**CAEPR Working Paper
#2016-004**

A Practical Approach to Testing Calibration Strategies

**Yongquan Cao
Indiana University**

**Grey Gordon
Indiana University**

January 3, 2018

This paper can be downloaded without charge from the Social Science Research Network electronic library at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2843814

The Center for Applied Economics and Policy Research resides in the Department of Economics at Indiana University Bloomington. CAEPR can be found on the Internet at: <http://www.indiana.edu/~caepr>. CAEPR can be reached via email at caepr@indiana.edu or via phone at 812-855-4050.

©2016 by Yongquan Cao and Grey Gordon. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

A Practical Approach to Testing Calibration Strategies*

Yongquan Cao[†]

Grey Gordon[‡]

January 3, 2018

Abstract

A calibration strategy tries to match target moments using a model's parameters. We propose tests for determining whether this is possible. The tests use moments at random parameter draws to assess whether the target moments are similar to the computed ones (evidence of existence) or appear to be outliers (evidence of non-existence). Our experiments show the tests are effective at detecting both existence and non-existence in a non-linear model. Multiple calibration strategies can be quickly tested using just one set of simulated data. Applying our approach to indirect inference allows for the testing of many auxiliary model specifications simultaneously. Code is provided.

JEL Codes: C13, C51, C52, C80, F34

Keywords: Calibration, GMM, Indirect Inference, Existence, Misspecification, Outlier Detection, Data Mining

*We thank Juan Carlos Escanciano, Pablo Guerron-Quintana, Amanda Michaud, Stefan Weiergraeber and two referees for helpful comments. All codes for this paper are available at <https://sites.google.com/site/greygordon/research>.

[†]Indiana University, yongcao@uemail.iu.edu.

[‡]Indiana University, greygordon@gmail.com. Corresponding author.

1 Introduction

Most calibration approaches are, in effect, generalized method of moments (GMM).¹ GMM assumes the existence of parameters θ^* such that a model’s moments $m(\theta)$ match target moments m^* , i.e., $m(\theta^*) = m^*$. The parameters θ^* are identified if $m(\theta) \neq m^*$ for all $\theta \neq \theta^*$. If these conditions are satisfied, θ^* can be found as the solution of

$$\min_{\theta \in \Theta} (m(\theta) - m^*)'W(m(\theta) - m^*) \quad (1)$$

where Θ is the parameter space and W is a positive definite weighting matrix (often diagonal). In this paper, our objective is to test whether a θ^* satisfying $m(\theta^*) = m^*$ exists without explicitly solving (1), a computationally costly and error-fraught process for many economic models. We use the term calibration strategy to reflect the particular choice of moments used for identification, noting that a researcher typically has multiple choices available to them.²

Our approach generates i.i.d. random draws of θ from a distribution G (chosen by the user). In many cases, Θ is a hypercube and G is the uniform distribution over it. Using the draws $\{\theta_i\}$ and associated moments $\{m(\theta_i)\}$ as data, we test whether m^* is an “outlier” with respect to the distribution of moments induced by G . If so, we take it as evidence that θ^* does not exist. For simplicity, we treat the moments as being without sampling error, but practically speaking this makes no difference for our proposed techniques.

We employ three tests, two straightforward statistical methods and one from the data-mining literature. The first uses regressions to determine the relationship between one component of m and all the other components. We then assess whether the m^* observation is an outlier using standardized residuals, leverage, and Cook’s distance. The second test measures m^* ’s Mahalanobis distance, a measure closely related to the log-density of the multivariate normal. Under an assumption that the moments are normally distributed, the probability of encountering a moment with a worse Mahalanobis distance is easily calculated. A low probability suggests that m^* is an outlier. The third test computes m^* ’s local outlier factor (LOF), a measure proposed by Breunig, Kriegel, Ng, and Sander (2000). The LOF is closely related to a ratio of densities estimated non-parametrically via a k-Nearest Neighbor (kNN) algorithm. A high LOF for m^* suggests it has a small density relative to its neighbors’ and hence is an outlier.

We assess our proposed tests using a real-life example of challenging existence issues given in Chatterjee and Eyigungor (2012). There, three parameters (the discount factor and two default cost parameters) are used to target three moments (average interest rate spreads, their standard deviation, and the debt-output ratio). This model provides a useful test case for four reasons. First, the moments are tightly connected to one another, which illustrates some of the difficulties

¹Hansen and Heckman (1996) provide an overview of the relationship between calibration as proposed by Kydland and Prescott (1982) and estimation.

²For concreteness, imagine there are 2 moment statistics $\{a, b\}$ known from the data and model but only 1 parameter. Then the researcher could take m to be any one of the three vectors $[a]$, $[b]$, or $[a, b]'$ with each one giving a distinct calibration strategy.

frequently encountered in calibration. Second, [Chatterjee and Eyigungor \(2012\)](#) show the long-term debt case gives existence for their calibration strategy, and our tests confirm this. Third, our results show the short-term debt case does not give existence, indicating our tests can accept or reject calibration strategies. Last, default decisions and a discrete state space create many local minima in (1). Consequently, solving directly for θ^* requires a thorough and time-consuming search of the parameter space, which highlights the usefulness of applying our low-cost tests before calibrating.

We also show how to quickly test multiple calibration strategies. This is particularly useful because researchers often use only a few moments as targets, reserving the rest for testing the model’s non-targeted predictions (a procedure [Hansen and Heckman, 1996](#), call calibration and verification).³ Hence, a researcher need not pick moments, run a costly minimization routine, and then see whether it found a θ^* such that $m(\theta^*) = m^*$. Rather, they can solve a model for multiple parameter values, check many different moment combinations all at once, and then attempt to solve (1) only for combinations that seem to give existence (or modify the model to better fit the data).

While we have motivated our approach in terms of GMM, it can also be useful in indirect inference as developed by [Smith \(1993\)](#); [Gourieroux, Monfort, and Renault \(1993\)](#); [Gallant and Tauchen \(1996\)](#); and others. With indirect inference, the moments to be matched are determined by an auxiliary model and may be parameters of the auxiliary model (like in [Smith, 1993](#)) or its score (like in [Gallant and Tauchen, 1996](#)). Given them, our method applies in straightforward way, and we explicitly show how in an example using vector autoregressions (VARs) and their score.⁴ We test for non-existence—an indication that the deep (non-auxiliary) model is misspecified—and find the long-term debt specification can give existence or not depending on which variables are included in the VAR. In contrast, the short-term debt does not give existence for any of the seven VAR specifications we consider.

