Limited Participation in International Business Cycle Models: A Formal Evaluation*

Xiaodan Gao† Viktoria Hnatkovska‡ Vadim Marmer†

November 2013

Abstract

In this paper, we argue that limited asset market participation (LAMP) plays an important role in explaining international business cycles. We show that when LAMP is introduced into an otherwise standard model of international business cycles, the performance of the model improves significantly, especially in matching cross-country correlations. To perform formal evaluation of the models we develop a novel statistical procedure that adapts the statistical framework of Vuong (1989) to DSGE models. Using this methodology, we show that the improvements brought out by LAMP are statistically significant, leading a model with LAMP to outperform a representative agent model. Furthermore, when LAMP is introduced, a model with complete markets is found to do as well as a model with no trade in financial assets – a well-known favorite in the literature. Our results remain robust to the inclusion of investment specific technology shocks.

JEL Classification: F3, F4

Keywords: international business cycles, incomplete markets, limited asset market participation

1 Introduction

A number of existing studies have shown that access to international borrowing and lending is important for international business cycles. We verify this result formally by means of a novel statistical procedure applied to several versions of a standard two-country two-good model. We show that a model with no cross-border asset trades (financial autarky) is the specification that outperforms other models in matching the data and that the differences in performance are statistically significant. We then propose a competing model that allows for within country household

*We thank our editor, Christopher Otrok, and an anonymous referee for helpful suggestions.

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

‡Vancouver School of Economics, University of British Columbia, 997 - 1873 East Mall, Vancouver, BC V6T 1Z1, Canada. E-mail addresses: hnatkovs@mail.ubc.ca (Hnatkovska) and vadim.marmer@ubc.ca (Marmer). Hnatkovska is the corresponding author.

heterogeneity in participation in asset markets. Using our procedure we show that this amended model does significantly better in matching the data than a representative agent benchmark. Furthermore, when limited participation is introduced, complete markets specification performs as well as financial autarky in fitting the data.

How does access to different financial assets affect the functioning of the economy? In a seminal work, Backus et al. (1995, 1994), BKK hereafter, document the key business cycle regularities in industrial countries related to volatilities of consumption, output, investment and their cross-country co-movements, and develop an international business cycles model with complete asset markets in an attempt to rationalize the data facts. They show that only some of the regularities can be explained by the model. The BKK model fails in three key dimensions. First, while cross-country consumption correlations tend to be similar to cross-country output correlations in the data, the model predicts consumption correlations far exceeding those for outputs. This is the so-called "quantity" puzzle (Backus et al., 1995). Second, investment and employment are positively correlated across countries while the model predicts a negative correlation. This data-model disconnect is usually referred to as the "international comovement" puzzle (Baxter, 1995). Third, the model generates significantly less volatility in the terms of trade and the real exchange rate relative to the data. The model also predicts a positive correlation between real exchange rate and the ratio of domestic to foreign consumption, again contrary to the data.[1]

To account for the disconnect between the model and data, Baxter and Crucini (1995), Kollmann (1996), Arvanitis and Mikkola (1996), Corsetti et al. (2008) study economies in which the only asset traded internationally is a non-contingent bond. They show that these economies admit different allocations from those arising under complete asset markets only if productivity shocks are very persistent and do not spill over across countries. Heathcote and Perri (2002) develop this argument further by considering an economy in which no international assets are traded. They call it financial autarky. They find that the equilibrium dynamics under financial autarky are similar to those in the data. Their conclusion, however, is based primarily on an "eyeball" comparison of various moments predicted by the model with those of competing models and with the data. In fact, such informal moment comparison is standard practice in the literature.

There are several important shortcomings of the "eyeball" approach. First, it does not inform whether differences in the model performance are statistically significant. Namely, an "eyeball" approach cannot credibly distinguish between the systematic differences in model performance (in the sense that the model uncovers important relationships between the variables at the population level) and, therefore, is likely to be found in other data sets; and differences arising due to random variations in the data. Second, often model comparison is hindered by the fact that one model performs better in matching some moments, but competing models perform better in matching other moments. Without a metric that aggregates across various moments of interest, informal

[1] A detailed recent discussion of various puzzles in the international business cycles models can be found in Mandelman et al. (2011).

model comparison remains inconclusive.

In this paper we utilize a testing procedure that allows the researcher to assess the statistical significance of results when comparing DSGE models to the data. The procedure builds upon Hnatkovska et al. (2012) and is a version of Vuong-type tests for misspecified models (Vuong, 1989) adopted for the DSGE framework. Suppose that the researcher is interested in evaluating whether a newly proposed economic structure is important for explaining some chosen data patterns. For that purpose, a Vuong-type procedure compares the empirical fit of the new model with that of a leading benchmark model, and tests a null hypothesis that they are equal. If the null hypothesis is accepted, then the researcher must conclude that there is no sufficient empirical evidence in favor of the new model. On the other hand, if the null hypothesis is rejected in favor of the new model, the researcher can credibly argue that the new economic structure provides a superior explanation to the data patterns over that of the benchmark model. The procedure does not require that either of the competing models be correctly specified, and therefore the conclusions are robust to misspecification.[2]

The procedure consists of several steps. In the first step we determine the values of the deep structural parameters in each of the competing models. This can be done either informally by setting the parameters to their values typically used in the literature or through formal estimation where the values for the parameters are chosen to match certain characteristics of the data. In the second step, we compute the weighted Euclidean distance between the vectors of model-predicted characteristics and their estimates from the data. We then obtain the test statistic as the difference between the estimated measures of fit of the two competing models as well as its standard error. The standard error has to take into account how the values for the structural parameters were obtained in the first step. Lastly, we reject the null hypothesis of equal fits if the studentized difference in fits exceeds a standard normal critical value.

We apply the methodology to a popular class of models in the international business cycles literature in order to determine which of the asset market structures used extensively in that literature has the strongest explanatory power for observed empirical regularities. More precisely, we compare three key models: financial autarky, single risk-free bond economy, and an economy with complete asset markets. Our comparison is based on a set of standard data characteristics: variances of key macroeconomic aggregates, such as consumption, investment, labor input, etc.; correlations of these aggregates with output, and their cross-country co-movements. Our procedure recognizes that different data characteristics have different scales (i.e. variances can take any non-negative values, while correlations are restricted to $[-1, 1]$ interval). This makes model comparison based on the equally-weighted aggregation of characteristics problematic. Instead, we propose a data-dependent weighting scheme which allows us to normalize various characteristics by their data counterparts and aggregate them easily. We show that based on both sets of moments (variances

[2] We define a structural model to be misspecified if it cannot predict the population values of the chosen data characteristics for any combination of the deep structural parameters. See Hnatkovska et al. (2012) for details.

and correlations) our test indeed picks financial autarky as the winning specification – consistent with the informal conclusion in Heathcote and Perri (2002).

We then propose a competing model that allows for agent heterogeneity. We focus on a simple dimension of heterogeneity – asset market participation. In our competing model there are two groups of agents in each country: those with access to international and domestic financial markets (participants) and non-participants. We characterize the business cycle properties of the model with limited asset market participation (LAMP) and then apply our test to evaluate the ability of this amended model relative to models with a representative agent in matching the properties of the data. As before, we consider three specifications for international asset markets: financial autarky, single risk-free bond economy, and an economy with complete asset markets, except that in the economy with LAMP these financial regimes apply to participants only. We show that in the setup with LAMP, financial autarky remains a preferred model if the comparison is based on volatilities of key macroeconomic aggregates. However, if the comparison is performed based on co-movements with output and cross-country correlations, then a complete markets economy is chosen as the winner. This is mainly due to the fact that LAMP improves the performance of the model for cross-country correlations: it significantly raises the cross-country correlation in hours of work and investment. Thus, it improves on the "international comovement" puzzle. Adding LAMP also raises the cross-country correlation of output, and lowers the corresponding correlation for consumption, thus bringing the two closer together. Therefore, our models with LAMP also improve on the "quantity" puzzle. Lastly, based on the overall performance (variances and correlations), we find that a complete markets model with LAMP performs no worse than financial autarky and outperforms all other models. In a majority of cases the improvements are statistically significant.

Adding LAMP alters the behavior of a representative agent benchmark in three key ways. First, non-participants are hand-to-mouth consumers whose consumption closely tracks their income. Therefore, LAMP raises the sensitivity of aggregate consumption to income shocks, in line with the data. Second, non-participants' only income is from their labor earnings, making their hours inelastic. Therefore, LAMP reduces the sensitivity of aggregate labor supply to productivity shocks. Third, with consumption responding more and labor supply responding less to shocks, investment becomes less sensitive to productivity shocks. These modifications improve the model's performance in some dimensions (i.e. cross-country correlations, volatility of consumption and its comovement with output), but worsen its performance in other dimensions (i.e. volatilities of output, investment, hours, etc.). Thus, ex-ante, the overall contribution of LAMP is ambiguous. This result highlights the need for a procedure that allows to compare models formally.

We contrast the overall performance of LAMP against the model with a representative agent by aggregating the fits across all three financial regimes. We find that LAMP class of models significantly outperforms the original BKK model class. We verify the robustness of our results to the presence of investment specific technology shocks, with respect to the elasticity of substitution between labor inputs of participants and non-participants, and for various values of the consumption share of non-participants. Overall, our results indicate that adding LAMP to a standard international business cycles model significantly improves its ability to match business cycle facts and can overturn the existing result that financial autarky provides a better fit to the data. To the best of our knowledge, ours is the first paper to perform a statistical model evaluation and comparison based on the agents' heterogeneity over a large set of international business cycle statistics.

