Supplement to "Limited Participation in International Business Cycle Models: A Formal Evaluation"

Xiaodan Gao[*], Viktoria Hnatkovska[†], and Vadim Marmer[†]

November 29, 2013

S.1 Introduction

This appendix contains the details of statistical methodology used for model comparison in Gao et al. (2012). In particular, we extend the test for potentially misspecified calibrated models proposed in Hnatkovska et al. (2012, 2011) along two key dimensions. In Section S.2, we show how to adjust the procedure to account for simulation uncertainty. Such adjustment becomes important when the model moments can not be computed exactly and instead simulations must be used. In Section S.3 we introduce a class-based test that allows us to compare classes of models with several models in each class. This becomes important when one is interested in evaluating the model's performance with different features, for a range of parameter values, or with different types of shocks. For instance, in the evaluations in the main paper we are interested in whether LAMP improves model's performance across all international asset market regimes. In this case, one needs a way to aggregate model fits across the different scenarios, which is what the class-based test does. In Section S.4, we describe our estimation procedure. Sections S.5 and S.6 contain

[*]Department of Strategy and Policy, NUS Business School, Mochtar Riady Building, #7-76, 15 Kent Ridge Drive, Singapore 119245. Email: bizgx@nus.edu.sg.

[†]Department of Economics, University of British Columbia, 997 - 1873 East Mall, Vancouver, BC V6T 1Z1, Canada. E-mail addresses: hnatkovs@mail.ubc.ca (Hnatkovska), vadim.marmer@ubc.ca (Marmer).

the derivations of the asymptotic variances and standard errors used in estimation procedure.

S.2 Pairwise comparison

We begin by assuming that data can be summarized using two mutually exclusive vectors of characteristics denoted by h_1 and h_2, where the first vector is used for estimation of unknown structural parameters, while the second vector is used to compare structural models. This reflects a standard practice in applied macroeconomics, when parameters are calibrated to one group of data characteristics, while models are evaluated on another. We assume that h_1 and h_2 can be estimated from data without employing a structural model. For example, in our case, h_1 consists of the estimated productivity shocks, while h_2 consists of volatilities and correlations between the variables of interest as described in Tables 3-5 in the main text.

Suppose that there are two structural models denoted $f(\theta)$ and $g(\beta)$, where θ and β are the corresponding structural parameters describing consumer's preferences, technology, etc. Here, $f(\theta)$ and $g(\beta)$ denote the value of h_2 predicted by models f and g, respectively. Naturally, vectors h_2, $f(\theta)$ and $g(\beta)$ must be of the same dimension; we assume that they are m-vectors. We allow for the competing models to be misspecified, i.e. it is possible that for all permitted values of θ and β, $h_2 \neq f(\theta)$ and $h_2 \neq g(\beta)$.

The models are allowed to share some of the parameters. Note, however, that θ and β contain only the parameters that must be estimated from data. We allow that some of the parameters may be assigned fixed values, for example, values that are commonly used in the literature. Such parameters are excluded from θ and β and absorbed into f and g.[1]

We are interested in testing a hypothesis that models f and g have equivalent fit to the data as described by h_2. For an $m \times m$ symmetric and positive definite weight matrix W_{h_2}, the null hypothesis of the models' equivalence is

$$H_0 : (h_2 - g(\beta))' W_{h_2} (h_2 - g(\beta)) - (h_2 - f(\theta))' W_{h_2} (h_2 - f(\theta)) = 0.$$

The notation indicates that the weight matrix W_{h_2} can depend on h_2. A simple choice

[1]In our application θ and β are the same and describe the productivity process.

for a weight matrix is to use the identity matrix. In that case, the weight matrix is independent of h_2, and the models are compared in terms of their squared prediction errors. Another example for W_{h_2} is a diagonal matrix with the reciprocals of the elements of h_2 on the main diagonal. With such a choice of the weight matrix, the models are compared in terms of the squares of their percentage prediction errors. In our application, we use a combination of the two. That is to evaluate the models, for some parameters, such as correlations, we use prediction errors, while for others, such as volatilities, we use percentage prediction errors.

The alternative hypotheses are

$$H_f : (h_2 - g(\beta))' W_{h_2} (h_2 - g(\beta)) - (h_2 - f(\theta))' W_{h_2} (h_2 - f(\theta)) > 0,$$
$$H_g : (h_2 - g(\beta))' W_{h_2} (h_2 - g(\beta)) - (h_2 - f(\theta))' W_{h_2} (h_2 - f(\theta)) < 0,$$

where f has a better fit according to H_f, and g has a better fit according to H_g.