Our approach is designed to test existence, i.e., whether there is a θ^* such that $m(\theta^*) = m^*$. However, it can also aid in identification. For concreteness, suppose existence holds for an “over-identified” calibration strategy m in the sense that $\dim(m) > \dim(\theta)$. In that case, existence also holds for any calibration strategy $\tilde{m}$ that requires only a subset of m ’s moment conditions. However, the parameters identified by m , namely $\{\theta | m(\theta) = m^*\}$, are a subset of the parameters identified by $\tilde{m}$, namely $\{\theta | \tilde{m}(\theta) = \tilde{m}^*\}$.⁵ Hence, m should be preferred over $\tilde{m}$ on the grounds of identification. In our [Chatterjee and Eyigungor \(2012\)](#) application with long-term debt, we found a calibration strategy with 6 moments for which, according to our tests, existence holds. If existence does hold, then since there are only three parameters, there would be 41 ($= \binom{6}{3} + \binom{6}{4} + \binom{6}{5}$) calibration strategies consisting of three or more moments that also give existence but are dominated in terms of identification (and one of these is the strategy [Chatterjee and Eyigungor, 2012](#), use).⁶ This shows

³E.g., this is done in [Kydlund and Prescott \(1982\)](#), where the correlations are non-targeted, and [Christiano and Eichenbaum \(1992\)](#), where labor market moments are left non-targeted. [Christiano and Eichenbaum \(1992\)](#) and [Hansen \(2008\)](#) discuss how to formally test the fit of the non-targeted moments.

⁴[Feldman and Sun \(2011\)](#) also validate econometric specifications by using simulated model data.

⁵If the model is identified under m and $\tilde{m}$, then both sets will be a singleton, but in general the parameters are only set-identified. The $\tilde{m}^*$ appearing in the $\{\theta | \tilde{m}(\theta) = \tilde{m}^*\}$ is the appropriate subset of moments in m^* .

⁶Specifically, the 6 moments we found give existence are the mean, standard deviation, and cyclical of the interest

that in addition to reducing the cost of calibration, testing multiple calibration strategies should also aid in identification.

To ease adoption of our proposed approach, we provide two types of code. The first is a Fortran subroutine that uses MPI to compute the $\{(\theta_i, m(\theta_i))\}$ draws in parallel. This allows for efficient testing even if the routine for computing $m(\cdot)$ is single-threaded (but it may be multiply-threaded with OpenMP). Optionally, it will also use these draws as part of a global optimization algorithm that can be used to solve (1). The second type of code is a Stata do-file that imports the Fortran results, runs all our tests, and produces reports and graphs indicating potential problems with the calibration strategy (or strategies in the case where testing multiple combinations is desired).⁷

1.1 Related literature

We test existence by asking whether the target moments are outliers. While we have considered three tests, there are in fact many others, all of which could be useful. A number of traditional statistical methods and references are described in [Bollen and Jackman \(1990\)](#) and [Ben-Gal \(2005\)](#). The data-mining literature also contains many other methods. Examples include [Kriegel, Kröger, Schubert, and Zimek \(2009\)](#)’s “Local Outlier Probabilities,” [Zhang, Hutter, and Jin \(2009\)](#)’s “Local Distance-based Outlier Factor,” [Ramaswamy, Rastogi, and Shim \(2000\)](#)’s selection of outliers according to kNN-distance ranking, and [Angiulli and Pizzuti \(2002\)](#)’s weighted kNN distance method.

Our focus on existence is complementary to an extensive literature on identification dating back to at least [Koopmans and Reiersol \(1950\)](#). Several tests have been proposed for checking identification problems, including those discussed in the survey [Stock, Wright, and Yogo \(2002\)](#) and proposed by [Canova and Sala \(2009\)](#). While our approach can eliminate some calibration strategies on the grounds of identification, once that has been done it has nothing else to say regarding identification. Consequently, tests like the ones discussed in these papers are indispensable.

A smaller literature examines the consequences of favoring certain observables over others in estimation, analogous to our examining combinations of moments that give existence. Building on [Sargan \(1958, 1959\)](#), [Hansen \(1982\)](#) characterizes GMM’s consistency and efficiency for many different types of weighting matrices. As a special case, these index all the moment combinations we consider.⁸ [Guerron-Quintana \(2010\)](#) considers how selection of observables impacts DSGE estimation, and [Canova, Ferroni, and Matthes \(2014\)](#) propose and analyze selection procedures.

rate spread, the mean debt-output ratio, and the relative standard deviation of net exports and of consumption. For these, the LOF is 0.99 and 1.02 when taking $k = 20$ and 40, respectively. Moreover, assuming normality, the probability of encountering a moment with Mahalanobis distance worse than the target moment’s distance is 72%. As we discuss later, these measures strongly suggest existence. [Chatterjee and Eyigungor \(2012\)](#) use the mean and standard deviation of the spread and the mean debt-output ratio.

⁷The tests we propose are simple enough to implement in other languages like Matlab or Fortran. We chose Stata because it is well-known, cross platform, and has built-in functions that are helpful in the statistical analysis.

⁸Specifically, [Hansen \(1982\)](#) considers what [Hansen \(2008\)](#) calls selection matrices A , having dimensions $r \times \dim(m)$ where $\dim(\theta) \leq r \leq \dim(m)$. A is assumed to have full (row) rank ([Hansen, 1982](#), Assumption 3.6, p. 1040). The weighting matrix W in (1) is then $A'A$ ([Hansen, 1982](#), p. 1041). Since A is only required to have full row rank, it may place zero weight on any of the $\dim(m(\theta)) - \dim(\theta)$ “extra” moments, thus selecting any of the combinations we consider.

Hansen (1982), Stock and Wright (2000), and Kleibergen (2005) all derive asymptotic distributions of functions of the optimal estimator that can be used to conduct hypothesis tests in finite samples. E.g., Hansen’s shows that, with the optimal GMM estimator, the GMM objective (1) times the sample size is distributed χ^2 at θ^* . With lack of existence, the same statistic explodes as the sample size grows.⁹ The key distinction between these tests of existence and ours is that they require finding a θ^* first—a procedure that is often costly and error-prone—and then using it to form a null hypothesis. Our approach circumvents this by looking at the distribution of $m(\theta)$ to infer whether a θ^* giving $m(\theta^*) = m^*$ is likely to exist.

2 Checking existence for a given strategy

In this section, we develop a multi-pronged approach to testing existence. We evaluate the performance of our tests using the sovereign default model of Chatterjee and Eyigungor (2012). We mention just a few aspects of it here in order to define the parameter space; a thorough description is given in Appendix B. In the model, a sovereign with stochastic endowment stream y discounts future utility at rate β . If he defaults, the endowment is exogenously reduced by $\phi(y) = \max\{0, d_0 y + d_1 y^2\}$ (where $d_1 \geq 0$). For long-term debt, Chatterjee and Eyigungor (2012) find $\beta = 0.954$, $d_0 = -0.188$ and $d_1 = 0.246$ deliver the targeted moments (the model period taken to be a quarter). We set the parameter space to be $(\beta, d_0, d_1) \in [0.94, 0.98] \times [-0.5, 0] \times [0, 0.5]$ and take G as the uniform distribution over it. Our methods only test the existence of $\theta^* \in \Theta$. For example, if $m(\theta^*) = m^*$ only for $\theta^* = (0.5, -0.2, 0.2)$, the tests should indicate non-existence since $\theta^* \notin \Theta$.¹⁰ Our sample is 1000 draws.

