Robustness Checks in Structural Analysis*

Sylvain Catherine † Mehran Ebrahimian‡
David Sraer§ David Thesmar¶

August 30, 2022

Abstract

Robustness checks, such as adding controls or sample splits, are a standard feature of reduced-form empirical research. Because of computational costs of reestimating 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 the NYU Stern School of Business and Berkeley.

†University of Pennsylvania
‡Stockholm School of Economics
§UC Berkeley, NBER and CEPR
¶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 channels can explain their findings and provide additional empirical analyses that control for such alternative channels. It is also standard to establish the robustness of findings across various sub-samples. When worried that the empirical results may depend on including particular years in the sample (e.g., a recession) or particular industries (e.g., service industries), economists will typically run additional regression analyses that exclude particular years or industries. Sample splits are also commonly used to assess the economic channels that underlie empirical findings. These analyses 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 evaluate the role of alternative mechanisms in explaining the main empirical finding. This is usually done by introducing additional control variables in a regression analysis. 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 estimating alternative structural models 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 show how it can be used to perform many robustness checks in the context of two standard applications in economics and finance.

Our approach works as follows. We consider the structural estimation of an economic model through a simulated method of moments. Given structural parameters $\boldsymbol{\theta}$, the model generates moments $\boldsymbol{m} = f(\boldsymbol{\theta})$. In most applications, as for instance with dynamic models, f is calculated numerically. Structural parameters $\boldsymbol{\theta}$ are then estimated by minimizing a distance between simulated moments $f(\boldsymbol{\theta})$ and empirical moments $\widehat{m}$. While simulating the economic model given $\boldsymbol{\theta}$ is in general computationally cheap, estimating $\boldsymbol{\theta}$ can be costly as it can require a large number of 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 rely on a parametric approximation of the *moment function* $f(\boldsymbol{\theta})$, $g(\boldsymbol{\theta}, \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a vector of parameters that characterize this approximation. Our approach boils down to estimating $\boldsymbol{\beta}$. In contrast to the numerical solution of the underlying model, which can be non-linear and therefore hard to approximate, the function $f()$, which maps structural parameters to moments, is smooth. This feature allows us to obtain good approximations of $f()$ at low computational costs (i.e. using low-dimensional parameters $\boldsymbol{\beta}$).

Our methodology proceeds in three steps. First, we draw a large number (N) of potential parameters $(\boldsymbol{\theta}_i)_{i\in 1,N}$ and simulate corresponding moments $\boldsymbol{m}_i = f(\boldsymbol{\theta}_i)$. This generates a dataset $\mathcal{D} = \{(\boldsymbol{\theta}_i, f(\boldsymbol{\theta}_i))\}_{i\in[1,N]}$, 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. Second, we use the dataset $\mathcal{D} = \{(\boldsymbol{\theta}_i, f(\boldsymbol{\theta}_i))\}_{i\in[1,N]}$ to fit a parametric approximation of $f()$. For instance, $g()$ can be a polynomial function. In this case, the coefficients $\boldsymbol{\beta}$ can be estimated at close to zero computational cost via regressions. More advanced statistical learning methods can be used if in-sample over-fitting is a concern (e.g., penalized regressions or tree-based methods). In our applications, this step – the estimation of β – is computationally cheap, as we show that a third-order polynomial function provides a precise approximation of $f()$. Third, once $\boldsymbol{\beta}$ has been estimated, the "approximate" moment function $g(\boldsymbol{\theta}, \widehat{\boldsymbol{\beta}})$ can be computed for any parameter θ without simulation. This feature allows us to minimize the distance between empirical moments $\widehat{m}$ and approximate simulated moments $g(\boldsymbol{\theta}, \widehat{\boldsymbol{\beta}})|$ at close to zero computational cost. This approach can be seen as a generalization of Andrews et al. (2017b), who focus on first-order local deviations and thus restrict $g()$ to be a linear function. As we show in our applications, focusing on linear functions can be restrictive. We discuss the relation of our approach to that of Andrews et al. (2017b) in Section 2.

How precise are the parameter estimates resulting from our approximation? We answer this question using simulated data in the context of two canonical models of

the literature: (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 generate the dataset $\mathcal{D}$ of parameters and simulated moments introduced above. We split $\mathcal{D}$ into a training and a validation sample. We use the training sample to estimate the parameters β of our approximation. For each set of moments in the validation sample $f(\boldsymbol{\theta}_i)$, we estimate structural parameters $\widehat{\boldsymbol{\theta}}$ using the approximation $g(\boldsymbol{\theta}, \widehat{\boldsymbol{\beta}})$ fit on the training sample. We then compare the actual structural parameters that generate the moments in the validation sample, $f(\boldsymbol{\theta}_i)$, with the parameter estimates $\widehat{\boldsymbol{\theta}}_i$. In both applications, when the structural parameters are identified in the true model, the estimates from the approximate model are almost identical to the true parameters: the correlation between true and estimated parameters across the validation sample is in most cases larger than .99, and never lower than .97.

Using the approximation moment function $g(\boldsymbol{\theta}, \widehat{\boldsymbol{\beta}})$ instead of the true moment function $f(\boldsymbol{\theta})$ 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 typically rely on the function $f(\theta)$ in the vicinity of $\widehat{\theta}$, i.e. how local variations in parameters affect simulated moments. This approach is computationally cheap as it only requires a few additional simulations of the underlying economic model. As pointed out by Andrews et al. (2017a), a preferable diagnostic tool is the relationship between parameter estimates $\widehat{\boldsymbol{\theta}}$ and empirical moments $\widehat{m}$: how variations in targeted moments affect parameter estimates? However, uncovering this relationship has a high computationally cost: it requires re-estimating the model for each alternative moment value. Andrews et al. (2017a) bypass this cost by consider a first-order local approximation. Their approach essentially involves inverting the Jacobian matrix – a linear approach that is valid for small changes in moments around their empirical value. Because our methodology reduces the cost of re-estimating the model on alternative moments by several orders of magnitude, it allows us to trace out the mapping from targeted moments to parameter estimates without relying on a first-order approximation. We show, in our two applications, that this linear approximation is restrictive for several moments and thus may fail to provide all the information required to discuss identification.

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 ($\mathcal{M}^{\text{used}}$) and non-targeted moments ($\mathcal{M}^{\text{non-target}}$) (2) compare non-targeted moments in the data to their simulated counterparts. This approach is not ideal. The selection of targeted vs. non-targeted moments is arbitrary. This approach also remains silent on how parameter estimates would change had non-targeted moments been included in the set of targeted moments. We leverage the low estimation cost offered by our methodology to develop two systematic tests of robustness to moment selection. We first consider how parameter estimates change when one of the moments in $\mathcal{M}^{\text{non-target}}$ is included in $\mathcal{M}^{\text{used}}$. A robust baseline estimation should not be sensitive to these inclusions. A second exercise re-estimates the model using *all possible combinations of moments in* $\mathcal{M}^{\text{non-target}}$ and plots the resulting distribution of parameter estimates. Because this exercise requires thousands of estimations, it would be infeasible with standard simulation-based methods. A robust estimation should exhibit a peaked distribution of alternative parameter estimates around their baseline value. When a parameter estimates is not robust to moment selection, this analysis serves as a useful diagnostic tool to identify features of the data missed by the model, and thus helps amend the model to better fit the data. In our two applications, we find that few parameter estimates are robust to moment selection.

A standard concern in empirical work is the robustness of parameter estimates to the sample used in estimation. In reduced-form analyses, it is for instance common to re-estimate regression models on different sub-samples of industries (e.g., manufacturing vs. services) or years (e.g., excluding recessions). Sample splits are also commonly used as additional tests to validate the robustness of the baseline findings (e.g., when the main effect is expected to be stronger for small vs. large firms). In structural work, the equivalent exercise would require to (1) recompute moments on different sub-samples (e.g., for each industry in the sample) (2) re-estimate the model using these alternative moments. Working with dozens or hundreds of sub-samples can make such robustness checks computationally prohibitive with standard simulation-based methods. In contrast, the low computational cost of estimating the model with the approximate moment function $g()$ makes these robustness checks feasible. As a result, researchers can estimate and report structural parameters for each year or industry in the sample. 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 misspecification, a standard concern with structural analysis (e.g., Andrews et al. (2017a)): economic forces not included in the model may alter the inference of struc-

tural parameters. In reduced-form work, robustness to misspecification can be addressed at low computational cost by adding variables to regression analyses (e.g., non-linear controls or omitted variables that may affect the outcome variable). In structural work, considering alternative models and how they affect estimated parameters of interest is computationally costly as it involves a new estimation for each alternative considered. Our methodology does not allow us to speed up the estimation of a new model. However, we can still address robustness to misspecification in the following way. Let $\widehat{\theta^{\text{baseline}}}$ be the set of structural parameters recovered using the baseline model. Consider an alternative to this baseline model, with structural parameters θ^a and moments $h(\theta^a)$. Using our approximation $g(\boldsymbol{\theta}, \widehat{\boldsymbol{\beta}})$, we can estimate how inference on θ is affected when we target the alternative moments $h(\theta^a)$ generated by the alternative model. Note that this methodology only requires to (1) solve the alternative models to generate simulated moments (2) estimate the approximate model using moments simulated under the alternative model. Both steps carry low computational cost. This allows us to consider a large number of alternative parameters θ^a. Note that we can perform similar robustness exercises for functions of parameters (e.g., welfare measures).

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. (2017a) derives a local linear approximation of the relationship between parameter estimates and the moments of the data they depend on. This measure serves as a diagnostic tool that allows to easily assess potential misspecification bias for a range of alternatives models. 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 methodology can be seen as a generalization of Andrews et al. (2017a) – as discussed in Section 2. Our approximation is global, not local. As we show below, non-linearity can be an important feature of the relationship between moments and parameters, so that local approximations sometime produce erroneous diagnostics. In addition, our numerical approach allows quantitative researchers to report many sub-sample estimates, which is another way in which structural work can be made more transparent. Our analysis more generally 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, the relation between moment and structural parameters – the function $f(\boldsymbol{\theta})$ – is usually the focal point. Table A.1 provides a review of recent structural papers in corporate and household finance. Out of 36 papers, one third show local comparative statics (i.e. plots of $m(\boldsymbol{\theta})$ around the baseline estimates, nine reports the Jacobian matrix ($J(\widehat{\theta}) = \nabla f(\widehat{\theta})$), and two thirds do not report any of these two. Reporting $f(\boldsymbol{\theta})$) or its derivatives around the parameter estimates is useful. It confirms the precision of the numerical solution. It also helps understand how the model works. However, it is not a sufficient tool to discuss the robustness of the estimates, in particular how they depend on the targeted moments. Four recent papers in Table A.1 reports the sensitivity matrix of Andrews et al. (2017a), which provides a linear approximation of this relation between parameter estimates and targeted moments. Our approach provides a low-cost method to evaluate how estimates depends on targeted moment that does not rely on this linear approximation.

Our paper is also related to the emerging literature trying to improve numerical solutions of *models* through approximations. Our goal is quite different, and, in a sense, simpler. We seek to approximate the model *moments* in order to speed up *estimation* and perform robustness analysis. Norets (2012) extends the state-space by adding the model parameters as "pseudo-states" to efficiently estimate finite-horizon, dynamic discrete choice models. He uses shallow artificial neural networks to approximate the dynamic programming solution as a function of both state variables and parameters prior to estimation. Chen et al. (2021) use an insight similar to ours but use it to solve the model instead of the relation between moments and parameters; their focus is on solution more than estimation. Like us, they draw a large set of parameters, and solve the model numerically for each parameter draw. They then train a Machine Learning (ML) algorithm to predict model outcomes as a function of parameters. The non-linearity of their model requires that they use highly flexible ML algorithms (e.g., deep Neural Networks (NN)). The models we consider in our applications also feature solutions that may be highly non-linear functions of state variables. However, having a flexible approximation is less crucial for our purpose, since e do not approximate policy functions but instead the relation between moments and parameters $f(\boldsymbol{\theta})$, which is smoother. Our focus on $f(\theta)$ is driven by our interest in estimation and robustness analyses, a focus different from theirs. Similarly, Duarte (2020) describes a new solution method combining ML algorithms and Gradient Descent Algorithm. The intuition is that value functions can be represented through

NNs rather than functions defined on a grid. Using value function iteration, the fixed point of the Bellman problem can be obtained through gradient descent. His method allows to quickly solve models with large state spaces. Overall, this literature builds on the burgeoning fied that develops approximate solutions to dynamic quantitative models in economics and finance using machine learning tools(e.g., Duarte (2020), Fernández-Villaverde et al. (2021a), Villa and Valaitis (2019),Maliar et al. (2019), Azinovic et al. (2019)). In contrast to these papers, we solve the model *exactly* to generate a training sample that we then use to approximate the mapping between simulated *moments* from the true economic model and parameters. Our paper also differs by its focus on robustness and the scope of its applications (corporate and household finance).

Finally, our paper is related to the vast literature that structurally estimate dynamic models of corporate and household finance (see Strebulaev and Whited (2012) for a survey of the corporate finance literature, and Gomes et al. (2021) for household finance). Our paper provides a new methodology to improve the transparency of identification and assess the robustness of estimated models in these literature.

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 $\boldsymbol{\theta}$. The model $\mathcal{S}$ generates a vector of moments $f(\boldsymbol{\theta}) \in \mathbb{R}^M$ 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 $\widehat{\boldsymbol{m}}$ be the vector of empirical counterparts to the model-based moments $f(\boldsymbol{\theta})$.

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

$$\widehat{\theta} = \arg\min_{\theta} \left(\widehat{m} - f(\theta)\right)' \Omega \left(\widehat{m} - f(\theta)\right) \tag{1}$$

where Ω is a weighting matrix.

The estimation proceeds in two-steps:

1. For a given θ, standard numerical methods are used to solve the model given θ and simulate model-based moments $f(\theta)$ (inner loop).

2. θ is selected to minimize the objective in 1 (outer loop).

Because θ 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 estimation: if the model has to be estimated against a different set of empirical moments $\widehat{m}'$ (e.g., moments estimated for different subsamples, or alternative moments not included in $\widehat{m}$), 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 $f()$. Our approach proceeds in four steps:

(i) We define ex ante bounds for each parameter in $\boldsymbol{\theta}$. These bounds are based on expert knowledge, and also have to be specified for standard estimation techniques in the literature. We note Θ the resulting set of admissible parameter vectors.

(ii) We use a Halton sequence[1] to generate N vectors of parameters $(\boldsymbol{\theta}_i)_{i\in[1,N]}$ with N large.

