

# Robustness Checks in Structural Analysis\*

Sylvain Catherine<sup>†</sup>  
David Sraer<sup>§</sup>

Mehran Ebrahimian<sup>‡</sup>  
David Thesmar<sup>¶</sup>

January 14, 2023

## Abstract

Robustness checks are a standard feature of reduced-form empirical research. Because of computational costs of re-estimating alternative models, they are much less common in structural research using simulation-based methods. We propose a simple methodology to bypass this computational cost. Our approach is based on estimating a flexible approximation of the relation between moments and parameters. It provides a computationally cheap way to run the potentially large number of structural estimations required for such robustness checks. We demonstrate the validity and usefulness of this methodology in the context of two standard applications in economics and finance: (1) dynamic corporate finance (2) portfolio choice over the life cycle.

---

\*For their useful comments, we thank seminar participants at Berkeley, NYU, Stockholm School of Economics and Wharton, as well as conference participants at the Minnesota Corporate Finance Conference, especially Dan Green for his discussion.

<sup>†</sup>University of Pennsylvania

<sup>‡</sup>Stockholm School of Economics

<sup>§</sup>UC Berkeley, NBER and CEPR

<sup>¶</sup>MIT, NBER and CEPR

# 1 Introduction

Robustness checks are a standard feature of applied empirical work in economics and finance. After establishing their main results, researchers commonly ask if alternative mechanisms can explain their findings and provide additional analyses that control for such channels. For instance, researchers reestimate their main specification across various subsamples, either because they expect results to be stable, or vary in a certain expected way. In other situations, they estimate different specifications that account for alternative channels. They expect their main predictions to hold in these amended specifications. These analyses (sample splits, changing the model) belong to the standard toolbox used by empirical researchers.

Such robustness checks are rare in structural research that uses simulation-based methods. The reason is mostly computational. Consider for instance robustness checks in reduced-form work that introduce additional control variables in regressions. Such an analysis entails close to zero computational cost. With simulation-based estimation, the equivalent exercise requires to (1) consider alternative models (2) generate data based on these alternative models (3) re-estimate the baseline model on these data, and (4) assess how estimates change across the various models. In many cases, the computational burden of reestimating the structural model implies that very few alternatives, if any, can be considered in practice. Similarly, sub-sample analysis, e.g., re-estimating a model for particular periods or groups of observations, can prove computationally prohibitive. These limitations harm the credibility of structural estimates, since they make it difficult to assess how sensitive structural findings are to particular assumptions, or particular features of the data. This paper develops a methodology to bypass these computational limitations, and shows how it can be used to perform robustness checks in the context of two standard economic models.

Our approach works as follows. We consider the structural estimation of an economic model through a simulated method of moments. Given structural parameters  $\theta$ , the model generates moments  $\mathbf{m} = f(\theta)$ . In most applications, as for instance with dynamic models,  $f$  is calculated numerically. Structural parameters  $\theta$  are then estimated by minimizing a distance between simulated moments  $f(\theta)$  and empirical moments  $\hat{\mathbf{m}}$ . While simulating the economic model given  $\theta$  is in general reasonably fast, estimating  $\theta$  can be costly as it can require a very large number of such simulations. This computational cost often limits the number of estimations (e.g., of alternative models, or sample splits) that are feasible for a given research project.

We propose to reduce this computational cost using an approximation of  $f$ . Importantly, we do not approximate the numerical solution of the *model*, an approach that has been explored in recent research in macroeconomics and finance (e.g., [Fernandez-Villaverde et al. \(2021b\)](#), [Fernández-Villaverde et al. \(2021a\)](#), [Duarte \(2020\)](#)). Instead, we directly rely on a parametric approximation of the *moment function*  $f(\theta)$ , which we call  $g(\theta, \beta)$ , where  $\beta$  is a vector of parameters that characterizes this approximation. Our approach boils down to estimating  $\beta$ . The numerical solution of the underlying model can be non-linear and therefore hard to approximate. In contrast, the function  $f()$ , which maps structural parameters to moments, is smooth. This feature allows us to obtain a good approximation of  $f()$  at low computational costs.

Our methodology proceeds in three steps. First, we draw a large number  $N$  of potential parameters  $(\theta_i)_{i \in [1, N]}$  and simulate corresponding moments  $m_i = f(\theta_i)$ . This generates a dataset  $\mathcal{D} = \{(\theta_i, f(\theta_i))\}_{i \in [1, N]}$ , which is fixed once and for all. This simulation stage is the computationally intensive step in our approach. However, it is not more costly than estimating the model *once* using standard estimation techniques: we select  $N$  to match the typical number of simulations required to estimate the model *once*. Second, we use the dataset  $\mathcal{D}$  to fit  $g(\theta, \beta)$ , the parametric approximation of  $f(\theta)$ .  $\beta$  can be estimated via regressions or more advanced statistical learning methods (like random forests, which we explore). This step is computationally cheap. Third, we can estimate parameters  $\hat{\theta}$  that match any given moment value  $\hat{m}$  using the approximate model  $g(\theta, \beta)$  at close to zero computational cost. As discussed in Section 2, this approach can be seen as a generalization of [Andrews et al. \(2017a\)](#), who restrict  $g(\theta, \beta)$  to be a linear function of  $\theta$ . In our applications, it is often the case that  $g()$  is non-linear.

We apply our methodology to two canonical models: (1) a dynamic corporate finance model similar to [Hennessy and Whited \(2007a\)](#), and (2) a life-cycle consumption and portfolio choice model similar to [Viceira \(2001\)](#) and [Cocco et al. \(2005\)](#). In each application, we estimate the approximate moment function  $g()$  on a “training sample” of parameters and associated moments. In a separate, “validation sample”, we then compare our parameter estimates with the true parameters that generated the moments. In both applications, the correlation between estimated and true parameters across is never lower than .95, and in nearly all cases larger than .99.

Using the approximate moment function  $g(\theta, \hat{\beta})$  instead of the true function  $f(\theta)$  allows us to run identification diagnostics and robustness checks that would otherwise be computationally prohibitive. We first consider the sensitivity of parameter estimates to targeted moments. Discussions of identification in structural work typ-

ically focus on the function  $f(\theta)$  in the vicinity of  $\hat{\theta}$ . But, as pointed out by Andrews et al. (2017b), a preferable diagnostic tool is parameter estimates  $\hat{\theta}$  as a function of empirical moments  $\widehat{m}$ . Uncovering this relationship requires re-estimating the model for each moment value, a computationally expensive approach. Our approximate moment function is a much more efficient way of doing this, even if  $g()$  is non-linear. In both our applications, this non-linearity turns out to be important.

Our methodology also makes it feasible to assess the robustness of parameter estimates to the *selection* of targeted moments. A common practice to assess a model's performance is to (1) divide the set of relevant moments into moments targeted in estimation and non-targeted moments (2) compare non-targeted moments in the data to their simulated counterparts. This approach is not ideal, because it does not *quantify* the effect of matching non-targeted moments on parameter estimates. A better approach is to re-estimate the model on various sets of moments, which is computationally expensive under the traditional approach. Our methodology does not have this problem. In our two applications, we explore thousands of possible moment combinations, and report the distribution of parameter values. We find that few parameter estimates are robust to moment selection.

Another standard concern in empirical work is the robustness of estimates to the estimation sample. In reduced-form analyses, econometricians often re-estimate regression models over different sub-samples. They check that estimates are stable, or vary in an expected directions. In structural work, the equivalent exercise would require to re-estimate the model on each subsample, a computationally expensive exercise if the number of sample splits is large. In contrast, the low computational cost of our approach makes these robustness checks easy to do. As we show in our corporate finance application, this type of analysis can bring interesting insights about the validity of the underlying model.

Finally, our methodology also allows us to test the robustness of estimation to model misspecification: Economic forces present in the data, but not included in the model, alter the inference of structural parameters. In reduced-form work, robustness to misspecification can be easily addressed by adding controls to regression analyses. In structural work, considering alternative models and how they affect estimated parameters of interest is costly as it involves a new estimation of the model for each alternative data-generating process considered. Our methodology offers a way out of this problem, by speeding up the estimation of our baseline model on data generated by alternative models. Let  $\widehat{\theta^{\text{baseline}}}$  be the set of parameters of the baseline model. Consider an alternative model, with structural parameters  $\theta^a$  and moments  $h(\theta^a)$ .

For each value of  $\theta^a$ , we can very quickly estimate the baseline parameters  $\hat{\theta}$  that best fit the alternative model-generate moment  $h(\theta^a)$ . This allows to explore how baseline model inference responds to a large set of misspecifications.

**Related literature.** Our paper mostly relates to a recent literature that tries to make structural estimation more transparent, with a particular focus on the sensitivity of policy predictions to moment or model misspecification. Andrews et al. (2020b) propose a formal definition of transparency in empirical research and apply it to structural estimation in economics. Andrews et al. (2017b) derives a local linear approximation of the relationship between parameter estimates and data moments. This measure is a diagnostic tool to assess misspecification bias. It does not require to “specify the misspecification”, but assumes that misspecification bias is small. Our approximation is global, but misspecification analysis requires to make assumptions about what alternative models are. In follow-up work, Andrews et al. (2020a) propose a way to formalize the relationship between descriptive analysis and structural estimation using a similar local approximation. Our approach also allows to explore additional descriptive statistics at low cost (non-targeted moments, model outputs). More generally, our analysis connects to the literature on robustness to model misspecification (e.g., Huber (2011); Armstrong and Kolesár (2021); Bonhomme and Weidner (2018)).

When papers in the structural literature discuss identification, they typically use the relation between moments and structural parameters – the function  $f(\theta)$ . Table A.1 provides a review of recent structural papers in corporate and household finance. Out of 43 papers, 12 show local comparative statics (i.e. plots of  $f(\theta)$  around  $\hat{\theta}$ ); 8 report the Jacobian matrix (derivatives of  $f()$  around  $\hat{\theta}$ ). 24 papers do not report any of these two. 4 recent papers in Table A.1 reports the sensitivity matrix of Andrews et al. (2017b), i.e. the local derivative of parameter estimates w.r.t. targeted moments. Our approach allows researchers to dispense with linear approximation which, as we show in our examples, is not always valid.

Our paper is more distantly related to the emerging literature trying to improve numerical solutions of *models* through approximations. For instance, Norets (2012) uses neural networks to approximate the solution of a finite-horizon, dynamic discrete choice model. Similarly, Duarte (2020) describes a new solution method combining ML algorithms and Gradient Descent Algorithm. Chen et al. (2021) use predictive algorithms to estimate the solution of models, after these are numerically solved on a

training sample.<sup>1</sup> Our focus is different from these papers. We seek to approximate *moments* as a function of parameters, not *the value or policy functions*. It is driven by our interest in estimation and robustness analyses.

Finally, our paper contributes to the vast literature that structurally estimate dynamic models of corporate and household finance (see [Strelbulaev and Whited \(2012\)](#) for a survey of the corporate finance literature, and [Gomes et al. \(2021\)](#) for household finance). We assess the robustness of these widely-used models.

## 2 General Approach and Notations

This section lays out our general approach.

### 2.1 Approximation

Let  $\mathcal{S}$  be a structural model with deep parameters  $\theta$ . The model  $\mathcal{S}$  generates a vector of moments  $f(\theta) \in \mathbb{R}^M$  that have empirical counterparts, where  $M$  is the number of moments targeted in estimation. In simulation-based estimation,  $f()$  does not admit a closed-form representation, and is obtained through simulations. For the sake of clarity, we ignore simulation error, and assume that  $f()$  can be exactly computed with a large number of simulations. Let  $\widehat{m}$  be the vector of empirical counterparts to the model-based moments  $f(\theta)$ .

The minimum distance estimator of  $\theta$  is obtained by minimizing:

$$\widehat{\theta} = \arg \min_{\theta} (\widehat{m} - f(\theta))' W (\widehat{m} - f(\theta)) \quad (1)$$

where  $W$  is a weighting matrix.

The estimation proceeds in two-steps:

1. For a given  $\theta$ , standard numerical methods are used to solve the model given  $\theta$  and simulate model-based moments  $f(\theta)$  (inner loop).
2.  $\theta$  is selected to minimize the objective in 1 (outer loop).

Because  $\theta$  is potentially high-dimensional, model estimation can be computationally costly, as it requires a large number of inner loops, and each inner loop requires to solve and simulate the model. In particular, there are no “economies of scope” in

---

<sup>1</sup>For further references in this space, see also [Fernández-Villaverde et al. \(2021a\)](#), [Villa and Valaitis \(2019\)](#), [Maliar et al. \(2019\)](#), [Azinovic et al. \(2019\)](#).

estimation: if the model has to be estimated against a different set of empirical moments  $\widehat{\mathbf{m}}'$  (e.g., moments estimated for different subsamples, or alternative moments not included in  $\widehat{\mathbf{m}}$ ), the second estimation carries the same computational cost as the first one.

The objective of our paper is to create such economies of scope using an approximation of the function  $f()$ . To do this, we first construct a dataset of moments and parameters to train our approximate moment function. To build this dataset, we implement the following three steps:

- (a) We define ex ante bounds for each parameter in  $\theta$ . These bounds are based on expert knowledge, and also have to be specified for standard estimation techniques in the literature. We note  $\Theta$  the resulting set of admissible parameter vectors.
- (b) We use a Halton sequence<sup>2</sup> to generate  $N$  vectors of parameters  $(\theta_i)_{i \in [1, N]}$  with  $N$  large.
- (c) For each vector  $\theta_i$ , we solve the model  $\mathcal{S}$  and simulate moments  $f(\theta_i)$ . This is computationally intensive as the true model has to be solved and simulated  $N$  times. This step results in a dataset  $\mathcal{D} = \{(\theta_i, f(\theta_i))\}_{i \in [1, N]}$  that contains  $N$  vectors of parameters and the corresponding  $N$  vector of moments.

Building this dataset is the computationally intensive phase, since it requires solving and simulating  $N$  models with different parameter values.

Now, we are set to estimate parameters using the approximate moment function. Imagine we are trying to find the model parameters which match a given set of moments  $\widehat{\mathbf{m}}$ . Estimating these parameters involves the following, very simple, estimation procedure:

- (i) For each draw  $(\theta_i, m_i) \in \mathcal{D}$ , we first compute the distance between  $m_i$  and targeted moment values  $\widehat{\mathbf{m}}$ :  $\Delta_i = (m_i - \widehat{\mathbf{m}})' \Omega (m_i - \widehat{\mathbf{m}})$ .  $\Omega$  is a scaling matrix. In our applications, we choose the inverse of the variance-covariance matrix of the

---

<sup>2</sup>Alternatively, one could sample points over a regular grid. Low discrepancy sequences present two advantages. First, if the researcher underestimates the sample size required to properly approximate  $g$ , the sequence can just be expanded to the next points. On the other hand, a grid-based approach requires to restart the computation from scratch. Second, in high dimension, a regular grid approach would result in most sample points falling on a few hyperplanes. For example, with a sample size 50,000 and 7 parameters, the moment function would be evaluated at only 4 or 5 coordinates for each parameter.

empirical moments in the overall sample, estimated using bootstrap. This matrix corresponds to the efficient weighting matrix for estimation in Equation 1. A draw  $(\theta_i, m_i)$  that has a large  $\Delta_i$  is far from the moments we seek to target.

- (ii) We estimate the approximate moment function  $g(\theta; \beta)$ . In our applications, we consider three classes of functions for  $g()$ : linear, third degree polynomials and neural nets. The parameters  $\beta$  are estimated via minimum-distance:

$$\hat{\beta} = \arg \min_{\beta} \left[ \sum_{i=1}^N \frac{1}{(\Delta_i)^k} (g(\theta_i; \beta) - f(\theta_i))' (g(\theta_i; \beta) - f(\theta_i)) \right]$$

$k$  measures the weight the approximate model puts on observations in  $\mathcal{D}$  with moments close to the targeted moments  $\widehat{m}$ . When  $k = 0$ , the approximation is estimated using all elements of  $\mathcal{D}$  equally. As  $k$  increases, the approximation puts increasing weights on elements of  $\mathcal{D}$  that are near the targeted moments  $\widehat{m}$  that are targeted in the estimation.

- (iii) We use the approximation  $g(\theta; \hat{\beta})$  to estimate the deep parameters  $\theta$  that match the baseline moments  $\widehat{m}$ :

$$\hat{\theta}(\widehat{m}; \hat{\beta}) = \arg \min_{\theta} (\widehat{m} - g(\theta; \hat{\beta}))' W (\widehat{m} - g(\theta; \hat{\beta})) \quad (2)$$

This estimation is fast since simulations are no longer required. Since  $g()$  is a closed-form, smooth function of  $\theta$ , it is fast to estimate  $\beta$ , and then to estimate  $\theta$ . We do not have formulas for standard errors of  $\hat{\theta}(\widehat{m}; \hat{\beta})$ , since the model is misspecified:  $g()$  is not the right model if  $f()$  is, so inference may be biased and noisy. In our applications below, we show that the error induced by the approximation is, in practice, small. Finally, note that the approximate moment function  $g()$  is estimated by giving more weight to observations near the targeted moments (if  $k > 0$ ). This approach ensures a better fit to the true moment function  $f()$  near the region that we are interested in exploring.

## 2.2 Relation with the Sensitivity Matrix

Andrews et al. (2017b) propose a tool to assess the robustness of parameter estimates to misspecification. Their analysis relies on a local linear approximation of the mapping from targeted moments to parameter estimates. Conceptually, their approach builds on the idea that, for small variations in moments around their empirical value,

the model can be costlessly estimated using a linear approximation. However, for the type of robustness exercises we consider in this paper, such a linear approximation may be too coarse. Our approach can thus be seen as a generalization of Andrews et al. (2017b) that relies on a higher-order approximation.

More precisely, consider any vector of parameter  $\theta \in \Theta$  (possibly, an SMM estimate), and  $m = f(\theta)$  the corresponding moments. Let  $\tilde{\theta}$  be in the vicinity of  $\theta$ . Then, provided  $f()$  is differentiable,

$$f(\tilde{\theta}) \approx \underbrace{m + J(\theta)(\tilde{\theta} - \theta)}_{=g^{\text{linear}}(\tilde{\theta})},$$

where  $J(\theta) = \nabla f(\theta)$  is the Jacobian matrix. With this linear expansion, the approximation  $g^{\text{linear}}(\cdot)$  does not need to be estimated: its parameters are simply given by the moments  $m$  and the Jacobian  $J(\theta)$ .

Assume now that the econometrician wants to estimate the model using the linear approximation  $g^{\text{linear}}$  of the true model  $f()$  by targeting moments  $\widehat{m}$ . The parameter estimates  $\widehat{\theta}$  are then defined by the following First-Order Condition (FOC) from the optimization program (2):

$$0 = J(\theta)'W(\widehat{m} - g^{\text{linear}}(\widehat{\theta})) = J'(\theta)'W(\widehat{m} - m - J(\theta)(\widehat{\theta} - \theta))$$

This formula can be rewritten to make explicit the link between our approximate parameter estimate and Andrews et al. (2017a)'s sensitivity matrix  $\Lambda$ :

$$(\widehat{\theta} - \theta) \approx \underbrace{-(J'(\theta)'WJ(\theta))^{-1}J'(\theta)'W}_{\Lambda}(\widehat{m} - m) \quad (3)$$

Assume for instance that  $\theta$  are parameters estimated using SMM and  $m$  corresponds to the empirical moments. Andrews et al. (2017a) argues that 3 can be used to explore how estimated parameters vary for alternative moments  $\widehat{m}$  that differ slightly (in a first-order sense) from  $m = f(\theta)$ . Conceptually, Andrews et al. (2017a) suggest that, for moments in the neighborhood of  $m$ , parameters can be estimated without directly through the linearization formula 3.

However, this approach only works locally (or if the relationship between moments and parameters is in fact linear). As alternative targeted moments get further away from the baseline moments, the quality of the approximation, and thus of the estimation, may deteriorate. Since most of the robustness checks we consider requires re-estimating the model using a wide range of alternative targeted moments, this is

a key issue for our approach. Section 3 and 4 below explore the relative precision of estimations that rely on non-linear approximations relative to linear approximations.

