Recursivity and the Estimation of Dynamic Games with Continuous Controls

João Granja
Giuseppe Forte

February 15, 2024

Abstract

We revisit the estimation of dynamic games with continuous control variables, such as investments in R&D, quality, and capacity. We show how to use the recursive characterization of Markov Perfect Equilibria (MPE) to develop estimators that exploit the structure of optimal policies. Our preferred estimator uses indirect inference in a two-step procedure that is common in the estimation of dynamic games. We use Monte Carlo experiments based on an empirically-relevant model of investment in R&D with entry and exit to compare the performance of that estimator with alternatives. We find that the indirect inference estimator outperforms the commonly-used inequality estimator of Bajari, Benkard, and Levin (2007) and alternative estimators that exploit the recursive characterization of MPE.

First draft: December 4, 2022.
Granja: University College London, joao.granja@ucl.ac.uk.
Forte: University College London and CeMMAP, uctpgfo@ucl.ac.uk.
We thank Dennis Kristensen, Robert Miller, Lars Nesheim, and Karl Schurter for helpful comments and suggestions.

1 Introduction

Many questions of interest to Industrial Organization economists involve firm choices that have persistent effects on market conditions. Such choices include investments in research and development, the choice of productive capacity, and the choice of product characteristics. Many other examples can be given. Decisions of this type are inherently dynamic and are often taken in industries with few firms. Therefore, their study necessitates the use of dynamic oligopoly models. Furthermore, many of these choices, such as the ones above, are naturally modeled as continuous variables.

This paper revisits the estimation of dynamic oligopoly models with continuous controls. The estimation of such models was made feasible by the seminal contribution of Bajari et al. (2007), henceforth BBL. We make the observation that the main estimator proposed by BBL does not use the full structure of the model, in that it does not exploit the structure of equilibrium policies. Estimators that exploit this structure should exhibit improved econometric performance. We propose estimators that do use the structure of equilibrium policies and conduct Monte Carlo exercises that compare their performance to multiple implementations of BBL.

Our Monte Carlo exercises are based on an extension of Hashmi and van Biesebroeck (2016) – henceforth HvB. HvB propose and estimate an equilibrium model of innovation in the automobile industry. They are interested in the equilibrium relationship between market structure and innovation. In their model, firms engage in R&D – measured by their patenting activity – to increase product quality. We extend the HvB model to allow for firm entry and exit. We base our simulation exercises on the Hashmi and van Biesebroeck (2016) model because it underpins an actual empirical application, and thus accurately represents models used in empirical applications by Industrial Organization economists.

The first step in our proposed estimation routine consists of estimating policy functions and state transitions from the data. This step is similar to BBL and estimators of dynamic games with discrete controls, such as the ones proposed by Aguirregabiria and Mira (2007), Pakes, Ostrovsky, and Berry (2007), and Pesendorfer and Schmidt-Dengler (2008). We depart from BBL in the second step. For a given guess of the structural parameters, we use the estimated policy functions and state transitions to form and solve the maximization problems in the right-hand side of firms' Bellman equations. This yields predicted decisions as a function of structural parameters. We then project these predicted decisions onto a space spanned by basis functions of state variables. Finally, we minimize a measure of the distance between the projections of predicted and observed firm choices. Our estimator thus combines elements of the two-step estimators that sprung from Hotz and Miller (1993) with Indirect Inference estimators à la Gourieroux, Monfort, and Renault (1993). We refer to this estimator as the Recursive Indirect Inference (Rec-II) estimator. We also consider alternative estimators based on recursive equilibrium conditions. Specifically, we consider a nonlinear least squares estimator (Rec-NLLS) and the continuous-control analog of Pesendorfer and Schmidt-Dengler (2008) (Rec-MD). The intuition underpinning the estimators based on recursive equilibrium conditions is simple: under the maintained assumption that the estimated policies constitute a Markov Perfect Equilibrium, solving the right-hand-side of firms' Bellman equations must return the same policy.

We find that the Rec-II estimator has desirable properties. In our Monte Carlo exercise, its finite-sample bias is small and the estimator is precise. We compare it to three implementations of BBL, each using different forms of policy deviations: additive, multiplicative, and what we term asymptotic. The first one uses additive perturbations to the estimated policy function, as in BBL's simulations and in Ryan (2012). The second one uses multiplicative deviations, as in HvB and recommended by Srisuma (2013). The third uses the asymptotic distribution of the empirical policies to construct deviations. We find that all implementations of BBL have substantial finite sample bias.

We relate these findings to the shape of the objective functions that define each estimator. Out estimator's objective function attains its minimum close to the truth and has large curvature. On the other hand, the BBL objective is sometimes minimized far away from the truth, and sometimes it displays no curvature at all around the true parameter. We find that the Rec-NLLS and Rec-MD estimators perform reasonably well, but are dominated by the Rec-II estimator and are more expensive to compute.

Besides its superior econometric performance, the Rec-II estimator also enjoys a practical advantage relative to BBL: it does not require the econometrician to choose policy deviations.[1] This is an advantage, as the performance of the BBL estimator may very well depend on the deviations chosen by the analyst and the literature provides little guidance on how to choose deviations.[2] This

[1] This advantage is shared by all estimators based on recursive equilibrium conditions.

[2] As noted above, some guidance is provided by Srisuma (2013). He suggests that multiplicative

concern is substantiated by the different performance of the three BBL alternatives we consider.

It would be remiss of us not to remind the reader that Bajari et al. (2007) do discuss a second estimator based on solving for firms' optimal policies, albeit in the context of a model without shocks to the cost of investment and with deterministic scrap values. Nevertheless, the empirical literature has converged to using the estimator that BBL discuss at greater length, based on value function inequalities. Indeed, a number of applications, including very recent ones, apply the inequality estimator. These include Ryan (2012), Hashmi and van Biesebroeck (2016), Fowlie, Reguant, and Ryan (2016), and Liu and Siebert (2022). We show how to construct estimators based on firms' optimal policies in environments with random investment cost shocks and random scrap values, and establish that those estimators are computationally tractable for a model that is representative of those used in empirical applications. An important paper that does use firms' optimal policies to estimate a dynamic model is Jofre-Bonet and Pesendorfer (2003). In a dynamic auction model, they show that firms' first-order conditions and the observed distribution of bids identify the distribution of firms' costs. We show that similar ideas extend to the estimation of parameters determining firms' flow profits.

The paper most closely related to our is Srisuma (2013). Srisuma also observes that the BBL inequalities may fail to identify the structural parameters and proposes an estimator that makes use of agents' optimization problems in a two-step procedure. Srisuma's estimator is based on minimizing a distance between the observed conditional distributions of agents' actions and the one implied by agents' (pseudo) maximization problems. This has practical drawbacks. The estimation of the implied distribution is done by simulation. A precise estimator may require the solution of many (static) optimization problems – many more than what we have to solve. Moreover, the implied objective is discontinuous in the structural parameters. These two features make the estimator potentially costly and difficult to compute for the models that practitioners take to data. Indeed, the Monte Carlo simulations in Srisuma (2013) are based on simple static models.

In light of the aforementioned discussion, we view our contributions as threefold. First, we show how to implement estimators based on recursive equilibrium conditions in an empirically relevant setting, including shocks to firms'

deviations have more identifying power than additive ones. We implement both and find their performance to be equally poor.

marginal costs of investment and random scrap values. Second, we show that these estimators can substantially outperform multiple implementations of the commonly-applied BBL inequality estimator. Third, we compare the performance of different estimators based on recursive equilibrium conditions and find support for an estimator based on Indirect Inference. We also show that one can do away with value function simulation and can instead solve for these objects, thus eliminating simulation error.

The rest of the paper is organized as follows. In section 2 we discuss a general model of dynamic competition in an oligopolistic industry. In section 3 we discuss estimators based on recursive equilibrium conditions. In section 4 we provide a brief review of the BBL inequality estimator. In section 5 we introduce the specifics of the Hashmi and van Biesebroeck (2016) model, discuss details of how we implement the different estimators, and present simulation results. Section 6 concludes.

2 The Economic Model

We model the dynamic interaction between oligopolistic competitors. There are $\bar{N}$ firms in the market, including N incumbents and $\bar{N} - N$ potential entrants. $\bar{N}$ is a parameter of the model whereas N is an endogenous variable. Each firm has characteristics $\xi_i \in \Xi$. The set of possible firm characteristics Ξ satisfies $\Xi \subset \mathbb{R} \cup \{-\infty\}$ and $-\infty \in \Xi$, where $-\infty$ represents the firm being inactive. Time is discrete and the horizon is infinite. The state of the industry at time t is $\boldsymbol{\xi}_t = (\xi_{1t}, \ldots, \xi_{\bar{N}t})$.[3]

At the beginning of the period firm i earns flow profit $\pi_i(\boldsymbol{\xi}_t)$. If $\xi_i = -\infty$, then $\pi_i(\boldsymbol{\xi}_t) = 0$. Flow profits are typically modeled as the outcome of competition in static variables such as prices or quantities. We do not need to specify the underlying model that generates π_i. Rather, we treat these functions as parameters of the dynamic game. We assume that the functions π_i are symmetric, i.e., that

$$\pi_i(\xi_i, \xi_2, \ldots, \xi_{i-1}, \xi_1, \xi_{i+1}, \ldots, \xi_{\bar{N}}) = \pi_1(\boldsymbol{\xi}) = \pi(\boldsymbol{\xi}) \quad \text{for all } i = 2, \ldots, \bar{N} \tag{1}$$

and

$$\pi(\boldsymbol{\xi}_1, \boldsymbol{\xi}_{-1}) = \pi(\boldsymbol{\xi}_1, \boldsymbol{\xi}_{p(-1)}) \tag{2}$$

[3]It is straightforward to accommodate exogenous states that capture, e.g., changing demand and/or cost conditions.

for any permutation $p(\cdot)$ of the indices $2, \ldots, \bar{N}$ – see, e.g., Doraszelski and Satterthwaite (2010).[4]

After firms earn profits, incumbents privately observe scrap values $\rho_{it} \in \mathbb{R}_+$ and potential entrants privately observe entry costs $\phi_{it} \in \mathbb{R}_+$. Scrap values and entry costs are iid draws from the distributions F_ρ and F_ϕ, respectively. Upon observing these random variables, firms simultaneously decide whether or not to be active in period $t+1$. We denote the decision to be active by $\alpha_{it} = 1$; choosing not to be active is represented by $\alpha_{it} = 0$. Firms who decide not to be active in the next period perish and are replaced by new potential entrants.

Besides entry and exit decisions, firms also invest to affect the evolution of their characteristics ξ_{it}. Investment choices are denoted by $x_{it} \in \mathbb{R}_+$. After entry and exit decisions are made, all firms that chose to be active in $t+1$ privately observe investment cost shocks $\nu_{it} \in \mathcal{S} \subseteq \mathbb{R}$.

Investment cost shocks are iid draws from the distribution F_ν. Those firms then simultaneously choose their levels of investment and incur investment costs $c(x_{it}, \nu_{it})$.[5]

We assume that the distribution of $\boldsymbol{\xi}_{t+1}$ conditional on $\boldsymbol{\xi}_t$ and firms' action profile $\boldsymbol{a}_t = (a_{1t}, \ldots, a_{\bar{N}t})$, where $a_{it} = (\alpha_{it}, x_{it})$, satisfies

$$F_{\boldsymbol{\xi}}(\boldsymbol{\xi}_{t+1} \mid \boldsymbol{\xi}_t, \boldsymbol{a}_t) = \prod_{i=1}^{\bar{N}} F_\xi(\xi_{t+1} \mid \xi_t, a_{it}) \ . \qquad \text{(Conditional Independence)}$$

Note that we assume that $F_\xi(\xi_{t+1} \mid \xi_t, a_{it})$ is the same for all firms. This assumption makes two restrictions. First, the lack of statistical dependence between firms' future qualities rules out common determinants of firm characteristics. It is possible to accommodate those. Second, the independence between a firm's future characteristics and competitors' current characteristics and actions implies firms cannot directly affect their competitors' characteristics. Such dependence would not raise conceptual difficulties, but we rule it out to simplify the exposition and align with the literature. Finally, note that firms enter/exit if and only if they choose to: for all $\xi_t \in \Xi$ and $x_t \in \mathbb{R}_+$, $F_\xi(-\infty \mid \xi_t, (0, x_t)) = 1$ and

[4]These conditions are sometimes called, respectively, symmetry and anonymity – see, e.g., Doraszelski and Pakes (2007). Doraszelski and Satterthwaite (2010) call a set of functions symmetric if they satisfy both conditions. We adopt their terminology.

[5]We restrict attention to scalar firm characteristics ξ_{it} and investment x_{it}. Accommodating multidimensional characteristics and actions is conceptually straightforward but may generate computational challenges. Perhaps for this reason, we are unaware of papers estimating or studying dynamic games with multiple continuous choices. We thus restrict attention to the scalar case as that covers most (perhaps all) of the literature and saves on notation.

$F_\xi(-\infty \mid \xi_t, (1, x_t)) = 0$. In what follows, we will write $F_\xi(\xi_{t+1} \mid \xi_t, x_t)$ instead of $F_\xi(\xi_{t+1} \mid \xi_t, (1, x_t))$.

2.1 Equilibrium Concept

Given the assumptions we have made imply *ex-ante* firm symmetry, our attention will be directed towards Symmetric Markov Perfect Equilibria (SMPE). The literature has largely focused on symmetric environments and SMPEs due to their computational convenience. The Markov restriction constrains firm behavior to only depend on payoff relevant variables: publicly observed firm characteristics $\boldsymbol{\xi}_t$ and private information $\varepsilon_{it} = (\rho_{it}, \phi_{it}, \nu_{it})$. The symmetry restriction imposes that value and policy functions satisfy conditions analogous to (1) and (2). Under condition (1), it suffices to compute policy and value functions from the perspective of firm 1. We thus focus on firm 1's dynamic programming problems without loss of generality. Moreover, under condition (2) we can compute value and policy functions on a reduced state space. Instead of considering the original state space $\boldsymbol{\Xi} := \Xi^{\bar{N}}$, we can map states that are equivalent from firm 1's perspective onto an arbitrary member of that equivalence class. For instance, suppose $\bar{N} = 3$ and $\Xi = \{-\infty, 1, 2\}$. Then $\boldsymbol{\xi} = (1, 1, 2)$ and $\tilde{\boldsymbol{\xi}} = (1, 2, 1)$ are equivalent from firm 1's perspective. It suffices to compute value and policy functions for one of these two states. We focus on the reduced state space $\boldsymbol{\Xi}^R := \{\boldsymbol{\xi} \in \boldsymbol{\Xi} : \xi_2 \leq \xi_3 \leq \ldots \leq \xi_{\bar{N}}\}$. Given a state $\boldsymbol{\xi} \in \boldsymbol{\Xi}^R$, we denote by $\boldsymbol{\xi}_j$ the corresponding state in $\boldsymbol{\Xi}^R$ from the perspective of firm j, i.e., $\boldsymbol{\xi}_j = (\xi_j, s(\boldsymbol{\xi}_{-j}))$, where ξ_j is the j-th coordinate of $\boldsymbol{\xi}$ and $s(\boldsymbol{\xi}_{-j})$ denotes $\boldsymbol{\xi}_{-j}$ sorted in an increasing order.

In what follows, we will denote a strategy by

$$\sigma(\boldsymbol{\xi}, \varepsilon) = \left(\alpha^I(\boldsymbol{\xi}, \rho), \alpha^E(\boldsymbol{\xi}, \phi), \sigma^x(\boldsymbol{\xi}, \nu)\right) ,$$

where $\varepsilon = (\rho, \phi, \nu)$ and α^I and α^E denote, respectively, the incumbent's and entrant's decision to be active in $t+1$.