(iii) For each vector $\boldsymbol{\theta}_i$, we solve the model $\mathcal{S}$ and simulate moments $f(\boldsymbol{\theta}_i)$. This is computationally intensive as the true model has to be solved and simulated N times. This steps results in a dataset $\mathcal{D} = \{(\boldsymbol{\theta}_i, f(\boldsymbol{\theta}_i))\}_{i\in[1,S]}$ that contains N vectors of parameters and the corresponding N vector of moments.

(iv) On the dataset $\mathcal{D}$, we estimate the parameters $\boldsymbol{\beta}$ of an approximation $g(\boldsymbol{\theta}, \boldsymbol{\beta})$ of the function $f(\boldsymbol{\theta})$ in the following way:

   (a) For each draw $(\theta_i, m_i) \in \mathcal{D}$, we compute the distance between m_i and the moment values of interest $\widehat{\boldsymbol{m}}$: $\Delta_i = (m_i - \widehat{\boldsymbol{m}})'W(m_i - \widehat{\boldsymbol{m}})$. W is a scaling matrix. In our applications, we choose the inverse of the empirical

[1] 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.

variance-covariance matrix of moments in the overall sample, estimated using bootstrap. This matrix corresponds to the efficient weighting matrix for estimation in Equation 1.

(b) The approximation is estimated to minimize the following objective:

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

k measures the weight the approximate model puts on observations in $\mathcal{D}$ with moments close to the empirical moments $\widehat{\boldsymbol{m}}$. 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 empirical moments $\widehat{\boldsymbol{m}}$ that are targeted in the estimation.

In our applications, we consider, three classes of functions for $g()$: linear, third degree polynomials and neural nets. Once $\boldsymbol{\theta}$ is estimated, we use the approximation $g(\boldsymbol{\theta}, \widehat{\beta})$ to estimate the deep parameters $\boldsymbol{\theta}$ by solving:

$$\widehat{\theta}(\widehat{\boldsymbol{m}}; \widehat{\boldsymbol{\beta}}) = \arg\min_{\theta} \left(\widehat{\boldsymbol{m}} - g(\theta; \widehat{\boldsymbol{\beta}}) \right)' \Omega \left(\widehat{\boldsymbol{m}} - g(\theta; \widehat{\boldsymbol{\beta}}) \right) \tag{2}$$

This last step is fast since simulations are no longer required, and $g()$ is a closed-form, smooth function of θ. We do not have formulas for standard errors of $\widehat{\theta}(\widehat{\boldsymbol{m}}; \widehat{\boldsymbol{\beta}})$, 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.

2.2 Relation with the sensitivity matrix

Andrews et al. (2017a) 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. (2017a) that relies on a higher-order approximation.

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

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

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

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

$$0 = J(\boldsymbol{\theta})'\Omega(\widehat{\boldsymbol{m}} - g^{\text{linear}}(\widehat{\boldsymbol{\theta}})) = J'(\boldsymbol{\theta})\Omega \left(\widehat{\boldsymbol{m}} - \boldsymbol{m} - J(\boldsymbol{\theta}) \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\right)$$

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

$$\left(\widehat{\boldsymbol{\theta}} - \theta\right) \approx \underbrace{-(J'(\boldsymbol{\theta})\Omega J(\boldsymbol{\theta}))^{-1} J'(\boldsymbol{\theta})\Omega}_{\Lambda} (\widehat{m} - m) \tag{3}$$

Assume for instance that $\boldsymbol{\theta}$ are parameters estimated using SMM and $\boldsymbol{m}$ corresponds to the empirical moments. Andrews et al. (2017b) 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 $\boldsymbol{m} = f(\boldsymbol{\theta})$. Conceptually, Andrews et al. (2017b) suggest that, for moments in the neighborhood of $\boldsymbol{m}$, parameter 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. At date t, the firm's EBIDTA, π_t, depends on the firm's capital stock and its productivity:

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

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

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

$$k_{t+1} = k_t + i_t - \delta k_t, \tag{5}$$

where δ is the depreciation rate. In period t, investing i_t entails a convex cost of $\frac{\gamma}{2}\frac{i_t^2}{k_t}$. 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 (δk_t) are taxed at a rate τ. 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_i 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 of debt issued at date t, the firm commits to repay $(1+r)d_{t+1}$ at date $t+1$. Finally, the interest rate the firm receives on cash is lower than the interest rate it has to pay on its debt. If the firm has negative net debt, it receives a positive cash inflow of $-(1+(1-m)r)d_{t+1}$ with $0 < m < 1$.

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:

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

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:

$$d_{t+1} \leq \lambda k_{t+1}. \tag{6}$$

The total collateral available to the creditor at the end of period $t+1$ consists of depreciated productive capital λk_{t+1}. λ, 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 statistics 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 $V^\star(k, d, z)$ the value of a firm with state variable (k, d, z) in the absence of financial frictions (i.e. when equity issuance cost is 0, $\zeta = 0$). We define the value loss as:

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

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 values $\delta \in [0; .2]$, $\gamma \in [0; .12]$, $\alpha \in [.4; .9]$, $\rho_z \in [.4; 1]$, $\sigma \in [0; .35]$, $\zeta \in [0; .2]$ and $\lambda \in [0; 1]$. This defines the set Θ of possible parameter values.

We then draw a Halton sequence of $N = 50,000$ vectors θ_i in Θ – we verify below that N is large enough to generate a good approximation. For each $\boldsymbol{\theta}_i$, we simulate the model and compute a vector of 17 moments $\boldsymbol{m}_i$, 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 correspond to "core" moments that are typically targeted in the literature: m_1 = mean(investment/assets), m_2 = mean(profit/assets), m_3 = mean(equity issuance/assets), m_4 = mean(net leverage), m_5 = autocorrelation of (investment/assets), m_6 = std(log growth sales), m_7 = std(log growth 5yr sales). These moments are our baseline moments, i.e. moments that we always target in our estimation.

The next 10 moments correspond to moments that have been less frequently targeted in the literature. We will explore the role of these moments for estimation in our robustness exercises below. These moments are: m_8 = var(investment/assets), m_9 = var(equity issuance/assets), m_{10} = frequency(equity issuance), m_{11} = coefficient of the regression of investment ratio on market to book ratio, m_{12} = coefficient of the regression of net leverage on market to book ratio, m_{13} = coefficient of the AR(1) regression of profit ratio with year fixed-effects, m_{14} = residual std of the AR(1) regression of profit ratio with year fixed-effects, m_{15} = var(leverage), m_{16} = mean(dividend/assets), m_{17} = var(dividend/assets),

The training dataset $\mathcal{D} = (\boldsymbol{\theta}_i, \boldsymbol{m}_i)_{i \in [1,50000]}$ is then used to fit several approximations $g()$ of the relationship between moments and parameters $\boldsymbol{m}_i = f(\boldsymbol{\theta}_i) \approx g(\boldsymbol{\theta}_i, \boldsymbol{\beta})$. 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 with our numerical setup, generating the training dataset takes about 140 hours. However, note that SMM estimations that use the Tik Tak algorithm with the same number of starting points require the exact same training step. Tik Tak is a multi-start algorithm that has been shown to have the strongest performance in both mathematics test functions and economic applications (Arnoud et al., 2019). 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

estimation that targets alternative moments *use the same training dataset*. The only remaining cost for additional estimations is to fit the approximation in the vicinity of the alternative targeted moments, which, in our applications, is minimal.

3.3 Validating the approximation approach

We start our analysis by assessing the precision of our approximation approach. We first draw a *validation* dataset $\mathcal{D}^{\text{validation}}$ made out of $1,000$ additional random draws of parameters $\left(\boldsymbol{\theta}_j^{\text{validation}}\right)_{j\in[1,1000]}$ within the bounds defined above, and their corresponding moments $\left(f\left(\boldsymbol{\theta}_j^{\text{validation}}\right)\right)_{j\in[1,1000]}$. For each of these draws, we estimate the model using the methodology described in Section 2.1 and targeting the vector of moment $f\left(\boldsymbol{\theta}_j^{\text{validation}}\right)$. We evaluate the precision of the estimation by comparing the resulting parameter estimates $\widehat{\boldsymbol{\theta}_j}$ with the vector of parameters that generated the moments, $\boldsymbol{\theta}_j$.

For this exercise, we consider a just-identified estimation of the seven structural parameters that targets only $(m_i)_{i\in\{1,\dots,7\}}$ described above. While the parameters $\boldsymbol{\theta}$ 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 δ. Mean(profit/assets) (m_2) pins down returns to scale α. Mean(equity issuance/assets) (m_3) informs the model about the cost of issuing equity while mean(leverage) (m_4) contributes to the estimation the collateral parameter λ. Estimates of adjustment cost γ are mostly sensitive to the autocorrelation of investment (m_5). Finally, the persistence of the investment ratio (m_6) and std(log growth 5yr sales) (m_7) help pin down the persistence and volatility of TFP shocks (ρ_z and σ_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 the moments $f(\boldsymbol{\theta}_j^{\text{validation}})$. However, since $f(\boldsymbol{\theta}_j^{\text{validation}})$ is not an empirical object, we cannot estimate the sampling variance of these simulated moments.

Instead, we approximate it using the inverse of the actual variance-covariance matrix of the baseline moments targeted in estimation, Ω. 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 error is 10 times larger than the standard error of the parameter estimates in the baseline estimates (i.e. targeting the baseline empirical moments).[2] 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. The remaining dataset of 158 observations constitute our validation sample.[3]

For each vector of moments $\boldsymbol{m}_j^{\text{validation}} = f(\boldsymbol{\theta}_j^{\text{validation}})$ in the validation sample, we then use the training dataset constructed in Section 3.2 to fit a local approximation of the relationship $\boldsymbol{m} = f(\boldsymbol{\theta})$. We follow similar steps to (iv)-(a) and (iv)-(b) in Section 2.1: (a) for each validation moment $\boldsymbol{m}_j^{\text{validation}}$, we compute the distance between each training moment $\boldsymbol{m}_i$ and $\boldsymbol{m}_j^{\text{validation}}$: $\Delta_i^j = (\boldsymbol{m}_i - \boldsymbol{m}_j^{\text{validation}})'\Omega(\boldsymbol{m}_i - \boldsymbol{m}_j^{\text{validation}})$ (b) we then fit the approximation by minimizing the following objective in the *training dataset*:

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

where we experiment with different levels of k. When $k = 0$, observations in the training sample are equal-weighted, and the approximation $\widehat{\beta}$ will be the same for each validation moment. As k increases, our approximation, for a given $\boldsymbol{m}_j^{\text{validation}}$, puts larger weights on observations "near" $\boldsymbol{m}_j^{\text{validation}}$ in the sense of the SMM objective function. As a result, when $k > 0$, $\widehat{\beta}$ depends on the moments $\boldsymbol{m}_j^{\text{validation}}$ targeted in the estimation and thus needs to be re-estimated for each set of moments in the validation sample. However, since fitting the approximation and estimating the approximate model are both computationally cheap, this step remains order of magnitudes cheaper than estimating the true model.

A difference between problem (7) and our general presentation in Section 2 is that we use log-transforms for moments that are positive: instead of matching their value m_i, we instead target $\log(m_i + \epsilon_i)$, where the constant ϵ is scaled by the size of the moment through the formula $\epsilon_i = \max\{10^{-6}, 10^{-2}m_i\}$. This steps improves the "gran-

[2]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.

[3]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.

ularity" of our estimation in situations where moment values are small. For instance, when equity issuance cost are large, average issuance is close to 0. Taking such a log-transform of average equity issuance helps to obtain a more accurate estimation of issuance costs.

We consider 5 different functional forms for g. The first one is a linear approximation $g(\boldsymbol{\theta}, \boldsymbol{\beta}) = \boldsymbol{\beta}'\boldsymbol{\theta}$. We also use a third-order polynomial approximation $g(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{\alpha_0+\alpha_1+\alpha_2+...+\alpha_7 \leq 3} \beta_{\alpha_0\alpha_1\alpha_2...\alpha_7} 1^{\alpha_0} \prod_{m=1}^{7} (\theta_m)^{\alpha_m}$. 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. We also 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.

For each one of these five possible functions, and various weights k, we then estimate the parameters β of the approximate model. As explained above, $\widehat{\beta}$ technically depends on $\boldsymbol{m}_j^{\text{validation}}$ since, as soon as $k > 0$, it is estimated using weights that increases close to the targeted moments $\boldsymbol{m}_j^{\text{validation}}$. However, we omit this dependence to keep notations light. Note also that β is estimated using the entire *training sample*. As is the case for the validation sample, the training sample includes draws of parameters for which the model is not identified. However, since the validation sample only contains moments where the true model is well-identified, the non-identified draws in the training datasets are presumably far enough from the validation moments that they do not affect the fit of the approximation. We thus keep all the 50,000 draws in the initial training dataset, except for 205 observations that are either undefined or extreme outliers.[4]

In a final step, we estimate parameters $\widehat{\boldsymbol{\theta}_j^{\text{validation}}}$ for each of the 158 vectors of moments in the validation dataset by minimizing the approximate-SMM objective:

$$\widehat{\boldsymbol{\theta}_j^{\text{validation}}} = \arg\min_{\theta} \left(\boldsymbol{m}_j^{\text{validation}} - g(\theta; \widehat{\boldsymbol{\beta}}) \right)' \Omega \left(\boldsymbol{m}_j^{\text{validation}} - g(\theta; \widehat{\boldsymbol{\beta}}) \right) \tag{8}$$

We measure the accuracy of the resulting estimates,$\widehat{\boldsymbol{\theta}_j^{\text{validation}}}$, by computing, for the seven dimensions i of the parameter vector, the share of the variance of the true parameters $\boldsymbol{\theta}_j^{\text{validation}}(i)$ left unexplained by the approximate parameters $\widehat{\boldsymbol{\theta}_j^{\text{validation}}}(i)$:

[4]More precisely, we drop cases with either (a) undefined auto-correlation of investment-asset ratio because of zero variance (this drops only one case) (b) mean and variance of investment-asset ratio of greater than 1 (this drops an additional 157 cases) or (c) mean and variance of equity issuance-asset ratio greater than 1 (this drops an additional 47 cases).

$$1 - R_i^2 = \frac{\mathbb{V}\text{ar}\left[\boldsymbol{\theta}_{\boldsymbol{j}}^{\text{validation}}(\boldsymbol{i}) - \widehat{\boldsymbol{\theta}_{\boldsymbol{j}}^{\text{validation}}}(i)\right]}{\mathbb{V}\text{ar}\left[\boldsymbol{\theta}_{\boldsymbol{j}}^{\text{validation}}(i)\right]}, \tag{9}$$

where variances are taken across the 158 draws in the validation sample. $1 - R_i^2$ is small when the approximate estimation yields parameter estimates that are, on average, close to the true parameters.

