Robustness Checks in Structural Analysis

Sylvain Catherine * Mehren Ebrahimian†
David Sraer‡ David Thesmar§

April 7, 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 relies on an approximation of the relationship 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) life-cycle savings/consumption.

*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 could explain the main findings, and provide additional empirical analyses that control for such alternative channels. Another standard question in empirical work is the robustness of findings across various sub-samples. When worried that the main finding 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 after dropping such years or industries.

Such robustness checks are rare in structural research using 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. The computational cost of doing so is close to zero. The equivalent analysis in structural estimation would be to (1) consider alternative models (2) re-estimate these alternative models on the data and (3) assess how the inference on the coefficient of interest is affected in these alternatives. In many cases, the computational burden of estimating alternative structural models imply that very few alternatives, if any, can be considered in practice. Similarly, sub-sample analysis, e.g., re-estimating a model for each individual year or industry in the sample, can prove computationally prohibitive. These limitations harm the credibility of structural estimates, since they make it difficult to assess how sensitive the findings are to particular assumptions, or particular feature of the data. This paper offers a simple approach to bypass these computational limitations. Our methodology makes it feasible to perform many robustness checks in simulation-based structural research.

Our methodology works as follows. We consider the structural estimation of an economic model through a simulated method of moments. Given structural parameters $\theta$, the model generates moments $m = f(\theta)$. In most applications (as for instance with dynamic models), $f$ does not admit a closed-form, so that the relationship between moments and structural parameters has to be calculated numerically, via simulations. The structural parameters $\theta$ are then estimated by minimizing a distance between the simulated moments $m$ and the empirical moments $\hat{m}$. While simulating the economic model given $\theta$ is computationally cheap, estimating $\theta$ can be quite costly as it requires a large number of simulations to ensure convergence to a global
optimum. This computational cost often limits the number of estimations (e.g., of alternative models, or on sample splits) that are feasible for a given research project.

We propose to reduce this computational cost using an approximation. Importantly, we do not approximate the numerical solution to the underlying economic model, which 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 $g(\theta, \beta)$ of the function $f(\theta)$, with $\beta$ a vector of parameters. In contrast to the numerical solution of the underlying model, the function $f(\cdot)$, which maps structural parameters to sample moments from simulated data, is smooth in most economic applications. This makes our approximation less computationally intensive. Our approach then boils down to estimating $\beta$.

This is done in three steps. First, we draw a large number of potential structural parameters $(\theta_i)_{i \in [1, S]}$ and simulate the corresponding moments $m_i = f(\theta_i)$. This results in a dataset $D = \{(\theta_i, f(\theta_i))\}_{i \in [1, S]}$, which is fixed once and for all. This simulation stage is the computationally intensive step in our approach. The computational time it uses is similar to the cost of estimating our model once using standard estimation techniques. Second, we use this dataset of moments and parameters to estimate our parametric approximation of $f(\cdot)$. For instance, $g$ can be a k-order polynomial, so that $\beta$ corresponds to the polynomial coefficients, which can be directly estimated by regressing $m_i$ on a flexible polynomial of the elements of $\theta_i$. Alternative methods from statistical learning can be used if in-sample over-fitting is a concern (e.g., penalized regressions or tree-based methods). Note that we allow this approximation to be “local” to the vector of empirical moments $\hat{m}$ in the sense that observations in $D$ that are “closer” to the empirical moments can have more weights in the estimation of $\beta$.

This step – the estimation of $\beta$ – is computationally cheap, especially since in our application, a third-degree polynomial approximation is found to provide a precise approximation. Third, once $\beta$ has been estimated, “approximate” simulated moments $g(\theta, \hat{\beta})$ can be computed for any parameter $\theta$ without simulation. As a result, once the approximation $g(\cdot)$ has been estimated, minimizing the distance between empirical moments $\hat{m}$ and approximate simulated moments $g(\theta, \hat{\beta})$ is orders of magnitude faster than when using the “true” moments $f(\theta)$. This approach can be seen as a generalization of Andrews et al. (2017b), who focus on first-order local deviations and thus restrict $g(\cdot)$ to be a linear function.

How good is this approximation quantitatively? We assess the validity of the approximation in 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), Cocco et al. (2005) and Catherine (2021). For each model, we generate a dataset of parameters and corresponding moments, which we split into a training and a validation sample. We use the training sample to estimate our approximation $g(\theta, \hat{\beta})$ of $f(\theta)$. Using the validation sample, we estimate, for each set of moments $f(\theta_i)$, the best parameters $\hat{\theta}_i$ that minimizes the distance between $f(\theta_i)$ and $g(\hat{\theta}_i, \hat{\beta})$. In other words, for each set of simulated moments generated by the true model, we use the approximation $g()$ to estimate structural parameters. We compare these estimated parameters with the true structural parameters $\theta_i$ used to generate the initial simulated moments $m_i$. In both of our applications, we find that, to the extent that the structural parameters are identified in the true model, the estimates from the approximate model are quantitatively close to the true estimates. Our corporate finance application uses seven structural parameters. Across these parameters, the correlation between true parameters $\theta_i$ and our estimates based on a third order polynomial is very high (.99 in most cases, never lower than .96).  

The approximation $g(\theta, \hat{\beta})$ allows us to run many useful robustness checks that are otherwise computationally prohibitive. We first consider sub-sample analyses. A standard concern in empirical work is the sensitivity of the baseline finding to the particular set of observations used in the estimation. In reduced-form work, it is common to re-estimate regression models using a sub-sample of industries or years. Sample splits are also commonly used as additional test to validate the robustness of the main empirical findings (e.g., when the main effect is expected to be stronger on a subset of firms relative to another). In structural work, the equivalent exercise is to (1) recompute moments on different sub-samples (for instance for each year in the sample) (2) re-estimate the model on these alternative moments. When working with dozens or hundreds of sub-samples, this is typically computationally prohibitive. The approximation $g(\theta, \hat{\beta})$ allows us to all these estimations in minutes. As a result, the researcher can easily plot year-by-year structural coefficient estimates. As we show in both of our example, interesting insights about the validity of the underlying model can be gained with such an analysis.

A usual approach to test the fit of a structurally estimated model is to consider how well the model does in matching non-targeted moments. Our methodology allows us

\footnote{Note that our validation sample contains 200 draws of structural parameters, so that this validation exercise requires 200 estimations. However, with the approximation $g()$, these 200 estimations take less than 1 minute. Using the simulation-based estimation $f()$ would require instead more than a week with standard numerical techniques.}
to perform a similar type of robustness checks in a much more systematic way. Let $M^\text{used}$ be the set of moments used in baseline estimation. Let $M^\text{non-target}$ be the set of moments non-targeted in the baseline estimation. A first robustness exercise is to consider how the baseline estimation is affected if one of the moments in $M^\text{non-target}$ is added to the set of targeted moments. A robust baseline estimation should not be sensitive to this addition. This is similar in spirit to checking ex post that the baseline model does well at matching the non-targeted moment. The difference is that the model is re-estimated for every possible added moments, which provides a better sense for the robustness of the various parameter estimates to the set of moments chosen in the baseline estimation. A second exercise reports how the estimated parameters vary as the model is re-estimated using all the possible combinations of moments in $M^\text{non-target}$. Such an approach is obviously infeasible when working with regular simulation-based methods $f()$, as it involves thousands of estimations. The approximation $g()$ allows to run this analysis in less than a day. By plotting the resulting distribution of parameters, it is straightforward to assess the sensitivity of parameter estimates to the moments selected in the estimation. Such an analysis also serves as a simple diagnostic tool that can be used to assess which moments the model has a hard time making sense of. This diagnostic tool is useful when thinking of model extensions.

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 structural parameters. In reduced-form work, robustness to misspecification can be addressed simply by adding additional variables in regression analyses (e.g., non-linear controls or additional omitted variable in the main specification). 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. While our methodology does not allow us to speed up the estimation of a new model, we can still address robustness to misspecification in the following way. Let $\hat{\theta}^\text{baseline}$ be the set of structural parameters recovered using the baseline model. We consider alternatives to this baseline model. These alternative models generates simulated moments $h(\theta^a)$, where $\theta^a$ is the set of structural parameters in the alternative model. $\theta^a$ contains $\theta^\text{baseline}$ and potentially additional parameters that govern missing forces in the baseline model. Using our approximation, we can estimate how inference on $\theta^b$ is affected when we re-estimate our baseline model using moments generated by this alternative model. Note that this methodology only requires to (1) solve the al-
ternative models to generate simulated moments (2) estimate the approximate model using the moments simulated under the alternative model. Both steps are computationally cheap, once $\beta$ has been estimated. This allows us to consider a large number of alternative parameters $\theta_a$. Note that while we focus here on parameter estimation, we can also perform similar robustness exercises for functions of parameters, such as welfare, productive efficiency, etc.