Let $\hat{h}_1$ and $\hat{h}_2$ denote the estimators of h_1 and h_2, respectively. We assume that $\hat{h}_1$ and $\hat{h}_2$ do not require the knowledge of the true structural model, are consistent and asymptotically normal as described in the following assumption:

$$\sqrt{n} \begin{pmatrix} \hat{h}_1 - h_1 \\ \hat{h}_2 - h_2 \end{pmatrix} \to_d N \left(0, \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12}' & \Lambda_{22} \end{pmatrix} \right), \tag{S.1}$$

where n denotes the sample size used in estimation of h_1 and h_2, Λ_{11} and Λ_{22} denote the asymptotic variance-covariance matrices of $\hat{h}_1$ and $\hat{h}_2$ respectively, and Λ_{12} denotes the asymptotic covariance between $\hat{h}_1$ and $\hat{h}_2$. Let $\hat{\Lambda}_{11}$, $\hat{\Lambda}_{22}$ and $\hat{\Lambda}_{12}$ denote consistent estimators of the corresponding elements in the above asymptotic variance-covariance matrix. In a typical time-series application, Λ_{11}, Λ_{22} and Λ_{12} are long-run variances and covariances and, therefore, require HAC-type estimators, see Newey and West (1987) and Andrews (1991).

Let $\hat{\theta}$ and $\hat{\beta}$ denote the estimators of θ and β respectively. We assume that the estimators are asymptotically linear in h_1:

$$\sqrt{n}(\hat{\theta} - \theta) = A\sqrt{n}(\hat{h}_1 - h_1) + o_p(1), \tag{S.2}$$
$$\sqrt{n}(\hat{\beta} - \beta) = B\sqrt{n}(\hat{h}_1 - h_1) + o_p(1), \tag{S.3}$$

where matrices A and B may depend on the elements of h_1. This specification is satisfied by most estimators used in practice. Appendix S.4 contains the derivations of equation (S.2) for our estimators.[2] We assume that A and B can be consistently estimated, and use $\hat{A}$ and $\hat{B}$ to denote their estimators.

When functions $f(\theta)$ and $g(\beta)$ are too complicated for analytical or even exact numerical calculations, we assume that they can be estimated by simulations. For example, as in our case, one can draw random shocks and solve the models as described in Section 3 in the main text using $\hat{\theta}$ for model f and $\hat{\beta}$ for model g, and obtain a set of random equilibrium values for the variables of interest. By repeating this process R times, one obtains a sample of R observations for the variables of interest, which can be used to estimate f and g by averaging across the simulations. Let $\hat{f}(\hat{\theta})$ and $\hat{g}(\hat{\beta})$ denote such estimators.

We assume that, at the true values θ and β, estimators $\hat{f}(\theta)$ and $\hat{g}(\beta)$ are independent of $\hat{h}_1$ and $\hat{h}_2$, and satisfy the following assumption:

$$\sqrt{R}\begin{pmatrix} \hat{f}(\theta) - f(\theta) \\ \hat{g}(\beta) - g(\beta) \end{pmatrix} \to_d N\left(0, \begin{pmatrix} \Lambda_{ff} & \Lambda_{fg} \\ \Lambda_{fg}' & \Lambda_{gg} \end{pmatrix}\right). \tag{S.4}$$

We use $\hat{\Lambda}_{ff}$, $\hat{\Lambda}_{gg}$ and $\hat{\Lambda}_{fg}$ to denote consistent estimators of the asymptotic variances and covariance in (S.4).

Our test is based on the difference between the estimated fits of the two models:

$$S = \big(\hat{h}_2 - \hat{g}(\hat{\beta})\big)' W_{\hat{h}_2} \big(\hat{h}_2 - \hat{g}(\hat{\beta})\big) - \big(\hat{h}_2 - \hat{f}(\hat{\theta})\big)' W_{\hat{h}_2} \big(\hat{h}_2 - \hat{f}(\hat{\theta})\big).$$

