Technological Heterogeneity and Corporate Investment

Theodosios Dimopoulos†  Stefano Sacchetto††

Abstract

We propose an importance-sampling procedure to improve the computational performance of the simulated method of moments (SMM) for the estimation of structural models with fixed parameter heterogeneity. The main advantage of the procedure is that it does not require to simulate observations every time that the structural parameters change during the minimization of the SMM criterion function. We illustrate the use of our method by estimating a neoclassical model of investment for a sample of US manufacturing companies, allowing the technological parameters to vary across firms.

JEL Classifications: C15, D92, E22

Keywords: Parameter Heterogeneity, Structural Estimation, Corporate Investment
1. Introduction

There is a growing body of research that estimates structural models of corporate investment and financing choices. The aim of many papers in this literature is to infer from the data the value of economic variables that are not directly observable by researchers, such as the costs of external financing (Hennessy and Whited, 2007), capital adjustment costs (Cooper and Haltiwanger, 2006), and the magnitude of agency conflicts between managers and shareholders (Nikolov and Whited, 2014). The approach followed by most papers in this literature is to parameterize an intertemporal model of firm-level investment and financing decisions, and estimate the parameters by matching a set of simulated moments from the model to their empirical counterparts. However, by assuming that all firms are described by the same set of parameters, this approach cannot account for the persistent differences in firm policies that have been documented in the empirical literature.

Our paper’s main contribution is to develop a tractable and robust estimation methodology, based on importance sampling, that accounts for parameter heterogeneity in the cross-section of firms. We demonstrate our approach in the context of a dynamic model of investment, allowing the parameters of the profit function and the capital adjustment costs to differ across firms. We estimate the cross-sectional distribution of the model’s parameters and quantify their value for each firm in a sample of US manufacturing companies.

Our methodology is based on the simulated method of moments (SMM). The standard SMM procedure solves numerically the firm’s intertemporal optimization problem, simulates a sample of firms, and computes a set of simulated moments. Minimizing the distance between simulated and empirical moments requires searching over the parameter space, which, in turn, requires numerically solving the firm’s problem for many different parameter combinations. In the presence of heterogeneity across firms, there is an additional computational challenge that makes the estimation of the cross-sectional distribution of parameters with the standard SMM approach intractable: At every step in the minimiza-

---

1See Strebulaev and Whited (2012) for a survey of this literature.
2For example, Lemmon, Roberts, and Zender (2008) show that firm fixed effects account for the majority of the explained variation in leverage regressions. Fuller, Netter, and Stegemoller (2002) provide evidence of differences in acquisition policies across firms. Studies of the determinants of investment typically include firm fixed effects to account for unobserved time-invariant heterogeneity across firms (see Fazzari, Hubbard, and Petersen, 1988), and Campello, Galvao, and Juhl (2013) show that there is considerable variation in investment–cash-flow and investment–Q sensitivities across firms.
tion, one needs to solve a separate intertemporal problem for each parameter combination that occurs in the simulated cross-section of firms. To alleviate the computational burden, we propose an algorithm based on adaptive importance sampling that iteratively estimates restricted local minima by only varying the distribution function of parameters, holding the simulated sample of firms constant, until no more improvement in the estimation criterion is possible. The output of the importance sampling algorithm is then used as an initial guess for the standard SMM estimator.

To illustrate our methodology, we employ a neoclassical model of investment that has become standard in the literature (see, for example, Adda and Cooper, 2003). The model is in discrete time, and the horizon is infinite. Firms are risk neutral and maximize the present value of future cash flows. In each period, cash flows are determined by operating profits, which are affected by persistent firm-specific profit shocks, the value of investment, and capital adjustment costs. The latter are incurred whenever a firm acquires or sells capital, which depreciates over time. The specification of the adjustment cost function accounts for asymmetric costs for investment and disinvestment, in the spirit of Abel and Eberly (1994) and Zhang (2005).

The key extension over the existing structural models in the literature is that we allow the technological parameters to vary across firms. In the model, firms are heterogeneous with respect to six technological parameters: the depreciation rate of capital, the persistence and volatility of the profit shocks, a profit curvature parameter that affects the marginal returns to capital, and two parameters that measure adjustment costs for positive and negative investment, respectively. Thus, in contrast to the existing structural models, different realizations of profits and investment across firms are not necessarily due to different firm-specific shock realizations, but they can be attributed to different firm-specific technological parameters.

We parameterize and estimate the cross-sectional distribution of technologies for a sample of 906 US manufacturing companies in Compustat for the period 1972 to 2012. The estimation results show that a considerable amount of technological heterogeneity exists across firms. Combining the distribution of technologies with firm-specific moments of investment and profits, we are also able to obtain parameter estimates at the firm level.
When we test the empirical performance of the model, we find that most simulated moments closely follow their empirical counterparts, but the discrepancies are statistically significant. This exercise shows the potential, and the limitations, of neoclassical investment models to explain persistent heterogeneity in firm profitability and corporate investment policies.

The use of importance sampling techniques for Monte-Carlo simulation has been studied extensively in the literature (see, for example, Robert and Casella, 2005). Recent application of importance sampling for simulation-based estimation are Richard and Zhang (2007), who study an efficient importance sampling procedure in the context of maximum likelihood estimation, and Ackerberg (2009), who proposes the application of importance sampling in method-of-moments estimators. Our paper contributes to this literature by formulating an importance sampling algorithm that alleviates the computational burden of SMM estimation in the context of a neoclassical model of investment.

Two recent papers estimate structural models of corporate decisions accounting for heterogeneity in firm parameters. Morellec, Nikolov, and Schurhoff (2012) estimate a dynamic capital structure model with agency costs. In their setting, closed form solutions for optimal policies enable estimation by simulated maximum likelihood at a feasible computational cost. To account for heterogeneity, Glover (2016) estimates firm-by-firm a trade-off model of capital structure, to quantify the cross-sectional distribution of financial-distress costs. Whereas these papers focus on financing policies, ours is the first paper, to the best of our knowledge, that estimates a neoclassical model of investment allowing for parameter heterogeneity across firms. Furthermore, our paper does not require closed form solutions of the firm’s policy functions, or to estimate the model separately for each firm in the data. The methodology employed in this paper can benefit researchers in macroeconomics, finance, or industrial organization who seek to incorporate parameter heterogeneity in the estimation of their structural models.

The paper is structured as follows. Section 2 discusses the estimation of models with fixed parameter heterogeneity using SMM. Section 3 presents the importance sampling algorithm. Section 4 applies the importance sampling algorithm to a dynamic model of firm investment. Section 5 concludes.
2. SMM for models with fixed parameter heterogeneity

In this section, we describe how the standard SMM procedure of McFadden (1989) and Pakes and Pollard (1989) can be used to estimate non-linear structural models with fixed parameter heterogeneity.\(^3\)

Consider a dataset \(\{y_{jt}\}\) of \(K\) variables observed for periods \(t = 1, ..., T\) and for \(j = 1, ..., N_f\) firms.\(^4\) At the firm level, the time series of observations \(y_j = (y_{j1}, ..., y_{jT}) \in Y\) is distributed according to the joint density \(f_y(y_j; \theta_j)\), where \(\theta_j\) is a \(N_{\theta} \times 1\) firm-specific parameter vector belonging to the parameter space \(\Theta\). The firm’s choices are summarized by a vector of firm-level moments

\[
h(\theta_j) = \int_Y u(y_j) f_y(y_j; \theta_j) dy_j,
\]

where \(u\) is an \(N_m \times 1\)-valued function. We assume that estimates of the firm-specific moments \(\hat{h}_j\) can be computed in the data.

Firms are heterogeneous with respect to the parameters \(\theta_j\) driving the distribution of the observables \(y_j\). These parameters are time-invariant, and identically and independently distributed across firms with density \(f_\theta(\theta; \lambda)\) for \(j = 1, ..., N_f\). The objective is to estimate and perform inference on the economy-wide parameter vector \(\lambda \in \Lambda\) characterizing the cross-sectional density of firm-specific parameters.

To that end, let \(\eta_i(\lambda)\) be a \(P \times 1\) vector of economy-wide moments describing the cross-section of the \(i\)-th firm-specific moment, \(i = 1, ..., N_m\). A necessary condition for identification is that the total number of moments \(N_m \times P\) exceeds the number of parameters in \(\lambda\). We assume that the \(p\)-th element of \(\eta_i(\lambda)\), \(p = 1, ..., P\), obtains as a result of minimizing the criterion function

\[
\eta_{i,p}(\lambda) = \arg \min_x \Psi_{i,p}(\lambda, x) \equiv \left[ \int_\Theta \psi_{i,p}(h_i(\theta), x) f_\theta(\theta; \lambda) d\theta \right]^2.
\]

\(^3\)See also Gouriéroux and Monfort (1997) for a discussion of SMM.
\(^4\)Throughout the paper we refer to the cross-sectional units as firms to be consistent with the corporate investment application of Section 4.