2.2 The Incumbent's Problem

Denote by $V_I(\boldsymbol{\xi}, \rho)$ the expected net present value (ENPV) of an incumbent faced with public state $\boldsymbol{\xi}$ and scrap value ρ. Denote by $V_I^A(\boldsymbol{\xi}, \nu)$ the ENPV of an incumbent that has chosen to be active and has observed investment cost shock

ν. Then

$$V_I^A(\boldsymbol{\xi}, \nu) = \max_{x \in \mathbb{R}_+} \left\{ \pi(\boldsymbol{\xi}) - c(x, \nu) + \beta \mathbb{E}\left[V_I(\boldsymbol{\xi}', \rho) \mid \boldsymbol{\xi}, x, \sigma \right] \right\} \tag{3}$$

where

$$\mathbb{E}[V_I(\boldsymbol{\xi}', \rho) \mid \boldsymbol{\xi}, x, \sigma] = \int_{\boldsymbol{\varepsilon}_{-1}} \int_{\boldsymbol{\xi}'} \int_{\rho} V_I(\boldsymbol{\xi}', \rho)\, \mathrm{d}F_\rho\, \mathrm{d}F(\boldsymbol{\xi}' \mid \boldsymbol{\xi}, (1, x), \sigma_{-1}(\boldsymbol{\xi}, \boldsymbol{\varepsilon}_{-1}))\, \mathrm{d}G_{\boldsymbol{\varepsilon}_{-1}} \tag{4}$$

and $\sigma_{-1}(\boldsymbol{\xi}, \boldsymbol{\varepsilon}_{-1}) = (\sigma(\boldsymbol{\xi}_2, \varepsilon_2), \ldots, \sigma(\boldsymbol{\xi}_{\bar{N}}, \varepsilon_{\bar{N}}))$.

Let $\bar{V}_I(\boldsymbol{\xi}) := \int_\rho V_I(\boldsymbol{\xi}, \rho)\, \mathrm{d}F_\rho$. By conditional independence

$$\int_{\boldsymbol{\xi}'} \bar{V}_I(\boldsymbol{\xi}')\, \mathrm{d}F(\boldsymbol{\xi}' \mid \boldsymbol{\xi}, (1, x), \sigma_{-1}(\boldsymbol{\xi}, \boldsymbol{\varepsilon}_{-1})) = \int_{\xi_1'} W(\xi_1' \mid \boldsymbol{\xi}, \boldsymbol{\varepsilon}_{-1}, \sigma)\, \mathrm{d}F(\xi_1' \mid \xi_1, x)$$

where

$$W(\xi_1' \mid \boldsymbol{\xi}, \boldsymbol{\varepsilon}_{-1}, \sigma) = \int_{\xi_2'} \cdots \int_{\xi_{\bar{N}}'} \bar{V}_I(\xi_1', \boldsymbol{\xi}_{-1}')\, \mathrm{d}F(\xi_{\bar{N}}' \mid \xi_{\bar{N}}, \sigma(\boldsymbol{\xi}_{\bar{N}}, \varepsilon_{\bar{N}})) \ldots \mathrm{d}F(\xi_2' \mid \xi_2, \sigma(\boldsymbol{\xi}_2, \varepsilon_2))$$

is the incumbent's ENPV of starting a period with characteristic ξ_1' given $\boldsymbol{\xi}$ and $\boldsymbol{\varepsilon}_{-1}$ and that its competitors behave according to σ.

Under the additional assumption of independence of the ε_i, the integral with respect to $\boldsymbol{\varepsilon}_{-1}$ can be written as a multiple integral. Changing the order of integration in (4), we have multiple terms of the form

$$\int_{\varepsilon_j} \int_{\xi_j'} \bar{V}_I(\boldsymbol{\xi}') \mathrm{d}F(\xi_j' \mid \xi_j, \sigma(\boldsymbol{\xi}_j, \varepsilon_j))\, \mathrm{d}G_{\varepsilon_j} = \int_{\xi_j'} \bar{V}_I(\boldsymbol{\xi}')\, \mathrm{d}F^\sigma(\xi_j' \mid \boldsymbol{\xi}_j) \tag{5}$$

where

$$F^\sigma(\xi_j' \mid \boldsymbol{\xi}_j) := \int_{\varepsilon_j} F(\xi_j' \mid \xi_j, \sigma(\boldsymbol{\xi}_j, \varepsilon_j))\, \mathrm{d}G_{\varepsilon_j}\ . \tag{6}$$

It is clear that that equation (5) holds when $F(\xi' \mid \xi, x)$ has a density for all (ξ, x) or Ξ is finite. Appendix A.1 establishes equation (5) for general $F(\xi' \mid \xi, x)$ under a technical condition. This implies that the ensuing results hold for continuous, discrete, and discrete-continuous public state transition processes.

In summary, we have

$$\mathbb{E}[V_I(\boldsymbol{\xi}', \rho) \mid \boldsymbol{\xi}, x, \sigma] = \int_{\xi_1'} W(\xi_1' \mid \boldsymbol{\xi}, F^\sigma)\, \mathrm{d}F(\xi_1' \mid \xi_1, x)\ ,$$

where

$$W(\xi_1' \mid \boldsymbol{\xi}, F^\sigma) = \int_{\xi_2'} \cdots \int_{\xi_{\bar{N}}'} \bar{V}_I(\xi_1', \boldsymbol{\xi}_{-1}') \, \mathrm{d}F^\sigma(\xi_{\bar{N}}' \mid \boldsymbol{\xi}_{\bar{N}}) \ldots \mathrm{d}F^\sigma(\xi_2' \mid \boldsymbol{\xi}_2) \tag{7}$$

and F^σ is given by (6).[6]

The first-order condition of the maximization problem on the right-hand side of (3) is thus

$$-\frac{\partial c(x, \nu)}{\partial x} + \beta \, \frac{\partial}{\partial x} \left(\int_{\xi_1'} W(\xi_1' \mid \boldsymbol{\xi}, F^\sigma) \, \mathrm{d}F(\xi_1' \mid \xi_1, x) \right) \leq 0 \, ,$$

with equality if the solution is interior. It is desirable for the investment first-order condition to be sufficient for an optimum. Sufficiency is useful both in equilibrium computation and in estimation based on firms' optimal policies. To establish sufficiency of the investment first-order condition, we will make the following assumption.

Assumption 1. Let Ξ be a compact subset of the real line, and denote its minimum and maximum by ξ_m and ξ_M. Let Ξ° be the interior of Ξ (understood in the discrete case as $\Xi \setminus \{\xi_m, \xi_M\}$). The family of distributions $F(\cdot \mid \xi, x)$ is such that, for all $\xi \in \Xi$ and $\xi' \in \Xi^\circ$,

(a) $F(\xi' \mid \xi, x)$ is strictly decreasing and strictly convex in x;

(b) $F(\xi_m \mid \xi, x)$ is decreasing and convex in x.

Assumption 1(a) states that for any current and future characteristics (other than the endpoints of Ξ), an increase in investment x causes the cumulative distribution function to decrease; i.e., an increase in x increases a firm's distribution of future quality in the first-order stochastic dominance sense. Moreover, investment has decreasing marginal returns in the sense that the reduction in the CDF is decreasing in investment. Assumption 1(b) allows for the continuous case, where $F(\xi_m \mid \xi, x) = 0$ for all ξ and x, and for the discrete case in which $F(\xi_m \mid \xi, x)$ may be strictly positive and depend on x.

Proposition 1. Assume

(a) $c(x, \nu)$ is convex in x;

[6] We abuse notation slightly by denoting the integral in (7) by $W(\xi_1' \mid \boldsymbol{\xi}, F^\sigma)$ when we have already defined $W(\xi_1' \mid \boldsymbol{\xi}, \boldsymbol{\varepsilon}_{-1}, \sigma)$ above. We will, however, have no further need for the latter, and we thus retain the former.

(b) $F(\xi' \mid \xi, x)$ is twice continuously differentiable in x;

(c) $W(\xi' \mid \boldsymbol{\xi}, F^{\sigma})$ is increasing in ξ' for all $\boldsymbol{\xi}$.

Then under Assumption 1

$$v(x; \boldsymbol{\xi}, \nu) = \pi(\boldsymbol{\xi}) - c(x, \nu) + \beta \int_{\xi_1'} W(\xi_1' \mid \boldsymbol{\xi}, F^{\sigma}) \, \mathrm{d}F(\xi_1' \mid \xi_1, x)$$

is strictly concave in x.

Proof. See Appendix A.2. □

We make three assumptions in addition to Assumption 1. First, we assume that the cost function is convex in investment. This is a standard restriction. Second, we impose twice continuous differentiability of the transition functions. This is a technical condition required in the proof, but has little economic content. It is also nonrestrictive in practice, as in applications the econometrician typically imposes a parametric restriction on $F(\xi' \mid \xi, x)$ that satisfy this condition. Finally, we assume that the ENPV of starting the following period at ξ' conditional on $\boldsymbol{\xi}$ is increasing in ξ': that is, firms would rather start the following period from a higher rather than lower ξ. This is an assumption on an equilibrium object, but we see it as a mild one.[7] Indeed, all equilibria of the environment in Section 5 that we have computed exhibit this property.

2.2.1 The Incumbent's Exit Decision

The incumbent commits to an exit decision before observing investment cost shock ν. Therefore, it must make its decision on the basis of its expected continuation value conditional on it being active:

$$\bar{V}_I^A(\boldsymbol{\xi}) := \int V_I^A(\boldsymbol{\xi}, \nu) \, \mathrm{d}F_{\nu} \,. \tag{8}$$

Recall $V_I(\boldsymbol{\xi}, \rho)$ denotes the ENPV of an incumbent with scrap value ρ. Then

$$V_I(\boldsymbol{\xi}, \rho) = \max \left\{ \pi(\boldsymbol{\xi}) + \rho, \bar{V}_I^A(\boldsymbol{\xi}) \right\} = \max_{\chi \in \{0,1\}} \chi \bar{V}_I^A(\boldsymbol{\xi}) + (1 - \chi)[\pi(\boldsymbol{\xi}) + \rho] \,. \tag{9}$$

[7] This is a mild assumption for models that consider ξ's with a "reasonable" payoff ordering, such as the quality ladder model considered in Section 5.

The implied conditional probability of an incumbent remaining active is

$$\mathbb{P}\big(\alpha^I(\boldsymbol{\xi},\rho) = 1 \mid \boldsymbol{\xi}\big) = F_\rho\big(\bar{V}_I^A(\boldsymbol{\xi}) - \pi(\boldsymbol{\xi})\big) \tag{10}$$

2.3 The Entrant's Problem

Denote by $V_E^A(\boldsymbol{\xi}_{-1},\nu)$ the ENPV of a potential entrant that enters under public state $\boldsymbol{\xi}_{-1}$ and draws investment cost shock ν. This function is characterized by

$$V_E^A(\boldsymbol{\xi}_{-1},\nu) = \max_{x\in\mathbb{R}_+}\left\{-c(x,\nu)+\beta\int_{\xi_1'} W(\xi_1' \mid (-\infty,\boldsymbol{\xi}_{-1}),F^\sigma)\,\mathrm{d}F(\xi_1' \mid \xi_e,x)\right\}, \tag{11}$$

where $\xi_e \in \Xi$ is an exogenously specified initial quality level for potential entrants. In our simulations below we assume that $\xi_e = \min(\Xi \setminus -\infty)$.

Potential entrants either enter the market or perish. Therefore, their ENPV given entry cost ϕ is

$$V_E(\boldsymbol{\xi}_{-1},\phi) = \max\left\{0, \bar{V}_E^A(\boldsymbol{\xi}_{-1}) - \phi\right\} = \max_{\chi\in\{0,1\}} \chi\big[\bar{V}_E^A(\boldsymbol{\xi}_{-1}) - \phi\big] \tag{12}$$

where we have normalized the value of entrants' outside option to zero and

$$\bar{V}_E^A(\boldsymbol{\xi}_{-1}) := \int V_E^A(\boldsymbol{\xi}_{-1},\nu)\,\mathrm{d}F_\nu\ . \tag{13}$$

The conditional probability of entry is

$$\mathbb{P}\big(\alpha^E(\boldsymbol{\xi}_{-1},\phi) = 1 \mid \boldsymbol{\xi}_{-1}\big) = F_\phi\big(\bar{V}_E^A(\boldsymbol{\xi}_{-1})\big) \tag{14}$$

2.4 Equilibrium

Definition 1. Let $\Xi_I := (\Xi \setminus \{-\infty\}) \times \Xi^{\bar{N}-1}$. A Symmetric Markov Perfect Equilibrium (SMPE) is a pair $(\bar{V}_I,\sigma)$ where $\bar{V}_I : \Xi_I \to \mathbb{R}$ and $\sigma = (\sigma^x,\alpha^E,\alpha^I)$ are such that

1. $\sigma^x(\boldsymbol{\xi},\nu)$ solves the right-hand side of

$$V^A(\boldsymbol{\xi},\nu) = \max_{x\in\mathbb{R}_+}\left\{\pi(\boldsymbol{\xi}) - c(x,\nu) + \beta\int_{\xi_1'} W(\xi_1' \mid \boldsymbol{\xi},F^\sigma)\,\mathrm{d}F(\xi_1' \mid \xi_1,x)\right\} \tag{15}$$

for all $\boldsymbol{\xi} \in \Xi^N$ and ν in the support of F_ν, where $W(\xi' \mid \boldsymbol{\xi},F^\sigma)$ is given by equation (7).

2. $\alpha^I(\boldsymbol{\xi}, \rho)$ solves problem (9) subject to (8) and (15), for all $\boldsymbol{\xi} \in \Xi_I$ and ρ in the support of F_ρ.

3. $\alpha^E(\boldsymbol{\xi}_{-1}, \phi)$ solves problem (12) subject to (13) and (11), for all $\boldsymbol{\xi}_{-1} \in \times\Xi^{\bar{N}-1}$ and ϕ in the support of F_ϕ.

4. For all $\boldsymbol{\xi} \in \Xi_I$, $\bar{V}_I(\boldsymbol{\xi}) = \int\{\alpha^I(\boldsymbol{\xi}, \rho)\bar{V}_I^A(\boldsymbol{\xi}) + (1 - \alpha^I(\boldsymbol{\xi}, \rho))[\pi(\boldsymbol{\xi}) + \rho]\}\ \mathrm{d}F_\rho$ where $\bar{V}_I^A(\boldsymbol{\xi})$ is given by (8).

We define an equilibrium only in terms of incumbents' integrated value functions. This is due to the assumption, common in this literature, that firms that choose to be inactive in the following period perish. As a result, when a firm's quality becomes $\xi' = -\infty$ (as a consequence of exit or no entry), that firm's continuation value is zero. Therefore, $\bar{V}_I$ alone is sufficient to determine firm behavior.[8]

We compute SMPEs as follows. We start with guesses for $\bar{V}_I(\boldsymbol{\xi})$ and $F^\sigma(\xi' \mid \boldsymbol{\xi})$. With these two objects we can compute $W(\xi' \mid \boldsymbol{\xi}, F^\sigma)$. We then solve for firms' optimal investment and entry and exit decisions, i.e., we perform the computations associated with conditions 1 to 3 in Definition 1. We update F^σ based on these decisions.[9] The update to $\bar{V}_I(\boldsymbol{\xi})$ follows from rewriting condition 4 of Definition 1 as

$$\bar{V}_I(\boldsymbol{\xi}) = F_\rho(\bar{V}_I^A(\boldsymbol{\xi}) - \pi(\boldsymbol{\xi}))[\bar{V}_I^A(\boldsymbol{\xi}) - \pi(\boldsymbol{\xi})] + \pi(\boldsymbol{\xi}) + \int_{\bar{V}_I^A(\boldsymbol{\xi}) - \pi(\boldsymbol{\xi})}^{\infty} \rho\ \mathrm{d}F_\rho\,. \tag{16}$$

For suitable distributions, e.g. the lognormal, the integral in this equation can be written in closed form. We iterate on these steps until both $\bar{V}_I$ and F^σ converge. As the firms' objective function depends solely and continuously on these objects, the Theorem of the Maximum implies that the policy functions associated with the firms' problem also converge. By the symmetry assumption, it is sufficient to compute $\bar{V}_I(\boldsymbol{\xi})$ and $F^\sigma(\xi' \mid \boldsymbol{\xi})$ for $\boldsymbol{\xi} \in \boldsymbol{\Xi}^R$. As shown in Pakes and McGuire (1994), the reduced state space $\boldsymbol{\Xi}^R$ grows in the number of firms as a polynomial of order $|\Xi|$ rather than exponentially. In the numerical examples presented in Section 5, symmetry reduces the cardinality of the state space from $16^5 = 1,048,576$ to $62,016$.

