Comparison of Misspecified Calibrated Models: The Minimum Distance Approach [DRAFT: preliminary and incomplete]

Viktoria Hnatkovska, Vadim Marmer, and Yao Tang*
University of British Columbia

December 4, 2007

Abstract

This paper presents testing procedures for comparison of misspecified calibrated models. The proposed tests are of the Vuong-type (Vuong, 1989; Rivers and Vuong, 2002). In our framework, an econometrician selects values for the parameters in order to match some characteristics of the data with those implied by the theoretical model. We assume that all competing models are misspecified, and suggest a test for the null hypothesis that all considered models provide equivalent fit to the data characteristics, against the alternative that one of the models is a better approximation. We consider both nested and non-nested cases. Our discussion includes the situation when parameters are estimated to match one set of moments and the model is evaluated by its ability to match another. We also relax the dependence of ranking of the models on the choice of weight matrix by suggesting averaged and sup procedures. The proposed method is applied to comparison of cash-in-advance and portfolio adjustment cost models.

JEL classification: C51; C52
Keywords: misspecified models; calibration; matching; minimum distance estimation; model selection

1 Introduction

This paper presents a method for comparison of calibrated models. While calibration is now an essential tool of quantitative analysis in macroeconomics, surprisingly, there is no

*We thank Don Andrews, Paul Beaudry, Atsushi Inoue, Eric Renault, Shinichi Sakata, Henry Siu, and Gregor Smith for the helpful discussions and suggestions. The second author gratefully acknowledges the research support of the Social Sciences and Humanities Research Council of Canada under grant number 410-2007-1998 and Hampton Research Endowment Fund under Small HSS Research Grant.

generally accepted definition of calibration, and calibration is rather viewed as a research style characterized by a certain attitude toward modelling, assigning parameters' values, and model assessment (Kim and Pagan, 1995). A number of authors define calibration as a sequence of steps allowing one to reduce the general theoretical framework to a quantitative relationship between variables. For instance, Cooley and Prescott (1995) outline three such steps: imposing parametric restrictions; constructing a set of measurements consistent with the parametric class of models; and assigning values to the model parameters. Canova and Ortega (1996) adopt a broader definition of calibration, by including model evaluation into the list of steps.

One of the characteristics of the calibration approach is an explicitly instrumental view of economic models: a calibrationist acknowledges that the model is false and will be rejected by the data (Canova, 1994). The objective of a calibrationist, therefore, is not an assessment of whether the model of interest is true, but rather which features of the data it can be used to capture. Furthermore, a calibrationist may be interested in learning which of the competing, but "false" models provides a better fit to the data.

This paper presents a formal test for comparison of two calibrated models. An important feature of our procedure is its focus on misspecified or "false" structural models. The test is quasi-likelihood ratio (QLR) and based on the difference of the minimum distance (MD) criterion functions corresponding to the competing models. It is argued that among two misspecified models, the econometrician should prefer the one that has a better match to the characteristics of the data.

In a typical calibration exercise, a researcher selects values for the parameters in order to match some characteristics of the observed data with those implied by the theoretical model. For example, a model can be calibrated to match empirical moments, cross-correlations, impulse responses, and stylized facts. Such characteristics will be referred as the properties of a reduced-form model or the reduced-form parameters, since they can be consistently estimated from the data regardless of the true data generating process (DGP). Calibrated parameters can be obtained using informal moment matching, the generalized method of moments (GMM), simulated method of moments (SMM), or maximum likelihood (ML) estimation (Kim and Pagan, 1995); calibration was also formalized as an example of MD estimation in Gregory and Smith (1990, 1993). If the structural model is correctly specified, the calibrated parameters are consistent and asymptotically normal estimators of the structural (or deep) parameters, and statistical inference about the calibrated parameters, as well as model comparisons can be performed in the usual fashion. If the structural model is misspecified, sampling distributions of calibrated parameters are no longer valid and the standard statistical methods no longer apply.

In this paper, we explicitly consider the case of misspecified structural models. Our

methodology uses a classical minimum distance (CMD) estimation procedure to calibrate model parameters. We then show that under some regularity conditions, the calibrated parameters (the CMD estimators) converge in probability to the value of the structural parameters that minimizes the distance between the population reduced from characteristics and those implied by the structural model (pseudo-true values). Further, the CMD estimator is asymptotically normal, however, due to misspecification some adjustments to the asymptotic variance matrix are required. Hall and Inoue (2003), HI hereafter, established such results for GMM estimators.

An important difference between the calibration methodology and GMM (or CMD) is that, in calibration, some parameters are not estimated but rather preset by the calibrationist before the formal estimation step (Gregory and Smith, 1990). A calibrationist distinguishes between free parameters that must be estimated, and other parameters with values chosen on the basis of what are considered to be reasonable values in the literature, or by other methods independent from the data used to calibrate the free parameters. While the existence of the parameters with values assigned using some extra-sample information is problematic if one accepts the structural model as a true DGP (see, for example a Monte Carlo experiment in Gregory and Smith (1990)), we argue that such parameters can be treated very naturally in the misspecified case; one includes such parameters as fixed constants in the function that relates the free structural and reduced-form parameters, and the inference step is performed conditional on the choice of preset values. Note that this function is already assumed to be misspecified, and misspecification has to be corrected in a valid statistical procedure regardless of presence of preset parameters in the model.

One of the prime aspects of calibration is evaluation of a structural model by comparing its implied population moments or other population characteristics with those of the actual data (Gregory and Smith, 1991, 1993; Cogley and Nason, 1995; Kim and Pagan, 1995). However, according to the calibrationist's approach, while evaluating a model one should keep in mind that it is only an approximation and should not be regarded as a null hypothesis to be statistically tested (Prescott, 1991). For example, Watson (1993) and Diebold, Ohanian, and Berkowitz (1998) propose measures of fit for calibrated models that take into account possible misspecification. Further, one can focus on evaluating a structural misspecified model against another misspecified benchmark model, the approach advocated in such papers as [1], Diebold and Mariano (1995), West (1996), Schorfheide (2000), White (2000), Corradi and Swanson (2003), and others.

In this paper, we compare misspecified models by the means of an asymptotic test. In this test, under the null hypothesis the two misspecified models provide an equivalent approximation to the data in terms of characteristics of the reduced-form model. Vuong

[1] As in Vuong (1989), it can be either mixed χ^2or normal for nested or non-nested cases respectively.

(1989) proposed such tests for misspecified models in the maximum likelihood framework, and Rivers and Vuong (2002), RV hereafter, discussed it in a more general setting that includes GMM.

The contribution of our paper relative to RV is threefold. First, RV focused on non-nested models; they provide generic conditions under which the difference between two sample lack-of-fit criteria is asymptotically normal under the null. Our CMD framework allows us to analyze both nested and non-nested models along the lines of Vuong (1989). Such analysis is possible because, in the CMD case, two non-nested structural models can be naturally nested in terms of their corresponding spaces for reduced-form parameters. Note that the nested case does not necessarily correspond to zero restrictions on some structural or deep parameters. Nested case is particularly important since, if the null is not rejected, it allows one to replace the bigger model with the one that is more parsimonious. Also, by considering the specific framework of MD estimation, we provide more specific assumptions and asymptotic results than in RV; and while, as a result, our treatment of the problem is less general than in RV, it covers the important case of calibration.

Second, we also analyze the situation when the models are estimated using one set of reduced-form characteristics and compared with respect to another, which is often the case in the calibration literature. The reason for this, as argued in the literature, is that a model should not be validated by the data that it is calibrated to match in the best possible way (Kleijnen, 1995). For instance, a structural model can be estimated to match the first moments of the data, and evaluated in terms of its ability to match the second moments. Unlike in the standard case, where the asymptotic null distribution of the QLR statistic depends on whether the competing models are nested or non-nested, we show that in the MD framework, when a model is estimated and evaluated on different sets of conditions, the asymptotic null distribution is always normal, regardless of whether the models are nested or non-nested. This fact substantially simplifies the testing procedure in such situations. Essentially, the reason for asymptotic normality in the case of nested models is that there is no selection criteria minimization when the models are estimated and evaluated on different sets of reduced-form parameters.

Third, we also discuss the issue of weighting of moments or reduced-form characteristics. When the model is misspecified, the pseudo-true value of parameters and ranking of the models depend on the choice of weighting scheme; the null hypothesis changes when one applies different weights. Those facts have been discussed in the literature; see, for example, HI and Hall and Pelletier (2007). In this paper, we relax to some extent dependence of ranking of the models on the choice of weighting scheme by suggesting procedures that take into account the models' relative performance for various choices of the weight matrix; we propose an average and sup procedures for model comparison. The averaged test

corresponds to the null hypothesis that the two models have equal lack-of-fit on average. The null hypothesis of the sup test says that one model cannot outperform another for any choice of the weighting matrix.

Comparison of misspecified calibrated models has been studied in the literature from the Bayesian perspective by Schorfheide (2000). His procedure uses vector autoregression (VAR) as a reference model, and allows for model and parameter uncertainty. The comparison of the two misspecified structural models is performed on the basis of overall posterior risk using different metrics chosen by the researcher. Our method can be viewed as a frequentist counterpart of the Schorfheide (2000) procedure. While in Schorfheide (2000), a structural model that achieves the lowest average posterior loss is named a winner, we follow the approach of Vuong (1989) and suggest a test for the null hypothesis that the two models have equal losses.

Corradi and Swanson (2003) designed a Kolmogorov-type test for comparison of misspecified calibrated models. Their approach is different from ours in several respects. In their paper, the models are compared in terms of the distances between the historic empirical cumulative distribution function (CDF) and the CDFs implied by the model or the data simulated from the model. Their model specifications must include the assumptions that allow one to simulate the data. With our approach, one can compare and evaluate structural models that do not necessarily imply a distribution for the data. The two approaches also differ in terms of measures of fit. The outcome of comparison of misspecified models depends on the choice of a measure of fit between the true DGP and its approximation, and different measures can lead to different ranking of the models. The approach of Corradi and Swanson (2003) is closer to that of Vuong (1989) and Kitamura (2000),[2] since the models are compared in terms of the distributional characteristics; we, on the other hand, focus on ability of a model to approximate some moment conditions, or impulse responses, and other such characteristics.

The issue of misspecified calibrated models was addressed by Dridi, Guay, and Renault (2007) using indirect inference.[3] The main focus their paper is consistent estimation of some deep parameters when the model is misspecified with respect to additional, nuisance, parameters. Dridi, Guay, and Renault (2007) also emphasize the necessity of correcting asymptotic variances formulas when there is a possibility of misspecification; such corrections are discussed in our paper for the MD estimators.

We apply our methodology to compare two standard monetary business cycle models. The first model is a cash-in-advance (CIA) model, while the second model is the Lucas (1990) and Fuerst (1992) model with the portfolio adjustment costs (PAC). The two models

[2] Vuong (1989) and Kitamura (2000) use the Kullback-Leibler Information Criterion as a measure of distance between a model and the true DGP.

[3] As a matter of fact, our method can be viewed as an example of indirect inference without simulations.

have the same underlying structure except in the information sets that agents possess when making their decisions. We judge the performance of the models based on their ability to replicate the dynamics of the business cycles in the US. As our comparison criterion or reduced-form characteristics, we use the response of the growth rate of output and inflation to an unanticipated monetary shock. The "true" impulse responses are obtained from the reference model which we choose to be a structural vector autoregression (SVAR). We identify structural shocks using the Blanchard and Quah (1989) decomposition under the restriction of the long-run monetary neutrality of output.

For each structural model we compute its impulse responses and calculate the structural parameters that minimize the deviations between these theoretical impulse responses and those implied by the data. We then calculate the implied distance functions for each model, and test the null that the two models have same lack of fit. In our applications we find that PAC model provides significantly better fit to the data along the chosen comparison criterion. This result is consistent with the finding of Schorfheide (2000) that the PAC model outperforms the CIA model in predicting the response of output growth to the innovation in monetary injection growth.

The paper proceeds as follows. Section 2 introduces the framework. Section 3 describes the asymptotic properties of CMD estimators under misspecification. Section 4 suggests a QLR statistic for model comparison. We discuss the distribution of the QLR statistic in the cases of nested, strictly non-nested and overlapping models. In Section 5, we consider the situation when a model is estimated using one set of reduced-form parameters and evaluated with respect to another. Section 6 discusses the averaged and sup procedures. Section 7 illustrates the technique with an empirical application. All proofs are in the Appendix.

2 Definitions

This section formally defines calibration as MD estimation and introduces the framework of comparison of two calibrated models. The definition of calibration is similar to that of Gregory and Smith (1990); it is viewed as an example of CMD estimation (for a discussion of CMD see Newey and McFadden (1994)). CMD estimation of optimization based models was considered recently in the econometrics literature by Moon and Schorfheide (2002).

Let $Y_n(\omega)$ be a data matrix of the sample size n defined on the probability space $(\Omega, \mathcal{F}, P)$. All random quantities in this paper are some functions of the data Y_n. We use h to denote an m-vector of parameters of some reduced-form model. Its true value, $h_0 \in R^m$ depends on the true unknown structural model of the economy and its parameters. For example, h can be a number of moments, cross-correlations, impulse responses, and etc. While the true structural model is unknown, we will assume that reduced-form parameter

h_0 can be estimated consistently from the data. Let $\hat{h}_n$ denote an estimator of h. We assume that $\hat{h}_n$ has the following properties.