## 3 Dynamic Corporate Finance Model

### 3.1 Model Layout

We use a standard model of firm dynamics with collateral constraints. It is similar to [Hennessy and Whited \(2007b\)](#) or [Ottonello and Winberry \(2020\)](#). The frictions are adjustment costs to capital, a tax shield for debt, costly equity issuance and a collateral constraint. The firm's shareholder is risk-neutral and her discount rate is  $r = .03$ . At date  $t$ , the firm's EBIDTA,  $\pi_t$ , depends on the firm's capital stock and its productivity:

$$\pi_t = e^{z_t(1-\alpha)} k_t^\alpha, \quad (4)$$

with  $z_t$  the firm's productivity, which follows an AR(1) process  $z_t = \rho_z z_{t-1} + \eta_t$ . The variance of the innovation term  $\eta_t$  is  $\sigma_z^2$ .

Capital accumulation is subject to depreciation, time to build, and adjustment costs:

$$k_{t+1} = k_t + i_t - \delta k_t, \quad (5)$$

where  $\delta$  is the depreciation rate. In period  $t$ , investing  $i_t$  entails a convex cost of  $\frac{\gamma i_t^2}{2 k_t}$ . Additionally, the firm pays in period  $t$  for capital that will only be used in production in period  $t+1$ . This one period time-to-build for capital is conventional in the macroeconomic literature ([Hall, 2004](#); [Bloom, 2009](#)) and acts as an additional adjustment cost. Firms' profits net of interest payments and capital depreciation ( $\delta k_t$ ) are taxed at a rate  $\tau = 1/3$ . This tax rate applies both to negative and positive income so that firms receive a tax credit when their accounting profits are negative.

The firm finances investment out of retained earnings, debt, and equity issuance to outside investors.  $d_t$  is net debt, so that  $d_t < 0$  means that the firm holds cash. We set up the model so that debt is risk-free and pays an interest rate  $r$ . As is standard in the structural corporate finance literature ([Hennessy and Whited, 2005](#)), we only consider short-term debt contracts with a one period maturity. For an amount  $d_{t+1}$  of debt issued at date  $t$ , the firm commits to repay  $(1+r)d_{t+1}$  at date  $t+1$ .

Financing frictions come from the combination of two constraints. First, issuing equity is costly. If pre-issuance cash-flows are  $x$ , cash-flows net of issuance costs are given by:

$$G(x) = x(1 + \zeta 1_{x < 0})$$

where  $\zeta > 0$  parameterizes the cost of equity issuance. Second, firms face a collateral constraint, which emanates from limited enforcement (Hart and Moore, 1994). We follow Liu et al. (2013) and adopt the following specification for the collateral constraint:

$$d_{t+1} \leq \lambda k_{t+1}. \quad (6)$$

The total collateral available to the creditor at the end of period  $t + 1$  consists of depreciated productive capital  $\lambda k_{t+1}$ .  $\lambda$ , the share of the collateral value realized by creditors, captures the quality of debt enforcement, but also the extent to which collateral can be redeployed and sold.

The firm is infinitely lived. Every period, physical capital and debt are chosen optimally to maximize a discounted sum of per-period cash flows, subject to the financing constraint. The firm takes as given its productivity, and forms rational expectations about future productivity. Firm behavior is represented by a Bellman equation whose solution is the present value of future cash-flows, maximized over capital  $k_{t+1}$  and debt  $d_{t+1}$ , under the collateral constraint. This value is a function of the elements of the state space:  $(k_t, d_t, z_t)$ . This Bellman equation is written in Catherine et al. (2022b).

Financial frictions in the model result in value losses. The model offers a simple statistic to gauge the economic importance of these frictions: The average value increase that constrained firms would experience if financial frictions were entirely removed. Precisely, let  $V^c(k, d, z)$  be the value of a firm with state variable  $(k, d, z)$  in the model with financial frictions. Define  $V^*(k, d, z)$  the value of a firm with state variable  $(k, d, z)$  in the absence of financial frictions (i.e. when equity issuance cost is zero,  $\zeta = 0$ ). We define the value loss as:

$$\text{Value loss} = \mathbb{E} [\log(V^*(k, d, z)) - \log(V^c(k, d, z))] ,$$

using the ergodic distribution of  $(k, d, z)$  in the model with financial frictions. Beyond structural parameters, we also report below how our approximation affects the estimation of this statistic.

## 3.2 Training Dataset

We consider a vector of seven structural parameters to be estimated:  $\boldsymbol{\theta} = (\delta, \gamma, \alpha, \rho_z, \sigma_z, \zeta, \lambda)$ . We restrict these parameters to  $\delta \in [0; .2]$ ,  $\gamma \in [0; .3]$ ,  $\alpha \in [.5; .9]$ ,  $\rho_z \in [.5; .98]$ ,  $\sigma \in [0.2; 1.5]$ ,  $\zeta \in [0; .3]$  and  $\lambda \in [0; .6]$ .

We then draw a Halton sequence of  $N = 50,000$  vectors  $\boldsymbol{\theta}_j$  – we verify below that  $N$  is large enough to generate a good approximation. For each  $\boldsymbol{\theta}_j$ , we simulate the model and compute a vector of 17 moments  $\mathbf{m}_j$ , which corresponds to the set of moments that have been used in the literature. The 17 moments are listed in Table 1. The first seven moments are typically targeted in the literature. The next 10 moments are less frequently targeted, so we will use them in our robustness exercises.

Table 1: Simulation Moments (Corporate Finance Model)

		Data	True SMM	Approximate	Simulation
$m_1$	<b>mean(investment/assets)</b>	.0760 (.0007)	.0761	.0760	.0747
$m_2$	<b>mean(profit/assets)</b>	.1343 (.0012)	.1343	.1343	.1342
$m_3$	<b>mean(equity issuance/assets)</b>	.0158 (.0007)	.0158	.0158	.0150
$m_4$	<b>mean(leverage)</b>	.1049 (.0030)	.1051	.1049	.1080
$m_5$	<b>autocorr(investment/assets)</b>	.3754 (.0067)	.3753	.3755	.3907
$m_6$	<b>std(log sales growth)</b>	.2270 (.0017)	.2270	.2270	.2252
$m_7$	<b>std(log sales growth 5yr)</b>	.5851 (.0052)	.5854	.5851	.5800
$m_8$	var(investment/assets)	.0033 (.0001)	.0167	.0170	.0156
$m_9$	var(equity issuance/assets)	.0071 (.0002)	.0024	.0021	.0021
$m_{10}$	frequency(equity issuance)	.1178 (.0015)	.1521	.1605	.1516
$m_{11}$	coeff. regr. investment ratio on market/book	.0122 (.0005)	.3190	.3039	.3081
$m_{12}$	coeff. regr. net leverage on market/book	-0.0348 (.0018)	-0.0325	-0.0490	-0.0246
$m_{13}$	coeff. AR(1) regr. of profit ratio	.5210 (.0057)	.5358	.5405	.5445
$m_{14}$	resid std AR(1) regr. of profit ratio	.0728 (.0006)	.0287	.0283	.0285
$m_{15}$	var(leverage)	.0266 (.0004)	.0002	.0004	.0001
$m_{16}$	mean(dividend/assets)	.0267 (.0004)	.0491	.0491	.0492
$m_{17}$	var(dividend/assets)	.0013 (.0001)	.0054	.0052	.0050

Notes. The ‘Data’ column reports moments in the data with standard errors in parenthesis. Column ‘True SMM’ reports simulated moments using the true economic model  $f(\boldsymbol{\theta})$  and parameters estimated using the true SMM. Column ‘Approximate’ reports moments calculated using the benchmark approximation  $g(\boldsymbol{\theta}, \boldsymbol{\beta})$  (third-order polynomials with  $k = 2$ ) and the parameters estimated using the approximate SMM. Column ‘Simulation’ reports moments calculated from the true economic model  $f(\boldsymbol{\theta})$  but using parameter estimates from the approximate SMM. Targeted moments in the SMM are shown in bold font.

The training dataset  $\mathcal{D} = (\boldsymbol{\theta}_j, f(\boldsymbol{\theta}_j))_{j \in [1, 50000]}$  is then used to fit several approximations  $g(\boldsymbol{\theta}, \boldsymbol{\beta})$  of the relationship between moments and parameters. Generating this

training dataset is computationally costly as it requires solving and simulating the model 50,000 times. Since solving and simulating the model takes 10 seconds, creating the training dataset takes about 140 hours. However, note that SMM estimations that use the Tik Tak algorithm (Arnoud et al., 2019) with the same number of starting points require the exact same number of simulations. Thus, this computational fixed cost has to be paid, whether one uses our approximation approach or standard estimation techniques. Note, however, that in our case this cost is paid *only once*: once the training dataset has been generated, additional estimations that target alternative moments *use the same training dataset*.

### 3.3 Validating the Approximation Approach

We start our analysis by assessing the precision of our approximation approach. Once we have estimated our approximate moment function  $g()$  on the training dataset  $\mathcal{D}$ , we draw a *validation* dataset  $\mathcal{D}^{\text{validation}}$  made out of 1,000 additional random draws of parameters  $(\theta_j^{\text{validation}})_{j \in [1, 1000]}$ . For each one of these draws, we use  $g()$  to estimate the parameter  $\hat{\theta}_j^{\text{validation}}$  that best fits the vector of moment  $\mathbf{m}_j^{\text{validation}} = f(\theta_j^{\text{validation}})$ . We then compare the estimated  $\hat{\theta}_j^{\text{validation}}$  with the true  $\theta_j^{\text{validation}}$ .

For this exercise, we consider a just-identified estimation of the seven structural parameters that targets only the first seven moments  $(m_1, \dots, m_7)$ . While the parameters  $\theta$  affect all the moments jointly, the intuition behind the model's identification is the following. Mean(investment/assets) ( $m_1$ ) is tightly connected to depreciation  $\delta$ . Mean(profit/assets) ( $m_2$ ) pins down returns to scale  $\alpha$ . Mean(equity issuance/assets) ( $m_3$ ) informs the model about the cost of issuing equity  $\zeta$ , while mean(leverage) ( $m_4$ ) contributes to the estimation of the collateral parameter  $\lambda$ . Estimates of adjustment cost  $\gamma$  are mostly sensitive to the autocorrelation of investment ( $m_5$ ). Finally, std(log growth sales) ( $m_6$ ) and std(log growth 5yr sales) ( $m_7$ ) help pin down the persistence and volatility of TFP shocks ( $\rho_z$  and  $\sigma_z^2$ ). Note that, while the model is just-identified in this exercise, this is not a requirement for our approach to work.

Because the draws that generate the validation sample  $\mathcal{D}^{\text{validation}}$  are random, many end up in a region where the model is not identified (with the set of targeted moments). For instance, draws with a sufficiently high equity issuance cost all lead to the same moment  $m_3 \approx 0$  since firms stop issuing equity. Thus, in this region, the specific cost of equity issuance used to generate the moments cannot be pinned down. In such cases, estimating the model is impossible, whether using the true SMM approach or our approximation. We thus discard these “non-identified” draws from our

validation dataset. To detect these “non-identified” draws, we would ideally want to calculate the standard errors that a standard SMM approach would estimate when targeting each moment  $m_j^{\text{validation}}$ . However, since  $m_j^{\text{validation}}$  does not come from the data, we cannot estimate its sampling variance. Instead, we take the inverse of the actual variance-covariance matrix of ‘baseline’ moment values from the data,  $\Omega$ . We then simply compute  $(J'(\boldsymbol{\theta}_j^{\text{validation}})\Omega J(\boldsymbol{\theta}_j^{\text{validation}}))^{-1}$  and say that a draw is not identified when one of the resulting standard errors is 10 times larger than the standard error of the parameter estimates in the baseline estimates (i.e. targeting the baseline empirical moments).<sup>3</sup> Among the 1,000 draws that constitute the initial validation dataset, there are 837 non-identified draws, i.e. draws of moments that cannot be identified by the true model.<sup>4</sup> We also drop 5 cases that are either undefined or extreme outliers.<sup>5</sup> The remaining dataset of 158 observations constitute our validation sample.

We now estimate the parameter that best match a given moment vector  $m_j^{\text{validation}}$ . As in Section 2.1, steps (i-ii), we first estimate the approximation function  $g()$ : (i) for each validation moment  $m_j^{\text{validation}}$ , we compute the distance between each training moment  $m_i$  and  $m_j^{\text{validation}}$ :  $\Delta_i^j = (m_i - m_j^{\text{validation}})' \Omega (m_i - m_j^{\text{validation}})$  (ii) we then fit the approximation by minimizing the following objective in the *training dataset*:<sup>6</sup>

$$\hat{\beta} = \arg \min_{\beta} \left[ \sum_{i=1}^N \frac{1}{(\Delta_i^j)^k} (g(\boldsymbol{\theta}_i; \beta) - f(\boldsymbol{\theta}_i))' (g(\boldsymbol{\theta}_i; \beta) - f(\boldsymbol{\theta}_i)) \right] \quad (7)$$

where we experiment with different levels of  $k$ . When  $k = 0$ , observations in the training sample are weighted equally, and the approximation  $\hat{\beta}$  will be the same for all  $m_j^{\text{validation}}$ . As  $k$  increases, the approximation  $g()$  puts larger weights on observations

---

<sup>3</sup>While this selection criterion is somewhat arbitrary, we have experimented with alternative definitions and found that this did not affect our assessment of the estimates’ precision.

<sup>4</sup>Whether equity issuance cost are small enough for firms to ever issue equity depends on other parameters. Even with small issuance costs, firms will rarely issue equity if TFP volatility is low or capital adjustment costs are large. On the other hand, firms may issue equity even when the cost of doing so is large if TFP is very volatile and capital adjustment costs are small. Overall, over most of the large parameter space we explore, firms do not issue equity.

<sup>5</sup>More precisely, we drop cases with either (a) undefined auto-correlation of investment-asset ratio because of zero variance (b) mean and variance of investment-asset ratio of greater than 1 or (c) mean and variance of equity issuance-asset ratio greater than 1.

<sup>6</sup>Some of the moments have natural boundaries, so that we perform a fit for log-transforms of moments rather than their actual values. This allows for better precision in estimation. Specifically, instead of matching their value  $m_i$ , we instead target  $\log(m_i + \epsilon)$ . Regarding the autocorrelation of investment ( $m_5$ ) that is by definition between  $-1$  and  $1$ , we use the transform  $\log\left(\frac{1+(m_i)_5+\epsilon}{1-(m_i)_5+\epsilon}\right)$ . For the case of mean equity issuance ( $m_3$ ) the constant  $\epsilon$  is scaled by the size of the targeted moment through the formula  $\epsilon = \max\{10^{-6}, 10^{-2}(m_j^{\text{validation}})_3\}$ . For all other bounded moments we use a fixed  $\epsilon = 10^{-2}$ .

“near”  $\mathbf{m}_j^{\text{validation}}$  in the sense of the SMM objective function. As a result, when  $k > 0$ ,  $\hat{\beta}$  depends on the validation moment  $\mathbf{m}_j^{\text{validation}}$  and thus needs to be re-estimated each time we seek to target a new moment. However, since fitting the approximation is computationally cheap, this step is very fast.

We consider 5 different functional forms for  $g$ . The first one is a linear function of parameters. We also use a third-order polynomial using up to all three-way interactions of parameters. Third, to directly account for the censoring at 0 of the equity issuance moment, we consider a third-order polynomial approximation augmented with a Tobit model for the equity issuance moment. Finally, we consider two neural net approximations: (a) a two-layer neural network with 10 neurons per layer and a hyperbolic tangent activation function and (b) a five-layer neural network with 10 neurons per layer and a hyperbolic tangent activation function.

Finally, as described in step (iii) in Section 2.1, we estimate parameter  $\hat{\theta}_j^{\text{validation}}$  to match moments  $\mathbf{m}^{\text{validation}}$  by minimizing the approximate-SMM objective:

$$\hat{\theta}_j^{\text{validation}} = \arg \min_{\theta} \left( \mathbf{m}_j^{\text{validation}} - g(\theta; \hat{\beta}) \right)' \Omega \left( \mathbf{m}_j^{\text{validation}} - g(\theta; \hat{\beta}) \right), \quad (8)$$

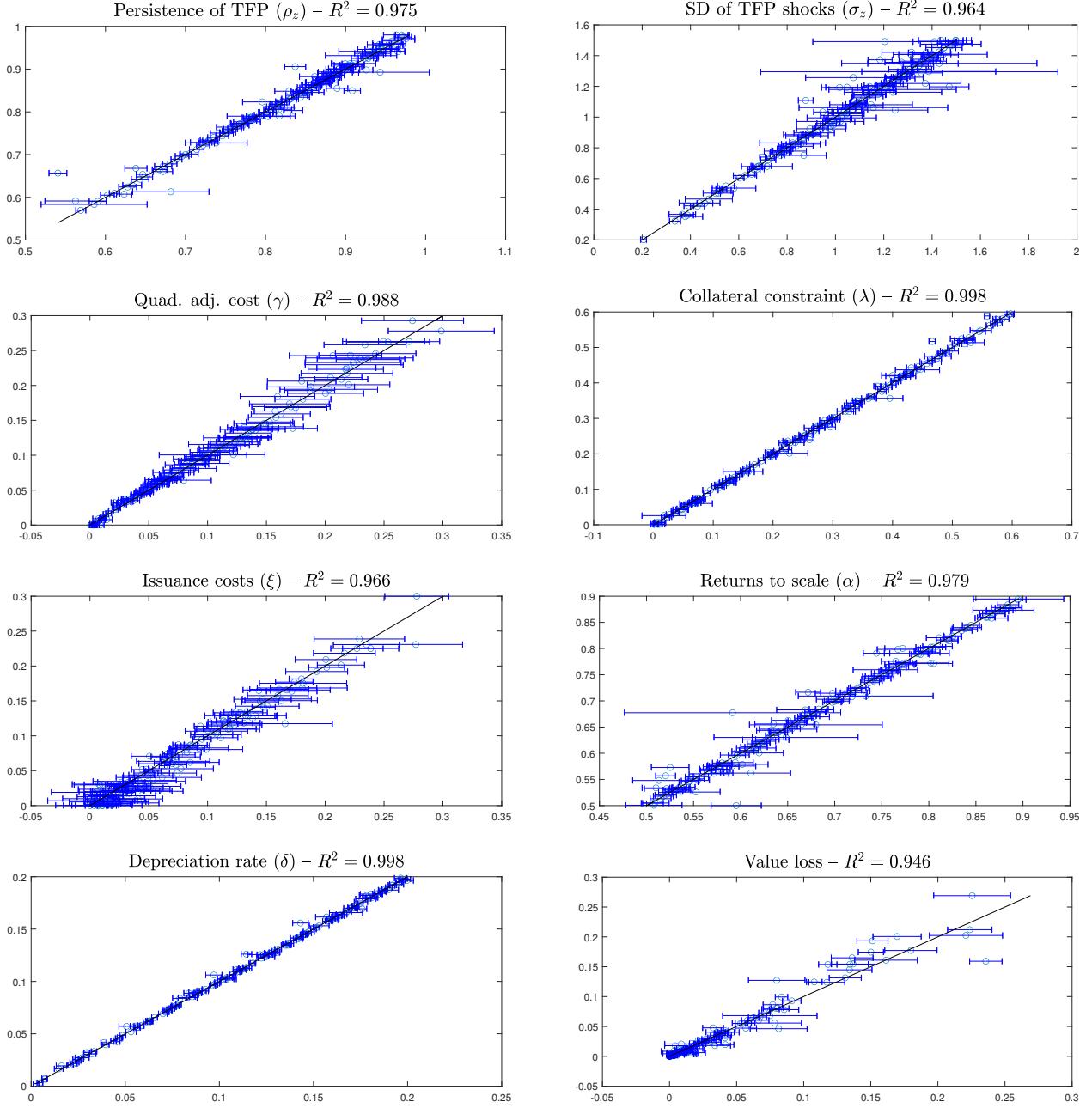
where  $\Omega$  is the inverse of the variance-covariance matrix of the empirical moments.