Figure 1 reports this measure $1 - R_i^2$ for the seven estimated parameters and the value loss statistic described in Section 3.1. The measure is computed for each of the five functional forms described above, and seven different weighting scalar k.

Figure 1 highlights that polynomial approximations improve the precision of estimation relative to linear functions. The linear fit leads to values for $1 - R^2$ that are at least twice larger than those computed for the third-order polynomial approximation. Neural nets – which can account for further non-linearity in $f(\theta)$ – are not preferable to the third-order polynomial approximation: they require significantly more computing time to be estimated,[5] but do not bring significant fit improvement. Figure 1 also makes it clear that weighting matters. The choice of the weight k involves the following trade-off. When k is low, all observations in the training dataset have a similar weight, and the approximation fails to detect the particular shape of $f(\boldsymbol{\theta})$ in the neighborhood of the targeted moments, leading to imprecise estimates. In contrast, when k is high, the approximation is calibrated on a small number of observations of the training dataset, which results in overfitting and imprecise estimates as well.

Given the findings in Figure 1, we use, in the rest of this section, the third-order polynomial approximation with weights $\frac{1}{(\Delta_i^j))^2}$ as our benchmark approximation model. The performance of this benchmark approximation is high on the validation sample: for most parameters, the approximation explains more than 99% of the variance of the parameters in the validation sample ($1 - R^2 < .01$). For the equity issuance parameter and the value loss – the two estimates with a weaker performance – $1 - R^2$ remains below 0.03.

Figure 2 provides additional details on the precision of our preferred approximate estimation (third-order polynomial with quadratic weights). Figure 2 plots, for each draw of the validation sample, the estimated parameters $\widehat{\boldsymbol{\theta}^{\text{validation}}}$ (y-axis) against the true parameters $\boldsymbol{\theta}^{\text{validation}}$. A perfect approximation would put all points on the 45 degree line. The horizontal bars in Figure 2 correspond to the square roots of the

[5] With our numerical setup, the SMM using the polynomial approximation takes 4 seconds while the two-layer neural net takes about 111 seconds

diagonal terms of the matrix $J'(\boldsymbol{\theta}^{\text{validation}})\Omega J\boldsymbol{\theta}^{\text{validation}}$, where Ω is the inverse of the variance-covariance matrix of the data moments used in our baseline estimation, and $J\boldsymbol{\theta}^{\text{validation}}$ is the Jacobian matrix of the true model around the value of the parameter that generated these moments, $\boldsymbol{\theta}^{\text{validation}}$. Wide bars imply that the parameter is poorly identified by the true model for this particular draw of the validation sample.

Across all seven parameters, the large mass of observations on the 45 degree line confirms the findings of Figure 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 lead to parameter estimates with large standard errors. Despite our attempt at weeding-out such poorly-identified draws, some remain in the validation sample. Unsurprisingly, for these draws, the approximate SMM – like the true SMM – has a harder time recovering the true parameter values. This issue appears clearly for high values of equity issuance costs, which all leads to similarly low values of m_3 the average equity issuance to asset ratio.

How does N – the number of draws in the training sample – affect the precision of the approximate estimates? For the seven parameters and three different approximations (third-order polynomial with no weight, third-order polynomial with $k = 2$, and linear weights with $k = 2$), Appendix Figure A.1 shows how $1 - R_i^2$ varies for values of N ranging from 10,000 to 50,000 (the value in our benchmark approximation). Figure A.1 first confirms the superior precision of our benchmark approximation, irrespective of the size of the training sample. In general, the precision of the approximate SMM does not increase much with the size of the training sample. For most parameters, $1 - R^2$ is below 5% when N=10,000. The one exception is equity issuance costs: since the approximation can only work well for low enough values of this parameter (otherwise the true model itself is poorly identified), the training sample needs to be large to contain enough well-identified draws, i.e. draws where equity issuance costs are small.

3.4 Actual Estimation with Approximate SMM

This section evaluates the precision of the approximate SMM using actual data. We simply compare the approximate estimates and the true-SMM estimates when targeting actual empirical moments.

3.4.1 Data

Our sample comes from COMPUSTAT. Our sample period is 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 the seven moments $(m_i)_{1,...,7}$ targeted in the baseline estimation. We described above how we compute these moments in the model. In the data, we compute moments as follows: m_1, the average investment to capital ration, is $\frac{\texttt{capx}}{\texttt{l.at}}$. m_2, the average profit to asset ratio, is $\frac{\texttt{oibdp}}{\texttt{l.at}}$. m_3, the average equity issuance to asset ratio, is computed net of repurchases: $\frac{\texttt{sstk}-\texttt{prstkc}}{\texttt{l.at}}$. m_4, mean net leverage, is $\frac{\texttt{dlc+dltt }-\texttt{che}}{\texttt{at}}$. m_5, the autocorrelation of investment rates, is measured by regressing $\frac{\texttt{capx}}{\texttt{l.at}}$ 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: $\log \texttt{sale} - \log \texttt{l.sale}$ and $\log \texttt{sale} - \log \texttt{l5.sale}$. 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 deviations of these moments in our sample.

3.4.2 Estimated Parameters

Table 2 presents parameters estimates of the model using two estimation techniques that target the seven moments described in Section 3.4.1:

1. A standard SMM, that 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. Standard errors are computed using the standard formula $J'\Omega J$, where the Jacobian J is computed at the SMM estimate, and Ω is the inverse of the variance-covariance matrix of targeted moments.

2. An approximate SMM that uses the benchmark approximation (third-order polynomial with $\frac{1}{\Delta^2}$ weights). We compute standard errors using the delta method with the approximate Jacobian. This estimation of standard errors underesti-

mate true standard errors as it neglects the error introduced by the approximation.

Table 2 reveals that the two resulting estimators are quantitatively close. For all parameters, both estimators lie within a few percentage points of each other. The difference in estimated parameter values across the two approaches is well within the estimated standard errors. Columns 2,3,4 of Table 1 shows that both estimators are equivalent in their ability to match the empirical moments. Column 2 of Table 1 reports the value of simulated moments using the true model $f()$ and the true-SMM estimate. Column 3 shows the value of approximate moments at the approximate estimate $g(\widehat{\theta^{\text{validation}}}, \widehat{\beta})$. Column 4 computes the values of simulated moments using the true model, but the approximate estimate $\widehat{\theta^{\text{validation}}}$: $f(\widehat{\theta^{\text{validation}}})$. For both targeted (first 7 lines) and non-targeted moments (last 10 lines), Column (4) shows that the approximate SMM leads to moments that are both close to the moments generated by the true-SMM (Column (2)) and to the data (Column (1)).

3.4.3 Computing time

How much computational cost is saved by using the approximate estimation vs. the true SMM? Figure 3 shows the estimated structural parameters as a function of computing time in our setup, for both the standard and approximate SMM using the benchmark specification. The computing times we report exclude the simulation of the training sample, which is required for both estimations. For the true SMM (blue line), it takes at least 17 minutes for the estimation to converge to its final value; for some parameters like productivity persistence or return to scale, full convergence requires as much as nine hours. In contrast, the approximation can be fit and the approximate model estimated in less than four seconds. This is why the approximate SMM (red line) jumps right away to its final value. Therefore, once the training dataset has been simulated, it takes at least 250 times longer for the true SMM to converge to its final value than for the approximate SMM. The convergence time of the approximate SMM – four seconds in our setup – also implies that it can be performed many times over to test robustness, as we discuss below.

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 $f(\boldsymbol{\theta})$ only provide partial information about the link from targeted moments $\boldsymbol{m}$ to estimated parameters $\widehat{\boldsymbol{\theta}}$.

Because of the low computational cost of an approximate SMM, our methodology allows to fully trace out the mapping from targeted moments $\boldsymbol{m}$ to estimated parameters $\widehat{\boldsymbol{\theta}}$, even far from the empirical moments $\widehat{\boldsymbol{m}}$. This is a in contrast to Andrews et al. (2017a), who rely on a local linear approximation around $\widehat{\boldsymbol{m}}$. We illustrate this analysis in Figure 4, 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 22%, 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 4 reports two additional curves: (1) the yellow line reports the link between parameter values $\boldsymbol{\theta}$ and simulated moments $f(\boldsymbol{\theta})$ 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. (2017a).

Figure 4 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 ρ_z: for most values of ρ_z around its estimated value, the simulated value of m_6 is close to its empirical value. However, Figure 4 clearly shows that this interpretation is incorrect: Small variations in m_6 would leads to significant variations in the estimation of ρ_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 σ_z. To keep matching the volatility of 5-year sales growth, a higher persistence of TFP shocks ρ_z is required. If m_6 further declines, the upper bound for ρ_z, $\rho_z = 1$, 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, and returns to scale to try and match the low volatility of one-year sales log-growth. This leads to non-linear variations in these parameters' estimate. Figure 4 also shows that the local linear approximation used in Andrews et al. (2017a) can fail to hold for moment values in a small vicinity of their empirical values. For instance, if m_6 was 17% instead of 22% (its empirical value), the

local linear approximation of Andrews et al. (2017a) would suggest an estimated ρ_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., σ_z, the volatility of TFP shocks), the linear approximation works well over a wide range of moment values.

3.6 Robustness to Selected Moments

A weakness of methods of moments is that there is no well-established theory of moment selection. Similar models can be estimated using different set of targeted moments. The selection of targeted moments is thus essentially arbitrary. Assessing the robustness of parameter estimates to moment selection is thus a key aspect of transparency in structural estimation. This is easy conceptually: 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, this exercise cannot be done with standard estimation techniques, as it is computationally prohibitive.

In contrast, 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. Thus, when generating the training dataset, researchers can include as many moments as possible, and then easily test the robustness of parameter estimates to moment selection.

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 $\{m_8, m_9, \ldots, m_17\}$ *in addition* to the seven baseline moments $\{m_1, m_2, \ldots, m_10\}$ targeted in the baseline estimation of Table 2. Each panel in Figure 5 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. The shaded line plots the 95% confidence interval. Each coordinate on the x-axis corresponds to one of the 10 moments in $\{m_8, m_9, \ldots, m_17\}$. The y-axis reports the parameter estimate when the estimation targets the seven baseline moment *and* the additional moment on the x-axis. We also report standard errors for each estimated parameters. These standard errors are calculated using the delta method with the approximate

Jacobian matrix.

Figure 5 offers a clear diagnostic tool for robustness to moment selection. For instance, the estimation of the collateral constraint parameter (Panel λ) 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}), or the autoregression coefficient of the profit ratio (m_{13})). However, the estimated collateral constraint parameter becomes significantly larger when the estimation also targets the coefficient of a regression of investment on the market-to-book ratio (m_{11}), the variance of leverage (m_{15}) or the average dividend to asset ratio (m_{16}). In particular, including the variance of the leverage ratio almost triples the estimated λ.

More generally, we see in Figure 5 that the baseline estimation is in general not robust to including dividend moments. 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 tighter financing constraints, which biases the estimate of λ. Figure 5 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 of profits. All the estimated parameters remain unchanged whether or not we target these moments.

The exercise in Figure 5 explores a small set of alternative moment selection. Given the low computational cost of an approximate estimation, we can explore more systematic tests of robustness to moment selection. In Figure 6, we re-estimate the model by targeting (a) the seven moments used in our baseline estimation $\{m_1, m_2, \ldots, m_7\}$ *and* (b) any possible subset of the other 10 additional moments $\{m_8, m_9, \ldots, m_17\}$. This represents $2^{10} = 1,064$ 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 all estimated parameters is more than 10 times larger than the standard errors of the baseline true SMM.[6] This leaves us with 987 estimates. Figure 6 reports the distribution of estimates across the 987 alternative sets of moments. The dashed line corresponds to the 95% confidence interval for the baseline

[6]As before, we compute standard errors through the classic SMM formula $J'\Omega J$, therefore using as weight matrix the inverse of the variance-covariance matrix of empirical moments.

estimate. A baseline estimate robust to moment selection would have a large share of alternative estimates close to its value.

Clearly, most parameter estimates for this model are not robust to moment selection. For instance, the estimated variance of innovations to z, σ_z, is distributed almost uniformly from .5 to 1.2, with a benchmark estimate around 1.1. Alternative estimates for the collateral constraint parameter λ or the depreciation rate δ are similarly uniformly distributed. We also see that a large share of estimates for equity issuance cost (ζ) and adjustment cost (γ) 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. In contrast, the returns to scale parameters appear to be robust to moment selection.

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 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 7. 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 7, 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 λ 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. λ 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 go 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 five broad industries (manufacturing, retail trade, services, transportation, and mining). We find that, relative to manufacturing, the value loss from financial frictions are much higher in the mining and transportation industries. This finding reflects the much higher equity issuance costs in these industries relative to manufacturing.

Figure 7 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.

3.8 Model Misspecification

Our approach can also be used to explore directly model misspecification. One approach to misspecification explored in Andrews et al. (2017a) 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 4.1 to account for the fact that the debt constraint may not only depend on the capital stock, but also on the firm's EBITDA:

$$d_t < \lambda k_t + \lambda_2 . \mathrm{E}[e^{z_t(1-\alpha)}] k_t^{\alpha} \tag{10}$$

so that the firm can not only pledge collateral but also a multiple λ_2 of its EBITDA. For the initial seven parameters, we use the baseline estimates of Table 2. We consider 40 linearly-spaced values for λ_2, ranging from 0 (no misspecification) to 2 (large misspecification). For each of these values of λ_2, we solve the model and simulate the seven baseline moments ($\{m_1, \ldots, m_7\}$).

In a second step, we use the approximate SMM to estimate the model targeting each of the 40 different vectors of moments. Figure 8 shows the effect of misspecification on our estimates of the value loss from financing constraints. The x-axis represents the value of λ_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 λ_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 8. First, the estimates of value loss is robust to low values of cash-flow pledgeability ($\lambda_2 < .6$). Second, as λ_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 λ_2 affects productive firms disproportionately, an effect that the misspecified model misses. This error affects two different moments.

First, as λ_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 λ_2.

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)[7] and a realistic Social Security system in retirement.

Macroeconomic environment The stock market log return in year t is

$$s_t = s_{1,t} + s_{2,t}, \tag{11}$$

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

$$s_{1,t} = \begin{cases} s^-_{1,t} \sim \mathcal{N}(\mu^-_s, \sigma^2_{s_1}) & \text{with probability} \quad p_s \\ s^+_{1,t} \sim \mathcal{N}(\mu^+_s, \sigma^2_{s_1}) & \text{with probability} \quad 1 - p_s \end{cases} \tag{12}$$

[7] See Catherine et al. (2022a) for reduced-form evidence of the effect of countercyclical risk on households' portfolio.