Assumption 2.1 **(a)** $\hat{h}_n \to_p h_0 \in R^m$.

(b) $n^{1/2}\left(\hat{h}_n - h_0\right) \to_d N(0, \Lambda_0)$, *where* Λ_0 *is positive definite* $m \times m$ *matrix.*

(c) *There is* $\hat{\Lambda}_n$ *such that* $\hat{\Lambda}_n \to_p \Lambda_0$.

According to Assumptions 2.1(a) and (b), $\hat{h}_n$ is a consistent and asymptotically normal estimator of h_0. Similarly to h_0, its asymptotic variance, Λ_0, depends on the unknown true structural model and its parameters. Part (c) of the assumption requires that Λ_0 also can be estimated consistently from the data. The above assumptions are of high level; they can be verified under more primitive conditions. For example, Assumption 2.1 holds when h_0 and $\hat{h}_n$ are functions of the first and second population and sample moments of Y_n respectively, $Y_n = (y_1', \ldots, y_n')'$, such that $\{y_t\}$ is a stationary mixing sequence with ϕ of size $-r/(r-1)$, $r \geq 2$, or α of size $-2r/(r-2)$, $r > 2$, $E\|y_t\|^{4r+\delta} < \infty$, where $\|\cdot\|$ denotes the Euclidean norm, and $Var\left(n^{-1/2}\sum_{t=1}^{n} y_t\right)$ is uniformly positive definite (White, 2001); alternatively, Assumption 2.1 can be verified using the linear processes conditions of Phillips and Solo (1992).

In our framework, m is fixed by the econometrician and independent of the data. We assume that the econometrician chooses h and m according to the economic importance of the reduced-form characteristics that a model is used to explain; note, however, that there are recent methods allowing one to choose m using the data and a statistical information criterion (Hall, Inoue, Nason, and Rossi, 2007).

Let $\theta \in \Theta \subset R^k$ be the vector of deep parameters corresponding to a structural model specified by the econometrician. We assume that one can compute analytically the value of the reduced-form parameters h given the model and a value of θ. The mapping from the space of θ to the space of reduced-form parameters is given by the function $f : \Theta \to R^m$, which we call the binding function using the terminology of indirect inference (Gouriéroux, Monfort, and Renault, 1993; Dridi, Guay, and Renault, 2007). In the remainder of the paper, structural models are referred by their binding functions.

The vector θ denotes only the free parameters that are estimated using the sample information; any preset parameters are included as constants into the binding function f. Such parameters usually are assigned values based on extra-sample information, and we assume that model comparison and evaluation is performed conditionally on the choice of preset parameters. As we discuss in the introduction, such a practice can be problematic if one wants to treat f as a true DGP; however, here it will be assumed that f is misspecified, so existence of preset parameters does not introduce additional complications.

The econometrician calibrates θ to minimize the distance between the sample reduced-form characteristics of the data, $\hat{h}_n$, and those implied by the chosen structural model, $f(\theta)$. Let A_n be a possibly random $m \times m$ weight matrix. The weight matrix can be nonrandom and chosen by the econometrician to put more weight on the relatively more important reduced-form parameters. Alternatively, it can be data dependent and, therefore, random. For example, in the spirit of GMM estimation, A_n can be set such that $A_n' A_n = \hat{\Lambda}_n^{-1}$, which exists with probability approaching one due to Assumption 2.1(c).

Assumption 2.2 *$A_n \to_p A$, where A is of full rank.*

The calibrated θ, or the CMD estimator of θ, is given by the value that minimizes the weighted distance function:

$$\hat{\theta}_n(A_n) = \arg\min_{\theta \in \Theta} \left\| A_n \left(\hat{h}_n - f(\theta) \right) \right\|^2. \tag{2.1}$$

The structural model is said to be correctly specified if for some value $\theta_0 \in \Theta$ the binding function f produces exactly the true value of the reduced-form parameter. The following definition is similar to Definitions 1 and 2 of HI.

Definition 2.1 *The structural model f is said to be correctly specified if there exists some $\theta_0 \in \Theta$ such that $f(\theta_0) = h_0$; f is said to be misspecified if $\inf_{\theta \in \Theta} \|(h_0 - f(\theta))\| > 0$.*

Naturally, the structural model chosen by the econometrician is correctly specified in the sense of Definition 2.1 in the unlikely situation that f is the true data generating process. Also, the model f is correctly specified according to Definition 2.1 in the case of exact identification, i.e. when $m = k$, even if the structural model and its binding function describe an incorrect DGP. Thus, the structural model is misspecified if it is overidentified and, for no value of θ, it can replicate the reduced-form characteristics. The requirement on overidentification is a crucial one since an exactly identified model is never misspecified according to Definition 2.1 (see HI for a discussion of overidentification and misspecification). Typically, the number of reduced-form parameters available for calibration exceeds k, and, therefore, the calibrationist can always chose the binding function and reduced-form parameters so that the model is overidentified.

We assume that the econometrician considers two competing structural models. The second structural model is given by the binding function g and the vector of deep parameters $\gamma \in \Gamma \subset R^l$. Let $\hat{\gamma}_n$ be the calibrated value of γ, where $\hat{\gamma}_n$ constructed in the similar fashion to $\hat{\theta}_n$ in (2.1):

$$\hat{\gamma}_n(A_n) = \arg\min_{\gamma \in \Gamma} \left\| A_n \left(\hat{h}_n - g(\gamma) \right) \right\|^2.$$

We assume that f and g are both overidentified and misspecified in the sense of Definition 2.1.

Assumption 2.3 *f and g are misspecified according to Definition 2.1.*

Next, we define the pseudo-true values of the structural parameters θ and γ. The pseudo-true value minimizes the distance between h_0 and the binding functions for a given weight matrix A.

Assumption 2.4 (a) *There exists a unique $\theta_0(A) \in \Theta$ such that for all $\theta \in \Theta$, $\|A(h_0 - f(\theta_0(A)))\| \leq \|A(h_0 - f(\theta))\|$.*

(b) *There exists a unique $\gamma_0(A) \in \Gamma$ such that for all $\gamma \in \Gamma$, $\|A(h_0 - g(\gamma_0(A)))\| \leq \|A(h_0 - g(\gamma))\|$.*

The pseudo-true value is written as a function of A to emphasize that different choices of weight matrix may lead to different minimizers of $\|A(h_0 - f(\theta))\|$ (see Maasoumi and Phillips (1982) and HI). For notational brevity, we may suppress the dependence on A if there is no ambiguity regarding the choice of A. The uniqueness of θ_0 and γ_0 is a strong assumption, which may not be true in the case of misspecified models. However, it is important for the inference procedure suggested in the paper and often assumed in the literature (see Assumption 3 of RV and Assumption 3 of HI). The uniqueness of the pseudo-true value can be verified with the probability approaching one since the binding functions are known, $A_n \to_p A$, and $\hat{h}_n$ is a consistent estimator of h_0 by Assumption 2.1(a). When the pseudo-true value lies in the interior of Θ, it has to be a unique solution to the following equation, provided that f is differentiable:

$$\frac{\partial f(\theta_0(A))'}{\partial \theta} A' A(h_0 - f(\theta_0(A))) = 0. \tag{2.2}$$

Due to Assumption 2.2 on A, $\partial f(\theta_0(A))/\partial \theta'$ must have the rank k for $\theta_0(A)$ to be unique.

The econometrician's objective is to choose between the two wrong models the one that provides a better A-weighted fit to the reduced-form parameters h_0. We suggest a testing procedure for the null hypothesis that the two models are equally wrong

$$H_0 : \|A(h_0 - f(\theta_0(A)))\| = \|A(h_0 - g(\gamma_0(A)))\|, \tag{2.3}$$

against the alternatives in which one of the models provides a better fit. The econometrician prefers the model f if the following alternative is true.

$$H_f : \|A(h_0 - f(\theta_0(A)))\| < \|A(h_0 - g(\gamma_0(A)))\|. \tag{2.4}$$

Similarly, the econometrician prefers the model g when

$$H_g : \|A\left(h_0 - f\left(\theta_0\left(A\right)\right)\right)\| > \|A\left(h_0 - g\left(\gamma_0\left(A\right)\right)\right)\|$$

is true. The hypotheses are analogous to those of Vuong (1989) and RV. Note that, in the current framework, the decision depends on the choice of the weight matrix A. Thus, under the null, the two structural models provide equivalent explanations of the reduced-form characteristics for a given weighting scheme A. Naturally, different weighting schemes may lead to different ranking of f and g.

In order to test the null hypothesis in (2.3), it is natural to consider a sample counterpart of the difference in fit between the two competing models which is given by the following QLR statistic

$$QLR_n\left(\hat{\theta}_n\left(A_n\right), \hat{\gamma}_n\left(A_n\right)\right) = -\left\|A_n\left(\hat{h}_n - f\left(\hat{\theta}_n\right)\right)\right\|^2 + \left\|A_n\left(\hat{h}_n - g\left(\hat{\gamma}_n\right)\right)\right\|^2. \qquad (2.5)$$

Given our assumptions, QLR_n consistently estimates the difference in the population measures of fit $-\|A\left(h_0 - f\left(\theta_0\left(A\right)\right)\right)\| + \|A\left(h_0 - g\left(\gamma_0\left(A\right)\right)\right)\|$, as implied by the results presented in the next section.

3 Properties of the CMD estimators of structural parameters

In this section, we discuss the asymptotic properties of the CMD estimators defined in the previous section. We make the following assumptions about the binding functions f and g and their parameters' spaces.Θ and Γ.

Assumption 3.1 **(a)** *Θ and Γ are compact.*

(b) *θ_0 lies in the interior of Θ; γ_0 lies in the interior of Γ.*

(c) *f is continuous on Θ; g is continuous on Γ.*

The following theorem gives consistency of the CMD estimators of θ and γ.

Theorem 3.1 *Suppose that Assumptions 2.1, 2.2, 2.4, and 3.1 hold. Then, $\hat{\theta}_n \to_p \theta_0$ and $\hat{\gamma}_n \to_p \gamma_0$.*

As usual, the asymptotic distribution of the CMD estimators centered around their pseudo-true values can be derived from the mean value expansion of the sample first order conditions for the minimization problem in (2.1). We make the following assumption.

Assumption 3.2 *The binding function f is twice continuously differentiable in the neighborhood of θ_0; the binding function g is twice continuously differentiable in the neighborhood of γ_0.*

It follows from Theorem 3.1 and Assumption 3.2 that the binding functions evaluated at the corresponding CMD estimators are twice continuously differentiable with probability approaching one. Thus, the CMD estimator of θ must satisfy the first order conditions:

$$\frac{\partial f\left(\hat{\theta}_n\right)}{\partial \theta}' A_n' A_n \left(\hat{h}_n - f\left(\hat{\theta}_n\right)\right) = 0.$$

Using the mean value theorem twice to expand $f\left(\hat{\theta}_n\right)$ around $f(\theta_0)$ and $\partial f\left(\hat{\theta}_n\right)/\partial\theta'$ around $\partial f(\theta_0)/\partial\theta'$, and taking into account the population first order conditions (2.2), we obtain the following equation determining the asymptotic distribution of the CMD estimators in the misspecified case:

$$\begin{aligned}\left(\hat{\theta}_n - \theta_0\right) &= F_n^{-1} \frac{\partial f\left(\hat{\theta}_n\right)}{\partial \theta}' \\ &\quad \times \left(A_n' A_n \left(\hat{h}_n - h_0\right) + \left(A_n' A_n - A'A\right)\left(h_0 - f(\theta_0)\right)\right), \end{aligned} \tag{3.1}$$

where

$$\begin{aligned} F_n &= \frac{\partial f\left(\hat{\theta}_n\right)}{\partial \theta}' A_n' A_n \frac{\partial f\left(\widetilde{\theta}_n\right)}{\partial \theta'} - M_{f,n}, \\ M_{f,n} &= \left(I_k \otimes \left(h_0 - f(\theta_0)\right)' A'A\right) \frac{\partial}{\partial \theta'} vec\left(\frac{\partial f\left(\overline{\theta}_n\right)}{\partial \theta'}\right), \end{aligned}$$

$\left\|\widetilde{\theta}_n - \theta_0\right\|$ and $\left\|\overline{\theta}_n - \theta_0\right\|$ are dominated by $\left\|\hat{\theta}_n - \theta_0\right\|$, and $vec(\cdot)$ denotes column vectorization of a matrix. The term $M_{f,n}$, which involves the second derivatives of f, reflects the fact that the model is misspecified. It is not present in the correctly specified case as well as the second summand in (3.1). The analogous result holds for g and $\hat{\gamma}_n$. The expansion is similar to equation (9) of HI, however, in the case of CMD it involves one term less than in the GMM case. This is due to the fact that f and g do not have the data Y_n as an argument: in the case of CMD, the data and the parameters are additively separated, the data enters through h, and the parameters through the binding function. The result in (3.1) is also similar to Assumptions 10 and 13 of RV.