We reproduce these steps for each moment in our validation sample. We end up with 158 cases where we can compare the true data-generating parameters with the ones estimated with our approach. As it turns out, among the different functional forms for  $g()$ , third order polynomials offer the best fit, along with a quadratic weight of  $k = 2$ . This is shown in Appendix Figure A.1, which, for several values of  $k$  and each functional form, reports one minus the correlation coefficient of true and estimated parameters.

Importantly, all linear fit specifications yield much lower  $R^2$ . This shows that the sensitivity matrix proposed by Andrews et al. (2017a) cannot be relied upon outside the vicinity of a given vector of parameter.

Because of its superior fit, we use third-order polynomials with quadratic weights in the remainder of this application. To visualize the quality of our estimation method, we report scatter plots of estimated versus true parameters in the validation sample in Figure 1. The horizontal lines show the “identifiability” of each parameter, measured through the square roots of the diagonal terms of the matrix  $J'(\theta^{\text{validation}}) \Omega J(\theta^{\text{validation}})$ , where  $\Omega$  is the inverse of the variance-covariance matrix of the ‘baseline’ data moments, and  $J(\theta^{\text{validation}})$  is the Jacobian matrix of the true model around the value of

**Figure 1: Out-of-sample Performance (Corporate Finance Model)**



*Notes.* This figure shows the precision of our benchmark approximate SMM across estimated parameters. For each draw  $(\theta_j^{\text{validation}}, f(\theta_j^{\text{validation}}))$  in the validation sample, we estimate parameters  $\widehat{\theta}_j^{\text{validation}}$  with an approximate SMM targeting moments  $f(\theta_j^{\text{validation}})$ . The x-axis reports the true parameters  $\theta_j^{\text{validation}}$ , while the y-axis reports the estimated parameters  $\widehat{\theta}_j^{\text{validation}}$ . The approximation  $g(\theta; \beta)$  we use in this plot is estimated on the training dataset  $\mathcal{D}$  as  $\widehat{\beta} = \arg \min_{\beta} \left[ \sum_{l \in \mathcal{D}} \frac{1}{(\Delta_l^j)^k} (g(\theta_l; \beta) - f(\theta_l))' (g(\theta_l; \beta) - f(\theta_l)) \right]$  where  $\Delta_l^j = (\mathbf{m}_l - \mathbf{m}_j^{\text{validation}})' \Omega (\mathbf{m}_l - \mathbf{m}_j^{\text{validation}})$ ,  $\Omega$  is the inverse of the variance-covariance matrix of the empirical moments,  $g()$  is a third-order polynomial approximation and  $k = 2$ .

the parameter in the validation set. Across all seven parameters, the large mass of observations on the 45 degree line confirms the findings of Figure A.1: the benchmark approximation does a good job at recovering the true parameters in the vast majority of validation draws. The estimated parameters that do not sit on the 45 degree line correspond mostly to draws in the validation sample that leads to poorly-identified parameters, i.e. moments such that, if estimated against the true model, would also lead to badly estimated parameters.

Appendix Figure A.2 shows how  $1 - R^2$  in the *validation* sample (one minus the share of the variance of true parameters explained by the estimated ones) varies as a function of the *training* dataset size  $N$ . In general, the precision of the approximate SMM does not increase much with the size of the training sample after 15,000 draws. The one exception is equity issuance costs, which are poorly identified for large values and thus benefit from more granularity in the training sample.

### 3.4 Actual Estimation with Approximate SMM

We now evaluate the precision and speed of our approximate SMM using actual data. Moments targeted are the first 7 moments of Table 1. They are estimated on COMPUSTAT over 1970-2019 (see Appendix A.1 for more details). Our SMM estimation uses the Tik Tak algorithm. We initialize the algorithm by evaluating the SMM objective at 50,000 starting points. We then run Nelder-Mead optimizations at the 50 best starting points using at most 200 function evaluations. Our approximate estimation uses the procedure described in the previous Section.

We start with precision. Estimates under both methods are reported in Table 2. For all parameters, both estimators lie within a few percentage points of each other, well within the estimated standard errors. Columns 3,4,5 of Table 1 shows that both estimators are equivalent in their ability to match the empirical moments. For both targeted (first 7 lines) and non-targeted moments (last 10 lines). Column 6 shows that the approximate SMM leads to moments that are both close to the moments generated by the true-SMM (Column 4) and to the data (Column 3).

Appendix Figure A.3 compares the convergence speeds of both approaches, and shows that our approach is faster by several orders of magnitude. The computing times we report exclude the simulation of the training sample, which is required for both estimations. For the true SMM, it takes at least 17 minutes—on top of the Tik-tak global optimization stage—for the estimation to converge to its final value. For some parameters like productivity persistence and volatility or return to scale, full

Table 2: Moment and Parameter Estimates: true vs. approximate SMM (Corporate Finance Model)

	$\rho_z$	$\sigma_z$	$\gamma$	$\lambda$	$\xi$	$\alpha$	$\delta$	value loss
true SMM	.7176	1.1601	.0412	.1090	.0381	.8161	.0671	.0253
- s.e., local deriv.	.0069	.0525	.0023	.0032	.0020	.0076	.0008	.0020
approximate SMM	.7269	1.1204	.0462	.1110	.0375	.8103	.0663	.0240
- s.e., local fit deriv.	.0062	.0465	.0025	.0031	.0017	.0070	.0007	.0018
estimation, lower bound	.5	.2	0	0	0	.5	0	
estimation, upper bound	.98	1.5	.3	.6	.3	.9	.2	

Notes. The table reports the parameter estimates and simulated moments of the corporate finance model presented in Section 3.1. The first line corresponds to parameter estimates using the true SMM. The second line shows parameter estimates using the benchmark approximate SMM (third-order polynomial with  $k = 2$ ).  $\rho_z$  is the persistence of the productivity process.  $\sigma_z$  is standard deviation of innovations to productivity.  $\gamma$  is the capital adjustment cost parameter.  $\lambda$  is the collateral constraint parameter.  $\xi$  is linear equity issuance cost.  $\alpha$  is the return to scale.  $\delta$  is the depreciation rate of capital. ‘s.e., local deriv.’ corresponds to the standard errors of the true SMM parameters, calculated using the Jacobian matrix of the true model. ‘s.e., local fit deriv.’ corresponds to the standard errors of the approximate SMM parameters, calculated using the approximate Jacobian matrix. ‘estimation, lower bound’ (resp. ‘estimation, upper bound’) indicates the lower bound (resp. upper bound) imposed ex ante on an estimated parameter.

convergence requires as much as four hours. In contrast, the approximation-based estimation converges in less than a second. Note that, while the Tiktak algorithm allows researchers to “recycle” the (140-hour long) global optimization stage, many commonly used routines (e.g. simulated annealing, particle swarm) do not. The efficiency gains from using our method relative to such routines would be larger by orders of magnitude.

### 3.5 Using the Approximation to Explore Identification

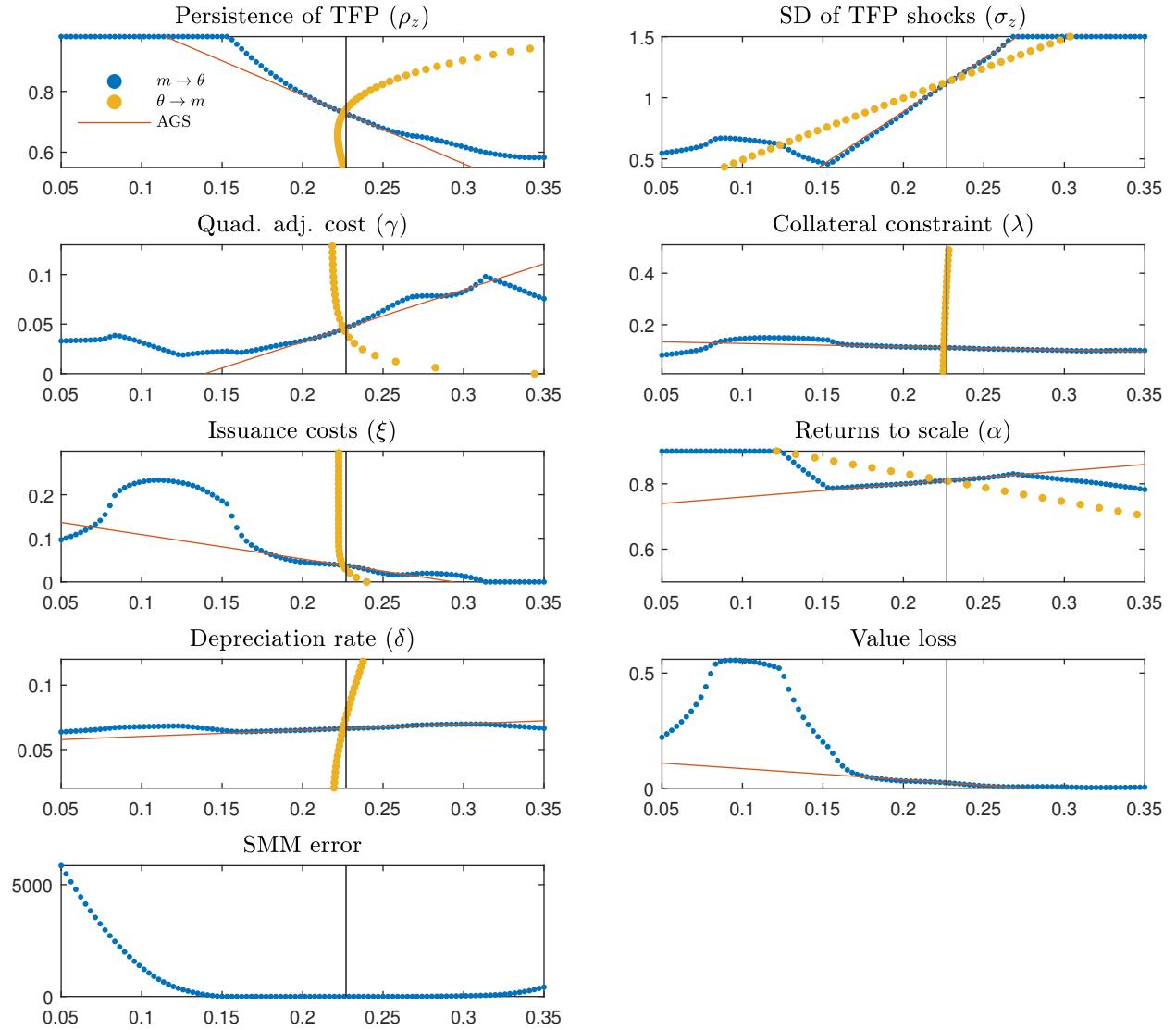
A standard practice in the structural literature is to present *local* comparative statics to discuss identification. Typically, after the model has been estimated, a researcher shows how variations of parameters *around* their estimated values affect simulated moments. This analysis allows researchers to get some intuition for the role of targeted moments in identifying the model’s parameters. However, such intuition is necessarily incomplete: local variations in  $f(\theta)$  only provide partial information about the link from targeted moments  $m$  to estimated parameters  $\hat{\theta}$ .

Because of the low computational cost of an approximate SMM, our methodology allows to fully trace out the mapping from targeted moments  $m$  to estimated param-

eters  $\hat{\theta}$ , even far from the empirical moments  $\widehat{m}$ . This is in contrast to Andrews et al. (2017b), who rely on a local linear approximation around  $\widehat{m}$ . We illustrate this analysis in Figure 2, where the blue line shows parameter estimates for alternative values of  $m_6$ , the standard deviation of 1-year sales log-growth. In the data,  $m_6$  is 23%, and we consider 100 alternative values ranging from 5% to 35%. Drawing this relationship thus requires 100 separate estimations. It is computationally costly with standard simulation-based techniques, but becomes cheap using the approximate SMM. Figure 2 reports two additional curves: (1) the yellow line reports the link between parameter values  $\theta$  and simulated moments  $f(\theta)$  for 100 possible parameter values; it corresponds to the standard local comparative statics shown in many structural papers (2) the red line corresponds to the local linear approximation of the mapping from empirical moments to parameter estimates proposed by Andrews et al. (2017b).

Figure 2 illustrates well how incomplete intuitions about identification can be when relying solely on local comparative statics. For instance, local comparative statics suggest that  $m_6$  is not very useful to identify  $\rho_z$ : for most values of  $\rho_z$  around its estimated value, the simulated value of  $m_6$  is close to its empirical value. However, Figure 2 clearly shows that this interpretation is incorrect: Small variations in  $m_6$  would lead to significant variations in the estimation of  $\rho_z$ . The wedge in interpretation comes from the fact that parameter estimates depend jointly on all the targeted moments. For instance, when the volatility of 1-year sales growth decreases, the estimation naturally finds a lower volatility of TFP shocks  $\sigma_z$ . To keep matching the volatility of 5-year sales growth, a higher persistence of TFP shocks  $\rho_z$  is required. If  $m_6$  further declines, the upper bound for  $\rho_z$  is reached. Below this value, since the persistence of TFP shocks can no longer be adjusted, we see that (1) the SMM error sharply increases (the model fit deteriorates) (2) the model uses a combination of adjustment costs, equity issuance costs, returns to scale, and volatility of TFP shock to try and match the low volatility of one-year sales log-growth. This leads to non-linear variations in these parameters' estimate. Figure 2 also shows that the local linear approximation used in Andrews et al. (2017b) can fail to hold for moment values in a small vicinity of their empirical values. For instance, if  $m_6$  was 15% instead of 23% (its empirical value), the local linear approximation of Andrews et al. (2017b) would suggest an estimated  $\rho_z$  of about .9, while its true estimated value would be close to 1. For parameters that have a more direct connection to  $m_6$  (e.g.,  $\sigma_z$ , the volatility of TFP shocks), the linear approximation works well over a wide range of moment values.

**Figure 2: Sensitivity of parameters to the volatility of sales growth (Corporate Finance Model)**



*Notes.* This figure plots parameter values on the y-axis and the value of the moment  $m_6$  on the x-axis.  $m_6$  is the standard deviation of the one year log sales growth. The yellow line draws local comparative statics, i.e. how variations in one parameter around its estimated value – holding other parameters fixed at their estimated value – affect the value of  $m_6$  obtained through simulations. The blue line plots how variations in the value of the empirical moment  $m_6$  – holding other moments fixed at their empirical value – affects the estimated parameter values. Each dot on the blue line corresponds to a separate estimation. Finally, the red line corresponds to the local linear approximation of the blue line around the parameter estimates, and represents the “sensitivity matrix” of Andrews et al. (2017b): it is a linear approximation of the mapping from moments to parameter estimates around the empirical value of the moments  $m_6$ .

### 3.6 Robustness to Selected Moments

A weakness of methods of moments is that there is no well-established theory of moment selection. The selection of targeted moments is thus in general somewhat arbitrary. Assessing the robustness of parameter estimates to moment selection is thus a key aspect of transparency in structural estimation. Conceptually, it is easy to do: the model needs to be re-estimated using a large set of alternative moments. Robustness is established if parameter values exhibit limited sensitivity to the set of moments targeted in estimation. In practice, however, this exercise cannot be done with standard estimation techniques, as it is computationally prohibitive.

Our approximation approach makes such robustness exercises feasible at low computational cost. An important reason is that, when we simulate the training dataset  $\mathcal{D}$ , including a large number of moments (beyond those used for estimation) comes at almost-zero marginal cost. The computationally intensive step is to solve the Bellman equation; adding an extra-moment computed on simulated data is cheap.

We present two tests of robustness to moment selection.

The first exercise reports how parameter estimates change when we target one of the 10 moments in  $\{m_8, m_9, \dots, m_{17}\}$  *in addition* to the seven baseline moments  $\{m_1, m_2, \dots, m_7\}$  targeted in the baseline estimation of Table 2. Each panel in Figure 3 corresponds to one of the seven estimated parameters (plus the average value loss from financing constraints). The solid black horizontal line corresponds to the baseline estimation in Table 2, which targets the seven baseline moments only. The dashed lines plot the 95% confidence interval. Each coordinate on the x-axis corresponds to one of the 10 moments in  $\{m_8, m_9, \dots, m_{17}\}$ . The y-axis reports the parameter estimate when the estimation targets the seven baseline moment *plus* the additional moment on the x-axis. We also report standard errors for each estimated parameter. These standard errors are calculated using the delta method with the approximate Jacobian matrix.

Figure 3 offers a clear diagnostic tool for robustness to moment selection. For instance, the estimation of the collateral constraint parameter (Panel  $\lambda$ ) is robust to including several additional moments to the set of baseline moments (e.g., the variance of investment ( $m_8$ ), the variance of equity issuance to asset ( $m_9$ ), the frequency of equity issuance ( $m_{10}$ ), the coefficient of a regression of net leverage on market to book ratio ( $m_{12}$ ), the autoregression coefficient of the profit ratio ( $m_{13}$ ), or the variance of dividend to asset ratio ( $m_{17}$ )). However, the estimated collateral constraint parameter becomes significantly larger when the estimation also targets the coefficient of a re-

gression of investment on the market-to-book ratio ( $m_{11}$ ), or the variance of leverage ( $m_{15}$ ). In particular, matching the variance of the leverage ratio almost triples the estimated  $\lambda$ .

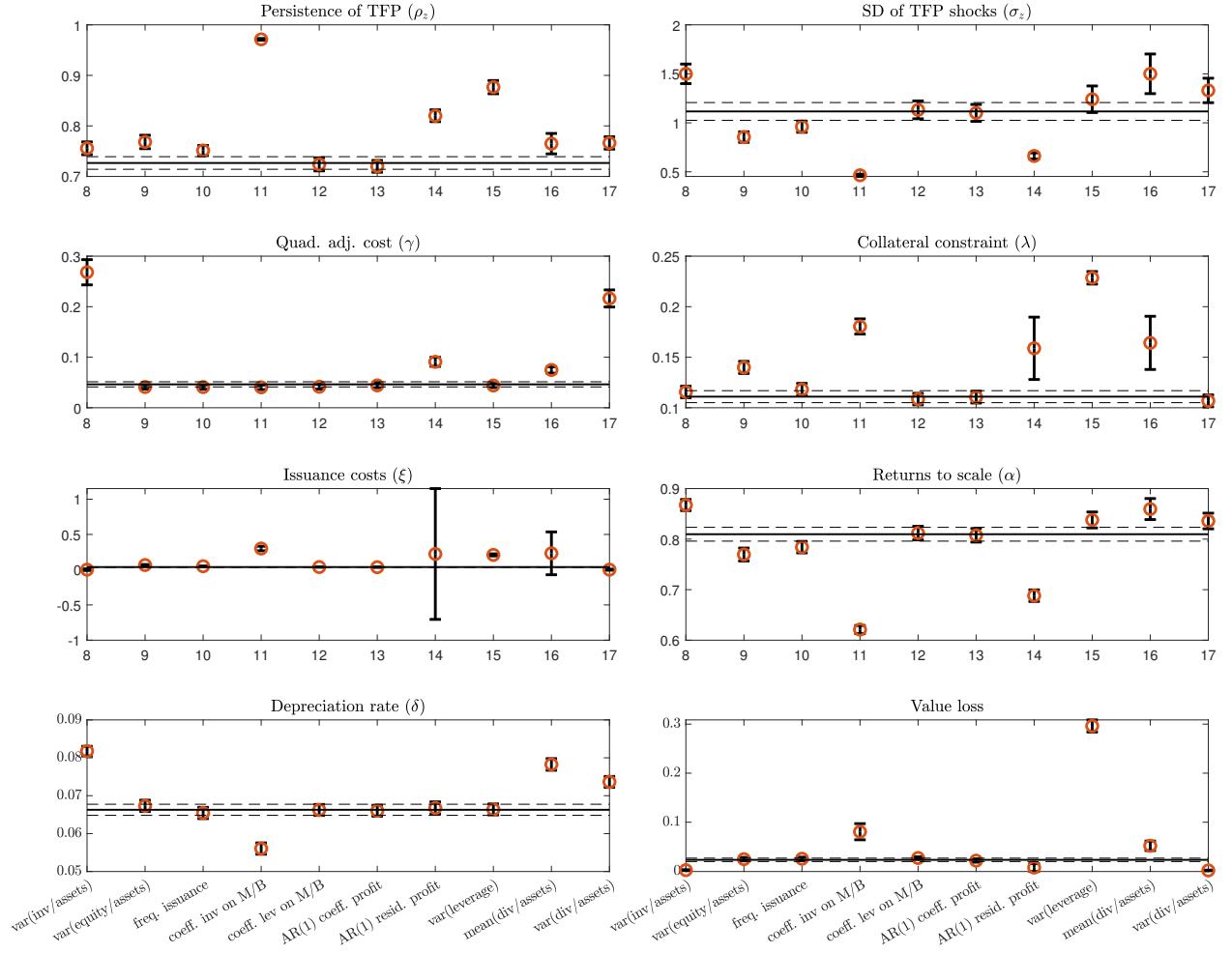
More generally, we see in Figure 3 that the baseline estimation is in general not robust to including the variance of leverage. This is intuitive as the model has a fixed level of pledgeability  $\lambda$ . This makes book leverage quite stable as  $d/k$  remains close to  $\lambda$ , as the firm seeks to maximize the tax shield of debt. Dividend related moments also affect estimates quite a bit. This is intuitive, as firm's dividend policy is poorly explained by traditional investment models in the literature. In the data, firms tend to smooth out dividends for many reasons explored in the literature (e.g., signaling), but not included in the model. To capture this prudent behavior, the estimation that targets dividend policy finds a larger adjustment cost ( $\gamma$ ) and collateral constraint ( $\lambda$ ). Figure 3 also allows us to see that the estimation is overall robust to certain moments, such as the sensitivity of net leverage to MB, or the autoregression coefficient of profits.