2.1 Joint-moment restrictions

Figure 1 shows pairwise plots of the three moments for long-term debt (the top panel) and short-term debt (the bottom panel). The figure reveals clear connections between moments, which we refer to as “joint-moment restrictions.” For long-term debt, the graphs visually reveal that the target moments appear to satisfy the joint-moment restrictions in that the target moments (indicated by the red squares) fall within the clouds of simulated data. However, because the moment conditions are only pairwise, this may be misleading in that matching 2 of the moments simultaneously does not imply matching 3 simultaneously will be possible.¹¹ What cannot be misleading is the short-term debt case where the target moments are far away from the pairwise joint-moment conditions in each case. Note that, for short-term debt, the target spread standard deviation and debt-output

⁹As we have assumed there is no sampling error in the moments, these hypothesis tests in our context have an infinite sample size and, with lack of existence, the test statistic is “infinite.”

¹⁰Chatterjee and Eyigungor (2012) do not estimate the short-term debt case. However, in a similar model Gordon and Guerron-Quintana (2017) find it is possible to match the same moments as they do when using $\beta = 0.44$. We consider this value implausibly low, and so restrict the parameter space to more conventional values.

¹¹A trivial example of this is $m^1, m^2 \sim N(0, 1)$ (a superscript denoting a component of m) and $m^3 = m^1 + m^2$ with $m^* = [0, 0, 1]$. With enough data, m^* would be in the cloud for each pairwise case but not when considering the three moments simultaneously.

ratio can likely be matched individually: The support of the debt-output ratio covers the target debt-output ratio, and likewise for the spread standard deviation. However, the two statistics cannot be matched simultaneously. While graphing all the pairwise cases is helpful here, as the number of moments grows, this becomes increasingly costly from a user’s perspective.

Figure 1: Joint-moment restrictions

2.2 Leverage, error, and Cook’s distance

To handle these failures of a purely graphical approach, we attempt to capture the link between a moment m^j (a superscript here and throughout denotes a vector component) and all the other moments using a flexible OLS specification. Specifically, we regress a single component of m^j on interactions of the all other components $m^1, \dots, m^{j-1}, m^{j+1}, \dots, m^{\dim(m)}$ and a constant. Our code lets the user specify the order of interactions desired, and in our example we use a second order, i.e., the regressors are $1, m^k, m^k m^l$ for $k, l \neq j$. However, in this regression, we include one “fake” data point, the components of m^* . We then formally test whether the target moments are an outlier from the viewpoint of the regression model. If so, we consider it evidence that $\theta^* \in \Theta$ does not exist.

We employ three measures—standardized residuals, leverage, and a combination of the two called Cook’s distance—to determine whether m^* is an outlier. First, we consider where the targets fall in terms of standardized residuals. That is, we take the error term e_i^j (where a subscript denotes an observation) and scale it by η so that $\varepsilon_i^j = \eta e_i^j$ has $Std(\varepsilon_i^j) = 1$. Then, assuming the errors are

roughly normally distributed, the familiar rules of thumb may be used to interpret ε^{*j} (the error form the target moment observation). A large value, say greater than 4 or 5 (in absolute terms), would occur with low probability and so suggest m^* is an outlier; a value less than 1 would suggest the opposite.

Figure 2 presents this measure for our example. For the long-term debt version, the standardized residuals for the target moments are within ± 1 standard deviations. This suggests that given the underlying joint-moment restrictions, the targets are not unreasonable. However, for the short-term debt version, the targets are unreasonable. E.g., the spread mean is 17 standard deviations too high. This indicates that the model, conditional on matching the target debt-output ratio and target spread standard deviation, would require an absurdly high error to match the target spread mean.

Figure 2: Outlier tests

Our second measure is leverage. Leverage is always between 0 and 1 and indicates how much influence a particular datapoint has on the regression coefficients (Bollen and Jackman, 1990, p. 262). Formally, the leverage of an observation i is defined as the i -th diagonal of the matrix H such that $\hat{m}^j = Hm^j$ where $\hat{m}^j$ are the predicted values and m^j the actual (Bollen and Jackman, 1990, p. 262).¹² If the observation corresponding to the target moments has large leverage and a large error, it means that it does not fit into the pattern implied by the other moments. If the target moments have high leverage and low error, this may indicate that the moments fit into the

¹²In terms of $y = Xb$, the H matrix is $X(X'X)^{-1}X'$ since $\hat{b} = (X'X)^{-1}X'y$.

underlying pattern, but the θ delivering $m(\theta) \approx m^*$ are unusual in some way. For instance, it could indicate the solution θ may not be in the interior of the parameter space. Alternatively, it could be that $G(\theta)$ places very low probability on θ near θ^* . [Bollen and Jackman \(1990\)](#) summarize the rules of thumb for when an observation has high leverage. For p the number of regressors, these range from p/n (the mean leverage value) to $3p/n$ (p. 262).

Figure 2 reports leverage on the vertical axis. In the example, $p = 6$ and $n = 1000$ so the rules of thumb range from 0.006 to 0.018, and the larger of these is displayed as a black, horizontal line. For the long-term debt moments, all the leverage values fall well below this cutoff. Hence, the target moments seem to satisfy the joint-moment restrictions. For the short-term debt version, the leverage exceeds the cutoff for all the statistics indicating the target moments do not satisfy them.

Our last regression-based measure is Cook's distance. Cook's distance combines leverage and the standardized residual into a single number. Formally, letting h_i^j and e_i^j denote the leverage and (non-standardized) error of observation i and component j , Cook's distance is $(1/p)h_i^j(1 - h_i^j)^{-2}(e_i^j)^2/s^2$ where s^2 is the regression's mean squared error ([Bollen and Jackman, 1990](#), p. 266). There are two rules of thumb for this measure. One says a Cook's distance greater than 1 is suspicious, and the other uses a cutoff of $4/n$ instead ([Bollen and Jackman, 1990](#), p. 268). For long-term debt, the largest value is 0.0006, and so the target moments seem to satisfy the joint-moment restrictions according to both cutoffs. The opposite is true for short-term debt where the smallest Cook's distance is 8.0 and the largest is 231.