The expansion in (3.1) shows that the convergence rates of the CMD of structural parameters in the misspecified case depends on that of the reduced-form parameters and the

weight matrices. In many situations, it is natural to assume that $n^{1/2}\left(\hat{h}_n - h_0\right)$ is asymptotically normal as we do in Assumption 2.1(b). Concerning the weight matrices, HI distinguish between several cases: (i) fixed weight matrices, (ii) $n^{1/2}vec\left(A_n'A_n - A'A\right)$ is asymptotically normal, and (iii) $A_n'A_n$ is the inverse of centered or uncentred HAC estimator. In the current framework, the case of fixed weight matrices plays the important role, since the weight matrix defines the relative importance of various reduced-form characteristics of the data. On the other hand, the weights introduced by $A_n'A_n = \hat{\Lambda}_n^{-1}$ can be motivated on efficiency grounds when the model is correctly specified.

Consider the correctly specified case $h_0 - f(\theta_0) = 0$. In this case, Assumptions 2.1, 2.2, 2.4, 3.1, and 3.2 and Theorem 3.1 imply that $n^{1/2}\left(\hat{\theta}_n - \theta_0\right)$ has asymptotically normal distribution with the variance matrix

$$\left(\frac{\partial f(\theta_0)'}{\partial\theta}A'A\frac{\partial f(\theta_0)}{\partial\theta'}\right)^{-1}\frac{\partial f(\theta_0)'}{\partial\theta}A'A\Lambda_0A'A\frac{\partial f(\theta_0)}{\partial\theta'}\left(\frac{\partial f(\theta_0)'}{\partial\theta}A'A\frac{\partial f(\theta_0)}{\partial\theta'}\right)^{-1}.$$

As usual, in the correctly specified case, the efficient CMD estimator corresponds to $A_n'A_n = \hat{\Lambda}_n^{-1}$. However, the such weights may be difficult to interpret. Furthermore, when $\hat{\Lambda}_n$ is a HAC estimator and the model is misspecified, $\hat{\theta}_n$ has a convergence rate slower that $n^{1/2}$ as shown in HI. Consequently, usage of HAC based weight matrices can potentially lead to less powerful tests, although note that each weigh matrix A corresponds to a different null hypothesis when the models are misspecified. In this paper, we focus on cases (i) and (ii).

Define

$$\begin{aligned} F_0 &= \frac{\partial f(\theta_0)'}{\partial\theta}A'A\frac{\partial f(\theta_0)}{\partial\theta'} - M_{f,0}, \text{ where} \qquad (3.2)\\ M_{f,0} &= \left(I_k \otimes (h_0 - f(\theta_0))'A'A\right)\frac{\partial}{\partial\theta'}vec\left(\frac{\partial f(\theta_0)}{\partial\theta'}\right), \text{ and}\\ G_0 &= \frac{\partial g(\gamma_0)'}{\partial\gamma}A'A\frac{\partial g(\gamma_0)}{\partial\gamma'} - M_{g,0}, \text{ where} \qquad (3.3)\\ M_{g,0} &= \left(I_l \otimes (h_0 - g(\gamma_0))'A'A\right)\frac{\partial}{\partial\gamma'}vec\left(\frac{\partial g(\gamma_0)}{\partial\gamma'}\right). \end{aligned}$$

Assumption 3.3 *F_0 and G_0 are non-singular.*

The above assumption is similar to Assumption 5 of HI. Theorem 3.1 implies that $F_n \to_p F_0$. We define further

$$\begin{aligned} V_{ff,0} &= F_0^{-1}\frac{\partial f(\theta_0)'}{\partial\theta}A'A\Lambda_0A'A\frac{\partial f(\theta_0)}{\partial\theta'}F_0'^{-1},\\ V_{fg,0} &= F_0^{-1}\frac{\partial f(\theta_0)'}{\partial\theta}A'A\Lambda_0A'A\frac{\partial g(\gamma_0)}{\partial\gamma'}G_0'^{-1}, \end{aligned}$$

$$
\begin{aligned}
V_{gg,0} &= G_0^{-1}\frac{\partial g\left(\gamma_0\right)'}{\partial\gamma}A'A\Lambda_0A'A\frac{\partial g\left(\gamma_0\right)}{\partial\gamma'}G_0'^{-1}, \text{ and} \\
V_0 &= \begin{pmatrix} V_{ff,0} & V_{fg,0} \\ V_{fg,0}' & V_{gg,0} \end{pmatrix}. \qquad (3.4)
\end{aligned}
$$

The following theorem describes the asymptotic distribution of the CMD estimators in the fixed weight matrix case.

Theorem 3.2 *Suppose that $A_n = A$ for all $n \geq 1$. Under Assumptions 2.1, 2.2, 2.4, and 3.1-3.3,*

$$
n^{1/2}\begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\gamma}_n - \gamma_0 \end{pmatrix} \to_d N\left(0_{(k+l)\times 1}, V_0\right).
$$

When the weight matrix depends on the data, we extend Assumption 2.1 with Assumption 3.4 below, which assumes that the elements of $A_n'A_n$ are root-n consistent and asymptotically normal estimators of the elements of $A'A$ and they can be correlated with $\hat{h}_n$.

Assumption 3.4 **(a)** $n^{1/2}\left(\left(\hat{h}_n - h_0\right)', vec\left(A_n'A_n - A'A\right)'\right)' \to_d N\left(0, \Lambda_0^A\right)$, *where Λ_0^A is a positive definite $m(m+1)\times m(m+1)$ matrix*

$$
\Lambda_0^A = \begin{pmatrix} \Lambda_0 & \Lambda_{0A} \\ \Lambda_{0A}' & \Lambda_{AA} \end{pmatrix}.
$$

(b) *There is $\hat{\Lambda}_0^A$ such that $\hat{\Lambda}_0^A \to_p \Lambda_0^A$.*

The assumption is similar to condition (12) of Theorem 2 in HI. The assumption allows $A_n'A_n$ to depend on $\hat{h}_n$, $\hat{\theta}_n$, and $\hat{\gamma}_n$; however, as we discuss above, Assumption 3.4 rules out HAC based estimators of $A'A$.

Now, in view of expansion (3.1) and a similar expansion for $\hat{\gamma}_n$, the asymptotic distribution of $\hat{\theta}_n$ and $\hat{\gamma}_n$ depends on that of $A_n'A_n$. Define $V_{ff,0}^A$, $V_{gg,0}^A$, and $V_{fg,0}^A$, the asymptotic variance of $\hat{\theta}_n$, the asymptotic variance of $\hat{\gamma}_n$, and the asymptotic covariance of $\hat{\theta}_n$ and $\hat{\gamma}_n$ respectively as following:

$$
\begin{aligned}
V_{ff,0}^A &= F_0^{-1}\frac{\partial f\left(\theta_0\right)'}{\partial\theta}D_{f,0}^A\Lambda_0^AD_{f,0}^{A\prime}\frac{\partial f\left(\theta_0\right)}{\partial\theta'}F_0'^{-1}, \text{ where} \\
D_{f,0}^A &= \begin{pmatrix} A'A & I_m\otimes\left(h_0 - f\left(\theta_0\right)\right)' \end{pmatrix}; \\
V_{gg,0}^A &= G_0^{-1}\frac{\partial g\left(\gamma_0\right)'}{\partial\gamma}D_{g,0}^A\Lambda_0^AD_{g,0}^{A\prime}\frac{\partial g\left(\gamma_0\right)}{\partial\gamma'}G_0'^{-1}, \text{ where} \\
D_{g,0}^A &= \begin{pmatrix} A'A & I_m\otimes\left(h_0 - g\left(\gamma_0\right)\right)' \end{pmatrix};
\end{aligned}
$$

$$
\begin{aligned}
V_{fg,0}^{A} &= F_0^{-1}\frac{\partial f(\theta_0)'}{\partial\theta}D_{f,0}^{A}\Lambda_0^{A}D_{g,0}^{A\prime}\frac{\partial g(\gamma_0)}{\partial\gamma'}G_0'^{-1}, \text{ and} \\
V_0^{A} &= \begin{pmatrix} V_{ff,0}^{A} & V_{fg,0}^{A} \\ V_{fg,0}^{A\prime} & V_{gg,0}^{A} \end{pmatrix}.
\end{aligned}
$$

The joint asymptotic distribution of $\hat{\theta}_n$ and $\hat{\gamma}_n$ is given by the next theorem.

Theorem 3.3 *Under Assumptions 2.2, 2.4, and 3.1-3.4,*

$$
n^{1/2}\begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\gamma}_n - \gamma_0 \end{pmatrix} \to_d N\left(0_{(k+l)\times 1}, V_0^{A}\right).
$$

The asymptotic variance of $\hat{\theta}_n$ and $\hat{\gamma}_n$ can be consistently estimated by the plug-in method, i.e. by replacing $h_0 - f(\theta_0)$ with $\hat{h}_n - f\left(\hat{\theta}_n\right)$ and so on.

4 Model comparison

The distribution of the QLR_n statistic in (2.5) depends on the relationship between the two models. Similarly to Vuong (1989), we consider the following three cases: nested, strictly non-nested, and overlapping models f and g. Define

$$
\begin{aligned}
\mathcal{F} &= \{h \in R^m : h = f(\theta), \theta \in \Theta\}, \\
\mathcal{G} &= \{h \in R^m : h = g(\gamma), \gamma \in \Gamma\}.
\end{aligned}
$$

The subsets of R^m, $\mathcal{F}$ and $\mathcal{G}$, represent the spaces for the reduced-form parameter h that are spanned by the structural models f and g respectively. The relationship between the two structural models can be defined in terms of $\mathcal{F}$ and $\mathcal{G}$.

Definition 4.1 *The two structural models f and g are said to be*

(a) *nested if $\mathcal{F} \subset \mathcal{G}$ or $\mathcal{G} \subset \mathcal{F}$,*

(b) *strictly non-nested if $\mathcal{F} \cap \mathcal{G} = \emptyset$,*

(c) *overlapping if $\mathcal{F} \cap \mathcal{G} \neq \emptyset$, $\mathcal{F} \not\subset \mathcal{G}$, and $\mathcal{G} \not\subset \mathcal{F}$.*

Note that the nested case does not necessarily correspond to zero restrictions on the elements of structural parameters. The two models can be totally different in terms of their construction so that θ and γ are not directly comparable, and still be nested with respect to the spaces they span for h. The asymptotic behavior of the QLR statistic and resulting inference procedure depend on whether f and g are nested, strictly non-nested, or overlapping.

4.1 Nested models

Suppose that $\mathcal{G} \subset \mathcal{F}$. In this case, the model g cannot provide a better fit than f. Thus, in this case the econometrician is interested in testing H_0 against H_f, i.e. whether the approximation to the reduced-form characteristics of the data obtained with the smaller model is equivalent to that of the bigger model. Since the models are nested, and under Assumption 2.4 of unique pseudo-true values, the null hypothesis can be equivalently stated as $f(\theta_0) = g(\gamma_0)$. Indeed, let $h_{f,0} = f(\theta_0)$ and $h_{g,0} = g(\gamma_0)$; we have that under the null of models equivalence $\|A(h_0 - h_{f,0})\| = \|A(h_0 - h_{g,0})\|$. However, since the models are nested, $h_{g,0} \in \mathcal{F}$, and there should be some $\tilde{\theta}_0 \in \Theta$ such that $h_{g,0} = f\left(\tilde{\theta}_0\right)$, which would violate Assumption 2.4 if $h_{f,0} \neq h_{g,0}$. We have the following result.

Lemma 4.1 *Suppose that Assumption 2.4 holds, and the models are nested according to Definition 4.1. Then, under the H_0 in (2.3), $f(\theta_0) = g(\gamma_0)$.*

The QLR statistic depends on the weight matrix explicitly and through the estimators $\hat{\theta}_n$ and $\hat{\gamma}_n$. The following theorem establishes the distribution of the QLR statistic in the case of fixed weight matrices.

Theorem 4.1 *Suppose that $A_n = A$ for all $n \geq 1$, A is of full rank, Assumptions 2.1, 2.3, 2.4, 3.1-3.3 hold, and $\mathcal{G} \subset \mathcal{F}$.*

(a) *Under H_0,*

$$nQLR_n\left(\hat{\theta}_n, \hat{\gamma}_n\right) \to_d Z'\Lambda_0^{'1/2}A'A\left(W_{g,0} - W_{f,0}\right)A'A\Lambda_0^{1/2}Z,$$

where

$$\begin{aligned}
Z &\sim N(0, I_m), \\
W_{f,0} &= W_{f,0}(1) - W_{f,0}(2) - W_{f,0}(3), \\
W_{f,0}(1) &= \frac{\partial f(\theta_0)}{\partial \theta'}F_0^{'-1}\frac{\partial f(\theta_0)}{\partial \theta}'A'A\frac{\partial f(\theta_0)}{\partial \theta'}F_0^{-1}\frac{\partial f(\theta_0)}{\partial \theta}', \\
W_{f,0}(2) &= \frac{\partial f(\theta_0)}{\partial \theta'}F_0^{'-1}\frac{\partial f(\theta_0)}{\partial \theta}' + \frac{\partial f(\theta_0)}{\partial \theta'}F_0^{-1}\frac{\partial f(\theta_0)}{\partial \theta}', \\
W_{f,0}(3) &= \frac{\partial f(\theta_0)}{\partial \theta'}F_0^{'-1}M_{f,0}' + M_{f,0}F_0^{-1}\frac{\partial f(\theta_0)}{\partial \theta}',
\end{aligned}$$

and $W_{g,0}$ defined similarly to $W_{f,0}$ with θ_0, $\partial f/\partial\theta$, F_0, and $M_{f,0}$ replaced by γ_0, $\partial g/\partial\gamma$, G_0, and $M_{g,0}$ respectively.