The exercise in Figure 3 explores a small set of alternative moment selection. Given the low computational cost of our approximate estimation, we can explore more systematic tests of robustness to moment selection. In Figure 4, we re-estimate the model by targeting (a) the seven moments used in our baseline estimation  $\{m_1, m_2, \dots, m_7\}$  and (b) any possible subset of the other 10 additional moments  $\{m_8, m_9, \dots, m_{17}\}$ . This represents  $2^{10} = 1,024$  separate estimations. Of these, we only consider estimations that correspond to reasonably well-identified sets of moments: We drop cases where the standard errors for a estimated parameter is more than 10 times larger than the standard error of the baseline true SMM.<sup>7</sup> This leaves us with 702 estimates. Figure 4 reports the distribution of estimates across the 702 alternative sets of moments. The solid black lines correspond to the baseline estimates and dashed lines show the 95% confidence interval. A baseline estimate robust to moment selection would have a large share of alternative estimates close to the baseline value.

---

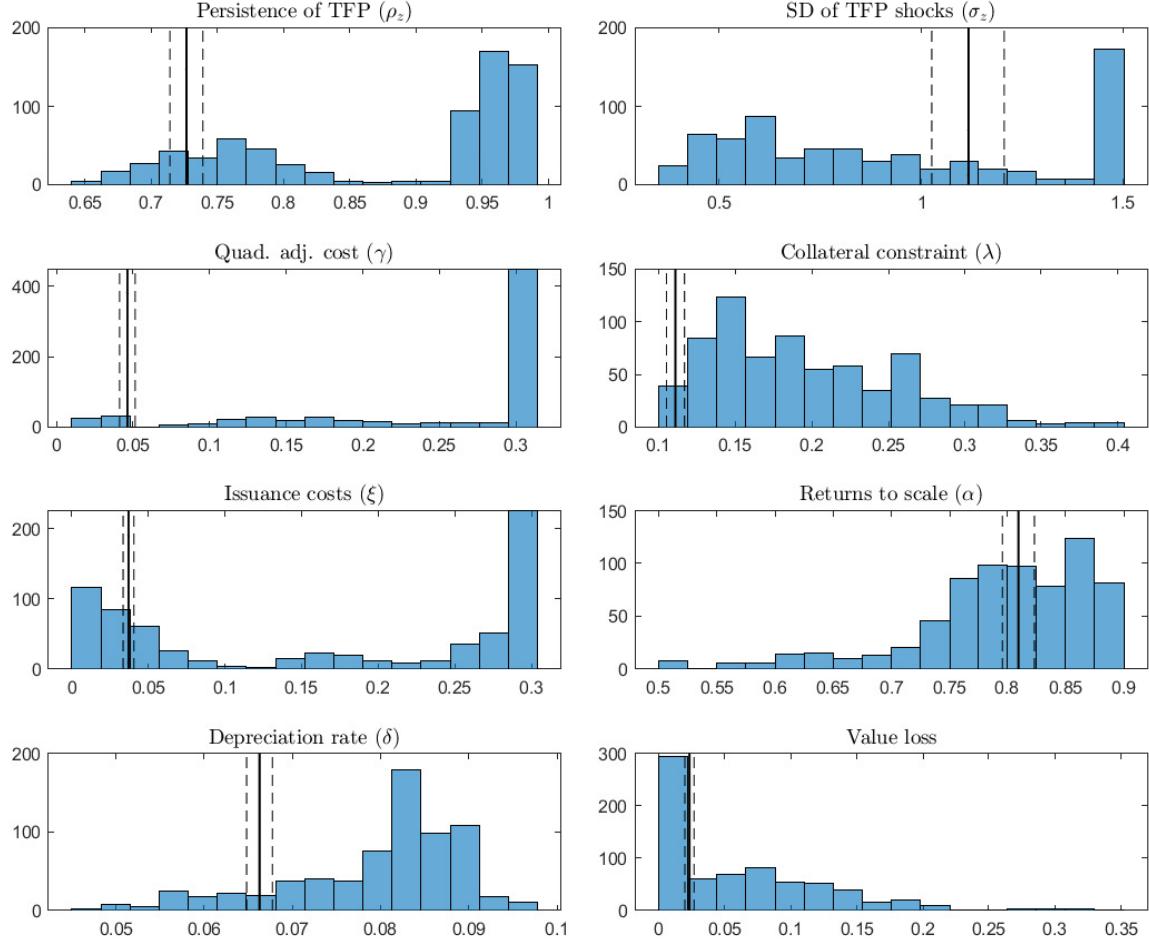
<sup>7</sup>As before, we compute standard errors through the classic SMM formula  $(J'\Omega J)^{-1}$ , using as weight matrix the inverse of the variance-covariance matrix of empirical moments.

Figure 3: Robustness to adding moments one by one (Corporate Finance Model)



*Notes.* This figure explores the sensitivity of parameter estimates to the inclusion of an additional moment in the set of targeted moments. Our baseline estimation targets seven moments  $(m_i)_{i \in \{1..7\}}$ . We consider a set of 10 additional moments used in the literature to estimate similar models:  $(m_i)_{i \in \{8..17\}}$  described in Section 3.2. Each point on the x-axis refers to one additional targeted moment in the list  $(m_i)_{i \in \{8..17\}}$ . The y-axis reports the parameter estimated with the benchmark approximate SMM that targets the baseline moments  $(m_i)_{i \in \{1..7\}}$  and one of the 10 additional moments, along with their 95% confidence interval. The black line and dashed lines show the baseline parameter estimates and their 95% confidence interval. Standard errors are derived using the standard formula  $(J'\Omega J)^{-1}$ , where  $\Omega$  is the inverse of the variance-covariance matrix of empirical moments and  $J$  is the approximate Jacobian matrix computed at the parameter estimates. We use the delta method to calculate SE of value loss.

**Figure 4: Histogram of estimates across 1,024 sets of targeted moments (Corporate Finance Model)**



*Notes.* This figure explores the sensitivity of parameter estimates to moment selection. Our baseline estimation targets seven moments  $(m_i)_{i \in \{1..7\}}$ . We consider a set of 10 additional moments used in the literature to estimate similar models:  $(m_i)_{i \in \{8..17\}}$  described in Section 3.2. We construct all possible sets of moments that contain the seven baseline moments and any combination of the 10 additional moments. These sets of moments are then used as targeted moments in an approximate SMM. These result in 1,024 sets of parameter estimates. After dropping cases where estimates are poorly-identified – where the standard errors for a estimated parameter is more than 10 times larger than the standard errors of the baseline estimate – we end up with 702 estimates. Each panel in the figure shows the distribution of parameters across these 702 estimations. The vertical black line and dashed lines show the baseline parameter estimates using the true SMM, together with their 95% confidence interval.

Clearly, most parameter estimates for this model are not robust to moment selection. For instance, the estimated variance of innovations to  $z$ ,  $\sigma_z$ , which has a benchmark estimate of 1.1., takes many values between 0.4 and 1.1, depending on mo-

ment selection. Many sets of moments also yield an estimate at the upper bound of 1.5. Alternative estimates for the collateral constraint parameter ( $\lambda$ ) or the returns to scale parameter ( $\alpha$ ) are similarly widely distributed. We also see that a large share of estimates for equity issuance cost ( $\zeta$ ) and adjustment cost ( $\gamma$ ) parameters are quite different from their baseline value, suggesting that the baseline estimates are quite specific to the seven moments selected in our baseline estimation.

### 3.7 Sample Splits

A traditional approach to analyzing robustness in reduced-form empirical work is to evaluate a regression model across various sub-samples. For instance, theory might predict that the baseline estimated effect should be stronger (or weaker) for groups with particular characteristics. Estimating a regression separately for different groups then provides a simple way to assess the validity of the findings' interpretation. Alternatively, researchers might be worried that a particular estimated effect is spuriously driven by a subset of the sample (e.g., particular years or particular regions). To evaluate the robustness of the main estimate, the regression can be re-estimated across different sub-samples to check whether the estimates remain similar in the restricted sample.

Such robustness checks can be costlessly performed in a regression setting. However, the equivalent exercise with a structural model can prove costly with simulation-based techniques as it may involve a large number of additional model simulations per estimation. In fact, it is uncommon for structural papers in the literature to report such “sample splits” analyses.

Our methodology provides a low-cost way to implement such exercises. We provide an example in Figure 5. For every year in the sample, we re-estimate the model by targeting the moments estimated over a 10-year rolling window. We report the resulting estimates in Figure 5, with their 95% confidence interval. For each date  $t$ , the targeted moments are calculated over the  $[t-5, t+4]$  window. The red solid and dashed line represents the baseline estimates, with their 95% confidence interval. While some estimated parameters remain relatively stable over time, some experience large trends. For instance, the collateral constraint parameter  $\lambda$  decreases from .2 in the 1970s to close to 0 in the 2000s. This results in a large *increase* in the value loss from financial constraints starting in the 2000. This finding can be simply understood through the lens of the contemporaneous increase in cash holdings over time (Bates et al., 2009). The corresponding reduction in net leverage leads the model to

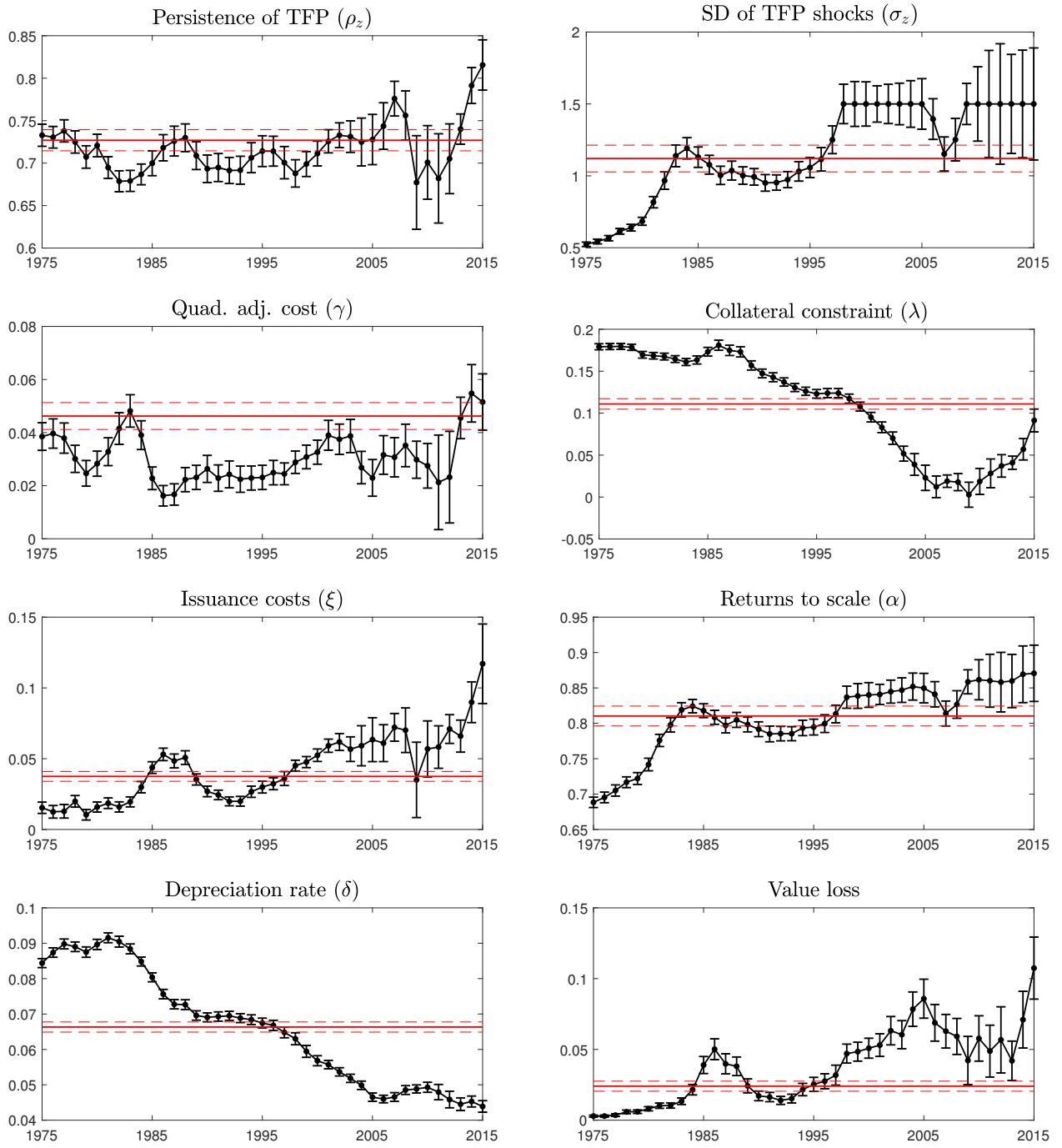
believe that firms are more financially constrained as collateral constraints become tighter (i.e.  $\lambda$  is smaller). This interpretation points to the difficulty of using leverage ratios to identify the severity of credit frictions, a point that echoes the analysis of Catherine et al. (2022b).

Similarly, we observe a large decline in the estimated depreciation rate over the sample period, which goes from .08 at the beginning of the sample period to .04 toward the end. While the actual depreciation of physical capital is unlikely to have changed over time, the rise in intangible capital may explain this apparently surprising result (Crouzet and Eberly, 2018). Since the model does not feature intangible capital, it interprets the reduction in physical capital expenditures as a decline in depreciation rate.

Another interesting exercise splits the sample across industries. Table 3 re-estimates the model by targeting moments calculated separately for four broad industries (manufacturing, retail trade, services, and transportation). We find that, relative to manufacturing, the value loss from financial frictions are much higher in the transportation. This finding reflects the much higher equity issuance costs in the transportation industry relative to manufacturing.

Figure 5 and Table 3 illustrate the power of our methodology to build transparent tools to evaluate the validity of structural estimates. By allowing to re-estimate the model on a large set of moments, we can easily test the mechanisms that underlie the structural model.

**Figure 5: Time Series Estimates (Corporate Finance Model)**



*Notes.* This figure shows the sensitivity of parameter estimates to the sample period used to compute the targeted moment. For each year  $t$  on the x-axis, we compute the seven baseline moments  $(m_i)_{i \in \{1..7\}}$  on the sample period  $[t - 5, t + 4]$ . We then re-estimate the model using the benchmark approximate SMM that targets the moments measured on this sub-period. The y-axis reports the resulting parameter estimates and their 95% confidence interval. The horizontal solid and dashed red lines corresponds to the baseline estimates obtained when computing moments on the entire sample, together with their 95% confidence interval.

Table 3: Subsample Estimates (Corporate Finance Model)

	$\rho_z$	$\sigma_z$	$\gamma$	$\lambda$	$\xi$	$\alpha$	$\delta$	value loss
benchmark (all)	.7269	1.1204	.0462	.1110	.0375	.8103	.0663	.0240
- se	.0062	.0465	.0025	.0031	.0017	.0070	.0007	.0018
manufacturing	.7055	.8488	.0367	.0735	.0172	.7534	.0545	.0095
- se	.0093	.0379	.0028	.0039	.0021	.0091	.0006	.0013
retail trade	.8626	.7911	.0821	.1837	.1577	.8294	.0941	.0439
- se	.0160	.0749	.0108	.0127	.0543	.0145	.0022	.0077
services	.8075	.8487	.0301	.0579	.0646	.7488	.0575	.0444
- se	.0145	.0626	.0061	.0090	.0059	.0172	.0019	.0068
transportation	.7397	1.5000	.0375	.3012	.0763	.8920	.0946	.0703
- se	.0547	.4611	.0069	.0203	.0221	.0306	.0043	.0145
lower bound	.5	.2	0	0	0	.5	0	
upper bound	.98	1.5	.3	.6	.3	.9	.2	

Notes. This table reports parameter estimates of the corporate finance models using the benchmark approximate SMM (third-order polynomial with k=2) that targets moments calculated separately for four broad industries: manufacturing, retail trade, services and transportation. Benchmark corresponds to the baseline approximate estimation in Table 2. Standard errors are calculated using the approximate Jacobian matrix and the efficient weight matrix. The standard error of the value loss statistics is calculated using the delta method.

### 3.8 Model Misspecification

Our approach can also be used to explore model misspecification. One approach to misspecification explored in Andrews et al. (2017b) considers only local sources of misspecification, which would affect moments only by a small amount. Our approach explores instead “larger” sources of misspecification, although at the cost of “specifying the misspecification” (Catherine et al., 2022b). The exercise we propose proceeds in two steps: (1) we solve (a potentially large number of) alternative models and simulate corresponding moments (2) we estimate our (misspecified) model with the approximate SMM using the moments generated in (1) as targeted moments. The model is robust to misspecification if the resulting parameter estimates are close to the baseline estimates.

We illustrate this exercise in the context of the recent corporate finance literature arguing that financial constraints often take the form of cash-flow constraints, rather than collateral constraints (e.g. Lian and Ma (2019), Greenwald (2019)). Our objective is to assess the robustness of the baseline estimates in Section 3.4 to this particular

source of misspecification.