**Related literature.** Our paper is mostly related to a recent literature that tries to make structural estimation more transparent, with a particular focus on the sensitivity of policy predictions in the presence of 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 the potential 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). Our approximation is global, not local. As we show below, non-linearity is an important feature of the relationship between moments and parameters, so that local approximations can end up producing 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)).

Our paper is also related to the emerging literature trying to improve numerical solutions of model through approximation. Our goal is quite different: We seek to approximate 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 in order to solve the model instead of the moments (their purpose is solution more than estimation). Like us, they draw a number of parameters, for which they solve the model numerically. Then, they train ML algorithms to predict the model outcome on these parameters. The non-
linearity of their model requires that they use highly flexible ML algorithms (deep NN). Even though our examples of models are non-linear too, flexible models are less crucial for us, as we seek to predict moments, which are by design more continuous. Our focus on moments is driven by our interest in estimation and robustness analysis, a focus different from theirs. In the same space of the literature, 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 field 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, our approach approximates the mapping between simulated moments from the true economic model and parameters, i.e. we solve the model exactly to generate our training sample. Since the relationship between moments and parameters is smooth, we find that this relationship can be efficiently approximated by polynomials, and neural networks do not improve the quality of the approximation. 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 assess several important dimensions of robustness of the findings uncovered in these literature.

2 General Approach and Notations

This Section lays out our general approach.

2.1 Approximation

Let $S$ be a structural model with deep parameters $\theta$. The model $S$ generates a vector of moments $f(\theta) \in \mathbb{R}^m$ that have empirical counterparts. 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
can be exactly computed through simulations. Let \( \hat{m} \) be the vector of empirical counterparts to the model-based moments \( f(\theta) \).

Using indirect inference, the minimum distance estimator of \( \theta \) is obtained by minimizing:

\[
\hat{\theta} = \arg \min_{\theta} (\hat{m} - f(\theta))^\prime \Omega (\hat{m} - f(\theta))
\]  

(1)

where \( \Omega \) is a weighting matrix.

The estimation proceeds in two-steps:

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

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

Because \( \theta \) is potentially high-dimensional, estimating a model against a set of empirical moment \( \hat{m} \) can end up being computationally costly, as it requires a large number of inner loops, and each inner loop requires itself 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 \( \hat{m}' \) (e.g., moments estimated for different subsamples, or alternative moments not included in \( \hat{m} \)), the second estimation will have the same computational cost as the first estimation.

The objective of our paper is to create such economies of scope using an approximation of the function \( f() \). Let \( \hat{m} \) be the value of empirical moments we are interested in using. Our approach uses the following steps:

1. We define bounds for each parameter in \( \theta \). These bounds are based on expert knowledge. Note that these bounds also have to be specified for standard estimations. We note \( \Theta \) the resulting set of admissible vectors of parameters.

2. We use a Halton sequence to generate \( N \) vectors of parameters \( (\theta_i)_{i \in [1,N]} \) with \( N \) is large.

3. For each vector of parameters \( \theta_i \), we solve the structural model \( S \) and simulate moments \( f(\theta_i) \). This is computationally intensive as the true structural model has to be solved and simulated \( N \) times. This steps results in a dataset \( D \) that contains \( N \) vectors of parameters and the corresponding \( N \) moments.

4. On the dataset \( D \), we estimate the parameters \( \beta \) of an approximation \( g(\theta, \beta) \) of the function \( f(\theta) \) in the following way:
(a) For each draw \((\theta_i, m_i) \in D\), we compute the distance between \(m_i\) and the moment values of interest \(\hat{m}\): \(\Delta_i = (m_i - \hat{m})' \Omega (m_i - \hat{m})\). \(\Omega\) is a scaling matrix. In our application we choose the empirical variance matrix of moments in the overall sample.

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

\[
\hat{\beta} = \arg \min_{\beta} \left[ \sum_{i=1}^{N} \frac{1}{(\Delta_i)^k} (g(\theta_i; \beta) - f(\theta_i))^2 \right]
\]

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

In our applications, we consider, three types of function \(g()\): linear functions, 3rd degree polynomials and neural nets.

Once we have the approximation \(g(\theta, \hat{\beta})\), we can use it to estimate the deep parameters. We simply need to solve:

\[
\hat{\theta}(\hat{m}; \hat{\beta}) = \arg \min_{\theta} \left( \hat{m} - g(\theta; \hat{\beta}) \right)' \Omega \left( \hat{m} - g(\theta; \hat{\beta}) \right)
\]

This last step is fast since simulations are no longer required, and \(g()\) is a closed-form, smooth function of \(\theta\). We do not have formulas for the standard error of \(\hat{\theta}(\hat{m}; \hat{\beta})\), since the model is misspecified: \(g()\) is not the right model if \(f()\) is, so approximate inference may be biased and noisy. We will show numerically later that the error induced by approximation is, in practice in the case of our models, small.

### 2.2 Relation with the sensitivity matrix

Andrews et al. (2017a) propose a tool to assess the robustness of parameter estimates to misspecification that builds on a local linear approximation of the relationship between moments and parameters. Their analysis relies on a linear approximation between parameters and moments. This is sensible if one is interested in (small) misspecification errors. But if we are interested in estimation for various sets of moments like we are, the linear approximation may be too coarse. This Section clarifies this.
Consider any vector of parameter $\theta \in \Theta$ (possibly, a SMM estimate), and $m = f(\theta)$ the corresponding moments. Consider any $\tilde{\theta}$ close to $\theta$. Then, provided $f()$ is differentiable,

$$f(\tilde{\theta}) \approx m + J(\theta) (\tilde{\theta} - \theta)$$

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

Assume now that the econometrician wants to fit a vector of moments $\hat{m}$ using the approximation $g^\text{linear}$ of the true model $f()$. In this case, the optimization program (2), using the linear approximation, yields an approximate estimate for deep parameters $\hat{\theta}$. They are defined by the following First-Order Condition (FOC):

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

$$\approx J'(\theta) \Omega \left( \hat{m} - m - J(\theta) (\hat{\theta} - \theta) \right)$$

where $\hat{\theta}$ is the approximate estimate of deep parameters. Then:

$$\left( \hat{\theta} - \theta \right) \approx -(J'(\theta) \Omega J(\theta))^{-1} J'(\theta) \Omega (\hat{m} - m)$$

(3)

where $\Lambda$ is the sensitivity matrix of Andrews et al. (2017b).

Assume for instance $\theta$ are parameters estimated by SMM. Andrews et al. (2017b) argue that one can explore the robustness of this estimate with this linear approximation. For instance, Equation 3 can be directly used to explore how estimated parameters would vary for alternative moments $\hat{m}$ that differ slightly (in a first-order sense) from $m = f(\theta)$. In a sense, Andrews et al. (2017b) suggest that one can estimate deep parameters for moments in the neighborhood of $m$ without running an actual SMM, but through the linearization formula 3.

While this result holds locally, the quality of the approximation may deteriorate as the researcher is interested in estimating the model for moments that are further away from the baseline estimate $\tilde{\theta}$. This is the problem we are interested in. In Section 3 and 4 below, we explore the precision of estimation that uses 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 very similar to Hennessy and Whited (2007b) or Ottonello and Winberry (2020): the frictions are adjustment costs, a tax shield for debt, a cost of equity issuance and a collateral constraint. The firm’s shareholder is risk-neutral and her time discount rate is \( r \). At date \( t \), the firm’s EBIDTA \( \pi_t \) depends on the amount of capital and the firm’s productivity as in:

\[
\pi_t = e^{z_t k_t^\alpha},
\]

(4)

with \( z_t \) the firm’s productivity which follows an AR(1) process \( z_t = \rho z_{t-1} + \eta_t \). \( \sigma^2 \) the variance of the innovation \( \eta_t \).

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

\[
k_{t+1} = k_t + i_t - \delta k_t,
\]

(5)

where \( \delta \) is the depreciation rate. In period \( t \), investing \( i_t \) entails a convex cost of \( \frac{\gamma i_t^2}{2 k_t} \). Additionally, the firm pays in period \( t \) for capital that will only be used in production in period \( t + 1 \): This one period time to build for capital is conventional in the macro literature (Hall, 2004; Bloom, 2009) and acts as an additional adjustment cost.