We believe that our model with LAMP provides a simple, but empirically important extension of the standard business cycle framework. The fact that only a small fraction of households participate in the stock market has been documented by Mankiw and Zeldes (1991), who showed that only 24% of US households owned equities in 1984; in 2007 this fraction was 51.1% based on the Survey of Consumer Finance.[3] Limited asset market participation has received attention in the theoretical asset pricing literature (see Polkovnichenko, 2004; Vissing-Jorgensen, 2002 and others). Chien et al. (2012) provide its quantitative evaluation and show that a model with LAMP (and incomplete markets) can account for high volatility of equity risk premium in the data. Chien et al. (2011) investigate the implications of LAMP and heterogeneous trading technologies for asset prices and wealth distribution and show that such a model matches well the high volatility of returns and the low volatility of the risk-free rate. Implications of LAMP for monetary policy have been studied by Grossman and Weiss (1983), Chatterjee and Corbae (1992), Alvarez et al. (2002), Bilbiie (2008). They show that LAMP improves model performance for nominal aggregates. van Wincoop (1996) studies the importance of LAMP and borrowing constraints for cross-country consumption correlations and welfare. Kollmann (2012) allows for deviations from the law of one price in the model and shows that LAMP can help resolve the consumption-real exchange rate anomaly (or Backus-Smith puzzle due to Backus and Smith (1993)).

Relative to the above papers, the key contribution of our work is to statistically examine the consequences of LAMP for a large set of business cycle moments as well as formally evaluate its performance relative to alternative models popular in the literature. Our results build the case for LAMP further by showing its importance for international business cycles.[4]

The remainder of the paper is organized as follows. Section 2 presents our model economies. We discuss calibration and model solution in Section 3. Section 4 presents our results, and Section

[3] The share of US households who own equities, while increasing dramatically since 1984, has remained relatively stable at around 50% in the past 15 years, based on the Survey of Consumer Finances. Thus, based on the Survey, the share was 31.6% in 1989, in 1992 – 36.7%, in 1995 – 40.4%, in 1998 – 48.9%, in 2001 – 52.2%, in 2004 – 50.2%, and in 2007 – 51.1%.

[4] The paper also makes methodological contributions by extending Hnatkovska et al. (2012) in two respects. First, we extend the procedure to account for simulation uncertainty. The complexity of DSGE models often makes exact calculations of model predicted moments very cumbersome. In such cases, it is convenient to resort to simulations as we do in this paper. We show how the standard error of the model comparison test statistic can be adjusted to account for simulation uncertainty, which can be used to ensure that no power is lost due to simulations. Second, we propose a class-based test that allows one to compare the overall performances of classes of model with several models in each class. Such an extension is useful when, as in our case, each model has several structurally different versions.

5 concludes. Our testing methodology is described in details in the online appendix to the paper (Gao et al., 2013).[5]

2 Model Economies

To study the role of asset market structure in capturing the properties of international business cycles, we consider a sequence of three economies: an economy in which there are no markets for international asset trades (we refer to it as financial autarky, FA); an economy in which a single non-contingent bond is traded – bond economy, BE; and an economy with complete markets, CM. The structure of these economies follows closely that proposed by Backus et al. (1995, 1994) and studied in Heathcote and Perri (2002). For completeness, we present it here as well. To study the role of investors heterogeneity we extend the three versions of the model to incorporate limited asset market participation. Aside from asset market structure and investors' heterogeneity, all our economies have common structure. We describe it next.

We consider the world consisting of two symmetric countries, H and F, each specializing in the production of its intermediate good. Each country is populated by a continuum of firms and households.

2.1 Firms

Firms are perfectly competitive and reside in two sectors: intermediate-goods sector and final-goods sector. Firms in the intermediate goods sector (i-firms) hire domestically-located capital, k^j, and labor, n^j, $j = \{\text{H,F}\}$, to produce intermediate goods. The i-firms in country H specialize in the production of good a, while i-firms in country F specialize in the production of good b. Period t production by a representative i-firm in country j is

$$F(z_t^j, k_t^j, n_t^j) = e^{z_t^j} \left(k_t^j\right)^{\theta} \left(n_t^j\right)^{1-\theta}, \tag{1}$$

with $\theta > 0$, and z_t^j being the exogenous state of productivity in country j. Let w_t^j and r_t^j denote the real wage and rental rate on capital in country j in period t, measured in terms of the domestic intermediate good. The problem facing i-firms in country j then becomes

$$\max \ F(z_t^j, k_t^j, n_t^j) - w_t^j n_t^j - r_t^j k_t^j,$$

subject to $n_t^j > 0$, $k_t^j > 0$, and equation (1). The intermediate goods produced by H and F i-firms can be freely traded in the international goods markets and can be costlessly transported between countries. Under these conditions, the law of one price must prevail to eliminate arbitrage opportunities. Households, who are the owners of the i-firms, sell their holdings of intermediate

[5] The online appendix is available from the authors' webpages.

goods to domestic final goods producing firms (f-firms), and use the proceeds for consumption, c_t^j and investment, x_t^j. Investment adds to the stock of physical capital available for production next period according to

$$k_{t+1}^j = (1-\delta)k_t^j + x_t^j,$$

where δ is the depreciation rate.

The f-firms are also perfectly competitive and produce final goods from the H and F intermediate goods using constant returns to scale (CRS) technology:

$$G(a_t^j, b_t^j) = \left[\omega^j \left(a_t^j\right)^{\frac{\sigma-1}{\sigma}} + \left(1-\omega^j\right)\left(b_t^j\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \tag{2}$$

where ω^j is the weight that f-firms from country j assigns to the intermediate goods produced in country H. When $\omega^j > 0.5$ there is home bias in the production of final goods in country j. The elasticity of substitution between H and F-produced intermediate goods is $\sigma > 0$. Let $q_{a,t}^j$ and $q_{b,t}^j$ denote the prices of intermediate goods a and b in country j in units of the final good produced in country j. Then, the problem facing f-firms in country j is

$$\max \ G(a_t^j, b_t^j) - q_{a,t}^j a_t^j - q_{b,t}^j b_t^j,$$

subject to $n_t^j > 0$, $k_t^j > 0$, and equation (2).

Productivity in intermediate good sectors is governed by an exogenous process. In particular, we assume that the vector $z_t \equiv [z_t^{\text{H}}, z_t^{\text{F}}]'$ follows an AR(1) process:

$$z_t = \alpha z_{t-1} + e_t, \tag{3}$$

where e_t is a (2×1) vector of independently normally distributed, mean zero shocks with covariance Ω_e.

2.2 Households

Each country is also populated by a continuum of households, whose preferences are defined over consumption and leisure. In particular, the preferences of households in country j are represented by

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t U(c_t^j, 1-n_t^j), \tag{4}$$

where $0 < \beta < 1$ is the discount factor, and $U(.)$ is a concave sub-utility function. Period utility function of the household in country j is given by $U(c_t^j, 1-n_t^j) = \frac{1}{\gamma}\left[\left(c_t^j\right)^{\mu}\left(1-n_t^j\right)^{1-\mu}\right]^{\gamma}$. Households choose consumption, c_t^j, and hours of work, $n_t^j \in [0,1]$, to maximize their lifetime expected utility subject to a sequence of budget constraints, which depend on the financial structure of the model economy. We consider three such structures. Under financial autarky (FA), households can not trade any international financial assets. Under bond economy (BE), households can hold a single non-state contingent internationally traded bond. The third case we consider is that of complete markets (CM). Here households have access to a complete set of Arrow securities. We now describe the budget constraints facing households under each of these different financial structures.

2.2.1 Financial autarky, FA

In the financial autarky, households do not have access to international financial assets. As a result, households consume and invest out of their factor income. The period$-t$ budget constraint of households in country j is

$$c_t^j + x_t^j = q_{a,t}^j \left(w_t^j n_t^j + r_t^j k_t^j \right).$$

Notice that FA rules out the possibility of international borrowing or lending, so neither country can have positive or negative trade balance.

2.2.2 Bond economy, BE

In the bond economy households only trade a single non-state-contingent international bond. We assume that bonds are denominated in the units of intermediate good a. Let B_t^j denote bond holdings of country j households and Q_t be the price of the bonds. Then the period$-t$ budget constraint of households in country j is

$$c_t^j + x_t^j + q_{a,t}^j Q_t B_t^j = q_{a,t}^j \left(w_t^j n_t^j + r_t^j k_t^j \right) + q_{a,t}^j B_{t-1}^j.$$

2.2.3 Complete markets, CM

Following Heathcote and Perri (2002) we assume that households complete the markets by trading in a complete set of Arrow securities denominated in units of intermediate good a. Thus the households' budget constraint can be written as

$$c_t^j + x_t^j + q_{a,t}^j \sum_{s_{t+1}} Q_t(s^t, s_{t+1}) B_t^j(s^t, s_{t+1}) = q_{a,t}^j \left(w_t^j n_t^j + r_t^j k_t^j \right) + q_{a,t}^j B_{t-1}^j(s^{t-1}, s_t),$$

where $s^t = (s_0, s_1, s_2, ..., s_t)$ denotes the entire state history of the economy till date t.

2.2.4 Equilibrium

An equilibrium in this economy consists of a set of goods' prices $\{q_{a,t}^j, q_{b,t}^j\}$, and asset prices (i.e. $\{Q_t\}$ under BE or $\{Q_t(s^t, s_{t+1})\}$ under CM) such that all markets clear when households optimally make their consumption, investment, and asset allocation decisions, taking goods and asset prices as given.