In a first step, we simply augment the baseline model of Section 3.1 to account for the fact that the debt constraint may not only depend on the capital stock, but also on the firm's EBITDA:

$$d_t < \lambda k_t + \lambda_2 \cdot E[e^{z_t(1-\alpha)}] k_t^\alpha \quad (9)$$

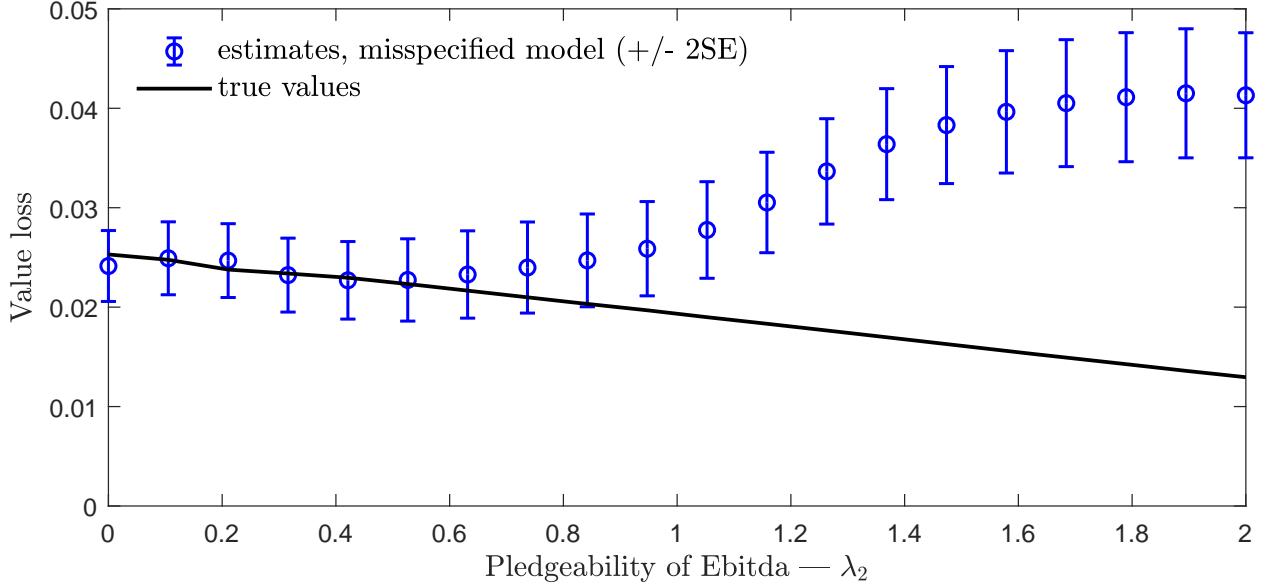
so that the firm can not only pledge collateral but also a multiple  $\lambda_2$  of its EBITDA. For the initial seven parameters, we use the baseline estimates of Table 2. We consider 40 linearly-spaced values for  $\lambda_2$ , ranging from 0 (no misspecification) to 2 (large misspecification). For each of these values of  $\lambda_2$ , we solve the model and simulate the seven baseline moments ( $\{m_1, \dots, m_7\}$ ).

In a second step, we use the approximate SMM to estimate the model targeting each of the 40 different vectors of moments. Figure 6 shows the effect of misspecification on our estimates of the value loss from financing constraints. The x-axis represents the value of  $\lambda_2$  used to simulate the moments. The black line corresponds to the actual value loss of financing constraints in the correctly-specified model with both collateral and cash-flow constraints. Intuitively, as  $\lambda_2$  increases, the value loss from financial constraints decrease since the firm is less and less constrained. The loss is about 2.5% if  $\lambda_2 = 0$  and goes down to about 1.2% for  $\lambda_2 = 2$ . The blue line shows the estimate of value loss when the misspecified model is used to estimate the model.

Two main results emerge from Figure 6. First, the estimates of value loss is robust to low values of cash-flow pledgeability ( $\lambda_2 < .6$ ). Second, as  $\lambda_2$  increases further, the misspecified model finds larger value loss while the true value loss decreases. One key mechanism explains this surprising result. In the cash-flow based model, an increase in  $\lambda_2$  affects productive firms disproportionately, an effect that the misspecified model misses. This error affects two different moments.

First, as  $\lambda_2$  rises, the autocorrelation of investment decreases. Looser cash-flow constraints make firms more responsive to productivity shocks – so that they smooth their investment policy less. The misspecified model, in which this effect is less pronounced, attributes the reduction in the autocorrelation of investment to lower capital adjustment costs. With lower capital adjustment costs, the value loss from financial constraints are amplified. This is because financial frictions are less likely to be binding in the presence of real frictions. Intuitively, if the firm needs to smooth investment because of capital adjustment costs, it's more likely to be able to finance its growth with its own cash flows. Thus, the misspecified model expects larger value losses when the data is generated with larger  $\lambda_2$ .

Figure 6: Model Mispecification: Estimating Model with Collateral Constraints on Data Generated by Cash-flow Constraints (Corporate Finance Model)



*Notes.* This figure explores the sensitivity of estimation to model misspecification. We consider an augmented version of our corporate finance model where firms can pledge a multiple  $\lambda_2$  of their EBITDA so that their debt constraint takes the form  $d_t < \lambda k_t + \lambda_2 \cdot \mathbb{E}[e^{z_t(1-\alpha)}] k_t^\alpha$ . We assume that the correctly specified model is the augmented model where the baseline parameters are set to their estimated value in Table 2 and  $\lambda_2$  takes various values from 0 (the baseline model) to 2. For each possible value of  $\lambda_2$ , we re-estimate the baseline parameter values using the mis-specified model that targets these moments (simulated with the correctly-specified model). The x-axis plots the values of  $\lambda_2$  used in each estimation. The blue circles report the value loss from financial constraint estimated with the approximation to the misspecified model while we target moments generated by the correctly-specified model. The black line reports the value loss from financial constraint in the correctly-specified model.

Second, as  $\lambda_2$  increases, equity issuance decreases. This is because firms that want to invest – i.e. the most productive firms – need less equity issuance as they can pledge their EBITDA. The collateral-only model, which does not have this feature, attributes this lower equity issuance to larger equity costs. Larger equity issuance costs increase the value loss from financial constraints.

## 4 Dynamic Household Finance Model

### 4.1 Model

Our second model is similar to Catherine (2021) but abstracts from housing choices to focus on consumption choices and the allocation of wealth between a risk-free asset and the stock market portfolio. Relative to seminal papers by Viceira (2001) and

Cocco et al. (2005), our model incorporates countercyclical income risk documented in Guvenen et al. (2014)<sup>8</sup> and a realistic Social Security system in retirement.

**Macroeconomic environment** The stock market log return in year  $t$  is

$$s_t = s_{1,t} + s_{2,t}, \quad (10)$$

where  $s_1$  denotes the component of returns that covaries with labor market conditions and follows a normal mixture distribution:

$$s_{1,t} = \begin{cases} s_{1,t}^- \sim \mathcal{N}(\mu_s^-, \sigma_{s_1}^2) & \text{with probability } p_s \\ s_{1,t}^+ \sim \mathcal{N}(\mu_s^+, \sigma_{s_1}^2) & \text{with probability } 1 - p_s \end{cases} \quad (11)$$

On the other hand,  $s_{2,t}$  is normally distributed with variance  $\sigma_{s_2}^2$ . We impose  $\mu_s^- < \mu_s^+$  and interpret  $\mu_s^-$  as the expected log return in stock market crash years, and  $p_s$  their frequency. The growth of the log national wage index  $l_1$  is:

$$l_{1,t} - l_{1,t-1} = \mu_l + \lambda_{ls}s_{1,t} + \varepsilon_{l,t}, \quad (12)$$

where  $\varepsilon_{l,t}$  follows  $\mathcal{N}(0, \sigma_l^2)$ ,  $\mu_l$  is the average growth rate, and  $\lambda_{ls}$  captures the correlation with stock returns.

**Income risk** Labor earnings can be decomposed as the product of the wage index and an idiosyncratic component  $L_{2,it}$ :

$$L_{it} = L_{1,t} \cdot L_{2,it}. \quad (13)$$

The idiosyncratic component is further decomposed into a deterministic function of age  $f_{it}$ <sup>9</sup>, a persistent component  $z_{it}$  and a transitory shock  $\eta_{it}$ :

$$L_{2,it} = e^{f_{it} + z_{it} + \eta_{it}}. \quad (14)$$

The persistent component follows an AR(1) process, with innovations drawn from a normal mixture. Specifically, the dynamics of  $z_i$  are given by

$$z_{it} = \rho_z z_{it-1} + \zeta_{it}, \quad (15)$$

---

<sup>8</sup>See Catherine et al. (2022a) for reduced-form evidence of the effect of countercyclical risk on households' portfolio.

<sup>9</sup>Specifically, we assume  $f$  to be a cubic polynomial function of age  $\theta_2 \text{age}^3/100 + \theta_1 \text{age}^2/10 + \theta_0$ .

where

$$\zeta_{it} = \begin{cases} \zeta_{it}^- \sim \mathcal{N}\left(\mu_{z,t}^-, \sigma_z^{-2}\right) & \text{with probability } p_z \\ \zeta_{it}^+ \sim \mathcal{N}\left(\mu_{z,t}^+, \sigma_z^{+2}\right) & \text{with probability } 1 - p_z \end{cases} \quad (16)$$

The values of  $p_z$ ,  $\mu_{z,t}^-$  and  $\mu_{z,t}^+$  control the degree of asymmetry in the distribution of income shocks. To capture the cyclicalities of skewness,  $\mu_{z,t}^-$  is an affine function of the log growth rate of the wage index:

$$\mu_{z,t}^- = \overline{\mu_z^-} + \lambda_{zl}(l_{1,t} - l_{1,t-1}). \quad (17)$$

where  $p_z\mu_{z,t}^- + (1 - p_z)\mu_{z,t}^+ = 0$  and  $p_z \leq 0.5$ . If  $\sigma_z^- \gg \sigma_z^+$ ,  $p_z$  represents the frequency of significant events in a worker's career. Finally, the transitory component of income is also modeled as a mixture of normals whose first and second components always coincide with the first and second components of the normal mixture governing the innovations to  $z_i$ .

$$\eta_{it} = \begin{cases} \eta_{it}^- \sim \mathcal{N}(0, \sigma_\eta^{-2}) & \text{if } \zeta_{it} = \zeta_{it}^- \\ \eta_{it}^+ \sim \mathcal{N}(0, \sigma_\eta^{+2}) & \text{if } \zeta_{it} = \zeta_{it}^+ \end{cases} \quad (18)$$

**Social Security** Social Security payroll taxes represent 12.4% of the agent's earnings below the maximum taxable earnings, which represents 2.5 times the national wage index.

$$T_{it} = .124 \cdot \min\{L_{it}, 2.5 \cdot L_{1,t}\}. \quad (19)$$

Retirement benefits depend on historical taxable earnings, adjusted for the growth in the national wage index. Specifically, the agent's Social Security benefits  $B$  are:

$$\frac{B_i}{L_{1,R}} = \begin{cases} .9 \cdot S_{iR} & \text{if } S_{iR} < .2 \\ .116 + .32 \cdot S_{iR} & \text{if } .2 \leq S_{iR} < 1 \\ .286 + .15 \cdot S_{iR} & \text{if } 1 \leq S_{iR}, \end{cases} \quad (20)$$

where  $R$  is the retirement age and  $L_{1,R}$  is the value of the wage index at that age. The variable  $S_{it}$  keeps track of a worker's average taxable idiosyncratic earnings:

$$S_{it} = \sum_{k=t_0}^t \frac{\min\{L_{2,ik}, 2.5\}}{t - t_0 + 1}, \quad (21)$$

where  $t_0$  denotes his first year of earnings.

**Agent** The agent maximizes lifetime expected utility:

$$V_{t_0} = E \sum_{t=t_0}^T \beta^{t-1} \left( \prod_{k=0}^{t-1} (1 - m_k) \right) \frac{C_{it}^{1-\gamma}}{1-\gamma}, \quad (22)$$

where  $\gamma$  is the coefficient of relative risk-aversion,  $m_k$  the mortality rate at age  $k$ ,  $\beta$  the subjective discount factor and  $T$  the maximum lifespan. After receiving his income, the agent determines his consumption and invests his remaining wealth in stocks or bonds. Wealth evolves as:

$$W_{it+1} = [W_{it} + L_{it} + B_{it} - T_{it} - C_{it} - c_{it}] \cdot [\pi_{it} e^{s_t} + (1 - \pi_{it}) e^r], \quad (23)$$

where  $\pi_{it}$  is the share of his wealth invested in equity. Owning stocks incurs a cost  $c_{it} = \Phi L_{1,t}$  if  $\pi_{it} > 0$ . Short selling or leveraging are not allowed, such that  $0 \leq \pi_{it} \leq 1$ .

**Preset parameters** We calibrate the labor income process and the distribution of stock market returns using estimates from [Catherine \(2021\)](#) reported in Appendix Table A.2.

## 4.2 Training Dataset and Validation

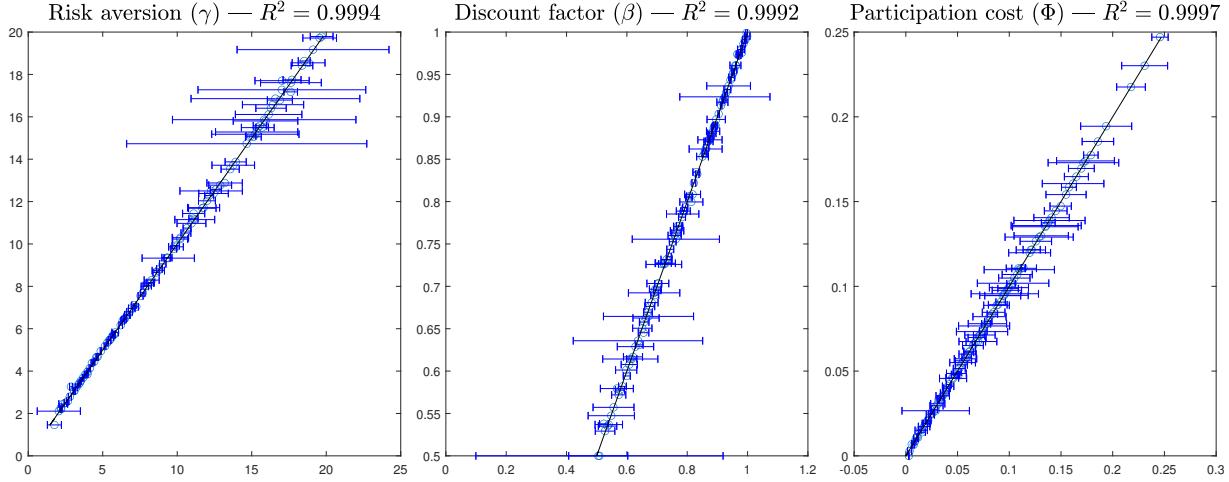
We estimate three structural parameters: risk-aversion, time preference and participation cost ( $\theta = (\gamma, \beta, \Phi)$ ). We restrict these parameters to values  $\gamma \in [1; 20]$ ,  $\beta \in [.5; 1]$ ,  $\Phi \in [0; .25]$ . We then draw a Halton sequence of  $N = 2,000$  vectors  $\theta_i$ . For each  $\theta_i$ , we simulate the model and compute three moments:  $m_1$  = average wealth, normalized by the wage index ( $E[W/L_1]$ ).  $m_2$  = stock market participation rate ( $E[\pi > 0]$ ).  $m_3$  = average equity share among stock market participants ( $E[\pi | \pi > 0]$ ).

These 2,000 parameter draws result in a training dataset  $\mathcal{D} = (\theta_i, m_i)_{i \in [1, 2000]}$ , which we use to estimate  $g()$ , the approximate relationship between moments and parameters. Like in Section 3.3, we also build a “validation” dataset of 200 additional draws of parameters and corresponding moments, that we use to assess the approximate SMM performance.

Like in the corporate finance application, the approximate moment relation  $g()$  is best captured with a third-order polynomial estimated using quadratically declining weights with  $k = 2$ . Figure A.4 shows the performance of various functional forms and weighting schemes. Compared to the corporate finance model, the precision of the approximate method is higher (share of the unexplained variance of true parameters

is lower), and neural networks perform better (on an unweighted basis).

Figure 7: Performance of Estimation using Benchmark Approximation (Household Finance Model)



*Notes.* This figure shows, for the household finance model, the precision of the benchmark approximate SMM across estimated parameters. For each draw  $(\theta_j^{\text{validation}}, f(\theta_j^{\text{validation}}))$  in the validation sample, we estimate parameters  $\widehat{\theta}_j^{\text{validation}}$  with an approximate SMM targeting moments  $f(\theta_j^{\text{validation}})$ . The x-axis reports the true parameters  $\theta_j^{\text{validation}}$ , while the y-axis reports the estimated parameters  $\widehat{\theta}_j^{\text{validation}}$ . The approximation  $g(\theta; \beta)$  we use in this plot is estimated on the training dataset  $\mathcal{D}$  as  $\widehat{\beta} = \arg \min_{\beta} \left[ \sum_{l \in \mathcal{D}} \frac{1}{(\Delta_l^j)^k} (g(\theta_l; \beta) - f(\theta_l))' (g(\theta_l; \beta) - f(\theta_l)) \right]$  where  $\Delta_l^j = (\mathbf{m}_l - \mathbf{m}_j^{\text{validation}})' \Omega (\mathbf{m}_l - \mathbf{m}_j^{\text{validation}})$ ,  $\Omega$  is the inverse of the variance-covariance matrix of the empirical moments,  $g()$  is a third-order polynomial approximation and  $k = 2$ .

Figure 7 is the analog of Figure 1. For each draw of the validation sample, it shows the true parameters  $\theta^{\text{validation}}$  on the x-axis, and, on the y-axis, the estimated parameters  $\widehat{\theta}^{\text{validation}}$ . The out-of-sample  $R^2$  is larger than 99.9% for all three parameters  $\gamma$ ,  $\beta$  and  $\Phi$ . Like in the corporate finance application, the few cases where the approximate SMM fails to recover the true parameters corresponds to cases where the true model is poorly identified, which can be seen by the large standard errors estimated with the true model for these parameter draws.<sup>10</sup>

Finally, Appendix Figure A.5 confirms that, as soon as the size of the training sample is beyond 500,  $1 - R^2$  is below 1%. So compared to the corporate finance model (which has more free parameters), our approximation is very good with a much smaller training sample.

<sup>10</sup>Like for the corporate finance application, standard errors are computed using the Delta method with the true model  $f(\theta)$  and the variance-covariance matrix of the baseline moments used in the main estimation. The choice of this matrix is by nature arbitrary, since the moments that we target in these estimations are not empirical objects – they are simulation-based.

### 4.3 Data and Approximate SMM

We compute the three core data moments ( $m_1, m_2, m_3$ ) using the 1989–2016 waves of the triennial Survey of Consumer Finances. We describe the construction of these moments in Appendix A.2. We then use these moments to estimate parameters using (1) a standard SMM and (2) our approximate moment function fitted on the training sample. First, as in the corporate finance model, note that our method is several order of magnitude faster than a standard SMM. Appendix Figure A.6 shows the speed of convergence for each one of the 3 parameters. The approximate SMM converges in about .2 seconds, while the true SMM takes between 15 minute to an hour to converge to its final value.

Panel A of Table 4 reports the parameter estimates using the two methods. Both approaches give similar values for risk-aversion  $\gamma$  (about 8.4) and the discount factor  $\beta$  (some .91). The estimate of  $\Phi$ , the participation cost, is however a bit lower in the approximate SMM, by about 15%: while  $\Phi$  is estimated at .0053 in the true SMM, the approximate SMM provides an estimate of .0045. Economically, this difference represents roughly \$50 a year. As in the corporate finance application, we also compute two additional, non-targeted moments:  $m_4$ , which is the median wealth and  $m_5$ , which is the median conditional equity share. Table 4, Panel B shows that these moments are matched, whether we use the true or the approximate SMM.

### 4.4 Identification Diagnostic

Figure 8 is the analog of Figure 2. It traces out the mapping from targeted moments to estimated parameters (blue line), which is only possible thanks to the low estimation cost of the approximate SMM. As with our corporate finance application, two salient features emerge from the figure. First, intuitions from local comparative statics (in yellow on the figure) can be misleading regarding the true relation from moments to estimated parameters. Second, the linear approximation used in Andrews et al. (2017b) can be rapidly imprecise for some parameters.