Under the assumptions in (S.1)-(S.4), S is asymptotically normal, and its standard error can be computed as $\hat{\sigma}/\sqrt{n}$, where[3]

$$\hat{\sigma}^2 = 4\hat{\sigma}_1^2 + 4\hat{\sigma}_2^2, \tag{S.5}$$

$$\hat{\sigma}_1^2 = \begin{pmatrix} \hat{A}' \frac{\partial \hat{f}(\hat{\theta})'}{\partial \theta} W_{\hat{h}_2} \left(\hat{h}_2 - \hat{f}(\hat{\theta})\right) - \hat{B}' \frac{\partial \hat{g}(\hat{\beta})'}{\partial \beta} W_{\hat{h}_2} \left(\hat{h}_2 - \hat{g}(\hat{\beta})\right) \\ W_{\hat{h}_2}\left(\hat{f}(\hat{\theta}) - \hat{g}(\hat{\beta})\right) + 0.5 \frac{\partial w(\hat{h}_2)'}{\partial h_2} J' K\left(\hat{h}, \hat{f}(\hat{\theta}), \hat{g}(\hat{\beta})\right) \end{pmatrix}' \begin{pmatrix} \hat{\Lambda}_{11} & \hat{\Lambda}_{12} \\ \hat{\Lambda}_{12}' & \hat{\Lambda}_{22} \end{pmatrix}$$
$$\times \begin{pmatrix} \hat{A}' \frac{\partial \hat{f}(\hat{\theta})'}{\partial \theta} W_{\hat{h}_2} \left(\hat{h}_2 - \hat{f}(\hat{\theta})\right) - \hat{B}' \frac{\partial \hat{g}(\hat{\beta})'}{\partial \beta} W_{\hat{h}_2} \left(\hat{h}_2 - \hat{g}(\hat{\beta})\right) \\ W_{\hat{h}_2}\left(\hat{f}(\hat{\theta}) - \hat{g}(\hat{\beta})\right) + 0.5 \frac{\partial w(\hat{h}_2)'}{\partial h_2} J' K\left(\hat{h}, \hat{f}(\hat{\theta}), \hat{g}(\hat{\beta})\right) \end{pmatrix}, \tag{S.6}$$

[2]In our application, because β and θ are the same, we do not use equation (S.3).
[3]The asymptotic variance formula is explained in Appendix S.5

$$\hat{\sigma}_2^2 = \frac{n}{R} \begin{pmatrix} W_{\hat{h}_2}\left(\hat{h}_2 - \hat{f}(\hat{\theta})\right) \\ -W_{\hat{h}_2}\left(\hat{h}_2 - \hat{g}(\hat{\beta})\right) \end{pmatrix}' \begin{pmatrix} \hat{\Lambda}_{ff} & \hat{\Lambda}_{fg} \\ \hat{\Lambda}'_{fg} & \hat{\Lambda}_{gg} \end{pmatrix} \begin{pmatrix} W_{\hat{h}_2}\left(\hat{h}_2 - \hat{f}(\hat{\theta})\right) \\ -W_{\hat{h}_2}\left(\hat{h}_2 - \hat{g}(\hat{\beta})\right) \end{pmatrix}. \quad \text{(S.7)}$$

In the expression for $\hat{\sigma}_1^2$,

$$K(h, f, g) = ((h - g) \otimes (h - g)) - ((h - f) \otimes (h - f)), \quad \text{(S.8)}$$

vector $w(h_2)$ collects the element of W_{h_2} without duplicates, and J denotes a known $m^2 \times m$ selection matrix of zeros and ones such that

$$\text{vec}(W_{h_2}) = Jw(h_2). \quad \text{(S.9)}$$

For example, when W_{h_2} is a diagonal matrix with the reciprocals of the elements of h_2 on the main diagonal, we have that $w_i(h) = 1/h_i$, $i = 1, \ldots, m$, and

$$J = \begin{pmatrix} J_1 \\ \vdots \\ J_m \end{pmatrix},$$

where, for $i = 1, \ldots, m$, J_i is an $m \times m$ matrix with 1 in position (i, i) and zeros everywhere else.

In (S.5), the first term, $\hat{\sigma}_1^2$, reflects the uncertainty due to estimation of θ, β, and h_2. For example, when comparing the models at some known fixed parameter values $\bar{\theta}$ and $\bar{\beta}$, matrices $\hat{A}$ and $\hat{B}$ should be replaced by zeros. Similarly, when comparing the models using a known fixed weight matrix (independent of h_2), the terms $0.5(\partial w(\hat{h})'/\partial h)J'K(\hat{h}, \hat{f}, \hat{g})$ in (S.6) should be replaced with zeros.

The second term in (S.5), $\hat{\sigma}_2^2$, is due to the simulations uncertainty in computation of $\hat{f}(\hat{\theta})$ and $\hat{g}(\hat{\beta})$. This term is zero when f and g can be evaluated numerically (without resorting to simulations). Uncertainty due to simulations can be ignored if one can select a large number of simulations R so that the ratio n/R is sufficiently small.