(b) *Under H_f, $nQLR_n\left(\hat{\theta}_n, \hat{\gamma}_n\right) \to +\infty$ with probability one.*

According to Theorem 4.1, under the null the re-scaled QLR statistic has a mixed χ^2 distribution. The result is similar to the one established by Vuong (1989) in the case of nested models. According to part (a) of the theorem, one should reject the null when

$$nQLR_n\left(\hat{\theta}_n, \hat{\gamma}_n\right) > c_\alpha,$$

where c_α is the critical value satisfying $P\left(nQLR_n\left(\hat{\theta}_n, \hat{\gamma}_n\right) > c_\alpha | H_0\right) \to \alpha$ as $n \to \infty$. Under the null, the distribution of the statistic is nonstandard and depends on the unknown parameters h_0, θ_0, γ_0, and Λ_0. However, its asymptotic distribution can be approximated by simulations using the consistent estimators of the unknown parameters. Let $\hat{W}_{f,n}$ and $\hat{W}_{g,n}$ be as in part (a) of Theorem 4.1, but with h_0, θ_0, γ_0, and Λ_0 replaced by $\hat{h}_n$, $\hat{\theta}_n$, $\hat{\gamma}_n$, and $\hat{\Lambda}_n$ respectively. First, one simulates a vector of $N(0, I_m)$ random variables, Z_r, and calculates

$$QLR_{nr} = Z_r' \hat{\Lambda}_n'^{1/2} A' A \left(\hat{W}_{g,n} - \hat{W}_{f,n}\right) A' A \hat{\Lambda}_n^{1/2} Z_r.$$

As $n \to \infty$, the asymptotic distribution of QLR_{nr} is that described in part (a) of Theorem 4.1. Repeating this for $r = 1, \ldots, R$ with Z_r being drawn independently across r's, the simulated critical value $c_{1-\alpha,n,R}$ is the $1-\alpha$ quantile of $\{QLR_{nr} : r = 1, \ldots, R\}$. Hence, in practice in the case of nested models, one rejects the null when

$$nQLR_n\left(\hat{\theta}_n, \hat{\gamma}_n\right) > c_{\alpha,n,R}.$$

Similarly to Vuong (1989), one can also consider the adjusted QLR statistic

$$\widetilde{QLR}_n\left(\hat{\theta}_n, \hat{\gamma}_n\right) = QLR_n\left(\hat{\theta}_n, \hat{\gamma}_n\right) + K_n(\mathcal{F}, \mathcal{G}), \tag{4.1}$$

where $K_n(\mathcal{F}, \mathcal{G}) = o_p\left(n^{-1}\right)$, and $K_n(\mathcal{F}, \mathcal{G})$ is the adjustment factor depending, for example, on the number of the structural parameters in f and g, such as $K_n(\mathcal{F}, \mathcal{G}) = -(k-l)c_n$ with $c_n = o_p\left(n^{-1}\right)$.

4.2 Strictly non-nested models

In the case of strictly non-nested models, the space of reduced-form parameters generated under f and g does not have any common points. Either one of the models can be preferred as providing a better fit to the reduced-form parameters. Again similarly to the results of Vuong (1989), under the null in the case of strictly non-nested models, the asymptotic distribution of re-scaled QLR statistic becomes normal, as shown in the following theorem

Theorem 4.2 *Suppose that $A_n = A$ for all $n \geq 1$. Under Assumptions Assumptions 2.1, 2.3, 2.4, 3.1-3.3, if $\mathcal{F} \cap \mathcal{G} = \emptyset$,*

(a) *Under H_0, $n^{1/2} QLR_n \left(\hat{\theta}_n, \hat{\gamma}_n\right) \to_d N\left(0, \omega_0^2\right)$, where*

$$\omega_0 = 2 \left\| \Lambda_0^{1/2} A' A \left(f\left(\theta_0\right) - g\left(\gamma_0\right) \right) \right\|.$$

(b) *Under H_f, $n^{1/2} QLR_n \left(\hat{\theta}_n, \hat{\gamma}_n\right) \to \infty$ with probability one; under H_g, $n^{1/2} QLR_n \left(\hat{\theta}_n, \hat{\gamma}_n\right) \to -\infty$ with probability one.*

The QLR statistic has an asymptotically normal distribution in the non-nested case because, in the asymptotic expansion of the QLR statistic, there is now the dominating $O_p\left(n^{-1/2}\right)$ term $\left(f\left(\theta_0\right) - g\left(\gamma_0\right)\right)' A' A \left(\hat{h}_n - h_0\right)$, as we show in the proof of Theorem 4.2. When the models are nested, this term equals zero since, under the null, $f\left(\theta_0\right) = g\left(\gamma_0\right)$ in the nested case.

In practice, the null should be rejected in favor of H_f when

$$n^{1/2} QLR_n \left(\hat{\theta}_n, \hat{\gamma}_n\right) / \hat{\omega}_n > z_{1-\alpha/2},$$

where z_α is the α quantile of the standard normal distribution, and

$$\hat{\omega}_n = 2 \left\| \hat{\Lambda}_n^{1/2} A' A \left(f\left(\hat{\theta}_n\right) - g\left(\hat{\gamma}_n\right) \right) \right\|.$$

The null should be rejected in favor of H_g when

$$n^{1/2} QLR_n \left(\hat{\theta}_n, \hat{\gamma}_n\right) / \hat{\omega}_n < -z_{1-\alpha/2}.$$

One can also base inference on the adjusted QLR statistic as in (4.1), in this case, with the correction factor

$$K_n\left(\mathcal{F}, \mathcal{G}\right) = o_p\left(n^{-1/2}\right).$$

One of the important issues discussed in the literature on misspecified model selection is describing the conditions under which the asymptotic variance of statistics similar to our QLR_n is strictly positive (see, for example, Assumption 7 of RV). In our case, as one can see from part (a) of Theorem 4.2, $\omega_0^2 > 0$ whenever $f\left(\theta_0\right) \neq g\left(\gamma_0\right)$; the last condition holds for non-nested models in the sense of Definition 4.1. When $\omega_0^2 = 0$, QLR_n has a mixed χ^2 distribution as described in the previous subsection. This conclusion is similar to that of RV, Section 6.

4.3 Overlapping models

The models are overlapping when the intersection of $\mathcal{F}$ and $\mathcal{G}$ is non-empty, however, neither model nests the other. One has to consider two possible situations when the models are overlapping. First, it is possible that $f(\theta_0) = g(\gamma_0)$; in this case $\omega_0^2 = 0$, and the QLR statistic re-scaled by n has an asymptotic mixed χ^2 distribution. In the second case, we have $f(\theta_0) \neq g(\gamma_0)$, but $\|A(h_0 - f(\theta_0(A)))\| = \|A(h_0 - g(\gamma_0(A)))\|$, and the QLR statistic, when re-scaled by $n^{1/2}$, is asymptotically normal. In order to test the null hypothesis, one has to determine which one of the two possible situation applies.

Vuong (1989) proposed the following sequential procedure for testing overlapping models. In the first step, one would test whether $f(\theta_0) = g(\gamma_0)$. In our framework, as discussed in Section 4.1, this can be achieved in one of the two ways. The hypothesis $f(\theta_0) = g(\gamma_0)$ should be rejected when $nQLR_n\left(\hat{\theta}_n, \hat{\gamma}_n\right)$ exceeds the critical value of the mixed χ^2 distribution, say c_{α_1}, where α_1 denotes the significance level used in the first step. If it is not rejected, one concludes that $f(\theta_0) = g(\gamma_0)$ and the two models have the same lack-of-fit; however, it can be rejected either because $f(\theta_0) \neq g(\gamma_0)$, but the models have same lack of fit (H_0 is true), or because one of the models actually has a better fit (H_f or H_g are true).

If the null is not rejected in the first step, one continues to the second step, in which $H_0 : \|A(h_0 - f(\theta_0(A)))\| = \|A(h_0 - g(\gamma_0(A)))\|$ is rejected when $n^{1/2}QLR_n\left(\widehat{\theta}_n, \widehat{\gamma}_n\right)/\widehat{\omega}_n > z_{1-\alpha_2/2}$ (f is the preferred model) or $n^{1/2}QLR_n\left(\widehat{\theta}_n, \widehat{\gamma}_n\right)/\widehat{\omega}_n < -z_{1-\alpha_2/2}$ (g is the preferred model). However, if H_0 is not rejected in the second step as well, one concludes that the two models are equivalent. If α_1 and α_2 denote the significance levels in steps one and two respectively, the asymptotic significance level of the sequential procedure, as shown by Vuong (1989) is $\max(\alpha_1, \alpha_2)$.

5 Model comparison with estimation and evaluation on different sets of reduced-form parameters

In the calibration literature, model parameters are often estimated or calibrated using one set of reduced-form characteristics, while the model evaluation is conducted on another. For example, a structural model can be estimated to match first moments, and evaluated with respect to second moments. Such a situation is discussed in this section; it is analogous to out-of-sample model evaluation in the forecasting literature[4]; it also corresponds to the case of model comparison without lack-of-fit minimization in RV.

The conclusion of this section is that, when a model is estimated and evaluated on different sets of reduced-form parameters, the QLR statistic has asymptotic normal distribution

[4] See, for example, West and McCracken (1998)

regardless of whether f and g are nested or non-nested. The reason for this is that, even when the models are nested, a bigger model does not necessarily provides a better fit, since the deep parameters are not calibrated to minimize the distance between the truth and the part of the model used for evaluation. This conclusion is in the agreement with the results of Section 6 of RV.

Next, we introduce the notation and assumptions of this section. We partition $h_0 = \left(h_{1,0}', h_{2,0}'\right)'$, where $h_{1,0}$ is an m_1-vector, and $h_{2,0}$ is an m_2-vector, $m_1 + m_2 = m$. Similarly, we partition $\hat{h}_n = \left(\hat{h}_{1,n}', \hat{h}_{2,n}'\right)'$, $f(\theta) = \left(f_1(\theta)', f_2(\theta)'\right)'$, and $g(\gamma) = \left(g_1(\gamma)', g_2(\gamma)\right)'$. Next, consider the weight matrices A_1 and A_2, where A_i is $m_i \times m_i$, $i = 1, 2$. The weight matrix A_1 is used at the estimation stage, and the model parameters are estimated using only the first m_1 moment restrictions:

$$\begin{aligned} \hat{\theta}_n(A_{1,n}) &= \arg\min_{\theta\in\Theta} \left\| A_{1,n}\left(\hat{h}_{1,n} - f_1(\theta)\right)\right\|^2, \text{ and} \\ \hat{\gamma}_n(A_{1,n}) &= \arg\min_{\gamma\in\Gamma} \left\| A_{1,n}\left(\hat{h}_{1,n} - g_1(\gamma)\right)\right\|^2. \end{aligned}$$

At the evaluation stage, the models are compared using the weight matrix A_2:

$$H_0 : \|A_2(h_{2,0} - f_2(\theta_0(A_1)))\| = \|A_2(h_{2,0} - g_2(\gamma_0(A_1)))\|. \tag{5.1}$$

$$H_f : \|A_2(h_{2,0} - f_2(\theta_0(A_1)))\| < \|A_2(h_{2,0} - g_2(\gamma_0(A_1)))\|. \tag{5.2}$$

$$H_g : \|A_2(h_{2,0} - f_2(\theta_0(A_1)))\| > \|A_2(h_{2,0} - g_2(\gamma_0(A_1)))\|. \tag{5.3}$$

Note that the pseudo-true parameters are defined by A_1. Consequently, the first-order condition (2.2) does not hold for f_2, g_2, h_2, and A_2, since $\theta_0(A_1)$ and $\gamma_0(A_1)$ do not minimize the CMD criterion for the last m_2 moment conditions, as described in part (d) of the assumption below. Also note that for testing it is important that the models are misspecified with respect to the second set of reduced-form parameters h_2.

Assumption 5.1 **(a)** *f_2 and g_2 are misspecified according to Definition 2.1.*

(b) *$A_{1,n} \to_p A_1$, $A_{2,n} \to_p A_2$; A_1 and A_2 have full ranks.*

(c) *Assumptions 2.4 and 3.3 hold with $A = A_1$.*

(d) $\frac{\partial f_2(\theta_0(A_1))'}{\partial\theta} A_2' A_2 (h_{2,0} - f_2(\theta_0(A_1))) \neq 0$; $\frac{\partial g_2(\gamma_0(A_1))'}{\partial\gamma} A_2' A_2 (h_{2,0} - g_2(\gamma_0(A_1))) \neq 0$.