On the other hand, $s_{2,t}$ is normally distributed with variance $\sigma_{s_2}^2$. We impose $\mu_s^- < \mu_s^+$ and interpret μ_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:

$$l_{1,t} - l_{1,t-1} = \mu_l + \lambda_{ls} s_{1,t} + \varepsilon_{l,t}, \tag{13}$$

where $\varepsilon_{l,t}$ follows $\mathcal{N}(0, \sigma_l^2)$, μ_l is the average growth rate, and λ_{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 $L_{2,it}$:

$$L_{it} = L_{1,t} \cdot L_{2,it}. \tag{14}$$

The idiosyncratic component is further decomposed into a deterministic function of age f_{it}, a persistent component z_{it} and a transitory shock η_{it}:

$$L_{2,it} = e^{f_{it} + z_{it} + \eta_{it}}. \tag{15}$$

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

$$z_{it} = \rho_z z_{it-1} + \zeta_{it}, \tag{16}$$

where

$$\zeta_{it} = \begin{cases} \zeta_{it}^- \sim \mathcal{N}\left(\mu_{z,t}^-, \sigma_z^{-2}\right) & \text{with probability} \quad p_z \\ \zeta_{it}^+ \sim \mathcal{N}\left(\mu_{z,t}^+, \sigma_z^{+2}\right) & \text{with probability} \quad 1 - p_z \end{cases} \tag{17}$$

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 cyclicality of skewness, $\mu_{z,t}^-$ is an affine function of the log growth rate of the wage index:

$$\mu_{z,t}^- = \overline{\mu_z^-} + \lambda_{zl}(l_{1,t} - l_{1,t-1}). \tag{18}$$

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.

$$\eta_{it} = \begin{cases} \eta_{it}^{-} \sim \mathcal{N}(0, \sigma_{\eta}^{-2}) & \text{if} \quad \zeta_{it} = \zeta_{it}^{-} \\ \eta_{it}^{+} \sim \mathcal{N}(0, \sigma_{\eta}^{+2}) & \text{if} \quad \zeta_{it} = \zeta_{it}^{+} \end{cases} \tag{19}$$

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

$$\mathrm{T}_{it} = .124 \cdot \min\left\{L_{it}, 2.5 \cdot L_{1,t}\right\}. \tag{20}$$

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:

$$\frac{B_i}{L_{1,R}} = \begin{cases} .9 \cdot S_{iR} & \text{if } S_{iR} < .2 \\ .116 + .32 \cdot S_{iR} & \text{if } .2 \leq S_{iR} < 1 \\ .286 + .15 \cdot S_{iR} & \text{if } 1 \leq S_{iR}, \end{cases} \tag{21}$$

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:

$$S_{it} = \sum_{k=t_0}^{t} \frac{\min\left\{L_{2,ik}, 2.5\right\}}{t - t_0 + 1}, \tag{22}$$

where t_0 denotes his first year of earnings.

Agent The agent maximizes lifetime expected utility:

$$V_{t_0} = \mathrm{E} \sum_{t=t_0}^{T} \beta^{t-1} \left(\prod_{k=0}^{t-1} (1 - m_k) \right) \frac{C_{it}^{1-\gamma}}{1-\gamma}, \tag{23}$$

where γ is the coefficient of relative risk-aversion, m_k the mortality rate at age k, β 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:

$$W_{it+1} = \left[W_{it} + L_{it} + B_{it} - \mathrm{T}_{it} - C_{it} - c_{it}\right] \cdot \left[\pi_{it} e^{s_t} + (1 - \pi_{it}) e^{r}\right], \tag{24}$$

where π_{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 $0 < \pi_{it} < 1$.

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 estimation

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

These 2,000 parameter draws result in a training dataset $\mathcal{D} = (\boldsymbol{\theta}_i, \boldsymbol{m}_i)_{i\in[1,S]}$, which we use to estimate $g()$, the approximate relationship between moments and parameters $\boldsymbol{m}_i = f(\boldsymbol{\theta}_i) \approx g(\boldsymbol{\theta}_i, \boldsymbol{\beta})$. 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's performance.

Figure 9 is the analog of Figure 1 for this household finance model. The figure illustrates the out-of-sample performance of the approximate SMM approach by comparing, on a validation sample, the estimated parameters to the true parameters. Like in the corporate finance application, the third-order polynomial approximation with $k = 2$ works best: $1 - R^2$ is smaller than 1% for the three estimated parameters. Unlike the corporate finance application, Figure 9 also shows that other approximations like an unweighted NN or a weighted linear model are precise. In what follows, we use the third-order polynomial approximation with $k = 2$ as our benchmark approximation.

Figure 10 is the analog of Figure 2. For each draw of the validation sample, it shows the true parameters $\boldsymbol{\theta}$ on the x-axis, and, on the y-axis, $\widehat{\boldsymbol{\theta}}$, the parameters estimated using the benchmark approximate SMM targeting $f(\boldsymbol{\theta})$, the moments generated by the true model with parameters $\boldsymbol{\theta}$. Figure 10 highlights the precision of the approximate SMM. The out-of-sample R^2 – the share of the variance of parameters in the validation sample explained by the estimated parameters – is larger than 99.9% for all three parameters γ, β and Φ. 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.[8]

Here again, the approximate SMM improves estimation speed by at least two orders of magnitudes. With our numerical setup, solving the agent's Bellman equation a single time takes 20 seconds whereas running a full approximate SMM only takes less than a second.

4.3 Data

We compute 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. We use three baseline moments to estimate the model (the equivalent of $(m_i)_{i\in\{1,..,7\}}$ in the corporate finance application): (a) m_1 is the mean wealth, which 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 (b) m_2 is the average participation rate, which is the share of households whose total holdings of stock (*equity*) is strictly positive (c) 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.

Panel A of Table 4 reports the true and approximate SMM estimations targeting these three moments. Both approaches give similar values for risk-aversion γ (about 8.4) and the discount factor β (some .91). For both parameters, estimates differ by about 1%, which is much lower than the estimates' standard error. The estimate of Φ, the participation cost, is however significantly lower in the approximate SMM, by about 15%: while Φ is estimated at.0053 in the true SMM, the approximate SMM provides an estimate of .0045. Figure 11 shows the speed of convergence of the true SMM and the approximate SMM when both target these empirical moments. With our numerical setup, the approximate SMM converge in about .2 seconds, while the true SMM takes between 15 minute to an hour to converge to its final value.

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,

[8]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 object – they are simulation-based.

whether we use the true or the approximate SMM.

4.4 Identification Diagnostic

Figure 12 is the analog of Figure 4. 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. (2017a) can be rapidly imprecise for some parameters. The global exercise to identification – made possible by the low estimation cost of the approximate SMM – is thus a useful tool to

Panel A shows how estimated parameters vary with mean wealth. Targeting a higher level of wealth results in a greater estimated discount factor β. 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 contrast, 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). When wealth increases, the model predicts a higher equity share; to keep matching the equity share, risk-aversion has to increase. 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 12 also makes clear that linear approximation in Andrews et al. (2017a) is sometime imprecise, even for moment value 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. (2017a) would suggest an estimated

risk-aversion of about 6, while its estimated value would in fact be close 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 ($m7$). 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 12 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-cycle profile of equity shares (m_{10})[9] 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 $\{m_4, \ldots, m_{10}\}$ in addition to the baseline moments $\{m_1, m_2, m_3\}$. Figure 13 reports our first robustness check, which is the equivalent of Figure 6 above: We target only one additional moment in $\{m_4, m_5, m_8, m_9, m_{10}\}$ and report the resulting parameter estimates.[10] 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 13 makes

[9]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.

[10]We omit m_6 and m_7 from this exercise since, conditional on targeting the mean participation rate, targeting the mean conditional or mean unconditional equity shares is equivalent.

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.

We also consider a more systematic exploration of robustness to moment selection that mirrors the exercise in Figure 6. We leverage the speed of the approximate SMM to re-estimate the model using 72 alternative sets of moments:

- 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_7 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_6). 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_6 (median conditional equity share using the alternative empirical definition). This adds another 24 sets of moments.

Figure 14 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.[11] The estimates of β, the discount factor, are clustered between .9 and .98, and the mode of the distribution is close to the baseline estimate (about .93 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.

[11] Note that this mode of 6-7 is consistent with Catherine (2021)'s estimates.

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 flexible relation between model parameters and moments. Once this relation has been estimated, it can be used to estimate structural parameters for any given set of moments. 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. Using model-based moments, we also show that, as the long as the model is identified, the approximate estimation is precise: its estimates are close to the true estimates. As a result, this approximate SMM is a useful tool for running robustness checks and identification diagnostics that require many alternative estimations.

We consider three such exercise. The first one assesses the robustness of parameter estimates to moment selection. Because our approximate SMM carries low computational cost, we can easily re-estimate the model using many different sets of targeted moments. This allows researchers to evaluate the influence of particular moments, or set of moments on parameter estimates, and thus serve as a useful identification diagnostic. The second one is sample-splits. Thanks to the low computational cost of the approximate SMM, we can easily re-estimate the model by targeting moments defined on different sub-samples (e.g., across different years or industries). This exercise allows researchers to evaluate how estimated parameters depend on the particular sub-sample used in estimation. If they do, the researcher can assess whether the variations in parameter estimates across the different samples conform to the economics of the underlying model. Finally, we use the approximate SMM to gauge the sensitivity of the baseline estimates to misspecification error. To do so, we generate a large set of new moments using alternative models and consider how baseline estimates would differ if the baseline model was used to estimate parameters against these alternative moments. Again, this exercise can be done at a low computational cost using the approximate SMM. We think these robustness checks are useful diagnostics to explore identification, misspecification and robustness of structural estimates. They are easy to implement for researchers seeking to make structural estimation more transparent.

References

Andrews, Isaiah, Matt Gentzkow, and Jesse Shapiro, "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *Quarterly Journal of Economics*, 2017, *132* (4).

_ , 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 _ , "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, St'ephane 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.

Hennessy, Christopher and Toni Whited, "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.

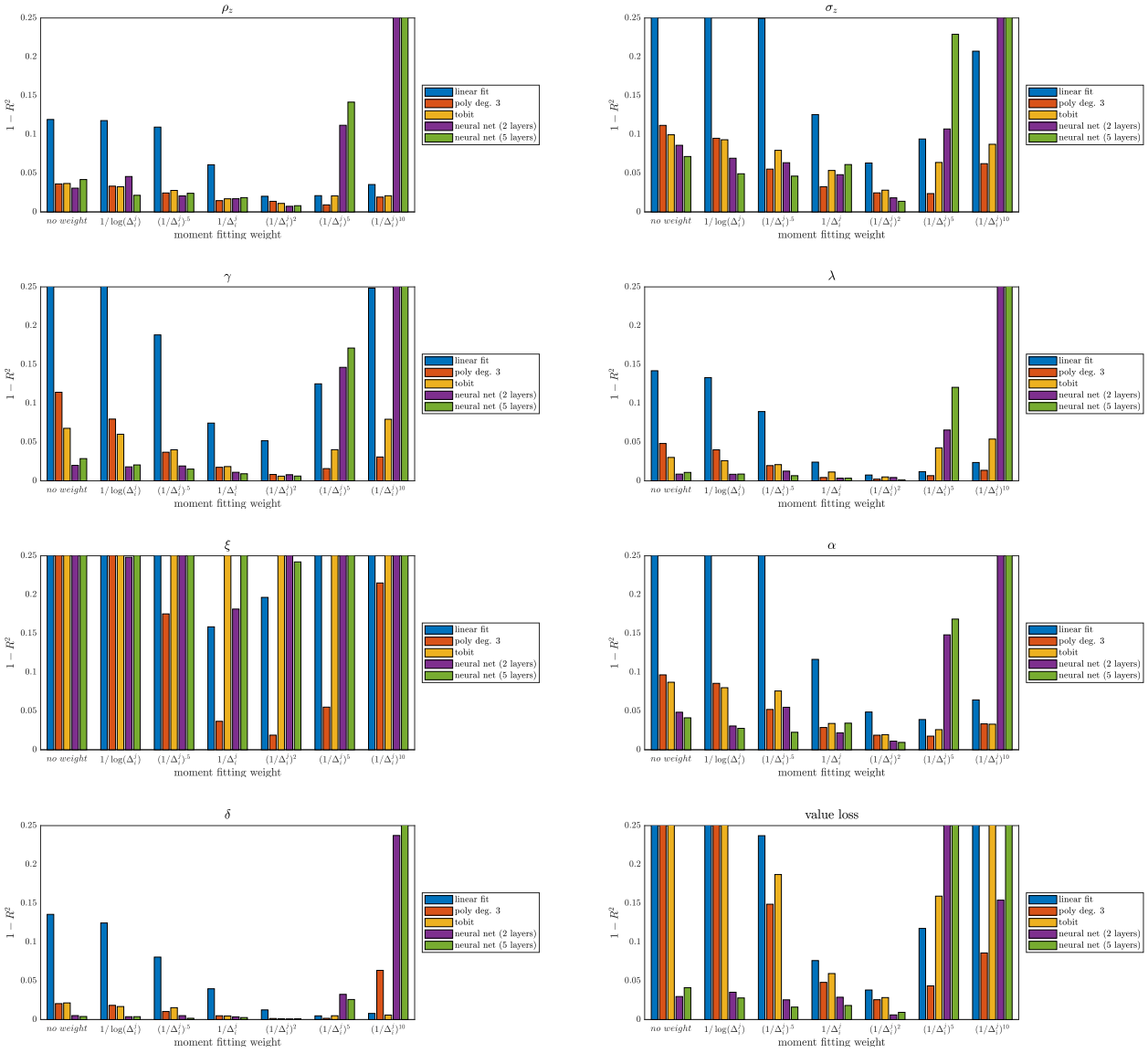
Strebulaev, Ilya A. and Toni M. Whited, "Dynamic Models and Structural Estimation in Corporate Finance," *Foundations and Trends(R) in Finance*, November 2012, *6* (1–2), 1–163.

Viceira, Luis M., "Optimal Portfolio Choice for Long-Horizon Investors with Nontradable 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.