[8]Note also that we do not define $\bar{V}_I$ on $\boldsymbol{\Xi}^R$ but rather on Ξ_I. Similarly we do not define the policy functions on the reduced state space. This is for the sake of precision, as strategies must be complete contingent plans. However, the discussion in section 2.1 applies. In particular, when computing equilibria we do exploit symmetry, as discussed in this section.

[9]Updating F^σ involves an integral with respect to F_ν – see equation (6). The choice of integral approximant will determine the values of ν for which we compute firms' optimal investment choices.

3 Estimation via Recursive Equilibrium Conditions

In this section we discuss estimators based on firms' optimal behavior, i.e. behaviour consistent with conditions 1 to 3 in Definition 1. For the purpose of discussing estimation we let the investment cost function depend on parameters $\boldsymbol{\theta}_x$, the distribution of scrap values depend on parameters $\boldsymbol{\theta}_\rho$, and the distribution of entry costs depend on parameters $\boldsymbol{\theta}_\phi$. We denote these parameters collectively by $\boldsymbol{\theta}$. We also use the notation $\boldsymbol{\theta}_{-\phi} := (\boldsymbol{\theta}_x, \boldsymbol{\theta}_\rho)$. Our goal is to estimate $\boldsymbol{\theta}$.

Suppose we can obtain an estimate (up to parameters) $\widehat{\bar{V}}_I(\boldsymbol{\xi};\boldsymbol{\theta}_{-\phi})$ of the ex-ante value function $\bar{V}_I(\boldsymbol{\xi})$.[10] Suppose we also have estimates of $F(\xi' \mid \xi, x)$ and $F^\sigma(\xi' \mid \boldsymbol{\xi})$, defined in equation (6). These estimates allow us to set up an empirical analog of firms' investment problem:

$$\max_{x\in\mathbb{R}_+}\left\{\pi(\boldsymbol{\xi}) - c(x,\nu;\theta_x) + \beta\int_{\xi_1'}\widehat{W}(\xi_1' \mid \boldsymbol{\xi},\boldsymbol{\theta}_{-\phi})\,\mathrm{d}\hat{F}(\xi_1' \mid \xi_1, x)\right\}, \tag{17}$$

where $\widehat{W}(\xi_1' \mid \boldsymbol{\xi},\boldsymbol{\theta}_{-\phi})$ is given by (7) substituting $\widehat{\bar{V}}_I(\boldsymbol{\xi};\boldsymbol{\theta}_{-\phi})$ for $\bar{V}_I(\boldsymbol{\xi})$ and $\hat{F}^\sigma$ for F^σ. The solution to problem (17) yields a predicted level of investment at state $(\boldsymbol{\xi},\nu)$ under $\boldsymbol{\theta}_{-\phi}$. Multiple solutions over the state space provide an estimate of $\bar{V}_I^A(\boldsymbol{\xi})$ and, via (10) and (14), the probability of exit by incumbents and entry by inactive firms.

All recursive estimators compute model-implied investment and probabilities of entry and exit in this way. They differ in how they exploit that information to estimate $\boldsymbol{\theta}$. Our numerical results have led us to favor an Indirect Inference estimator, which we now discuss.

3.1 Indirect Inference

3.1.1 Investment

We predict investment as the solution to (17) for $\boldsymbol{\xi}$'s observed in the data and randomly drawn $\nu \sim F_\nu$. We take K draws from F_ν for each of the M observations, for a total of KM simulated investment decisions. Denote these by

[10] We discuss alternative estimators of $\bar{V}_I(\boldsymbol{\xi})$ in Section 3.2. Observe that our notation indicates that these value-function estimates do not depend on $\boldsymbol{\theta}_\phi$. We discuss why below.

$\sigma^x(\boldsymbol{\xi}_j, \nu_j; \boldsymbol{\theta}_{-\phi})$, $j = 1, \ldots, KM$. We estimate the linear specification

$$\sigma^x(\boldsymbol{\xi}_j, \nu_j; \boldsymbol{\theta}_{-\phi}) = \sum_{k=1}^{B^x} \lambda_k^x \Psi_k^x(\boldsymbol{\xi}_j) + \zeta_j^x \qquad j = 1, \ldots, KM. \tag{18}$$

for a set of basis functions chosen by the econometrician $\{\Psi_k^x\}_{k=1}^{B^x}$. This yields

$$\hat{\boldsymbol{\lambda}}^x(\boldsymbol{\theta}_{-\phi}) := \begin{bmatrix} \hat{\lambda}_1^x(\boldsymbol{\theta}_{-\phi}) & \ldots & \hat{\lambda}_{B^x}^x(\boldsymbol{\theta}_{-\phi}) & S_{\zeta^x}(\boldsymbol{\theta}_{-\phi}) \end{bmatrix}$$

where $S_\zeta^x = \left[(N - B^x)^{-1} \sum_{j=1}^M (\sigma^x(\boldsymbol{\xi}_j, \nu_j; \boldsymbol{\theta}_{-\phi}) - \sum_{k=1}^{B^x} \hat{\lambda}_k^x \Psi_k^x(\boldsymbol{\xi}_j))^2\right]^{1/2}$ is an estimate of the standard deviation of ζ_j^x.

Turning to investment decisions observed in the data $\{x_i\}_{i=1}^M$, we estimate the linear specification

$$x_i = \sum_{k=1}^{B^x} \gamma_k^x \Psi_k^x(\boldsymbol{\xi}_i) + \eta_i^x \quad i = 1, \ldots, M. \tag{19}$$

to obtain $\hat{\boldsymbol{\gamma}}^x := \left[\hat{\gamma}_1^x, \ldots, \hat{\gamma}_{B^x}^x, S_\eta^x\right]$.

3.1.2 Exit

We compute the model-implied probability that firms choose to exit in state $\boldsymbol{\xi}$ when the structural parameters are $\boldsymbol{\theta}_{-\phi}$:

$$\mathbb{P}(\text{Exit} \mid \boldsymbol{\xi}; \boldsymbol{\theta}_{-\phi}) = 1 - F_\rho\left(\widehat{\overline{V}}_I^A(\boldsymbol{\xi}; \boldsymbol{\theta}_{-\phi}) - \pi(\boldsymbol{\xi}); \boldsymbol{\theta}_\rho\right) . \tag{20}$$

where $\widehat{\overline{V}}_I^A(\boldsymbol{\xi}; \boldsymbol{\theta}_{-\phi})$, the integral of $\hat{V}_I^A(\boldsymbol{\xi}, \nu; \boldsymbol{\theta}_{-\phi})$ with respect to ν, is estimated by averaging over problem (17) solved on a carefully selected vector of ν's.

We use these conditional probabilities to estimate the linear specification

$$\mathbb{P}(\text{Exit} \mid \boldsymbol{\xi}; \boldsymbol{\theta}_{-\phi}) = \sum_{k=1}^{B^E} \gamma_k^E \Psi_k^E(\boldsymbol{\xi}_i) + \zeta_i^E \ , \tag{21}$$

for basis functions $\{\Psi_k^E(\boldsymbol{\xi})\}_{k=1}^{B^E}$. This yields

$$\hat{\boldsymbol{\lambda}}^E(\boldsymbol{\theta}_{-\phi}) = \left[\hat{\lambda}_1^E(\boldsymbol{\theta}_{-\phi}) \ldots, \hat{\lambda}_{B^E}^E(\boldsymbol{\theta}_{-\phi})\right]$$

Turning to the data, we estimate the linear probability model

$$\mathbb{1}\{\xi_i' = -\infty\} = \sum_{k=1}^{B^E} \gamma_k^E \Psi_k^E(\boldsymbol{\xi}_i) + \eta_i^E \ , \tag{22}$$

using data on active firms only. This yields $\hat{\boldsymbol{\gamma}}^E = \begin{bmatrix} \hat{\gamma}_1^E & \dots & \hat{\gamma}_{B^E}^E \end{bmatrix}$. Note that here we do not keep track of an estimate of the standard deviation of the error term. We use a linear probability model because it is simple to estimate.

It may seem more natural to simulate exit decisions and use those to estimate (22) on simulated data. However, the discreteness of the outcome variable would cause $\hat{\boldsymbol{\lambda}}^E(\boldsymbol{\theta}_{-\phi})$ to be discontinuous in $\boldsymbol{\theta}_{-\phi}$. Moreover, we know from the theory of linear regression that both estimators have the same probability limit.[11]

3.1.3 Entry

Our treatment of entry decisions is analogous to how we handle exit decisions. Having estimated $\widehat{\overline{V}}_E^A(\boldsymbol{\xi}_{-1})$ by solving (17) repeatedly to integrate $\widehat{V}_E^A(\boldsymbol{\xi}_{-1}, \nu)$ with respect to ν, we compute

$$\mathbb{P}(\text{Entry} \mid \boldsymbol{\xi}, \boldsymbol{\theta}) = F_\phi\left(\widehat{\overline{V}}_E^A(\boldsymbol{\xi}_{-1}; \boldsymbol{\theta}_{-\phi}); \boldsymbol{\theta}_\phi\right) \tag{23}$$

and estimate the linear specification

$$\mathbb{P}(\text{Entry} \mid \boldsymbol{\xi}; \boldsymbol{\theta}) = \sum_{k=1}^{B^N} \gamma_k^N \Psi_k^N(\boldsymbol{\xi}_i) + \zeta_i^N \ , \tag{24}$$

for basis functions $\{\Psi_k^N(\boldsymbol{\xi})\}_{k=1}^{B^N}$ (the superscript N indicating entry) to obtain

$$\hat{\boldsymbol{\lambda}}^N(\boldsymbol{\theta}) = \left[\hat{\lambda}_1^N(\boldsymbol{\theta}) \dots \hat{\lambda}_{B^N}^N(\boldsymbol{\theta})\right]$$

[11] Indeed, let $e = Y - \mathbb{E}[Y \mid X]$. Then

$$(X'X)^{-1}X'Y = (X'X)^{-1}X'[\mathbb{E}[Y \mid X] + e] = (X'X)^{-1}X'\mathbb{E}[Y \mid X] + o_p(1) \ ,$$

where the last equality follows from $\mathbb{E}[e \mid X] = 0$. Also see Collard-Wexler (2013) for a use of conditional expectations of discrete choices in Indirect Inference estimation of a dynamic discrete game.

Turning to the data, we estimate the linear probability model

$$\mathbb{1}\{\xi_i' > -\infty \mid \xi_i = -\infty\} = \sum_{k=1}^{B^E} \gamma_k^E \Psi_k^E(\boldsymbol{\xi}_i) + \eta_i^E \ , \tag{25}$$

to obtain $\hat{\boldsymbol{\gamma}}^N = \begin{bmatrix} \hat{\gamma}_1^N & \dots & \hat{\gamma}_{B^N}^N \end{bmatrix}$.

3.1.4 Estimation Problem

For a positive definite weight matrix Ω, the Indirect Inference estimator is

$$\hat{\boldsymbol{\theta}}_{II} = \arg\min_{\boldsymbol{\theta} \in \Theta} [\hat{\boldsymbol{\lambda}}(\boldsymbol{\theta}) - \hat{\gamma}]' \Omega [\hat{\boldsymbol{\lambda}}(\boldsymbol{\theta}) - \hat{\gamma}] \tag{26}$$

where

$$\hat{\boldsymbol{\lambda}}(\boldsymbol{\theta}) = \begin{bmatrix} \hat{\boldsymbol{\lambda}}^x(\boldsymbol{\theta}_{-\phi})' & \hat{\boldsymbol{\lambda}}^E(\boldsymbol{\theta}_{-\phi})' & \hat{\boldsymbol{\lambda}}^N(\boldsymbol{\theta})' \end{bmatrix}' \quad \text{and} \quad \hat{\gamma} = \begin{bmatrix} (\hat{\boldsymbol{\gamma}}^x)' & (\hat{\boldsymbol{\gamma}}^E)' & (\hat{\boldsymbol{\gamma}}^N)' \end{bmatrix}'$$

Note that the objective function and the estimator both depend on $\hat{\gamma}$, Ω, and the first-stage estimates $\widehat{\overline{V}}_I(\boldsymbol{\xi}; \boldsymbol{\theta}_{-\phi})$, $\hat{F}(\xi' \mid \xi, x)$, and $\hat{F}^\sigma(\xi' \mid \boldsymbol{\xi})$. We do not highlight that dependence in our notation as this and subsequent sections do not concern themselves with the dependence of the estimator on those objects. We refer to the estimator defined by (26) as the Recursive Indirect Inference estimator, or Rec-II for short.[12]

The intuition for the Rec-II estimator is the following.[13] If the policies observed in the data are a Markov Perfect Equilibrium, they must satisfy the recursive equilibrium conditions in Definition 1. Therefore, if we solve the right-hand side of the Bellman Equation using transitions and value functions implied by $\hat{\sigma}$, we must obtain $\hat{\sigma}$ back. One could try to estimate the structural parameters by directly matching the model-implied policies and $\hat{\sigma}$. Along these lines, Srisuma (2013) proposes matching the observed conditional distribution of investment and that implied by firms' optimality conditions. Unfortunately, computing the model-implied conditional distribution is computationally very demanding. Moreover, the resulting estimator is based on a non-smooth objective function because the simulated distribution is not smooth for a finite

[12] As usual, using the analytical gradient of the objective function in problem (26) significantly accelerates the computation of the solution to that problem. We provide the required calculations in Appendix A.4.

[13] As discussed in the introduction, this intuition underpins not only the Rec-II estimator but all estimators based on recursive equilibrium conditions.

number of simulated draws. The Rec-II estimator instead matches features of observed and model-implied distributions, which yields a smooth objective. The estimator is sufficiently cheap to compute in empirically-relevant models and, as we show in section 5, can substantially outperform commonly-used alternatives.

We close this section by discussing alternative recursive estimators. Focusing on investment, a simple alternative is the nonlinear least squares estimator

$$\hat{\boldsymbol{\theta}}_{NLLS} := \arg\min_{\theta} \sum_{i=1}^{N} \left(x_i - \mathbb{E}_\nu[\sigma^x(\boldsymbol{\xi}, \nu; \boldsymbol{\theta}_x) \mid \boldsymbol{\xi}]\right)^2 . \tag{27}$$

Yet another alternative estimator would minimize the distance between the conditional expectation of investment and its model-predicted equivalent, i.e.

$$\hat{\boldsymbol{\theta}}_{ALS} := \arg\min_{\theta} \sum_{\boldsymbol{\xi}} \left(\mathbb{E}[x_i \mid \boldsymbol{\xi}] - \mathbb{E}_\nu[\sigma^x(\boldsymbol{\xi}, \nu; \boldsymbol{\theta}_x) \mid \boldsymbol{\xi}]\right)^2 , \tag{28}$$

which can be seen as the continuous-control analog of Pesendorfer and Schmidt-Dengler (2008). Furthermore, interpreting (28) as a method of moments estimator allows for extensions to entry and exit by adding conditions that match empirical and model-implied entry and exit moments.[14]

We prefer $\hat{\boldsymbol{\theta}}_{II}$ for two reasons. First, $\hat{\boldsymbol{\theta}}_{II}$ is computationally cheaper than $\hat{\boldsymbol{\theta}}_{NLLS}$ and $\hat{\boldsymbol{\theta}}_{ALS}$, as it requires fewer evaluations of $\sigma^x(\boldsymbol{\xi}, \nu; \boldsymbol{\theta}_x)$ per observation than would be needed to approximate the integral $\mathbb{E}_\nu[\sigma^x(\boldsymbol{\xi}, \nu; \boldsymbol{\theta}_x) \mid \boldsymbol{\xi}]$ to satisfactory precision. Second we found $\hat{\boldsymbol{\theta}}_{II}$ outperforms both $\hat{\boldsymbol{\theta}}_{NLLS}$ and $\hat{\boldsymbol{\theta}}_{ALS}$ in a model without entry and exit. We report results for this simpler model in Appendix D.

3.2 Estimating Integrated Value Functions

In this section we present a closed form solution for $\bar{V}_I(\boldsymbol{\xi}; \boldsymbol{\theta}_{-\phi})$ under the assumption that $|\Xi| < \infty$. We use this closed form in recursive estimators as an alternative to the forward simulation used in BBL.