The QLR statistic is now defined as

$$\begin{aligned} & QLR_n\left(\hat{\theta}_n(A_{1,n}), \hat{\gamma}_n(A_{1,n}), A_{2,n}\right) \\ = & -\left\|A_{2,n}\left(\hat{h}_{2,n} - f_2\left(\hat{\theta}_n(A_{1,n})\right)\right)\right\|^2 + \left\|A_{2,n}\left(\hat{h}_{2,n} - g_2\left(\hat{\gamma}_n(A_{1,n})\right)\right)\right\|^2, \end{aligned}$$

and its behavior in the large samples under null and alternatives (5.1)-(5.3) in the case of fixed weight matrices is described in the following theorem. Define

$$\begin{aligned} J_{f,0} &= \left(-\frac{\partial f_2(\theta_0(A_1))}{\partial \theta'} F_{1,0}^{-1} \frac{\partial f_1(\theta_0(A_1))'}{\partial \theta} A_1' A_1 \quad I_{m_2} \right), \\ J_{g,0} &= \left(-\frac{\partial g_2(\gamma_0(A_1))}{\gamma'} G_{1,0}^{-1} \frac{\partial g_1(\gamma_0(A_1))'}{\partial \gamma} A_1' A_1 \quad I_{m_2} \right), \end{aligned}$$

where $F_{1,0}$ and $G_{1,0}$ are defined similarly F_0 and G_0 in (3.2) and (3.3) respectively, but using A_1, $h_{1,0}$, f_1, and g_1.

Theorem 5.1 *Suppose that Assumptions 2.1 and 5.1 hold, and $A_{1,n} = A_1, A_{2,n} = A_2$ for all n.*

(a) *Under H_0 in (5.1), $n^{1/2} QLR_n \left(\hat{\theta}_n(A_1), \hat{\gamma}_n(A_1), A_2 \right) \to_d N\left(0, \omega_{21,0}^2\right)$, where*

$$\omega_{21,0} = 2 \left\| \Lambda_0^{1/2} \left(J_{g,0}' A_2' A_2 \left(h_{2,0} - g_2(\gamma_0(A_1)) \right) - J_{f,0}' A_2' A_2 \left(h_{2,0} - f_2(\theta_0(A_1)) \right) \right) \right\|.$$

(b) *Under the alternative H_f in (5.2), $n^{1/2} QLR_n \left(\hat{\theta}_n(A_1), \hat{\gamma}_n(A_1), A_2 \right) \to \infty$ with probability one; under the alternative H_g in (5.3), $n^{1/2} QLR_n \left(\hat{\theta}_n(A_1), \hat{\gamma}_n(A_1), A_2 \right) \to -\infty$ with probability one.*

The QLR statistic is remains asymptotically normal when the models are nested, because in the asymptotic expansion of the CMD criterion function, now appears a dominating term of order $O_p\left(n^{-1/2}\right)$; this term is not present in the case of lack-of-fit minimization due to population first-order condition (2.2). Thus, when the models are estimated using one set of reduced-form parameters and evaluated using another, one follows the same testing procedure regardless of whether the models are nested, non-nested, or overlapping: reject the null of equivalent models when $n^{1/2} \left| QLR_n \left(\hat{\theta}_n(A_1), \hat{\gamma}_n(A_1), A_2 \right) \right| / 2\hat{\omega}_{21,n} > z_{1-\alpha/2}$, where $\hat{\omega}_{21,n}$ is a consistent estimator of $\omega_{21,0}$. A consistent estimator of $\omega_{21,0}$ can be obtained by the plug-in method, since all the elements of $\omega_{21,0}$ can be consistently estimated. Note that, when $f_2(\theta_0(A_1)) = g_2(\gamma_0(A_1))$, which can occur if the models are nested or overlapping, the columns corresponding to I_{m_2} in $J_{f,0}$ and $J_{g,0}$ do not contribute to the asymptotic variance; however, this will be reflected automatically by any consistent estimator $\hat{\omega}_{21,n}$.

6 Averaged and sup tests for model comparison

The choice of a weight matrix A plays a crucial role when the models are misspecified: the null hypothesis (2.3) changes when one considers a different weight matrix, and as a

result different weighing schemes can lead to different ranking of the models. One way to relax dependence on the choice of weight matrix to some extent, is to consider a procedure that takes into account the models' performance for various weighting schemes.

In this section, we consider again the situation when a model is estimated and evaluated on the same set of reduced-form parameters. Let $\mathbb{A}$ be a sub-space of $m \times m$ full-rank matrices, $\|A\| = tr\left(A'A\right)^{1/2}$, $\mathcal{B}\left(\mathbb{A}\right)$ be a σ-field generated by open subsets of $\mathbb{A}$, and π be a probability measure on $\mathcal{B}\left(\mathbb{A}\right)$. In this section, we discuss averaged and sup procedures that compare two misspecified models. We make the following assumption.

Assumption 6.1 **(a)** *$\mathbb{A}$ is compact.*

(b) *Assumption 2.4 holds for all $A \in \mathbb{A}$.*

The null hypothesis of the averaged procedure is stated as

$$H_0^a : \int_{\mathbb{A}} \left(\left\| A\left(h_0 - g\left(\gamma_0\left(A\right)\right)\right)\right\|^2 - \left\| A\left(h_0 - f\left(\theta_0\left(A\right)\right)\right)\right\|^2 \right) \pi\left(dA\right) = 0.$$

According to H_0^a, the two models f and g provide equivalent approximations to the truth h_0 on average, where the average is taken in the class $\mathbb{A}$ with respect to the probability measure π. For example, $\mathbb{A}$ may consist of a finite number of matrices A, and π assign equal weights to all A's. Note that the pseudo-true values $\theta_0\left(A\right)$ and $\gamma_0\left(A\right)$ continue to depend on A. The null hypothesis H_0^a will be tested against alternatives

$$\begin{aligned} H_f^a &: \int_{\mathbb{A}} \left(\left\| A\left(h_0 - g\left(\gamma_0\left(A\right)\right)\right)\right\|^2 - \left\| A\left(h_0 - f\left(\theta_0\left(A\right)\right)\right)\right\|^2 \right) \pi\left(dA\right) > 0, \text{ or} \\ H_g^a &: \int_{\mathbb{A}} \left(\left\| A\left(h_0 - g\left(\gamma_0\left(A\right)\right)\right)\right\|^2 - \left\| A\left(h_0 - f\left(\theta_0\left(A\right)\right)\right)\right\|^2 \right) \pi\left(dA\right) < 0. \end{aligned}$$

The null hypothesis of the sup procedure is given by

$$H_0^s : \sup_{A \in \mathbb{A}} \left(\left\| A\left(h_0 - g\left(\gamma_0\left(A\right)\right)\right)\right\|^2 - \left\| A\left(h_0 - f\left(\theta_0\left(A\right)\right)\right)\right\|^2 \right) \leq 0.$$

According to H_0^s, the model f cannot outperform the model g for any considered weight matrix $A \in \mathbb{A}$. Thus, H_0^s imposes a much stronger restriction than H_0^a. The null H_0^s will be tested against

$$H_f^s : \sup_{A \in \mathbb{A}} \left(\left\| A\left(h_0 - g\left(\gamma_0\left(A\right)\right)\right)\right\|^2 - \left\| A\left(h_0 - f\left(\theta_0\left(A\right)\right)\right)\right\|^2 \right) > 0,$$

which says that there is a weight matrix A such that the model f outperforms g.

Again, we consider the QLR statistic defined previously in (2.5), however, now it is

explicitly indexed by A:

$$QLR_n\left(\hat{\theta}_n(A), \hat{\gamma}_n(A), A\right) = \left\|A\left(\hat{h}_n - g\left(\hat{\gamma}_n(A)\right)\right)\right\|^2 - \left\|A\left(\hat{h}_n - f\left(\hat{\theta}_n(A)\right)\right)\right\|^2.$$

The averaged and sup statistics are given below:

$$\begin{aligned} AQLR_n &= \int_{\mathbb{A}} QLR_n\left(\hat{\theta}_n(A), \hat{\gamma}_n(A), A\right)\pi(dA), \\ SQLR_n &= \sup_{A\in\mathbb{A}} QLR_n\left(\hat{\theta}_n(A), \hat{\gamma}_n(A), A\right). \end{aligned}$$

The asymptotic null distributions and ranking of the models according to $AQLR_n$ or $SQLR_n$ depend on the choice of measure π. The asymptotic null distributions of the averaged and sup statistics depend on whether f and g are nested or non-nested, since the asymptotic distribution of QLR_n for any fixed A depends on the relationship between the models. Note that when the models are nested, $\mathcal{G} \subset \mathcal{F}$, the model g cannot outperform the model f, and the inequality in H_0^s holds as an equality. When the models are nested, we have the following result.

Theorem 6.1 *Suppose that Assumptions 2.1, 2.3, 3.1-3.3, 6.1 hold, and $\mathcal{G} \subset \mathcal{F}$. Let $Z \sim N(0, I_m)$, and, for a given $A \in \mathbb{A}$, define the matrices $W_{f,0}(A)$, $W_{g,0}(A)$ as $W_{f,0}$, $W_{g,0}$ in Theorem 4.1.*

(a) *Under H_0^a, $nAQLR_n \to_d Z'\Lambda_0^{'1/2}\left(\int A'A\left(W_{g,0}(A) - W_{f,0}(A)\right)A'A\pi(dA)\right)\Lambda_0^{1/2}Z$. Under H_f^a, $nAQLR_n \to \infty$ with probability one; under H_g^a, $nAQLR_n \to -\infty$ with probability one.*

(b) *Under H_0^s, $nSQLR_n \to_d \sup_{A\in\mathbb{A}}\left(Z'\Lambda_0^{'1/2}A'A\left(W_{g,0}(A) - W_{f,0}(A)\right)A'A\Lambda_0^{1/2}Z\right)$. Under H_f^s, $nSQLR_n \to \infty$ with probability one.*

According to the above theorem, the asymptotic distribution of the averaged statistic in the case of nested models remains mixed χ^2. However, the weights are now given by the average of matrices $W_{f,0}$ and $W_{g,0}$. Note that $W_{f,0}$, $W_{g,0}$, and the matrices they composed of, $\partial f(\theta_0(A))/\partial\theta'$, F_0, $M_{f,0}$, and etc., all depend on A. Since $W_{f,0}(A)$ and $W_{g,0}(A)$ can be estimated consistently by the plug-in method, the critical values of the mixed χ^2.distribution can be computed by simulations as described in Section 4.1. The asymptotic null distribution of the sup statistic depends on the sup transformation of the mixed χ^2 distribution. The critical values can be obtained by simulations as well.

In the case of non-nested models, the asymptotic null distribution is a functional of Gaussian process. Note that when the models are non-nested, H_0^s does not determine uniquely the null distribution; it is a composite hypothesis, and the distribution depends

on whether the restriction is binding. However, the least favorable alternative, as usual, corresponds to the case when the restriction is binding.

Theorem 6.2 *Suppose that Assumptions 2.1, 2.3, 3.1-3.3, 6.1 hold, and* $\mathcal{F} \cap \mathcal{G} = \emptyset$. *Let* $\{X(A) \in R : A \in \mathbb{A}\}$ *be a mean zero Gaussian process such that the covariance of* $X(A_1)$ *and* $X(A_2)$, $A_1, A_2 \in \mathbb{A}$, *is* $\omega_0(A_1, A_2)$, *where*

$$\omega_0(A_1, A_2) = 4\left(f(\theta_0(A_1)) - g(\gamma_0(A_1))\right)' A_1' A_1 \Lambda_0 A_2' A_2 \left(f(\theta_0(A_2)) - g(\gamma_0(A_2))\right).$$

(a) *Under* H_0^a, $n^{1/2} AQLR_n \to_d N\left(0, \int_{\mathbb{A}} \int_{\mathbb{A}} \omega_0(A_1, A_2)\, \pi(dA_1)\, \pi(dA_2)\right)$. *Under* H_f^a, $n^{1/2} AQLR_n \to \infty$ *with probability one; under* H_g^a, $n^{1/2} AQLR_n \to -\infty$ *with probability one.*

(b) *Under* H_0^s, $\lim_{n\to\infty} P\left(n^{1/2} SQLR_n > c\right) \leq P\left(\sup_{A\in\mathbb{A}} X(A) > c\right)$. *Under* H_f^s, $n^{1/2} SQLR_n \to \infty$ *with probability one.*

According to Theorem 6.2, the averaged statistic has a normal distribution. The variance is given by averaged variances and covariances of the QLR statistics for different A's; it can be estimated consistently by the plug-in method. For the sup statistic, the asymptotic distribution is that of the sup of the Gaussian process, and the critical values for a test based on $SQLR_n$ can be obtained by simulations.

In the case of overlapping models, one use a sequential procedure similar to the one discussed in Section 4.3.

7 Application

In this section we applied our proposed test to two monetary macroeconomic models, the CIA and PAC models. The setups here follow Schorfheide (2000) and Christiano and Eichenbaum (1992). The vector h, as defined in Section 2, is the impulse responses of output and inflation to a monetary growth shock. It is assumed that h can be estimated consistently by a Structural Vector Auto-Regression (SVAR) model of GDP and inflation. The identification scheme employed for the SVAR follows Blanchard and Quah (1989). The particular restriction applied to identify the SVAR model is that money is neutral in the long run, which is satisfied by both the CIA and PAC models.