2.3 Mahalanobis distance

A conceptually simple parametric multivariate outlier detection method is given by considering the Mahalanobis distance. The Mahalanobis distance d is defined as

$$d(m) = ((m - \mu)' \Sigma^{-1} (m - \mu))^{1/2}$$

where μ is the sample mean and Σ is the sample variance-covariance matrix.¹³ Assuming that $m \sim N(\mu, \Sigma)$, $d(m)^2$ is distributed $\chi_{\dim(m)}^2$.¹⁴ Consequently, the probability of $d(m)^2$ being greater than $d(m^*)^2$ is given by

$$Pr(d(m)^2 \geq d(m^*)^2) = 1 - \chi_{\dim(m)}^2(m^*),$$

which also equals $Pr(d(m) \geq d(m^*))$. Hence, the probability of encountering an observation with greater Mahalanobis distance than the target moments' distance is $1 - \chi_{\dim(m)}^2(m^*)$. We refer to this statistic as the M-test value. For our example, long-term (short-term) debt has a M-test value

¹³[Penny and Jolliffe \(2001\)](#) and [Ben-Gal \(2005\)](#) provide examples of using the Mahalanobis distance for detecting outliers. [Gnanadesikan \(1997\)](#) calls d^2 the squared generalized distance (p. 48).

¹⁴[Gnanadesikan \(1997, p. 48\)](#) makes a general claim of this. Formally, the result may be had by converting m to a standard normal z via $z = A^{-1}(m - \mu)$ for AA' the Cholesky decomposition of Σ . Then $d^2(m) = (m - \mu)' \Sigma^{-1} (m - \mu) = (m - \mu)' (A^{-1})' A^{-1} (m - \mu) = z(m)' z(m) = \sum_{j=1}^{\dim(m)} z_j(m)^2$. So, d^2 is the sum of $\dim(m)$ squared independent normals, with independence of z 's components following from their being uncorrelated ([Billingsley, 1995](#), pp. 384-385). Hence, $d^2 \sim \chi_{\dim(m)}^2$.

of 0.59 (6×10^{-63}). This suggests the long-term debt target moments could plausibly be generated by a $\theta \in \Theta$ but clearly not for short-term debt.

2.4 Local Outlier Factors

The main drawback of the Mahalanobis distance test in the preceding section is the assumption of normality. Breunig et al. (2000)’s local outlier factor (LOF) improves on this by non-parametrically estimating something akin to a density. Formally, consider some set X and some metric $d(.,.)$. Let $N_k(x)$ denote the k -nearest neighbors of $x \in X$. Here, we assume $N_k(x)$ is unique for each x to simplify exposition (Breunig et al., 2000, handle the general case). Define the “ k -distance” as $\epsilon_k(x) = \max_{o \in N_k(x)} d(x, o)$, which is the radius of the smallest closed ball containing the k -nearest neighbors of x . A “reachability distance” from x to some point p is defined as $r_k(x, p) = \max\{\epsilon_k(p), d(x, p)\}$ (note that generally $r_k(x, p) \neq r_k(p, x)$).

Only one more definition is needed before defining the LOF. The “local reachability density” is defined as

$$lrd_k(x) = \left(\frac{1}{k} \sum_{p \in N_k(x)} r_k(x, p) \right)^{-1},$$

the inverse of the average reachability distance. A relatively high value means x can be easily reached ($r_k(x, p)$ is low) from x ’s k -nearest neighbors (and hence in some sense has a large density). The LOF is defined as

$$LOF_k(x) = \frac{1}{k} \sum_{p \in N_k(x)} \frac{lrd_k(p)}{lrd_k(x)},$$

which is the mean of the local reachability densities of x ’s k -nearest neighbors relative to that of point x . Hence, if LOF_k is large, x is further from its neighbors ($lrd_k(x)$ is small) than x ’s neighbors are from *their* neighbors. An LOF_k value close to or less than 1 indicates that x is not an outlier while a value much greater than 1 suggests x is an outlier.

Table 1 reports LOF values for the long and short-term debt cases. There is no consensus on the best value of k , but there is some evidence to suggests k between $n^{1/4}$ to $n^{1/2}$ are reasonable.¹⁵ Because LOF values greater but not much greater than 1 are hard to interpret, we also computed the target moments’ LOF percentile among all the moments.¹⁶ The percentile is computed as $100 \times (rank - 1)/(n - 1)$ where $rank = n$ means the observation had the highest LOF in the sample.

¹⁵Loftsgaarden and Quesenberry (1965) propose a consistent non-parametric density function estimator that uses the distance from a point to its kNN. They suggest $k = n^{1/2}$ gives “good results” empirically (p. 1051). Enas and Choi (1986) quantitatively investigate optimal values of k in the context of kNN classification (assigning observations to groups based on their kNN distance) and conclude values of $n^{2/8}$ to $n^{3/8}$ are best (p. 244). The results in Kriegel et al. (2009) can be sensitive to the value of k , but for the range we consider (5 to 40) they seem to be robust.

¹⁶An advantage of LOF is that one need not compute the lrd_k or r_k for all points, only for the neighbors of a given point and their neighbors. For our 1000 observations and $k = 20$, we can compute the 112 (56 twice) moment combinations used in section 3 in 1.2 minutes when only computing the target moments’ LOF. For computing the percentile, we need to compute LOF for all the points, which takes 3.2 minutes total. Our approach here is qualitatively similar to Ramaswamy et al. (2000) who rank observations according to their kNN distance to find outliers.

For long-term debt, the LOF values and percentiles suggest the target moments are not outliers as the largest LOF value is 1.05 (and hence the “density” is only 5% percent lower at the target moments) and the percentiles are not large. However, for short-term debt, all the LOF values are greater than 16 and the percentiles are all 100% (the latter implying that the target moments have the largest LOF of all the moments). This forcefully suggests that m^* is an outlier for the short-term debt model.

	LOF values			LOF percentile		
	$k = 5$	$k = 20$	$k = 40$	$k = 5$	$k = 20$	$k = 40$
Long-term debt	1.05	1.02	1.02	58.2	40.0	37.2
Short-term debt	32.1	21.6	16.6	100.0	100.0	100.0

Note: A percentile of x means in it is the $x\%$ largest LOF value.