Electronic copy available at: https://ssrn.com/abstract=2625694
For example, if \( \eta_{i,p}(\lambda) \) is the \( q \)-th percentile of the cross-sectional distribution of \( h_i(\theta) \), then

\[
\psi_{i,p}(h_i(\theta), x) = I\{h_i(\theta) \leq x\} - q,
\]

and if \( \eta_{h,p}(\lambda) \) is the \( q \)-th central moment of \( h_i(\theta) \), then

\[
\psi_{i,p}(h_i(\theta), x) = [h_i(\theta) - \mathbb{E}_\theta(h_i(\theta))]^q - x.
\]

Let \( \eta(\lambda) = (\eta_1(\lambda)', \ldots, \eta_{Nm}(\lambda)')' \) be the \( PN_m \times 1 \) vector of all economy-wide moments. Furthermore, let \( \hat{\eta}_{N_f} \) be a consistent estimate of \( \eta(\lambda) \) in the actual data. If analytical expressions for \( \eta(\lambda) \) exist, estimation can be performed using a GMM approach (see Hansen, 1982). The need for simulation-assisted estimation arises when closed-form expressions for \( f_y(y_j; \theta_j), h(\theta_j) \), and therefore \( \eta(\lambda) \) do not exist. In this case, the researcher can simulate a cross-section of firm-specific parameters \( \theta_s \sim f_\theta(\theta; \lambda), s = 1, \ldots, S \). If, in addition, the researcher can simulate data \( \tilde{y}(\theta_s) \) given the parameters drawn, then it is possible to obtain simulations of firm-specific moments \( \tilde{h}(\theta_s) \) and economy-wide moments:

\[
\tilde{\eta}_{i,p}(\lambda) = \arg \min_x \left[ \frac{1}{S} \sum_{s=1}^{S} \tilde{\psi}_{i,p}(\tilde{h}_i(\theta_s), x) \right]^2.
\]

For example, for the \( q \)-th percentile of \( \tilde{h}_i \),

\[
\tilde{\psi}_{i,p}(\tilde{h}_i(\theta_s), x) = I\{\tilde{h}_i(\theta_s) \leq x\} - q,
\]

and for the \( q \)-th central moment of \( \tilde{h}_i \)

\[
\tilde{\psi}_{i,p}(\tilde{h}_i(\theta_s), x) = \left[ \tilde{h}_i(\theta_s) - \frac{1}{S} \sum_{s=1}^{S} \tilde{h}_i(\theta_s) \right]^q - x.
\]

Equations 6 and 7 represent the simulation estimators of Equation 3 and Equation 4, respectively.
The SMM estimator is given by

$$\hat{\lambda}_{SMM} = \arg \min_{\lambda \in \Lambda} \left[ \tilde{\eta}(\lambda) - \hat{\eta}_{Nf} \right]' W \left[ \tilde{\eta}(\lambda) - \hat{\eta}_{Nf} \right], \quad (8)$$

where $W$ is a symmetric positive definite weighting matrix. To obtain efficiency, the weighting matrix is set to $\hat{W}_{Nf} = \frac{1}{N_f} \text{Var} \left( \hat{\eta}_{Nf} \right)^{-1}$. The asymptotic distribution of the (optimal) SMM estimator is:

$$\sqrt{N_f} \left( \hat{\lambda}_{SMM} - \lambda \right) \overset{d}{\to} N \left( 0, \Omega_{SMM}(\lambda) \right), \quad (9)$$

where

$$\Omega_{SMM}(\lambda) = \left( 1 + \frac{1}{S} \right) \left[ \frac{\partial \eta'(\lambda)}{\partial \lambda} \hat{W} \frac{\partial \eta(\lambda)}{\partial \lambda'} \right]^{-1} \quad (10)$$

and $\hat{W} = \text{plim}_{N_f \to \infty} \hat{W}_{Nf}$.

Solving the minimization problem in Equation 8 can be computationally very cumbersome. For example, the application that we examine in Section 4 requires to solve numerically, for a given value of the structural parameter vector $\lambda$, an intertemporal optimization problem for each draw $s = 1, ..., S$ of the firm-specific parameter vector $\theta_s$, to obtain firm-specific moments $\tilde{h}(\theta_s)$. This procedure must be repeated for each value of $\lambda$ visited in the minimization of the SMM criterion: If the minimization of Equation 8 requires $M$ iterations, the dynamic programming problem will need to be solved numerically in total $M \times S$ times. Hence, this estimation procedure can be extremely time consuming even for relatively simple structural models. In the next section, we propose an algorithm based on importance sampling that can drastically reduce the computational burden of SMM when estimating structural models with fixed heterogeneity.

### 3. Importance sampling algorithm

In this section, we discuss a procedure that alleviates the computational burden of the SMM estimation by replacing the standard Monte Carlo simulation estimator $\tilde{\eta}(\lambda)$ with one based on importance sampling, which we denote by $\bar{\eta}(\lambda)$.

Let $g(\theta)$ be an alternative distribution to $f_\theta(\theta; \lambda)$, with the same domain but invariant
to λ. The importance-sampling estimator for the p-th element of \( h_i \) is:

\[
\eta_{i,p}(\lambda) = \arg \min_x \left[ \frac{1}{S} \sum_{s=1}^S \tilde{\psi}_{i,p}(\tilde{h}_i(\theta_s), x) \frac{f_\theta(\theta_s; \lambda)}{g(\theta_s)} \right]^2,
\]

(11)

where \( \theta_s \sim g(\theta) \) and S is the number of random draws from \( g(\theta) \). This importance-sampling estimator rests on computing firm-specific moments on parameters \( \theta_s \) drawn from a distribution that is potentially different from the true distribution \( f_\theta(\theta; \lambda) \). Unlike the draws in the standard Monte Carlo approach described in Section 2, the draws \( \theta_s \) in Equation 11 do not change with λ. Therefore, the firm-specific moments \( \tilde{h}(\theta_s) \) need not be recomputed as λ changes. This represents a major advantage of the importance sampling procedure compared to the standard Monte Carlo approach, and results in very large savings in computation time.

It is important to notice that the variance of the simulation estimator in Equation 11 increases the more dissimilar the distributions \( f_\theta(\theta; \lambda) \) and \( g(\theta) \) are (see Robert and Casella, 2005). For this reason, we follow an adaptive procedure to periodically update the importance-sampling distribution \( g \). Accordingly, we initiate this distribution at \( g^0(\theta) = f_\theta(\theta; \lambda^0) \) for an arbitrary \( \lambda^0 \in \Lambda \). We then draw firm-specific parameters \( \theta^0_s \), for \( s = 1, \ldots, S \), and compute the corresponding moments \( \tilde{h}(\theta^0_s) \). Allowing the structural parameters \( \lambda \) to vary in a narrow neighborhood \( \Lambda^0 \) of \( \lambda^0 \), we minimize the SMM criterion to obtain an update of the parameter vector

\[
\lambda^1 = \arg \min_{\lambda \in \Lambda^0} \left[ \tilde{\eta}^0(\lambda) - \tilde{\eta}_N \right]^T \tilde{W}_N \left[ \tilde{\eta}^0(\lambda) - \tilde{\eta}_N \right].
\]

(12)

Notice that \( g^0(\theta) \) and \( h(\theta^0_s) \) are invariant to the argument of optimization, \( \lambda \). At the new value \( \lambda^1 \), we draw new firm-specific parameters \( \theta^1_s \), \( s = 1, \ldots, S \). Rather than discard, however, the firm-specific moments from the previous step, we use the simulation estimator

\[
\eta^1_{i,p}(\lambda) = \arg \min_x \left( \frac{1}{S} \sum_{s=1}^S \frac{1}{2} \left[ \tilde{\psi}_{i,p}(\tilde{h}_i(\theta^0_s), x) \frac{f_\theta(\theta^0_s; \lambda)}{f_\theta(\theta^0_s; \lambda^0)} + \tilde{\psi}_{i,p}(\tilde{h}_i(\theta^1_s), x) \frac{f_\theta(\theta^1_s; \lambda)}{f_\theta(\theta^0_s; \lambda^1)} \right] \right)^2,
\]

(13)

which uses all the simulated moments in iterations 0 and 1 and attaches different importance-
sampling weights depending on the distribution from which the simulated parameters are drawn. This procedure is iterated so that, at iteration $L$,

$$
\eta_{i,p}^L(\lambda) = \arg \min_x \left( \frac{1}{S} \sum_{s=1}^S \sum_{l=0}^L \frac{1}{L+1} \tilde{v}_{i,p} \left( \tilde{h}_i^{l}(\theta_{s}^l), x \right) \frac{f_{\theta}(\theta_{s}^l; \lambda)}{\hat{f}_{\theta}(\theta_{s}^l; \lambda)} \right)^2,
$$

(14)

and $\lambda^{L+1}$ is obtained from minimizing the SMM criterion

$$
\lambda^{L+1} = \arg \min_{\lambda \in \Lambda^L} \left[ \eta^L(\lambda) - \hat{\eta}_N \right]' \hat{W}_N \left[ \eta^L(\lambda) - \hat{\eta}_N \right].
$$

(15)

To limit the simulation variance of $\eta^L(\lambda)$ it is important to restrict the range over which $\lambda$ varies in each iteration $L$. A natural choice is to require that $\Lambda^L$ is such that each element of $\lambda$ can vary at most by a fraction $\delta_\lambda$ from the corresponding value in $\lambda^L$.