7.1 CIA Model

The production function in this economy is $Y_t = K_t^{\alpha}(\mathcal{A}_t N_t)^{1-\alpha}$, where K_t, N_t, and $\mathcal{A}_t$ are capital stock, labour, and technology respectively. Technology $\mathcal{A}_t$ and money growth rate

$m_t = M_{t+1}/M_t$ follow stochastic processes

$$\begin{aligned}
\ln \mathcal{A}_t &= \psi + \ln \mathcal{A}_{t-1} + \epsilon_{\mathcal{A},t}, \\
\ln m_t &= (1-\rho)\ln m_{ss} + \rho \ln m_{t-1} + \epsilon_{M,t}, \\
\epsilon_{\mathcal{A},t} &\sim N\left(0, \sigma^2_{\mathcal{A}}\right), \\
\epsilon_{M,t} &\sim N\left(0, \sigma^2_{M}\right),
\end{aligned}$$

where m_{ss} is the steady state inflation rate. The representative household's problem is

$$\begin{aligned}
&\max_{\{C_t, H_t, M_{t+1}, D_t\}} E_0\left[\sum_{t=0}^{\infty} \beta^t \left[(1-\phi)\ln C_t + \phi \ln(1-H_t)\right]\right] \\
s.t. \quad & P_t C_t \leq M_t - D_t + W_t H_t, \\
& 0 \leq D_t, \\
& M_{t+1} = (M_t - D_t + W_t H_t - P_t C_t) + R_{H,t} D_t + F_t + B_t,
\end{aligned}$$

where E_t denotes expectation at date t, β the discount factor, and ϕ the share of leisure in per period utility; $C_t, H_t, P_t, M_t, D_t, W_t, R_{H,t}, F_t$, and B_t denote real household consumption, labour supply, price, nominal money holding, nominal money deposit, nominal wage rate, nominal return on deposit, nominal profit paid by the firm, and nominal profit paid by the bank respectively. The representative firm's problem is

$$\begin{aligned}
&\max_{F_t, K_{t+1}, N_t, L_t} E_t\left[\sum_{t=0}^{\infty} \beta^{t+1} \frac{F_t}{C_{t+1} P_{t+1}}\right] \\
s.t. \quad & F_t \leq L_t + P_t\left[K_t^{\alpha}(\mathcal{A}_t N_t)^{1-\alpha} - K_{t+1} + (1-\delta)K_t\right] - W_t N_t - L_t R_{F,t}, \\
& W_t N_t \leq L_t.
\end{aligned}$$

There is a financial intermediary in the economy and its objective is simple. At the beginning of each period, it loans out the household's deposit D_t and the money injection X_t from the central bank to the firm. At the end of period, it collects the loan plus interest $L_t R_{F,t}$ and pays the amount to the household, who is the owner. The household, firm and financial intermediary all take prices as given.

Given the first order conditions, we detrend the variables and solve the model by log-linearizing the equilibrium conditions. The real variables, prices and nominal variables are detrended respectively by technology level $\mathcal{A}_t$, $M_t/\mathcal{A}_t$, and M_t. The state space representation of the model is

$$\hat{y}_{t+1} = \psi_1 \hat{a}_t + \psi_2 \hat{\mu}_t + \psi_3 \hat{y}_t + \psi_4 \epsilon_{\mathcal{A},t+1} + \psi_5 \epsilon_{M,t+1},$$

$$
\begin{aligned}
\hat{a}_{t+1} &= \hat{a}_t + \epsilon_{\mathcal{A},t+1}, \\
\hat{\mu}_{t+1} &= \rho\hat{\mu}_t + \epsilon_{M,t+1},
\end{aligned}
$$

where $\hat{y}_t, \hat{a}_t, \hat{\mu}_t$ are the log deviations of $Y_t, \mathcal{A}_t$, and m_t from steady state. For small deviations, they can be interpreted as percentage deviations. The coefficients $\psi_1, \psi_2, \psi_3, \psi_4$, and ψ_4 are functions of model parameters $\mathcal{A}_0, m_{ss}, \alpha, \beta, \delta, \psi, \phi$, and ρ. Given one set of parameter values, we can iterate the state space representation to calculate the theoretical impulse responses of the model. In the minimum distance estimation, we look for the set of $\mathcal{A}_0, m_{ss}, \alpha, \beta, \delta, \psi, \phi$, and ρ values that minimize the distance between theoretical impulse responses and impulse responses generated by the data.

7.2 PAC Model

The production function and stochastic processes governing technology and money growth are the same as in the CIA model. The household's problem is

$$
\begin{aligned}
& \max_{\{C_t, H_t, M_{t+1}, Q_{t+1}\}} E_0 \left[\sum_{t=0}^{\infty} \beta^t [(1-\phi)\ln C_t + \phi \ln(1 - H_t - G_t)] \right] \\
s.t. \quad & P_t C_t \leq Q_t + W_t H_t, \\
& Q_t \leq M_t, \\
& M_{t+1} = Q_t + W_t H_t - P_t C_t + R_{H,t}(M_t - Q_t) + F_t + B_t,
\end{aligned}
$$

where Q_t is defined as $M_t - D_t$ for convenience in derivation. The main difference from the CIA model is the introduction of an adjustment cost G_t given by

$$
G_t = \alpha_1 \left[\exp\left(\alpha_2 (\frac{Q_t}{Q_{t-1}} - m_{ss}) \right) + \exp\left(-\alpha_2 (\frac{Q_t}{Q_{t-1}} - m_{ss}) \right) - 2 \right].
$$

The firm's problem and the financial intermediary's problem are identical to the CIA model. This model is also solved by log-linearization to obtain the state-space representation

$$
\begin{aligned}
\hat{y}_{t+1} &= \pi_1 \hat{a}_{t+1} + \pi_2 \hat{a}_t + \pi_3 \hat{\mu}_{t+1} + \pi_4 \hat{\mu}_t + \pi_5 \hat{\mu}_{t-1} + \pi_6 \hat{q}_{t+1} + \pi_7 \hat{q}_t + \pi_8 \hat{y}_t, \\
\hat{q}_{t+1} &= -\beta\hat{\mu}_t + f_{22}\hat{\mu}_{t-1} - \hat{\mu}_{t-2} + (1+2\beta)\hat{q}_t - (2+\beta)\hat{q}_{t-1} + \hat{q}_{t-2}, \\
\hat{a}_{t+1} &= \hat{a}_t + \epsilon_{\mathcal{A},t+1}, \\
\hat{\mu}_{t+1} &= \rho\hat{\mu}_t + \epsilon_{M,t+1},
\end{aligned}
$$

where $\hat{q}_t$ is the log deviation of Q_t from steady state; the coefficients $\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6, \pi_7$, and π_8 are functions of model parameters $\mathcal{A}_0, m_{ss}, \alpha, \alpha_1, \alpha_2, \beta, \delta, \psi, \phi$, and ρ. Again for a given set of parameter values, we can employ the state space representation to calculate the

theoretical impulse responses of the model.

7.3 Model Estimation and Comparison Results

The CIA and PAC models both provide predictions for evolution for multiple time series. In this application we focus on the growth rates of GDP per capita and price. The h vector we choose is 40 by 1, consisting of 20 period output growth impulse responses and 20 period inflation impulse responses to a shock to inflation rate. The data used here are the US GDP per capita growth rate and inflation rate available from the Basic Economics database produced by DRI/McGraw-Hill. The GDP per capita series are obtained by dividing the GDP series (*GDP215* in the DRI database) by the population (*POP* in the DRI database). For the price level, we choose the GDP deflator series (*GDPD15* in the DRI database). Then we take logs and first difference of the series to get GDP per capita growth rate and inflation rate. The range of the series covers the second quarter of 1947 to the third quarter of 2003.

We estimate a SVAR model on the GDP per capita growth and inflation series and obtain estimates for 20 period impulse responses of GDP per capita and inflation to a one-unit money growth shock. Then we look for the set of parameter values of θ, where $\theta = (\mathcal{A}_0, m_{ss}, \alpha, \beta, \delta, \psi, \phi, \rho)$, that would best match the theoretical impulse responses in CIA model $f(\theta)$ with the estimated impulse responses $\hat{h}_n$. Similarly, we look for the set of parameter values of γ where $\gamma = (\mathcal{A}_0, m_{ss}, \alpha, \alpha_1, \alpha_2, \beta, \delta, \psi, \phi, \rho)$, to minimize the distance between the theoretical impulse responses in PAC model $g(\gamma)$ and the estimated impulse responses $\hat{h}_n$.

In our calibration exercise we fix the values of capital share α and quarterly discount factor β, which significantly reduces computation times. Following Christiano and Eichenbaum (1992) the values are set to be 0.36 and 0.9926, which are similar to those used by macroeconomists in practice. The ranges of parameters for ρ, ϕ, and δ in CIA and PAC model are specified as in Schorfheide (2000), while the ranges of ψ, m, α_1, and α_2 differ. While Schorfheide (2000) allows the steady sate GDP growth rate ψ and money growth rate m to take any real value, we restrict them to economically plausible ranges of $[0, 0.2]$ and $[0.5, 1.5]$. This is because our procedure requires that the vectors of model parameters come from compact sets. Drawing on Christiano and Eichenbaum (1992) which sets $\alpha_1 = 0.00005$ and $\alpha_2 = 1000$ in their calibration, we set the upper bound of these parameters to be ten times of these values. Compared to Schorfheide (2000), estimation of our model also involves the initial technology level $\mathcal{A}_0$. Using a Cobb-Douglas production function with capital share of 0.36, the imputed initial technology level for US in the first quarter of 1947 is 0.07. We set the upper bound to be ten times this value, and thus the range of $[0, 0.7]$. Table 1 summarizes the range of parameters and their estimates with the

weighting matrix being an identity ($A_n = A = I$).

Figures 1 and 2 plot impulse responses of inflation and output implied by the two structural models together with the SVAR impulse responses. From the graphs we can see, while the CIA model can approximate the inflation response to money growth shock well it does poorly in replicating the output impulses. This is because in the CIA model the household can always rebalance its money holding to nullify the real effect of money growth shock. In comparison, the PAC model generates a positive output response to money shock although the magnitude is smaller than that of SVAR. From the eye-ball comparison, we can see PAC model provides a better fit. The better fit provided by the PAC model is not surprising though, since it nests the CIA model. If $\alpha_2 = 0$, i.e. there is no adjustment cost in the PAC model, then it is reduced to the CIA model.

In order to determine whether the PAC superior performance in terms of approximating the impulse responses is statistically significant, we compute the test statistic proposed in Section 4.1. The key ingredients are the parameter estimates $\hat{\theta}_n$ and $\hat{\gamma}_n$, the impulse responses implied by the SVAR and the two structural models, and the partial derivatives of theoretical impulse responses with respect to the model parameters. To compute the QLR_n statistic, we obtained analytical expressions for the partial derivatives of f and g, and evaluated them at the estimated values of the deep parameters.

The value of the QLR_n statistic obtained is 0.204. Since the distribution of the test statistic has a mixed χ^2_{40} distribution, we simulate its critical values. The 5% and 1% critical values are respectively 0.040 and 0.165, both smaller than 0.204. Therefore we reject that the CIA model fits the data as well as the PAC model, and conclude that the PAC model provides a statistically better fit.

8 Appendix

Proof of Theorem 3.1. For consistency of $\hat{\theta}_n$, it is sufficient to show uniform convergence of $\left\|A_n\left(\hat{h}_n - f(\theta)\right)\right\|^2$ to $\|A(h_0 - f(\theta))\|^2$ on Θ. The desired result will follow from Assumptions 2.4 and 3.1 by the usual argument for extremum estimators (see, for example, Theorem 2.1 in Newey and McFadden (1994)).

$$\begin{aligned}
\left\|A_n\left(\hat{h}_n - f(\theta)\right)\right\|^2 - \|A(h_0 - f(\theta))\|^2 &= R_{1,n} - 2R_{2,n}(\theta) + R_{3,n}(\theta), \text{ where} \\
R_{1,n} &= \hat{h}_n A_n' A_n \hat{h}_n - h_0' A_n' A_n h_0, \\
R_{2,n}(\theta) &= \left(\hat{h}_n - h_0\right)' A_n' A_n f(\theta) \\
R_{3,n}(\theta) &= (h_0 - f(\theta))' \left(A_n' A_n - A' A\right)(h_0 - f(\theta)).
\end{aligned}$$

By Assumption 2.1(a) and 2.2, $|R_{1,n}| \to_p 0$. Let $\|A\| = tr\left(A'A\right)^{1/2}$. Due to Assumption 3.1 (a) and (c), f is bounded on Θ (Davidson, 1994, Theorem 2.19), and, therefore,

$$\begin{aligned}\sup_{\theta\in\Theta}\left|R_{2,n}\left(\theta\right)\right| &\leq \left\|A_n\right\|^2 \left\|\left(\hat{h}_n - h_0\right)\right\| \sup_{\theta\in\Theta}\left\|f\left(\theta\right)\right\| \\ &\to_p 0,\end{aligned}$$

by Assumptions 2.1(a) and 2.2.

$$\begin{aligned}\sup_{\theta\in\Theta}\left|R_{3,n}\left(\theta\right)\right| &\leq \left\|A_n'A_n - A'A\right\| \sup_{\theta\in\Theta}\left\|h_0 - f\left(\theta\right)\right\|^2 \\ &\leq \left\|A_n'A_n - A'A\right\| \left(\left\|h_0\right\| + \sup_{\theta\in\Theta}\left\|f\left(\theta\right)\right\|\right)^2 \\ &\to_p 0.\end{aligned}$$