Let $P(\xi' \mid \xi, a)$ denote the probability that a firm's quality in $t+1$ is ξ' conditional on its current quality being ξ and its action being $a = (\alpha, x)$, where the notation is analogous to that in Section 2. Moreover, let Ξ_I^R denote the set of

[14] This is related to the "aggregate moments" estimator described in Bajari et al. (2007, p. 1363), though (28) matches conditional average investment whereas BBL suggest pooling across states.

states in the reduced state space in which firm 1 is active, i.e., $\Xi_I^R := \{\boldsymbol{\xi} \in \Xi^R : \xi_1 > -\infty\}$. Let $\bar{\boldsymbol{V}}_I = [\bar{V}_I(\boldsymbol{\xi}) : \boldsymbol{\xi} \in \Xi_I^R]$ be a vector stacking the incumbents' integrated value function across states in Ξ_I^R. We show in appendix A.3 that $\bar{\boldsymbol{V}}_I$ satisfies

$$\bar{\boldsymbol{V}}_I = \boldsymbol{\pi} - \boldsymbol{K}(\boldsymbol{\theta}_x) + \boldsymbol{\Sigma}(F_\rho) + \beta \boldsymbol{M}(\boldsymbol{P})\bar{\boldsymbol{V}}_I \tag{29}$$

where

$$\boldsymbol{K}(\boldsymbol{\theta}_x) = \left[\mathbb{P}_I^A(\boldsymbol{\xi}) \int c(\sigma^x(\boldsymbol{\xi}, \nu), \nu; \boldsymbol{\theta}_x)\, \mathrm{d}F_\nu \ : \ \boldsymbol{\xi} \in \Xi_I^R\right] \tag{30}$$

$$\boldsymbol{\Sigma}(F_\rho) = \left[[1 - \mathbb{P}_I^A(\boldsymbol{\xi})]\mathbb{E}[\rho \mid \rho > F_\rho^{-1}(\mathbb{P}_I^A(\boldsymbol{\xi})); \boldsymbol{\theta}_\rho] \ : \ \boldsymbol{\xi} \in \Xi_I^R\right] \tag{31}$$

where $\mathbb{P}_I^A(\boldsymbol{\xi}) = \mathbb{P}(\alpha^I(\boldsymbol{\xi}, \rho) = 1 \mid \boldsymbol{\xi})$ is the probability that the incumbent chooses to be active in state $\boldsymbol{\xi}$ and $\boldsymbol{M}(\boldsymbol{P})$ is the transition matrix implied by the policy function σ, i.e.,[15]

$$\boldsymbol{M}(\boldsymbol{P}) = \left[\mathbb{P}^\sigma(\boldsymbol{\xi}_l \mid \boldsymbol{\xi}_k) \ : \ 1 \le l, k \le |\Xi_I^R|\right] \tag{32}$$

where

$$\mathbb{P}^\sigma(\boldsymbol{\xi}'|\boldsymbol{\xi}) = \prod_{j=1}^{N} P^\sigma(\xi_j' \mid \boldsymbol{\xi}) = \prod_{j=1}^{N} \int P(\xi_j' \mid \xi_j, \sigma(\boldsymbol{\xi}_j, \varepsilon))\, \mathrm{d}G_\varepsilon \ . \tag{33}$$

Equations (29) to (33) imply that we can estimate $\bar{\boldsymbol{V}}_I$ up to parameters by estimating the probabilities $\mathbb{P}_I^A(\boldsymbol{\xi})$ and $P^\sigma(\xi' \mid \boldsymbol{\xi})$ and the investment policy function $\sigma^x(\boldsymbol{\xi}, \nu)$. The probabilities are directly estimable from the data on firms characteristics. To identify and estimate $\sigma^x(\boldsymbol{\xi}, \nu)$ we follow Bajari et al. (2007). Their argument implies, in the case in which $\sigma^x(\boldsymbol{\xi}, \nu)$ is decreasing in ν, that

$$\sigma^x(\boldsymbol{\xi}, \nu) = F_X^{-1}(1 - F_\nu(\nu) \mid \boldsymbol{\xi}) \ , \tag{34}$$

where $F_X(x \mid \boldsymbol{\xi})$ is the distribution of investment conditional on $\boldsymbol{\xi}$, which is identified. That is, the policy function is identified by the quantiles of the conditional distribution of investment.[16] Using this, the integral in equation (30)

[15] We order states in Ξ^R and Ξ_I^R lexicographically, where firm 1 takes precedence over firm 2, who takes precedence over firm 3, so on and so forth. We do so by interpreting $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_{\bar{N}})$ as a number in base $|\Xi|$.

[16] This argument rests on the maintained assumption that F_ν is known. If F_ν is known up to parameters θ_ν, (34) must account for that dependence. This in turn implies that $\boldsymbol{K}(\boldsymbol{\theta}_x)$ in equation (30) also depends on $\boldsymbol{\theta}_\nu$. It can then be seen from equation (29) that the integrated value function ceases to be linear in the structural parameters *even when* the investment cost function is linear in $\boldsymbol{\theta}_x$ and the scrap value is deterministic (as in Bajari et al. (2007)).

can be approximated by $\sum_{i=1}^{N} \omega_i c(F_X^{-1}(1 - F_\nu(\nu_i) \mid \boldsymbol{\xi}), \nu_i; \boldsymbol{\theta}_x)$, for judiciously chosen weights ω_i and nodes ν_i. Note that in principle this argument requires estimating investment quantiles at each element of Ξ_I^R, which is infeasible in practice. We estimate $F_X^{-1}(1 - F_\nu(\nu_i) \mid \boldsymbol{\xi})$ as the predicted values from quantile regressions of investment on features of $\boldsymbol{\xi}$.

Equation (29) allows us to solve efficiently for $\bar{\boldsymbol{V}}_I(\boldsymbol{\xi}; \boldsymbol{\theta}_{-\phi})$. That system of equations can be essentially solved only once, as we can compute a decomposition of $I - \beta \boldsymbol{M}(\boldsymbol{P})$ and store it in memory. Then, as we vary $\boldsymbol{\theta}_{-\phi}$, we only need to recompute the expected flow profits and solve the resulting system using the stored matrix decomposition.[17]

4 A Review of the BBL Inequality Estimator

In this section we review the inequality estimator proposed by Bajari et al. (2007), which is the main point of comparison for the Rec-II estimator proposed in Section 3.

The expected discounted stream of profits of an incumbent playing strategy $\tilde{\sigma}^I = (\tilde{\sigma}^x, \tilde{\alpha}^I)$ when all its competitors play the strategy $\sigma = (\sigma^x, \alpha^I, \alpha^E)$ is given by

$$\bar{V}_I(\boldsymbol{\xi}; \tilde{\sigma}^I, \sigma, \boldsymbol{\theta}_{-\phi}) = \mathbb{E}\left\{ \sum_{t=0}^{\tau_e} \beta^t \left[\pi(\boldsymbol{\xi}_t) - c(\tilde{\sigma}^x(\boldsymbol{\xi}_t, \nu_t), \nu_t; \boldsymbol{\theta}_x)\right] + \beta^{\tau_e} \rho_{\tau_e} \middle| \boldsymbol{\xi}_0 = \boldsymbol{\xi} \right\} \tag{35}$$

where τ_e is the incumbent's endogenous and potentially infinite exit date and the public state evolves according to the probability distribution induced by $(\tilde{\sigma}^I, \sigma)$. The expected discounted stream of profits of a potential entrant playing strategy $\tilde{\sigma}^E = (\tilde{\sigma}^x, \tilde{\alpha}^I, \tilde{\alpha}^E)$ when all its competitors play strategy σ is

$$\bar{V}_E(\boldsymbol{\xi}; \tilde{\sigma}^E, \sigma, \boldsymbol{\theta}) = \int \int \tilde{\alpha}^E(\boldsymbol{\xi}_{-1}, \phi) v(\phi, \nu, \boldsymbol{\xi}_{-1}; \boldsymbol{\theta}_{-\phi}) \, \mathrm{d}F_\nu \, \mathrm{d}F_\phi(\boldsymbol{\theta}_\phi)$$

[17] If the cost of investment is linear in parameters and either the scrap value is deterministic (as in Bajari et al. (2007)) or there is no exit decision (as in Hashmi and van Biesebroeck (2016) and Appendix D), equation (29) implies that $\bar{V}_I$ is linear in parameters, i.e., $\bar{V}_I = [I - \beta \boldsymbol{M}(\boldsymbol{P})]^{-1} \boldsymbol{X} \boldsymbol{\theta}_{-\phi}$ for some matrix $\boldsymbol{X}$. In this case there are additional computational savings, as one can store the solution A to $[I - \beta \boldsymbol{M}(\boldsymbol{P})]A = \boldsymbol{X}$ in memory and only compute $A\boldsymbol{\theta}_{-\phi}$ as the parameters change.

where

$$v(\phi, \nu, \boldsymbol{\xi}_{-1}; \boldsymbol{\theta}_{-\phi}) := -\phi - c(\tilde{\sigma}^x(-\infty, \boldsymbol{\xi}_{-1}, \nu), \nu; \boldsymbol{\theta}_x) \\ + \beta \int \int \bar{V}_I(\boldsymbol{\xi}'; \tilde{\sigma}^E, \sigma, \boldsymbol{\theta}_{-\phi}) \, \mathrm{d}F^\sigma(\boldsymbol{\xi}'_{-1} \mid (-\infty, \boldsymbol{\xi}_{-1})) \, \mathrm{d}F(\xi'_1 \mid \xi_1, \tilde{\sigma}^x(-\infty, \boldsymbol{\xi}_{-1}, \nu))$$

A symmetric strategy profile $(\sigma, \ldots, \sigma)$ is a Symmetric Markov Perfect Equilibrium only if, for all $\boldsymbol{\xi}$ and σ',

$$\bar{V}_E(\boldsymbol{\xi}; \sigma, \sigma, \boldsymbol{\theta}) \geq \bar{V}_E(\boldsymbol{\xi}; \sigma', \sigma, \boldsymbol{\theta}) \quad \text{and} \quad \bar{V}_I(\boldsymbol{\xi}; \sigma, \sigma, \boldsymbol{\theta}_{-\phi}) \geq \bar{V}_I(\boldsymbol{\xi}; \sigma', \sigma, \boldsymbol{\theta}_{-\phi}) \,. \tag{36}$$

Bajari et al. (2007) base their estimator on the equilibrium conditions (36).[18] Let H be a distribution over the space of pairs of the form $(\boldsymbol{\xi}, \sigma')$. Define

$$Q(\boldsymbol{\theta}, \sigma) := \int \left(\min \left\{ g(\boldsymbol{\xi}, \sigma'; \sigma, \boldsymbol{\theta}), 0 \right\} \right)^2 \mathrm{d}H(\boldsymbol{\xi}, \sigma') \,, \tag{37}$$

where $g(\boldsymbol{\xi}, \sigma'; \sigma, \boldsymbol{\theta}) := \bar{V}(\boldsymbol{\xi}; \sigma, \sigma, \boldsymbol{\theta}) - \bar{V}(\boldsymbol{\xi}; \sigma', \sigma, \boldsymbol{\theta})$.

Let $\mathcal{E}(\boldsymbol{\theta})$ be the set of SMPEs when the parameters of the model are given by $\boldsymbol{\theta}$ and let $\boldsymbol{\theta}_0$ denote the true parameter value. If $\sigma \in \mathcal{E}(\boldsymbol{\theta}_0)$, the equilibrium conditions above imply that $Q(\boldsymbol{\theta}_0, \sigma) = 0$.

Assumption 2 (Identification)**.** For any $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$, $\mathcal{E}(\boldsymbol{\theta}) \cap \mathcal{E}(\boldsymbol{\theta}') = \emptyset$.

Under assumption 2, $\sigma \in \mathcal{E}(\boldsymbol{\theta}_0) \Rightarrow \sigma \notin \mathcal{E}(\boldsymbol{\theta}')$. Therefore, if $\boldsymbol{\theta}' \neq \boldsymbol{\theta}_0$, then there must exist $(\boldsymbol{\xi}, \sigma')$ for which $g(\boldsymbol{\xi}, \sigma'; \sigma, \boldsymbol{\theta}') < 0$. It follows that, for an appropriate choice of H, $Q(\boldsymbol{\theta}', \sigma) > 0$.[19]

Bajari et al. (2007) propose estimating the structural parameters of the model by minimizing a sample analog of (37). In particular, given a set of $\{(\boldsymbol{\xi}_i, \sigma'_i)\}_{i=1}^{n_I}$

[18] Condition (36) is slightly weaker than Markov Perfect Equilibrium as it allows violations of optimality at seats of measure zero (according F_ρ and F_ϕ).

[19] To be more precise, $Q(\boldsymbol{\theta}', \sigma) > 0$ requires that $g(\boldsymbol{\xi}, \sigma'; \sigma, \boldsymbol{\theta}') < 0$ on a set of positive H-measure for all $\boldsymbol{\theta}' \neq \boldsymbol{\theta}_0$. We can attach this condition to our definition of MPE. Given a measure μ on the set of tuples $(\boldsymbol{\xi}, \sigma')$, say that $(\sigma, \ldots, \sigma)$ is a symmetric MPE if $g(\boldsymbol{\xi}, \sigma'; \sigma, \theta_0) < 0$ with zero μ-measure. Then choose H such that μ is absolutely continuous with respect to H. If $\theta' \neq \theta_0$, assumption 2 implies that $g(\boldsymbol{\xi}, \sigma'; \sigma, \theta') < 0$ with positive μ-measure. This implies that $g(\boldsymbol{\xi}, \sigma'; \sigma, \theta') < 0$ with positive H-measure, otherwise absolute continuity of μ with respect to H would be violated. Thus, $Q(\theta', \sigma) > 0$. This hints at difficulties with the BBL approach: the measure H has to be rich, in the sense of $\mu \ll H$, where μ is itself rich enough that we are willing to define MPE on its basis. If H is not sufficiently rich, the equilibrium conditions may be violated at a set of positive μ-measure that is neglected by H. In this case $Q(\boldsymbol{\theta}', \sigma) = 0$.

pairs and an estimate of the strategy profile $\hat{\sigma}$, they propose minimizing

$$\hat{Q}(\boldsymbol{\theta}, \hat{\sigma}) := \frac{1}{n_I} \sum_{i=1}^{n_I} \left(\min \left\{ g(\boldsymbol{\xi}_i, \sigma_i'; \hat{\sigma}, \boldsymbol{\theta}), 0 \right\} \right)^2 \quad i = 1, \ldots, n_I \, .$$

Evaluating this objective requires estimates of $\bar{V}(\boldsymbol{\xi}, \sigma', \hat{\sigma}, \boldsymbol{\theta})$. Bajari et al. (2007) propose obtaining these estimates by forward simulation. As they note, linearity of the value function with respect to $\boldsymbol{\theta}$ significantly reduces the computational burden of forward simulation. In fact, under linearity there exists a function $\bar{\Lambda}(\boldsymbol{\xi}; \sigma', \sigma)$ such that $\bar{V}(\boldsymbol{\xi}; \sigma', \sigma, \boldsymbol{\theta}) = \bar{\Lambda}(\boldsymbol{\xi}; \sigma', \sigma)\boldsymbol{\theta}$ and forward simulation need not be repeated as $\boldsymbol{\theta}$ varies. As established by equation (29), however, value functions are generally not linear in $\boldsymbol{\theta}$ when scrap values are random. Nonetheless, we will show in Section 5.3 that nonlinearity arising from exit behavior does not add substantive computational burden.

5 Monte Carlo Simulations

In this section we first introduce the model we use as a testing ground for the estimators based on recursive equilibrium conditions and BBL. The model is inspired by and extends the Hashmi and van Biesebroeck (2016) model of R&D in the automobile industry to allow for entry and exit. After presenting the model parameterization we adopt in simulation exercises, we provide implementation details for the different estimators. We end by presenting results from a numerical study.

5.1 The Model

The model we simulate is a special case of the model discussed in Section 2. The market is populated by single-product firms characterized by product quality ξ. Each period, firms set prices and invest to improve product quality.