Market clearing in the intermediate goods markets requires

$$a_t^{\text{H}} + a_t^{\text{F}} = F(z_t^{\text{H}}, k_t^{\text{H}}, n_t^{\text{H}}),$$
$$b_t^{\text{H}} + b_t^{\text{F}} = F(z_t^{\text{F}}, k_t^{\text{F}}, n_t^{\text{F}}).$$

Market clearing in the final goods markets requires

$$c_t^j + x_t^j = G(a_t^j, b_t^j), \qquad j = \{\text{H}, \text{F}\}.$$

The market clearing conditions in financial markets vary according to the financial structure of the economy. Under BE, the bond market clearing condition requires

$$0 = B_t^{\text{H}} + B_t^{\text{F}}.$$

Under CM, a similar condition applies for every s_{t+1}:

$$0 = B_t^{\text{H}}(s^t, s_{t+1}) + B_t^{\text{F}}(s^t, s_{t+1}).$$

2.3 Limited asset market participation

Next, we introduce LAMP in our model economy. This feature is used to capture the empirical observation that a large fraction of population does not hold any financial assets. Thus, we assume that each country is populated by two types of households: non-participants and participants. Participants hold all of the capital stock in the economy and can borrow and lend at the international markets (if the model specification allows it). They also supply labor services to the intermediate goods producing firms and make all investment decisions. We assume that there is a fraction λ of such households in each country. Non-participants do not own any capital, do not have access to international markets, and only choose how much time to work and how much to consume. Such behavior may arise due to lack of access to capital markets, lack of knowledge about intertemporal borrowing/lending opportunities, households' myopia, etc. We capture these features in an arguably extreme way. However, we believe that our representation is a simple and parsimonious way to account for the existing empirical evidence on households' consumption-saving behavior: (i) The fact that current income and consumer spending are highly correlated; (ii) The fact that many people have net worth near zero. See Campbell and Mankiw (1989) and Mankiw (2000) for some aggregate evidence. Micro-level evidence can be found in Hall and Mishkin (1982), Shea (1995), Parker (1999), Souleles (1999), Souleles et al. (2006), Agarwal et al. (2007) and others.

The problem facing non-participants (N) is

$$\max \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t U(c_{\text{N},t}^j, 1 - n_{\text{N},t}^j),$$

subject to

$$c^j_{\text{N},t} = q^j_{a,t} w^j_t n^j_{\text{N},t},$$

where subscript N is used to denote the variables pertinent to non-participants. Note that the non-participants' problem remains the same independent of the assumed asset market structure.

The problem facing participants (P) is the same as in the economy with a representative agent:

$$\max \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t U(c^j_{\text{P},t}, 1 - n^j_{\text{P},t}),$$

subject to a budget constraint. Here subscript P is used to denote the variables specific to asset market participants. For participants the exact form of the budget constraint varies with the financial structure of the economy. For instance, the budget constraint of participants in the financial autarky is

$$c^j_{\text{P},t} + x^j_t = q^j_{a,t} \left(w^j_t n^j_{\text{P},t} + r^j_t k^j_t \right)$$

In the bond economy the budget constraint becomes

$$c^j_{\text{P},t} + x^j_t + q^j_{a,t} Q_t B^j_t = q^j_{a,t} \left(w^j_t n^j_{\text{P},t} + r^j_t k^j_t \right) + q^j_{a,t} B^j_{t-1},$$

while under complete markets, it is

$$c^j_{\text{P},t} + x^j_t + q^j_{a,t} \sum_{s_{t+1}} Q_t(s^t, s_{t+1}) B^j_t(s^t, s_{t+1}) = q^j_{a,t} \left(w^j_t n^j_{\text{P},t} + r^j_t k^j_t \right) + q^j_{a,t} B^j_{t-1}(s^{t-1}, s_t).$$

Note that in this case, the asset markets are complete internationally for participants only. The optimization problems solved by i-firms and f-firms remain unchanged.

Aggregate labor input in the economy consists of labor inputs of participants and non-participants and is defined as:

$$n^j_t = \left[\lambda \left(n^j_{\text{P},t} \right)^{\frac{\upsilon-1}{\upsilon}} + (1-\lambda) \left(n^j_{\text{N},t} \right)^{\frac{\upsilon-1}{\upsilon}} \right]^{\frac{\upsilon}{\upsilon-1}},$$

where υ is the elasticity of substitution between the two types of labor.

The market clearing conditions in the goods markets remain the same, while the market clearing conditions in the asset markets apply to participants only.

2.4 Investment-specific technology (IST) shocks

Several recent papers have emphasized the role played by investment-specific technology (IST) shocks in the international business cycles (IBC). In a framework similar to ours, Raffo (2010) shows that IST shocks can help account for a number of puzzles in the business cycles literature. He emphasizes the Backus-Smith puzzle – the fact that consumption and real exchange rate tend to be negatively correlated in the data, while a standard IBC framework predicts the opposite; and the "price" puzzle – the fact that models generate far lower volatility of international relative prices relative to the data. At the same time, Mandelman et al. (2011) show that an IBC model with IST shocks estimated from the data fails to reproduce the moments emphasized in Raffo (2010). Our interest in IST shocks is motivated by their potentially important interactions with LAMP. When only a segment of population has access to capital and asset markets IST shocks will have differential effects on the participants and non-participants, leading to important distributional effects between them. We investigate the role of IST shocks by incorporating them in our models as in Greenwood et al. (2000) and Raffo (2010), but using the properties of these shocks as estimated in Mandelman et al. (2011). In what follows we highlight the new model features introduced by IST shocks.

The problem facing non-participants does not change when IST shocks are introduced. Objective functions of participants and their budget constraints also remain unchanged. In the presence of IST shocks, capital accumulation equation becomes

$$k^j_{t+1} = (1-\delta)k^j_t + e^{v^j_t} x^j_t,$$

where e^{v^j} is the IST shock in country j. As shown in Greenwood et al. (2000), in a competitive equilibrium, e^{-v^j} is interpreted as the relative price of capital goods in terms of consumption goods. We assume that IST shocks, $v_t \equiv [v^{\mathrm{H}}_t, v^{\mathrm{F}}_t]'$ follow an AR(1) process:

$$v_t = \alpha_v v_{t-1} + \zeta_t, \tag{5}$$

where ζ_t is a (2×1) vector of independently normally distributed, mean zero shocks with covariance Ω_ζ. All other model equations remain unchanged.

2.5 Definitions

There are several variables of interest that we define here. Gross domestic product in country j expressed in terms of final consumption goods is given by $y^j_t = q^j_{a,t}F(z^j_t, k^j_t, n^j_t)$. Net exports are $nx^{\mathrm{H}}_t = q^{\mathrm{H}}_{a,t}a^{\mathrm{F}}_t - q^{\mathrm{H}}_{b,t}b^{\mathrm{H}}_t$. Imports ratio for home country is defined following Heathcote and Perri (2002), as the ratio of imports to domestically consumed intermediate goods, both measured at the steady state prices which are symmetric under the benchmark calibration, giving $ir^{\mathrm{H}}_t = b^{\mathrm{H}}_t/a^{\mathrm{H}}_t$. Terms of trade in H country are defined as the price of imports divided by the price of exports, $p^{\mathrm{H}}_t = q^{\mathrm{H}}_{b,t}/q^{\mathrm{H}}_{a,t}$, while the real exchange rate is defined as the relative price of foreign consumption goods to domestic consumption goods, giving $rer^{\mathrm{H}}_t = q^{\mathrm{H}}_{a,t}/q^{\mathrm{F}}_{a,t}$.

3 Calibration, model solution, and econometric methodology

In calibrating the model we assign some parameters their values commonly used in the literature, while we estimate other parameters from the data. Such an approach has become standard in the literature.

In the calibration we consider the world economy as consisting of two countries: country 1 matching the properties of the US economy in quarterly data, and country 2 as the rest of the world. Most of the parameter values are borrowed from Heathcote and Perri (2002). We summarize them in Table 1. We set discount factor to 0.99, which implies annual real interest rate of 4 percent. Risk aversion coefficient is set at 2. As in Heathcote and Perri (2002), we fix consumption share parameter at $\mu = 0.34$. We assume that capital income share, θ is 0.34; and depreciation rate δ of 2.5 percent. Parameter ω, which controls the consumption home bias in household's preferences is set to match the observed import share in the U.S. equal to 15 percent of GDP. We set the elasticity of substitution between domestic and imported intermediate goods at 0.9, which is the value estimated in Heathcote and Perri (2002). This is above the value of this parameter used in Raffo (2010) and Mandelman et al. (2011), but more along the lines of the values used in the IBC literature.[6]

In the model with LAMP a new parameter, λ, is introduced. In the model, $1 - \lambda$ captures the share of nonparticipants, which we calibrate to match the average consumption share of US households who did not hold any equity as reported in the Survey of Consumer Finance. This share is estimated at about 50% of total consumption (see Campbell and Mankiw (1989), Mankiw (2000), Galí et al. (2007), Kollmann (2012)). Therefore, we set $\lambda = 0.5$ in our benchmark model parameterization.[7] We also conduct a sensitivity analysis with respect to this parameter in section 4.3.

The only remaining parameter is υ which equals the elasticity of substitution between labor input of participants and non-participants in the model. For simplicity and given the lack of estimates of this parameter in the literature, we assume that the two types of labor are perfectly substitutable. In what follows we check the robustness of our results with respect to this parameter.