Crucially, the optimization of the right side of Equation 15 does not require new solutions to the firm’s intertemporal optimization problem as $\lambda$ changes. New solutions are only needed once the parameter vector $\lambda^{L+1}$ is obtained. Moreover, this importance-sampling approach makes use of the entire history of firm-specific moments, calculated at all iterations, and therefore its accuracy improves as new iterations take place.

Below we outline an algorithm that describes the steps needed to implement this adaptive importance-sampling simulated method of moments (IS-SMM) procedure.

**Algorithm 1 (IS-SMM)**

1. **Initiate the structural parameter vector** $\lambda^0 \in \Lambda$.

2. **Fix** $\delta_\lambda > 0$.

3. **In iteration** $L \geq 0$:

   (a) **Update the parameter space**: $\Lambda^L = \left( \lambda \in \Lambda : \underline{\lambda}^L \leq \lambda \leq \bar{\lambda}^L \right)$, where

   $$
   \underline{\lambda}^L = \min \left\{ (1 - \delta_\lambda)\lambda^L, (1 + \delta_\lambda)\lambda^L \right\} \text{ and } \bar{\lambda}^L = \max \left\{ (1 - \delta_\lambda)\lambda^L, (1 + \delta_\lambda)\lambda^L \right\}.
   $$

   (b) **Draw** $S$ independent values $\theta_{s}^L$ from $f_{\theta}(\theta; \lambda^L)$.

   (c) **For each** $s = 1, \ldots, S$ simulate a $K \times T$ matrix of data $\tilde{y}(\theta_{s}^L)$ according to the postulated structural model.
(d) Compute firm-specific moments \( \tilde{h}(\theta^j_s) \) for each simulated parameter.

(e) Given the history of simulated moments \( \{\tilde{h}(\theta^i_l)\}_{l=0,s=1}^{L,S} \) and parameters drawn \( \{\theta^i_l\}_{l=0,s=1}^{L,S} \), minimize the criterion in Equation 15 starting at \( \lambda = \lambda^L \) to obtain \( \lambda^{L+1} \).

(f) Update the history of economy wide parameters \( \{\lambda^l\}_{l=0}^{L+1} \).

4. Iterate step 3 until \( \|\lambda^{L+1} - \lambda^L\| < \epsilon \), where \( \epsilon \) is a tolerance parameter, to obtain the final parameter vector \( \tilde{\lambda}_{IS-SMM} \).

The savings in computational time compared to the standard Monte Carlo approach described in Section 2 are generated in step 3(e) of Algorithm 1: The histories \( \{\tilde{h}(\theta^i_l)\}_{l=0,s=1}^{L,S} \) and \( \{\theta^i_l\}_{l=0,s=1}^{L,S} \) do not change with the argument of the criterion function (Equation 15) and, therefore, minimization is very fast.

Once Algorithm 1 converges, the final parameter vector \( \tilde{\lambda}_{IS-SMM} \) is used as an initial guess to minimize the SMM criterion in Equation 8 using the standard Monte Carlo approach and obtain \( \tilde{\lambda}_{SM} \), for which inference can be performed using the asymptotic distribution in Equation 9.

For firm \( j \) with vector of moments \( \hat{h}_j \) estimated from the actual data, the expected firm-specific parameter vector is

\[
\text{E} \left( \theta \mid \hat{h}_j \right) = \frac{\int \theta \xi \left( \theta, \hat{h}_j \right) d\theta}{\int \xi \left( \theta, \hat{h}_j \right) d\theta},
\]

where \( \xi(\theta, h) \) is the joint density of firm-specific parameters and moments. To obtain firm-specific estimates of the structural parameters, we use the following procedure. Let \( \theta_1, \ldots, \theta_S \) be random draws from \( f_\theta \left( \theta; \tilde{\lambda}_{SM} \right) \). If \( \xi \) is given in a known closed form expression, we can estimate \( \text{E} \left( \theta \mid \hat{h}_j \right) \) by Monte Carlo integration:

\[
\hat{\theta}_j = \frac{\sum_{s=1}^{S} \theta_s \xi \left( \theta_s, \hat{h}_j \right)}{\sum_{s=1}^{S} \xi \left( \theta_s, \hat{h}_j \right)}.
\]

If such an expression is unavailable, we can obtain an estimate of the joint density of parameters and moments, \( \hat{\xi}(\theta, h) \), by fitting a Gaussian copula on \( \left( \theta_1, \tilde{h}(\theta_1) \right), \ldots, \left( \theta_S, \tilde{h}(\theta_S) \right) \).
Then, the simulation estimator of the firm-specific parameter vector becomes

$$\hat{\theta}_j = \frac{\sum_{s=1}^{S} \hat{\theta}_s \hat{\xi}(\hat{\theta}_s, \hat{h}_j)}{\sum_{s=1}^{S} \hat{\xi}(\hat{\theta}_s, \hat{h}_j)}.$$  

(18)

4. Application: corporate investment with heterogeneous technologies

In this section, we present a concrete application of our methodology, using a neoclassical corporate investment model with parameter heterogeneity. First, we describe the model, discuss the choice of the moments to match, and provide a step-by-step analysis of the estimation algorithm. Then, we describe the data sample and present the results of the estimation. Finally, we discuss the cross section of firm-level technology parameter estimates.

4.1. Model

We study a standard neoclassical dynamic model of investment (see Adda and Cooper, 2003). The economy is populated by risk-neutral firms that discount future cash flows at rate $r$. Time is discrete, and the horizon is infinite. Firm $j$’s operating profits in period $t$ are

$$\pi(z_{j,t}, k_{j,t}; \alpha_j) \equiv z_{j,t} (k_{j,t})^{\alpha_j},$$  

(19)

where $k_{j,t}$ is the capital stock, $\alpha_j \in (0, 1)$ is the firm-specific curvature parameter of the profit function, and $z_{j,t}$ is a random profit shock that follows a first-order Markov chain in $[\underline{z}_j, \overline{z}_j]$, which approximates the process

$$\log(z_{j,t+1}) = \rho_j \log(z_{j,t}) + \sigma_j \varepsilon_{j,t},$$  

(20)
where \( \rho_j \in (0, 1) \), \( \sigma_j \geq 0 \) and \( \varepsilon_{j,t} \) follows a standard normal distribution.\(^5\) The dynamics of capital are determined by the firm’s investment choices,

\[
i_{j,t} = k_{j,t+1} - (1 - \delta_j)k_{j,t},
\]

(21)

where \( \delta_j \in (0, 1) \) denotes the depreciation rate of capital. Investment results in adjustment costs,

\[
A(i_{j,t}, k_{j,t}, \phi_j^-, \phi_j^+) = \frac{\phi_j^- \mathbb{I}\{i_{j,t} < 0\} + \phi_j^+ \mathbb{I}\{i_{j,t} \geq 0\}}{2} \left( \frac{i_{j,t}}{k_{j,t}} \right)^2 k_{j,t},
\]

(22)

where \( \mathbb{I}\{\cdot\} \) is an indicator function, \( \phi_j^- \geq 0 \) and \( \phi_j^+ \geq 0 \). The specification of adjustment costs in Equation 22 accounts for costly reversibility of investment decisions, and it allows for different levels of adjustment costs for investment and disinvestment (see Zhang, 2005).\(^6\)

We define the firm’s technology by the parameter vector \( \theta_j \equiv (\alpha_j, \rho_j, \sigma_j, \phi_j^-, \phi_j^+, \delta_j) \).

We assume that \( \theta_j \) is time invariant. Therefore, unlike most prior papers in the literature, firm heterogeneity in the model results not only from transitory profit shocks, but also from permanent differences in investment policies across firms.

Using Bellman’s principle of optimality, we characterize firm \( j \)'s dynamic problem as the solution to

\[
V(z_{j,t}, k_{j,t}; \theta_j) = \max_{k_{j,t+1}, i_{j,t}} \pi(z_{j,t}, k_{j,t}; \alpha_j) - i_{j,t} - A(i_{j,t}, k_{j,t}; \phi_j^-, \phi_j^+) \\
+ \frac{1}{1 + r} \int_{z_{j,t}} V(z_{j,t+1}, k_{j,t+1}; \theta_j) dF(z_{j,t+1} \mid z_{j,t}; \rho_j, \sigma_j),
\]

(23)

where \( F(\cdot \mid z_{j,t}; \rho_j, \sigma_j) \) is the profit shock’s transition c.d.f. implied by Equation 20, subject

\(^5\)The period profit function in Equation 19 is commonly used in the literature (e.g., Hennessy and Whited, 2005). Cooper and Ejarque (2003) show that this specification can be derived in a setting in which a firm with market power faces a demand function with constant elasticity, has a production function with constant returns to scale over capital and a flexible input, and is affected by shocks to profit, demand, and input prices. See also Abel and Eberly (2011).