Panel A shows how estimated parameters vary with mean wealth. Targeting a higher level of wealth results in a greater estimated discount factor  $\beta$ . At the same time, locally, an increase in the discount factor leads to higher wealth. Thus, for this parameter-moment pair, local comparative statics provides the correct intuition for identification. This is not the case for the other two parameters. For instance, local comparative statics tell us that an increase in the participation cost leads to a small reduction in wealth, as households invest less in the risky asset (yellow line). In con-

Table 4: Moment and Parameter Estimates: true vs. approximate SMM (Household Finance Model)

(a) Parameter Estimates

	$\gamma$	$\beta$	$\Phi$
true SMM	8.328	.9106	.0053
- s.e., local deriv.	.064	.0021	.0002
approximate SMM	8.455	.9060	.0045
- s.e., local fit deriv.	.060	.0020	.0002
estimation, lower bound	1.01	.5	0
estimation, upper bound	20	1	.25

(b) Estimated Moments

	Data	True SMM	Approximate	Simulation
<b>mean(wealth)</b>	5.63 (.03)	5.63	5.63	5.60
<b>participation rate</b>	.557 (.002)	.557	.557	.568
<b>mean(cond. equity share)</b>	.345 (.003)	.345	.345	.337
median(wealth)	2.00 (.01)	2.92	2.94	2.9
median(cond. equity share)	.224 (.002)	.269	.261	.259

*Notes.* This table reports the parameter estimates and simulated moments of the household finance model presented in Section 4.1. Panel 4a reports parameter estimates using the true SMM (first line) and the benchmark approximate SMM (second line). The benchmark approximation is a third-order polynomial with  $k = 2$ .  $\gamma$  is risk-aversion.  $\beta$  is time discount factor.  $\Phi$  is the participation cost. Panel 4b reports the moments targeted in estimation in bold fonts and untargeted moments representing median of statistics in regular fonts. Column “Data” shows the empirical moments, with standard errors in parenthesis. Column “True SMM” corresponds to the simulated moments for the parameters using the true SMM estimation. Column “Approximate” show the approximate moments at the approximate parameter estimates ( $g(\hat{\theta}^{\text{approx}}, \hat{\beta})$ ). Column “Simulation” reports the true simulated moments at the approximate parameter estimates ( $f(\theta^{\text{approx}})$ ). The moments used in the estimation are defined in Section 4.2.

trast, we see that lower mean wealth leads to a lower estimation of the participation cost (blue line). As mean wealth decreases, the model predicts lower participation; to keep matching the participation rate, the cost of participation has to decrease. Thus, local comparative statics do not provide a valid intuition for the role of mean wealth in identifying the participation cost. The same is true for risk-aversion. Because of precautionary savings, wealth is a steeply increasing function of risk-aversion (yellow line). However, in estimation, an increase in mean wealth leads to a small reduction in *estimated* risk-aversion (blue line). Locally, when wealth increases, the model predicts a slightly lower equity share; to keep matching the equity share, risk-aversion has to decrease. Here again, local comparative statics do not provide useful intuitions for identification. We reach a similar conclusion in Panel B, which looks at how estimated parameters depend on the participation rate, and Panel C (conditional equity share).

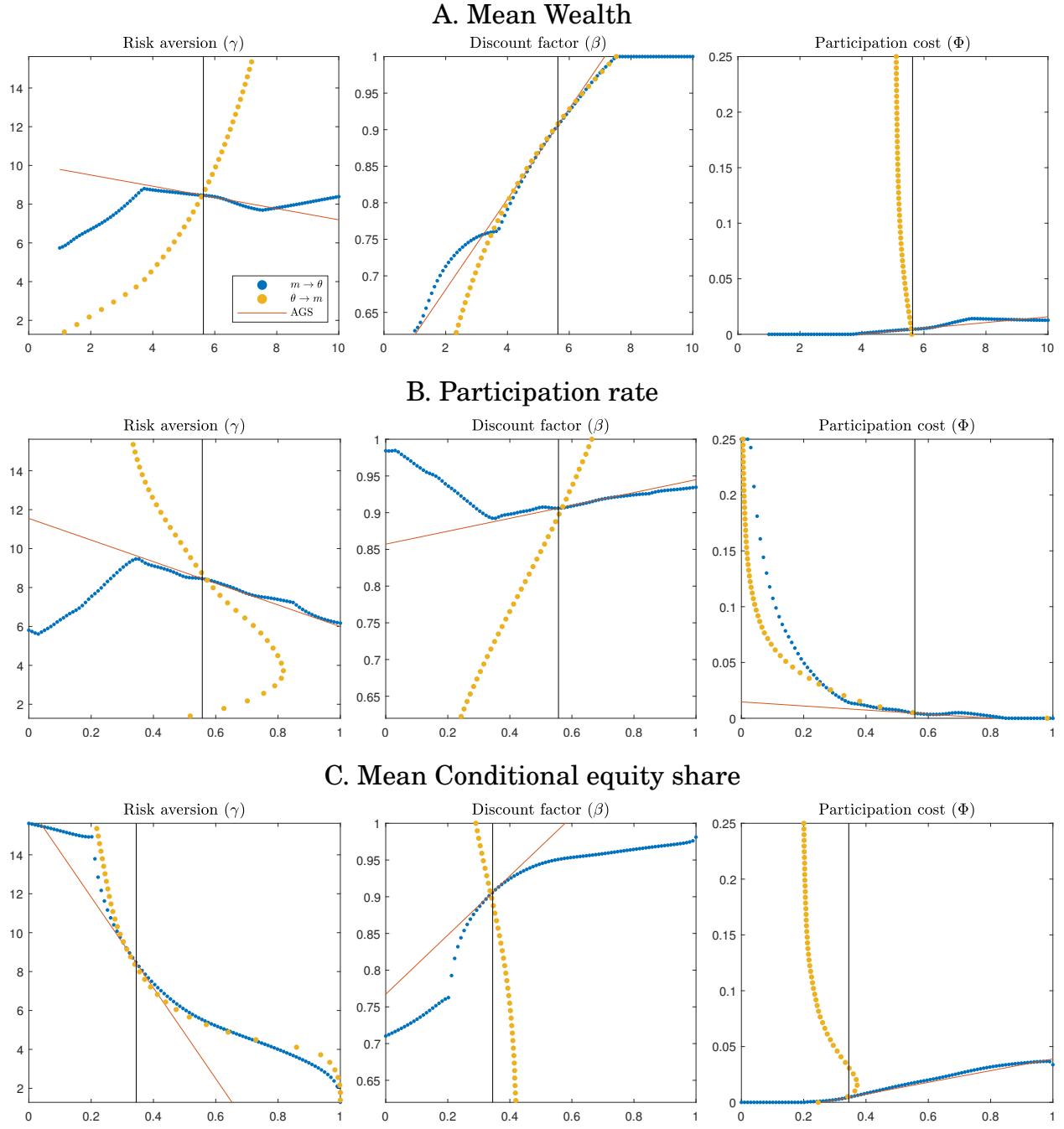
Figure 8 also makes clear that linear approximation in Andrews et al. (2017b) is sometime imprecise, even for moment values close to their empirical counterparts. For instance, the conditional equity share in the data is about 35%. If it was 45% instead, the local linear approximation of Andrews et al. (2017b) would suggest an estimated risk-aversion of about 6, while its estimated value would in fact be closer to 7.

## 4.5 Robustness to Moment Selection

We use the approximate SMM to explore robustness to moment selection. We first consider the following set of moments in addition to the baseline moments: (1) median wealth ( $m_4$ ) (2) median conditional equity share ( $m_5$ ) (3) an alternative definition of the conditional equity share, where we normalize stock holdings by financial wealth instead of net worth ( $m_6$ ) (4) the median of this alternative conditional equity share ( $m_7$ ).  $m_4$  and  $m_5$  provide additional information when stock holdings and total wealth have a fat upper-tail, as is the case in the model.  $m_6$  and  $m_7$  recognize the fact that, absent housing, net worth and financial wealth are the same in the model, but not in the data. There is no consensus in the literature on the proper denominator, but as Panel C of Figure 8 illustrates, this choice is consequential as a modest change in the target conditional equity share has substantial effect on parameter estimates.

We also consider the robustness of estimation to targeting *groups* of moments that describe the age profile of the baseline moments ( $m_1, m_2, m_3$ ): (i) the life-cycle profile of mean wealth ( $m_8$ ) (ii) the life-cycle profile of participation rates ( $m_9$ ) (iii) the life-

Figure 8: Sensitivity of Parameters to Moments (Household Finance Model)



*Notes.* This figure plots parameter values on the y-axis and the value of the moment on the x-axis. Panel A (resp. B and C) corresponds to  $m_1$  (mean wealth) (resp.  $m_2$  (participation rate) and  $m_3$  (conditional equity share)). The yellow line draws local comparative statics, i.e. how variations in one parameter around its estimated value – holding other parameters fixed at their estimated value – affect the value of  $m_i$  obtained through simulations. The blue line plots how variations in the value of the empirical moment  $m_i$  – holding other moments fixed at their empirical value – affects the estimated parameter values. Each dot on the blue line corresponds to a separate estimation. Finally, the red line corresponds to the local linear approximation of the blue line around the parameter estimates, and represents the “sensitivity matrix” of Andrews et al. (2017b): it is a linear approximation of the mapping from moments to parameter estimates around the empirical value of the moments  $m_i$ .

cycle profile of equity shares ( $m_{10}$ ).<sup>11</sup> These groups of moments are computed using the procedure explained in Catherine (2021), which implements the Deaton-Paxson method to filter out cohort and year effects from the data. We use 20 age groups, going from 23-25 to 80-82 years old, by increments of 3 years, so  $m_8$ ,  $m_9$  and  $m_{10}$  each contains 20 moments.

We explore robustness of parameter estimates to targeting some of the additional moments in  $\{m_4, \dots, m_{10}\}$  in addition to the baseline moments  $\{m_1, m_2, m_3\}$ . Figure 9 reports our first robustness check, which is the equivalent of Figure 3 above: We target only one additional moment in  $\{m_4, m_5, m_8, m_9, m_{10}\}$  and report the resulting parameter estimates. Our estimates of risk-aversion and participation cost are robust to including  $m_4$  and  $m_5$  in the set of targeted moments. The estimate of the discount factor, however, becomes lower when targeting median wealth ( $m_4$ ), from about .9 to .86. Since median wealth is much lower in the data than mean wealth, the model requires more impatience to match this additional moment. Figure 9 makes clear that our baseline estimation is not robust to the inclusion of the life-cycle profile of equity shares in the set of targeted moments. Targeting this group of moments leads to a significantly higher participation cost and discount rate, and a significantly lower risk-aversion.

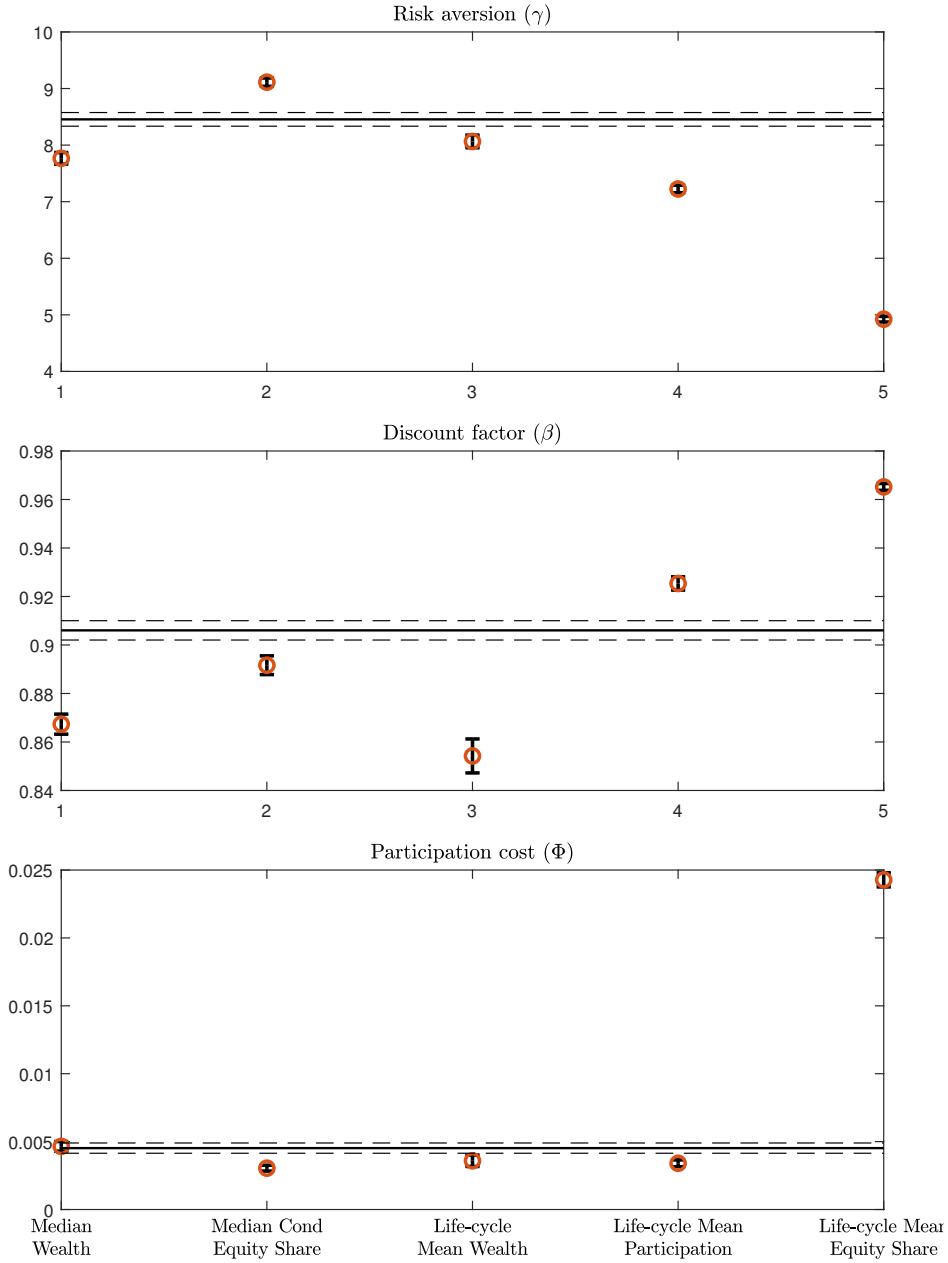
Some studies measure equity shares in the data differently, by dividing stock holdings by financial wealth rather than net worth, effectively replacing  $m_3$  by  $m_6$  (or  $m_5$  by  $m_7$ ). In the model, these moments are equivalent but not in the data. Hence, the effect of replacing  $m_3$  by  $m_6$  on estimated parameters can be analyzed using Panel C of Figure 8. Using financial wealth as the denominator of the conditional equity share would increase the target moment from .345 to .451. Intuitively, this leads to a lower estimated risk aversion. But the blue curves of Panel C show that a 10 pp increase in the targeted conditional equity share also results in substantially higher estimates for the discount factor and participation costs. Traditional comparative statics (in yellow) show that way the conditional share is measured in the data will affect the estimate of  $\gamma$  but incorrectly suggest that they will have little effect on the estimates of  $\beta$  and  $\Phi$ .

We also consider a more systematic exploration of robustness to moment selection that mirrors the exercise in Figure 4. We leverage the speed of the approximate SMM to re-estimate the model using 72 alternative sets of moments:

---

<sup>11</sup>Note that we compute the life-cycle of *unconditional* equity share, instead of conditional equity share. This is because the model tends to return no participation in the stock market for some age groups in the simulation, which makes the conditional mean equity share not well-defined.

Figure 9: Robustness to adding moments one by one (Household Finance Model)



*Notes.* This figure explores the sensitivity of parameter estimates to the inclusion of an additional moment in the set of targeted moments. Our baseline estimation targets three moments  $(m_i)_{i \in \{1..3\}}$ . We consider a set of five (group of) additional moments: (1) median wealth ( $m_4$ ), (2) median conditional equity share ( $m_5$ ), (3) life-cycle mean wealth ( $m_8$ ) but excluding overall mean wealth in this case, (4) life-cycle mean participation ( $m_9$ ) but excluding overall mean participation, and (5) life-cycle mean equity share ( $m_{10}$ ) but excluding the overall mean conditional share. Each point on the x-axis refers to one of the additional targeted moment. The y-axis reports the parameter estimated with the benchmark approximate SMM that targets the baseline moments  $\{m_1, m_2, m_3\}$  and one of the additional (group of) moments from the set  $\{m_4, m_5, m_8, m_9, m_{10}\}$ , along with their 95% confidence interval. The black line and dashed lines show the baseline parameter estimates and their 95% confidence interval. Standard errors are derived using the standard formula  $(J'\Omega J)^{-1}$ , where  $\Omega$  is the inverse of the variance-covariance matrix of empirical moments and  $J$  is the approximate Jacobian matrix computed at the parameter estimates.

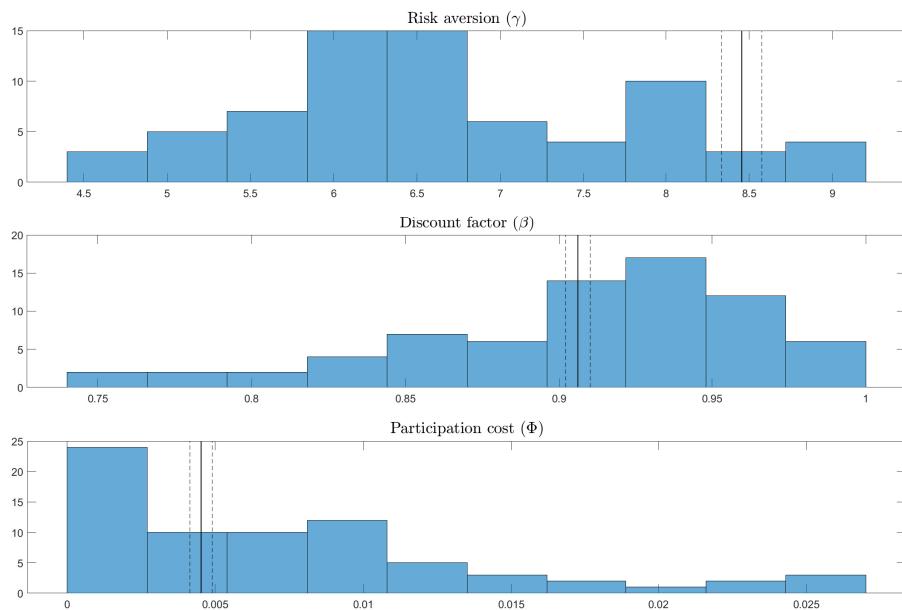
- We consider all possible combinations of three moments drawn from (a)  $m_1$  or  $m_8$ , i.e. moments related to mean wealth (overall or life-cycle)  $\times$  (b)  $m_2$  or  $m_9$ , i.e. moments related to mean participation (overall or life-cycle)  $\times$  (c)  $m_3$ ,  $m_6$  and  $m_{10}$ , i.e. moments related to the mean equity share (conditional, or conditional using the alternative empirical definition, or life cycle unconditional share). This leads to 12 different possible combinations.
- the 12 set of moments above, to which we add either median wealth ( $m_4$ ) or median conditional equity share ( $m_5$ ) or median conditional equity share using the alternative definition ( $m_7$ ). This adds 36 sets of moments.
- any of the 12 combinations above to which we add either  $m_4$  (median wealth) and  $m_5$  (median conditional equity share) or  $m_4$  (median wealth) and  $m_7$  (median conditional equity share using the alternative empirical definition). This adds another 24 sets of moments.

Figure 10 reports the distribution of parameter estimates across all 72 sets of targeted moments. A large share of moment combinations lead to consistent risk-aversion estimates: The distribution of risk-aversion estimates peaks at around 6-7. However, this mode of 6-7 is substantially below the baseline estimate for this parameter, which is about 8.5.<sup>12</sup> The estimates of  $\beta$ , the discount factor, are clustered between .9 and .98, and the mode of the distribution is close to the baseline estimate (about .92 vs. .91). The distribution of participation costs estimates peaks at 0. A large share of these estimates is thus significantly lower than the baseline estimate of about .005. These histograms illustrate how the approximate SMM can help build useful tools to analyze the robustness of structural analysis to moment selection. They also explain why, depending on moment selection, different papers in the literature can legitimately obtain significantly different parameter estimates.

---

<sup>12</sup>Note that this mode of 6-7 is consistent with Catherine (2021)'s estimates.