Our asymptotic test with significance level α is:

Reject H_0 in favor of H_f when $\sqrt{n}S/\hat{\sigma} > z_{1-\alpha/2}$,
Reject H_0 in favor of H_g when $\sqrt{n}S/\hat{\sigma} < -z_{1-\alpha/2}$,

where $z_{1-\alpha/2}$ denotes a standard normal critical value.

S.3 Comparison of model classes

When there are general classes of models with each class containing several sub-models, the researcher may be interested in overall comparison of classes instead of pairwise comparison of each sub-model. We discuss such a procedure in this section.

Suppose that we have two classes of models with k models in each class: $\mathcal{F} = \{f_1(\theta), \ldots, f_k(\theta)\}$ and $\mathcal{G} = \{g_1(\beta), \ldots, g_k(\beta)\}$. We are interested in comparing the overall performances of $\mathcal{F}$ and $\mathcal{G}$. More specifically, we are testing whether $\mathcal{F}$ and $\mathcal{G}$ have the same distance from the moments vector h_2. Here we adopt the von Mises-type (or average) distance between a set $\mathcal{F}$ and a point h_2:

$$D_M(\mathcal{F}, h_2) = \sum_{j=1}^{k} d(f_j(\theta), h_2; W_{h_2}),$$

where $d(f_j(\theta), h_2; W_{h_2})$ denotes the previously used weighted Euclidean distance between vectors $f_j(\theta)$ and h_2:

$$d(f_j(\theta), h_2; W_{h_2}) = (h_2 - f_j(\theta))' W_{h_2} (h_2 - f_j(\theta)).$$

Note that, alternatively, one could use a Kolmogorov-type distance between $\mathcal{F}$ and h_2: $D_{min}(\mathcal{F}, h_2) = \min_{j=1,\ldots,k} d(f_j(\theta), h_2; W_{h_2})$ or $D_{max}(\mathcal{F}, h_2) = \max_{j=1,\ldots,k} d(f_j(\theta), h_2; W_{h_2})$. While with a Kolmogorov-type distance each class is represented by its best (or worst) performer, the von Mises-type distance measures the average performance of a class of models, and we find it more appropriate when the object of interest is the overall performance of a class.

Thus, our null hypothesis of interest can now be stated as

$$H_0 : D_M(\mathcal{F}, h_2) = D_M(\mathcal{G}, h_2), \tag{S.10}$$

and a test can be based on the difference of sample analogues of $D_M(\mathcal{F}, h_2)$ and

$D_M(\mathcal{G}, h_2)$:[4]

$$S^M = \sum_{j=1}^{k} \left(d(\hat{g}_j(\hat{\beta}), \hat{h}_2; W_{\hat{h}_2}) - d(\hat{f}_j(\hat{\theta}), \hat{h}_2; W_{\hat{h}_2}) \right).$$

Let $\hat{\sigma}_M$ denote the standard error of S^M. As before, the null hypothesis in (S.10) should be rejected when the studentized statistic $\sqrt{n}S^M/\hat{\sigma}_M$ exceeds standard normal critical values. The standard error can be computed as follows.[5] Define

$$Q_j = \begin{pmatrix} A' \frac{\partial f_j(\theta)'}{\partial \theta} W_{h_2}(h_2 - f_j(\theta)) - B' \frac{\partial g_j(\beta)'}{\partial \beta} W_{h_2}(h_2 - g_j(\beta)) \\ W_{h_2}(f_j(\theta) - g_j(\beta)) + 0.5 \frac{\partial w(h_2)'}{\partial h} J' K(h_2, f_j(\theta), g_j(\beta)) \end{pmatrix},$$

and let $\hat{Q}_j$ denote a consistent estimator of Q_j. Ignoring the simulation uncertainty, the standard error of S^M is given by the square-root of

$$\hat{\sigma}_M^2 = 4 \left(\sum_{j=1}^{k} \hat{Q}_j \right)' \begin{pmatrix} \hat{\Lambda}_{11} & \hat{\Lambda}_{12} \\ \hat{\Lambda}_{12}' & \hat{\Lambda}_{22} \end{pmatrix} \left(\sum_{j=1}^{k} \hat{Q}_j \right). \quad \text{(S.11)}$$

The expression in (S.11) can be easily adjusted to account for simulation uncertainty. Note that the formula will depend on whether each model is simulated independently or if the same simulated structural shocks used in all models. In our case, the number of simulations is sufficiently large for the simulation uncertainty to be ignored.

S.4 Estimation details

In this section, we describe our estimation procedure, and show how it corresponds with the asymptotic linearization in (S.2) and (S.3).