TFP shocks are assumed to be persistent, but temporary. We estimate the process for TFP shocks as in Heathcote and Perri (2002). Namely, we compute productivity sequences for the US and the rest of the world during 1973:1-2007:4 period, where the rest of the world is identified with the aggregate of 21 major trade partners for the U.S..[8] In our estimation, we impose the symmetry restrictions $\rho_{11} = \rho_{22}$ and $\rho_{12} = \rho_{21}$.

Our estimation results for productivity process are presented in Table 2 and they are very similar

[6] For instance, Backus et al. (1995, 1994) use a value of 1.5. Kollmann (2006) uses traded elasticity values as low as 0.6; Chari et al. (2002) and Engel and Matsumoto (2009) use 1.5.

[7] Since the investment rate is 25% in the model, this also implies that the income share of non-participants is equal to 35% of total income, in line with the data.

[8] Details on sample construction and data sources are provided in the Appendix A.1.

Table 1: Benchmark parameter values without estimation step

PREFERENCES		
discount factor	β	0.99
risk-aversion	$1-\gamma$	2
consumption share	μ	0.34
TECHNOLOGY		
capital income share	θ	0.36
depreciation rate	δ	0.025
import share	is (ω)	0.15
elasticity of subst, b/n goods a and b	σ	0.9
share of P households	λ	0.5
IST SHOCKS		
transition matrix	$\alpha^I = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix}$	$\begin{bmatrix} 0.975 & 0.024 \\ 0.024 & 0.975 \end{bmatrix}$
std. dev. of innovations	$\sigma^I_{e_1} = \sigma^I_{e_2}$	0.0066
corr. of innovations	$\sigma^I_{e_1 e_2}$	0.1955

to the estimates in Heathcote and Perri (2002). Namely, our estimates of productivity persistence ρ_{11} and spill-over ρ_{12} are almost the same, while the standard deviation of productivity innovations σ_{e_1} and the correlation between domestic and foreign productivity innovations $\sigma_{e_1 e_2}$ are somewhat smaller than their values.[9]

Table 2: Estimated productivity process

productivity transition matrix	$\alpha = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix}$	$\begin{bmatrix} 0.975 & 0.024 \\ (0.009) & (0.009) \\ 0.024 & 0.975 \\ (0.009) & (0.009) \end{bmatrix}$
std. dev. of productivity innovations	σ_{e_1}	0.0066
	σ_{e_2}	0.0039
corr. of productivity innovations	$\sigma_{e_1 e_2}$	0.1955

Note: Following Heathcote and Perri (2002), we estimate productivity shock process using: $\begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} = \begin{bmatrix} \rho_{11} & \rho_{21} \\ \rho_{12} & \rho_{22} \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$ with the symmetry restriction imposed, $\rho_{11} = \rho_{22}$ and $\rho_{12} = \rho_{21}$. Coefficient estimates and their standard errors are reported in the table.

In calibrating IST shocks, we follow the findings of Mandelman et al. (2011) who show that IST processes for the U.S. and the rest of the world are very persistent and exhibit no spill-overs across countries. Importantly, Mandelman et al. (2011) show that the variance of these shocks is of the same magnitude as the variance of TFP shocks. Motivated by these results, and to facilitate the comparison of the models with and without IST shocks, we assume that IST shocks are fully symmetric to TFP shocks, with no spillovers across the two types of shocks.

Each model is solved by linearizing the sequence of equilibrium conditions and solving the

[9] When simulating the models we use $\sigma_{e_1} = \sigma_{e_2} = 0.0066$.

resulting system of linear difference equations. We derive the second moments of model's variables by simulating the model over 100 periods. The statistics based on which the model comparison is conducted are derived from 10000 simulations. All series, except net exports, are logged and Hodrick-Prescott (HP) filtered with a smoothing parameter of 1600.

For a formal statistical comparison of the considered models, we rely on a Vuong-type (Vuong, 1989) test for potentially misspecified calibrated models proposed in Hnatkovska et al. (2012, 2011). Relative to this work we develop two key extensions. First, we adjust the procedure to account for simulation uncertainty. This becomes important when the model moments can not be computed exactly and instead simulations must be used.[10] Second, 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 below we will ask 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 our proposed class-based test does.

Below, we provide an outline of the econometric framework. A detailed description of the procedure including the aforementioned extensions are provided in the online appendix to the paper (Gao et al., 2013).

Suppose 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 Section 4. The econometrician is interested in comparing between two structural models denoted $f(\theta)$ and $g(\beta)$, where θ and β are the corresponding structural parameters describing consumer's preferences, technology, and etc. Here, $f(\theta)$ and $g(\beta)$ denote the value of h_2 predicted by models f and g, respectively. 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 $h_2 \neq g(\beta)$.

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 for a weight

[10] Using simulations to obtain model implied moments is a common practice in the business cycles literature.

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.

Let $\hat{h}_1$ and $\hat{h}_2$ denote consistent and asymptotically normal estimators of h_1 and h_2, respectively. Recall that the structural parameters θ and β are estimated using only the information in $\hat{h}_1$. 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: $\sqrt{n}(\hat{\theta}-\theta) = A\sqrt{n}(\hat{h}_1 - h_1) + o_p(1)$, where n denotes the sample size, with a similar assumption for $\hat{\beta}$. [11]

In many applications, structural functions f and g are unknown, and the econometrician may resort to simulations in order to estimate them. Let $\hat{f}$ and $\hat{g}$ denote such estimators. Our test is based on the difference between the estimated fits of the two models:

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

The null hypothesis is rejected in favour of model f when $\sqrt{n}S/\hat{\sigma}$ exceeds a standard normal critical value, where $\hat{\sigma}$ denotes an estimator of the asymptotic variance of S. Calculations of the asymptotic variance are discussed in details in the online supplement.

4 Empirical results

In this section we present the findings from the numerical solutions of our models and model comparisons. We conduct model comparisons based on two sets of moments: volatilities of endogenous variables and correlations, which include co-movements of key macroeconomic aggregates with output and cross-country correlations. To perform the comparison, we estimate the corresponding moments in the U.S. quarterly data over the period of 1973:1-2007:4. Details on data sources and calculations are provided in the Appendix A.1. We begin by presenting the results for the BKK and LAMP economies under the benchmark calibration.

4.1 Benchmark case

In this section we present the results from our simulations of BKK and LAMP models under the benchmark parameterization. Table 3 presents the volatilities of various macroeconomic aggregates in the data and in different versions of our models. Thus, panel (a) reports the statistics from the

[11] This specification is satisfied by most estimators used in practice.

original BKK model specification. Panel (b) reports the corresponding statistics in the model with LAMP under perfect substitutability in labor inputs of participants and non-participants.

Table 3: Volatilities: Benchmark calibration

	% std dev	% std dev / % std dev of y			% std dev					
	y	c	x	n	ex	im	nx	ir	p	rx
U.S. Data	1.49	0.62	2.92	0.68	3.93	4.98	0.50	3.84	2.64	3.55
(a) BKK										
FA	0.98	0.54	1.86	0.25	1.07	1.07	0.00	1.37	1.53	1.00
BE	1.01	0.54	2.71	0.29	0.81	0.81	0.20	0.75	0.83	0.54
CM	1.01	0.55	2.73	0.29	0.82	0.82	0.21	0.70	0.78	0.51
(b) LAMP										
FA	0.95	0.59	1.67	0.20	1.03	1.03	0.00	1.31	1.46	0.95
BE	0.97	0.62	2.47	0.23	0.79	0.79	0.19	0.67	0.75	0.49
CM	0.97	0.63	2.49	0.22	0.80	0.80	0.20	0.62	0.69	0.45

Note: This Table presents actual and simulated percent standard deviations for the U.S. economy. The data statistics are for the period of 1973:1-2007:4. Details on the data are available in the Appendix A.1. Model-based statistics are obtained from 10000 simulations, 100 periods long, each. All series, except net exports (nx), are logged and HP-filtered. The following models are considered: (a) original BKK; (b) BKK with LAMP. FA, BE and CM refer, respectively, to financial autarky, bond economy and complete markets economy.

As in Heathcote and Perri (2002), financial autarky model generates significantly higher volatilities of exports, imports and especially relative prices, in comparison with the complete markets and bond economies; but implies lower volatilities of output, consumption, investment and employment relative to bond economy and complete markets economy. These results are driven by the inability of agents in the environment of financial autarky to run trade imbalances. In such a framework, following productivity shocks, it is impossible to shift final goods production to the country that has comparative advantage in doing so. As a result, a larger adjustment in relative prices, such as terms of trade, is needed to clear the markets. Such larger movements in the terms of trade under financial autarky partially offset the productivity changes (as in Cole and Obstfeld, 1991), thus reducing the incentives to work and invest. Consequently, employment, investment, output and consumption all become less volatile when no access to financial assets is available.

When agents become heterogeneous in terms of their access to financial instruments, there are two key changes in the volatility characteristics of our economies. First, volatility of consumption increases across all financial regimes; second, the volatility of all other variables declines across all financial regimes. In our setup, introducing LAMP implies that asset markets become incomplete within a country. Namely, the non-participants can not trade any assets (neither financial, nor real, like capital) and only consume their labor income. Their consumption, as a result becomes more volatile, thus raising the volatility of aggregate consumption in the country. On the other hand, employment is the only source of income for non-participants, as a result, their labor supply is inelastic. This implies that aggregate employment, output and investment, all become less volatile relative to the economy with no LAMP.

Next, we evaluate the performance of our model in terms of co-movements with output. The results are summarized in Table 4. As before, the top row of the table reports the co-movements in the data, while panels (a) and (b) report them, respectively, in the original BKK model and in the economy with LAMP.