5.1.1 Static Price Competition

Suppose there are J firms in the market, indexed by $j = 1, \ldots, J$. Consumer i derives conditional indirect utility u_{ij} from purchasing firm j's product, where

$$u_{ij} = \begin{cases} \epsilon_i^{\text{out}} + (1 - \varsigma)\varepsilon_{i0} & \text{if } j = 0 \\ \alpha p_j + \xi_j + \epsilon_i^{\text{in}} + (1 - \varsigma)\varepsilon_{ij} & \text{if } j = 1, \ldots, J \end{cases} . \tag{38}$$

In (38), $j = 0$ denotes the no-purchase option and p_j denotes good j's price.[20] Goods are grouped into two nests, one containing all inside goods (i.e, those produced by one of the J firms) and one containing the no-purchase option. The ε_{ij}'s are independent and identically distributed Type 1 Extreme Value random variables. The nest-level disturbances ϵ_i^{out} and ϵ_i^{in} follow the unique distribution such that $\epsilon_i^{\text{g}} + (1-\varsigma)\varepsilon_{ij}$, for g $= \{\text{in}, \text{out}\}$, is also Type 1 Extreme Value distributed – see Cardell (1997). This nested-logit specification yields market share formulas

$$s_j(\boldsymbol{p}, \boldsymbol{\xi}) = \frac{\exp\left(\frac{\alpha p_j + \xi_j}{1-\varsigma}\right)}{D^{\varsigma}(1 + D^{1-\varsigma})} ,$$

where $D = \sum_{j=1}^{J} \exp\left(\frac{\alpha p_j + \xi_j}{1-\varsigma}\right)$. Prices are modeled as the outcome of Nash-Bertrand competition. That is, firm j solves

$$\max_{p_j} \pi_j(p_j, \boldsymbol{p}_{-j}, \boldsymbol{\xi}) := M(p_j - mc(\xi_j))s_j(\boldsymbol{p}, \boldsymbol{\xi}) \tag{39}$$

taking $\boldsymbol{p}_{-j}$ and $\boldsymbol{\xi}$ as given. In problem (39), $mc(\xi_j)$ denotes the firm's constant marginal cost of production, specified as

$$mc(\xi_j) = \exp(\theta_{c1} + \theta_{c2}\xi_j) ,$$

and M is the number of consumers in the market. For a fixed vector $\boldsymbol{\xi}$, uniqueness and existence of the price equilibrium follow from results in Caplin and Nalebuff (1991). The equilibrium $\boldsymbol{p}^*(\boldsymbol{\xi})$ satisfies the system of first-order conditions

$$\frac{\partial s_j(\boldsymbol{p}, \boldsymbol{\xi})}{\partial p_j}(p_j^* - mc(\xi_j)) + s_j(\boldsymbol{p}, \boldsymbol{\xi}) = 0 \qquad j = 1, \ldots, J$$

inducing profits $\pi(\boldsymbol{\xi}) := M(p_j^*(\boldsymbol{\xi}) - mc(\xi_j))s_j(\boldsymbol{p}^*(\boldsymbol{\xi}), \boldsymbol{\xi})$.

5.1.2 The Investment Decision

Firms invest to affect the quality of their product in the following period. Quality satisfies the conditions laid out in Section 2. It belongs to the set $\Xi = \{-\infty, \xi_m, \xi_m + \delta, \ldots, \xi_M - \delta, \xi_M\}$. If $\xi_m < \xi < \xi_M$, ξ can increase by δ, remain unchanged, or decrease by δ. Quality only transitions to $-\infty$ as a result of exit. Each firm's quality is affected by two shocks, one positive and one negative, which are independent from one another and across firms. The negative shock lowers quality by δ with exogenous probability $\theta_{t1} \in (0, 1)$. The positive shock

[20] As all firms offer a single product, we let j denote interchangeably a firm and its product.

instead increases quality by δ with probability $\text{up}(\xi, x)$. This can be interpreted as the probability of R&D success, which is allowed to depend on a firm's investment and quality (but not on competitors'). For instance, in the parameterization below $\text{up}(\xi, x)$ is an increasing function of x and a decreasing function of ξ. This captures the notions that investment increases the probability of an R&D success and that it is harder to improve on a high-quality product.

Given independence of the shocks, quality transition probabilities satisfy[21]

$$P(\xi' \mid \xi, x) = \begin{cases} \theta_{t1}[1 - \text{up}(\xi, x)] & \text{if } \xi' = \xi - \delta \\ 1 - \theta_{t1} - \text{up}(\xi, x)(1 - 2\theta_{t1}) & \text{if } \xi' = \xi \\ (1 - \theta_{t1})\text{up}(\xi, x) & \text{if } \xi' = \xi + \delta \\ 0 & \text{otherwise} \end{cases} \tag{40}$$

Given this structure, firms choose investment to maximise the present-discounted stream of profits. They balance expected higher product quality with an immediate cost of investment $c(x, \nu)$, where ν denotes the private investment cost shock.

We note that the structure of quality transitions in (40) implies that the investment first-order condition can be written as

$$\frac{\partial c(x, \nu)}{\partial x} = \beta \frac{\partial \text{up}(\xi, x)}{\partial x} \Delta W(\boldsymbol{\xi}) \tag{41}$$

where $\Delta W(\boldsymbol{\xi})$ is the expected increase in the continuation value due to a positive shock, i.e.,[22]

$$\Delta W(\boldsymbol{\xi}) := (1 - \theta_{t1})[W(\xi + \delta \mid \boldsymbol{\xi}) - W(\xi \mid \boldsymbol{\xi})] + \theta_{t1}[W(\xi \mid \boldsymbol{\xi}) - W(\xi - \delta \mid \boldsymbol{\xi})] \ .$$

Fix $\boldsymbol{\xi}$ and suppose that there exists ν such that $\sigma(\boldsymbol{\xi}, \nu) > 0$ – which, of course, is implied by observing strictly positive investment in the data. Assume further that $\partial_x c(x, \nu) > 0$ for all (x, ν). Because $\sigma(\boldsymbol{\xi}, \nu) > 0$, the first-order condition must hold with equality – i.e., equation (41) must hold. Therefore, it must be the case that $\Delta W(\boldsymbol{\xi}) > 0$. This has two implications. First, this condition is a restriction on the equilibrium value function and by itself provides restrictions on the structural parameters. Second, under the additional assumption that

[21] To ensure the bounds of Ξ are satisfied, the probability that $\xi' = \xi$ at maximum (minimum) quality is defined to be the complement of the probability of ξ decreasing (increasing).

[22] Again, this expression needs to be adjusted slightly if the focal firm is at either end of the set of possible quality levels.

up(ξ, x) is strictly concave in x, the firm's investment problem is globally strictly concave at that $\boldsymbol{\xi}$ for all ν. This in turn implies that the investment first-order conditions uniquely determine optimal investment. We thus obtain sufficiency of the first-order condition without the assumption that $W(\xi \mid \boldsymbol{\xi})$ is increasing in ξ, which is required in Proposition 1.

The local structure of the transitions (40) also has positive computational implications. It implies that the transition matrix $\boldsymbol{M}(\boldsymbol{P})$ in equation (32) is banded. A banded matrix is a sparse matrix whose non-zero entries are confined to a band around the main diagonal. Importantly, the LU decomposition of a banded matrix has banded components. This implies substantial computational savings in both the computation of the LU decomposition and the substitutions used to compute the solution to the linear system (29).[23]

We adopt a parameterization very similar to that in Hashmi and van Biesebroeck (2016). Specifically, we let

$$\text{up}(\xi, x) := \exp(-\exp(-\theta_{t2}\log(1+x) - \theta_{t3}\xi - \theta_{t4}\xi^2)) \tag{42}$$

where x denotes the firm's investment in R&D. This is exactly as in HvB. The functional form (42) allows for the notion that it is harder to improve on a high-quality product. Indeed, with the parameter values in Table 1, the quadratic term $\theta_{t3}\xi + \theta_{t4}\xi^2$ decreases in ξ. Other forms of dependence are possible. Moreover, we parameterize the cost of investment as

$$c(x, \nu) = \theta_{x1}x + \theta_{x2}x^2 + \theta_{x3}x\nu \ .$$

HvB have an additional cubic term that we omit. Note that the cost shock directly interacts with the investment level and hence affects the optimal investment choice. This rationalizes that two firms facing the same quality vector $\boldsymbol{\xi}$ may optimally choose different levels of investment. We further assume, following Hashmi and van Biesebroeck (2016), that $\nu \sim N(0, 1)$.[24]

[23] Note that it is important for this observation that equation (29) refers only to incumbent states. The matrix of transitions over all states includes non-zero entries in its first few columns (i.e., those that pertain to transitions to $\xi = -\infty$), making its (lower) bandwidth large, and reducing the computational savings. The full computational savings are thus a consequence of both the local nature of transitions and the assumption that exiting firms perish.

[24] Observe that the parameterization in HvB violates both the positivity of marginal costs and the concavity of the up$(\cdot)$ function. However, these violations are mild. First, at the levels of investment we simulate and at the true parameters, negative marginal cost of investment only happens for extreme negative values of ν, which occur with very low probability. Second, HvB's functional form for the up$(\cdot)$ function violates concavity only at very low levels of investment. Nevertheless, better parametric choices can be made to ensure that the invest-

5.2 Parameterization

Table 1 reports parameters governing the data generating process. We consider a rather data-rich environment, with data from 100 separate markets recorded for 40 periods. We fix $\bar{N} = 5$ in each simulated market, in line with simulations in HvB.[25] Incumbent product quality can take on fifteen values, from -1.4 to 1.4 in increments of 0.2. Firms are characterized by constant marginal costs $mc_j = \exp(2.47)$ – namely, $\theta_{c1} = 2.47, \theta_{c2} = 0$. The probability of a quality downgrade shock is $\theta_{t1} = 0.347$, and the probability of a quality upgrade shock is increasing in own investment ($\theta_{t2} = 0.162$) but decreasing (at an increasing rate) in own quality level ($\theta_{t3}, \theta_{t4} < 0$). Investment costs are convex ($\theta_{x1}, \theta_{x2} > 0$) in investment and increasing in the 'shock' term ($\theta_{x3} > 0$). Our parameterization differs somewhat from that in Hashmi and van Biesebroeck (2016). For instance, we assume a lower probability of quality downgrade shock and a higher elasticity of quality improvement to own quality. We also consider different values for investment cost parameters. For instance, HvB report a negative θ_{x2}. We impose a positive value to ensure convexity of the investment cost function. The changes in parameter values are not instrumental for the qualitative results we report in section 5.4.

Finally, we have to take a stance on entry cost and scrap value distributions, which are not included in HvB. We assume both are lognormal distributions. Scrap values have lower mean and standard deviation than entry costs.[26]

5.3 Implementation Details for Different Estimators

We compare the performance of four estimators. The first is Indirect Inference, the estimator proposed in Section 3, equation (26). The others are three variants of the BBL estimator presented in Section 4.

ment cost and up functions retain desirable properties. For instance, one may assume that $\log \nu \sim N(0,1)$ and that $\text{up}(\cdot)$ is derived from an exponential, rather than a Gumbel, CDF. We plan to adopt that parameterization in future revisions of this paper.

[25] In their empirical analysis, Hashmi and van Biesebroeck (2016) aggregate data to have a single market with 14 firms for the 1982-2006 period. When computing the equilibrium of the dynamic game, however, they restrict the number of firms to 5 to reduce computational burden. See Hashmi and van Biesebroeck (2016, Footnote 30). As a second point of comparison, Ryan (2012) collected a total of 517 market-year pairs (an unbalanced panel of 27 markets over 19 years), with the number of firms in a market-year ranging from 1 to 20 and averaging 4.75. We repeat the estimator comparison for a data structure akin to Ryan's in Appendix C.

[26] Table 1 details the distribution of scrap values and entry costs *before* they are scaled to have the same order of magnitude as firm profits. Scaling is implemented by multiplying these shocks by the average of $\pi(\boldsymbol{\xi})/(1-\beta)$ across states.

Table 1: Exercise Parameters

Parameter	Value
Data Structure	
Number of Markets	100
Number of Periods	40
Maximum Number of Firms	5
Own State Space	[$-\infty$, -1.4, -1.2, ... , 1.2, 1.4]
Model Parameters	
Number of Households	1.0e8
Discount Factor	0.925
Nesting Parameter	0.5
Marginal Utility of Income	-0.222
Marginal Cost Parameters	[2.47, 0.0]
Investment Cost Parameters	[2.625, 1.624, 0.5096]
Transition Probability Parameters	[0.347, 0.162, -1.0, -0.285]
Scrap Value Distribution	$\log$ N(-1.0, 0.75)
Entry Cost Distribution	$\log$ N(0.625, 0.5)
II Estimator	
Number of Investment Repetitions	5
Weight Matrix	Bootstrap
Number of Bootstrap Samples	1000
Number of Inequalities	5000
BBL Estimators	
Number of Simulated Paths	500
Simulation Horizon	80

This table reports parameters used in our Monte Carlo exercise.
Data Structure
The number of markets and periods define the size of the simulated balanced panel.
"Own State Space" denotes the values ξ_j can take for each firm.
II Estimator
Number of investment repetitions: number of simulated investment cost shocks per choice.
Weight matrix: we use a bootstrap estimator of the inverse of the variance-covariance matrix of the estimation moments.
BBL Estimators
Number of inequalities: number of equilibrium conditions included in each BBL objective.
Number of simulated paths: number of simulated histories used in simulating value functions.
Simulation horizon: number of periods for which paths are simulated.
Deviation details are discussed in the main text.

Both Rec-II and the BBL inequality estimator requires estimates of policy functions. We obtain those as follows. From equation (34) we know that the investment policy function satisfies $\sigma^x(\boldsymbol{\xi}, \nu) = F_X^{-1}(1 - F_\nu(\nu) \mid \boldsymbol{\xi})$. Due to this result, we estimate the investment policy function by estimating a set of quantile regressions of investment on functions of $\boldsymbol{\xi}$. Formally, let $\tau \in (0, 1)$ and let ν_τ be the τ-th quantile of F_ν. Denote by $\hat{\chi}_\tau^x$ the quantile regression coefficient estimate associated with the τ-th conditional quantile. Our estimate of the investment policy function is

$$\hat{\sigma}^x(\boldsymbol{\xi}, \nu_\tau) = f^x(\boldsymbol{\xi})' \hat{\chi}_\tau^x \ ,$$

where $f^x(\boldsymbol{\xi})$ includes dummies for the firm's own quality, the number of active firms in the market, the rank of the firm's quality in the market, the mean quality in the market, and the maximum quality in the market.[27] With regards to entry and exit decisions, we estimate conditional probabilities of firms choosing to be active. We do so by estimating logistic regressions of those decisions on functions of $\boldsymbol{\xi}$. Let $\mathbb{P}^E(\boldsymbol{\xi}_{-1}) := \mathbb{P}(\alpha^E(\boldsymbol{\xi}_{-1}, \phi) = 1)$ and $\mathbb{P}^I(\boldsymbol{\xi}) := \mathbb{P}(\alpha^I(\boldsymbol{\xi}, \rho) = 1)$ denote the conditional probabilities that entrants and incumbents choose to be active, respectively. Our estimates of these probabilities are

$$\hat{\mathbb{P}}^E(\boldsymbol{\xi}_{-1}) = \Lambda\big(f^E(\boldsymbol{\xi})' \hat{\chi}^E\big) \qquad \hat{\mathbb{P}}^I(\boldsymbol{\xi}) = \Lambda\big(f^I(\boldsymbol{\xi})' \hat{\chi}^I\big) \ .$$

where Λ denotes the logistic CDF and $\hat{\chi}^E$ and $\hat{\chi}^E$ denote the logistic regression coefficient estimates. The vector $f^E(\boldsymbol{\xi})$ includes dummies for the number of firms and the average quality in the market, whereas $f^I(\boldsymbol{\xi})$ includes dummies for the number of firms and the firm's own quality.