\(^6\)Cooper and Haltiwanger (2006) use a specification that assumes firms pay a fixed adjustment cost every time they invest or disinvest assets. Such fixed costs give rise in equilibrium to lumpy investment behavior: Periods of high investment are followed by periods of inactivity. Whereas such inactivity is present at the plant level, it is rarely observed for large firms. For this reason, we model investment frictions by means of convex adjustment costs.
to the capital accumulation constraint in Equation 21.

4.2. Technological heterogeneity and optimal investment

In this subsection, we derive the optimality conditions for investment, and describe the effects of technological heterogeneity. Using the first-order condition of Equation 23 and the envelope condition with respect to capital, the optimality condition for investment is

\[
1 + \frac{i_{j,t}}{k_{j,t}} \left[ \phi_j^- I\{i_{j,t} < 0\} + \phi_j^+ I\{i_{j,t} \geq 0\} \right] = \frac{1}{1 + r} \int_{z_j}^{z_{j,t+1}} \alpha_j z_{j,t+1} (k_{j,t+1})^{\alpha_j - 1} + (1 - \delta_j) \\ \text{d}F(z_{j,t+1}|z_{j,t}; \rho_j, \sigma_j),
\]

\[
+ \frac{i_{j,t+1}}{k_{j,t+1}} \left[ (1 - \delta_j) + \frac{1}{2} \left( \frac{i_{j,t+1}}{k_{j,t+1}} \right) \right] \left[ \phi_j^- I\{i_{j,t+1} < 0\} + \phi_j^+ I\{i_{j,t+1} \geq 0\} \right] dF(z_{j,t+1}|z_{j,t}; \rho_j, \sigma_j).
\]

This investment Euler equation has the following interpretation: The marginal cost of investing in capital today (left-hand side) is balanced against its marginal benefit (right-hand side). The marginal cost is due to the direct cost of installing one additional unit of capital, and the associated adjustment costs. The marginal benefit results from increases in future cash flows, from the availability of future capital, which will be deployed for productive use, and from the reduction in future adjustment costs.

The fact that technology, as described by the parameter vector \((\alpha_j, \rho_j, \sigma_j, \phi_j^-, \phi_j^+, \delta_j)\), is heterogeneous across firms means that, in the model, profits (Equation 19) and investment (Equation 24) exhibit persistent cross-sectional variation. The aim of the empirical analysis in the next sections is to estimate the structural parameters using the cross-sectional variation in investment and profitability across firms in the data. The structural parameters capture important determinants of firms’ technology and competitive environment. For example, a firm’s degree of market power, which affects its profitability, will be reflected in the curvature of the profit function, \(\alpha\). Moreover, a firm may invest more aggressively than others because it has, for technological or competitive reasons, a shorter product life cycle (larger \(\delta\)), a more loyal clientele (higher \(\rho\)), it faces less volatile demand and input

---

7The model satisfies assumptions 9.4–9.7 in Stokey, Lucas, and Prescott (1989). Therefore, the value function exists and is unique.

8See discussion of the profit function in footnote 5.
cost conditions (lower $\sigma$), it has lower fire-sales pressures when disinvesting (lower $\phi^-$), or because it can assimilate the technology of new assets with lower cost (lower $\phi^+$).

4.3. Selection of moments

We now discuss the moments chosen for estimation. We set the real risk-free rate $r$ to 1%, the average annualized difference between the three-month US Treasury bill rate and the quarterly growth rate of the Consumer Price Index over our sample period. We select moments on the basis of their informativeness about the structural parameters. For each firm, we compute a vector of six moments: the mean and standard deviation of operating profits scaled by assets, defined in the model as $zk^\alpha/k$; and the mean, standard deviation, skewness, and serial correlation of the investment rate, $i/k$. We then aggregate these firm-specific moments and choose as moments to match the deciles of their cross-sectional distributions. We leave out the first and ninth deciles to reduce the impact of outliers, which can be large given the short time series available for each firm in the data.

Each component of the parameter vector $\theta$ affects essentially all firm-level moments, often in a nonlinear way. However, to gain intuition about what guides our choice of moments to match for estimation, we highlight some key relations between firm-level parameters and moments. We do so by means of a comparative statics analysis, plotting firm-level simulated moments as a function of firm-specific parameters in Figure 1.

A higher curvature parameter, $\alpha$, for the profit function leads to lower average operating profits as a fraction of capital. The autocorrelation, $\rho$, and variance, $\sigma$, of the profit shock are positively related to the autocorrelation of the investment rate and the variance of operating profits, respectively. The depreciation parameter, $\delta$, positively affects the average investment rate. Finally, the standard deviation of the investment rate decreases in both adjustment cost parameters. However, $\phi^+$ reduces the skewness of investment, whereas $\phi^-$ increases it.
Figure 1: Comparative statics of firm-level moments. This figure shows the relation between simulated moments of investment, $i/k$, and operating income, $zk^\alpha/k$, as a function of the firm-specific parameters.
4.4. Computational advantages of the estimation procedure

Before describing the implementation of the importance-sampling algorithm, it is worth highlighting its computational advantages, compared to the standard SMM approach, in estimating our model with fixed firm heterogeneity.

Consider, first, the case in which there is no parameter heterogeneity, that is, all firms are characterized by the same parameter vector \( \theta \equiv (\alpha, \rho, \sigma, \phi^-, \phi^+, \delta) \). Under this assumption, which is common in the literature, investment and profits vary across firms only because of different profit-shock realizations. In this case, the object to be estimated is the parameter vector \( \theta \). Instead, in our model, each firm \( j \) is characterized by its own parameter vector \( \theta_j \), and what needs to be estimated is \( \lambda \), the economy-wide parameter vector that determines the cross-sectional distribution \( f_\theta(\theta; \lambda) \) of the firm-specific parameters. For example, one element of the vector \( \theta \) is the parameter \( \alpha \), which determines the curvature of the profit function. If all firms had the same technology, we would need to estimate one number for \( \alpha \). However, with parameter heterogeneity, firms have different \( \alpha \)'s, and we need to estimate the parameters that describe the distribution of \( \alpha \)'s across firms.

When firms are heterogeneous in parameters, the standard SMM approach described in Section 2 requires to solve the firm optimization problem for different draws of \( \theta \) each time a particular value of \( \lambda \) is considered in the minimization of the SMM criterion (Equation 8). When \( \lambda \) changes, the cross-sectional distribution of \( \theta \) changes and, as a consequence, new values of \( \theta \) need to be drawn and the corresponding firm optimization problems need to be solved again.

The main computational advantage of the importance-sampling algorithm is that it allows to vary locally \( \lambda \) and \( f_\theta(\theta; \lambda) \), but keep the values of \( \theta \) for the sample of simulated firms fixed. To see this, notice that the importance-sampling estimator in Equation 11 has three parts: (i) the moments of the simulated firm sample given a draw of \( \theta \); (ii) the auxiliary distribution \( g(\theta) \); and (iii) the density \( f_\theta(\theta; \lambda) \). The importance sampling algorithm varies only (iii), keeping fixed (i) and (ii), which implies considerable savings in computation time, as one does not need to solve the firm optimization problem and obtain the corresponding simulated moments for new draws of \( \theta \).
4.5. Estimation algorithm

In this section, we describe the algorithm that we use to estimate the structural model of corporate investment. We assume that each element \( m \in \{ \alpha, \rho, \sigma, \phi^-, \phi^+, \delta \} \) of the parameter vector \( \theta \) is independently distributed across firms according to a beta distribution with parameters \( \lambda_m = (\beta_m, \gamma_m) \) over the domain \([m, \overline{m}]\). The beta distribution offers a high degree of shape flexibility: depending on parameter values, it can be symmetric, positively or negatively skewed, or bimodal. Moreover, this distribution comes as a natural choice for random variables with compact support, such as the autocorrelation coefficient of the profitability shock, \( \rho \), the depreciation rate, \( \delta \), and the profit curvature parameter, \( \alpha \). We assume that the support for the parameters has lower bounds \( \theta^l \equiv [\alpha, \rho, \sigma, \phi^-, \phi^+, \delta] = [0.02, 0.02, 0, 0, 0, 0, 0.05] \) and upper bounds \( \theta^u \equiv [\pi, p, \sigma, \phi^-, \phi^+, \delta] = [0.98, 0.98, 2, 5, 5, 0.8] \).