Table 4: Correlations with output: Benchmark calibration

	correlation between								
	c, y	x, y	n, y	ex, y	im, y	nx, y	p, y	rx, y	$rx, c_1 - c_2$
U.S. Data	0.82	0.94	0.85	0.42	0.82	-0.37	-0.16	0.16	-0.17
(a) BKK									
FA	0.89	0.99	0.98	1.00	0.07	0.01	0.64	0.64	0.96
BE	0.93	0.95	0.96	0.55	0.80	-0.65	0.64	0.64	0.99
CM	0.94	0.95	0.96	0.50	0.84	-0.65	0.64	0.64	0.99
(b) LAMP									
FA	0.93	0.99	0.99	1.00	0.08	0.01	0.64	0.64	0.99
BE	0.97	0.95	0.97	0.54	0.83	-0.64	0.63	0.63	0.97
CM	0.97	0.95	0.97	0.48	0.87	-0.64	0.62	0.62	0.97

Note: See notes to Table 3.

As was the case for volatilities, the financial autarky economy is the most distinct among our three financial regimes. The fact that all trades in this economy must be *quid pro quo* implies that net exports are acyclical. Financial autarky also generates more procyclical exports and less procyclical imports relative to the bond and complete markets economies. In terms of these co-movements financial autarky economy departs from the data relative to the other two financial regimes. When LAMP is introduced, the comovement properties of the model do not change much. The only exception is the comovement of consumption with output, which increases when LAMP is introduced. The main reason is again the behavior of non-participants, whose work hours are inelastic, which in turn makes their wage income and thus consumption more sensitive to productivity changes. Consumption of non-participants, therefore, is more strongly procyclical than consumption of participants. This makes aggregate consumption move more closely with output relative to the original BKK framework. Both BKK and LAMP economies fail to replicate the negative correlation between real exchange rate and relative consumption of domestic to foreign economies that is observed in the data. This mismatch of theory and data is a well-known Backus-Smith puzzle due to Backus and Smith (1993) and Kollmann (1996). Adding LAMP reduces this correlation, but only marginally.

Lastly, we summarize the model performance based on cross-country co-movements of various macroeconomic aggregates. Table 5 reports our results. The top row reports the estimates in the data, the second panel summarizes them in the BKK economies, and the bottom panel - in the economies with LAMP. There are several puzzles associated with the cross-country correlations, and they can be seen clearly from Table 5. First, is the fact that consumption is less correlated than output across countries in the data, while models predict the opposite ("quantity" puzzle). Second, in the data the correlations of investment and employment across countries are positive, while complete markets and bond economy models predict negative correlations ("international comovement" puzzle). Financial autarky, on the other hand, generates investment and employment across countries that are positively correlated, consistent with the data. So, as was the case with volatilities, financial autarky model seems to provide a better match to the data even when it comes to the cross-country co-movements.

Table 5: Cross-country correlations: Benchmark calibration

	correlation between			
	y_1, y_2	c_1, c_2	x_1, x_2	n_1, n_2
U.S. Data	0.58	0.43	0.41	0.45
(a) BKK				
FA	0.16	0.86	0.29	0.01
BE	0.08	0.68	-0.42	-0.32
CM	0.09	0.64	-0.43	-0.29
(b) LAMP				
FA	0.17	0.79	0.37	0.10
BE	0.12	0.57	-0.39	-0.22
CM	0.13	0.53	-0.40	-0.18

Note: See notes to Table 3.

Adding agents' heterogeneity in asset market access works towards resolving these puzzles. In particular, LAMP reduces the cross-country correlation of consumption, while simultaneously increasing it for output; and does so for all three financial regimes considered. It also significantly increases the cross-country correlation in investment and employment.

To understand these results, consider what happens to employment, investment, consumption and output in the economy with a representative households following a positive productivity shock. The country experiencing a productivity improvement (say, home country) sees its real wages rise, leading to an increase in labor supply, output and investment. At the same time, following the shock, the terms of trade depreciate in the home country, thus making foreign households relatively wealthier.[12] As a result, they reduce their labor supply, lowering real output. For consumption in the foreign country to go up, investment must fall. When markets are complete or a single non-contingent bond is available these adjustments imply a negative correlation of employment and investment between home and foreign economies. In the financial autarky, where shifting production across countries is not an option, terms of trade must adjust to eliminate the incentives to do so. These terms of trade movements are larger than in the bond or complete market economies as was argued before. By offsetting some of the productivity improvement in the home country, terms of trade adjustment implies that output, consumption, investment and employment in this country increase by less under financial autarky than under bond or complete market regimes. Correspondingly, in the foreign country, these macroeconomic aggregates increase by more as foreign households take advantage of larger favorable terms of trade movements. These adjustments imply positive cross-country correlations under financial autarky.

Adding LAMP changes these dynamics. With LAMP non-participants allocate all their time endowment to work. Thus, only households participating in the asset and capital markets adjust their labor supply following the shock. Consequently, aggregate labor supply in both countries

[12] There are several channels through which wealth effect in the foreign country arises following productivity improvement in the home country. First is the fact that productivity shocks spill over across countries. Second, is the terms of trade effect mentioned in the text. Third effect works through the world interest rate (whenever any assets are traded across countries). In particular, interest rate in the country experiencing a productivity improvement rises, creating an additional positive wealth effect for foreign households, who want to lend following the shock.

responds to shocks less relative to the economy with a representative agent. This results in larger cross-country correlation of hours and output in the LAMP economy.[13] In the bond and complete market economies this reduces the incentives to shift production across countries following the shocks and increases the cross-country correlation in investment. With investment responding less, so does consumption, thus lowering consumption correlation across countries. This result highlights how the absence of risk-sharing within a country spills into lower international risk-sharing.

The results above show that different versions of our model perform better in matching different data characteristics. Financial autarky economy does best in matching volatilities of macroeconomic aggregates, but can not account for the cyclical properties of trade variables. Complete markets and bond economies do better in accounting for the cyclical properties of the data, but under-perform in terms of volatilities and cross-country correlations. Adding LAMP has three key effects in the model: (i) it raises the sensitivity of aggregate consumption to income shocks; (ii) it reduces the sensitivity of aggregate labor supply to productivity shocks; (iii) it makes investment less responsive to productivity shocks. These effects improve models performance primarily in matching cross-country correlations of consumption, output, investment and employment, but worsen their performance in matching volatilities. Given these results, a formal statistical test is necessary to aggregate various characteristics and pick a winner among our model variants. We turn to this next.

4.2 Comparison results

To determine which version of the model described above provides the best fit to the data, we apply our test and its extensions described in Gao et al. (2013) to all possible pair-wise model comparisons. Our null hypothesis is that any two models considered provide equivalent fit to the data. We evaluate each model's performance based on three criteria: (i) its ability to match volatilities; (ii) its ability to match co-movements with output and cross-country correlations; and (iii) on its overall performance which aggregates all of the aforementioned characteristics. Aggregation of model characteristics is equivalent to choosing the weight matrix W_{h_2} defined above and deserves a special note. The simplest approach would be to assign equal weights to all model characteristics, that is to use an identity weighting matrix. Such an approach, however, may not be very informative if different data characteristics have significantly different scales. For instance, in our case, variances can take any non-negative values, while correlations are restricted to $[-1, 1]$ interval. Thus, if we use an identity weighting matrix to aggregate across such variances and correlations, the overall model performance will be heavily influenced by its performance for the moments that are larger - variances in our case.

[13] While adding LAMP moderates the dynamics of hours, it is quantitatively distinct from lowering labor supply elasticity in the representative agent BKK economy. In fact, our experiments show that reducing labor supply elasticity in the BKK economy leads to a larger negative correlation of labor inputs and investment across countries; as well as larger positive cross-country correlation of consumption. This deepens the puzzles.

To account for the differences in scale we utilize a data-dependent weighting scheme in which various characteristics of interest are brought to a common base, thus facilitating their aggregation. According to our weighting scheme, prediction errors are assigned weights that are inversely related to the values of corresponding moments in the data. This way, instead of measuring absolute distance between the model and data as with the identity matrix, we measure the percentage error made by the model relative to the data. In other words, we compute the ratio of the distance between the model and data over the data moment, $(h - f)/h$ where f is the model-predicted moment and h is its data counterpart. Such a weighting scheme allows us to construct a scale-free measure of fit.[14] We use the aforementioned percentage errors in the case of volatilities. Since correlations are unit-free, we aggregate them using simple prediction errors.[15]

The comparison results for BKK and LAMP models with various financial structures are presented in Table 6. Three panels in the table identify the set of characteristics based on which we conduct the comparisons: variances (panel (a)), co-movements with output and cross-country correlations (panel (b)), and overall performance (panel (c)). The test statistic is computed as the difference between the loss function of the model in the row (model g) and the loss function of the model in the column (model f). Therefore, a positive sign of the test statistic implies that the model in the row does worse in matching data moments as compared with the model in the column. In addition, the larger the test statistic, the worse the model in the row performs. We report p-values in parenthesis below the test statistics.

First, consider the original BKK models. Among the three financial regimes our test picks financial autarky as the winning specification based on volatilities, correlations and the overall performance. This result is in accord with the informal findings in the literature. Extending comparisons to include LAMP economies, our test results show that financial autarky with no LAMP outperforms all other models based on volatilities. Its superior performance for volatilities is statistically significant in all five pair-wise comparisons. Based on co-movements with output and cross-country correlations, our test picks complete market economy with LAMP as the model that matches data best. This result is also statistically significant in three pair-wise comparisons out of five possible. Finally, based on the overall performance, complete markets economy with LAMP outperforms all other models, with most comparisons being highly statistically significant. The only exception is autarky economies, but the difference in their overall performance is not statistically significant.