Consistently with the corporate finance literature, we also assume firms’ profits net of interest payments and capital depreciation, \( \delta k_t \), are taxed at rate \( \tau \). 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. \( d_{it} \) is net debt, so that \( d_{it} < 0 \) 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 \( d_{it} \) of debt issued at date \( t \), the firm commits to repay \( (1 + r)d_{it+1} \) 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 \(- (1 + (1 - m)r)d_{it+1} \) with \( 0 < m < 1 \).
Financing frictions come from the combination of two constraints. First, equity issuance is costly. If pre-issuance cash-flows are $x$, cash-flows net of issuance costs are given by:

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

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

$$ d_{it+1} \leq \lambda k_{it+1}. $$

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

The firm is subject to a death shock with probability $D$, but infinitely lived otherwise. 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 $k_{it+1}$ and debt $d_{it+1}$, under the collateral constraint. This value is a function of the elements of the state space: $(k_{it}, d_{it}, z_{it})$. This Bellman equation is written in Catherine et al. (2021).

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 increase in value that firms in the constrained economy would experience if financial frictions were entirely removed. Precisely, let $V_c(k, d, z)$ be the value of a firm with state variable $(k, d, z)$. Define $V^*(k, d, z)$ the value of a firm with state variable $(k, d, z)$ in the absence of financial frictions (i.e. when equity issuance cost is 0, $\zeta = 0$). Then, we define value loss as:

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

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: $\theta = (\delta, \gamma, \alpha, \rho, \sigma, \zeta, \lambda)$. We restrict these parameters to values $\delta \in [0; .2]$, $\gamma \in [0; .12]$, $\alpha \in [.4; .9]$, $\rho \in [.4; 1]$, $\sigma \in [0; .35]$, $\zeta \in [0; .2]$ and $\lambda \in [0; 1]$. This defines the set $\Theta$ of possible parameter values.

We then draw a Halton sequence of $N = 50,000$ vectors $\theta_i$ in $\Theta$. For each $\theta_i$, we simulate the model and compute 17 moments $m_i$, which corresponds to moments that have been used in the literature to estimate this type of model.\footnote{The 17 moments we simulate are: $m_1 = \text{mean}(\text{investment/asset})$, $m_2 = \text{var}(\text{investment/asset})$, $m_3 = \text{mean}(\text{profit/asset})$, $m_4 = \text{mean}(\text{equity issuance/asset})$, $m_5 = \text{var}(\text{equity issuance/asset})$, $m_6 = \text{frequency}(\text{equity issuance})$, $m_7 = \text{coefficient of the regression of investment ratio on market to book ratio}$, $m_8 = \text{coefficient of the regression of net leverage on market to book ratio}$, $m_9 = \text{coefficient of the AR(1) regression of profit ratio with year fixed-effects}$, $m_{10} = \text{residual std of the AR(1) regression of profit ratio with year fixed-effects}$, $m_{11} = \text{mean}(\text{leverage})$, $m_{12} = \text{var}(\text{leverage})$, $m_{13} = \text{mean}(\text{dividend/asset})$, $m_{14} = \text{var}(\text{dividend/asset})$, $m_{15} = \text{autocorrelation of (investment/asset)}$, $m_{16} = \text{std(log growth sales)}$, $m_{17} = \text{std(log growth 5yr sales)}$.}

This results in a training dataset $D = (\theta_i, m_i)_{i \in [1, 50000]}$, which will be used below to estimate various approximations $g()$ of the relationship between moments and parameters $m_i = f(\theta_i) \approx g(\theta_i, \beta)$. 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, so that generating the training dataset takes about 140 hours. However, note that the same training step is performed in SMM estimations that use the Tik Tak algorithm with $N$ starting points. Tik Tak is a multistart algorithm that has been shown to have the strongest performance in both math test functions and economic applications (Arnoud et al. (2019)). Note also that this step is done only once: once the training dataset has been generated, estimation using our approximate methodology takes less than 5 seconds for any set and values of targeted moments. This makes robustness exploration fast.

3.3 Validating the approximation approach

We start our analysis by validating our methodology using model-generated moments. More precisely, we consider the performance of various approximations in estimating structural parameters by targeting moments generated by the model. This methodology allows us to control the data-generating process.

For this exercise, we consider a just-identified estimation of the seven structural parameters that targets the following seven moments: (1 and 2) the s.d. of 1-year and 5 year log sales growth (3) the average investment to asset ratio (4) the average
profit to asset ratio (5) the average equity issuance to asset ratio (6) the average debt to asset ratio and (7) the autocorrelation of investment rates. While the parameters $\theta$ affect all these moments, intuitively, (1) and (2) inform the dynamics of productivity ($\sigma$ and $\rho$), (3) is tightly connected to depreciation $\delta$, (4) characterizes in particular decreasing returns to scale $\alpha$, (5) and (6) are related to equity issuance cost ($\zeta$) and debt enforcement ($\lambda$), and (7) is related to the capital adjustment cost $\gamma$. Our estimation here is just identified (as many moments as parameters), but it does not need to be.

In a first step, we construct a validation sample made out of 1,000 additional random draws of parameters $(\theta_{\text{validation}}^j)_{j\in[1,1000]}$ within the bounds defined above, and their corresponding moments $(m_{\text{validation}}^j)_{j\in[1,1000]}$. Of these, 788 draws end up in a region of the moment space where the model is not identified: the equity issuance to asset ratios is close to 0 (i.e. lower than .001), which is consistent with wide interval of equity issuance cost parameters. We thus exclude these draws and work with a validation dataset of 218 observations.

For each validation moment $m_{\text{validation}}^j$, we first compute the distance between each training moment $m_i$ and $m_{\text{validation}}^j$: $\Delta_i^j = (m_i - m_{\text{validation}}^j)'\Omega(m_i - m_{\text{validation}}^j)$. The distance matrix we use, $\Omega$, is the inverse of the variance-covariance matrix of sample moments. This choice is a bit arbitrary, but there is not obvious way of computing the variance-covariance matrix around $m_{\text{validation}}^j$ since this is not an empirical object. The advantage of this metric is that it allows to scale moments reasonably well.

We then compute the approximation used to estimate the model against the validation moments $m_{\text{validation}}^j$ by minimizing the following objective:

$$\hat{\beta} = \arg \min_{\beta} \left[ \sum_{i=1}^{s} \frac{1}{(\Delta_i^j)^k} (g(\theta_i; \beta) - f(\theta_i))^2 \right]$$

where we experiment different levels of $k$. When $k = 0$, there observations in the training sample are equal-weighted, and the approximation will be the same for each validation moment. As $k$ increases, our approximation, for a given $m_{\text{validation}}^j$, puts larger weights on observations “near” $m_{\text{validation}}^j$ in the sense of the SMM objective function.

We consider 5 different approximations. The first one is a linear approximation $g(\theta, \beta) = \beta' \theta$. We also use a third-order polynomial approximation $g(\theta, \beta) = \sum_{\alpha_0+\alpha_1+\alpha_2+...+\alpha_7 \leq 3} \beta_{\alpha_0\alpha_1\alpha_2...\alpha_7} \prod_{m=1}^{7} (\theta_m)^{\alpha_m}$. For both of these approximations, we need to apply an additional filter to the training dataset: a significant share of the points
in the training dataset ends up with an equity issuance ratio close to 0; as a result, for the parameters corresponding to these points, the model is not identified – different structural parameters $\theta_i$ lead to similar moments (intuitively, when firms never issue equity, equity issuance cost is hard to identified). Training a linear or polynomial model over this region of the training sample will automatically result in a poor fit. For these two approximations, we thus exclude points with an equity issuance to assets ratio below .001 from the training dataset. For the portfolio choice model, we do not have this problem true?

The next three approximations we use do not require this additional filter. 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 of the 218 moments $\left(m_{j_{\text{validation}}}\right)_{j\in[1,218]}$ in our validation set, we can now recover structural parameters by minimizing the SMM objective using the estimated approximation:

$$\theta_{\text{validation}}^j = \arg \min_{\theta} \left( m_{j_{\text{validation}}} - g(\theta; \hat{\beta}) \right)^T \Omega \left( m_{j_{\text{validation}}} - g(\theta; \hat{\beta}) \right)$$

(7)

We evaluate the accuracy of our estimates on the validation sample by computing two measures of fit:

- the fit-weighted $R^2$ for each parameter. This measure captures the performance of the approximate SMM to fit the true data generating process, but gives more weight to validation draws that are better identified in the true SMM. Conceptually, if a particular draw of the validation set corresponds to a set of moments that is not identified in the true model, we do not expect the approximation to recover the true parameters. The fit-weighted $R^2$ is thus defined as:

$$\text{fit-weighted } R^2(k) = 1 - \frac{\text{Var} \left[ \left( \theta_{\text{validation}}(k) - \hat{\theta}_{\text{validation}}(k) \right) \right]}{\text{Var} \left[ \theta_{\text{validation}}(k) \right]},$$

(8)

where the variances are weighted by the inverse of $\text{SE}(\hat{\theta}_{\text{validation}}(k))$, the standard error of the true model estimated at each validation parameter $\theta_{\text{validation}}^j(k)$.\textsuperscript{3}

\textsuperscript{3}We estimate $\text{SE}(\hat{\theta}_{\text{validation}}(k))$ as the squared $k^{th}$ diagonal elements of
the Mean Relative Absolute Error:

\[
\frac{1}{128} \sum_{l=1}^{218} \frac{|\hat{\theta}_l^{\text{validation}}(k) - \theta_l^{\text{validation}}(k)|}{\text{SE}(\hat{\theta}_l^{\text{validation}}(k))},
\]

Figure 1 shows the fit-weighted \( R^2 \) obtained using different approximations for the seven structural parameters, as well as the value loss statistics described in Section 3.1. Figure 2 displays the same information using the MRAE to measure the performance of estimation using the approximate SMM.