We solve Equation \(23\) by value function iteration. To that end, we discretize the state space for both the profit shock and the level of capital. Since firms’ technologies are heterogeneous, the ergodic sets for \( z \) and \( k \) differ depending on the parameters. To compute expectations over future profit shocks, we follow Tauchen (1986). For each simulated parameter vector \( s = 1, \ldots, S \), we construct an equally-spaced grid with \( n_z = 10 \) points for \( \log(z) \) that spans eight standard deviations of the ergodic distribution; that is,

\[
\log(z^G_s) = \left( -\frac{4\sigma_s}{\sqrt{1-\rho_s^2}}, \ldots, \frac{4\sigma_s}{\sqrt{1-\rho_s^2}} \right)'.
\]

Capital lies on the grid

\[
k^G_s = \left( k^*_s(1-\delta_s)^{\frac{n_k-1}{s}}, \ldots, k^*_s(1-\delta_s)^{\frac{-n_k^{-1}}{s}} \right)'.
\]

where \( k^*_s = \left( \frac{\alpha_s}{\gamma_s+\delta_s} \right)^{1/\alpha_s} \) is the firm’s steady-state level of capital assuming that \( z_s = 1 \) permanently (see Nikolov and Whited, 2014), and \( n_k = 401 \) is the number of points of the capital grid.

We now provide a detailed analysis of the computational algorithm based on the impor-

\footnote{Our estimation results in Subsection 4.8 show that these constraints do not bind, that is, the density at the boundaries is zero (see Figure 4).}
tance sampling acceleration method discussed in Section 3. We use this algorithm to obtain a starting point \( \tilde{\lambda}_{IS-SMM} \) for the standard SMM estimation. For each step, we indicate the input needed by the procedure and output generated.

Algorithm 2 (IS-SMM for corporate investment model)

1. Initiate the structural parameter vector \( \lambda^0 \in \Lambda \).

2. Fix \( \delta_{\lambda} = 0.2 \).

3. In iteration \( L \geq 0 \):
   
   (a) Update the parameter space by setting \( \Lambda^L = \left( \lambda \in \Lambda : \lambda^L \leq \lambda \leq \tilde{\lambda}^L \right) \) where \( \lambda^L = (1 - \delta_{\lambda}) \lambda^L \) and \( \tilde{\lambda}^L = (1 + \delta_{\lambda}) \lambda^L \).

   (b) Draw, from \( \text{Beta}(\lambda^L_m) \), \( \lambda^L_m = (\beta^L_m, \gamma^L_m) \), \( S = 1000 \) independent values for each element \( m \in \{\alpha, \rho, \sigma, \phi, \phi^+, \phi^-\} \) of the parameter vector \( \theta \).

   Input: \( \theta, \bar{\theta}, \) and \( \lambda^L \). Output: \( 1000 \) draws of \( 6 \times 1 \) parameter vectors, \( \{\theta^L_s\}_{s=1}^S \).

   (c) For each \( s = 1, \ldots, S \), simulate the time series of operating profits and capital, \( \tilde{y}(\theta^L_s) \), from the dynamic investment model. To do so:

      i. Construct a grid for productivity \( z^G_s \) and capital \( k^G_s \) according to Equation 25 and Equation 26, respectively.

      ii. Construct the \( n_z \times n_z \) transition matrix \( P_s \) for the grid \( \log(z^G_s) \) given parameters \( \rho_s \) and \( \sigma_s \) according to Tauchen (1986).

      iii. Compute the period payoff matrix \( U_s \) of \( n_z n_k \) states by \( n_k \) controls. To that end, let \( 1_n \) denote a \( n \times 1 \) vector of ones. Set \( Z_s = 1_{n_k} \otimes z^G_s \), \( K_s = k^G_s \otimes 1_{n_z} \), \( I_s = 1_{n_z n_k} \left( k^G_s \right)' - (1 - \delta_s) K_s 1'_{n_z n_k} \), and

      \[
      U_s = \pi (Z_s, K_s; \alpha_s) 1'_{n_z n_k} - I_s - A \left( I_s, K_s 1'_{n_z n_k}; \phi^-_s, \phi^+_s \right).
      \]

      Input: \( \theta^L_s, z^G_s, k^G_s \). Output: \( U_s \).

   iv. Initiate the value function \( V^{(0)}_s \) as an \( n_z n_k \times 1 \) vector of ones.

   v. At iteration \( n \geq 0 \), update the value function according to

   \[
   V^{(n+1)}_s = \max \left[ U_s + 1_{n_k} \otimes \left( P_s V^{(n)}_1 \right) \right],
   \]

   Electronic copy available at: https://ssrn.com/abstract=2625694
where the maximization is performed column-wise for every row, and $V_{1s}^{(n)}$ is $V_s^{(n)}$ reshaped into a $n_z \times n_k$ matrix. Iterate until convergence:

$$\max \left( \| V_s^{(n+1)} - V_s^{(n)} \| \right) < \nu,$$

where $\nu$ is a tolerance parameter. Collect the resulting value function $V_s^*$ and next-period capital policy $K_s^*$.

Input: $P_s$, $U_s$. Output: $V_s^*$, $K_s^*$.

vi. For each $s = 1, \ldots, S$, simulate 10,000 independent paths of log-productivity innovations for 200 periods. Linearly interpolate the value function and next period log-capital on current period values of log($z$) and log($k$) using the results of step v. Discard the first 180 periods.

Input: $\{ \theta_s^L \}_{s=1}^S$, $\{ V_s^* \}_{s=1}^S$, $\{ K_s^* \}_{s=1}^S$. Output: data on the time series of operating profits and capital for each simulated firm.

(d) Compute for each simulated firm the moments of investment and operating profits $\tilde{h}(\theta_s^L)$, $s = 1, \ldots, S$.

Input. Simulated data from step (c). Output: $S$ vectors of moments $\tilde{h}(\theta_s^L)$, $s = 1, \ldots, S$.

(e) Update the history of moments $\{ \tilde{h}(\theta_s^l) \}_{l=0,s=1}^{L,S}$ and parameters $\{ \theta_s^l \}_{l=0,s=1}^{L,S}$ drawn.

(f) Use the Nelder-Mead simplex method to minimize the criterion in Equation 15 using Equation 14 and starting at $\lambda = \lambda^L$, to obtain $\lambda^{L+1}$.

Inputs: $\tilde{h}$, $\hat{W}_{N}$, $\{ \tilde{h}(\theta_s^l) \}_{l=0,s=1}^{L,S}$, $\{ \theta_s^l \}_{l=0,s=1}^{L,S}$. Output: $\lambda^{L+1}$.

(g) Update the history of economy-wide parameters $\{ \lambda^l \}_{l=0}^{L+1}$.

4. Iterate step 3 until $\| \lambda^{L+1} - \lambda^L \| < \epsilon$, where $\epsilon$ is a tolerance parameter, to obtain the final parameter vector $\lambda_{IS-SMM}$.

In a given iteration $L$, the minimization of the IS-SMM criterion leads to a solution $\lambda^{L+1}$ which may be an interior or a boundary point of $\Lambda^L$. The parameter space in the next iteration, $\Lambda^{L+1}$, is updated so that parameters vary within a 20% range of their optimum value in the current iteration, $\lambda^L$. Thus, successive iterations allow the algorithm to search for an optimum of the IS-SMM criterion over different regions of $\Lambda$. By step 4 of Algorithm 2, convergence necessitates that $\lambda_{IS-SMM}$ is an interior point of the parameter space in
To illustrate graphically how the algorithm updates the value of \( \lambda \) across iterations, we initialize Algorithm 2 at a particular value \( \lambda^0 \) and plot in Figure 2 the IS-SMM criterion (Equation 15) as a function of each element of \( \lambda \) within a 20% range of the corresponding element of \( \lambda^0 \).

We now discuss the computational performance of the IS-SMM algorithm in the context of our corporate investment application. We run Algorithm 2 on an Exxact desktop with two 12-core 2.3 gigahertz processors, 64 gigabytes RAM and a NVIDIA Tesla Kepler K40C GPU card with 12 gigabytes of memory. At each iteration \( L \), we use parallel computing on the GPU card to solve for optimal firm policies across draws \( \theta^L_s, s = 1, \ldots, S \). Executing steps 3(c) to 3(e) of Algorithm 2 requires 84 seconds on our machine. The same computational time is also required in the standard Monte Carlo approach of Section 2 for the same number of draws per iteration, \( S = 1,000 \). However, each evaluation of the criterion function during the minimization of Equation 15 requires only 0.072 seconds. This allows the algorithm to make fast progress in matching the economy-wide moments \( \hat{\eta}_{N_I} \) without resolving for optimal firm policies for every value \( \lambda \in \Lambda^L \) used in step 3(f) of iteration \( L \).

4.6. Comparison with other estimation methods

In this subsection, we compare SMM estimation to alternative methods, namely, Generalized Method of Moments (GMM) based on the investment Euler equations, and Maximum Simulated Likelihood (MSL).