The proof of $\hat{\gamma}_n \to_p \gamma_0$ is identical with f and θ replaced by g and γ. □

Proof of (3.1). First, applying the mean value expansion to $f\left(\hat{\theta}_n\right)$,

$$\begin{aligned}0 &= \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial\theta} A_n'A_n\left(\hat{h}_n - f\left(\hat{\theta}_n\right)\right) \\ &= \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial\theta} A_n'A_n\left(\hat{h}_n - f\left(\theta_0\right) - \frac{\partial f\left(\widetilde{\theta}_n\right)}{\partial\theta'}\left(\hat{\theta}_n - \theta_0\right)\right) \\ &= \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial\theta}\left(A_n'A_n\left(\hat{h}_n - h_0\right) + \left(A_n'A_n - A'A\right)\left(h_0 - f\left(\theta_0\right)\right)\right) \\ &\quad + \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial\theta} A'A\left(h_0 - f\left(\theta_0\right)\right) - \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial\theta} A_n'A_n \frac{\partial f\left(\widetilde{\theta}_n\right)}{\partial\theta'}\left(\hat{\theta}_n - \theta_0\right),\end{aligned}$$

where $\widetilde{\theta}_n$ is the mean value. Next,

$$\begin{aligned}&\frac{\partial f\left(\hat{\theta}_n\right)'}{\partial\theta} A'A\left(h_0 - f\left(\theta_0\right)\right) \\ =\ &\left(I_k \otimes \left(h_0 - f\left(\theta_0\right)\right)' A'A\right) Vec\left(\frac{\partial f\left(\hat{\theta}_n\right)}{\partial\theta'}\right) \\ =\ &\frac{\partial f\left(\theta_0\right)'}{\partial\theta} A'A\left(h_0 - f\left(\theta_0\right)\right) \\ &+ \left(I_k \otimes \left(h_0 - f\left(\theta_0\right)\right)' A'A\right) \frac{\partial}{\partial\theta'} Vec\left(\frac{\partial f\left(\overline{\theta}_n\right)}{\partial\theta'}\right)\left(\hat{\theta}_n - \theta_0\right)\end{aligned}$$

$$= \quad M_{f,n}\left(\hat{\theta}_n - \theta_0\right), \tag{8.1}$$

where $\overline{\theta}_n$ is the mean value; note that the last equality follows from the population first-order condition (2.2). □

Proof of Theorem 3.2. One can expand the first order conditions for $\hat{\gamma}_n$ similarly to that of $\hat{\theta}_n$, equation (3.1). Taking into account that $A_n = A$ for all n,

$$n^{1/2}\begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\gamma}_n - \gamma_0 \end{pmatrix} = \begin{pmatrix} F_n^{-1}\frac{\partial f(\hat{\theta}_n)'}{\partial \theta} \\ G_n^{-1}\frac{\partial g(\hat{\gamma}_n)'}{\partial \gamma} \end{pmatrix} A'An^{1/2}\left(\hat{h}_n - h_0\right),$$

where

$$\begin{aligned} G_n &= \frac{\partial g\left(\hat{\gamma}_n\right)'}{\partial \gamma} A_n' A_n \frac{\partial g\left(\widetilde{\gamma}_n\right)}{\partial \gamma'} - M_{g,n}, \\ M_{g,n} &= \left(I_l \otimes \left(h_0 - g\left(\gamma_0\right)\right)' A'A\right) \frac{\partial}{\partial \gamma'} vec\left(\frac{\partial g\left(\overline{\gamma}_n\right)}{\partial \gamma'}\right), \end{aligned}$$

and $\widetilde{\gamma}_n, \overline{\gamma}_n$ are between $\hat{\gamma}_n$ and γ_0. The result follows from Theorem 3.1, Assumptions 2.1(b) and 3.3. □

Proof of Theorem 3.3. The result follows immediately from (3.1), a similar expansion for $\hat{\gamma}_n$, and the assumptions of the theorem by writing

$$\begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\gamma}_n - \gamma_0 \end{pmatrix} = \begin{pmatrix} F_n^{-1}\frac{\partial f(\hat{\theta}_n)'}{\partial \theta} \begin{pmatrix} A_n' A_n & I_m \otimes \left(h_0 - f\left(\theta_0\right)\right)' \end{pmatrix} \\ G_n^{-1}\frac{\partial g(\hat{\gamma}_n)'}{\partial \gamma} \begin{pmatrix} A_n' A_n & I_m \otimes \left(h_0 - g\left(\gamma_0\right)\right)' \end{pmatrix} \end{pmatrix} \begin{pmatrix} \hat{h}_n - h_0 \\ vec\left(A_n' A_n - A'A\right) \end{pmatrix}.$$

□

Proof of Theorem 4.1. In the case of the fixed weight matrix, using (3.1) the following expansion is obtained.

$$\begin{aligned} & \left\| A\left(\hat{h}_n - f\left(\hat{\theta}_n\right)\right)\right\|^2 \\ = \quad & \left\| A\left(\hat{h}_n - f\left(\theta_0\right)\right)\right\|^2 + \left(\hat{h}_n - h_0\right)' A'AW_{f,n}A'A\left(\hat{h}_n - h_0\right) + o_p\left(n^{-1}\right), \end{aligned} \tag{8.2}$$

where

$$\begin{aligned} W_{f,n} &= W_{f,n}(1) - W_{f,n}(2) - W_{f,n}(3), \\ W_{f,n}(1) &= \frac{\partial f\left(\hat{\theta}_n\right)}{\partial \theta'} F_n'^{-1} \frac{\partial f\left(\widetilde{\theta}_n\right)}{\partial \theta}' A'A \frac{\partial f\left(\widetilde{\theta}_n\right)}{\partial \theta'} F_n^{-1} \frac{\partial f\left(\hat{\theta}_n\right)}{\partial \theta}', \end{aligned}$$

$$
\begin{aligned}
W_{f,n}(2) &= \frac{\partial f\left(\hat{\theta}_n\right)}{\partial \theta'} F_n^{\prime -1} \frac{\partial f\left(\widetilde{\theta}_n\right)'}{\partial \theta} + \frac{\partial f\left(\widetilde{\theta}_n\right)}{\partial \theta'} F_n^{-1} \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta}, \\
W_{f,n}(3) &= \frac{\partial f\left(\hat{\theta}_n\right)}{\partial \theta'} F_n^{\prime -1} M_{f,n}' + M_{f,n} F_n^{-1} \frac{\partial f\left(\hat{\theta}_n\right)'}{\partial \theta}.
\end{aligned}
$$

To show (8.2), write

$$
\left\| A\left(\hat{h}_n - f\left(\hat{\theta}_n\right)\right)\right\|^2 = \left\| A\left(\hat{h}_n - f\left(\theta_0\right)\right)\right\|^2 + S_{1,n} + S_{2,n} + S_{3,n}, \tag{8.3}
$$

where

$$
\begin{aligned}
S_{1,n} &= \left(f\left(\hat{\theta}_n\right) - f\left(\theta_0\right)\right)' A'A \left(f\left(\hat{\theta}_n\right) - f\left(\theta_0\right)\right), \\
S_{2,n} &= -2\left(\hat{h}_n - h_0\right)' A'A \left(f\left(\hat{\theta}_n\right) - f\left(\theta_0\right)\right), \\
S_{3,n} &= -2\left(h_0 - f\left(\theta_0\right)\right)' A'A \left(f\left(\hat{\theta}_n\right) - f\left(\theta_0\right)\right).
\end{aligned}
$$

Now, one obtains (8.2) by expanding $f\left(\hat{\theta}_n\right)$ in $S_{1,n}$, $S_{2,n}$, and $S_{3,n}$ around $f\left(\theta_0\right)$ and using (3.1); in the case of $S_{3,n}$, after expanding $f\left(\hat{\theta}_n\right)$, one can apply the result in (8.1) to $\left(h_0 - f\left(\theta_0\right)\right)' A'A \left(\partial f\left(\widetilde{\theta}_n\right)/\partial \theta'\right)$, which leads to $M_{f,n}$ in the expression for $W_{f,n}(3)$.

An expansion similar to (8.2) is available for $\left\| A\left(\hat{h}_n - g\left(\hat{\gamma}_n\right)\right)\right\|^2$ with f, θ, and F replaced by g, γ, and G. Hence,

$$
\begin{aligned}
QLR_n\left(\hat{\theta}_n, \hat{\gamma}_n\right) &= -\left\| A\left(\hat{h}_n - f\left(\theta_0\right)\right)\right\|^2 + \left\| A\left(\hat{h}_n - g\left(\gamma_0\right)\right)\right\|^2 \\
&\quad + \left(\hat{h}_n - h_0\right)' A'A \left(W_{g,n} - W_{f,n}\right) A'A \left(\hat{h}_n - h_0\right).
\end{aligned} \tag{8.4}
$$

Under the null, the first summand on the right-hand side of (8.4) is zero by Lemma 4.1, and, due to Assumption 2.1(c) and Theorem 3.2, the second summand, when multiplied by n, converges in distribution to the random variable defined in part (a) of the theorem.

Since under H_f, $\left\|A\left(h_0 - f\left(\theta_0\right)\right)\right\|^2 \leq \left\|A\left(h_0 - g\left(\gamma_0\right)\right)\right\|^2$, part (b) of the theorem follows. □

Proof of Theorem 4.2. From (8.2) we obtain

$$
\left\| A\left(\hat{h}_n - f\left(\hat{\theta}_n\right)\right)\right\|^2 = \left\|A\left(h_0 - f\left(\theta_0\right)\right)\right\|^2 + 2\left(h_0 - f\left(\theta_0\right)\right)' A'A \left(\hat{h}_n - h_0\right) + O_p\left(n^{-1}\right),
$$

with a similar expression for $\left\|A\left(\hat{h}_n - g\left(\hat{\gamma}_n\right)\right)\right\|^2$. Hence,

$$\begin{aligned} QLR_n\left(\hat{\theta}_n, \hat{\gamma}_n\right) &= -\left\|A\left(h_0 - f\left(\theta_0\right)\right)\right\|^2 + \left\|A\left(h_0 - g\left(\gamma_0\right)\right)\right\|^2 \\ &\quad +2\left(f\left(\theta_0\right) - g\left(\gamma_0\right)\right)' A'A\left(\hat{h}_n - h_0\right) + O_p\left(n^{-1}\right). \end{aligned} \tag{8.5}$$

Since $\mathcal{F} \cap \mathcal{G} = \emptyset$, we have that $f\left(\theta_0\right) \neq g\left(\gamma_0\right)$, and the result follows from Assumption 2.1(b). □

Proof of Theorem 5.1. From (8.3), $\left\|A_2\left(\hat{h}_{2,n} - f_2\left(\hat{\theta}_n\left(A_1\right)\right)\right)\right\|^2$ can be expanded as

$$\begin{aligned} & \left\|A_2\left(h_{2,0} - f_2\left(\theta_0\left(A_1\right)\right)\right)\right\|^2 + 2\left(h_{2,0} - f_2\left(\theta_0\left(A_1\right)\right)\right)' A_2' A_2\left(\hat{h}_{2,n} - h_{2,0}\right) \\ & -2\left(h_{2,0} - f_2\left(\theta_0\left(A_1\right)\right)\right)' A_2' A_2 \frac{\partial f_2\left(\theta_0\left(A_1\right)\right)}{\partial \theta'}\left(\hat{\theta}_n\left(A_1\right) - \theta_0\left(A_1\right)\right) + o_p\left(n^{-1/2}\right) \\ = & \left\|A_2\left(h_{2,0} - f_2\left(\theta_0\left(A_1\right)\right)\right)\right\|^2 \\ & -2\left(h_{2,0} - f_2\left(\theta_0\left(A_1\right)\right)\right)' A_2' A_2 \frac{\partial f_2\left(\theta_0\left(A_1\right)\right)}{\partial \theta'} F_{1,0}^{-1} \frac{\partial f_1\left(\theta_0\left(A_1\right)\right)'}{\partial \theta} A_1' A_1\left(\hat{h}_{1,n} - h_{1,0}\right) \\ & +2\left(h_{2,0} - f_2\left(\theta_0\left(A_1\right)\right)\right)' A_2' A_2\left(\hat{h}_{2,n} - h_{2,0}\right) \\ & +o_p\left(n^{-1/2}\right) \\ = & \left\|A_2\left(h_{2,0} - f_2\left(\theta_0\left(A_1\right)\right)\right)\right\|^2 + 2\left(h_{2,0} - f_2\left(\theta_0\left(A_1\right)\right)\right)' A_2' A_2 J_{f,0}\left(\hat{h}_n - h_0\right) + o_p\left(n^{-1/2}\right). \end{aligned}$$

□

Proof of Theorem 6.1. First, note that in the case of nested models for all $A \in \mathbb{A}$, $\left\|A\left(h_0 - g\left(\gamma_0\left(A\right)\right)\right)\right\|^2 \geq \left\|A\left(h_0 - f\left(\theta_0\left(A\right)\right)\right)\right\|^2$, and thus, under H_0^a, we have that for all $A \in \mathbb{A}$, $\left\|A\left(h_0 - g\left(\gamma_0\left(A\right)\right)\right)\right\|^2 = \left\|A\left(h_0 - f\left(\theta_0\left(A\right)\right)\right)\right\|^2$.