Table 1: Target moments’ Local Outlier Factor (LOF) values

While the results here are intuitive and agree with the other tests, the claim that the LOF is non-parametrically estimating ratios of densities may seem suspicious. Based on the work of Loftsgaarden and Quesenberry (1965), we can formalize the connection as follows. Assume that the distribution F induced by G , i.e., $F(M) = G(\{\theta \in \Theta | m(\theta) \in M\})$, is absolutely continuous. Given a sequence $k(n)$ such that $\lim k(n) = \infty$ and $\lim k(n)/n = 0$, a consistent estimator of the density f at a given point m is

$$\hat{f}(m) = \frac{k(n) - 1}{n} \frac{1}{A_{\epsilon_{k(n)}(m)}}$$

where $A_{\epsilon_{k(n)}(m)}$ is the volume of a closed ball with radius $\epsilon_{k(n)}(m)$ (the k -distance from above). The volume is $2r^p \pi^{p/2} p^{-1} \Gamma(p/2)$ where $p = \dim(m)$ (Loftsgaarden and Quesenberry, 1965, p. 1049). Consequently, for two points m_1 and m_2 ,

$$\left(\frac{\hat{f}(m_2)}{\hat{f}(m_1)} \right)^{1/\dim(m)} = \frac{\epsilon_{k(n)}(m_2)^{-1}}{\epsilon_{k(n)}(m_1)^{-1}}.$$

Hence, the ratio of the densities (raised to $1/\dim(m)$) is equal to the ratio of inverse distances of the k -th nearest neighbor. This expression is analogous to the LOF, the average ratio of the inverse of the average distances $r_k = \max\{\epsilon_k, d\}$.

3 Checking existence for many strategies

In this section, we consider one of the first problems encountered in calibration: Which moments should be targeted? Suppose a researcher knows q moments in the model and data with the model having p parameters. If exact identification is desired, there are a total of $\binom{q}{p} = \frac{q!}{p!(q-p)!}$ mo-

ment combinations (i.e., calibration strategies) that could be used in estimation. Similarly, if over-identification using $l > p$ moments is desired, there are $\binom{q}{l}$ possible calibration strategies. (In the next section we will consider indirect inference, which provides an alternative approach to moment selection.) We propose using the M-test values and LOFs to determine which of these many strategies are likely to give existence.¹⁷

To generate a large, but not overly large, list of pairs, we consider two methods of generation. In our code, the user specifies an l . Then, $\binom{q}{l}$ combinations of moments are generated. If $l = p$, this exhausts all the moments giving exact identification, but a user-selectable l prevents there from being too many combinations.¹⁸ The second method of generation is by taking the list of q moments and removing l , which also gives $\binom{q}{l}$ pairs. This exhausts all the moments giving exact identification if $q - l = p$ since $\binom{q}{l} = \binom{q}{q-l} = \binom{q}{p}$. Note that the two methods typically produce different moments. For instance, if the moments are $\{a, b, c\}$ and $l = 1$, then the first method produces $\{a\}, \{b\}, \{c\}$ while the second produces $\{a, b\}, \{a, c\}, \{b, c\}$.

In our example, we add to the three moments already used five additional moments (the volatilities of consumption and net exports and the cyclicalities of consumption, net exports, and spreads). Since there are only three parameters, we set $l = 3$ (alternatively, we could set $l = 5$ implying $q - l = p$) and enumerate all possible moment combinations giving exact identification. Since there are 56 combinations, we do not present them all, but just list in table 2 the best and worst three ranked according to the M-test value. Our code takes around 1.7 seconds to evaluate each combination.¹⁹

According to the M-test values, both models have some target moment combinations that are not outliers and some combinations that definitely are. While the LOF values correlate somewhat well with the M-test values, there are notable exceptions. For instance, long-term debt's best combination as measured by the M-test has a high LOF percentile at 84%. Hence, a researcher choosing moments to calibrate to (and indifferent over which moments should be used) might favor the second or third best combination as a calibration strategy rather than the first.

The discrepancy for the best short-term debt combination (the debt-output ratio μ_b , the volatil-

¹⁷Having determined which calibration strategies give existence, we do not take a stand on how one should select a particular strategy. Many times, some moments are more essential for research questions than other ones, and so the best choice will be obvious. Of course, one should avoid any strategies that might result in weak or under-identification; Stock et al. (2002) and Canova and Sala (2009) provide ways to check for identification problems.

¹⁸To see why, note that for a given l there will be $\binom{q}{l}$ moment combination, which is bounded by $q!/l!$. While this bound grows extremely quickly (faster than exponential) in q , it also shrinks extremely quickly in l . So the approach can still handle an arbitrarily large number of moments q , but it may require l close to q to prevent the generation of too many moment combinations. E.g., if $l = q$ ($l = q - 1$), so that of the q moments available in the data q ($q - 1$) of them should be selected, there is only 1 (q) possible moment combination(s).

An alternative way to reduce the number of moments is to redefine the q moments into groups. For concreteness, suppose the moments consist of the mean μ and standard deviation σ for spreads r , consumption c , and output y (so that $q = 6$). Then one could require that the mean and standard deviation must both be included if either one of them is: E.g., if μ_r is included, then σ_r must be as well. This essentially reduces the choice space from $q = 6 - \{\mu_r, \sigma_r, \mu_c, \sigma_c, \mu_y, \sigma_y\}$ to $q = 3 - \{(\mu_r, \sigma_r), (\mu_c, \sigma_c), (\mu_y, \sigma_y)\}$. The next section does essentially this, selecting observables and then generating moments from those observables.

¹⁹The times are for $k = 20$ running in Stata SE (and hence not parallel) on a quad core 2.7 GHz processor. If the percentile is not computed, the time drops to 0.6 seconds.

Long-term debt, best			Long-term debt, worst		
Moments	M-test	LOF value (pctile)	Moments	M-test	LOF value (pctile)
μ_b, μ_r, ρ_r	0.89	1.14 (83.7)	$\sigma_{nx}, \rho_c, \rho_r$	3×10^{-4}	2.03 (98.8)
$\mu_r, \sigma_c, \sigma_{nx}$	0.86	1.05 (52.1)	$\sigma_r, \sigma_{nx}, \rho_c$	3×10^{-4}	1.94 (99.4)
μ_b, σ_r, ρ_r	0.85	1.08 (74.2)	$\sigma_{nx}, \rho_c, \rho_{nx}$	7×10^{-5}	2.92 (100)
Short-term debt, best			Short-term debt, worst		
Moments	M-test	LOF value (pctile)	Moments	M-test	LOF value (pctile)
$\mu_b, \sigma_{nx}, \rho_c$	0.95	2.30 (100)	$\mu_r, \sigma_r, \sigma_c$	1×10^{-57}	19.1 (100)
$\mu_b, \sigma_c, \sigma_{nx}$	0.21	2.42 (99.9)	μ_b, μ_r, σ_r	6×10^{-63}	21.6 (100)
μ_b, σ_c, ρ_r	0.18	3.01 (99.9)	$\mu_r, \sigma_r, \rho_{nx}$	2×10^{-63}	21.3 (100)

Note: M-test is the Mahalanobis test value discussed in section 2.3; LOF is for $k = 20$; μ, σ , and ρ denote mean, standard deviation, and correlation with log output, respectively; and b, c, nx , and r are debt-output, log consumption, net exports-output, and spreads, respectively. The volatilities for c and nx are relative to log output volatility.