The moment-matching approach we follow requires solving the Bellman equation in 23 numerically, which is computationally expensive. To circumvent this problem, other papers perform GMM estimation based on the orthogonality conditions implied by the investment Euler equation in 24 (see, for example, Whited, 1992, Whited and Wu, 2006). Estimation on the basis of the investment Euler equation, however, does not present a solution to the computational problem if parameters are heterogeneous across firms. The reason is that the orthogonality conditions implied by the investment Euler equation hold only conditional on parameters, which, in our model, vary randomly across firms. The need arises, then, to in-

\[^{10}\text{To limit the possibility that the estimate } \hat{\lambda}_{IS-SMM} \text{ reflects a local optimum, in our estimation we repeat the algorithm initialization at different values of } \lambda^0.\]

Electronic copy available at: https://ssrn.com/abstract=2625694
**Figure 2: IS-SMM criterion as a function of the economy-wide parameters.** This figure shows the IS-SMM criterion (Equation 15) after one iteration of Algorithm 2 (stopping at step 3.e). Each element \( m \in \{\alpha, \rho, \sigma, \phi^+, \phi^-, \delta\} \) is independently distributed across firms according to a beta distribution with parameters \((\beta_m, \gamma_m)\) over the domain \([m, \bar{m}]\). The structural parameter vector is initiated at \( \beta_m^0 = [30, 3, 14, 170, 12, 4] \) and \( \gamma_m^0 = [40, 2, 70, 130, 70, 50] \). The lower and upper bounds are, respectively, \( \bar{m} = [0.02, 0.1, 0.02, 0, 0, 0.005] \) and \( \bar{m} = [0.98, 0.98, 2, 5, 5, 0.8] \). \( \alpha \) is the curvature of the profit function. \( \rho \) is the autocorrelation coefficient of the log-profitability shock. \( \sigma \) is the volatility of the innovation to the log-profitability shock. \( \phi^+ \) and \( \phi^- \) are adjustment-cost parameters for investment and disinvestment, respectively. \( \delta \) is the depreciation rate of capital. Estimation is performed using a sample of 906 publicly traded manufacturing firms between 1972 and 2012 from the Compustat Industrial Annual database. The details of the sample construction can be found in Subsection 4.7, the definitions of the variables in Table 1, and the moments in Table 2.
egrate out, for each firm, the parameter heterogeneity, conditional on the firm’s observable quantities (e.g., the investment rate). Because there is no closed-form solution for the conditional density of parameters given firm-specific observables, one would need to estimate this density by means of simulation. In turn, this requires solving for the firm-specific investment policy using the Bellman equation, which renders the Euler-equation approach unhelpful to alleviate the computational burden. A second alternative approach would be to perform MSL estimation using investment-rate data for the firms in our sample. Such a method would effectively use all moments of the distribution of investment in the estimation. Our approach, however, offers the advantage of matching moments that are not based on investment, such as those describing firm profitability. Moreover, from a computational perspective, the SMM approach is superior. This is because, in our model, investment policy does not admit closed-form solutions, unless the capital adjustment-cost parameters are set to zero. MSL, therefore, would require numerically finding how the observed investment rate and capital pairs map to profit shocks for each simulated firm-specific parameter vector, which would substantially slow down the estimation process.

4.7. Data

Our source of data is the Compustat Industrial Annual database for the 1972–2012 period. We start from the full sample of US manufacturing firms (primary Standard Industrial Classification [SIC] code between 2000 and 3999). We delete firm–year observations with missing values, and observations with negative sales (Compustat item SALE) or gross property plant and equipment (PPEGT).

We also drop firms with negative average operating income, as the model does not account for negative operating profits, firms with total assets (AT) less than 10 million real 2009 dollars, and those with growth in real sales or total

\[ \text{In our setting, firm-specific heterogeneity cannot be removed by simple transformations of the data, such as first-differencing, which is commonplace in GMM estimation of linear dynamic panel data (see Arellano and Bond, 1991).} \]

\[ \text{Notice the contrast with the model of leverage policy with firm heterogeneity in Morellec, Nikolov, and Schurhoff (2012), for which closed-form solutions exist. There, MSL can be employed at a lower computational cost.} \]

\[ \text{We replace missing values of sale of property (SPPE) with zeros. According to Frank and Goyal (2003), Compustat records this variable “as missing when a firm does not report a particular item or combines it with other data items.”} \]

\[ \text{Variables are deflated to constant 2009 dollars using the GDP deflator. Source: Bureau of Economic Analysis [www.bea.gov], NIPA Table 1.1.9.} \]
assets greater than 100% in a given year. The latter two filters are common in the empirical literature on corporate investment (see, for example, Almeida and Campello, 2007), and are applied to eliminate firms that are very small or likely to be involved in large mergers or reorganizations. We then choose, for each firm, the longest consecutive time series of data and drop firms with less than 10 observations. Our final sample consists of 19,326 yearly observations for 906 firms, with an average of about 21 observations per firm. Variable definitions are reported in Table 1. Finally, to reduce the effect of outliers, we winsorize the variables at the top and bottom 1%.

Table 1: Definition of variables. The source of data is the Compustat Industrial Annual database. Sample construction is discussed in Subsection 4.7. Compustat item names are in parentheses.

| Variable          | Definition                                                                 |
|-------------------|-----------------------------------------------------------------------------|
| Investment        | (Capital Expenditures (CAPX) – Sale of Property (SPPE)) / Beginning-of-period Assets - Total (AT). Missing values of SPPE are replaced with zeros. |
| Operating income  | Operating Income Before Depreciation (OIBDP) / Beginning-of-period Assets - Total (AT). |

The variables used for estimation are investment and operating income, scaled by beginning-of-period assets. An analysis of variance shows that between-firm variation represents 35% of total variation in investment, and 38% in operating income. Moreover, a test of equality in means for these variables across firms is rejected at the 1% level\(^{15}\) This rejects the notion that the observed variation in profits and investment is due only to transitory firm-specific shocks. Hence, there is significant cross-sectional heterogeneity in investment and profits among firms in our sample. To further illustrate the degree of firm heterogeneity, we provide histograms of the firm-specific empirical moments of investment and operating income in Figure 3 and the deciles of the moments’ distributions in Panel A of Table 2. All moments exhibit substantial cross-sectional dispersion.

It is important to point out that the persistent heterogeneity in firm characteristics found in the data cannot be explained by structural models that assume a unique value of parameters across firms. The typical procedure when estimating such homogeneous-

\(^{15}\) We perform a multivariate test on means for investment and operating income allowing for heterogeneous covariance matrices across groups.
Figure 3: Cross-sectional distribution of empirical moments. This figure shows histograms of the firm-specific moments of investment and operating income used for estimation. The sample consists of 906 firms in Compustat for the 1972–2012 period. We provide the details of sample construction in Subsection 4.7, the definitions of the variables in Table 1, and the summary statistics of the cross-sectional distribution of the moments in Panel A of Table 2.
Table 2: Actual and simulated moments. Each column reports the $i$-th deciles, denoted by $p(i)$, of the cross-sectional distribution of firm-specific moments of investment and operating profits. Panel A presents the target cross-sectional moments, computed from the actual data, used in the estimation. The sample consists of 906 firms between 1972 and 2012 from the Compustat Industrial Annual database. The details of the sample construction can be found in Subsection 4.7 and variable definitions are in Table 1. Panel B shows the simulated moments according to the parameters in Panel A of Table 3. The definition of each variable according to the model is reported in parentheses. Estimation follows the SMM procedure described in Subsection 4.5. Panel C presents $t$-statistics for the difference between actual and simulated moments.

### Panel A: Actual moments

| Moment                                      | $p(2)$ | $p(3)$ | $p(4)$ | $p(5)$ | $p(6)$ | $p(7)$ | $p(8)$ |
|---------------------------------------------|--------|--------|--------|--------|--------|--------|--------|
| Average operating income                    | 0.107  | 0.124  | 0.138  | 0.151  | 0.163  | 0.177  | 0.202  |
| Standard deviation of operating income      | 0.040  | 0.049  | 0.056  | 0.063  | 0.071  | 0.081  | 0.097  |
| Average investment                          | 0.035  | 0.042  | 0.049  | 0.056  | 0.064  | 0.070  | 0.083  |
| Standard deviation of investment            | 0.019  | 0.023  | 0.028  | 0.032  | 0.036  | 0.042  | 0.051  |
| Skewness of investment                      | 0.171  | 0.373  | 0.557  | 0.730  | 0.915  | 1.109  | 1.364  |
| Autocorrelation of investment               | 0.125  | 0.225  | 0.322  | 0.381  | 0.452  | 0.533  | 0.624  |

### Panel B: Simulated moments

| Moment                                      | $p(2)$ | $p(3)$ | $p(4)$ | $p(5)$ | $p(6)$ | $p(7)$ | $p(8)$ |
|---------------------------------------------|--------|--------|--------|--------|--------|--------|--------|
| Average operating income ($zk^{\alpha}/k$) | 0.098  | 0.111  | 0.124  | 0.136  | 0.149  | 0.165  | 0.185  |
| Standard deviation of operating income ($zk^{\alpha}/k$) | 0.041  | 0.048  | 0.054  | 0.060  | 0.068  | 0.076  | 0.088  |
| Average investment ($i/k$)                  | 0.035  | 0.040  | 0.046  | 0.051  | 0.057  | 0.064  | 0.072  |
| Standard deviation of investment ($i/k$)    | 0.018  | 0.022  | 0.027  | 0.032  | 0.038  | 0.046  | 0.056  |
| Skewness of investment ($i/k$)              | 0.300  | 0.372  | 0.460  | 0.566  | 0.688  | 0.835  | 1.016  |
| Autocorrelation of investment ($i/k$)       | 0.292  | 0.347  | 0.393  | 0.433  | 0.473  | 0.515  | 0.562  |