[14] To bring the variances to a (0,1) base that is more comparable with the correlations, we apply a logistic re-scaling to them.

[15] Another possibility is to use a weight matrix that is inversely related to the asymptotic variance-covariance matrix of data moments. Such an approach gives a scale-free measure of fit which is reminiscent of the GMM approach, i.e. moments that are more precisely estimated are assigned greater weights. Note, however, that in the time-series context where the asymptotic variance-covariance matrix is estimated by HAC methods, the uncertainty in the estimation of the weight matrix will dominate the uncertainty in estimation of parameters and moments. This is because HAC estimators converge at a slower than square root-n rate (see, for example, Hall and Inoue (2003) for details). As a result, such a test may have poor power in finite samples.

Table 6: Test results from benchmark models comparisons

Model g	Model f BKK FA	BKK BE	BKK CM	LAMP FA	LAMP BE	LAMP CM
(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	0.03***	-0.16***	-0.17***	0		
	(0.00)	(0.00)	(0.00)			
LAMP, BE	0.23***	0.04***	0.03***	0.20***	0	
	(0.00)	(0.00)	(0.00)	(0.00)		
LAMP, CM	0.25***	0.06***	0.05***	0.22***	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	-0.12	-0.25	-0.17	0		
	(0.40)	(0.65)	(0.76)			
LAMP, BE	-0.22	-0.35***	-0.27**	-0.10	0	
	(0.57)	(0.00)	(0.03)	(0.84)		
LAMP, CM	-0.31	-0.45***	-0.36***	-0.19	-0.09***	0
	(0.42)	(0.00)	(0.01)	(0.68)	(0.00)	
(c) Overall	FA	BE	CM	FA	BE	CM
BKK, FA	0					
BKK, BE	0.32	0				
	(0.47)					
BKK, CM	0.25	-0.07***	0			
	(0.57)	(0.00)				
LAMP, FA	-0.09	-0.41	-0.34	0		
	(0.52)	(0.46)	(0.54)			
LAMP, BE	0.01	-0.31***	-0.24**	0.10	0	
	(0.97)	(0.01)	(0.05)	(0.83)		
LAMP, CM	-0.06	-0.38***	-0.31**	0.03	-0.08***	0
	(0.87)	(0.01)	(0.02)	(0.95)	(0.00)	
LAMP - BKK			**-0.71***			
class comparison			**(0.07)**			

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.

The pair-wise model comparisons discussed above are informative in isolating the combinations of model features that produce the closest fit to the data (e.g. LAMP and complete markets, as above). But does the asset participation margin improve model performance independent of the assumed international financial regime? To provide such an evaluation we apply our class-based test presented in Gao et al. (2013). The test evaluates the overall performance of LAMP class of models relative to a representative agent BKK class of models by aggregating the fits across the three financial regimes within each class. We test whether these two model classes have the same distance from the data moments. We find that LAMP specification provides a better fit to the data relative to a representative agent benchmark, and the difference is statistically significant. More precisely, the resulting test statistic, reported at the bottom of Table 6, is -0.71 in favor of LAMP with the standard error of 0.39 and the resulting p-value of 0.07.

4.3 Robustness

Next we consider the robustness of our results with respect to parameter υ, to the presence of IST shocks, and to alternative values of parameter λ. The results under alternative calibrations are presented in Table 7 for volatilities, Table 8 for correlations with output, and Table 9 for cross-country correlations. In each exercise we change only the parameter of interest and keep all remaining parameters unchanged.

Table 7: Volatilities: Robustness

	% std dev	% std dev / % std dev of y			% std dev					
	y	c	x	n	ex	im	nx	ir	p	rx
U.S. Data	1.49	0.62	2.92	0.68	3.93	4.98	0.50	3.84	2.64	3.55
(a) LAMP, $\upsilon = 0.5$										
FA	0.90	0.65	1.49	0.13	0.98	0.98	0.00	1.24	1.38	0.90
BE	0.92	0.69	2.25	0.15	0.76	0.76	0.18	0.60	0.67	0.44
CM	0.91	0.70	2.27	0.15	0.78	0.78	0.19	0.55	0.61	0.40
(b) BKK with IST										
FA	1.00	0.61	2.27	0.38	1.09	1.09	0.00	1.41	1.57	1.02
BE	1.04	0.57	3.86	0.45	1.42	1.43	0.33	1.03	1.15	0.75
CM	1.04	0.57	3.90	0.45	1.48	1.49	0.35	0.89	1.15	0.75
(c) LAMP with IST										
FA	0.96	0.63	1.97	0.30	1.04	1.04	0.00	1.34	1.48	0.97
BE	0.99	0.62	3.48	0.36	1.44	1.44	0.33	1.04	1.16	0.76
CM	0.98	0.63	3.52	0.35	1.50	1.50	0.34	0.68	1.17	0.77
(d) LAMP, $\lambda = 0.7$										
FA	0.96	0.56	1.76	0.22	1.05	1.05	0.00	1.34	1.49	0.97
BE	0.99	0.58	2.59	0.26	0.80	0.80	0.19	0.71	0.79	0.51
CM	0.99	0.59	2.61	0.25	0.81	0.81	0.20	0.66	0.73	0.48
(e) LAMP, $\lambda = 0.3$										
FA	0.92	0.64	1.53	0.16	1.01	1.00	0.00	1.27	1.41	0.92
BE	0.94	0.67	2.30	0.19	0.78	0.78	0.19	0.62	0.69	0.45
CM	0.94	0.69	2.32	0.18	0.80	0.80	0.20	0.57	0.63	0.41

Note: This Table presents actual and simulated volatilities for the U.S. economy. All data statistics are for the period of 1973:1-2007:4. Details on the data are available in the Appendix A.1. Model-based statistics are obtained from 10000 simulations, 100 periods long, each. All series, except net exports (nx), are logged and HP-filtered. The following models are considered: (a) LAMP with imperfectly substitutable labor input of participants and non-participants; (b) BKK with IST shocks; (c) LAMP with IST shocks; (d) LAMP with $\lambda = 0.7$; (e) LAMP with $\lambda = 0.3$. FA, BE and CM refer, respectively, to financial autarky, bond economy and complete markets economy.

Consider first the scenario where labor inputs of participants and non-participants are imperfectly substitutable with elasticity $\upsilon = 0.5$. These results are presented in panel (a) of each table. In this case, the distinction between the original BKK and LAMP models becomes quantitatively sharper. In particular, relative to the case of perfect substitutability between two labor types reported in panel (b) of Tables 3, 4, and 5, volatility of consumption rises further, while volatilities of all other aggregates fall. Reducing elasticity of substitution in labor has the largest effect on cross-country correlations. In particular, it significantly lowers cross-correlation of consumption, and raises the cross-correlation of employment and investment. These changes are reflected in the formal model comparison. We find that while qualitatively, our test results remain unchanged, quantitatively they become stronger and more significant.[16]

[16] Given that test results do not change qualitatively in this case, we do not report them in the paper. These

Table 8: Correlations with output: Robustness

	correlation between								
	c, y	x, y	n, y	ex, y	im, y	nx, y	p, y	rx, y	$rx, c_1 - c_2$
U.S. Data	0.82	0.94	0.85	0.42	0.82	-0.37	-0.16	0.16	-0.17
(a) LAMP, $\upsilon = 0.5$									
FA	0.95	0.99	0.99	1.00	0.10	0.01	0.63	0.63	1.00
BE	0.98	0.95	0.97	0.52	0.86	-0.63	0.61	0.61	0.95
CM	0.99	0.95	0.98	0.46	0.89	-0.63	0.61	0.61	0.95
(b) BKK with IST									
FA	0.69	0.89	0.75	1.00	0.06	0.02	0.65	0.65	0.26
BE	0.78	0.78	0.74	0.17	0.59	-0.52	0.35	0.35	0.98
CM	0.81	0.78	0.74	0.13	0.60	-0.52	0.30	0.30	0.95
(c) LAMP with IST									
FA	0.80	0.90	0.74	1.00	0.08	0.02	0.64	0.64	0.55
BE	0.90	0.77	0.72	0.18	0.56	-0.48	0.31	0.31	0.78
CM	0.92	0.76	0.72	0.15	0.57	-0.48	0.26	0.26	0.67
(d) LAMP, $\lambda = 0.7$									
FA	0.91	0.99	0.99	1.00	0.08	0.03	0.64	0.64	0.98
BE	0.95	0.95	0.96	0.55	0.82	-0.64	0.63	0.63	0.98
CM	0.96	0.95	0.96	0.49	0.85	-0.64	0.63	0.63	0.98
(e) LAMP, $\lambda = 0.3$									
FA	0.94	0.99	0.99	1.00	0.10	0.03	0.63	0.63	1.00
BE	0.98	0.95	0.97	0.52	0.85	-0.63	0.61	0.62	0.95
CM	0.99	0.95	0.97	0.46	0.89	-0.63	0.61	0.61	0.95

Note: See notes to Table 7.

Next, we turn to IST shocks. The simulated moments for the original BKK models with IST shocks are shown in panel (b) of Tables 7, 8, 9; while those for the LAMP model with IST shocks are in panel (c) of the same three tables. Not surprisingly, when IST shocks are introduced, all volatilities go up, especially for investment, international trade variables and relative prices. This increase is particularly pronounced in the bond economy and complete markets economy. Correlations with output, on the other hand, decline. Cross-country correlations of output and consumption also fall, while those of investment and employment turn more negative. These changes are characteristic of both BKK and LAMP economies.