Table 2: Best and worst moment combinations

ity of net exports relative to output σ_{nx} , and the cyclical of consumption ρ_c) is perhaps more interesting. There, the M-test value is close to 1 but the LOF percentile indicates that the target moments have the highest LOF. To see why this is, consider figure 3 that plots ρ_c against σ_{nx} . The data exhibit clear joint-moment restrictions, which the target moments do not appear to satisfy. However, the target moments lie almost on top of the best fit line. The M-test value, which assumes a normal distribution, is fooled because of this as the target is not an outlier with respect to $N(\mu, \Sigma)$. However, the LOF correctly detects that the target is relatively far from its nearest neighbors.

Because we set $l = 3$, our code also produces the 5-moment combinations constructed by taking 8 moments and removing 3 at a time. For short-term debt, these have a maximum M-test value of 2×10^{-5} , a minimum LOF of 1.5, and a minimum percentile of 98%. Hence, all 56 combinations are likely outliers. However, for long-term debt, the best six moment combinations (ranked by the M-test value) have an M-test value of 0.8 or more, an LOF of 1.06 or less, and percentiles between 30% and 63%. Hence, it is likely that the long-term debt model has multiple calibration strategies that give existence, even in the over-identified case.

4 Application to indirect inference

In testing existence for calibration strategies, our method connects to the literature on indirect inference (including Smith, 1993; Gouriou et al., 1993; Gallant and Tauchen, 1996, and many others). In indirect inference, there are two models. One is the “deep” model whose parameters θ are

Figure 3: Short-term debt's best moment calibration

of interest. Typically, the likelihood for this model is not easy to derive or compute. The other model is an “auxiliary” model with parameters γ , which may capture some properties of the deep model, but typically has a likelihood function that is easier to compute. The basic principle of indirect inference is that the model’s deep parameters θ are adjusted until the auxiliary model estimates using simulated model data resemble the auxiliary model estimates using actual data. The most straightforward way to do this is by choosing θ so that parameter estimates from simulated data—call them $\gamma(\theta)$ —match parameter estimates on the actual data—call them γ^* . In our setup, this is equivalent to having “moments” $m(\theta) := \gamma(\theta)$ and target moments $m^* := \gamma^*$. A less straightforward way, but one which can be very efficient both computationally and econometrically, is the approach of [Gallant and Tauchen \(1996\)](#). There, the auxiliary model’s score—i.e., the derivative of the log likelihood with respect to γ (the gradient)—determines the moment conditions. Specifically, the score using simulated model data gives $m(\theta)$ and the score using actual data gives m^* .

If the deep model is properly specified in that the data is generated from the deep model using some parameter θ^* , then indirect inference always gives existence: Simulated model data from θ^* has exactly the same distribution as the actual data and so $m(\theta^*)$ must, under mild conditions, be m^* . However, it is also the case that existence does *not* necessarily hold if the model is misspecified. We provide an example showing this in [Appendix A](#). In terms of matching parameters $\gamma(\theta)$ and γ^* , a lack of existence means the model cannot replicate the data’s properties as summarized by the auxiliary model parameters. With the [Gallant and Tauchen \(1996\)](#) approach, it means the model cannot replicate the data’s properties as summarized by the auxiliary model’s likelihood. In either case, lack of existence is economically important as it implies the deep model is not the data generating process.

To illustrate how our approach can be useful for testing existence, we again consider the [Chatterjee and Eyigungor \(2012\)](#) model while focusing on a VAR specification of the auxiliary model like in [Smith \(1993\)](#). In particular, we consider VAR(1) specifications of the form

$$\mathbf{x} = \mu + A\mathbf{x}_{-1} + \mathbf{e} \quad (2)$$

with $\mathbf{e} \sim N(0, SS')$ and S lower triangular. Letting n denote the dimensionality of $\mathbf{x}$, there are a total of $n + n^2 + n(n + 1)/2$ auxiliary model parameters $\gamma = (\mu, A, S)$.²⁰ We define our moments using the auxiliary model scores as proposed in [Gallant and Tauchen \(1996\)](#).

We consider seven total auxiliary models / calibration strategies corresponding to seven different definitions of the observables $\mathbf{x}$, namely, $[r]$, $[c]$, $[y]$, $[r, c]'$, $[r, y]'$, $[c, y]'$, and $[r, c, y]'$ where r is spreads, c is log consumption, and y is log output (the latter two linearly detrended in the data). Using one set of simulated data, we test existence for all of these auxiliary model specifications. Like in the previous section, we do so by looking at the M-test and LOF values, which are presented for long- and short-term debt in table 3.

VAR variables $\mathbf{x}$	Long-term debt		Short-term debt	
	M-test	LOF value (pctile)	M-test	LOF value (pctile)
$[r]$	0.578	2.55 (97.2)	$0 \times 10^{-\infty}$	28.7 (31.2)
$[c]$	0.731	4.12 (99.4)	2×10^{-4}	6.72 (100.0)
$[y]$	8×10^{-4}	2.08 (98.1)	3×10^{-18}	3.77 (99.7)
$[r, c]'$	4×10^{-11}	2.10 (95.3)	$0 \times 10^{-\infty}$	10.2 (100.0)
$[r, y]'$	$0 \times 10^{-\infty}$	4.66 (99.8)	$0 \times 10^{-\infty}$	12.3 (100.0)
$[c, y]'$	7×10^{-29}	4.10 (99.5)	10×10^{-23}	6.03 (100.0)
$[r, c, y]'$	$0 \times 10^{-\infty}$	3.79 (96.2)	$0 \times 10^{-\infty}$	37.9 (100.0)

Note: the LOF value is computed for $k = 20$; $0 \times 10^{-\infty}$ means the value is zero to numerical precision.

Table 3: Existence in indirect inference

For long-term debt, VARs with $[r]$ or $[c]$ (i.e., ARs) both seem to give existence according to the M-test while the LOF values and percentiles are high and seem to indicate nonexistence. For $[y]$, the M-test and LOF agree that there is probably non-existence.²¹ The VAR specifications with more than one variable all indicate lack of existence. As discussed above, this is an indication that

²⁰Note that for indirect inference one must always have the order condition $\dim(\gamma) \geq \dim(\theta)$ satisfied ([Gallant and Tauchen, 1996](#), p. 664). Here, since $\dim(\theta) = 3$, this is satisfied even with $n = 1$ (in which case the VAR is an AR).