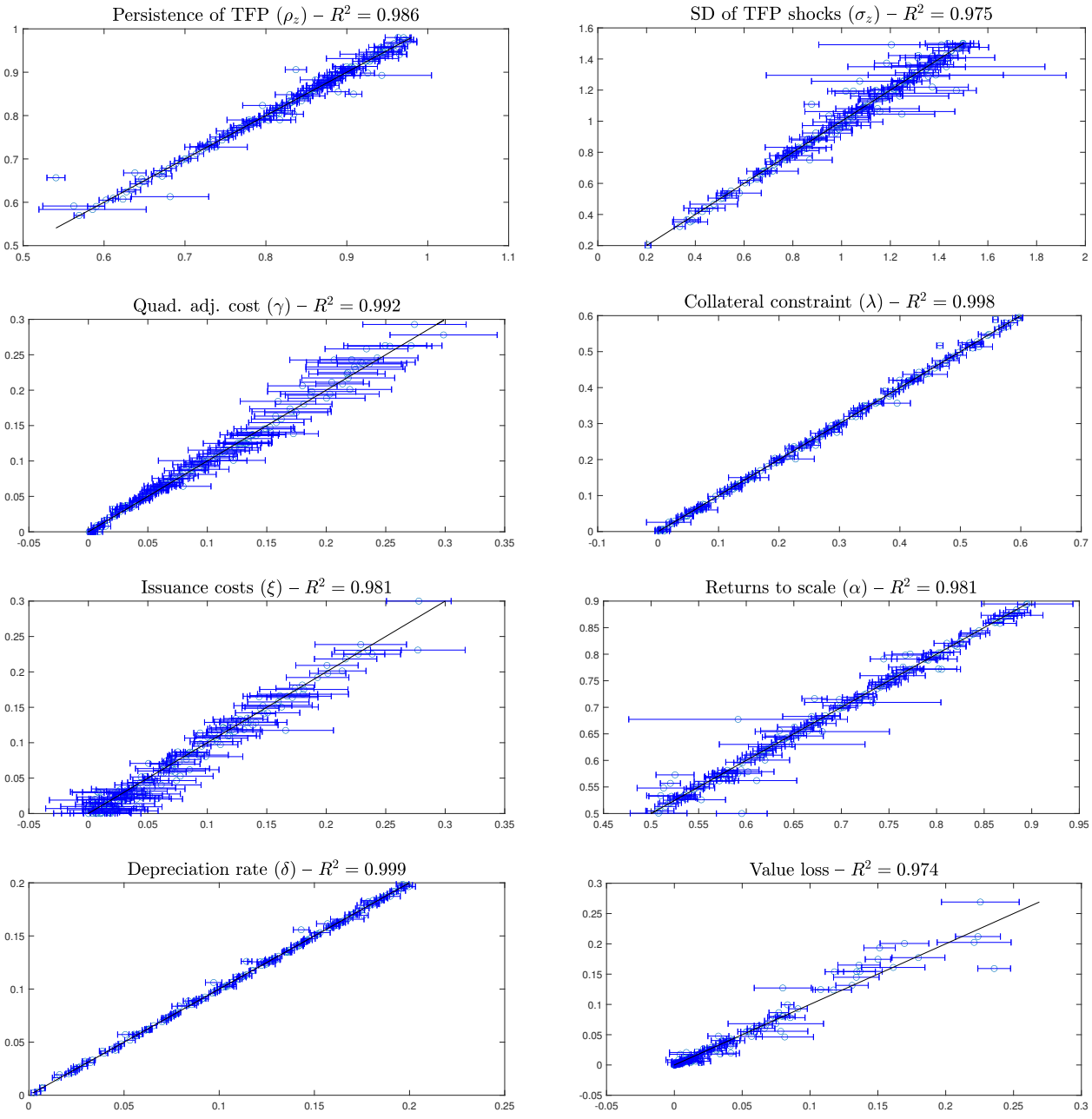
Figures

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



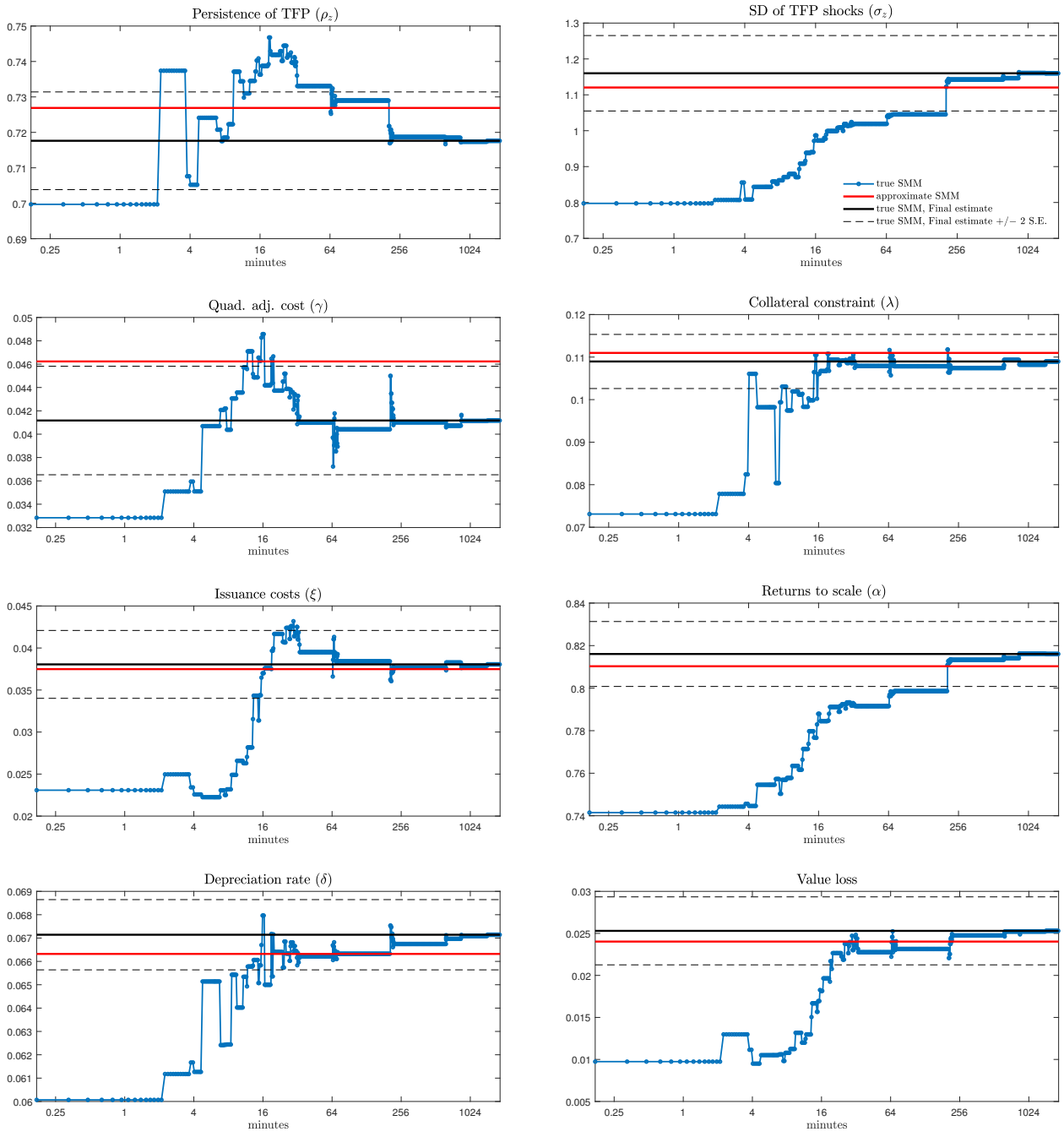
Notes. This figure reports a measure of the estimation error from the approximate SMM for different approximations and weighting schemes. For each draw $(\boldsymbol{\theta}_j^{\text{validation}}, f(\boldsymbol{\theta}_j^{\text{validation}}))$ in the validation sample, we estimate parameters $\widehat{\boldsymbol{\theta}_j^{\text{validation}}}$ with an approximate SMM targeting moments $f(\boldsymbol{\theta}_j^{\text{validation}})$. For each parameter i and approximation used, the figure reports $1 - R^2(i) = \frac{\mathbb{V}\text{ar}\left[\boldsymbol{\theta}_j^{\text{validation}}(i) - \widehat{\boldsymbol{\theta}_j^{\text{validation}}}(i)\right]}{\mathbb{V}\text{ar}\left[\boldsymbol{\theta}_j^{\text{validation}}(i)\right]}$ 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 two layers, and a neural net with 5 layers. The approximation $g(\boldsymbol{\theta}; \boldsymbol{\beta})$ is estimated on the training dataset $\mathcal{D}$ as $\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left[\sum_{l \in \mathcal{D}} \frac{1}{(\Delta_l^j)^k} \left(g(\boldsymbol{\theta}_l; \boldsymbol{\beta}) - f(\boldsymbol{\theta}_l)\right)' \left(g(\boldsymbol{\theta}_l; \boldsymbol{\beta}) - f(\boldsymbol{\theta}_l)\right)\right]$ where $\Delta_l^j = (\boldsymbol{m}_l - \boldsymbol{m}_j^{\text{validation}})' \Omega (\boldsymbol{m}_l - \boldsymbol{m}_j^{\text{validation}})$ and Ω 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 2: Out-of-sample Performance (Corporate Finance Model)



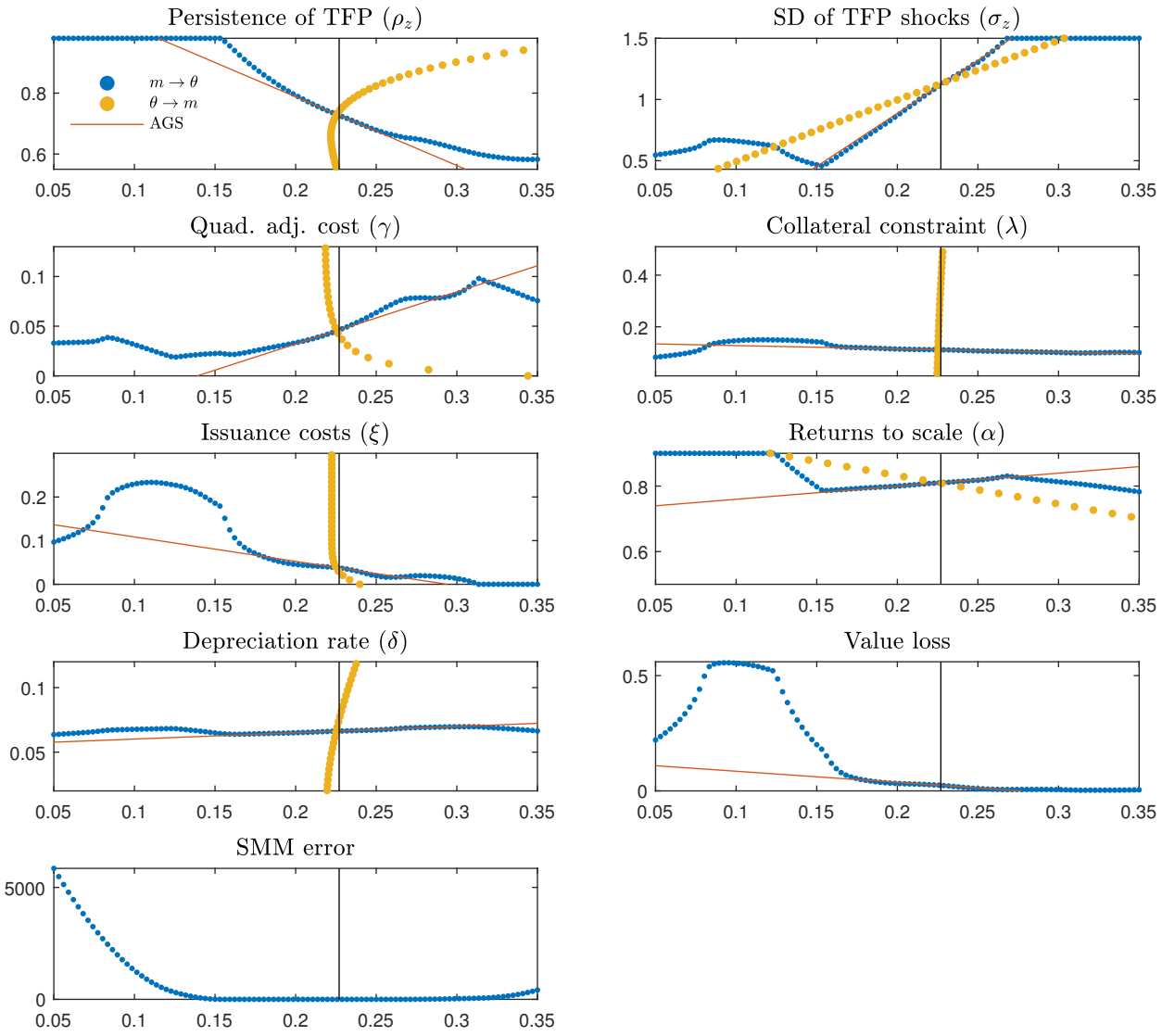
Notes. This figure shows the precision of our benchmark approximate SMM across estimated parameters. For each draw $(\boldsymbol{\theta}_j^{\text{validation}}, f(\boldsymbol{\theta}_j^{\text{validation}}))$ in the validation sample, we estimate parameters $\widehat{\boldsymbol{\theta}_j^{\text{validation}}}$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

Figure 3: True SMM Estimates, Convergence Speed (Corporate Finance Model)