What is behind these results? As in Raffo (2010) and Mandelman et al. (2011), IST shocks in our setup act as demand shocks. For instance, consider a positive IST shock in the domestic economy. Following this shock, domestic investment demand goes up, appreciating home terms of trade, on impact. To accommodate higher investment demand, domestic households must reduce their consumption. In bond and complete market economies, imports from abroad also rise to finance domestic investment boom, leading to trade deficit. Domestic households also increase their labor supply in response to the shock. As home output goes up and investment demand subsides (with temporary IST shocks), domestic terms of trade begin to depreciate. So does the real exchange rate. The impact appreciation of the terms of trade and real exchange rate, followed by depreciation some quarters later helps understand the higher volatility of these variables in the economy with IST shocks.

Foreign economy, on the other hand, being relatively less productive, cuts down its investment and employment. Released resources are used for temporarily higher consumption by foreign

results are available from the authors upon request.

Table 9: Cross-country correlations: Robustness

	correlation between			
	y_1, y_2	c_1, c_2	x_1, x_2	n_1, n_2
U.S. Data	0.58	0.43	0.41	0.45
(a) LAMP, $\upsilon = 0.5$				
FA	0.18	0.73	0.44	0.18
BE	0.15	0.48	-0.38	-0.12
CM	0.15	0.44	-0.38	-0.06
(b) BKK with IST				
FA	0.14	0.54	0.11	-0.17
BE	0.05	0.59	-0.64	-0.48
CM	0.06	0.57	-0.64	-0.45
(c) LAMP with IST				
FA	0.17	0.63	0.21	-0.08
BE	0.09	0.56	-0.62	-0.39
CM	0.10	0.53	-0.62	-0.36
(d) LAMP, $\lambda = 0.7$				
FA	0.17	0.83	0.33	0.06
BE	0.11	0.62	-0.40	-0.27
CM	0.11	0.58	-0.41	-0.23
(e) LAMP, $\lambda = 0.3$				
FA	0.19	0.75	0.43	0.17
BE	0.14	0.50	-0.38	-0.14
CM	0.15	0.46	-0.39	-0.08

Note: See notes to Table 7.

households. These dynamics imply low (for consumption) or negative (for output, employment and investment) cross-country correlations after IST shocks.

Overall, adding temporary IST shocks to our benchmark economies helps improve their performance on some dimensions, such as volatilities and some correlations. However, the models fit also worsens in some other dimensions, such as cross-country co-movements of investment and employment. As a result, a formal statistical method of model comparison is again warranted. Our results from comparison of models with IST shocks are presented in Table 10, where as before, to measure overall performance we aggregate variances and covariances using data-dependent weighting matrix.

In the presence of IST shocks our test picks BKK bond economy as the preferred model specification among all pair-wise comparisons when the objective is to match volatilities. The results are statistically significant in all but one pair. If the objective is to match correlations, out test implies that BKK autarky with IST shocks comes out at the top. When the overall performance (variances and correlations) is considered, BKK autarky economy with IST shocks is chosen as the winner, although this superior performance is statistically significant in only two out of five possible pairs.

Turning to the comparison between BKK and LAMP models classes, we find that LAMP with IST shocks outperforms the original BKK representative agent model with IST shocks and that this superior performance is highly statistically significant. More precisely, the test statistic for the overall test between LAMP and BKK model classes is -1.34 in favor of LAMP, with the standard error of 0.29 and implied p-value of 0.00 (see the bottom of Table 10). These results imply that also in the presence of IST shocks, LAMP delivers a better match to the data.

Lastly, we investigate the robustness of our results with respect to parameter λ which cap-

Table 10: Test results from the comparison of models with IST shocks

Model g	Model f					
		BKK			LAMP	
(a) Volatilities	FA	BE	CM	FA	BE	CM
BKK, FA	0					
BKK, BE	-0.01	0				
	(0.65)					
BKK, CM	-0.00	0.01**	0			
	(0.89)	(0.03)				
LAMP, FA	0.03***	0.05*	0.04	0		
	(0.00)	(0.09)	(0.14)			
LAMP, BE	-0.01	0.01*	-0.00	-0.04	0	
	(0.83)	(0.10)	(0.66)	(0.18)		
LAMP, CM	0.03	0.04***	0.03***	-0.01	0.03***	0
	(0.29)	(0.00)	(0.00)	(0.86)	(0.00)	
(b) Correlations (with output and cross-country)						
BKK, FA	0					
BKK, BE	1.19***	0				
	(0.00)					
BKK, CM	1.02**	-0.17***	0			
	(0.02)	(0.00)				
LAMP, FA	0.12	-1.08**	-0.90*	0		
	(0.54)	(0.03)	(0.07)			
LAMP, BE	0.48	-0.71***	-0.54***	0.37	0	
	(0.20)	(0.00)	(0.00)	(0.42)		
LAMP, CM	0.21	-0.99***	-0.81***	0.09	-0.28***	0
	(0.60)	(0.00)	(0.00)	(0.85)	(0.00)	
(c) Overall						
BKK, FA	0					
BKK, BE	1.18***	0				
	(0.00)					
BKK, CM	1.02***	-0.17***	0			
	(0.01)	(0.00)				
LAMP, FA	0.15	-1.03**	-0.87*	0		
	(0.43)	(0.03)	(0.08)			
LAMP, BE	0.48	-0.70***	-0.54***	0.33	0	
	(0.20)	(0.00)	(0.00)	(0.47)		
LAMP, CM	0.23	-0.95***	-0.78***	0.08	-0.24***	0
	(0.54)	(0.00)	(0.00)	(0.86)	(0.00)	
LAMP - BKK			**-1.34*****			
class comparison			**(0.00)**			

Note: See notes to Table 6.

tures the share of asset market participants in the economy. One may argue that this share is larger than 0.5 that we assumed in our benchmark calibration since even those individuals who do not participate in the stock markets are likely to have a bank account, a credit card, or receive government transfers. So, we consider a calibration in which $\lambda = 0.7$. In this case, both the consumption and income shares of participants are significantly higher than the corresponding shares for non-participants. The results under this calibration are presented in panel (d) of Tables 7, 8, 9. Notice that when the share of participants increases, the LAMP economies begin to look more like the BKK economies. We confirm this using our formal statistical comparison and find that LAMP economies still outperform the BKK economies, although the difference is smaller (in absolute value), as summarized by the test statistic for the class-based test (not reported). In particular, we find that the test statistic is now equal to -0.41, and is statistically significant at 5% level. These results and pair-wise model comparisons are available in the online appendix to the paper. An alternative interpretation of the parameter λ is that it captures the share of participants in the *international* asset markets. In this case, given the robust empirical evidence on portfolio home bias (see, for instance, Lewis (1999) for a survey of the literature on home bias in equities), the value of λ should be lower. We set it equal to 0.3 and recompute the models dynamics. The moments obtained under this calibration of the LAMP economies are presented in panel (e) of Tables 7, 8, 9. It is easy to see that the differences with the BKK specifications grow relative to the benchmark calibration. This is neatly summarized by a class-based test statistic which we find to be equal to -1.18, with the p-value of 0.09. Thus, our finding that LAMP model can replicate international business cycle dynamics better than the BKK model remains robust to alternative parameterizations of the share of asset market participants.

5 Conclusion

In this paper we propose a novel statistical test to conduct evaluation and formal comparison of DSGE models. Our procedure explicitly accounts for the possibility that a DSGE model might be misspecified. It also accounts for simulation uncertainty, the fact that some model parameters are estimated rather than calibrated, and allows for both pair-wise comparison of models and comparison of model classes. We apply our test to a standard international business cycles model with three specifications for asset markets structure: financial autarky, single risk-free bond economy, and an economy with complete asset markets. We find that financial autarky economy indeed fits the data best, in line with the informal findings in the literature. We then allow for domestic asset market incompleteness by introducing hand-to-mouth consumers that do not participate in the domestic or foreign financial markets. With limited asset market participation (LAMP), the models' performance is improved in matching cross-country correlations, but worsened in matching volatilities. Formal statistical comparison finds that the improvements brought out by LAMP are statistically significant, allowing economies with LAMP to outperform the representative agent benchmark economies. The superior performance of LAMP is robust to lower substitutability in labor inputs of participants and non-participants, to the presence of the investment-specific productivity shocks, and to variations in the share of asset market participants.

References

Agarwal, S., C. Liu, and N. S. Souleles (2007, December). The reaction of consumer spending and debt to tax rebates-evidence from consumer credit data. *Journal of Political Economy 115*(6), 986–1019.

Alvarez, F., A. Atkeson, and P. J. Kehoe (2002, February). Money, interest rates, and exchange rates with endogenously segmented markets. *Journal of Political Economy 110*(1), 73–112.

Arvanitis, A. V. and A. Mikkola (1996). Asset-market structure and international trade dynamics. *American Economic Review 86*(2), 67–70.

Backus, D., P. J. Kehoe, and F. E. Kydland (1995). International business cycles: Theory and evidence. In T. F. Cooley (Ed.), *Frontiers of Business Cycle Research*, pp. 331–356. Princeton: Princeton University Press.

Backus, D. K., P. J. Kehoe, and F. E. Kydland (1994). Dynamics of the trade balance and the terms of trade: The J-curve? *American Economic Review 84*(1), 84–103.

Backus, D. K. and G. W. Smith (1993). Consumption and real exchange rates in dynamic economies with non-traded goods. *Journal of International Economics 35*(3-4), 297–316.