²¹There are two reasons for this. First, we are using the output process parameter estimates from [Chatterjee and Eyigungor \(2012\)](#) which (a) include an i.i.d. shock in the estimation and (b) use 12 extra quarters of data, 1980Q1:1982Q4. We start our estimation in 1983Q1 because that is the first quarter with spreads data available, but 1980Q1:1982Q4 was a volatile period with annualized real GDP growth ranging from -17% to +16%. Second, is that we exclude from the sample the 74 periods following each default. Consequently, the estimates in (2) are biased due to the sample selection of model data. Since the model's spreads are not defined while in default/autarky, excluding at least some periods post default is necessary. See Appendix B for details.

(1) the model could be the data generating process for r or for c , but not both; and (2) the model is not the data generating process for y and not for combinations of r , c , and y , either. For short-term debt, each specification is soundly rejected.²² Consequently, the short-term debt model seems to not be the data generating process for r , c , or y .

5 Conclusion

We developed a battery of tests for evaluating calibration strategies. In our quantitative experiments, the methods unanimously suggested existence holds for long-term debt and does not hold for short-term debt. We also showed how multiple calibration strategies can be quickly evaluated. Additionally, we showed how our techniques can be applied to test existence in indirect inference. While no single test is fool proof, our multi-pronged approach should help researchers avoid unnecessary and costly attempts to match target moments that simply cannot be matched.

References

- F. Angiulli and C. Pizzuti. *Fast Outlier Detection in High Dimensional Spaces*, pages 15–27. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002.
- C. Arellano. Default risk and income fluctuations in emerging economies. *American Economic Review*, 98(3):690–712, 2008.
- I. Ben-Gal. Outlier detection. In *Data mining and knowledge discovery handbook*, pages 131–146. Springer, 2005.
- P. Billingsley. *Probability and Measure (Third Edition)*. John Wiley & Sons, Inc., Hoboken, NJ, 1995.
- K. A. Bollen and R. W. Jackman. *Modern Methods of Data Analysis*, chapter Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases. Sage, Newbury Park, CA, 1990.
- M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. LOF: Identifying density-based local outliers. *Proceedings of the ACM SIGMOD 2000 International Conference on Management of Data*, pages 93–104, 2000.
- F. Canova and L. Sala. Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics*, 56(4):431–449, 2009.
- F. Canova, F. Ferroni, and C. Matthes. Choosing the variables to estimate singular DSGE models. *Journal of Applied Econometrics*, 29(7):1099–1117, 2014.

²²The short-term debt’s high LOF but low LOF percentile for spreads is a testament to the model’s volatile behavior with short-term debt. In particular, spreads tend to be zero or close to zero up until and then drastically increase. See figure 5 of [Arellano \(2008\)](#) on p. 707.

- S. Chatterjee and B. Eyigungor. Maturity, indebtedness, and default risk. *American Economic Review*, 102(6):2674–2699, 2012.
- L. J. Christiano and M. Eichenbaum. Current real-business-cycle theories and aggregate labor-market fluctuations. *American Economic Review*, 82(3):430–450, 1992.
- G. G. Enas and S. C. Choi. Choice of the smoothing parameter and efficiency of the k-nearest neighbor classification. *Computers & Mathematics with Applications*, 12A(2):235–244, 1986.
- T. Feldman and Y. Sun. Econometrics and computational economics: an exercise in compatibility. *International Journal of Computational Economics and Econometrics*, 2(2):105–114, 2011.
- A. R. Gallant and G. Tauchen. Which moments to match? *Econometric Theory*, 12(4):657–681, 1996.
- R. Gnanadesikan. *Methods for statistical data analysis of multivariate observations*. John Wiley & Sons, Inc., New York, NY, 1997.
- G. Gordon and P. Guerron-Quintana. Dynamics of investment, debt, and default. *Review of Economic Dynamics*, Forthcoming, 2017.
- G. Gordon and S. Qiu. A divide and conquer algorithm for exploiting policy function monotonicity. *Quantitative Economics*, Forthcoming, 2017.
- C. Gourieroux, A. Monfort, and E. Renault. Indirect inference. *Journal of Applied Econometrics*, 8:S85–S118, 1993.
- P. A. Guerron-Quintana. What you match does matter: The effects of data on DSGE estimation. *Journal of Applied Econometrics*, 25(5):774–804, 2010.
- L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982.
- L. P. Hansen. Generalized method of moments estimation. In S. N. Durlauf and L. E. Blume, editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke, 2008.
- L. P. Hansen and J. J. Heckman. The empirical foundations of calibration. *Journal of Economic Perspectives*, 10(1):87–104, 1996.
- F. Kleibergen. Testing parameters in GMM without assuming that they are identified. *Econometrica*, 73(4):1103–1123, 2005.
- T. C. Koopmans and O. Reiersol. The identification of structural characteristics. *The Annals of Mathematical Statistics*, 21(2):165–181, 1950.

- H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek. LoOP: Local outlier probabilities. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM '09*, pages 1649–1652, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-512-3.
- F. E. Kydland and E. C. Prescott. Time to build and aggregate fluctuations. *Econometrica*, 50(6):1345–70, 1982.
- D. O. Loftsgaarden and C. P. Quesenberry. A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics*, 36:1049–1051, 1965.
- K. I. Penny and I. T. Jolliffe. A comparison of multivariate outlier detection methods for clinical laboratory safety data. *The Statistician*, 50(3):295–308, 2001.
- S. Ramaswamy, R. Rastogi, and K. Shim. Efficient algorithms for mining outliers from large data sets. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, SIGMOD '00, pages 427–438, New York, NY, USA, 2000. ACM.
- J. D. Sargan. The estimation of economic relationships using instrumental variables. *Econometrica*, pages 393–415, 1958.
- J. D. Sargan. The estimation of relationships with autocorrelated residuals by the use of instrumental variables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 21(1):91–105, 1959.
- A. A. Smith, Jr. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8:S63–S84, 1993.
- J. H. Stock and J. H. Wright. GMM with weak identification. *Econometrica*, 68(5):1055–1096, 2000.
- J. H. Stock, J. H. Wright, and M. Yogo. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business Economics & Statistics*, 20(4):518–529, 2002.
- K. Zhang, M. Hutter, and H. Jin. *A New Local Distance-Based Outlier Detection Approach for Scattered Real-World Data*, pages 813–822. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

A Lack of existence in indirect inference with misspecification

In this appendix, we show indirect inference does not guarantee existence for misspecified models.

Consider the following indirect inference setup. Let the auxiliary model be $y = \gamma_1 + \gamma_2\epsilon$ with $\epsilon \sim N(0, 1)$. Then the maximum likelihood estimates (MLE) for γ_1 and γ_2 are the mean and square root of the sample variance. Let γ_1^* and γ_2^* denote the MLE estimates from the actual data and $\gamma_1(\theta)$ and $\gamma_2(\theta)$ the MLE estimates using simulated model data. For clarity, suppose there is an

infinite amount of both simulated and actual data. Then $\gamma_1^*, \gamma_1(\theta)$ are the population mean from the data and model, respectively, and $\gamma_2^*, \gamma_2(\theta)$ are the population standard deviation from the data and model.