The BBL variants we consider differ in the construction of the deviations from these policies, which we denote as $\{\tilde{\sigma}^x(\boldsymbol{\xi}, \nu_\tau), \tilde{\mathbb{P}}^E(\boldsymbol{\xi}_{-1}), \tilde{\mathbb{P}}^I(\boldsymbol{\xi})\}$. Let i index a deviation. In what we term Asymptotic BBL we draw $\tilde{\boldsymbol{\chi}}_i \sim \mathrm{N}(\hat{\boldsymbol{\chi}}, \hat{\Sigma}_{\boldsymbol{\chi}})$ for each i. We then form deviations as

$$\tilde{\sigma}_i^x(\boldsymbol{\xi}, \nu_\tau) = f^x(\boldsymbol{\xi})' \tilde{\chi}_{i\tau}^x; \qquad \tilde{\mathbb{P}}_i^E(\boldsymbol{\xi}_{-1}) = \Lambda\big(f^E(\boldsymbol{\xi}) \tilde{\chi}_i^E\big); \qquad \tilde{\mathbb{P}}_i^I(\boldsymbol{\xi}) = \Lambda\big(f^I(\boldsymbol{\xi}) \tilde{\chi}_i^I\big).$$

[27] This is similar to the specification that HvB adopt to estimate the investment policy function. We adopt a specification that is more flexible with respect to own quality and omits higher order moments of the distribution of quality in a market, as we simulate a dataset with fewer firms per market.

In what we term Multiplicative BBL we form deviations as

$$\tilde{\sigma}_i^x(\boldsymbol{\xi}, \nu_\tau) = \iota_i f^x(\boldsymbol{\xi})' \hat{\chi}_\tau^x; \qquad \tilde{\mathbb{P}}_i^E(\boldsymbol{\xi}_{-1}) = \Lambda\big(\iota_i f^E(\boldsymbol{\xi}) \hat{\chi}^E\big); \qquad \tilde{\mathbb{P}}_i^I(\boldsymbol{\xi}) = \Lambda\big(\iota_i f^I(\boldsymbol{\xi}) \hat{\chi}^I\big) ,$$

where $\iota_i \in \{.95, .975, 1.025, 1.05, 1.075\}$ and we use all of these values of ι_i for each public state included in the BBL inequalities. This is HvB's approach.[28] Lastly, in what we term Additive BBL we draw $o_{is} \sim N(0, 0.5)$ for each inequality i and simulated decision s. We then form deviations as

$$\tilde{\sigma}_{is}^x(\boldsymbol{\xi}, \nu_\tau) = f^x(\boldsymbol{\xi})' \hat{\chi}_\tau^x + o_{is}^x; \qquad \tilde{\mathbb{P}}_{is}^E(\boldsymbol{\xi}_{-1}) = \Lambda\big(f^E(\boldsymbol{\xi}) \hat{\chi}^E + o_{is}^E\big); \qquad \tilde{\mathbb{P}}_{is}^I(\boldsymbol{\xi}) = \Lambda\big(f^I(\boldsymbol{\xi}) \hat{\chi}^I + o_{is}^I\big).$$

Given estimated policy functions and deviations, typical implementations of the BBL inequality estimator compute value functions by forward simulation. As shown by equation (29), the incumbent value function depends on the conditional expectation of scrap values which causes the value function to be non-linear in model parameters. This nonlinearity increases the computational cost of the forward simulation routine, as the discussion at the end of section 4 no longer applies.

However, in the model we consider here, it is still the case that the forward simulation routine can be performed essentially once. To see this, note that we can define the exit policy as a function of a $U[0,1]$ random variable by means of a change of variable: $\check{\sigma}^I(\boldsymbol{\xi}, \tau) := \sigma^I(\boldsymbol{\xi}, F_\rho^{-1}(\tau))$. It follows that $\check{\sigma}^I(\boldsymbol{\xi}, \tau) = \mathbf{1}\{\tau \leq \mathbb{P}^I(\boldsymbol{\xi})\}$.[29] We thus take $\tau \sim U[0,1]$ draws and use those to simulate exit decisions: incumbents remain active if and only if $\tau \leq P^I(\boldsymbol{\xi})$. As we vary structural parameters, these simulations do not need to be repeated. All that needs to be recomputed is the scrap value that accrues to firms when they do decide to exit, which is $F_\rho^{-1}(\tau; \boldsymbol{\theta}_\rho)$, when $\tau > \mathbb{P}^I(\boldsymbol{\xi})$. The cost of repeatedly calling F_ρ^{-1} is typically dwarfed by the cost of repeating the simulation.

5.4 Monte Carlo Results

This section compares the performance of the II estimator and the three BBL alternatives outlined in section 5.3. We fix the computational budget across estimators to be the same. Namely, we set the number of inequalities in the BBL estimators to equalize their runtime to the runtime of the II estimator.

[28] Because this approach generates 5 policy deviations per initial state, we consider $nI/5$ initial states to obtain nI inequalities.

[29] Indeed, $\check{\sigma}^I(\boldsymbol{\xi}, \tau) = \sigma^I(\boldsymbol{\xi}, F_\rho^{-1}(\tau)) = \mathbb{1}\{F_\rho^{-1}(\tau) \leq \bar{V}_I^A(\boldsymbol{\xi}) - \pi(\boldsymbol{\xi})\} = \mathbb{1}\{\tau \leq F_\rho(\bar{V}_I^A(\boldsymbol{\xi}) - \pi(\boldsymbol{\xi}))\} = \mathbb{1}\{\tau \leq \mathbb{P}^I(\boldsymbol{\xi})\} = \mathbb{P}^I(\boldsymbol{\xi})$.

Table 2 lists true parameter values, along with averages of estimates and standard deviations for the four estimators over 500 simulations.

Of the four estimators, the II estimator evidently performs the best: averages are close to the true parameters, with smaller standard deviation. All BBL estimators display substantial bias. Figures 1 and 2 present the distribution of parameter estimates for each algorithm; vertical dotted lines represent true parameter values. The II estimates are correctly centered around the true values. The distribution of $\boldsymbol{\theta}_x$ is close to Gaussian. The distributions of the remaining parameters display some skewness.[30] BBL estimators, on the other hand, are centered around wrong values, displaying substantial bias. BBL performs poorly for all parameters, but seems to have a particularly difficult time estimating $\boldsymbol{\theta}_x$.[31]

The relative performance of the different estimators is a consequence of the differences in the objective functions that define them. We illustrate this by analysing the shape of each objective function in a neighbourhood of the true parameters for a single simulated dataset. To do so, we plot slices of the objective function by varying one parameter at a time while holding the others fixed at their true values. To render the shape of different objective functions comparable, we normalize objective values on the parameter grid by dividing each by the objective value at the true parameter. Vertical dashed lines represent true parameter values. Figures 3 and 4 display the results.

Figure 3 shows that the objective function of the II estimator generally features pronounced local convexity around each true parameter, with minimal distance between the local objective minimum and the objective value at the true parameter. This is reflected in the good overall performance of the estimator. On the other hand, the objective slice for σ_ρ is a reminder that good performance is not guaranteed in a single sample.

The picture is very different for the BBL estimators in Figure 4. The objective slices are generally not minimized close to the true parameters. In many cases, the objective function is monotonic over the parameters, which is consistent with the degenerate distributions observed in Figure 2. Moreover, the II objective is typically much more responsive to deviations from its minimum than

[30] We note that these distributions are closer to Gaussian than those we obtain with a smaller sample size in Appendix C.

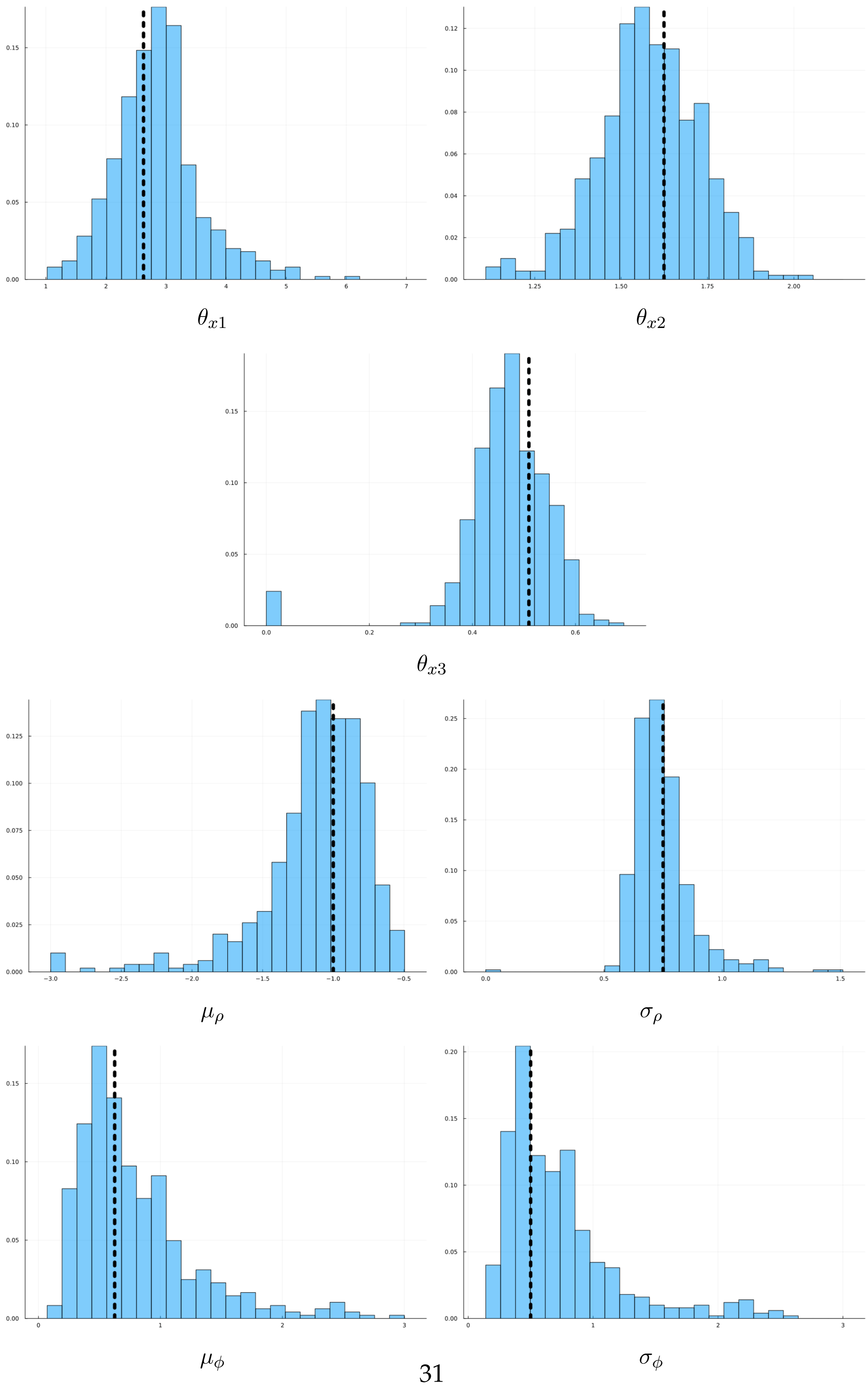
[31] A cursory glance at these figures may understate the precision of the II estimator relative to BBL, because the range of these figures varies across estimators. The ranges in the distributions of II estimates are much shorter than those in the distributions of BBL estimates.

Table 2: Summary of Parameter Estimates

	Value	Indirect Inference	BBL		
			Asymptotic	Multiplicative	Additive
θ_{x1}	2.625	2.872	6.888	1.944	0.08
		0.734	2.635	2.21	0.442
θ_{x2}	1.624	1.581	10.0	5.481	0.055
		0.148	0.004	0.813	0.335
θ_{x3}	0.5096	0.466	5.714	9.611	9.067
		0.097	2.694	1.142	2.003
μ_{ρ}	-1.0	-1.127	-1.329	-1.103	-2.0
		0.39	0.101	0.081	0.001
σ_{ρ}	0.75	0.746	0.69	0.589	1.038
		0.13	0.036	0.031	0.167
μ_{ϕ}	0.625	0.864	1.809	1.851	1.898
		0.616	0.074	0.063	0.268
σ_{ϕ}	0.5	0.735	2.998	2.998	3.0
		0.468	0.015	0.019	0.0

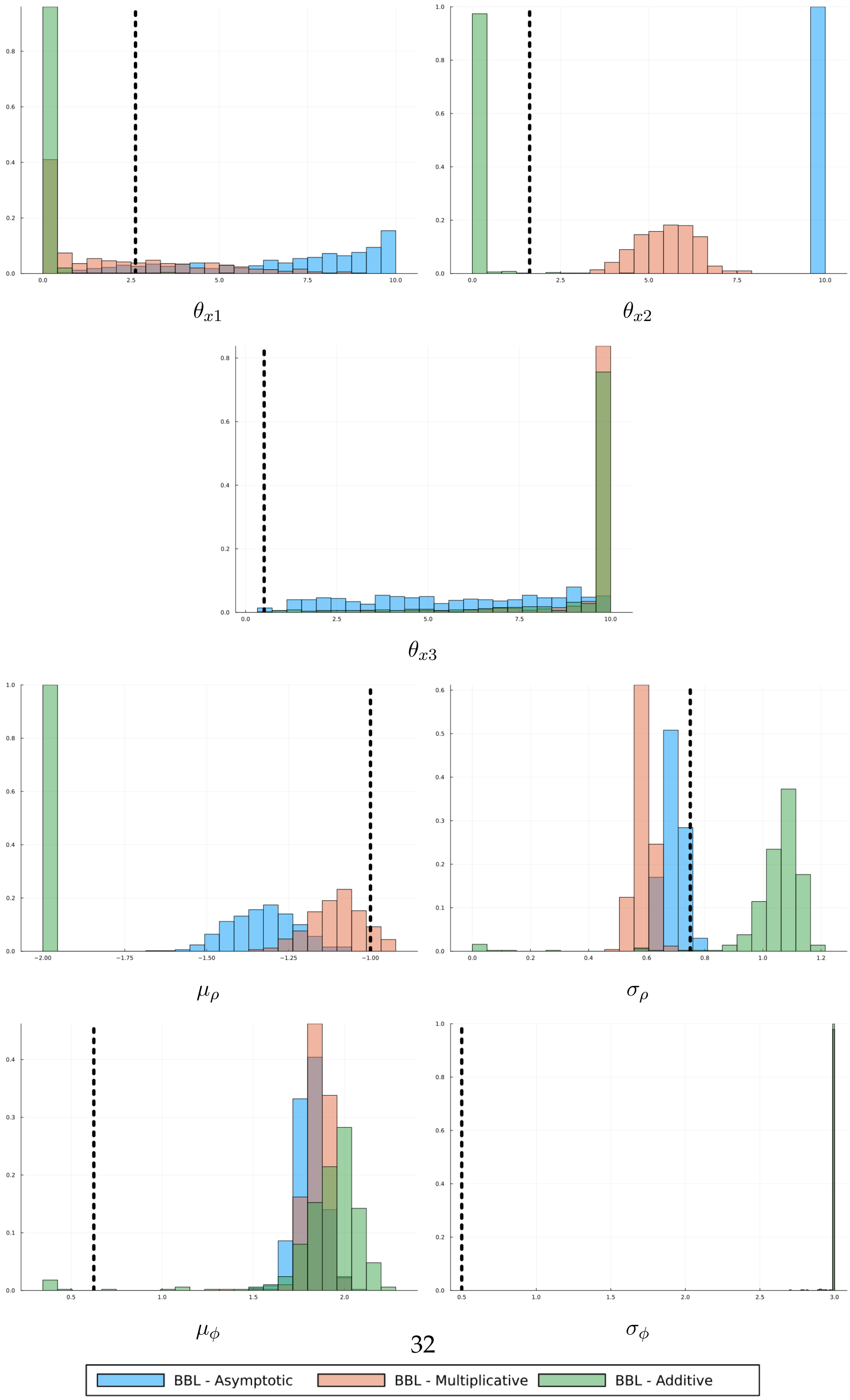
This table summarizes the results of our Monte Carlo experiment. The first column shows the value of the investment cost, scrap value, and entry cost parameters used in the data generating process. Each subsequent column shows the mean and standard deviation for estimates across 500 Monte Carlo replications. The column labeled “II” shows the results of the estimator we propose in this paper. The columns labeled “Asymptotic”, “Multiplicative”, and “Additive” display estimates from the three BBL alternatives we consider. See the main text for details.