These figures yield the following insights. First, using the weights \( \Delta_i^j \) to estimate the approximation results in a significantly higher \( R^2 \) and significantly lower MRAE for most parameters. This result is intuitive: this weighting scheme improves the quality of the approximation in the relevant region, i.e. the region where moments of the training sample are close to the moments being targeted. Across parameters, we see that the best weighting scheme is \((\Delta_i^j)^2\), i.e. using the inverse of the squared SMM objective to weight points in the training dataset.

Second, for several parameters, non-linear fits improve significantly the quality of the estimation relative to a linear fit. This finding is clear in Figure 2. Relative to the polynomial approximation, the linear fit leads to a MRAE that is at least twice larger, for all parameters but depreciation. For quadratic adjustment costs and value loss, the estimation using a linear approximation is on average, across the validation sample, two standard deviation away from the true, data-generating parameter. The two neural nets we consider bring marginal fit improvement relative to the 3rd order polynomial approximation. However, they require significantly more computing time: SMM using the polynomial approximation takes 4 seconds while the two-layer neural net takes 111 seconds. Based on Figure 1 and figure 2, we use, in the rest of this section, the polynomial approximation with weights \( \frac{1}{(\Delta_j^i)^2} \) as our benchmark approximation model. The performance of this benchmark approximation is high on the validation sample: for most parameters, the fit-adjusted \( R^2 \) is above .99; for the equity issuance parameter and the value loss – the two parameters with a weaker performance – the fit-adjusted \( R^2 \) remains above .95.

We then offer a visual representation of the fit in Figure 3. There, we focus on the benchmark approximation (weighted third order polynomials). We plot, on the
x-axis, the true parameters $\theta_j^{\text{validation}}(k)$ and on the y-axis the estimated parameters using the baseline approximation, $\hat{\theta}_j^{\text{validation}}(k)$. A perfect fit would put all points on the 45 degree line. The horizontal bars correspond to $SE(\hat{\theta}_i^{\text{validation}}(k))$, the standard errors that would be estimated using the true SMM for these validation moments. Wide bars imply that the parameter is poorly identified for this particular draw of the validation sample. The large mass of observations on the 45 degree line for all parameters confirms the findings of Figure 1 and 2: the baseline approximation does a good job at recovering true parameters. A caveat applies: the estimated parameters that do not sit on the 45 degree line tend to correspond to draws in the validation sample that leads to poorly identified parameters. If a point in the validation sample is drawn in a region of the parameter space where the model is poorly identified (i.e. where the true SMM would lead to estimating a parameter with a large standard error), it is not surprising that the approximation is more imprecise. This issue appears mostly for the estimation of the issuance cost parameter. Figure 3 shows that wide standard errors for this parameter. This is intuitive: large values of equity issuance costs all lead to similarly low value for the equity issuance to asset ratio. As a result, pinning down the true value of equity issuance costs for high values of equity issuance costs is difficult with this moment, be it with the true SMM or the approximate SMM.

### 3.4 Estimation with Approximate SMM

In this Section, we evaluate the performance of the surrogate model in estimation using actual data.

#### 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 targeted in the baseline estimation using the following variables: (1) the 1-year and 5 year log sales growth is $\text{sale}_l$ and $\text{sale}_{5,\text{sale}}$ (2) the average investment to lagged asset ratio is $\text{capx}_l$ (4) the average profit to asset ratio corresponds uses operating income $\text{oibdp}_l$ (5) the average equity issuance to asset ratio is the ratio of net equity issuance to lagged assets $\frac{\text{sstk} - \text{prstkc}}{\text{at}}$ (6) the average debt to
asset ratio is calculated using net debt \( \frac{dlc+dltt - che}{at} \) and (7) the investment rate is \( \frac{capx}{at} \). 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. Finally, the autocorrelation of investment rates is the regression coefficient of a regression of investment rates on lagged investment rates with year fixed-effects. Column Data in Table 1b provides the mean and standard deviation of these moments in our sample.

### 3.4.2 Estimated Parameters

Table 1 presents parameters estimates of the model using two estimation techniques that targets 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.

2. An approximate SMM that uses the benchmark approximation (third-order polynomial with \( \frac{1}{smm^2} \) weights). We compute standard errors using the delta method with the approximate Jacobian. This estimation of standard errors underestimate true standard errors as it neglects the error introduced by the approximation. However, we show below that this approximation error is minimal and can be neglected.

As can be seen on the table, the estimated coefficients of these two models are quantitatively close: most parameters estimated using the approximate SMM are within 1% of the “true” SMM estimator. Two notable exceptions are (a) the adjustment cost parameter \( \gamma \), where the estimation using the approximation is 2% away from the SMM estimate and (b) the equity issuance cost \( \zeta \), which is 8% away from the SMM estimate. However, Table 1b shows that these differences in parameter estimates lead to small differences in simulated moments. The column “Moments with approx. \( \theta \)” shows the simulated moments using the approximate parameter estimates. These moments are all well within 1% of the the simulated moments using the SMM estimates, except for the equity issuance to assets ratio, which is 1.58% when simulated using the SMM estimates, and 1.64% when using the approximate SMM.
3.4.3 Computing time

Figure 4 shows the estimated structural parameters as a function of computing time, for both the standard SMM and our approximate SMM using our benchmark approximation. The computing times we report exclude the simulation of the training sample, which is required for both estimation. In the case of the approximate SMM, the approximation can be fit and the model estimated with the resulting approximation in four seconds. In the case of the true SMM, it takes at least 17 minutes for the estimation to converge, although full convergence for some parameters requires nine hours (e.g., productivity persistence or return to scale parameters). Therefore, the approximate SMM runs at least 250 times faster than the true SMM once the training dataset has been simulated.

3.5 Using the Approximation to Explore Identification

A standard practice in simulated method of moments is to present local comparative statics to discuss identification. Typically, after the model has been estimated, a researcher will show how variations of parameters around their estimated values affect the simulated moments. This type of local numerical comparative static allows researchers to understand intuitively how local variations in moments would have affected estimated parameters. However, such analysis is necessarily heuristic: in general, all estimated parameters depend simultaneously on all targeted moments, making it impossible from such a local analysis to precisely map variations in moments to variations in structural parameters.

Our approximation approach allows us to run much more systematic analyses of identification, that (a) do not have to be local and (b) can explore directly how variations in empirical moment would affect estimated parameters, even far away from their estimated values. We provide such an analysis in Figure 5. In this figure, we consider how different values for $m_{16}$, the standard deviation of 1-year sales log-growth, would affect the estimation of structural parameters and value losses from financial friction. We consider 100 values for $m_{16}$ ranging from 5% to 35%. For each of these values, we estimate structural parameters using our benchmark approximation, and assuming that all other targeted moments remain at their empirical value. While such an analysis would require several days using a standard SMM approach, it can be done in less than 10 minutes using our approximate SMM. Figure 5 shows the resulting parameter estimates.

The analysis in Figure 5 allows to clearly see how estimated parameters depend on
this particular moment of the data. For instance, while $\lambda$, $\alpha$ and $\delta$ do not vary much with $m_{16}$, we see that the parameters of the log-productivity process ($\sigma_z$ and $\rho_z$), as well as the quadratic adjustment cost ($\gamma$) and the equity issuance cost ($\zeta$) move sharply with $m_{16}$. This figure also allows us to see that the relationship between moments and estimated parameters is not linear. For some moment value, this relationship even becomes non-monotonic. For instance, we see that as the volatility of 1-year sales growth becomes really low, the estimation naturally finds a lower volatility of TFP shocks $\sigma_z$. To keep matching the volatility of 5-year sales growth, a higher persistence of TFP shocks $\rho_z$ is required. At some point, the upper bound $\rho_z = 1$ is reached (for $m_{16} \approx .15$). Below this value of $m_{16}$, since the persistence of TFP shocks can no longer be adjusted, we see that (1) the model fit decreases (the SMM error sharply increases) (2) the model uses a combination of adjustment costs, equity issuance costs, and returns to scale to match the reduced $m_{16}$, leading to non-linear variations in these parameters. Thus, such an analysis allows researchers to understand precisely how identification is achieved globally, as opposed to the usual local approach.