### Panel C: $t$-Statistics for the difference between actual and simulated moments

| Moment                                      | $p(2)$ | $p(3)$ | $p(4)$ | $p(5)$ | $p(6)$ | $p(7)$ | $p(8)$ |
|---------------------------------------------|--------|--------|--------|--------|--------|--------|--------|
| Average operating income                    | -4.23  | -5.65  | -6.94  | -6.43  | -5.19  | -4.87  | -4.12  |
| Standard deviation of operating income      | 0.99   | -0.85  | -1.86  | -2.02  | -2.26  | -3.29  | -3.76  |
| Average investment                          | -0.35  | -1.75  | -2.44  | -4.15  | -6.24  | -4.65  | -6.90  |
| Standard deviation of investment            | -2.79  | -2.19  | -0.92  | 0.04   | 3.07   | 3.52   | 4.41   |
| Skewness of investment                      | 4.90   | -0.04  | -3.79  | -6.05  | -8.41  | -9.91  | -9.54  |
| Autocorrelation of investment               | 10.24  | 9.32   | 5.85   | 5.21   | 1.87   | -1.12  | -5.04  |
parameters models is to filter out heterogeneity in the data by removing firm fixed effects in the variables of interest (see discussion in Strebulaev and Whited, 2012). Instead, in this paper, we aim precisely to understand the drivers of persistent variation in firm policies. We do so by allowing firms to have heterogeneous technological parameters, which we estimate in the following section.

4.8. Parameter estimates

The estimates of the structural parameters \( \hat{\lambda}_m = (\hat{\beta}_m, \hat{\gamma}_m) \), for \( m \in \{ \alpha, \rho, \sigma, \phi^-, \phi^+, \delta \} \), and their associated standard errors are shown in Panel A of Table 3. Figure 4 plots the resulting cross-sectional densities of the firm-level technological parameters. Notice that a technological parameter \( m \) distributed over the domain \([m, \bar{m}]\) according to a beta distribution with parameters \((\beta_m, \gamma_m)\) has mean \( E(m) = \frac{m + \beta_m}{\beta_m + \gamma_m} \) and variance \( \text{Var}(m) = \frac{\beta_m \gamma_m}{(\beta_m + \gamma_m)^2 (\beta_m + \gamma_m + 1)} \). In Panel B of Table 3, we test whether firms are homogeneous with respect to each parameter, that is \( \text{Var}(m) = 0 \), using the null hypothesis \( H_0 : \beta_m \gamma_m = 0 \). We perform this test based on a \( t \)-statistic constructed by the delta method (see Greene, 2003, p. 914). The assumption of homogeneity is rejected at least at the 5% level for all technological parameters, except the disinvestment adjustment-cost parameter, \( \phi^- \). However, when we test that \( E(\phi^-) = E(\phi^+) \), the hypothesis is rejected at the 1% level against the alternative \( E(\phi^-) > E(\phi^+) \). Hence, we conclude that disinvestment adjustment costs exist, their average value is higher than positive-investment adjustment costs, but they display a lower degree of heterogeneity than other technological parameters.

It is informative to examine the performance of the model in matching the cross-sectional moments of investment and operating income. The goal here is not to test a new model of investment; it is to provide a novel empirical test of a standard investment model that accounts for firm-specific heterogeneity. In other words, we examine how well the neoclassical model fits the investment behavior of a heterogeneous cross section of firms. In the first two panels of Table 2, we show the empirical moments used for estimation and the corresponding simulated moments. Statistically, most simulated moments differ significantly from the

\[ \frac{\beta_{\phi^-}}{\beta_{\phi^-} + \gamma_{\phi^-}} = \frac{\beta_{\phi^+}}{\beta_{\phi^+} + \gamma_{\phi^+}}. \]
Table 3: Parameter estimates. Panel A shows the estimates of the structural parameters and their standard errors. Each element $m \in \{\alpha, \rho, \sigma, \phi^-, \phi^+, \delta\}$ is independently distributed across firms according to a beta distribution with parameters $(\beta_m, \gamma_m)$ over the domain $[\underline{m}, \bar{m}]$. The lower and upper bounds are, respectively, $[\underline{\alpha}, \underline{\rho}, \underline{\sigma}, \underline{\phi}^-, \underline{\phi}^+, \underline{\delta}] = [0.02, 0.1, 0.02, 0, 0, 0.005]$ and $[\bar{\alpha}, \bar{\rho}, \bar{\sigma}, \bar{\phi}^-, \bar{\phi}^+, \bar{\delta}] = [0.98, 0.98, 2, 5, 5, 0.8]$. $\alpha$ is the curvature of the profit function. $\rho$ is the autocorrelation coefficient of the log-profitability shock. $\sigma$ is the volatility of the innovation to the log-profitability shock. $\phi^+$ and $\phi^-$ are adjustment-cost parameters for investment and disinvestment, respectively. $\delta$ is the depreciation rate of capital. Estimation follows the SMM procedure described in Subsection 4.5 and is performed using a sample of 906 publicly traded manufacturing firms between 1972 and 2012 from the Compustat Industrial Annual database. The details of the sample construction can be found in Subsection 4.7, and the moments in Table 2. The last row reports the $\chi^2$ statistic for the test of the overidentifying restrictions. Panel B report the $t$-statistics, constructed by the delta method (see Greene, 2003), for the test of parameter homogeneity using the null hypothesis $H_0 : \beta_m \gamma_m = 0$.

Panel A: Structural parameter estimates

| Parameter | Value | Standard error |
|-----------|-------|----------------|
| $\beta_\alpha$ | 35.340 | 6.807 |
| $\beta_\rho$ | 3.571 | 0.163 |
| $\beta_\sigma$ | 15.791 | 2.959 |
| $\beta_{\phi^-}$ | 181.462 | 298.523 |
| $\beta_{\phi^+}$ | 13.859 | 2.351 |
| $\gamma_\alpha$ | 4.465 | 0.224 |
| $\gamma_\rho$ | 43.583 | 8.363 |
| $\gamma_\sigma$ | 2.729 | 0.108 |
| $\gamma_{\phi^-}$ | 78.220 | 14.885 |
| $\gamma_{\phi^+}$ | 142.888 | 223.702 |
| $\gamma_\delta$ | 83.171 | 15.347 |
| $\gamma_\delta$ | 69.124 | 3.544 |
| $\chi^2$ | 814.459 |

Panel B: Parameter homogeneity test

| $t$-statistic of homogeneity test | $\alpha$ | $\rho$ | $\sigma$ | $\phi^-$ | $\phi^+$ | $\delta$ |
|---------------------------------|-------|-------|-------|-------|-------|-------|
| $t$-statistic of homogeneity test | 2.61 | 12.62 | 2.65 | 0.31 | 2.82 | 10.02 |
Figure 4: Cross-sectional distribution of technological parameters. Each element \( m \in \{\alpha, \rho, \sigma, \phi^-, \phi^+\} \) is independently distributed across firms according to a beta distribution with parameters \((\beta_m, \gamma_m)\) over the domain \([m, \bar{m}]\). The estimates of \((\beta_m, \gamma_m)\) are reported in Panel A of Table 3. The lower and upper bounds are, respectively, \([\alpha, \rho, \sigma, \phi^-, \phi^+] = [0.02, 0.1, 0.02, 0, 0, 0.005]\) and \([\bar{\alpha}, \bar{\rho}, \bar{\sigma}, \bar{\phi}^-, \bar{\phi}^+] = [0.98, 0.98, 2, 5, 5, 0.8]\). Estimation follows the SMM procedure described in Subsection 4.7 and is performed using a sample of 906 publicly traded manufacturing firms between 1972 and 2012 from the Compustat Industrial Annual database. The details of the sample construction can be found in Subsection 4.7 and the moments in Table 2.
empirical ones (Panel C). Overall, the $\chi^2$ statistic for the model’s goodness of fit (Panel A of Table 3) allows us to reject the hypothesis that the empirical moments are matched. Perhaps, though, this is not surprising in our case, as there is a large difference between the number of moments we target (42) and the number of structural parameters (12). A closer look at the moment comparison reveals that discrepancies between empirical and simulated moments may be statistically significant, but are often economically small. For example, the discrepancy in average operating income is less than 1.5% across most deciles. This is also the case when we examine the match for the standard deviation of operating income, average investment and the standard deviation of investment. Most simulated moments follow closely their empirical counterparts, yet the remaining discrepancies are statistically significant. From an economic perspective, we argue, therefore, that the model is informative about the heterogeneity in firm profitability and investment. The largest deviations emerge in the skewness of investment and in the lower deciles of the autocorrelation of investment. We attribute this difference to the relatively lower weight that these higher-order moments receive in the SMM objective function.