Figure 1: II Estimator Parameter Estimates



This figure plots the distribution of parameter estimates obtained using the indirect inference estimator defined in Equation (26) over 500 Monte Carlo replications. The vertical dashed red line indicates the value of the corresponding parameter in the data generating process.

Figure 2: BBL Estimator Parameter Estimates



This figure plots the distribution of parameter estimates obtained using all BBL estimators over 500 Monte Carlo replications. 'Asymptotic' BBL in blue, 'Multiplicative' in red, 'Additive' in green. The vertical dashed red line indicates the value of the corresponding parameter in the data generating process.

the BBL objectives.[32] These figures also illustrate the fact that the performance of the BBL estimator depends on the deviations chosen by the econometrician. Even when the BBL objective is minimized close to the true parameter (holding other parameters fixed at the truth), as is the case for the μ_ρ slices, the degree of convexity of different BBL implementations differs markedly. This highlights one of the advantages of estimators based on recursive equilibrium conditions: they free the analyst from specifying policy deviations.

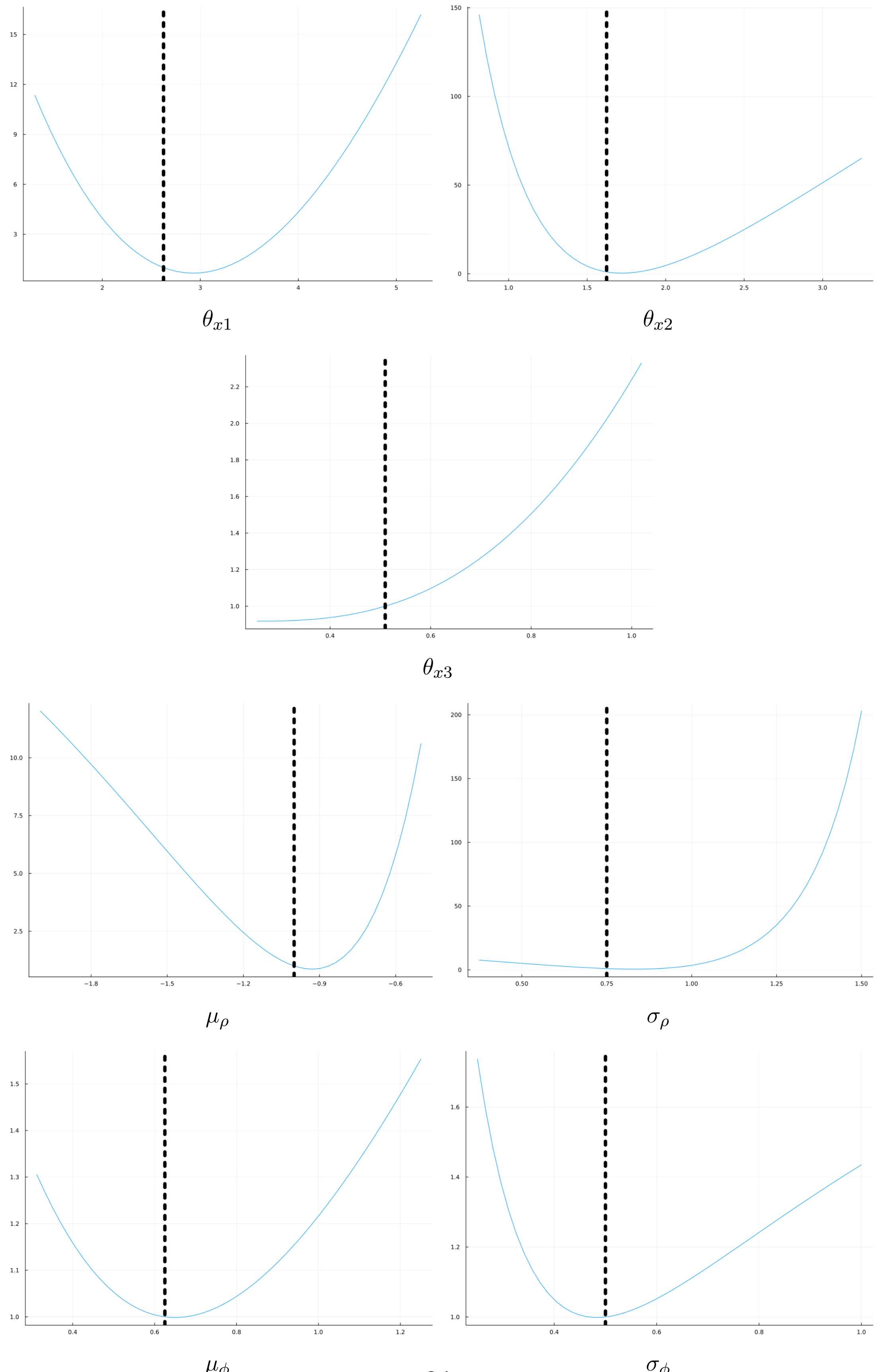
6 Conclusion

We have revisited the estimation of dynamic games with continuous controls. We note that the commonly applied inequality estimator of Bajari et al. (2007) does not fully exploit the structure of optimal policies and propose an estimator that does so. Our estimator combines two-step methods common in the estimation of dynamic models with firms' Bellman equations and indirect inference ideas. Importantly, our estimator applies to models with shocks to the marginal cost of the continuous control and random scrap values and entry costs. We conduct a Monte Carlo experiment based on an extension of Hashmi and van Biesebroeck (2016) that allows for entry and exit. We find that the estimator we propose significantly outperforms multiple implementations of the BBL inequality estimator at a fixed computational budget. In the context of a simpler model without entry and exit, we find that the indirect inference estimator outperforms alternative estimators that also exploit the structure of optimal policies.

Bajari et al. (2007) themselves propose an estimator that does use firms' optimal behavior, albeit in the context of a simpler model. However, the empirical literature has converged to applying exclusively their inequality estimator. We show how to use the optimality conditions characterizing firm behavior for estimation of a model with shocks to firms' costs of investment, entry, exit and random entry costs and scrap values. We show that incorporating that information brings about substantial econometric benefits relative to the BBL inequality estimator. We hope that these results will steer empirical researchers towards exploiting those conditions whenever possible.

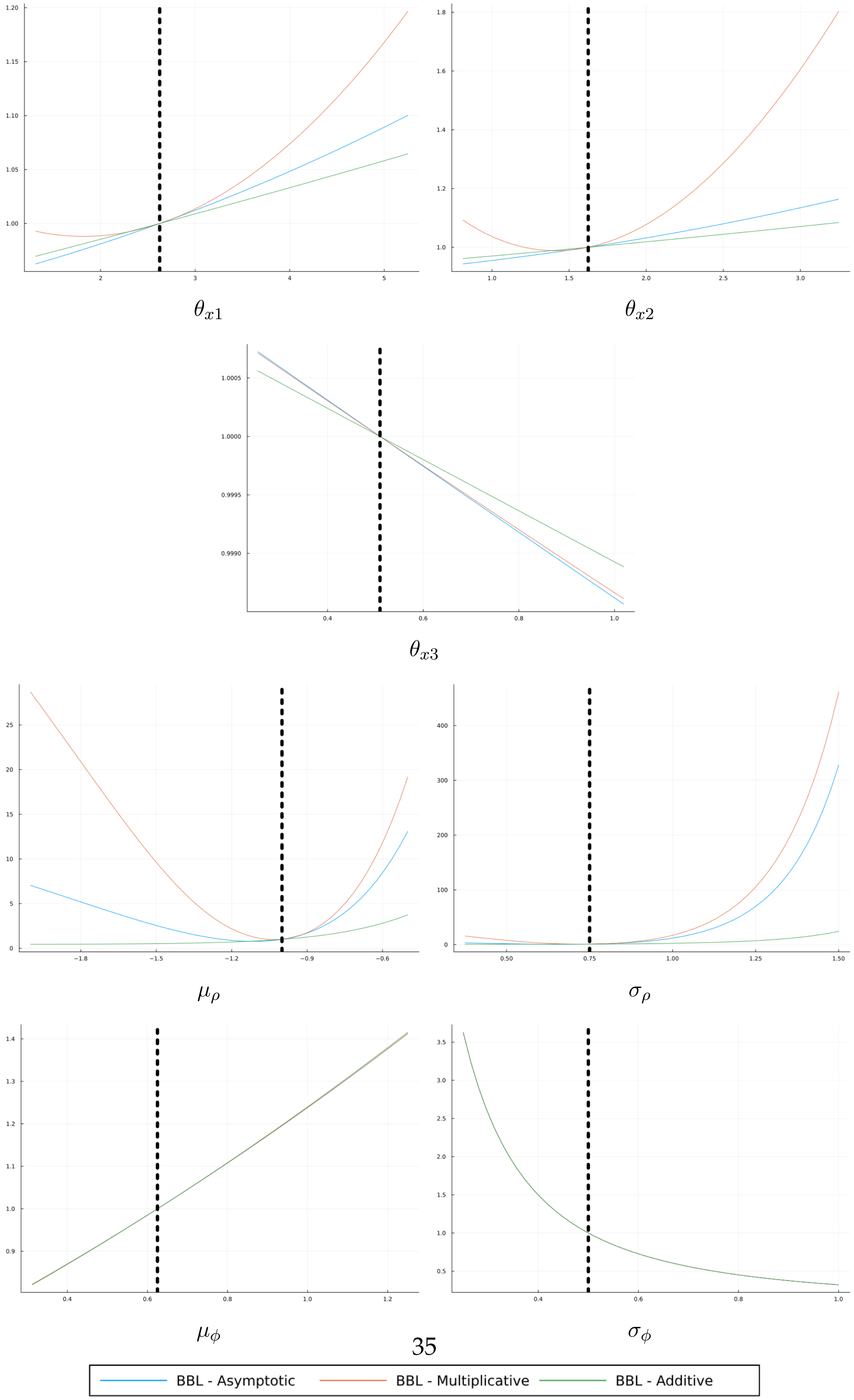
[32]The μ_ρ and σ_ρ slices for the multiplicative implementation of BBL are an exception to this rule.

Figure 3: II Estimator Objective Slices



This figure plots the value of the objective function of the indirect inference estimator varying one parameter at a time while holding the other parameters fixed at their true values. The vertical dashed line indicates the value of the corresponding parameter in the data generating process.

Figure 4: BBL Estimator Objective Slices



This figure plots the value of the objective function of the BBL estimators varying one parameter at a time while holding the other parameters fixed at their true values. 'Asymptotic' BBL in blue, 'Multiplicative' in red, 'Additive' in green. The vertical dashed line indicates the value of the corresponding parameter in the data generating process.

References

Aguirregabiria, V., Collard-Wexler, A., & Ryan, S. P. (2021). Dynamic games in empirical industrial organization. In *Handbook of industrial organization* (Vol. 4, pp. 225–343).

Aguirregabiria, V., & Mira, P. (2007). Sequential estimation of dynamic discrete games. *Econometrica*, *75*, 1-53.

Bajari, P., Benkard, C. L., & Levin, J. (2007). Estimating dynamic models of imperfect competition. *Econometrica*, *75*, 1331-1370.

Bartle, R. G. (1964). *The elements of real analysis* (Vol. 2). Wiley New York.

Caplin, A., & Nalebuff, B. (1991). Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica*, *59*, 25.

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, *13*(2), 185–213.

Collard-Wexler, A. (2013). Demand fluctuations in the ready-mix concrete industry. *Econometrica*, *81*(3), 1003–1037.

Doraszelski, U., & Pakes, A. (2007). A framework for applied dynamic analysis in io. In M. P. R. Armstrong (Ed.), .

Doraszelski, U., & Satterthwaite, M. (2010). Computable markov-perfect industry dynamics. *The RAND Journal of Economics*, *41*(2), 215–243.

Fowlie, M., Reguant, M., & Ryan, S. P. (2016). Market-based emissions regulation and industry dynamics. *Journal of Political Economy*.

Gourieroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, *8*, S85-S118.

Hashmi, A. R., & van Biesebroeck, J. (2016). The relationship between market structure and innovation in industry equilibrium: A case study of the global automobile industry. *Review of Economics and Statistics*, *98*, 192-208.

Hotz, V. J., & Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, *60*, 497-529.

Jofre-Bonet, M., & Pesendorfer, M. (2003). Estimation of a dynamic auction game. *Econometrica*, *71*(5), 1443–1489.

Liu, A. H., & Siebert, R. B. (2022, 1). The competitive effects of declining entry costs over time: Evidence from the static random access memory market. *International Journal of Industrial Organization*, *80*.

Manski, C. F., & Tamer, E. (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica*, *70*(2), 519–546.

Pakes, A., & McGuire, P. (1994). Computing markov-perfect nash equilibria : Numerical implications of a dynamic differentiated product model. *The RAND Journal of Economics*, *25*, 555-589.

Pakes, A., Ostrovsky, M., & Berry, S. (2007). Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). *RAND Journal of Economics*, *38*, 373-399.

Pesendorfer, M., & Schmidt-Dengler, P. (2008). Asymptotic least squares estimators for dynamic games. *The Review of Economic Studies*, *75*(3), 901–928.

Ryan, S. P. (2012). The costs of environmental regulation in a concentrated industry. *Econometrica*, *80*, 1019-1061.

Srisuma, S. (2013). Minimum distance estimators for dynamic games. *Quantitative Economics*, *4*(3), 549–583.

Appendices

Appendix A Proofs and Derivations

A.1 Establishing equations (5) and (6)

We repeat the two equations here for the reader's convenience:

$$\underbrace{\int_{\varepsilon_j}\int_{\xi_j'} \bar{V}_I(\boldsymbol{\xi}')\,\mathrm{d}F(\xi_j' \mid \xi_j, \sigma(\boldsymbol{\xi}_j, \varepsilon_j))\,\mathrm{d}G_{\varepsilon_j}}_{A} = \underbrace{\int_{\xi_j'} \bar{V}_I(\boldsymbol{\xi}')\,\mathrm{d}F^{\sigma}(\xi_j' \mid \boldsymbol{\xi}_j)}_{B} \quad (5 - \text{repeated})$$

where

$$F^{\sigma}(\xi_j' \mid \boldsymbol{\xi}_j) := \int_{\varepsilon_j} F(\xi_j' \mid \xi_j, \sigma(\boldsymbol{\xi}_j, \varepsilon_j))\,\mathrm{d}G_{\varepsilon_j}\ . \quad (6 - \text{repeated})$$

We establish (5) under the following condition.

Definition 2. A function $g : [a, b] \to \mathbb{R}$ is said to be uniformly Riemann-Stieltjes integrable with respect to the family of functions $\mathcal{F}$, where all $F \in \mathcal{F}$ map $[a, b]$ into $\mathbb{R}$, if for all $\eta > 0$ there exists a partition $P = (x_0, x_1, \ldots, x_n)$ of $[a, b]$ such that

$$\left| \int g\ \mathrm{d}F - \sum_{i=1}^{n} g(x_i)[F(x_i) - F(x_{i-1})] \right| < \eta$$

for all $F \in \mathcal{F}$.

Proposition 2. Let $\bar{V}_I(\xi, j, \boldsymbol{\xi}) := \bar{V}_I(\xi_1, \ldots, \xi_{j-1}, \xi, \xi_{j+1}, \ldots, \xi_{\bar{N}})$. Suppose that $\bar{V}_I(\xi, j, \boldsymbol{\xi})$ is uniformly Riemann-Stieltjes integrable with respect to the family of distributions $\{F(\cdot \mid \xi, x) : \xi \in \Xi, x \in \mathbb{R}_+\}$. Then equation (5) holds.