### 3.6 Robustness to Selected Moments

A weakness of methods of moments is that the selection of moments targeted in estimation is arbitrary. There is no well-established theory of moment selection. Assessing how estimated parameters are robust to the particular set of moments targeted in estimation is thus a key aspect of transparency in structural estimation. However, computational costs usually prevent such robustness checks. The initial set of moments that can be matched in estimation is large, so that a thorough check requires to re-estimate the structural model a large number of times. With standard SMMs, this makes such robustness check computationally prohibitive.

In contrast, our approximation approach makes this type of exercise computationally feasible, thus improving the transparency of estimation. Conceptually, this is feasible because when simulating the training dataset $D$, including a large number of moments comes at almost-zero marginal cost: the computationally intensive step is to solve the model; adding an extra-moment in the simulation is cheap. Thus, researchers simply need to include as many moments as possible when drawing the training dataset. In the context of our corporate finance model, we include the 17 moments described in Footnote 2. In addition to the seven baseline moments described in Section 3.4.1, these moments correspond to: (1) the variance of the investment to assets ratio ($m_2$) (2) the variance of the equity issuance to asset ratio ($m_5$) (3) the
frequency of equity issuance \((m_6)\) (4) the coefficient of a regression of investment on market to book using year fixed-effects \((m_7)\) (5) the coefficient of a regression of net leverage on market to book ratio with year fixed-effects \((m_8)\) (6) the coefficient of an AR(1) regression of profit ratios using year fixed-effects \((m_9)\) (7) the standard deviation of the residual of an AR(1) regression of profit ratios using year fixed-effects \((m_{10})\) (8) the variance of the debt to asset ratio \((m_{12})\) (9) the average dividend to asset ratio \((m_{13})\) (10) the variance of the dividend to asset ratio \((m_{14})\).

With this training dataset, it is then possible to explore robustness to moment selection in different ways. We present here two exercises. First, we consider how the estimation of each of the seven structural parameters is affected by adding an additional moment. We start from the estimation of Section 3.4, which targets the seven moments constructed in Section 3.4.1. We then consider 10 alternative estimations that target these seven moments and one of the other 10 moments simulated in the training dataset. Figure 6 reports the findings. Each panel corresponds to one of the seven estimated parameters. The solid black horizontal line corresponds to the baseline estimation that targets the seven moments in Section 3.4. The shaded line plots the 95\% confidence interval. Each coordinate on the x-axis corresponds to one of the additional 17 moments. On the y-axis, we report the estimated parameter when the set of targeted moments include the initial seven moments 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 6 presents a clear diagnostic tool for robustness to moment selection. We see that, for each parameter, the estimation is left unchanged by the addition of an additional moment. For instance, the estimation of the collateral constraint parameters is robust to including the variance of the investment to asset ratio \((m_2)\), the variance of the equity issuance to asset ratio \((m_5)\), the frequency of equity issuance \((m_6)\), the coefficient of a regression of net leverage on market to book ratio \((m_7)\), the coefficient of an AR(1) regression of profit ratios \((m_8)\), the standard deviation of the residual of an AR(1) regression of profit ratios \((m_9)\), the variance of leverage \((m_{12})\), the average dividend to asset ratio \((m_{13})\), and the variance of the dividend to asset ratio \((m_{14})\). However, the estimated collateral constraint parameter is significantly larger when the estimation also targets the coefficient of a regression of investment ratio on market to book ratio \((m_7)\), the standard deviation of the residuals of an AR(1) regression of the profit ratio with year fixed-effects \((m_9)\). However, the estimated collateral constraint parameter is significantly larger when the estimation also targets the coefficient of a regression of investment ratio on market to book ratio \((m_7)\), the standard deviation of the residuals of an AR(1) regression of the profit ratio \((m_9)\), the variance of the leverage ratio \((m_{12})\), or the average dividend to asset ratio \((m_{13})\). In particular, including the variance of the leverage ratio almost triple the estimated \(\lambda\): the baseline estimation results in a variance of leverage that is about 10 times lower than in the data; to match this moment, the estimation requires
a much larger collateral constraint parameter.

More generally, we see in Figure 6 that the baseline estimation is not robust to both the average dividend payout and the variance of leverage. Including these moments result in significantly different estimate for most structural parameters. This is intuitive. For instance, the model is not designed to replicate firm’s actual dividend policy, which is poorly explained by traditional investment models. Figure 6 also allows us to see that the estimation is overall robust to certain moments, such as the coefficient of a regression of net leverage on market to book ratio (m8), or the coefficient of an AR(1) regression of the profit ratio with year fixed-effects (m9). All the estimated parameters remain unchanged whether or not we include these additional moments in the estimation.

The exercise in Figure 6 explores a small set of alternative moment selection: adding one moment to the seven moments used in the baseline estimation. Given the speed at which we can estimate an additional model (about 4 seconds), we can easily explore robustness to moment selection in a much more systematic way. In Figure 7, we re-estimate the model by targeting the seven moments used in our baseline estimation and each possible subset of the other 10 additional moments: this corresponds to \(2^{10} = 1,024\) separate estimations. Of these, we only consider estimations that result in “reasonable” standard errors for parameter estimates: we drop cases where the standard errors for all estimated parameters is more than 10 times larger than the standard error estimated in the baseline estimation. This leaves us with 987 estimates. Figure 7 reports the results of this exercise in the following way: each panel corresponds to one of the seven structural parameters or the value loss from financial frictions; for each of the 1,024 estimations, we rank the resulting estimated parameter in ascending order and report the rank of the parameter, normalized between 0 and 1, on the x-axis; on the y-axis, we report the corresponding parameter estimate in blue and the 95% confidence interval of this estimate in red. The black solid line corresponds to the baseline estimate targeting the initial seven moments. The dashed line reports 95% confidence interval for this baseline estimate. An estimate robust to moment selection would have a large share of observations close to the baseline estimate.

Clearly, most parameter estimates for this model are not robust to moment selection. For instance, the estimated persistence of log-productivity (\(\rho_z\)) varies almost uniformly from .5 to 1, its upper bound: 20% of the estimations find a persistence below .6, while 20% of the estimations find a persistence above .8. While the baseline
estimate is close to the median estimate among the 987, we also see that there is a large spread in the estimated parameters across all the set of moments targeted in these estimations. We reach similar conclusion regarding the robustness of the estimates of the volatility of log-TFP shocks ($\sigma_z$), quadratic adjustment costs ($\gamma$), equity issuance cost ($\zeta$) and returns to scale ($\alpha$).

The estimates of the collateral constraint parameter ($\lambda$) and depreciation rate ($\delta$) appear more robust to moment selection. For instance, for about 50% of the potential set of moments used in estimation, the estimate of $\lambda$ is between .05 and .15. As a result, the estimation of value loss from financial constraint also appears quite robust to moment selection.

Overall, this section illustrates the benefits of our approximation approach to increase the transparency of estimation. The low computational burden of the approximate SMM we use here allows us to check how estimates vary across many sets of targeted moments. This type of robustness analysis is not feasible with standard SMM approaches.

3.7 Sample Splits

A traditional approach to 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 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, hoping that the resulting estimates remain similar.

While such robustness checks can be performed costlessly in a regression setting, the equivalent exercises in a simulated method of moments framework can prove very costly using standard estimation techniques. For instance, over 50 years of data, re-estimating the model every year can take up to several days, if not weeks. As a result, it is uncommon for researchers using such estimation technique to report this type of robustness checks.

The approximation methodology we develop in this paper provides a low cost way to implement such exercises. We provide an example in Figure 8. For every year
in the sample, we re-estimate the model by targeting the moments estimated over a 10-year window. We report the resulting estimates in Figure 8, with their 95% confidence interval. The reported estimated in year t corresponds to the estimation targeting moments 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 estimates decreases from .2 in the 1970s to close to 0 in the 2000s. This results in a large increase in the value loss from financial constraints starting in the 2000. This finding can be simply understood through the lens of the contemporaneous increase in cash holdings over time (Bates et al., 2009). The corresponding reduction in net leverage leads the model to believe that firms are more financially constrained as collateral constraints become tighter (i.e. \(\lambda\) is smaller). This interpretation points to the difficulty of using leverage ratios to identify the severity of credit frictions, a point that echoes the analysis of Catherine et al. (2021).

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 is unlikely to have changed over time, the nature of the capital stock has likely changed: intangible capital has risen at the expense of physical capital (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.

Figure 8 illustrates the power of our methodology to build diagnostic tools: the time-series variations in estimated parameters are useful to understand sources of misspecification in the baseline estimates. This simple exercise, which becomes computationally feasible with our approximation approach, is key in assessing the robustness of the baseline findings.