First, note that in our case, $\theta = \beta = (\rho_{11}, \rho_{12}, \sigma_{e_1}, \sigma_{e_1 e_2})'$. The parameters are estimated using the following estimating equations:

$$\begin{pmatrix} z_{1,t} \\ z_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{12} & \rho_{11} \end{pmatrix} \begin{pmatrix} z_{1,t-1} \\ z_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}, \quad \text{(S.12)}$$

[4]We assume here, as in our case, that the same estimator of structural parameters is used inside each class of models. A generalization allowing for model-specific estimators inside each class is straightforward.

[5]The details of the derivation are provided in Appendix S.6.

$$\sigma_{e_1} = \sqrt{E\varepsilon_{1,t}^2}, \tag{S.13}$$

$$\sigma_{e_1 e_2} = \frac{E\varepsilon_{1,t}\varepsilon_{2,t}}{\sqrt{E\varepsilon_{1,t}^2 E\varepsilon_{2,t}^2}}. \tag{S.14}$$

Define $y_{1,t} = z_{1,t}$ and $y_{2,t} = z_{2,t}$ for $t = 2, \ldots, n$, and let Y_1 and Y_2 denote the corresponding $(n-1)$-vectors of observations. Let $X_t = (1, z_{1,t-1}, z_{2,t-1})'$ for $t = 2, \ldots, n$, and let X denote the corresponding $(n-1) \times 3$ matrix of observations. Let ε_1 and ε_2 denote the $(n-1)$-vectors of observations on the error terms. We have the following SUR system:

$$Y_* = (I_2 \otimes X)\, \gamma_* + \varepsilon_*,$$

where $Y_* = (Y_1', Y_2')'$, $\varepsilon_* = (\varepsilon_1', \varepsilon_2')'$, and $\gamma_* = (\mu_1, \rho_{11}, \rho_{12}, \mu_2, \rho_{12}, \rho_{11})'$, and note that γ_* is restricted by $R\gamma_* = 0_{2\times 1}$, where

$$R = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 \end{pmatrix}.$$

Define Σ as the variance-covariance matrix of $(\varepsilon_1, \varepsilon_2)'$:

$$\Sigma = \begin{pmatrix} \sigma_{e_1} & \sigma_{e_1 e_2}\sigma_{e_1}\sigma_{e_2} \\ \sigma_{e_1 e_2}\sigma_{e_1}\sigma_{e_2} & \sigma_{e_2} \end{pmatrix},$$

and let $\hat{\Sigma}$ denote its consistent estimator. For example, $\hat{\Sigma}$ can be constructed using the residuals obtained from OLS equation-by-equation estimation of (S.12). The restricted (FGLS) efficient SUR estimator of γ_* is given by:

$$\hat{\gamma}_* = \tilde{\gamma}_* - \left(\hat{\Sigma}^{-1} \otimes (X'^{-1}\right) R' \left(R\left(\hat{\Sigma}^{-1} \otimes (X'^{-1}\right) R'\right)^{-1} R\tilde{\gamma}_*,$$

where $\tilde{\gamma}_*$ denotes the unrestricted OLS equation-by-equation estimator of γ_*.[6]

Let $\hat{\sigma}_{e_1}$ and $\hat{\sigma}_{e_1 e_2}$ denote the estimators of σ_{e_1} and $\sigma_{e_1 e_2}$ respectively constructed by replacing the expectations in (S.13) and (S.14) with sample averages and ε's with fitted residuals from the SUR system above. We need additional notation to describe

[6] Since the two equations have the same set of regressors, the unrestricted efficient SUR estimator is the equation-by-equation OLS estimator.

the linearization of the estimator of β in (S.3). Define:

$$H = I_6 - \left(\Sigma^{-1} \otimes \left(\text{plim}_{n\to\infty} \frac{X'X}{n} \right)^{-1} \right) R' \left(R \left(\Sigma^{-1} \otimes \left(\text{plim}_{n\to\infty} \frac{X'X}{n} \right)^{-1} \right) R' \right)^{-1} R,$$

and let $H_{2;3}$ denote the second and third rows of H. In this case, $\sqrt{n}(\hat{\beta} - \beta)$, B, and $\sqrt{n}(\hat{h}_1 - h_1)$ in (S.3) are given by the corresponding terms in the following expression:

$$\sqrt{n} \left(\begin{array}{c} \left(\begin{array}{c} \hat{\rho}_{11} - \rho_{11} \\ \hat{\rho}_{12} - \rho_{12} \end{array} \right) \\ \hat{\sigma}_{e_1} - \sigma_{e_1} \\ \hat{\sigma}_{e_1 e_2} - \sigma_{e_1 e_2} \end{array} \right) = \left(\begin{array}{cccc} H_{2;3} & 0_{2\times 1} & 0_{2\times 1} & 0_{2\times 1} \\ 0_{1\times 6} & \frac{1}{2\sigma_{e1}} & 0 & 0 \\ 0_{1\times 6} & -\frac{\sigma_{e_1 e_2}}{2\sigma_{e_1}^2} & -\frac{\sigma_{e_1 e_2}}{2\sigma_{e_2}^2} & \frac{1}{\sigma_{e_1}\sigma_{e_2}} \end{array} \right) \frac{1}{\sqrt{n}} \sum_{t=2}^{n} \left(\begin{array}{c} \left(\begin{array}{c} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{array} \right) \otimes X_t \\ \varepsilon_{1,t}^2 - \sigma_{e_1}^2 \\ \varepsilon_{2,t}^2 - \sigma_{e_2}^2 \\ \varepsilon_{1,t}\varepsilon_{2,t} - \sigma_{e_1 e_2}\sigma_{e_1}\sigma_{e_2} \end{array} \right)$$
$$+ o_p(1),$$

where $\hat{\rho}_{11}$ and $\hat{\rho}_{12}$ denote the second and third elements of the efficient SUR estimator $\hat{\gamma}_*$.

To estimate Λ_{11} and B, one should replace the population parameters in the above expression with their sample counterparts and ε's with fitted residuals from SUR estimation. To estimate Λ_{22} and Λ_{12}, one can use a linearization (similar to that of $\hat{\sigma}_{e_1}$ and $\hat{\sigma}_{e_1 e_2}$ above) for $\hat{h}_2$.

S.5 Derivation of the asymptotic variances formulas in (S.5)-(S.7)

When H_0 is true, S can be written as

$$S = \left[\left(\hat{h}_2 - \hat{g}(\hat{\beta}) \right)' W_{\hat{h}_2} \left(\hat{h}_2 - \hat{g}(\hat{\beta}) \right) - (h_2 - g(\beta))' W_{h_2} (h_2 - g(\beta)) \right]$$
$$- \left[\left(\hat{h}_2 - \hat{f}(\hat{\theta}) \right)' W_{\hat{h}_2} \left(\hat{h}_2 - \hat{f}(\hat{\theta}) \right) - (h_2 - f(\theta))' W_{h_2} (h_2 - f(\theta)) \right]. \quad \text{(S.15)}$$

Next,

$$\left(\hat{h}_2 - \hat{g}(\beta) \right)' W_{\hat{h}_2} \left(\hat{h}_2 - \hat{g}(\beta) \right) - (h_2 - g(\beta))' W_{h_2} (h_2 - g(\beta)) \quad \text{(S.16)}$$

$$
\begin{aligned}
&= \left(\hat{h}_2 - \hat{g}(\beta)\right)' \left(W_{\hat{h}_2} - W_{h_2}\right) \left(\hat{h}_2 - \hat{g}(\beta)\right) \\
&\quad + \left(\hat{h}_2 - \hat{g}(\beta) + h_2 - g(\beta)\right)' W_{h_2} \left(\hat{h}_2 - h_2\right) \\
&\quad - \left(\hat{h}_2 - \hat{g}(\beta) + h_2 - g(\beta)\right)' W_{h_2} \left(\hat{g}(\beta) - g(\beta)\right) \\
&= \left((h_2 - g(\beta)) \otimes (h_2 - g(\beta))\right)' J \left(w(\hat{h}_2) - w(h_2)\right) \\
&\quad + 2\left(h_2 - g(\beta)\right)' W_{h_2} \left(\hat{h}_2 - h_2\right) \\
&\quad - 2\left(h_2 - g(\beta)\right)' W_{h_2} \left(\hat{g}(\beta) - g(\beta)\right) + o_p(1/\sqrt{n}),
\end{aligned}
$$

where the last equality holds by $\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec}(B)$ (see Magnus and Neudecker (1999), equation (5) on page 30), (S.4), and (S.9). With a similar expression for the second term in (S.15) and a first-order Taylor expansion for $w(\hat{h}_2)$, we obtain that (S.16) multiplied by $\sqrt{n}$ is equal to

$$
\begin{pmatrix}
2W_{h_2}(f(\theta) - g(\beta)) + \frac{\partial w(h_2)'}{\partial h} J' K(h_2, f(\theta), g(\beta)) \\
2W_{h_2}(h_2 - f(\theta)) \\
-2W_{h_2}(h_2 - g(\beta))
\end{pmatrix}'
\begin{pmatrix}
\sqrt{n}\big(\hat{h}_2 - h_2\big) \\
\sqrt{\frac{n}{R}}\sqrt{R}\big(\hat{f}(\theta) - f(\theta)\big) \\
\sqrt{\frac{n}{R}}\sqrt{R}\big(\hat{g}(\beta) - f(\beta)\big)
\end{pmatrix}
+ o_p(1) \tag{S.17}
$$