**Figure 10: Histogram of estimates across 72 sets of targeted moments (Household Finance Model)**



*Notes.* This figure explores the sensitivity of parameter estimates to moment selection. Our baseline estimation targets three moments  $(m_i)_{i \in \{1..3\}}$ . We consider 72 alternative set of moments described in Section 4.5. These sets are combinations of variants of the baseline moments (overall average, median, life-cycle, changing normalization). Note that none of these 72 cases are poorly identified by the notion defined in Figure 4, meaning the standard errors for all estimated parameters is no more than 10 times larger than the standard errors of the baseline estimate for all of these 72 cases. Each panel in the figure shows the distribution of parameters across these 72 estimations. The vertical black line and dashed lines report the baseline parameter estimates using the true SMM, together with their 95% confidence interval.

## 5 Conclusion

This paper provides a fast and simple way to run robustness checks and identification diagnostics in structural estimation. The approach consists of estimating a relation between model parameters and moments using a training dataset of parameters and associated moments. Once this relation has been estimated, it can be used to estimate structural parameters. In the two applications we consider – workhouse corporate finance and household finance models – we show that this “approximate SMM” carries computational costs that are several orders of magnitude lower than standard SMM estimations, with little to no precision loss.

Given the speed gains of our approximate estimation method, three exercises that seemed out of reach can easily be carried out to improve estimation transparency. The first one consists of estimating parameter robustness to moment selection. The second one is sample-splits: Estimation can be quickly carried out on various subsamples, to assess model robustness and validity. Finally, we use the approximate SMM to gauge the sensitivity of the baseline estimates to misspecification bias. We simulate many alternative models, and show how baseline model estimation is affected.

## References

- Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro**, “Measuring the Sensitivity of Parameter Estimates to Estimation Moments,” *Quarterly Journal of Economics*, 2017, 132 (4), 1553–1592.
- , —, and —, “Measuring the Sensitivity of Parameter Estimates to Estimation Moments,” *Quarterly Journal of Economics*, 2017, 132 (4).
- , —, and —, “On the Informativeness of Descriptive Statistics for Structural Estimates,” *Econometrica (Matthew Gentzkow’s Fisher-Schultz Lecture)*, 2020, 88 (6), 2231–2258. Working Paper.
- , —, and —, “Transparency in Structural Research,” *Journal of Business and Economic Statistics (invited discussion paper)*, 2020, 38 (4), 711–722.
- Armstrong, Timothy B. and Michal Kolesár**, “Sensitivity analysis using approximate moment condition models,” *Quantitative Economics*, January 2021, 12 (1), 77–108.
- Arnoud, Antoine, Fatih Guvenen, and Tatjana Kleineberg**, “Benchmarking Global Optimizers,” NBER Working Papers 26340, National Bureau of Economic Research, Inc October 2019.

- Azinovic, Marlon, Luca Gaegauf, and Simon Scheidegger**, “Deep equilibrium nets,” 2019.
- Bates, Thomas, Kathleen Kahle, and René Stulz**, “Why Do U.S. Firms Hold So Much More Cash than They Used To?,” *Journal of Finance*, 2009, 64, 1985–2021.
- Bloom, Nicholas**, “The Impact of Uncertainty Shocks,” *Econometrica*, 2009, 77 (3), 623–685.
- Bonhomme, Stéphane and Martin Weidner**, “Minimizing Sensitivity to Model Misspecification,” Papers 1807.02161, arXiv.org July 2018.
- Catherine, Sylvain**, “Countercyclical Labor Income Risk and Portfolio Choices over the Life Cycle,” *The Review of Financial Studies*, 12 2021. hhab136.
- , **Paolo Sodini, and Yapei Zhang**, “Countercyclical Income Risk and Portfolio Choices: Evidence from Sweden,” *Working Paper*, 2022.
- , **Thomas Chaney, Zongbo Huang, David Sraer, and David Thesmar**, “Quantifying Reduced-Form Evidence on Collateral Constraints,” *Journal of Finance*, 2022.
- Chen, Hui, Antoine Didisheim, and Simon Scheidegger**, “Deep Structural Estimation:With an Application to Option Pricing,” Cahiers de Recherches Economiques du Département d'économie 21.14, Université de Lausanne, Faculté des HEC, Département d'économie February 2021.
- Cocco, João F., Francisco J. Gomes, and Pascal J. Maenhout**, “Consumption and Portfolio Choice over the Life Cycle,” *The Review of Financial Studies*, 02 2005, 18 (2), 491–533.
- Crouzet, Nicolas and Janice Eberly**, “Intangibles, Investment, and Efficiency,” *American Economic Review: AEA Papers and Proceedings*, 2018, 108, 426–431.
- Duarte, Victor**, “Machine Learning for Continuous-Time Finance,” 2020.
- Fernández-Villaverde, Jesús, Galo Nuño, George Sorg-Langhans, and Maximilian Vogler**, “A Deep Learning Algorithm for High-Dimensional Dynamic Programming Problems,” 2021.
- Fernandez-Villaverde, Jesus, Mahdi Ebrahimi Kahou, Jesse Perla, and Arnav Sood**, “Exploiting Symmetry in High-Dimensional Dynamic Programming,” 2021.
- Gomes, Francisco, Michael Haliassos, and Tarun Ramadorai**, “Household Finance,” *Journal of Economic Literature*, September 2021, 59 (3), 919–1000.
- Greenwald, Dan**, “Firm Debt Covenants and the Macroeconomy: The Interest Coverage Channel,” Technical Report 2019.

- Guvenen, Fatih, Serdar Ozkan, and Jae Song**, “The Nature of Countercyclical Income Risk,” *Journal of Political Economy*, 2014, 122 (3), 621 – 660.
- Hall, Robert**, “On the Nature of Capital Adjustment Costs,” *Quarterly Journal of Economics*, 2004, 119 (3), 899–927.
- Hart, Oliver and John Moore**, “A Theory of Debt Based on the Inalienability of Human Capital,” *The Quarterly Journal of Economics*, 1994, 109 (4), pp. 841–879.
- Hennessy, Christopher A. and Toni M. Whited**, “Debt Dynamics,” *Journal of Finance*, 06 2005, 60 (3), 1129–1165.
- **and** — , “How Costly Is External Financing? Evidence from a Structural Estimation,” *The Journal of Finance*, 2007, 62 (4), 1705–1745.
  - **and** — , “How Costly is External Financing? Evidence From a Structural Estimation,” *Journal of Finance*, 2007.
- Huber, Peter J.**, *Robust Statistics*, Berlin, Heidelberg: Springer Berlin Heidelberg,
- Lian, Chen and Yueran Ma**, “Anatomy of Corporate Borrowing Constraints,” Technical Report 2019.
- Liu, Zheng, Pengfei Wang, and Tao Zha**, “Land-Price Dynamics and Macroeconomic Fluctuations,” *Econometrica*, 2013, 81 (3), 1147–1184.
- Maliar, Lilia, Serguei Maliar, and Pablo Winant**, “Will Artificial Intelligence Replace Computational Economists Any Time Soon?,” CEPR Discussion Papers 14024, C.E.P.R. Discussion Papers September 2019.
- Norets, Andriy**, “Estimation of Dynamic Discrete Choice Models Using Artificial Neural Network Approximations,” *Econometric Reviews*, 2012, 31 (1), 84–106.
- Ottonello, Pablo and Thomas Winberry**, “Financial Heterogeneity and the Investment Channel of Monetary Policy,” *Econometrica*, 2020, 88 (6), 2473–2502.
- Strebulaev, Ilya A. and Toni M. Whited**, “Dynamic Models and Structural Estimation in Corporate Finance,” *Foundations and Trends(R) in Finance*, November 2012, 6 (1–2), 1–163.
- Viceira, Luis M.**, “Optimal Portfolio Choice for Long-Horizon Investors with Non-tradable Labor Income,” *The Journal of Finance*, 2001, 56 (2), 433–470.
- Villa, Alessandro T. and Vytautas Valaitis**, “Machine Learning Projection Methods for Macro-Finance Models,” *International Political Economy: Investment & Finance eJournal*, 2019.

# APPENDIX – FOR ONLINE PUBLICATION

## A Additional Material

### A.1 Construction of Corporate Finance Moments

We start with a COMPUSTAT extract over 1970-2019. We only keep firms that appear at least twice in the sample. We drop firms in the financial (SIC code 6) or regulated (SIC code 49) sectors. We also drop observations with total assets that are less than 10 million real 1982 dollars, or sales or book assets that grow by more than 200%. This results in a sample of 117,976 firm-year observations and 11,198 unique firms. We compute moments targeted in the baseline estimation in the data as follows:  $m_1$ , the average investment to capital ration, is  $\frac{\text{capx}}{\text{l.at}}$ .  $m_2$ , the average profit to asset ratio, is  $\frac{\text{oibdp}}{\text{l.at}}$ .  $m_3$ , the average equity issuance to asset ratio, is computed net of repurchases:  $\frac{\text{sstk\_prstkc}}{\text{l.at}}$ .  $m_4$ , mean net leverage, is  $\frac{\text{dlc} + \text{dltt} - \text{che}}{\text{at}}$ .  $m_5$ , the autocorrelation of investment rates, is measured by regressing  $\frac{\text{capx}}{\text{l.at}}$  on its lag, with year fixed-effects. Last,  $m_6$  and  $m_7$ , are the sample standard deviations of 1-year and 5 year log sales growth:  $\log \text{sale} - \log \text{l.sale}$  and  $\log \text{sale} - \log \text{l5.sale}$ . All ratios are winsorized at the median +/- five times the interquartile range. We also remove firm fixed-effects from all the variables used in the empirical analysis, as the model does not feature any source of fully persistent heterogeneity across firms: for each variable, we subtract the within-firm average and add back the overall sample average. The bold lines in Table 1, in Column 1, provide the means and standard errors of these moments in our sample.

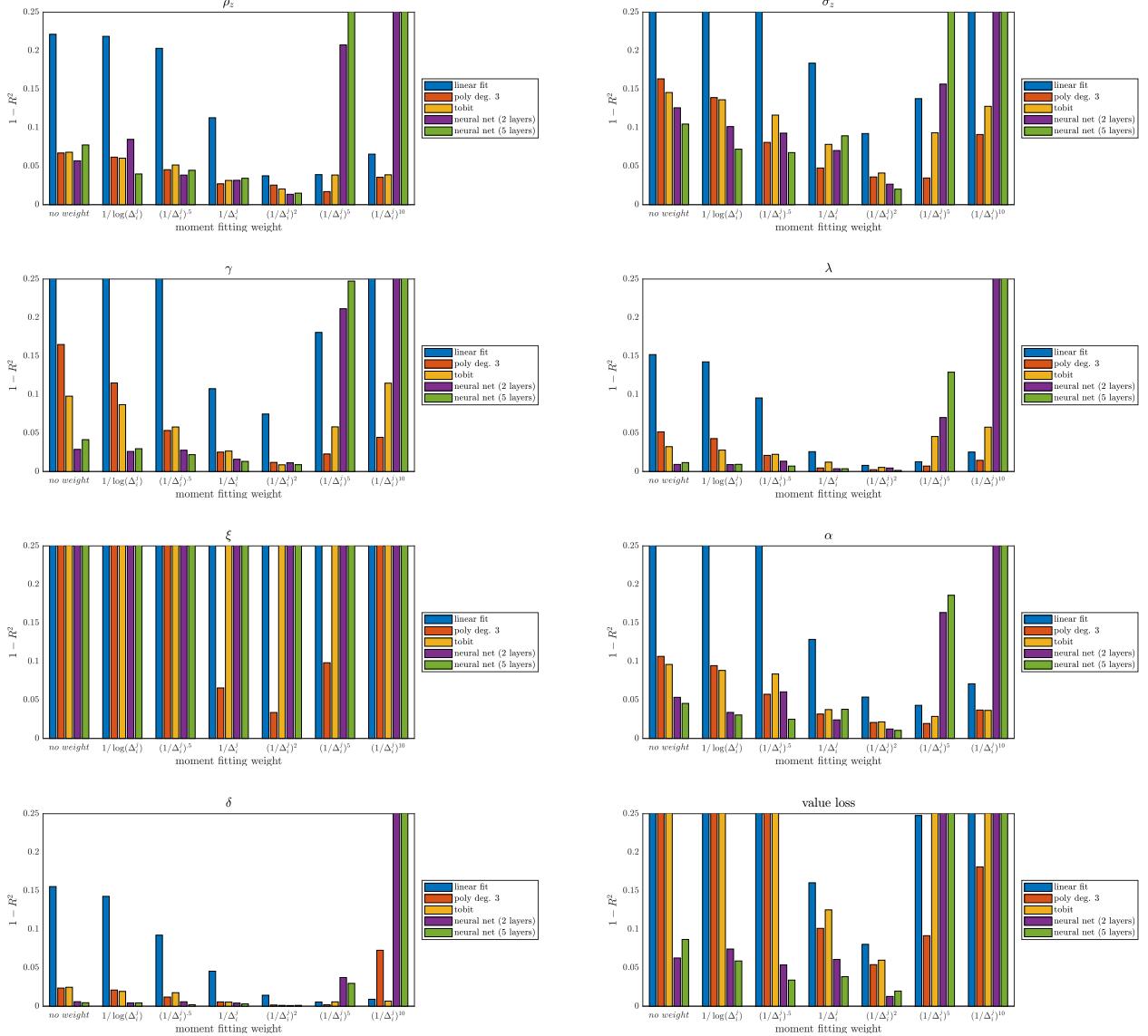
### A.2 Construction of Household Finance Moments

We compute the three core data moments using the 1989–2016 waves of the triennial Survey of Consumer Finances (SCF). We restrict the sample to households whose head is between age 22 and 99 and have positive net worth. The three “core” moments are estimated as:

- $m_1$ , the mean wealth, is measured using the net worth variable (*networth*) from the SCF summary extract public data; note that to improve comparability across survey years, we scale wealth by the average wage income (*wageinc*) of each survey year
- $m_2$  is the average participation rate, which is the share of households whose total holdings of stock (*equity*) is strictly positive
- $m_3$  the mean conditional equity share, which is the total holdings of stock (*equity*) divided by net worth, excluding vehicles (*vehic*), and is only computed for households with strictly positive holdings of stocks.

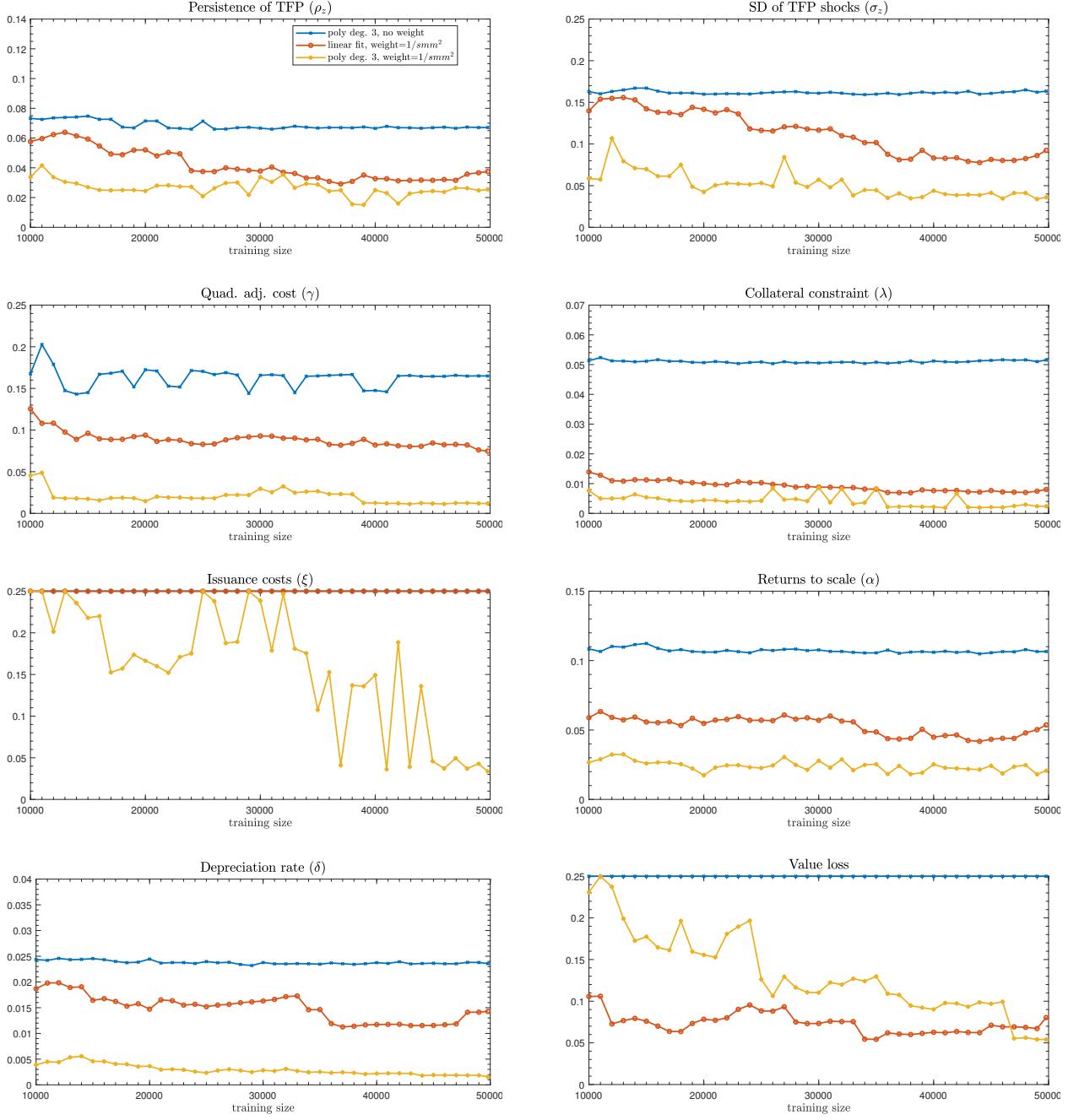
## B Appendix Figures

**Figure A.1: Out-of-sample precision of approximate estimation, for different approximations and weighting schemes (Corporate Finance Model)**