Another interesting sample split is along the industry dimension. In Table 2, we re-estimate the model using the baseline approximation and targeting moments calculated separately for five broad industries (manufacturing, retail trade, services, transportation, and mining). This is another useful diagnostic tool for the validity of the model. For instance, we find that the value loss from financial frictions are much higher in the mining and transportation industries relative to manufacturing. This finding reflects the much higher equity issuance costs in these industries relative to manufacturing. This can be tied to the higher equity issuance observed in these
4 Dynamic Household Finance Model

4.1 Model

Our model is similar to Catherine (2021) but abstracts from housing choices to focus on consumption choice and 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 the countercyclical income risk documented in Guvenen et al. (2014b) and a realistic of Social Security system in retirement.

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

$$s_t = s_{1,t} + s_{2,t},$$

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

$$s_{1,t} = \begin{cases} s_{1,t}^- & \sim \mathcal{N}(\mu_{s_1}^-, \sigma_{s_1}^2) \quad \text{with probability} \quad p_s \\ s_{1,t}^+ & \sim \mathcal{N}(\mu_{s_1}^+, \sigma_{s_1}^2) \quad \text{with probability} \quad 1 - p_s \end{cases}$$ (10)

On the other hand, $s_{2,t}$ is normally distributed with variance $\sigma_{s_2}^2$. We impose $\mu_{s_1}^- < \mu_{s_1}^+$ and interpret $\mu_{s_1}$ and 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:

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

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

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

$$L_{it} = L_{1,t} \cdot L_{2,it}.$$ (12)
The idiosyncratic component is further decomposed into a deterministic function of age \( f_{it} \), a persistent component \( z_{it} \) and a transitory shock \( \eta_{it} \):

\[
L_{2,it} = e^{f_{it} + z_{it} + \eta_{it}}.
\] (13)

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

\[
z_{it} = \rho_z z_{i,t-1} + \zeta_{it},
\] (14)

where

\[
\zeta_{it} = \begin{cases} 
\zeta^-_{it} \sim \mathcal{N}\left(\mu^-_{z,t}, \sigma^-_{z}^2\right) & \text{with probability } p_z \\
\zeta^+_{it} \sim \mathcal{N}\left(\mu^+_{z,t}, \sigma^+_{z}^2\right) & \text{with probability } 1 - p_z
\end{cases}
\] (15)

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:

\[
\mu^-_{z,t} = \mu^- + \lambda_{zt}(l_{1,t} - l_{1,t-1}).
\] (16)

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

\[
\eta_{it} = \begin{cases} 
\eta^-_{it} \sim \mathcal{N}(0, \sigma^-_{\eta}^2) & \text{if } \zeta_{it} = \zeta^-_{it} \\
\eta^+_{it} \sim \mathcal{N}(0, \sigma^+_{\eta}^2) & \text{if } \zeta_{it} = \zeta^+_{it}
\end{cases}
\] (17)

**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.

\[
T_{it} = 0.124 \cdot \min\{L_{it}, 2.5 \cdot L_{1,t}\}.
\] (18)

The agents receives a retirement benefit which depends on his historical taxable earnings, adjusted for the growth in the national wage index. Specifically, his Social
Security benefit $B$ is:

$$
\frac{B_i}{L_{1,R}} = \begin{cases} 
0.9 \cdot S_{iR} & \text{if } S_{iR} < 0.2 \\
0.116 + 0.32 \cdot S_{iR} & \text{if } 0.2 \leq S_{iR} < 1 \\
0.286 + 0.15 \cdot S_{iR} & \text{if } 1 \leq S_{iR}, 
\end{cases}
$$

(19)

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

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

(20)

where $t_0$ denotes his first year of earnings.

**Agent** The agent has constant relative risk aversion and maximizes his expected utility, given by

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

(21)

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

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

(22)

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

**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.1.

### 4.2 Training dataset and estimation

We consider a vector of three structural parameters: $\theta = (\gamma, \beta, \Phi)$. We restrict these parameters to values $\gamma \in [0; 20]$, $\beta \in [0.5; 1]$, $\Phi \in [0; 0.25]$. This defines the set $\Theta$ of possible parameter values. We then draw a Halton sequence of $S = 2,000$ vectors $\theta_i$ in $\Theta$. For each $\theta_i$, we simulate the model and compute the three moments $m_i$: $m_1$ 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 \( \mathbb{E}[\pi | \pi > 0] \).

We compute these moments using the 1989–2016 waves of the triennial Survey of Consumer Finances (SCF). Specifically, we measure wealth using the net worth variable (\textit{networth}) from the SCF summary extract public data. The equity share is computed as the total holdings of stock (\textit{equity}) divided by net worth, excluding vehicles (\textit{vehic}). We restrict the sample to households whose head is between age 22 and 99 and has positive net worth. To improve comparability across survey years, we scale wealth by the average wage income (\textit{wageinc}) of each survey year.

This results in a training dataset \( D = (\theta_i, m_i)_{i \in [1,S]} \), which we use to estimate \( g() \), the relationship between moments and parameters \( m_i = f(\theta_i) \approx g(\theta_i, \beta) \). Our preferred specification for \( g \) is a third-order polynomial whose coefficients are estimated by OLS using the inverse of the SMM loss function as weights.

Figure 9 illustrates the out-of-sample performance of the approximate SMM approach. Clearly, our method is able to match true parameter values very well. The out-of-sample \( R^2 \) for \( \gamma, \beta \) and \( \phi \) are .994, .986 and .993 respectively. Standard errors are much larger in some parts of the parameter space, which implies substantial variations in the \( J \) matrix or, in other words, that the relationship between moments and parameters is not linear.

Once the training dataset has been built, the approximate SMM approach improves estimation speed by many orders of magnitudes. Indeed, solving the agent’s Bellman equation a single time takes 20 seconds using a high-end GPU whereas running a full approximate SMM using \( g \) takes only a second on a single CPU core.

### 4.3 Identification

Figure 10 illustrates the relationship between targeted moments and estimated parameters. By contrast to our benchmark estimation of the dynamic corporate finance model, we find the relationship between targeted moments and estimated parameters to be highly non-linear, and in fact not even monotonic. Hence, our global approach presents clear advantages over Andrews et al. (2017a). This is somewhat surprising as, ex ante, our estimation strategy seems straightforward: we expect the mean wealth \( m_1 \) to pin down the discount factor (\( \beta \)), the participation rate \( m_2 \) the participation cost (\( \Phi \)) and the equity share of stockholders \( m_3 \) the coefficient of relative risk aversion (\( \gamma \)). Ex post, the mapping between moments and parameters appears more
complex.

Panel A shows how estimated parameters evolve with $m_1$. As expected, targeting a higher level of wealth results in a greater estimated discount factor $\beta$. However, if households are wealthier, preventing them from participating in the stock market require a higher fixed cost, which is why the estimated value of $\Phi$ also increases such that the model still matches the same participation rate $m_2$. More interestingly, the estimated risk aversion $\gamma$ is not a monotonic function of $m_1$. Catherine (2021) shows that, in the presence of countercyclical tail income risk, the optimal equity share is not a monotonic function of $W$. For low wealth individuals, equity shares increase with $W$. Hence, for them, a higher $m_1$ must be offset by an increase in $\gamma$ for the model to match the same mean conditional equity share $m_2$. Above a certain level of wealth, the optimal equity share decreases in $W$. Then, an increase in $m_1$ must be offset by a decrease in $\gamma$ to keep matching $m_3$.

Panel B shows how estimated parameters depend on the stock market participation rate $m_2$. As expected, matching a low participation rates requires a higher fixed participation cost $\Phi$. Excluding more households from participating changes the mean equity share of participants $m_3$ through a composition effect. If the marginal participant has an equity share lower than $m_3$, reducing the participation rate $m_2$ causes $m_3$ to rise, which must be offset by an increase in risk aversion. This seems to be the case around our point estimate but the opposite is true for lower values of $m_2$. As a consequence, our estimate of $\gamma$ is not monotonic in $m_2$ either. Changes $\gamma$ affects wealth accumulation in ways that must be corrected by offsetting adjustments of $\beta$, which causes the relationship between $\beta$ and $m_2$ to be non-monotonic as well. In fact, the evolution of $\beta$ perfectly mirrors that of $\gamma$.

Panel C shows that the role of $m_3$, the mean conditional equity share, can be understood more easily. As long as estimated parameters do not hit the limits of the parameter space, our estimates are monotonic functions of $m_3$. Matching a higher $m_3$ requires a lower risk aversion. A lower risk aversion reduces precautionary savings, and thus calls for a higher discount factor $\beta$ to keep matching $m_1$. Finally, a lower risk aversion makes stocks more attractive and calls for higher stock market participation costs to keep matching $m_2$.

