIRSC}	
		bias	rej. rate	bias	rej. rate
0.4	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

Table 3(b). Monte Carlo results for the structural VAR(2) case

H		Bias			Rej. rates		
		$\alpha\delta$	δ	β_0	$\alpha\delta$	δ	β_0
1	IRFME	-0.036	0.025	-0.023	0.066	0.059	0.051
	IRFME_{RIRSC}	-0.036	0.025	-0.023	0.057	0.058	0.047
2	IRFME	-0.024	0.018	-0.024	0.065	0.061	0.060
	IRFME_{RIRSC}	-0.036	0.025	-0.024	0.053	0.057	0.048
6	IRFME	-0.028	0.021	-0.030	0.403	0.356	0.238
	IRFME_{RIRSC}	-0.035	0.026	-0.025	0.053	0.050	0.056
12	IRFME	-0.031	0.026	-0.036	0.646	0.565	0.460
	IRFME_{RIRSC}	-0.034	0.025	-0.024	0.050	0.052	0.070

Table 3(c). Monte Carlo results for the Burnside et al.'s (2003) model

		α	δ	$\ln Y$	$\ln\gamma$	$\ln G$	ρ_g	σ_{ε_a}	
	True param.	0.6552	0.0208	8.5721	0.0040	6.9487	0.9819	0.0068	
H					Bias				Average Bias
2	IRFME	0.1207	0.0968	1.1534	0.0572	0.7026	0.0505	0.0690	0.1556
	IRFME$_{RIRSC}$	0.1207	0.0968	1.1534	0.0572	0.7026	0.0505	0.0690	0.1556
10	IRFME	0.2374	0.4029	0.7414	0.3297	0.4606	0.0823	0.1093	0.2965
	IRFME$_{RIRSC}$	0.1150	0.0202	1.0121	0.0197	0.5853	0.0505	0.0679	0.1281
15	IRFME	0.1702	0.2858	0.6388	0.2517	0.3519	0.1305	0.1320	0.2801
	IRFME$_{RIRSC}$	0.1160	0.0212	1.1017	0.0216	0.6794	0.0485	0.0660	0.2934

Table 3(d). Monte Carlo results for the Christiano-Eichenbaum (1992) model

		γ	θ	δ	λ	ρ	
	True param.	0.003	0.6552	0.0208	0.0037	0.9571	
H				Bias			Average Bias
3	IRFME	0.057	0.3174	0.0531	0.0508	0.0420	0.0938
	IRFME$_{RIRSC}$	0.057	0.3174	0.0531	0.0508	0.0420	0.0938
10	IRFME	0.057	0.3560	0.3253	0.3091	0.0635	0.2119
	IRFME$_{RIRSC}$	0.055	0.3163	0.0212	0.0212	0.0935	0.0915
15	IRFME	0.006	0.3554	0.3690	0.3519	0.0781	0.2320
	IRFME$_{RIRSC}$	0.006	0.3386	0.0472	0.0436	0.1096	0.1090

Note to Tables 3(a,b). The tables reports bias (i.e. true parameter value minus estimated value) and rejection rates of 95% nominal confidence intervals for examples AR(1) and (4).

Note to Tables 3(c,d). The tables report mean absolute bias (defined as the absolute value of the difference between the true parameter value and the estimated value) for selected parameters, as well as their average, for two estimators: the IRFME with H lags for each impulse response, and the IRFME with lags chosen according to the RIRSC (4).