Next, by a first-order Taylor expansion,

$$
\begin{aligned}
\left(\hat{h}_2 - \hat{g}(\hat{\beta})\right)' W_{\hat{h}_2} \left(\hat{h}_2 - \hat{g}(\hat{\beta})\right) &= \left(\hat{h}_2 - \hat{g}(\beta)\right)' W_{\hat{h}_2} \left(\hat{h}_2 - \hat{g}(\beta)\right) \\
&\quad - 2\left(\hat{h}_2 - \hat{g}(\beta)\right)' W_{\hat{h}_2} \frac{\partial g(\beta)}{\partial \beta'} \big(\hat{\beta} - \beta\big) + o_p(1/\sqrt{n}).
\end{aligned}
$$

By combining this result (and a similar expansion for model f) with the results in (S.17) and (S.15), and using (S.2)-(S.3), we obtain:

$$
\begin{aligned}
\sqrt{n} S = 2
&\begin{pmatrix}
A' \frac{\partial f(\theta)'}{\partial \theta} W_{h_2}(h_2 - f(\theta)) - B' \frac{\partial g(\beta)'}{\partial \beta} W_{h_2}(h_2 - g(\beta)) \\
W_{h_2}(f(\theta) - g(\beta)) + 0.5 \frac{\partial w(h_2)'}{\partial h} J' K(h_2, f(\theta), g(\beta)) \\
W_{h_2}(h_2 - f(\theta)) \\
-W_{h_2}(h_2 - g(\beta))
\end{pmatrix}' \\
&\times
\begin{pmatrix}
\sqrt{n}\big(\hat{h}_1 - h_1\big) \\
\sqrt{n}\big(\hat{h}_2 - h_2\big) \\
\sqrt{\frac{n}{R}}\sqrt{R}\big(\hat{f}(\theta) - f(\theta)\big) \\
\sqrt{\frac{n}{R}}\sqrt{R}\big(\hat{g}(\beta) - f(\beta)\big)
\end{pmatrix}
+ o_p(1).
\end{aligned} \tag{S.18}
$$

The results in (S.5)-(S.7) now follow by (S.1) and (S.4).

S.6 Derivation of the standard error formula in (S.11)

When H_0 is true, one can write

$$S^M = \sum_{j=1}^{k} \left(d(\hat{g}_j(\hat{\beta}), \hat{h}_2; W_{\hat{h}_2}) - d(\hat{f}_j(\hat{\theta}), \hat{h}_2; W_{\hat{h}_2}) - d(g_j(\beta), h_2; W_{h_2}) + d(f_j(\theta), h_2; W_{h_2}) \right).$$

Assuming that the contribution of simulation uncertainty is negligible, it follows from (S.18) that

$$\sqrt{n} S^M = 2 \sum_{j=1}^{k} Q_j \sqrt{n} \begin{pmatrix} \hat{h}_1 - h_1 \\ \hat{h}_2 - h_2 \end{pmatrix} + o_p(1),$$

where note that Q_j is the same as the first two row-blocks of the multiplication matrix appearing in (S.18). The result in (S.11) follows.

S.7 Robustness with respect to parameter λ

This section presents the pair-wise model comparisons and class-based comparison results of the BKK models and LAMP models with various values for the consumption share of participants, parameter λ. In particular, we consider two scenarios: (i) $\lambda = 0.3$, the results for which are given in Table 1; and (ii) $\lambda = 0.7$, the results for which are in Table 2.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Gao, X., V. Hnatkovska, and V. Marmer (2012). Limited participation in international business cycle models: A formal evaluation. Working Paper 2012-1, University of British Columbia.

Hnatkovska, V., V. Marmer, and Y. Tang (2011). Supplement to comparison of misspecified calibrated models. Working Paper, University of British Columbia.

Hnatkovska, V., V. Marmer, and Y. Tang (2012). Comparison of misspecified calibrated models: The minimum distance approach. *Journal of Econometrics 169*(1), 131–138.

Magnus, J. R. and H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics* (Revised ed.). Chichester: Wiley.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariane matrix. *Econometrica 55*, 703–708.