When one of the estimated parameter hits a boundary of the parameter space, such as $\beta = 1$ or $\Phi = 0$, the model can no longer match all the moments and the SMM loss function can have several local minima. Then, the relationship between moments and estimated parameters is no longer differentiable everywhere, as a small change in a target moment can cause the ranking of local minima to change, resulting in
jumps in parameter estimates.

5 Conclusion
References


Cocco, João F., Francisco J. Gomes, and Pascal J. Maenhout, “Consumption


Figures
Approximate SMM estimate fit performance for different weight schemes and models on 218 validation moments. For each set of validation moments $m_{\text{validation}}$, we use the training dataset to approximate the relationship between moments and parameters using five approximations: linear, $3^{rd}$ order polynomial, tobit ($3^{rd}$ order polynomial with equity issuance censored at zero), neural net with two layers, and neural net with 5 layers. The approximation is estimated using different weights on the training sample: unity, $1/\log(smm)$, $1/smm$, $1/smm^2$, $1/smm^5$, and $1/smm^{10}$, where $smm$ is the SMM objective computed at the validation moments. We then use the estimated approximation to estimate parameters $\theta_{\text{validation}}$. The figure shows how the estimated parameters $\theta_{\text{validation}}$ compare to the parameters used to generate the validation moments $\theta_{\text{validation}}$. Chart bars show fit-weighted $R^2$ of structural parameters and value loss estimates, defined in Equation 8.
Approximate SMM estimate fit performance for different weight schemes and models on 218 validation moments. For each set of validation moments $m^{\text{validation}}$, we use the training dataset to approximate the relationship between moments and parameters using five approximations: linear, $3^\text{rd}$ order polynomial, tobit ($3^\text{rd}$ order polynomial with equity issuance censored at zero), neural net with two layers, and neural net with 5 layers. The approximation is estimated using different weights on the training sample: unity, $1/\log (\text{smm})$, $1/\text{smm}$, $1/\text{smm}^2$, $1/\text{smm}^5$, and $1/\text{smm}^{10}$, where $\text{smm}$ is the SMM objective computed at the validation moments. We then use the estimated approximation to estimate parameters $\hat{\theta}^{\text{validation}}$. The figure shows how the estimated parameters $\hat{\theta}^{\text{validation}}$ compare to the parameters used to generate the validation moments $\theta^{\text{validation}}$. Chart bars show MRAE of structural parameters and value loss estimates, defined as: $\frac{1}{218} \sum_{l=1}^{218} \frac{|\theta_l^{\text{validation}}(k) - \hat{\theta}_l^{\text{validation}}(k)|}{\text{SE}(\theta_l^{\text{validation}}(k))}$.
For each set of 218 validation moments \( m^{\text{validation}} \), we use the training dataset to estimate the benchmark approximation and then estimate parameters \( \hat{\theta}^{\text{validation}} \) using the resulting approximation. The approximation we use is a third-order polynomial, using \( 1/smm^2 \) as weighting scheme. The figure reports the estimated parameters against the data-generating parameters. Horizontal bars correspond to \( SE(\hat{\theta}_l^{\text{validation}}(k)) \), the standard errors that would be estimated using the true SMM for these validation moments.
Figure 4: Computing time: Benchmark Approximation vs. SMM

$\rho_z$ is the persistence of the productivity process. $\sigma_z$ is standard deviation of innovations to productivity. $\gamma$ is the capital adjustment cost parameter. $\lambda$ is the collateral constraint parameter. $\xi$ is linear equity issuance cost. $\alpha$ is return to scale. $\delta$ is the capital depreciation rate. The figure shows the estimated parameters as a function of computing time (in minutes, on a logarithmic scale) for both the true SMM and the approximate SMM. Both methods target seven moments described in Section 3.4.1. The true SMM (blue line) is estimated using a Tiktok algorithm (Guvenen et al. (2014a)), with 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 approximate SMM (red line) uses a third-order polynomial approximation with $1/s_{\text{smm}}^2$ weights. The shaded lines correspond to 95% confidence intervals for the SMM estimates. Computing times do not include simulation of the training sample, which is required for both estimations. In the approximate SMM case, fitting and approximation and estimating the model with the approximation takes 4 seconds.
Figure 5: How $m_{16} = \text{std}(1\text{-year sales log-growth})$ affects estimated parameters

This figure shows how estimated parameters vary as a function of $m_{16}$, the standard deviation of 1-year log-growth rate of sales. In the data, $m_{16} = .227$, which corresponds to the solid black vertical line. We consider alternative values for $m_{16}$ ranging from 5% to 35%. For each alternative value, we re-estimate the model using the benchmark approximation, and assuming all other moments remain at their empirical value. The figure plots the resulting estimated parameters.
Each panel corresponds to one of the seven structural parameters or the value loss from financial friction. The x-axis corresponds to one of the following 10 additional moments: $m_2 = \text{var}(\text{investment/assets})$, $m_5 = \text{var}(\text{equity issuance/assets})$, $m_6 = \text{frequency(\text{equity issuance})}$, $m_7 = \text{coefficient of the regression of investment ratio on market to book ratio}$, $m_8 = \text{coefficient of the regression of net leverage on market to book ratio}$, $m_9 = \text{coefficient of the AR(1) regression of profit ratio with firm and year level fixed effects}$, $m_{10} = \text{residual std of the AR(1) regression of profit ratio with firm and year fixed-effects}$, $m_{12} = \text{var(\text{leverage})}$, $m_{13} = \text{mean(\text{dividend/assets})}$, $m_{14} = \text{var(\text{dividend/assets})}$. The y-axis corresponds to the parameter estimated by targeting the initial seven moments and the additional moment on the x-axis using the approximate SMM. The error bars report 95% confidence intervals. The solid black line reports the baseline estimate that only targets the initial seven moments. The dashed lines plots 95% confidence interval for the baseline estimate.
We estimate our model using the baseline approximation (third-order polynomial with $\frac{1}{smm^2}$ weights) and targeting 1,024 different moments, which include the seven moments used in our baseline estimation and all possible subset of the 10 additional moments described in Table 6. For each parameter, we rank the resulting estimates. The figure plots, for each parameter and value loss estimate, the estimates (y-axis) against their ranks, normalized between 0 and 1 (x-axis). We also report 95% confidence interval in red. Note that we drop sets of moments that lead to standard errors that are more than 10 times larger than the baseline standard errors for all parameters, which leaves 987 sets of moments. The black solid line and black dashed lines report baseline estimates and their 95% confidence intervals.
We calculate the seven baseline moments used in the estimation in Section 3.4 using a 10-year rolling window. We then re-estimate the model using the baseline approximation (third-order polynomial with $1/smm^2$ weights) and targeting each set of moments. The estimate using the window $[t-5, t+4]$ is reported in year $t$. The red solid and dashed lines corresponds to the whole sample baseline estimates with their 95% confidence interval.
For each set of 200 validation moments $m^{\text{validation}}$, we use the training dataset to estimate the benchmark approximation and then estimate parameters $\hat{\theta}^{\text{validation}}$ using the resulting approximation. The approximation we use is a third-order polynomial, using $1/\text{smm}^2$ as weighting scheme. The figure reports the estimated parameters against the data-generating parameters. Horizontal bars correspond to $SE(\hat{\theta}^{\text{validation}(k)})$, the standard errors that would be estimated using the true SMM for these validation moments.
Each panel corresponds to one of three moments ($m(k)$=average wealth, participation rate, average conditional equity share) and one of three structural parameters ($\theta(k')$=risk-aversion $\gamma$, discount factor $\beta$, and participation cost $\Phi$). For each $(m(k), \theta(k'))$, we show: (blue solid line) how variations around the measured moments $m(k)$ would affect the estimation of $\theta(k')$, using our baseline approximation (third-order polynomial with $1/smm^2$ weights) (black dashed line) a local linear approximation of the relationship between $m(k)$ and $\theta(k')$ using AGS (orange dashed line) how variations in the structural parameter around its estimated value ($\theta(k')$) affect the simulated moments $m(k)$. 