Baxter, M. (1995). International trade and business cycles. In G. M. Grossman and K. Rogoff (Eds.), *Handbook of International Economics* (1 ed.), Volume 3, Chapter 35, pp. 1801–1864. Elsevier.

Baxter, M. and M. J. Crucini (1995). Business cycles and the asset structure of foreign trade. *International Economic Review 36*(4), 821–854.

Bilbiie, F. O. (2008, May). Limited asset markets participation, monetary policy and (inverted) aggregate demand logic. *Journal of Economic Theory 140*(1), 162–196.

Campbell, J. Y. and N. G. Mankiw (1989). Consumption, income and interest rates: Reinterpreting the time series evidence. In *NBER Macroeconomics Annual 1989, Volume 4*, NBER Chapters, pp. 185–246. National Bureau of Economic Research, Inc.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2002). Can sticky price models generate volatile and persistent real exchange rates? *Review of Economic Studies 69*(3), 533–63.

Chatterjee, S. and D. Corbae (1992, June). Endogenous market participation and the general equilibrium value of money. *Journal of Political Economy 100*(3), 615–46.

Chien, Y., H. Cole, and H. Lustig (2011). A multiplier approach to understanding the macro implications of household finance. *Review of Economic Studies 78*(1), 199–234.

Chien, Y., H. Cole, and H. Lustig (2012, October). Is the volatility of the market price of risk due to intermittent portfolio rebalancing? *American Economic Review 102*(6), 2859–96.

Cole, H. L. and M. Obstfeld (1991, August). Commodity trade and international risk sharing : How much do financial markets matter? *Journal of Monetary Economics 28*(1), 3–24.

Corsetti, G., L. Dedola, and S. Leduc (2008). International risk sharing and the transmission of productivity shocks. *Review of Economic Studies 75*(2), 443–473.

Engel, C. and A. Matsumoto (2009, July). The international diversification puzzle when goods prices are sticky: It's really about exchange-rate hedging, not equity portfolios. *American Economic Journal: Macroeconomics 1*(2), 155–88.

Galí, J., J. D. López-Salido, and J. Vallés (2007, 03). Understanding the effects of government spending on consumption. *Journal of the European Economic Association 5*(1), 227–270.

Gao, X., V. Hnatkovska, and V. Marmer (2013). Supplement to limited participation in international business cycle models: A formal evaluation. Working Paper, University of British Columbia.

Greenwood, J., Z. Hercowitz, and P. Krusell (2000, January). The role of investment-specific technological change in the business cycle. *European Economic Review 44*(1), 91–115.

Grossman, S. and L. Weiss (1983, December). A transactions-based model of the monetary transmission mechanism. *American Economic Review 73*(5), 871–80.

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*(2), 361–394.

Hall, R. E. and F. S. Mishkin (1982, March). The sensitivity of consumption to transitory income: Estimates from panel data on households. *Econometrica 50*(2), 461–81.

Heathcote, J. and F. Perri (2002). Financial autarky and international business cycles. *Journal of Monetary Economics 49*(3), 601–627.

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.

Kollmann, R. (1996). Incomplete asset markets and the cross-country consumption correlation puzzle. *Journal of Economic Dynamics and Control 20*(5), 945–961.

Kollmann, R. (2006, February). International portfolio equilibrium and the current account. CEPR Discussion Papers 5512, C.E.P.R. Discussion Papers.

Kollmann, R. (2012, May). Limited asset market participation and the consumption-real exchange rate anomaly. *Canadian Journal of Economics 45*(2), 566–584.

Lewis, K. (1999). Trying to explain home bias in equities and consumption. *Journal Of Economic Literature 37*, 571–608.

Mandelman, F., P. Rabanal, J. F. Rubio-Ramirez, and D. Vilan (2011, January). Investment specific technology shocks and international business cycles: An empirical assessment. *Review of Economic Dynamics 14*(1), 136–155.

Mankiw, N. G. (2000, May). The savers-spenders theory of fiscal policy. *American Economic Review 90*(2), 120–125.

Mankiw, N. G. and S. P. Zeldes (1991). The consumption of stockholders and nonstockholders. *Journal of Financial Economics 29*(1), 97–112.

Parker, J. A. (1999, September). The reaction of household consumption to predictable changes in social security taxes. *American Economic Review 89*(4), 959–973.

Polkovnichenko, V. (2004). Limited stock market participation and the equity premium. *Finance Research Letters 1*(1), 24 – 34.

Raffo, A. (2010). Technology shocks: novel implications for international business cycles. International Finance Discussion Papers 992, Board of Governors of the Federal Reserve System.

Shea, J. (1995, August). Myopia, liquidity constraints, and aggregate consumption: A simple test. *Journal of Money, Credit and Banking 27*(3), 798–805.

Souleles, N. S. (1999, September). The response of household consumption to income tax refunds. *American Economic Review 89*(4), 947–958.

Souleles, N. S., J. A. Parker, and D. S. Johnson (2006, December). Household expenditure and the income tax rebates of 2001. *American Economic Review 96*(5), 1589–1610.

van Wincoop, E. (1996, June). A multi-country real business cycle model with heterogeneous agents. *Scandinavian Journal of Economics 98*(2), 233–51.

Vissing-Jorgensen, A. (2002, August). Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy 110*(4), 825–853.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-Nested hypotheses. *Econometrica 57*(2), 307–333.

A Appendix

A.1 Data sources and calculations

We collect data from OECD Main Economic Indicator (MEI) and OECD Quarterly National Accounts (QNA) for the period 1973-2007 and construct variables using the definitions summarized in Table A1.

Table A1: Data sources and calculations

Variable	Definition	Source
The U.S.		
Output (y_1)	Gross Domestic Product (at constant price 2000)	OECD MEI
Consumption (c_1)	Private plus Government Final Consumption Expenditure (at constant price 2000)	OECD MEI
Investment (x_1)	Gross Fixed Capital Formation (at constant price 2000)	OECD MEI
Employment (n_1)	Civilian Employment Index	OECD MEI
Real exchange rate (rx)	Price-adjusted Broad Dollar Index	Board of Governors
Import price	imports at current prices/imports at constant prices	OECD QNA
Export price	exports at current prices/exports at constant prices	OECD QNA
Terms of trade (p)	import price/export price	
Net exports ratio (nx)	(import-p*export)/y_1 (all at current prices)	
Real imports ratio (ir)	import/(GDP-export)	
Rest of the World		
Output (y_2)	Aggregate of Canada, Japan and 19 European Counties (aggregate with PPP exchange rates in 2000)	OECD MEI
Consumption (c_2)	Aggregate of Canada, Japan and 19 European Counties (aggregate with PPP exchange rates in 2000)	OECD MEI
Investment (x_2)	Aggregate of Canada, Japan and 19 European Counties (aggregate with PPP exchange rates in 2000)	OECD MEI
Employment (n_2)	Aggregate of Canada, Japan and 8 European Counties (weighted with populations in 2000)	OECD MEI

However, since OECD no longer reports aggregate data series on GDP, consumption and investment for European 15 which Heathcote and Perri (2002) used to compute variables for the rest of the world and since consistent series for each of those 15 European counties are not available either, instead, we used 19 European countries, including Austria, Belgium,Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain, Sweden and United Kingdom, Iceland, Luxembourg, Switzerland and Turkey. The employment series for the rest of the world, because of data unavailability, is computed as the weighted aggregate of Canada, Japan and 8 European countries (Austria, Finland, Germany, Italy, Norway, Spain, Sweden and UK). These differences in the sample may be contribute to the differences between the estimates of productivity shock process in Heathcote and Perri (2002) and in this paper.

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

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

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:

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

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

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]

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

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

[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

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

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

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

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

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

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: $\mathcal{F} = \{f_1(\theta), \dots, f_k(\theta)\}$ and $\mathcal{G} = \{g_1(\beta), \dots, g_k(\beta)\}$. 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: $D_{min}(\mathcal{F}, h_2) = \min_{j=1,\dots,k} d(f_j(\theta), h_2; W_{h_2})$ or $D_{max}(\mathcal{F}, h_2) = \max_{j=1,\dots,k} d(f_j(\theta), h_2; W_{h_2})$. 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

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

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:

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

[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:

$$\sqrt{n}\begin{pmatrix} \begin{pmatrix} \hat{\rho}_{11} - \rho_{11} \\ \hat{\rho}_{12} - \rho_{12} \end{pmatrix} \\ \hat{\sigma}_{e_1} - \sigma_{e_1} \\ \hat{\sigma}_{e_1e_2} - \sigma_{e_1e_2} \end{pmatrix} = \begin{pmatrix} H_{2;3} & 0_{2\times1} & 0_{2\times1} & 0_{2\times1} \\ 0_{1\times6} & \frac{1}{2\sigma_{e1}} & 0 & 0 \\ 0_{1\times6} & -\frac{\sigma_{e_1e_2}}{2\sigma_{e_1}^2} & -\frac{\sigma_{e_1e_2}}{2\sigma_{e_2}^2} & \frac{1}{\sigma_{e_1}\sigma_{e_2}} \end{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=2}^{n} \begin{pmatrix} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \otimes X_t \\ \varepsilon_{1,t}^2 - \sigma_{e_1}^2 \\ \varepsilon_{2,t}^2 - \sigma_{e_2}^2 \\ \varepsilon_{1,t}\varepsilon_{2,t} - \sigma_{e_1e_2}\sigma_{e_1}\sigma_{e_2} \end{pmatrix}$$
$$+ o_p(1),$$

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 $\hat{\sigma}_{e_1e_2}$ 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

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

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.