Table 1: Test results from the comparison of models with λ=0.3

	Model f					
Model g		BKK			LAMP	
(a) Volatilities	FA	BE	CM	FA	BE	CM
BKK, FA	0					
BKK, BE	0.19***	0				
	(0.00)					
BKK, CM	0.20***	0.02***	0			
	(0.00)	(0.00)				
LAMP, FA	0.05***	-0.13***	-0.15***	0		
	(0.00)	(0.00)	(0.00)			
LAMP, BE	0.26***	0.08***	0.06***	0.21***	0	
	(0.00)	(0.00)	(0.00)	(0.10)		
LAMP, CM	0.28***	0.10***	0.08***	0.23***	0.02***	0
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
(b) Correlations (with output and cross-country)						
BKK, FA	0					
BKK, BE	0.13	0				
	(0.77)					
BKK, CM	0.05	-0.08***	0			
	(0.91)	(0.00)				
LAMP, FA	-0.21	-0.34	-0.26	0		
	(0.34)	(0.58)	(0.68)			
LAMP, BE	-0.45	-0.58***	-0.50**	-0.24	0	
	(0.21)	(0.01)	(0.02)	(0.62)		
LAMP, CM	-0.55	-0.69***	-0.60**	-0.34	-0.10***	0
	(0.12)	(0.01)	(0.02)	(0.46)	(0.00)	
(c) Overall						
BKK, FA	0					
BKK, BE	0.32	0				
	(0.47)					
BKK, CM	0.25	-0.07***	0			
	(0.57)	(0.00)				
LAMP, FA	-0.16	-0.48	-0.41	0		
	(0.47)	(0.44)	(0.51)			
LAMP, BE	-0.18	-0.50**	-0.44**	-0.03	0	
	(0.60)	(0.03)	(0.05)	(0.95)		
LAMP, CM	-0.27	-0.59**	-0.52**	-0.11	-0.08***	0
	(0.45)	(0.02)	(0.04)	(0.81)	(0.01)	
LAMP - BKK				**-1.18***		
class comparison				**(0.09)**		

Note: This Table reports the test statistics for comparison of the model in the row (model g) against the model in the column (model f). Positive numbers for the test statistic indicate that, compared with the model in the column, the model in the row provides a worse fit to the data moments. P-values are in the parentheses. * p-value≤0.10, ** p-value≤0.05, *** p-value≤0.01.

Table 2: Test results from the comparison of models with λ=0.7

	Model f					
Model g		BKK			LAMP	
(a) Volatilities	FA	BE	CM	FA	BE	CM
BKK, FA	0					
BKK, BE	0.19***	0				
	(0.00)					
BKK, CM	0.20***	0.02***	0			
	(0.00)	(0.00)				
LAMP, FA	0.01***	-0.17***	-0.19***	0		
	(0.00)	(0.00)	(0.00)			
LAMP, BE	0.21***	0.02***	0.01***	0.19***	0	
	(0.00)	(0.00)	(0.00)	(0.10)		
LAMP, CM	0.23***	0.04***	0.02***	0.21***	0.02***	0
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
(b) Correlations (with output and cross-country)						
BKK, FA	0					
BKK, BE	0.13	0				
	(0.77)					
BKK, CM	0.05	-0.08***	0			
	(0.91)	(0.00)				
LAMP, FA	-0.05	-0.19	-0.10	0		
	(0.49)	(0.71)	(0.84)			
LAMP, BE	-0.07	-0.20***	-0.12**	-0.02	0	
	(0.86)	(0.00)	(0.04)	(0.97)		
LAMP, CM	-0.16	-0.29***	-0.21***	-0.11	-0.09***	0
	(0.70)	(0.00)	(0.00)	(0.82)	(0.00)	
(c) Overall						
BKK, FA	0					
BKK, BE	0.32	0				
	(0.47)					
BKK, CM	0.25	-0.07***	0			
	(0.57)	(0.00)				
LAMP, FA	-0.04	-0.36	-0.29	0		
	(0.61)	(0.48)	(0.57)			
LAMP, BE	0.14	-0.18***	-0.11**	0.18	0	
	(0.74)	(0.00)	(0.05)	(0.70)		
LAMP, CM	0.07	-0.25***	-0.19***	0.11	-0.07***	0
	(0.87)	(0.00)	(0.01)	(0.82)	(0.00)	
LAMP - BKK			**-0.41****			
class comparison			**(0.05)**			

Note: This Table reports the test statistics for comparison of the model in the row (model g) against the model in the column (model f). Positive numbers for the test statistic indicate that, compared with the model in the column, the model in the row provides a worse fit to the data moments. P-values are in the parentheses. * p-value$\leq$0.10, ** p-value$\leq$0.05, *** p-value$\leq$0.01.