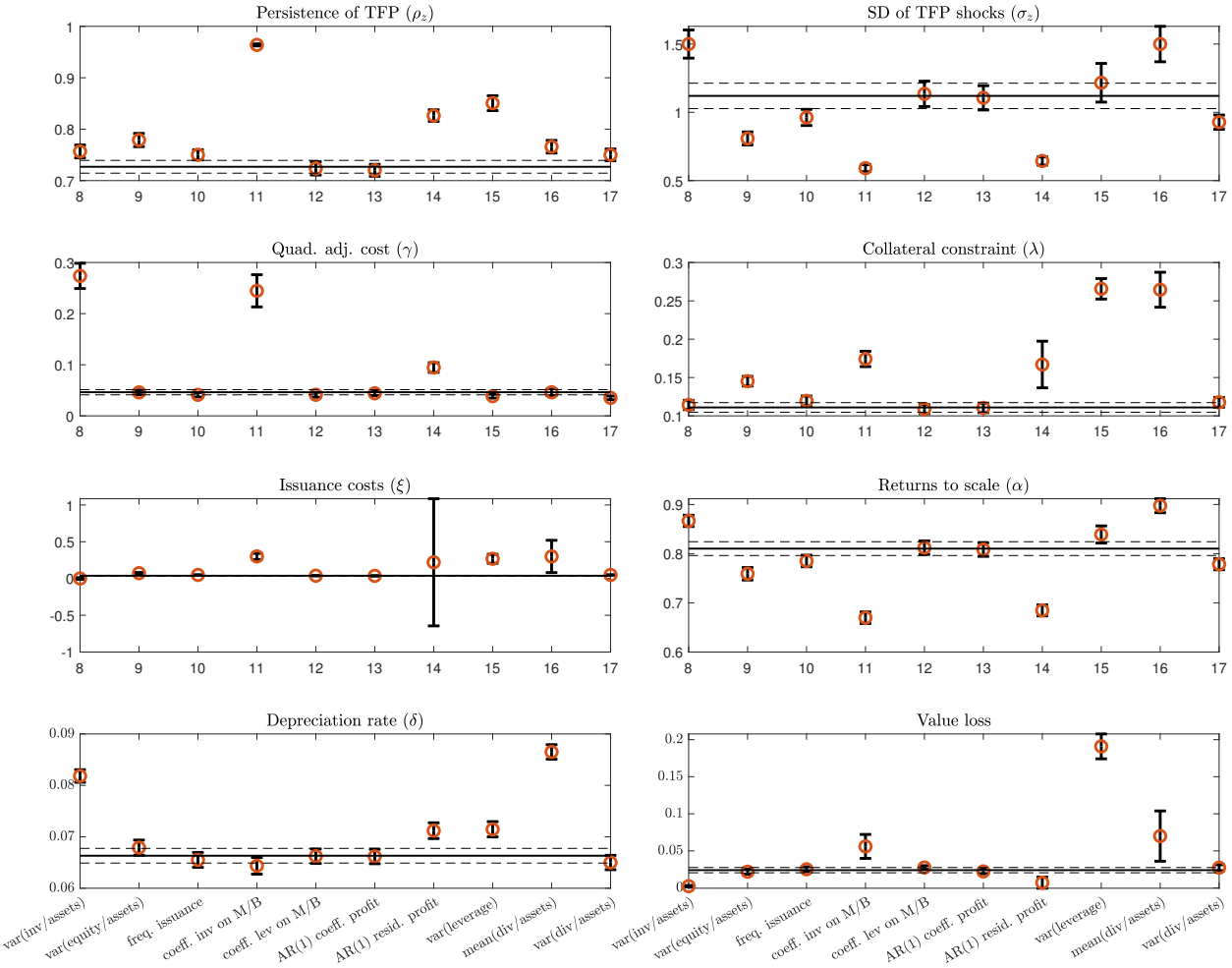
Notes. We report parameter values estimated as a function of time taken via two different algorithms in our numerical setup. This estimation corresponds to the baseline estimation described in Section 3.4. The blue line corresponds to the true SMM, i.e. the minimization of the distance of empirical moments 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 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 1.1 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.

Figure 4: Sensitivity of parameters to the volatility of sales growth



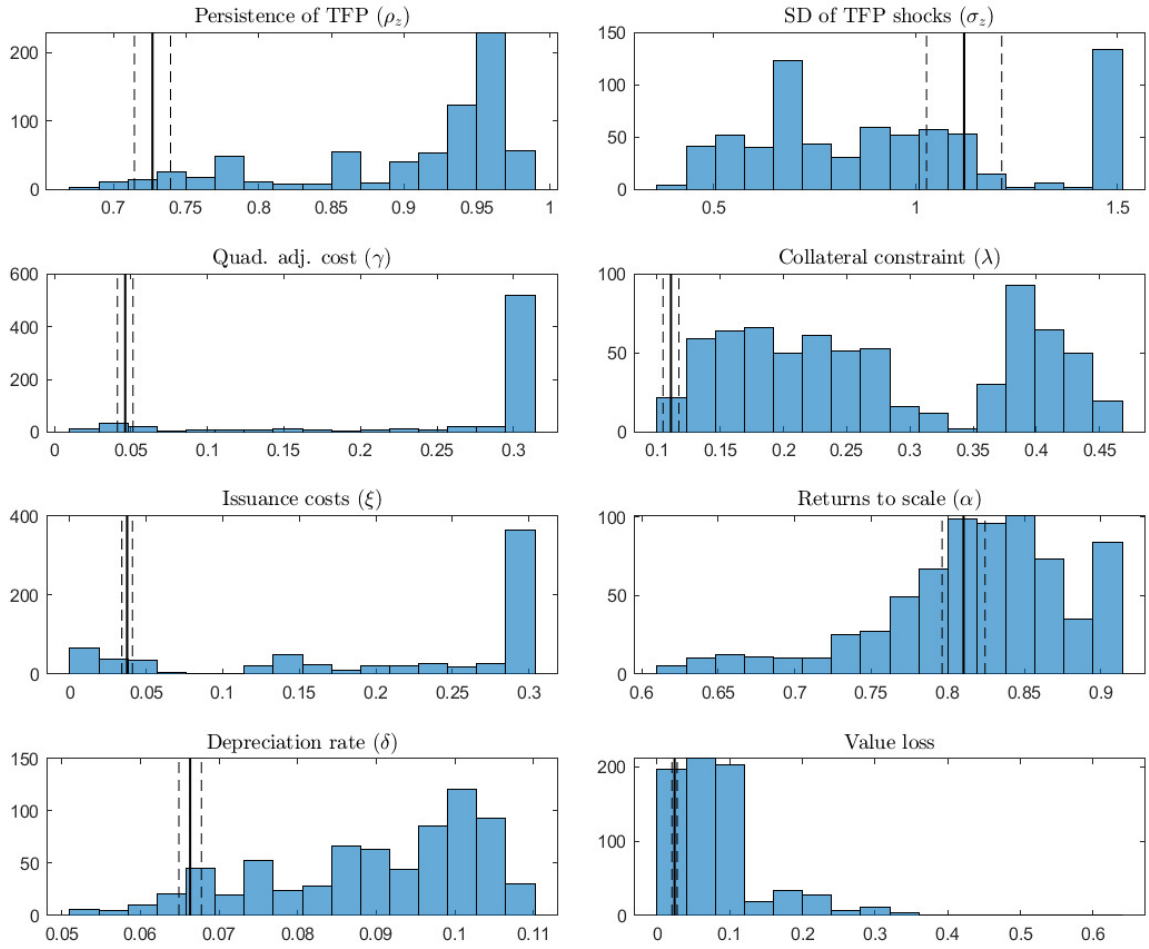
Notes. This figure plots parameters values on the y-axis and the value of the moment m_6 on the x-axis. m_6 is the standard deviation of the one year log sales growth. The yellow line draws local comparative statics, i.e. how variations in one parameter around its estimated value – holding other parameters 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. (2017a): it is a linear approximation of the mapping from moments to parameter estimates around the empirical value of the moments m_6.

Figure 5: Robustness to adding moments one by one (Corporate Finance Model)



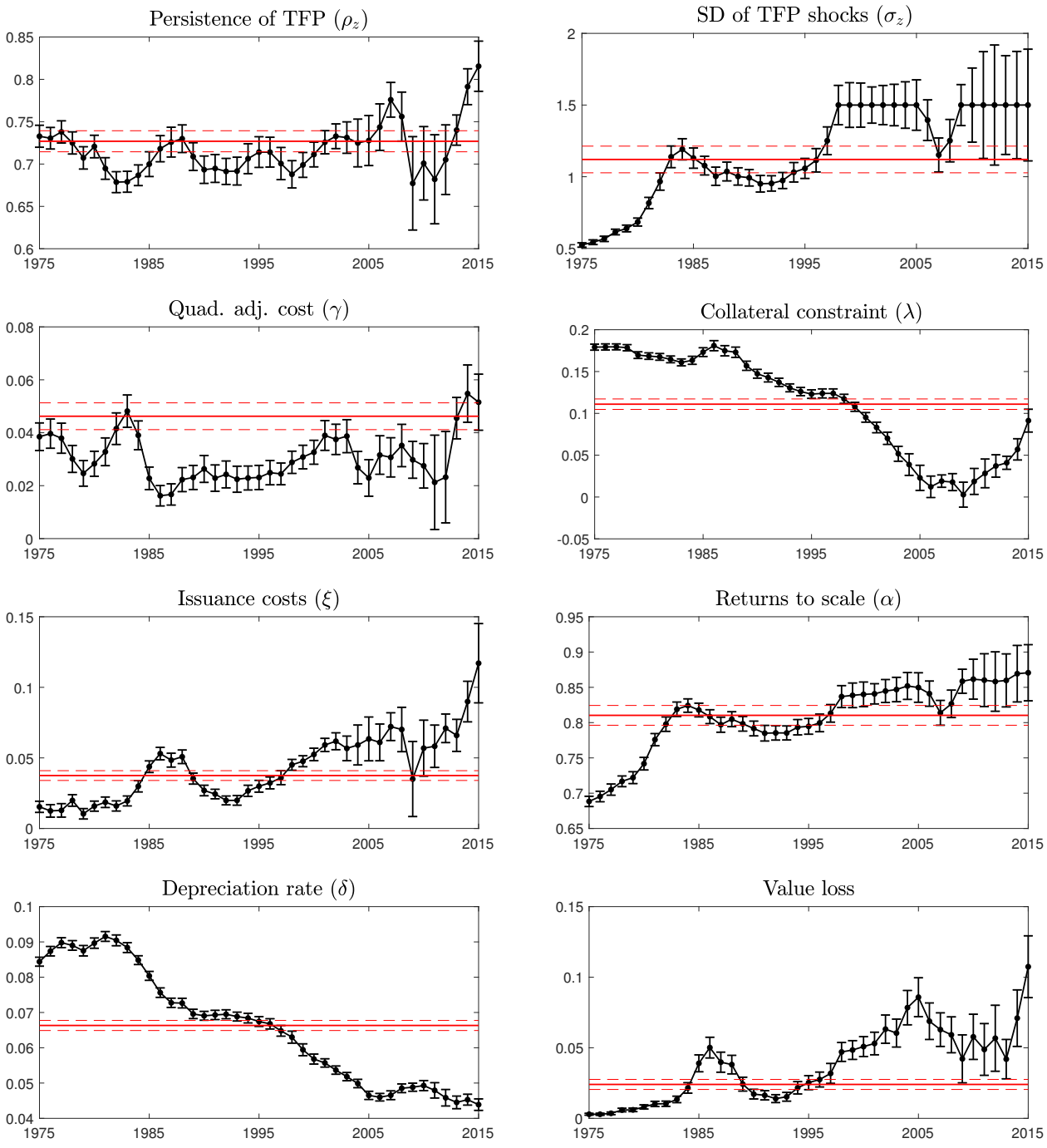
Notes. This figure explores the sensitivity of parameter estimates to the inclusion of an additional moment in the set of targeted moments. Our baseline estimation targets seven moments $(m_i)_{i\in\{1..7\}}$. We consider a set of 10 additional moments used in the literature to estimate similar models: $(m_i)_{i\in\{8..17\}}$ 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 $(J'WJ)^{-1}$, where W is the SMM weight matrix and J the approximate Jacobian matrix computed at the parameter estimates. We use the delta method to calculate SE of value loss.

Figure 6: Histogram of estimates across 1,024 sets of targeted moments (Corporate Finance Model)



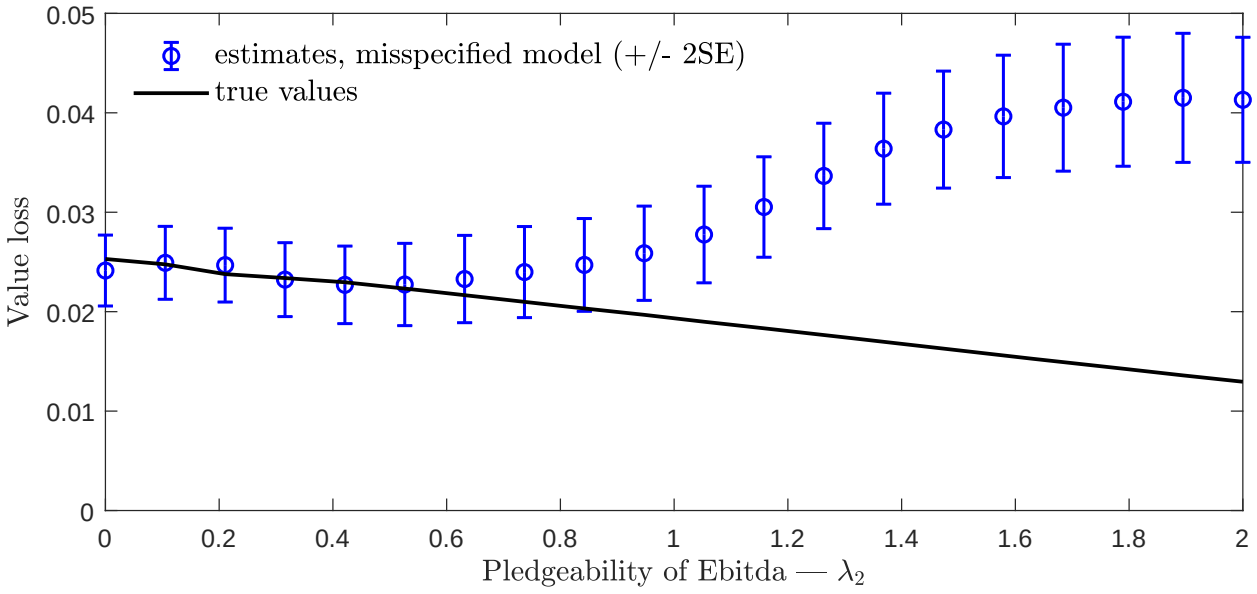
Notes. This figure explores the sensitivity of parameter estimates to moment selection. Our baseline estimation targets seven moments $(m_i)_{i\in\{1..7\}}$. We consider a set of 10 additional moments used in the literature to estimate similar models: $(m_i)_{i\in\{8..17\}}$ described in Section 3.2. 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 1,064 sets of parameter estimates. After dropping cases where estimates are poorly-identified – where the standard errors for all estimated parameters is more than 10 times larger than the standard errors of the baseline true SMM – we end up with 987 estimates. Each panel in the figure shows the distribution of parameters across these 987 estimations. The vertical black line and dashed lines show the baseline parameter estimates using the true SMM, together with their 95% confidence interval.