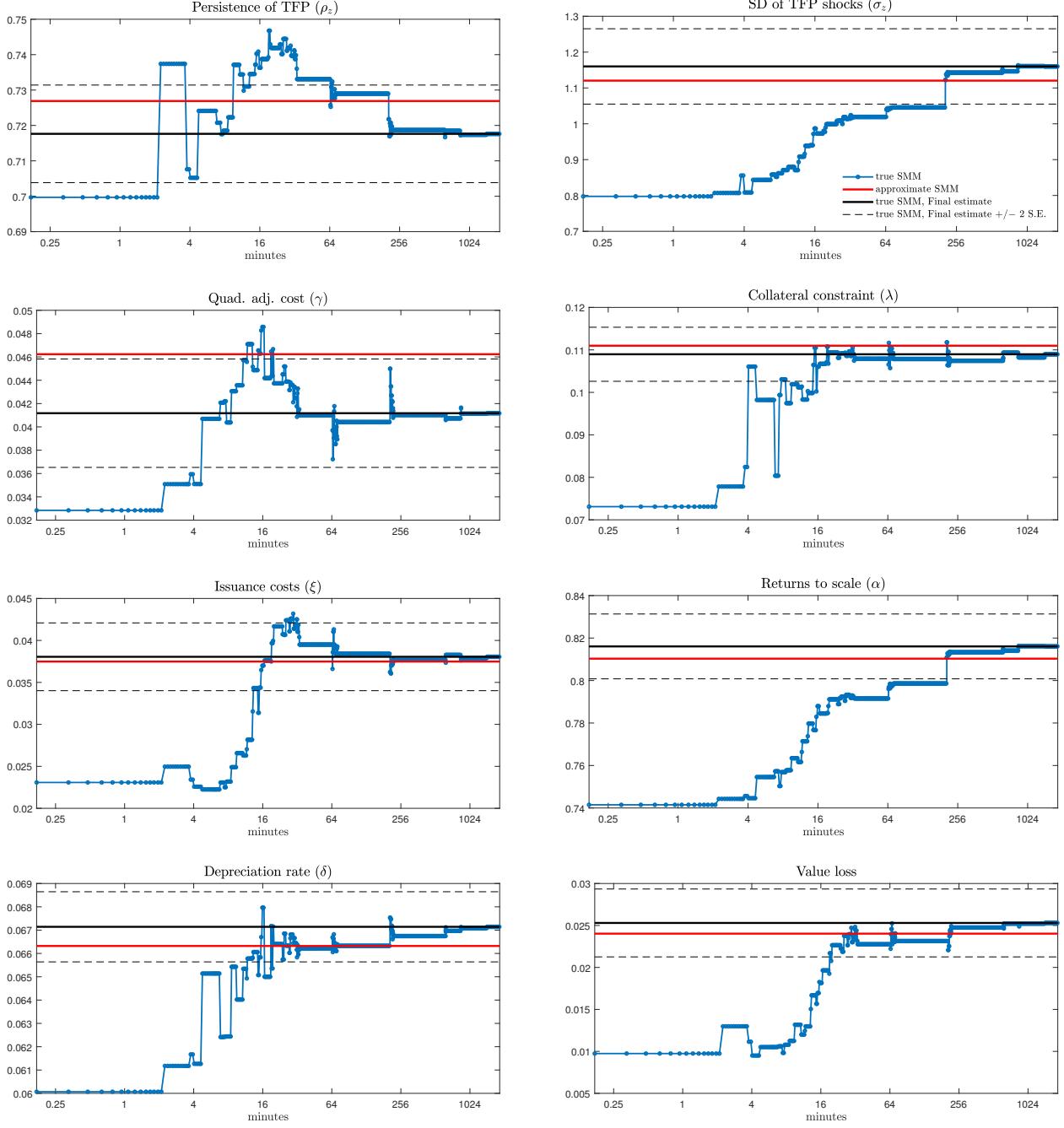
*Notes.* This figure reports a measure of the estimation error from the approximate SMM for different approximations and weighting schemes. For each draw  $(\theta_j^{\text{validation}}, f(\theta_j^{\text{validation}}))$  in the validation sample, we estimate parameters  $\widehat{\theta_j^{\text{validation}}}$  with an approximate SMM targeting moments  $f(\theta_j^{\text{validation}})$ . For each parameter  $i$  and approximation used, the figure reports  $1 - R^2(i) = \frac{\text{Var}[\theta_j^{\text{validation}}(i) - \widehat{\theta_j^{\text{validation}}}(i)]}{\text{Var}[\theta_j^{\text{validation}}(i)]}$ . We consider the following specification for the approximation  $g()$ : linear, third-order polynomial, tobit (third-order polynomial with equity issuance censored at zero), a neural net with 2 layers, and a neural net with 5 layers. The approximation  $g(\theta; \beta)$  is estimated on the training dataset  $\mathcal{D}$  as  $\widehat{\beta} = \arg \min_{\beta} \left[ \sum_{l \in \mathcal{D}} \frac{1}{(\Delta_l^j)^k} (g(\theta_l; \beta) - f(\theta_l))' (g(\theta_l; \beta) - f(\theta_l)) \right]$  where  $\Delta_l^j = (\mathbf{m}_l - \mathbf{m}_j^{\text{validation}})' \Omega (\mathbf{m}_l - \mathbf{m}_j^{\text{validation}})$  and  $\Omega$  is the inverse of the variance-covariance matrix of the empirical moments. We consider  $k = 0, .5, 1, 2, 5$  and  $10$ , as well as  $1/\log \Delta_l^j$  weights.

**Figure A.2: Out-of-sample Performance with respect to Training Sample Size (Corporate Finance Model)**



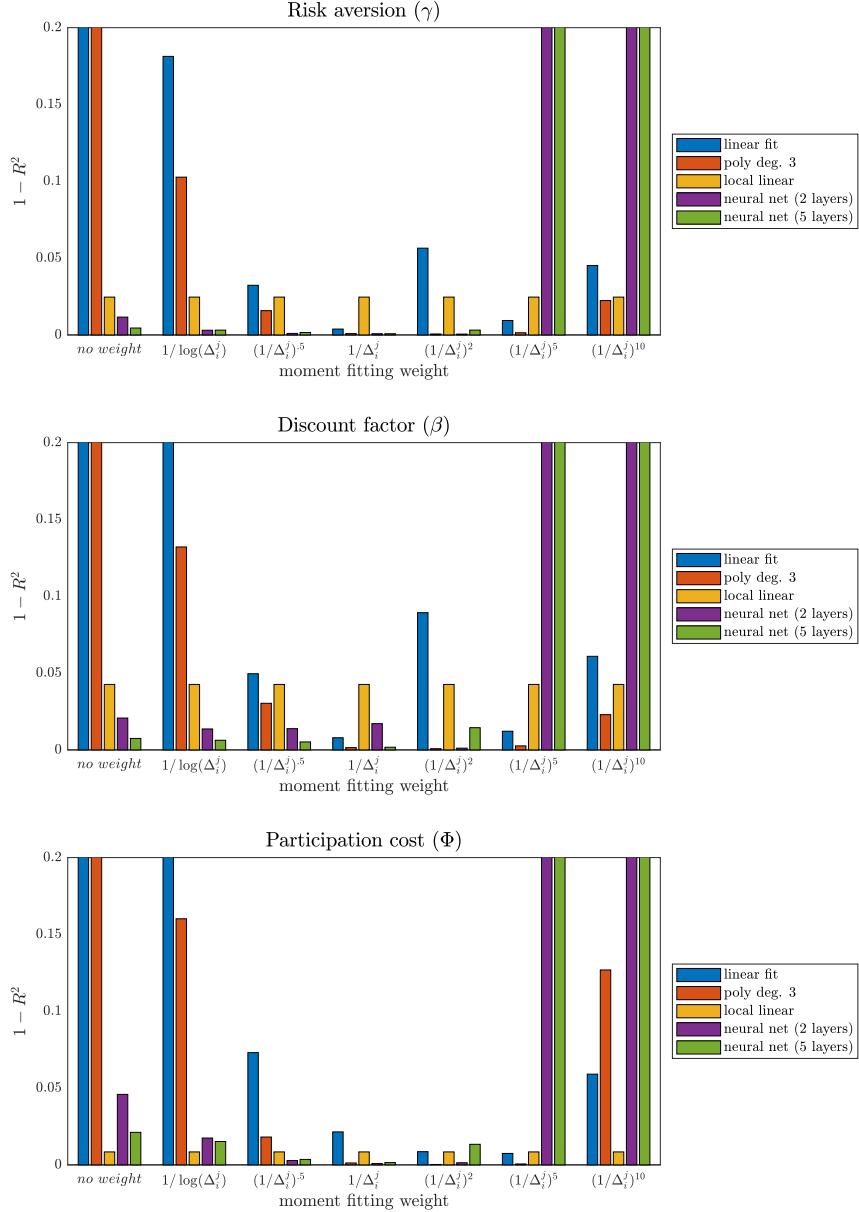
*Notes.* This figure explores the precision of the approximate SMM for the corporate finance model as a function of the size of the training sample used to fit the approximation. The y-axis reports a measure of the approximation error for estimated parameters,  $1 - R^2$ , described in fig. A.1. The x-axis corresponds to  $n$ , the size of the training dataset  $\mathcal{D}$  used to estimate the approximation. We use the first  $n$  points of the training sample, for  $n$  starting from 1,000 to 50,000 (our benchmark approximation) in increments of 1,000. We consider three approximations: (a) a third-order polynomial approximation with no weight (blue line) (b) a linear approximation with weight  $k = 2$  (red line) (c) the benchmark approximation (third-order polynomial with  $k = 2$ ). We limit the illustration on the graph to a max of .25 on the y-axis.

**Figure A.3: True SMM Estimates, Convergence Speed of Local Optimization Stage (Corporate Finance Model)**



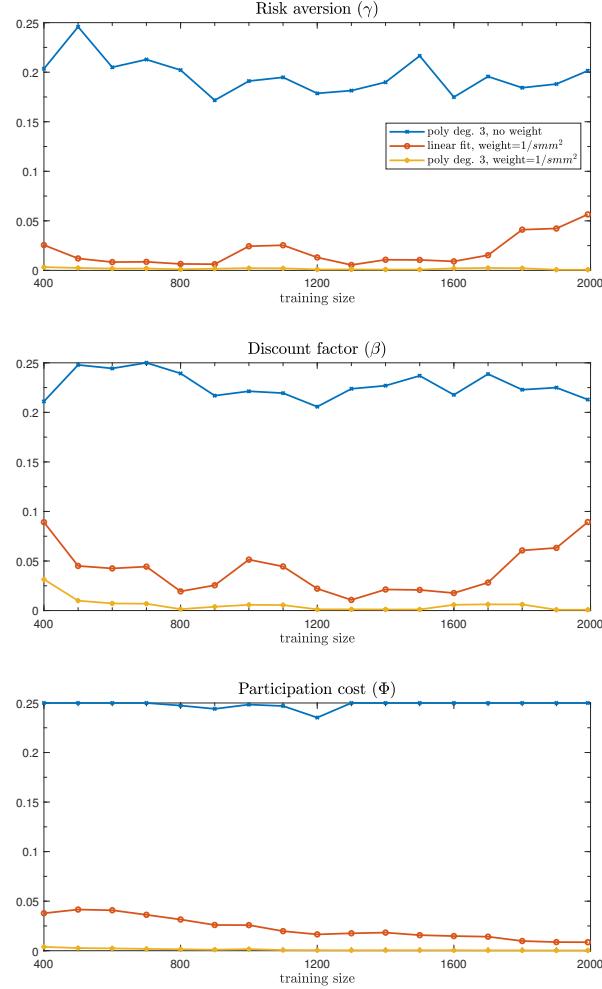
*Notes.* We report parameter values estimated as a function of time taken via two different algorithms in our numerical setup. This estimation corresponds to the baseline estimation described in Section 3.4. The blue line corresponds to the local optimization stage of the true SMM, i.e. the minimization of the distance of empirical moments to the true model  $f(\theta)$ . The optimization algorithm used in this case is Tiktak, using 50 starting points selected from a training set of 50,000 cases and Nelder-Mead algorithm for optimization per starting point with 200 max function iteration. The red line corresponds to the benchmark approximate SMM, which uses a third-order polynomial approximation and a weight  $k = 2$ . The approximate estimation requires .9 seconds. The black line corresponds to the true SMM estimate and the dashed lines represent the confidence interval around it.

**Figure A.4: Out-of-sample Performance across models and weight schemes (Household Finance Model)**



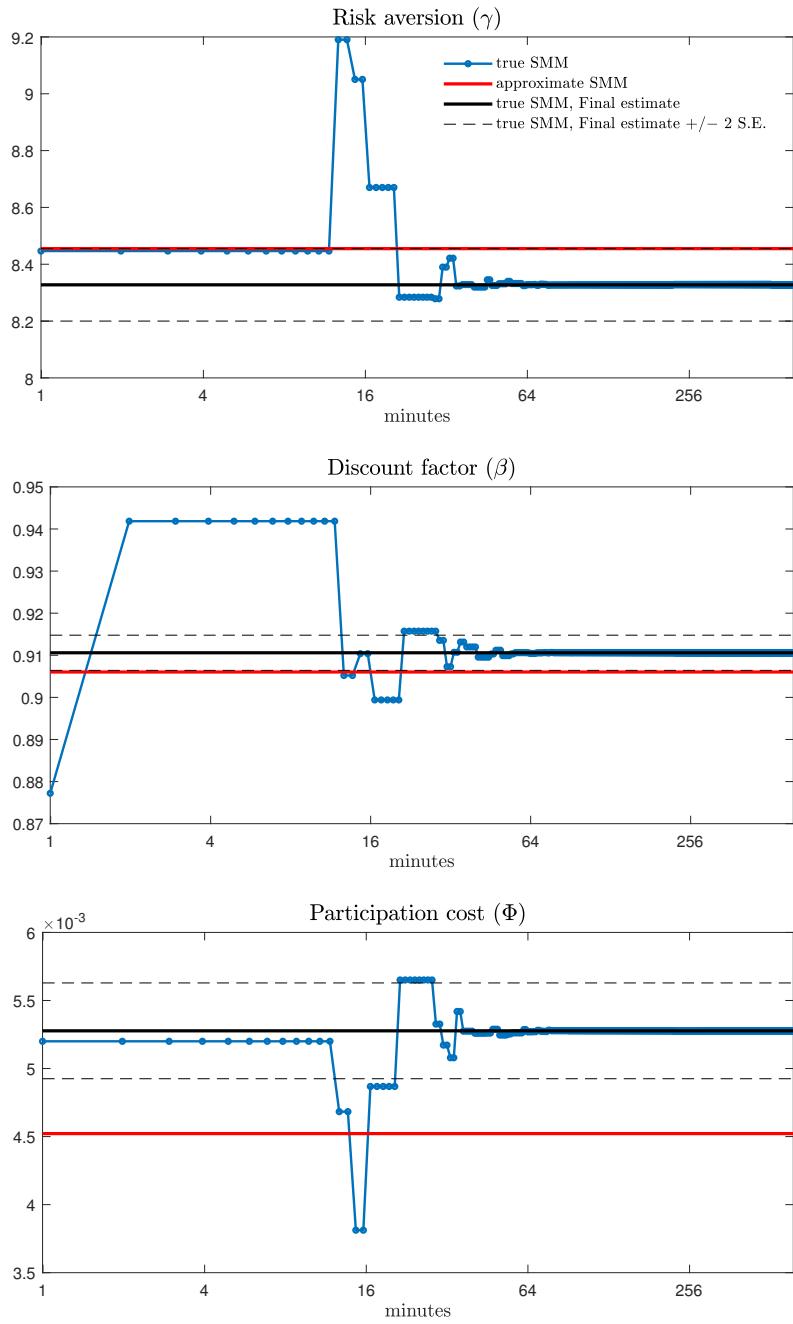
*Notes.* This figure reports, for the household finance model, a measure of the estimation error from the approximate SMM for different approximations and weighting schemes. For each draw  $(\theta_j^{\text{validation}}, f(\theta_j^{\text{validation}}))$  in the validation sample, we estimate parameters  $\widehat{\theta}_j^{\text{validation}}$  with an approximate SMM targeting moments  $f(\theta_j^{\text{validation}})$ . For each parameter  $i$  and approximation used, the figure reports  $1 - R^2(i) = \frac{\text{Var}[\theta_j^{\text{validation}(i)} - \theta_j^{\widehat{\text{validation}}(i)}]}{\text{Var}[\theta_j^{\text{validation}(i)}]}$ . We consider the following specification for the approximation  $g()$ : linear, third-order polynomial, a linear fit based on neighborhood points, a neural net with 2 layers, and a neural net with 5 layers. The approximation  $g(\theta; \beta)$  is estimated on the training dataset  $\mathcal{D}$  as  $\widehat{\beta} = \arg \min_{\beta} \left[ \sum_{l \in \mathcal{D}} \frac{1}{(\Delta_l^j)^k} (g(\theta_l; \beta) - f(\theta_l))' (g(\theta_l; \beta) - f(\theta_l)) \right]$  where  $\Delta_l^j = (\mathbf{m}_l - \mathbf{m}_j^{\text{validation}})' \Omega (\mathbf{m}_l - \mathbf{m}_j^{\text{validation}})$  and  $\Omega$  is the inverse of the variance-covariance matrix of the empirical moments. We consider  $k = 0, .5, 1, 2, 5$  and  $10$ , as well as  $1/\log \Delta_l^j$  weights. For the local linear fit, we only use 4 closest points in the parameter space in terms of norm-2 distance, and discard the rest, irrespective of  $k$ .

**Figure A.5: Out-of-sample Performance with respect to Training Sample Size (Household Finance Model)**



*Notes.* This figure explores the precision of the approximate SMM for the household finance model as a function of the size of the training sample used to fit the approximation. The y-axis reports a measure of the approximation error for estimated parameters,  $1 - R^2$ , described in fig. A.4. The x-axis corresponds to  $n$ , the size of the training dataset  $\mathcal{D}$  used to estimate the approximation. We use the first  $n$  points of the training sample, for  $n$  starting from 400 to 2,000 (our benchmark approximation) in increments of 100. We consider three approximations: (a) a third-order polynomial approximation with no weight (blue line) (b) a linear approximation with weight  $k = 2$  (red line) (c) the benchmark approximation (third-order polynomial with  $k = 2$ ). We limit the illustration on the graph to a max of .25 on the y-axis, considering all cases of  $R^2$  less than .75 as poor performance cases.

**Figure A.6: True SMM Estimates, Convergence Speed of Local Optimization Stage (Household Finance Model)**



*Notes.* We report, for the household finance model, parameter values estimated as a function of time taken via two different algorithms in our numerical setup. This estimation corresponds to the baseline estimation described in Section 4.3. The blue line corresponds to the local optimization stage of the true SMM, i.e. the minimization of the distance of empirical moments to the true model  $f(\theta)$ . The optimization algorithm used in this case is Tiktak, using 5 starting points selected from a training set of 2,000 cases and Nelder-Mead algorithm for local optimization per starting point with 200 max function iteration. The red line corresponds to the benchmark approximate SMM, which uses a third-order polynomial approximation and a weight  $k = 2$ . The approximate estimation requires .2 seconds – hence the red line jumps to its final value at the origin. The black line corresponds to the true SMM estimate and the dashed lines represent the confidence interval around the true SMM estimate.

## C Appendix Tables

Table A.1: Literature

Sub-field	Year	Reference	Comparative statics	Jacobian	AGS	N/A
Investment, capital structure, and financing	1992	Whited JF			✓	
	2003	Love RFE			✓	
	2005	Hennessy et al. JF			✓	
	2007	Hennessy et al. JF		✓		
	2011	DeAngelo et al. JFE	✓			
		Lin et al. JFE			✓	
	2013	Matvos RFS			✓	
	2014	Nikolov et al. JF	✓			
	2016	Warusawitharana et al. RFS	✓		✓	
		Li et al. RFS				
	2017	Bakke et al. JFE	✓			
		Gu JFE			✓	
Corporate governance	2018	Wu RFS		✓	✓	
	2019	Nikolov et al. JFE	✓			
		Frank et al. RFS				
	2021	Begnauet al. JFE			✓	
	(Forthcoming)	Catherine et al. JF	✓	✓	✓	
	2009	Gayle et al. AER			✓	
Bankruptcy	2010	Taylor JF			✓	
	2012	Kang et al. JFE			✓	
	2013	Coles et al. JFE		✓		
	2017	Taylor JFE			✓	
	2018	Jung et al. JFE		✓		
	2022	Page JFE			✓	
Banking	2016	Bertomeu JFE			✓	
Corporate control	2014	Glover JFE			✓	
Entrepreneurship	2015	Schroth et al. JFE			✓	
	2018	Dimopoulos et al. JFE	✓			
	2020	Albuquerque et al.JF	✓			
		Li et al. JFE			✓	
	2022	Wang JFE				✓
Household finance	2018	Wang JFE	✓			
	2019	Jones et al. AER			✓	
	2020	Ewens et al. JFE	✓			
	2022	Catherine JFE	✓			
Real estate	2018	Pagel Econometrica			✓	
	2019	Sun et al. JFE	✓			
	2020	Ameriks et al. JPE			✓	
	2022	Catherine RFS			✓	

**Table A.2: Preset Parameters of Life-cycle Model**

$p_z$	.136	$r$	.02	$p_s$	.146
$\rho_z$	.967	$\theta_1$	.1237	$\mu_s^-$	-.245
$\mu_z^-$	-.086	$\theta_2$	-.0125	$\mu_s^+$	.115
$\lambda_{zl}$	4.291	$\theta_0$	-3.015	$\sigma_{s1}$	.077
$\sigma_z^-$	.562	$t_0$	23	$\sigma_{s2}$	.114
$\sigma_z^+$	.037	$R$	65	$\mu_l$	.008
$\sigma_\eta^-$	.895	$T$	100	$\lambda_{ls}$	.161
$\sigma_\eta^+$	.089			$\sigma_l$	.017

*Notes.* This table shows the calibrated parameters used in the estimation of the life-cycle model introduced in Section 4.