If indirect inference is done by trying to find a θ^* such that $\gamma_i(\theta^*) = \gamma_i^*$, then evidently the deep model must be able to simultaneously match the data’s mean and variance. Under misspecification, this is not always possible. E.g., suppose the data generating process (DGP) is $\tilde{y} = 1 + 2\epsilon$ with $\epsilon \sim N(0, 1)$ —so the auxiliary model in fact nests the DGP. Then if the deep model is an exponential with parameter θ , existence will not hold because the mean is θ^{-1} (which requires $\theta^* = 1$) and the variance is θ^{-2} (which does not equal 4 at $\theta^* = 1$). Hence, indirect inference does not guarantee existence when the model is misspecified.

Unsurprisingly then, this is also true with the [Gallant and Tauchen \(1996\)](#) approach. For instance, letting $\phi(y; \gamma_1, \gamma_2)$ denote the normal density with mean γ_1 and standard deviation γ_2 , the expected score of the auxiliary model is

$$\mathbb{E} \left[\begin{array}{c} \frac{\partial \log(\phi(y; \gamma_1, \gamma_2))}{\partial \gamma_1} \\ \frac{\partial \log(\phi(y; \gamma_1, \gamma_2))}{\partial \gamma_2} \end{array} \right] = \mathbb{E} \left[\begin{array}{c} \frac{y - \gamma_1}{\gamma_2} \\ -\frac{1}{\gamma_2} + \frac{(y - \gamma_1)^2}{\gamma_2^3} \end{array} \right] = \left[\begin{array}{c} \frac{\mathbb{E}(y) - \gamma_1}{\gamma_2} \\ -\frac{1}{\gamma_2} + \frac{1}{\gamma_2^3} \mathbb{E}(y - \gamma_1)^2 \end{array} \right]. \quad (3)$$

For existence, there must be a θ such that this expectation—taken with respect to model data—equals zero when evaluated at γ_1^*, γ_2^* . From the first component of the vector, this requires $\mathbb{E}(y) = \gamma_1$, i.e., the deep model must be able to reproduce the mean in the data. Supposing θ is chosen to make this hold, the second component then requires that the variance of deep-model-generated data at θ equals γ_2^{*2} .²³ So if the deep model cannot simultaneously reproduce the mean and variance in the data, existence will not hold. The example DGP and deep model discussed above provide an example where this is the case.

B The Chatterjee and Eyigungor (2012) model

In this appendix, we more thoroughly describe the [Chatterjee and Eyigungor \(2012\)](#) model.

A sovereign seeks to maximize $\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t)$ where u is the period utility function and β the time discount factor. The sovereign’s “output” is a stochastic endowment consisting of a persistent component y that evolves according to a Markov chain plus an i.i.d. shock $m \sim U[-\bar{m}, \bar{m}]$. As [Chatterjee and Eyigungor \(2012\)](#) discuss, the m shock’s role is simply to aid convergence.²⁴ A default triggers an entry to autarky, which entails (1) a direct loss in output $\phi(y) = \max\{0, d_0 y + d_1 y^2\}$; (2) exclusion from borrowing; and (3) m being replaced with $-\bar{m}$. The sovereign returns from autarky with probability ξ .

²³The second component requires $-1/\gamma_2^* + \mathbb{E}(y - \gamma_1^*)^2/\gamma_2^{*3} = 0$ or $\mathbb{E}(y - \gamma_1^*)^2 = \gamma_2^{*2}$. If θ is chosen to have $\mathbb{E}(y) = \gamma_1^*$, then $\mathbb{E}(y - \gamma_1^*)^2$ is the variance, which must equal γ_2^{*2} .

²⁴[Chatterjee and Eyigungor \(2012\)](#) use a 250 point grid for debt and find m helps in obtaining convergence of the value and price functions. Our approach is slightly different, using a 2500 point grid to help convergence rather than the m shock (i.e., we take $\bar{m} = 0$). With standard methods, our approach could be a hundred times slower, but we employ binary monotonicity as proposed by [Gordon and Qiu \(2017\)](#) to vastly speed the computation.

Every unit of debt matures probabilistically at rate $\lambda \in (0, 1]$ so that $\lambda = 1$ is short-term debt and $\lambda \in (0, 1)$ is long-term debt. In keeping with the literature, let b denote assets so that $-b$ is debt. The total stock of assets evolves according to $b' = x + (1 - \lambda)b$ with $-x$ being new debt issuance. The price paid on debt issuance is $q(b', y)$, which depends only on b' and y as they are sufficient statistics for determining future repayment rates. For each unit of debt that does not mature, the sovereign must pay a coupon z . The sovereign's budget constraint when not in autarky is $c = y + m + (\lambda + (1 - \lambda)z)b + q(b', y)(b' - (1 - \lambda)b)$. In autarky, the sovereign's budget constraint is $c = y - \phi(y) + (-\bar{m})$.

When the sovereign is not in autarky, he compares the value of repaying debt with the value of default and optimally chooses to default, $d = 1$, or not, $d = 0$. Let the default policy function be denoted $d(b, y, m)$ and the next-period asset policy be denoted $a(b, y, m)$. With r_f the risk-free world interest rate and risk-neutral (foreign) purchasers of sovereign debt, bond prices must satisfy

$$q(b', y) = \mathbb{E}_{y', m' | y} \left[(1 - d(b', y', m')) \frac{\lambda + (1 - \lambda)(z + q(a(b', y', m'), y'))}{1 + r_f} \right]. \quad (4)$$

Equilibrium is a fixed point in which (1) q satisfies the above functional equation given d and a ; and (2) d and a are optimal given q .

The mapping of the model to data for consumption and output is the obvious one. Less obvious is the definition of spreads and net exports. Following [Chatterjee and Eyigungor \(2012\)](#), we define spreads for a given (b', y) pair as the difference between “internal rate of return” r and the risk-free rate r . Specifically, r is defined as the solution to²⁵

$$q(b', y) = \frac{\lambda + (1 - \lambda)(z + q(b', y))}{1 + r(b', y)} \quad (5)$$

and spreads are $(1 + r)^4 - (1 + r^f)^4$ since the model is calibrated to quarterly time periods. Net exports is more simply defined as output less consumption, which is $y + m - c$ when not in autarky and $y - \phi(y) + (-\bar{m}) - c$ in autarky.

²⁵[Chatterjee and Eyigungor \(2012\)](#) define it implicitly as $q(b', y) = (\lambda + (1 - \lambda)z)/(\lambda + r(b', y))$, which is equivalent to the definition in (5).