Figure 7: Time Series Estimates (Corporate Finance Model)



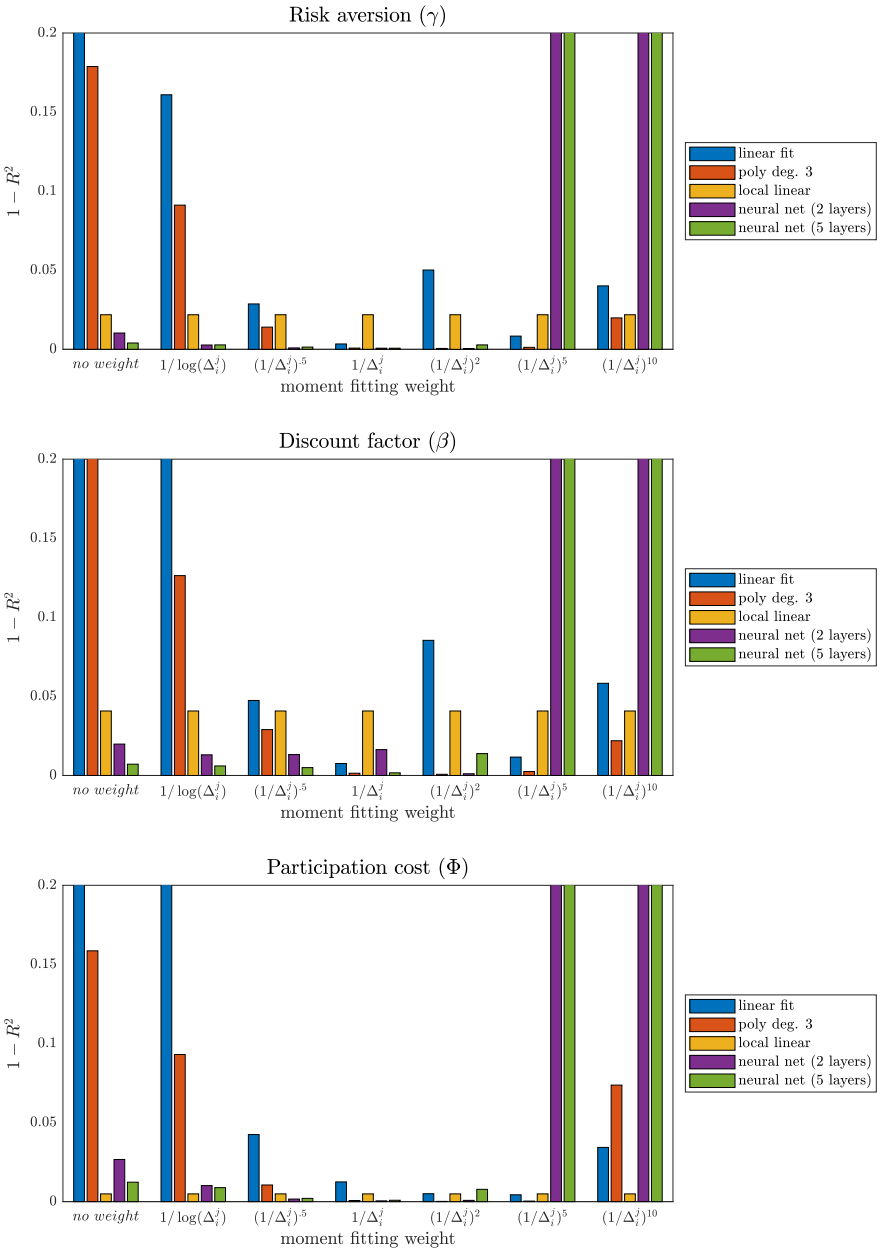
Notes. This figure shows the sensitivity of parameter estimates to the sample period used to compute the targeted moment. For each year t on the x-axis, we compute the seven baseline moments $(m_i)_{i\in\{1..7\}}$ on the sample period [t-5,t+4]. We then re-estimate the model using the benchmark approximate 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.

Figure 8: Model Mispecification: Estimating Model with Collateral Constraints on Data Generated by Cash-flow Constraints (Corporate Finance Model)



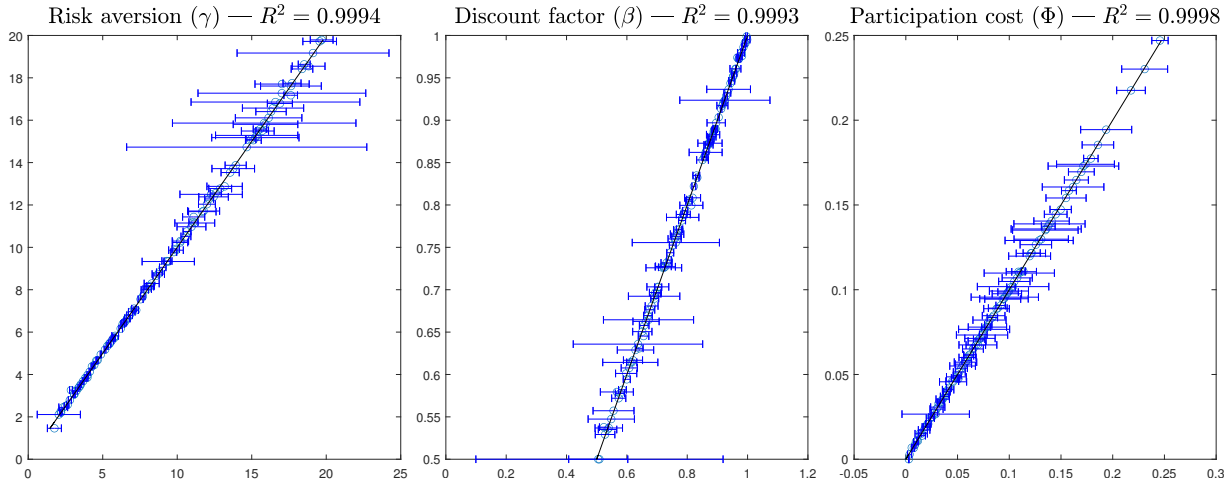
Notes. This figure explores the sensitivity of estimation to model misspecification. We consider an augmented version of our corporate finance model where firms can pledge a multiple λ_2 of their EBITDA so that their debt constraint takes the form $d_t < \lambda k_t + \lambda_2 . \mathrm{E}[e^{z_t(1-\alpha)}] k_t^{\alpha}$. 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 λ_2 takes various values from 0 (the baseline model) to 2. For each possible value of λ_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 λ_2 used in each estimation. The blue circles report the value loss from financial constraint estimated with the misspecified model estimated by targets moments generated by the correctly-specified model. The black line reports the value loss from financial constraint in the correctly-specified model.

Figure 9: Out-of-sample Performance across models and weight schemes (Household Finance Model)



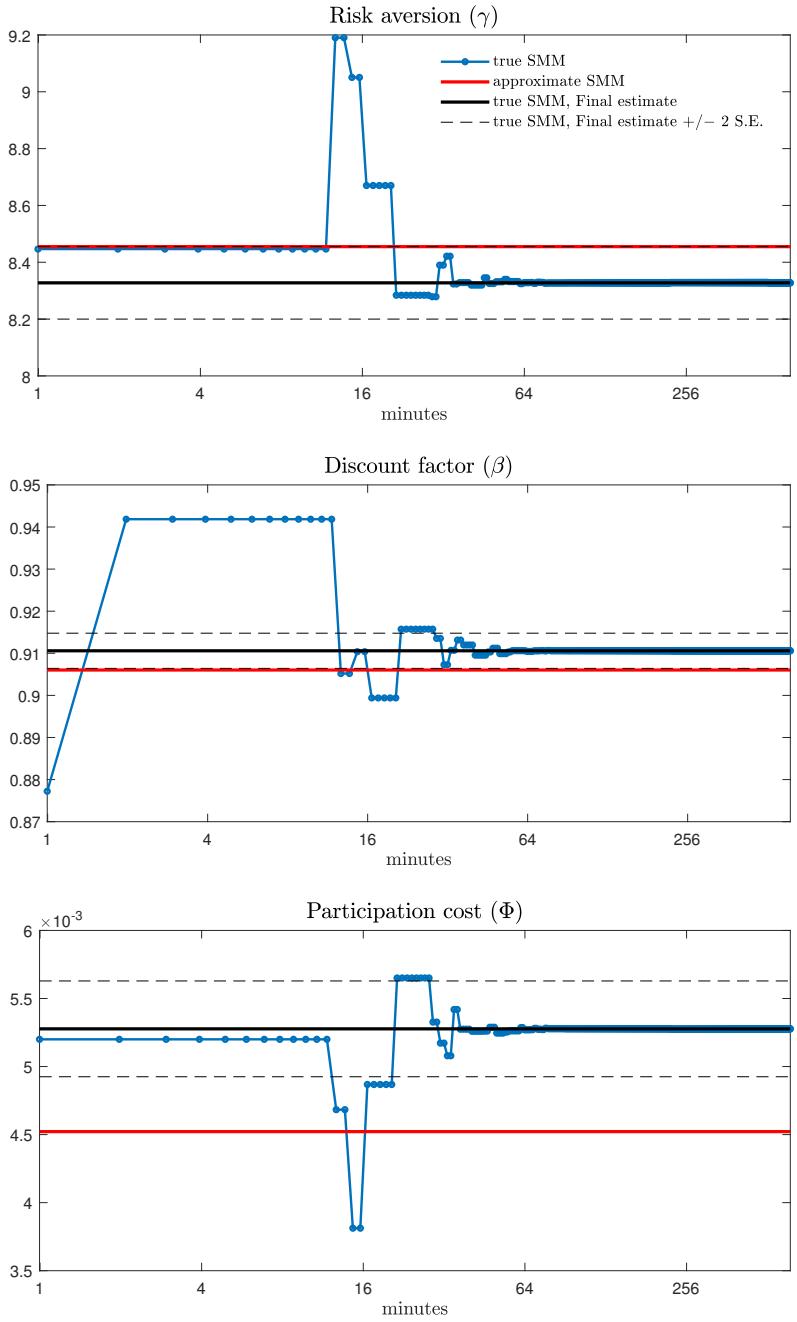
Notes. This figure reports, for the household finance model, a measure of the estimation error from the approximate SMM for different approximations and weighting schemes. For each draw $(\boldsymbol{\theta}_j^{\text{validation}}, f(\boldsymbol{\theta}_j^{\text{validation}}))$ in the validation sample, we estimate parameters $\widehat{\boldsymbol{\theta}_j^{\text{validation}}}$ with an approximate SMM targeting moments $f(\boldsymbol{\theta}_j^{\text{validation}})$. For each parameter i and approximation used, the figure reports $1 - R^2(i) = \frac{\mathbb{V}\text{ar}\left[\boldsymbol{\theta}_j^{\text{validation}}(i) - \widehat{\boldsymbol{\theta}_j^{\text{validation}}}(i)\right]}{\mathbb{V}\text{ar}\left[\boldsymbol{\theta}_j^{\text{validation}}(i)\right]}$ 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 two layers, and a neural net with 5 layers. The approximation $g(\boldsymbol{\theta}; \boldsymbol{\beta})$ is estimated on the training dataset $\mathcal{D}$ as $\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left[\sum_{l \in \mathcal{D}} \frac{1}{(\Delta_l^j)^k} \left(g(\boldsymbol{\theta}_l; \boldsymbol{\beta}) - f(\boldsymbol{\theta}_l)\right)' \left(g(\boldsymbol{\theta}_l; \boldsymbol{\beta}) - f(\boldsymbol{\theta}_l)\right)\right]$ where $\Delta_l^j = (\boldsymbol{m}_l - \boldsymbol{m}_j^{\text{validation}})' \Omega (\boldsymbol{m}_l - \boldsymbol{m}_j^{\text{validation}})$ and Ω 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 10: Performance of Estimation using Benchmark Approximation (Household Finance Model)



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

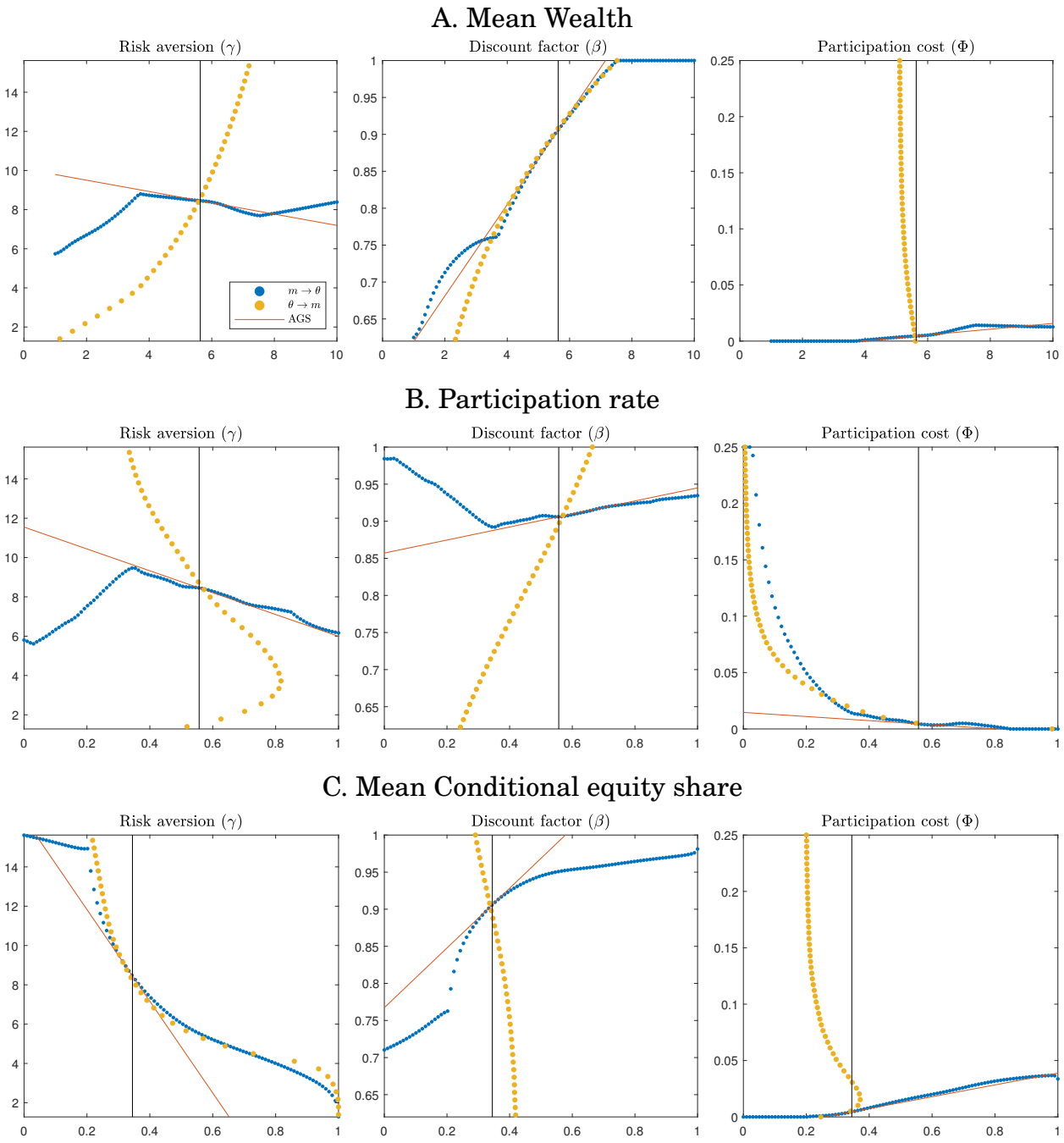
Figure 11: True SMM Estimates, Convergence Speed (Household Finance Model)



Notes.