We show next that under H_0^a, $nQLR_n\left(\hat{\theta}_n\left(A\right), \hat{\gamma}_n\left(A\right), A\right)$ converges weakly to a stochastic process indexed by A. According to Theorem (10.2) of Pollard (1990), for weak convergence one needs to show finite dimensional convergence and stochastic equicontinuity of $nQLR_n\left(\hat{\theta}_n\left(A\right), \hat{\gamma}_n\left(A\right), A\right)$ with respect to A.

Finite dimensional convergence follows by the same arguments as in the proof of Theorem 4.1. For stochastic equicontinuity, from (8.2) one can show that

$$\begin{aligned} & n\left|QLR_n\left(\hat{\theta}_n\left(A_1\right), \hat{\gamma}_n\left(A_1\right), A_1\right) - QLR_n\left(\hat{\theta}_n\left(A_2\right), \hat{\gamma}_n\left(A_2\right), A_2\right)\right| \\ \leq & n\left\|\hat{h}_n - h_0\right\|^2 K_n \left\|A_1 - A_2\right\|^{\delta} + o_p\left(1\right), \end{aligned}$$

where $\delta > 0$, $K_n = O_p\left(1\right)$ and independent of $\left(A_1 - A_2\right)$, and $o_p\left(1\right)$ term is uniform in A; this is because $W_{f,n}$ and $W_{g,n}$ are continuous in A, and $o_p\left(n^{-1}\right)$ term is uniform in

A. Stochastic equicontinuity of $nQLR_n\left(\hat{\theta}_n(A), \hat{\gamma}_n(A), A\right)$ follows from Lemma 2(a) of Andrews (1992).

The results of the theorem follow now from weak convergence by the continuous mapping theorem (CMT). □

Proof of Theorem 6.2. Convergence of finite dimensional distributions and stochastic equicontinuity can be established from (8.5). The results of the theorem will follow by the CMT. □

References

Andrews, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8(2), 241–257.

Blanchard, O. J., and D. Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79(4), 655–73.

Canova, F. (1994): "Statistical Inference in Calibrated Models," *Journal of Applied Econometrics*, 9, 123–144.

Canova, F., and E. Ortega (1996): "Testing Calibrated General Equilibrium Models," Working Paper 166, Universitat Pompeu Fabra.

Christiano, L. J., and M. Eichenbaum (1992): "Liquidity Effects and the Monetary Transmission Mechanism," *American Economic Review*, 82(2), 346–53.

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

Cooley, T. F., and E. C. Prescott (1995): "Economic Growth and Business Cycles," in *Frontiers of Business Cycle Research*, ed. by T. F. Cooley, pp. 1–38. Princeton University Press.

Corradi, V., and N. R. Swanson (2003): "Evaluation of Dynamic Stochastic General Equilibrium Models Based on Distributional Comparison of Simulated and Historic Data," *Journal of Econometrics*, 127(2), 699–723.

Davidson, J. (1994): *Stochastic Limit Theory*. Oxford University Press, New York.

Diebold, F. X., and R. S. Mariano (1995): "Comparing predictive accuracy," *Journal of Business & Economic Statistics*, 13(3), 253–263.

Diebold, F. X., L. E. Ohanian, and J. Berkowitz (1998): "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data," *Review of Economic Studies*, 65(3), 433–51.

Dridi, R., A. Guay, and E. Renault (2007): "Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models," *Journal of Econometrics*, 136(2), 397–430.

Fuerst, T. S. (1992): "Liquidity, Loanable Funds, and Real Activity," *Journal of Monetary Economics*, 29(1), 3–24.

Gouriéroux, C., A. Monfort, and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S118.

Gregory, A. W., and G. W. Smith (1990): "Calibration as Estimation," *Econometric Reviews*, 9(1), 57–89.

——— (1991): "Calibration as Testing: Inference in Simulated Macroeconomic Models," *Journal of Business & Economic Statistics*, 9(3), 297–303.

Gregory, A. W., and G. W. Smith (1993): "Statistical Aspects of Calibration in Macroeconomics," in *Handbook of Statistics*, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod, vol. 11, chap. 25, pp. 703–719. North-Holland, Amsterdam.

Hall, A., A. Inoue, J. M. Nason, and B. Rossi (2007): "Information Criteria For Impulse Response Function Matching Estimation of DSGE Models," Federal Reserve Bank of Atlanta Working Paper 2007-10.

Hall, A. R., and A. Inoue (2003): "The Large Sample Behavior of the Generalized Method of Moments Estimator in Misspecified Models," *Journal of Econometrics*, 114, 361–394.

Hall, A. R., and D. Pelletier (2007): "Non-Nested Testing in Models Estimated via Generalized Method of Moments," Working Paper 011, North Carolina State University.

Kim, K., and A. Pagan (1995): "The Econometric Analysis of Calibrated Macroeconomic Models," in *Handbook of Applied Econometrics: Macroeconomics*, ed. by M. H. Pesaran, and M. R. Wickens, chap. 7, pp. 356–390. Blackwell, Cambridge, Massachusetts.

Kitamura, Y. (2000): "Comparing Misspecified Dynamic Econometric Models Using Nonparametric Likelihood," Working Paper, University of Pennsylvania.

Kleijnen, J. P. C. (1995): "Verification and Validation of Simulation Models," *European Journal of Operational Research*, 82(1), 145–162.

Lucas, R. J. (1990): "Liquidity and Interest Rates," *Journal of Economic Theory*, 50(2), 237–264.

Maasoumi, E., and P. C. B. Phillips (1982): "On the Behavior of Inconsistent Instrumental Variables Estimators," *Journal of Econometrics*, 19, 183–201.

Moon, H. R., and F. Schorfheide (2002): "Minimum Distance Estimation of Nonstationary Time Series Models," *Econometric Theory*, 18, 1385–1407.

Newey, W. K., and D. L. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. IV, chap. 36, pp. 2111–2245. Elsevier, Amsterdam.

Phillips, P. C. B., and V. Solo (1992): "Asymptotics For Linear Processes," *Annals of Statistics*, 20, 971–1001.

Pollard, D. (1990): *Empirical Processes: Theory and Applications*, vol. 2 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, California.

Prescott, E. C. (1991): "Real Business Cycle Theory: What Have We Learned?," *Revista de Análisis Económico*, 6(2), 3–19.

Rivers, D., and Q. Vuong (2002): "Model Selection Tests For Nonlinear Dynamic Models," *The Econometrics Journal*, 5(1), 1–39.

Schorfheide, F. (2000): "Loss Function-Based Evaluation of DSGE Models," *Journal of Applied Econometrics*, 15, 645–670.

Vuong, Q. H. (1989): "Likelihood Ratio Tests For Model Selection and Non-nested Hypotheses," *Econometrica*, 57, 307–333.

Watson, M. W. (1993): "Measures of Fit for Calibrated Models," *Journal of Political Economy*, 101(6), 1011–1041.

West, K. D. (1996): "Asymptotic Inference About Predictive Ability," *Econometrica*, 64, 1067–1084.

West, K. D., and M. W. McCracken (1998): "Regression-Based Tests of Predictive Ability," *International Economic Review*, 39, 817–840.

White, H. (2000): "A Reality Check For Data Snooping," *Econometrica*, 68, 1097–1126.

——— (2001): *Asymptotic Theory For Econometricians*. Academic Press, San Diego.

Figure 1: Inflation impulse responses: CIA (top) and PAC (bottom)

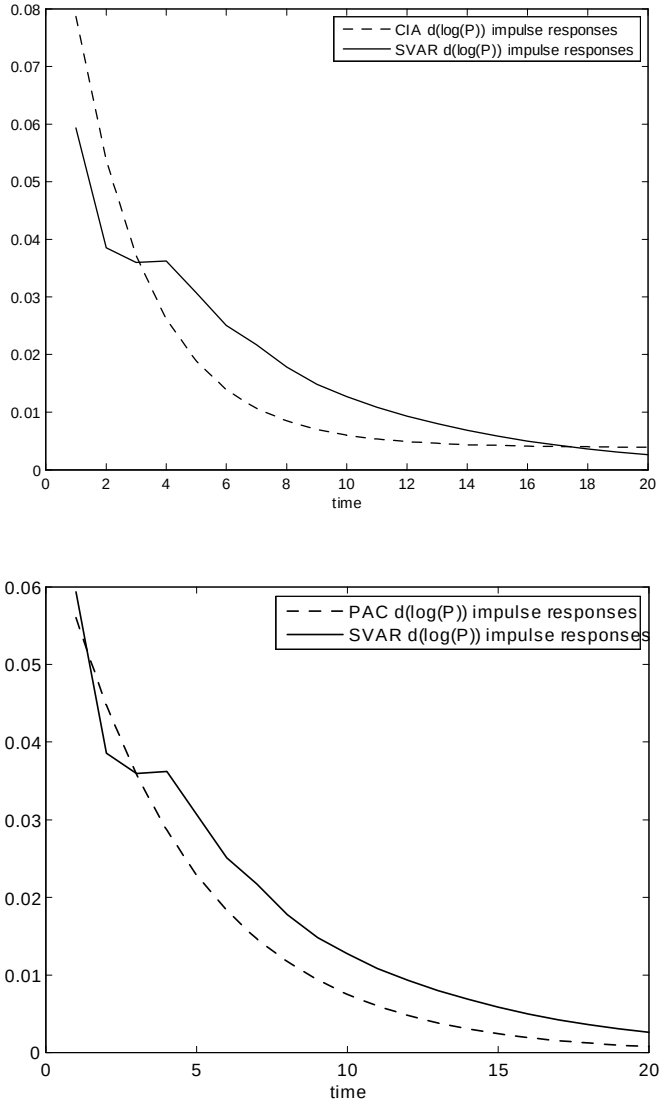

Figure 2: Output impulse responses: CIA (top) and PAC (bottom)

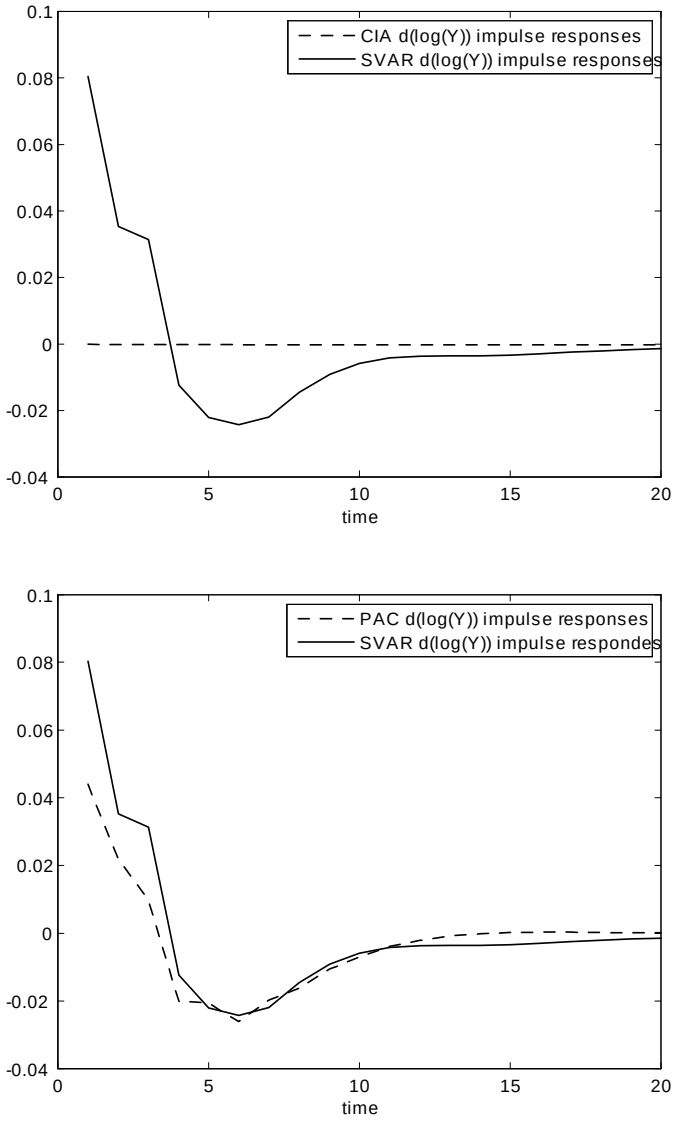

Table 1: CIA and PAC parameters' estimates and their standard errors

Name	Meaning	Range	CIA estimates	PAC estimates
α	capital share	-	0.36	0.36
			fixed	fixed
α_1	adjustment cost parameter	[0,0.0005]	-	0.0002
				(0.0003)
α_2	adjustment cost parameter	[0,10000]	-	8931
				(5632)
β	discount factor	-	0.9926	0.9926
			fixed	fixed
δ	depreciation	[0,1]	0.12	0.16
			(0.03)	(0.11)
ψ	steady state growth	[0,0.2]	0.05	0.03
			(0.04)	(0.05)
ρ	money shock persistence	[0,1]	0.37	0.45
			(0.08)	(0.10)
ϕ	leisure share in utility	[0,1]	0.38	0.54
			(0.33)	(0.25)
A_0	initial technology level	[0,1]	0.33	0.13
			(0.21)	(0.08)
m_{ss}	steady state money growth	[0.5,1.5]	1.13	1.21
			(0.37)	(0.24)