Figure 10: Robustness, Intensive Margin
# Tables:

## Table 1: Moment and Parameter Estimates: true vs. approximate SMM

### (a) Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>$\rho_z$</th>
<th>$\sigma_z$</th>
<th>$\gamma$</th>
<th>$\lambda$</th>
<th>$\xi$</th>
<th>$\alpha$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>true SMM</td>
<td>.7182</td>
<td>.2133</td>
<td>.0414</td>
<td>.1086</td>
<td>.0379</td>
<td>.8152</td>
<td>.0671</td>
</tr>
<tr>
<td>- s.e., local deriv.</td>
<td>.0066</td>
<td>.0018</td>
<td>.0025</td>
<td>.0031</td>
<td>.0020</td>
<td>.0075</td>
<td>.0008</td>
</tr>
<tr>
<td>approximate SMM</td>
<td>.7130</td>
<td>.2141</td>
<td>.0423</td>
<td>.1085</td>
<td>.0348</td>
<td>.8149</td>
<td>.0670</td>
</tr>
<tr>
<td>- s.e., local deriv.</td>
<td>.0067</td>
<td>.0018</td>
<td>.0025</td>
<td>.0032</td>
<td>.0022</td>
<td>.0076</td>
<td>.0008</td>
</tr>
<tr>
<td>estimation, lower bound</td>
<td>.5</td>
<td>.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.4</td>
<td>0</td>
</tr>
<tr>
<td>estimation, upper bound</td>
<td>.98</td>
<td>.3</td>
<td>.1</td>
<td>1</td>
<td>.2</td>
<td>.9</td>
<td>.2</td>
</tr>
</tbody>
</table>

### (b) Estimated Moments

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Approx. Moments</th>
<th>Moments with Approx. $\theta$</th>
<th>Moments with True SMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_1$</td>
<td>mean($\frac{investment}{assets}$)</td>
<td>.0760 (.0007)</td>
<td>.0760</td>
<td>.0760</td>
</tr>
<tr>
<td>$m_3$</td>
<td>mean($\frac{profit}{assets}$)</td>
<td>.1343 (.0012)</td>
<td>.1343</td>
<td>.1344</td>
</tr>
<tr>
<td>$m_4$</td>
<td>mean($\frac{eq. issuance}{assets}$)</td>
<td>.0158 (.0007)</td>
<td>.0158</td>
<td>.0158</td>
</tr>
<tr>
<td>$m_{11}$</td>
<td>mean(leverage)</td>
<td>.1049 (.0030)</td>
<td>.1049</td>
<td>.1049</td>
</tr>
<tr>
<td>$m_{15}$</td>
<td>autocorr($\frac{investment}{assets}$)</td>
<td>.3754 (.0067)</td>
<td>.3754</td>
<td>.3771</td>
</tr>
<tr>
<td>$m_{16}$</td>
<td>std(log sales growth)</td>
<td>.2270 (.0017)</td>
<td>.2270</td>
<td>.2269</td>
</tr>
<tr>
<td>$m_{17}$</td>
<td>std(log sales growth 5yr)</td>
<td>.5851 (.0052)</td>
<td>.5851</td>
<td>.5849</td>
</tr>
</tbody>
</table>

The table reports the parameter estimates and simulated moments of the corporate finance model presented in Section 3.1. Panel 1a reports the parameter estimates using the true SMM (first line) and the approximate SMM using a third-order polynomial with $1/smm^2$ as weights (second line). $\rho_z$ is the persistence of the productivity process. $\sigma_z$ is standard deviation of innovations to productivity. $\gamma$ is the capital adjustment cost parameter. $\lambda$ is the collateral constraint parameter. $\xi$ is linear equity issuance cost. $\alpha$ is the return to scale. $\delta$ is the depreciation rate of capital. Panel 1b reports the moments targeted in estimation. Column “Data” shows the empirical moments, with standard errors in parenthesis. Column “Approx. Moments” show the approximate moments at the approximate parameter estimates ($\hat{\theta}_{\text{approx}}$). Column “Moments with approx. $\theta$” reports the true simulated moments at the approximate parameters ($f(\hat{\theta}_{\text{approx}})$). “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 3.4.1.
Table 2: Subsample Estimates (Industries)

<table>
<thead>
<tr>
<th></th>
<th>$\rho_z$</th>
<th>$\sigma_z$</th>
<th>$\gamma$</th>
<th>$\lambda$</th>
<th>$\xi$</th>
<th>$\alpha$</th>
<th>$\delta$</th>
<th>Value Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>benchmark (all)</td>
<td>.7130</td>
<td>.2141</td>
<td>.0423</td>
<td>.1085</td>
<td>.0348</td>
<td>.8149</td>
<td>.0670</td>
<td>.0650</td>
</tr>
<tr>
<td>- std</td>
<td>.0067</td>
<td>.0018</td>
<td>.0025</td>
<td>.0032</td>
<td>.0022</td>
<td>.0076</td>
<td>.0008</td>
<td>.0062</td>
</tr>
<tr>
<td>manufacturing</td>
<td>.6921</td>
<td>.2102</td>
<td>.0403</td>
<td>.0730</td>
<td>.0115</td>
<td>.7512</td>
<td>.0542</td>
<td>.0146</td>
</tr>
<tr>
<td>- std</td>
<td>.0089</td>
<td>.0024</td>
<td>.0037</td>
<td>.0039</td>
<td>.0026</td>
<td>.0092</td>
<td>.0007</td>
<td>.0035</td>
</tr>
<tr>
<td>retail trade</td>
<td>.8637</td>
<td>.1347</td>
<td>.0846</td>
<td>.1870</td>
<td>.1696</td>
<td>.8295</td>
<td>.0939</td>
<td>.1174</td>
</tr>
<tr>
<td>- std</td>
<td>.0140</td>
<td>.0042</td>
<td>.0108</td>
<td>.0096</td>
<td>.0490</td>
<td>.0145</td>
<td>.0022</td>
<td>.0250</td>
</tr>
<tr>
<td>services</td>
<td>.7995</td>
<td>.2161</td>
<td>.0277</td>
<td>.0563</td>
<td>.0635</td>
<td>.7488</td>
<td>.0575</td>
<td>.1130</td>
</tr>
<tr>
<td>- std</td>
<td>.0166</td>
<td>.0044</td>
<td>.0045</td>
<td>.0096</td>
<td>.0077</td>
<td>.0176</td>
<td>.0020</td>
<td>.0150</td>
</tr>
<tr>
<td>transportation</td>
<td>.7374</td>
<td>.1673</td>
<td>.0390</td>
<td>.3102</td>
<td>.0893</td>
<td>.9000</td>
<td>.0955</td>
<td>.2246</td>
</tr>
<tr>
<td>- std</td>
<td>.0253</td>
<td>.0075</td>
<td>.0072</td>
<td>.0139</td>
<td>.0122</td>
<td>.0238</td>
<td>.0037</td>
<td>.0634</td>
</tr>
<tr>
<td>mining</td>
<td>.6945</td>
<td>.2575</td>
<td>.0194</td>
<td>.3776</td>
<td>.1218</td>
<td>.9000</td>
<td>.1088</td>
<td>.4835</td>
</tr>
<tr>
<td>- std</td>
<td>.0362</td>
<td>.0170</td>
<td>.0036</td>
<td>.0181</td>
<td>.0110</td>
<td>.0363</td>
<td>.0059</td>
<td>.1910</td>
</tr>
</tbody>
</table>

The table reports the parameter estimates of the corporate finance model presented in Section 3.1. The estimation uses the baseline approximate SMM (third-order polynomial with $1/\text{smm}^2$ as weights) and targets the seven moments used in Table 1b. The first line corresponds to the baseline estimates, for which the targeted moments are estimated on the whole sample. The next lines report estimates when the moments are calculate for five broad industries separately (manufacturing, retail trade, services, transportation, and mining). $\rho_z$ is the persistence of the productivity process. $\sigma_z$ is standard deviation of innovations to productivity. $\gamma$ is the capital adjustment cost parameter. $\lambda$ is the collateral constraint parameter. $\xi$ is linear equity issuance cost. $\alpha$ is the return to scale. $\delta$ is the depreciation rate of capital. The moments used in the estimation are defined in Section 3.4.1.
APPENDIX – FOR ONLINE PUBLICATION

A  Appendix Figures

B  Appendix Tables

Table A.1: Preset parameters of life-cycle model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_z$</td>
<td>0.136</td>
</tr>
<tr>
<td>$r$</td>
<td>0.02</td>
</tr>
<tr>
<td>$\rho_z$</td>
<td>0.967</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>0.1237</td>
</tr>
<tr>
<td>$\mu_z$</td>
<td>-0.086</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>-0.0125</td>
</tr>
<tr>
<td>$\lambda_{zl}$</td>
<td>4.291</td>
</tr>
<tr>
<td>$\theta_0$</td>
<td>-3.165</td>
</tr>
<tr>
<td>$\sigma_{z}^-$</td>
<td>0.562</td>
</tr>
<tr>
<td>$t_0$</td>
<td>23</td>
</tr>
<tr>
<td>$\sigma_{z}^+$</td>
<td>0.037</td>
</tr>
<tr>
<td>$R$</td>
<td>65</td>
</tr>
<tr>
<td>$\sigma_{\eta}^-$</td>
<td>0.895</td>
</tr>
<tr>
<td>$T$</td>
<td>100</td>
</tr>
<tr>
<td>$\sigma_{\eta}^+$</td>
<td>0.089</td>
</tr>
</tbody>
</table>

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