We obtain firm-specific estimates of the technological parameter vector $\hat{\theta}_j$ using Equation 18. To do so for a given firm $j$, the corresponding firm-specific moments $\hat{h}_j$ need to belong to a relatively tight neighborhood of the set of moments simulated under $\hat{\lambda}$. This is the case for 822 out of 906 firms in our sample. Table 4 presents summary statistics of the technology parameter estimates for these firms. All firm-specific estimates present considerable cross-sectional dispersion. This becomes even more remarkable recalling that our sample is constructed on a relatively homogeneous set of manufacturing firms that are in the sample for at least 11 years.

Overall, the results in this section show that there exists a large degree of persistent heterogeneity in corporate investment policies and firm profitability. Our methodology shows that it is possible to empirically test the potential of dynamic models of investment to explain this heterogeneity at a computationally feasible cost.
Table 4: Summary statistics of firm-specific parameters. This table reports descriptive statistics of the firm-level parameter estimates, \( \hat{\theta}_j \equiv (\hat{\alpha}_j, \hat{\rho}_j, \hat{\sigma}_j, \hat{\phi}_j^-, \hat{\phi}_j^+, \hat{\delta}_j) \). These are obtained, for each firm \( j \), using Equation 18 the cross-sectional parameter estimates in Panel A of Table 3 and the firm-specific empirical moments. \( \alpha \) is the curvature of the profit function. \( \rho \) is the autocorrelation coefficient of the log-profitability shock. \( \sigma \) is the volatility of the innovation to the log-profitability shock. \( \phi^+ \) and \( \phi^- \) are adjustment-cost parameters for investment and disinvestment, respectively. \( \delta \) is the depreciation rate of capital. The sample consists of 822 US manufacturing firms in Compustat for the 1972–2012 period, for which we obtain estimates of firm-specific technological parameters. Details of sample construction are in Subsection 4.7, and variable definitions are provided in Table 1.

| Variable | Mean  | Median | SD    | Min.  | Max.  | Obs. |
|----------|-------|--------|-------|-------|-------|------|
| \( \hat{\alpha}_j \) | 0.453 | 0.452  | 0.078 | 0.267 | 0.653 | 822  |
| \( \hat{\rho}_j \)  | 0.573 | 0.584  | 0.199 | 0.152 | 0.952 | 822  |
| \( \hat{\sigma}_j \) | 0.357 | 0.340  | 0.123 | 0.144 | 0.735 | 822  |
| \( \hat{\phi}_j^- \) | 2.807 | 2.804  | 0.129 | 2.333 | 3.118 | 822  |
| \( \hat{\phi}_j^+ \) | 0.719 | 0.687  | 0.263 | 0.243 | 1.520 | 822  |
| \( \hat{\delta}_j \) | 0.059 | 0.057  | 0.025 | 0.012 | 0.172 | 822  |

5. Conclusion

In this paper, we propose an adaptive importance-sampling method to improve the computational performance of SMM estimation in structural models with fixed cross-sectional parameter heterogeneity. To illustrate our method, we estimate a standard neoclassical model of corporate investment with firm-specific technological heterogeneity for a sample of US manufacturing companies. We find that there is a high degree of dispersion across firms in parameters describing the returns to scale of the profit function, the persistence and volatility of shocks to profits, the rate of economic depreciation, and capital adjustment costs.

The importance-sampling approach we employ shows how the daunting challenges of structural estimation of models with parameter heterogeneity and no analytical solutions can be overcome. The neoclassical model of investment that we consider is a building block on which further dynamic theories of firms’ decisions have been developed. Therefore, additional topics can be explored on the basis of the methodology used in this paper. Two particularly interesting areas are the empirical analysis of heterogeneity in costs of external financing, and the effects of technological heterogeneity on the aggregate price of risk. Although such models are characterized by more state variables than the model in this paper, recent developments in numerical integration methods (Judd, Maliar, and Maliar, 2011) and the use of parallel computing (Aldrich, Fernandez-Villaverde, Gallant,
and Rubio-Ramirez, 2011) alleviate the computational burden required for estimation.

References

Abel, Andrew B., and Janice C. Eberly, 1994, A unified model of investment under uncertainty, *American Economic Review* 84, 1369–1384.

———, 2011, How q and cash flow affect investment without frictions: An analytic explanation, *Review of Economic Studies* 78, 1179–1200.

Ackerberg, Daniel A., 2009, A new use of importance sampling to reduce computational burden in simulation estimation, *Quantitative Marketing and Economics* 7, 343–376.

Adda, Jérôme, and Russell W. Cooper, 2003, *Dynamic Economics: Quantitative Methods and Applications*, MIT Press, Cambridge, MA.

Aldrich, Eric M., Jesus Fernandez-Villaverde, A. Ronald Gallant, and Juan F. Rubio-Ramirez, 2001, Tapping the supercomputer under your desk: Solving dynamic equilibrium models with graphics processors, *Journal of Economic Dynamics and Control* 35, 386–393.

Almeida, Heitor, and Murillo Campello, 2007, Financial constraints, asset tangibility, and corporate investment, *Review of Financial Studies* 20, 1429–1460.

Arellano, Manuel, and Stephen Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies* 58, 277–297.

Campello, Murillo, Antonio Galvao, and Ted Juhl, 2013, Policy heterogeneity in empirical corporate finance, *Working paper*.

Cooper, Russell, and Joao Ejarque, 2003, Financial frictions and investment: Requiem in Q, *Review of Economic Dynamics* 6, 710–728.

Cooper, Russell W., and John C. Haltiwanger, 2006, On the nature of capital adjustment costs, *Review of Economic Studies* 73, 611–633.
Fazzari, Steven M., R. Glenn Hubbard, and Bruce C. Petersen, 1988, Financing constraints and corporate investment, *Brookings Papers on Economic Activity* 1988, 141–195.

Frank, Murray Z., and Vidhan K. Goyal, 2003, Testing the pecking order theory of capital structure, *Journal of Financial Economics* 67, 217–248.

Fuller, Kathleen, Jeffry Netter, and Mike Stegemoller, 2002, What do returns to acquiring firms tell us? Evidence from firms that make many acquisitions, *Journal of Finance* 57, 1763–1793.

Glover, Brent, 2016, The expected cost of default, *Journal of Financial Economics* 119, 284–299.

Gouriéroux, Christian, and Alain Monfort, 1997, *Simulation-Based Econometric Methods*, Oxford University Press, Oxford, UK.

Greene, William H., 2003, *Econometric Analysis*, Fifth Edition, Prentice Hall, Englewood Cliffs, NJ.

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, *Econometrica* pp. 1029–1054.

Hennessy, Christopher A., and Toni M. Whited, 2005, Debt dynamics, *Journal of Finance* 60, 1129–1165.

———, 2007, How costly is external financing? Evidence from a structural estimation, *Journal of Finance* 62, 1705–1745.

Judd, Kenneth L., Lilia Maliar, and Serguei Maliar, 2011, Numerically stable and accurate stochastic simulation approaches for solving dynamic economic models, *Quantitative Economics* 2, 173–210.

Lemmon, Michael L., Michael R. Roberts, and Jaime F. Zender, 2008, Back to the beginning: Persistence and the cross-section of corporate capital structure, *Journal of Finance* 63, 1575–1608.
McFadden, Daniel, 1989, A method of simulated moments for estimation of discrete response models without numerical integration, *Econometrica* 57, 995–1026.

Morellec, Erwan, Boris Nikolov, and Norman Schurhoff, 2012, Corporate governance and capital structure dynamics, *Journal of Finance* 67, 803–848.

Nikolov, Boris, and Toni M. Whited, 2014, Agency conflicts and cash: Estimates from a dynamic model, *Journal of Finance* 69, 1883–1921.

Pakes, Ariel, and David Pollard, 1989, Simulation and the asymptotics of optimization estimators, *Econometrica* 57, 1027–1057.

Richard, Jean-Francois, and Wei Zhang, 2007, Efficient high-dimensional importance sampling, *Journal of Econometrics* 141, 1385–1411.

Robert, Christian P., and George Casella, 2005, *Monte Carlo Statistical Methods*, Springer Texts in Statistics, Springer, New York, NY.

Stokey, Nancy L., Robert E. Lucas, and Edward C. Prescott, 1989, *Recursive Methods in Economic Dynamics*, Harvard University Press, Cambridge, MA.

Strebulaev, Ilya A., and Toni M. Whited, 2012, Dynamic models and structural estimation in corporate finance, *Foundations and Trends in Finance* 6, 1–163.

Tauchen, George, 1986, Finite state Markov-chain approximations to univariate and vector autoregressions, *Economics Letters* 20, 177–181.

Whited, Toni M., 1992, Debt, liquidity constraints, and corporate investment: Evidence from panel data, *Journal of Finance* 47, 1425–1460.

———, and Guojun Wu, 2006, Financial constraints risk, *Review of Financial Studies* 19, 531–559.

Zhang, Lu, 2005, The value premium, *Journal of Finance* 60, 67–103.