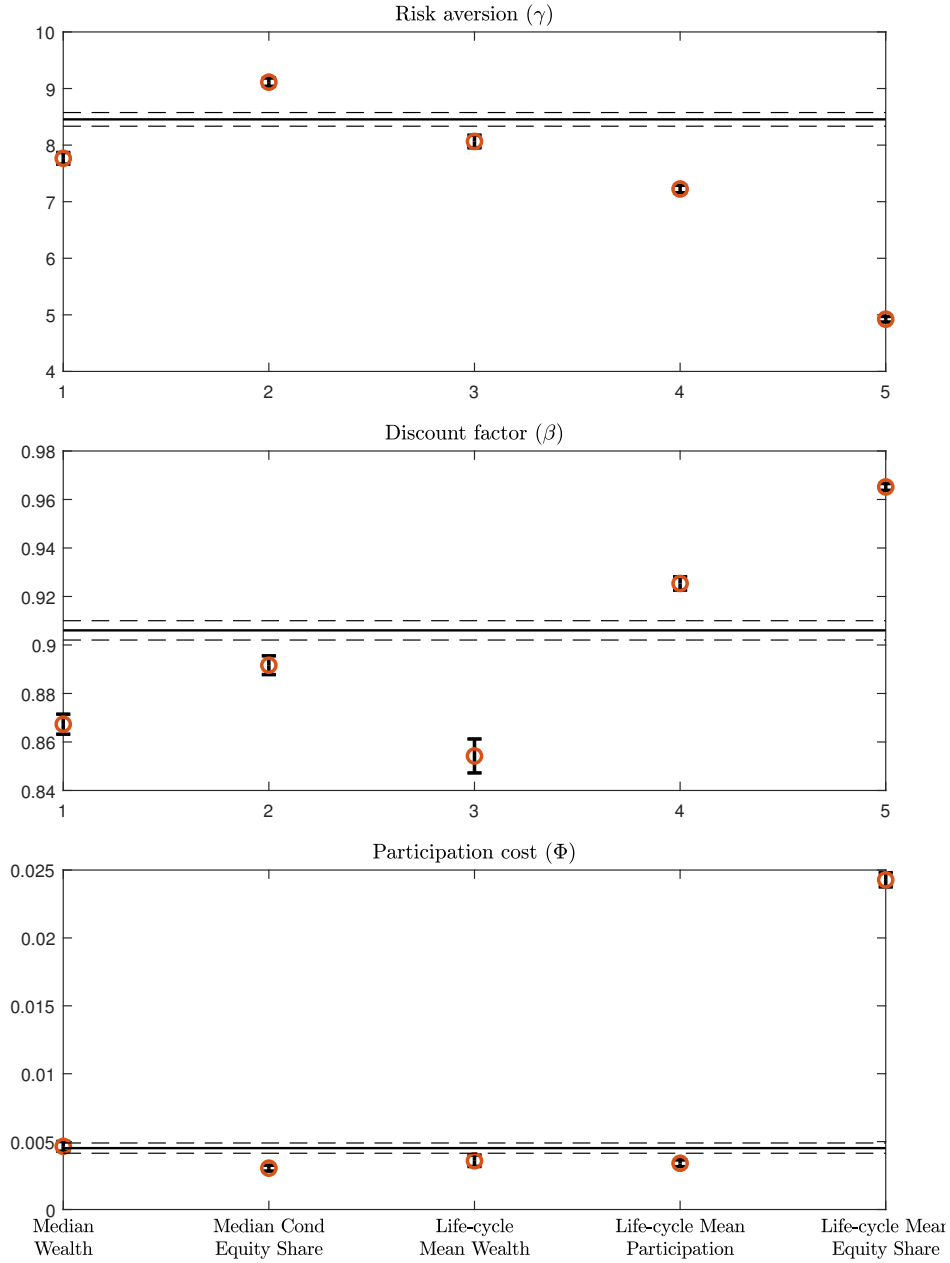
We report, for the household finance model, parameter values estimated as a function of time taken via two different algorithms in our numerical setup. This estimation corresponds to the baseline estimation described in Section 4.2. The blue line corresponds to the true SMM, i.e. the minimization of 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.

Figure 12: Sensitivity of Parameters to Moments (Household Finance Model)



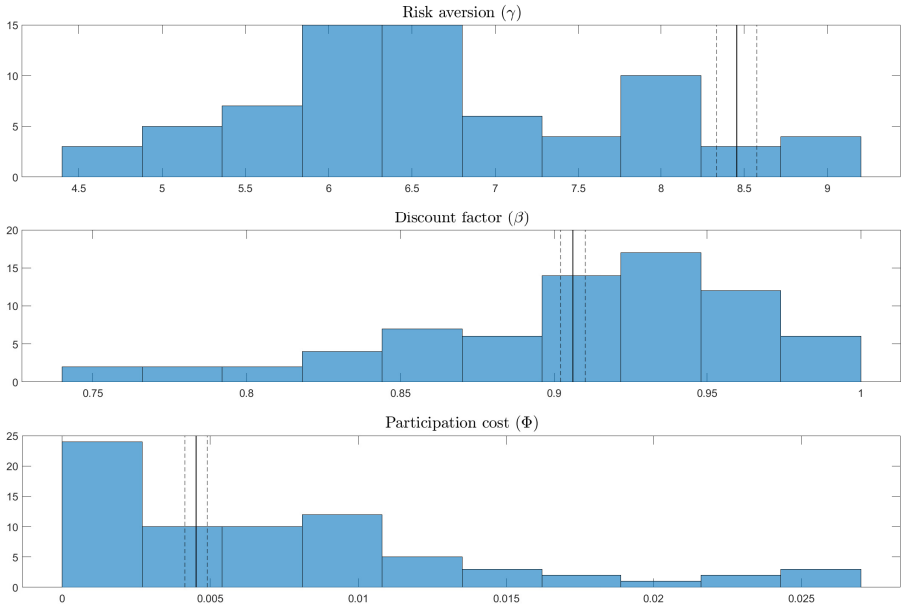
Notes. This figure plots parameters values on the y-axis and the value of the moment on the x-axis. Panel A (resp. B and C) corresponds to m_1 (mean wealth) (resp m_2 (participation rate) and m_3 (conditional equity share)). The yellow line draws local comparative statics, i.e. how variations in one 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_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. (2017a): it is a linear approximation of the mapping from moments to parameter estimates around the empirical value of the moments m_i.

Figure 13: 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 $(m_i)_{i\in\{1..7\}}$ *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 $(J'WJ)^{-1}$, where W is the SMM weight matrix and J the approximate Jacobian matrix computed at the parameter estimates. We use the delta method to calculate SE of value loss.

Figure 14: Histogram of estimates across 72 sets of targeted moments (Household Finance Model)



Notes. This figure shows the histogram of parameter estimates via approximate SMM that targets the 72 sets of moments described in Section 4.5. These sets are combinations of variants of the baseline moments (overall average, median, life-cycle, changing normalization). We only keep locally identified models: models with invertible matrix $J'\Omega J$ where Ω is the inverse of data moment variance matrix and J is derivative of moments wrt parameters, and models with standard error of all estimates being at most 10 times larger than the standard error of estimates in the benchmark estimation (i.e. with only 3 targeted moments). These two criteria leave us with the entire 72 models out of the total 72 different sets of moments. Blue charts show the histogram of estimates. The vertical black line and dashed lines show estimates plus/minus two standard errors in the benchmark estimation.

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). 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.

Tables:

Table 1: Simulation Moments (Corporate Finance Model)

	Data	True SMM	Approximate	Simulation
mean(investment/assets)	.0760 (.0007)	.0761	.0760	.0747
mean(profit/assets)	.1343 (.0012)	.1343	.1343	.1342
mean(equity issuance/assets)	.0158 (.0007)	.0158	.0158	.0150
mean(leverage)	.1049 (.0030)	.1051	.1049	.1080
autocorr(investment/assets)	.3754 (.0067)	.3753	.3755	.3907
std(log sales growth)	.2270 (.0017)	.2270	.2270	.2252
std(log sales growth 5yr)	.5851 (.0052)	.5854	.5851	.5800
var(investment/assets)	.0033 (.0001)	.0167	.0170	.0156
var(equity issuance/assets)	.0071 (.0002)	.0024	.0021	.0021
frequency(equity issuance)	.1178 (.0015)	.1521	.1605	.1516
coeff. regr. investment ratio on market to book ratio	.0122 (.0005)	.3190	.3039	.3081
coeff. regr. net leverage on market to book ratio	-0.0348 (.0018)	-0.0325	-0.0490	-0.0246
coeff. AR(1) regr. of profit ratio	.5210 (.0057)	.5358	.5405	.5445
resid std AR(1) regr. of profit ratio	.0728 (.0006)	.0287	.0283	.0285
var(leverage)	.0266 (.0004)	.0002	.0004	.0001
mean(dividend/assets)	-0.0031 (.0008)	.0491	.0491	.0492
var(dividend/assets)	.0073 (.0002)	.0054	.0052	.0050

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

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

	ρ_z	σ_z	γ	λ	ξ	α	δ	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	

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 estimated using the benchmark approximate SMM (third-order polynomial with $k = 2$). ρ_z is the persistence of the productivity process. σ_z is standard deviation of innovations to productivity. γ is the capital adjustment cost parameter. λ is the collateral constraint parameter. ξ is linear equity issuance cost. α is the return to scale. δ 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.

Table 3: Subsample Estimates (Corporate Finance Model)

	ρ_z	σ_z	γ	λ	ξ	α	δ	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.

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

(a) Parameter Estimates

	γ	β	Φ
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	Approx. Moments	Moments with Approx. θ	Moments with True SMM
mean(wealth)	5.63 (.03)	5.63	5.60	5.63
participation rate	.557 (.002)	.557	.568	.557
mean(cond. equity share)	.345 (.003)	.345	.337	.345
median(wealth)	2.00 (.01)	2.94	2.9	2.92
median(cond. equity share)	.224 (.002)	.261	.259	.269

The 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. γ is risk-aversion. β is time discount factor. Φ 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 "Approx. Moments" show the approximate moments at the approximate parameter estimates ($g(\widehat{\boldsymbol{\theta}}^{\text{approx}}, \widehat{\boldsymbol{\beta}})$). Column "Moments with approx. θ" reports the true simulated moments at the approximate parameters ($f(\boldsymbol{\theta}^{\text{approx}})$). "Moments with true SMM" corresponds to the simulated moments for the parameters using the SMM estimation. The moments used in the estimation are defined in Section 4.3.

APPENDIX – FOR ONLINE PUBLICATION

A Appendix Tables

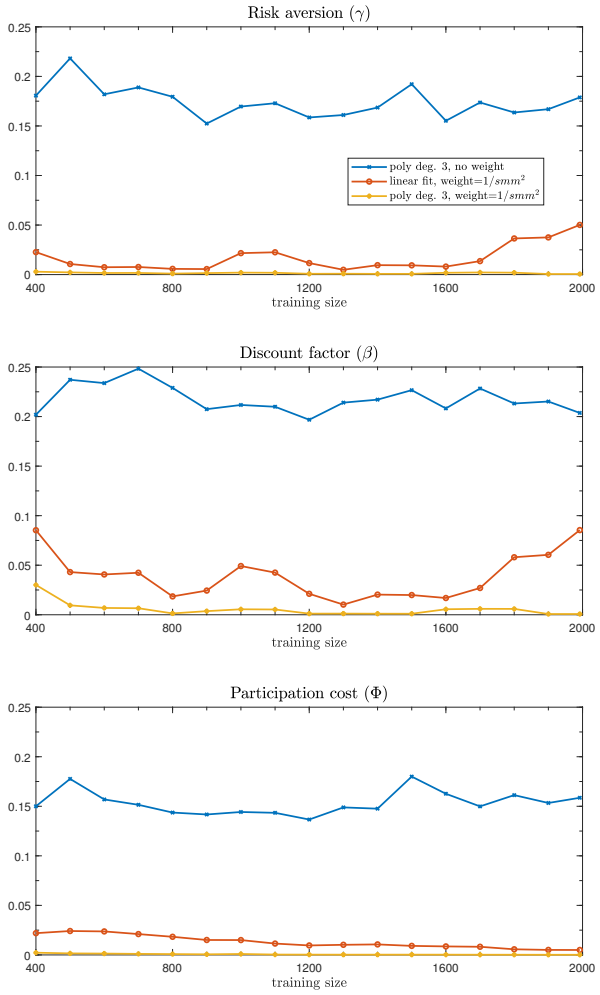
Table A.1: Literature

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			✓	
	2018	Wu RFS		✓	✓	
	2019	Nikolov et al. JFE	✓			
	2021	Frank et al. RFS				✓
		Begenauet al. JFE				✓
	(Forthcoming)	Catherine et al. JF	✓	✓	✓	
Corporate governance	2009	Gayle et al. AER				✓
	2010	Taylor JF				✓
		Kang et al. JFE				✓
	2012	Coles et al. JFE		✓		
	2013	Taylor JFE				✓
	2017	Jung et al. JFE		✓		
	2018	Page JFE			✓	
	2022	Bertomeu JFE				✓
Bankruptcy	2016	Glover JFE				✓
Banking	2014	Schroth et al. JFE		✓		
Corporate control	2014	Dimopoulos et al. JFE	✓			
	2015	Albuquerque et al.JF	✓			
	2018	Li et al. JFE		✓		
		Wang JFE				✓
	2020	Wang JFE	✓			
Entrepreneurship	2020	Jones et al. AER		✓		
	2022	Ewens et al. JFE	✓			
		Catherine JFE	✓			
Household finance	2018	Pagel Econometrica				✓
	2019	Sun et al. JFE	✓			
	2020	Ameriks et al. JPE				✓
	2022	Catherine RFS				✓
Real estate	2015	Corbae et al. JPE				✓
	2017	Landvoigt RFS				✓
	2020	Oh et al. JFE				✓
	2021	Ghent JFE				✓

Figure A.1: 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, $1 - R^2$, calculated over an identified out-of-sample validation set described in fig. 2. 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, considering all cases of R^2 less than .75 as poor performance cases.

Figure A.2: 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, $1 - R^2$, calculated over an identified out-of-sample validation set described in fig. 10. 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).

Table A.2: Preset parameters of life-cycle model

p_z	.136	r	.02
ρ_z	.967	θ_1	.1237
μ_z^-	-.086	θ_2	-.0125
λ_{zl}	4.291	θ_0	-3.165
σ_z^-	.562	t_0	23
σ_z^+	.037	R	65
σ_η^-	.895	T	100
σ_η^+	.089		

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