Proof. For some partition $P = (\xi_0, \xi_1, \ldots, \xi_n)$ of $[\xi_m, \xi_M]$, define the quantity

$$C(P) = \sum_{i=1}^{n} \bar{V}_I(\xi_i, j, \boldsymbol{\xi}) \left(\int_{\varepsilon} F(\xi_i \mid \xi_j, \sigma(\boldsymbol{\xi}_j, \varepsilon))\,\mathrm{d}G_{\varepsilon} - \int_{\varepsilon} F(\xi_{i-1} \mid \xi_j, \sigma(\boldsymbol{\xi}_j, \varepsilon))\,\mathrm{d}G_{\varepsilon} \right)$$

Fix $\eta > 0$. By the definition of B above, we can choose P so that $|C(P) - B| < \eta/2$. Moreover, $C(P)$ is also

$$C(P) = \int_{\varepsilon} \underbrace{\sum_{i=1}^{n} \bar{V}_I(\xi_i, j, \boldsymbol{\xi}) \left[F(\xi_i \mid \xi_j, \sigma(\boldsymbol{\xi}_j, \varepsilon)) - F(\xi_{i-1} \mid \xi_j, \sigma(\boldsymbol{\xi}_j, \varepsilon)) \right]}_{D(P,\varepsilon)}\ \mathrm{d}G_{\varepsilon}$$

Let $I(\varepsilon)$ denote the inner integral in A. By the uniform-integrability condition we can choose P' such that $|I(\varepsilon) - D(P', \varepsilon)| < \eta/2$ for all ε. Then

$$\begin{aligned} |A - C(P')| &= \left| \int_\varepsilon I(\varepsilon)\, \mathrm{d}G_\varepsilon - \int_\varepsilon D(P', \varepsilon)\, \mathrm{d}G_\varepsilon \right| \\ &\leq \int_\varepsilon |I(\varepsilon) - D(P', \varepsilon)|\, \mathrm{d}G_\varepsilon \\ &< \eta/2 \end{aligned}$$

Let P^* be a common refinement of P and P'. Then $|A - B| \leq |A - C(P^*)| + |C(P^*) - B| < \eta$. Since $\eta > 0$ is arbitrary, $A = B$. □

A.2 Proof of Proposition 1

Proof. By Integration by Parts (Bartle (1964, Theorem 22.7)),

$$\begin{aligned} \int_{\xi_i'} W(\xi_i' \mid \boldsymbol{\xi}, \sigma)\, \mathrm{d}F(\xi_i' \mid \xi_i, x) = &- \int F(\xi' \mid \xi, x)\, \mathrm{d}W(\xi' \mid \boldsymbol{\xi}, \sigma) \\ &+ W(\xi_M \mid \boldsymbol{\xi}, \sigma) \underbrace{F(\xi_M \mid \xi, x)}_{=1} - W(\xi_m \mid \boldsymbol{\xi}, \sigma) F(\xi_m \mid \xi, x) \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\partial^2}{\partial x^2} \left(\int_{\xi_i'} W(\xi_i' \mid \boldsymbol{\xi}, \sigma)\, \mathrm{d}F(\xi_i' \mid \xi_i, x) \right) &= - \int \frac{\partial^2}{\partial x^2} F(\xi' \mid \xi, x)\, \mathrm{d}W(\xi' \mid \boldsymbol{\xi}, \sigma) \\ &\quad - W(\xi_m \mid \boldsymbol{\xi}, \sigma) \frac{\partial^2}{\partial x^2} F(\xi_m \mid \xi, x) \\ &< 0 \end{aligned}$$

The differentiation under the integral sign is valid due to the twice continuous differentiability of $F(\xi' \mid \xi, x)$ in x. The inequality is due to Assumption 1 and the monotonicity of W. This inequality, coupled with the strict convexity of $c(x, \nu)$, implies the desired property. □

A.3 Characterizing EV

This section derives equation (29).[33] We start from $V_I(\boldsymbol{\xi}, \rho)$:

$$V_I(\boldsymbol{\xi}, \rho) = \max \left\{ \pi(\boldsymbol{\xi}) + \rho, \bar{V}_I^A(\boldsymbol{\xi}) \right\} = \max_{\chi \in \{0,1\}} \left\{ \chi \bar{V}_I^A(\boldsymbol{\xi}) + (1 - \chi)(\pi(\boldsymbol{\xi}) + \rho) \right\} \quad (43)$$

[33] Related calculations appear e.g. in Jofre-Bonet and Pesendorfer (2003); Pakes et al. (2007).

Letting $\alpha^I(\boldsymbol{\xi},\rho)$ denote the optimal policy, we have

$$V_I(\boldsymbol{\xi},\rho) = \alpha^I(\boldsymbol{\xi},\rho)\bar{V}_I^A(\boldsymbol{\xi}) + (1-\alpha^I(\boldsymbol{\xi},\rho))(\pi(\boldsymbol{\xi})+\rho)$$

and, integrating over ρ,

$$\begin{aligned}\bar{V}_I(\boldsymbol{\xi}) &:= \int V_I(\boldsymbol{\xi},\rho)\,\mathrm{d}F_\rho \\ &= \mathbb{P}_I^A(\boldsymbol{\xi})\bar{V}_I^A(\boldsymbol{\xi}) + [1-\mathbb{P}_I^A(\boldsymbol{\xi})]\pi(\boldsymbol{\xi}) \\ &+ [1-\mathbb{P}_I^A(\boldsymbol{\xi})]\mathbb{E}[\rho \mid \rho > \bar{V}_I^A(\boldsymbol{\xi}) - \pi(\boldsymbol{\xi})] \end{aligned} \tag{44}$$

where $\mathbb{P}_I^A(\boldsymbol{\xi}) := \mathbb{P}(\alpha^I(\boldsymbol{\xi},\rho)=1)$ is the probability that an incumbent chooses to be active when its initial state is $\boldsymbol{\xi}$.

Next, letting $\sigma^x(\boldsymbol{\xi},\nu)$ denote the optimal investment policy, we have that

$$V^A(\boldsymbol{\xi},\nu) = \pi(\boldsymbol{\xi}) - c(\sigma^x(\boldsymbol{\xi},\nu),\nu;\theta_x) + \beta \sum_{\xi_1'\in\Xi} W(\xi_1' \mid \boldsymbol{\xi},\theta_{-\phi})P(\xi_1' \mid \xi_1, \sigma^x(\boldsymbol{\xi},\nu)) \tag{45}$$

where, with slight abuse of notation, $P(\xi_1' \mid \xi, x)$ denotes the probability of the firms own characteristic evolving from ξ_1 to ξ_1' when the firm invests x. Integrating (45) we get

$$\bar{V}_I^A(\boldsymbol{\xi}) = \pi(\boldsymbol{\xi}) - \int c(\sigma^x(\boldsymbol{\xi},\nu),\nu;\theta_x)\,\mathrm{d}F_\nu + \beta \sum_{\xi_1'\in\Xi} W(\xi_1' \mid \boldsymbol{\xi},\theta_{-\phi})\mathbb{P}^A(\xi_1' \mid \boldsymbol{\xi}) \tag{46}$$

where $\mathbb{P}^A(\xi_1' \mid \boldsymbol{\xi}) := \int P(\xi_1' \mid \xi_1, \sigma^x(\boldsymbol{\xi},\nu))\,\mathrm{d}F_\nu$ is the ex-ante probability of ξ_1' given that the firm chooses to be active and investments optimally in state $\boldsymbol{\xi}$. Then, using the definition of $W(\xi_1' \mid \boldsymbol{\xi},\theta_{-\phi})$, we have that

$$\begin{aligned}\bar{V}_I^A(\boldsymbol{\xi}) &= \pi(\boldsymbol{\xi}) - \int c(\sigma^x(\boldsymbol{\xi},\nu),\nu;\theta_x)\,\mathrm{d}F_\nu \\ &+ \beta \sum_{\xi_1'\in\Xi}\left(\sum_{\boldsymbol{\xi}_{-1}'} \bar{V}_I(\xi_1',\boldsymbol{\xi}_{-1}') \prod_{k>1} \mathbb{P}^\sigma(\xi_k' \mid \boldsymbol{\xi})\right)\mathbb{P}^A(\xi_1' \mid \boldsymbol{\xi})\ .\,, \end{aligned} \tag{47}$$

where $P^\sigma(\xi_k' \mid \boldsymbol{\xi})$ is defined in equation (33).

We now plug (47) into (44) to obtain

$$\begin{aligned}\bar{V}_I(\boldsymbol{\xi}) &= \pi(\boldsymbol{\xi}) - \mathbb{P}_I^A(\boldsymbol{\xi}) \int c(\sigma^x(\boldsymbol{\xi},\nu),\nu;\theta_x)\,\mathrm{d}F_\nu \\ &\quad + [1-\mathbb{P}_I^A(\boldsymbol{\xi})]\mathbb{E}[\rho \mid \rho > \bar{V}_I^A(\boldsymbol{\xi}) - \pi(\boldsymbol{\xi})] \\ &\quad + \beta \sum_{\{\boldsymbol{\xi}':\xi_1' > -\infty\}} \bar{V}_I(\boldsymbol{\xi}') \prod_{k=1}^N \mathbb{P}^\sigma(\xi_k' \mid \boldsymbol{\xi}) \end{aligned} \tag{48}$$

where the last line uses the fact that $\mathbb{P}^\sigma(\xi_1' \mid \boldsymbol{\xi}) = \mathbb{P}^A(\xi_1' \mid \boldsymbol{\xi})\mathbb{P}_I^A(\boldsymbol{\xi})$. Note that we can also make the sum in the last line over all $\boldsymbol{\xi}'$ if we define $\bar{V}_I(\boldsymbol{\xi}') = 0$ when $\xi_1' = -\infty$. With this convention it is more accurate to interpret $\bar{V}_I(\boldsymbol{\xi})$ as the ENPV of landing in state $\boldsymbol{\xi}$ (given our assumption that firms that exit perish), rather than starting a period from state $\boldsymbol{\xi}$.

Observe that $\bar{\boldsymbol{V}}_I := [\bar{V}_I(\boldsymbol{\xi}) : \boldsymbol{\xi} \in \boldsymbol{\Xi}^R, \xi_1 > -\infty]$ enters equation (48) in a non-linear fashion through the $\mathbb{E}[\rho \mid \rho > \bar{V}^A(\boldsymbol{\xi}) - \pi(\boldsymbol{\xi})]$ term. Fortunately, assuming that F_ρ is strictly increasing, we can deal with that term as follows:

$$\begin{aligned}\mathbb{E}[\rho \mid \rho > \bar{V}^A(\boldsymbol{\xi}) - \pi(\boldsymbol{\xi})] &= \mathbb{E}[\rho \mid F_\rho(\rho) > F_\rho(\bar{V}^A(\boldsymbol{\xi}) - \pi(\boldsymbol{\xi}))] \\ &= \mathbb{E}[\rho \mid F_\rho(\rho) > \mathbb{P}_I^A(\boldsymbol{\xi})] \\ &= \mathbb{E}[\rho \mid \rho > F_\rho^{-1}(\mathbb{P}_I^A(\boldsymbol{\xi}))]\,, \end{aligned} \tag{49}$$

which does away with $\bar{V}^A(\boldsymbol{\xi})$. We can now plug (49) into (48) and stack across states with $\boldsymbol{\xi}_1 > -\infty$:

$$\bar{\boldsymbol{V}}_I = \boldsymbol{\pi} - \boldsymbol{K}(\boldsymbol{\theta}_x) + \boldsymbol{\Sigma}(F_\rho) + \beta \boldsymbol{M}(\boldsymbol{P})\bar{\boldsymbol{V}}_I \tag{29 – Repeated}$$

where the terms of this equation are defined in (30) to (33).

Finally, note that (29) does uniquely define $\bar{V}_I$ because the matrix $I - \beta\boldsymbol{M}(\boldsymbol{P})$ is invertible. Indeed, assume otherwise. Then there exists $x \in \boldsymbol{\Xi}_I^R \setminus \{\boldsymbol{0}\}$ such that $[I - \beta\boldsymbol{M}(\boldsymbol{P})]x = 0$, or $x = \beta\boldsymbol{M}x$. This implies $\|x\|_\infty = \|\beta\boldsymbol{M}x\|_\infty$. However, letting $(\boldsymbol{M}x)_i$ denote the i-th coordinate of $\boldsymbol{M}x$, we have

$$|(\boldsymbol{M}x)_i| = \left|\sum_{j=1}^{|\boldsymbol{\Xi}_I^R|} M_{ij}x_j\right| \le \sum_{j=1}^{|\boldsymbol{\Xi}_I^R|} M_{ij}|x_j| \le \|x\|_\infty \sum_{j=1}^{|\boldsymbol{\Xi}_I^R|} M_{ij} \le \|x\|_\infty\,,$$

where we have used that $\boldsymbol{M}$ is a sub-stochastic matrix, i.e. its rows sum to at most one.[34] Therefore $\|x\|_\infty = \|\beta\boldsymbol{M}x\|_\infty = \beta\|\boldsymbol{M}x\|_\infty \le \beta\|x\|_\infty$, a contradic-

[34]The rows of $\boldsymbol{M}(\boldsymbol{P})$ need not sum to one because $\boldsymbol{M}(\boldsymbol{P})$ is the matrix of transitions between

tion.

A.4 The Gradient of $Q(\theta)$

Let $Q(\boldsymbol{\theta}) = [\boldsymbol{\lambda}(\boldsymbol{\theta}) - \hat{\gamma}]'\Omega[\boldsymbol{\lambda}(\boldsymbol{\theta}) - \hat{\gamma}]$. The gradient of $Q(\theta)$ is

$$\nabla Q(\theta) = 2[\boldsymbol{\lambda}(\theta) - \hat{\gamma}]'\Omega D_\theta \boldsymbol{\lambda}(\theta) \tag{50}$$

We thus need to compute $D_\theta \boldsymbol{\lambda}(\theta)$, the jacobian of $\boldsymbol{\lambda}(\theta)$. We split this calculation in three parts, one for each subvector. We start with a few necessary preliminaries. In what follows, let $d(\boldsymbol{x})$ denote the dimension of $\boldsymbol{x}$.

A.4.1 Preliminaries

The calculation of the jacobians of $\boldsymbol{\lambda}^x, \boldsymbol{\lambda}^E$ and $\boldsymbol{\lambda}^N$ will require the jacobians of $\bar{\boldsymbol{V}}_I(\theta_{-\phi})$, the gradients of $W(\xi' \mid \boldsymbol{\xi}, \theta_{-\phi})$, and the jacobians of $\bar{V}^A(\boldsymbol{\xi}; \theta_{-\phi})$. We discuss those calculations in this subsection.

First, $\bar{\boldsymbol{V}}_I(\theta_{-\phi})$ satisfies equation (29), repeated here for convenience with the difference that we make it explicit that F_ρ depends on a (finite-dimensional) parameter vector θ_ρ:

$$[\boldsymbol{I} - \beta \boldsymbol{M}(\boldsymbol{P})]\bar{\boldsymbol{V}}_I(\theta_{-\phi}) = \boldsymbol{\pi} - \boldsymbol{K}(\boldsymbol{\theta}_x) + \boldsymbol{\Sigma}(F_\rho(\theta_\rho)) \tag{29 – repeated}$$

where

$$\boldsymbol{K}(\theta_x) = \left[\mathbb{P}_I^A(\boldsymbol{\xi}) \int c(\sigma^x(\boldsymbol{\xi}, \nu), \nu; \theta_x)\, \mathrm{d}F_\nu \ : \ \boldsymbol{\xi} \in \Xi_I^R\right] \tag{30 – repeated}$$

and

$$\boldsymbol{\Sigma}(F_\rho) = \left[[1 - \mathbb{P}_I^A(\boldsymbol{\xi})]\mathbb{E}[\rho \mid \rho > F_\rho^{-1}(\mathbb{P}_I^A(\boldsymbol{\xi}))] \ : \ \boldsymbol{\xi} \in \Xi_I^R\right] \tag{31 – repeated}$$

Therefore, to compute the jacobian of $\bar{\boldsymbol{V}}_I(\theta_{-\phi})$ we need to compute the jacobian of the vector of expected flow profits, the right-hand side of equation (29). This jacobian is

$$\left[D_{\theta_x}\boldsymbol{K}(\theta_x) \quad D_{\theta_\rho}\boldsymbol{\Sigma}(F_\rho(\theta_\rho)) \quad \mathbf{0}_{|\Omega| \times d(\theta_\phi)}\right] .$$

states in $\Xi_I^R = \{\boldsymbol{\xi} \in \Xi^R : \xi_1 > -\infty\}$ rather than Ξ^R.

In this expression, the first term is

$$D_{\theta_x}\boldsymbol{K}(\theta_x) = \left[\mathbb{P}_I^A(\boldsymbol{\xi}) \int \nabla_{\theta_x} c(\sigma^x(\boldsymbol{\xi},\nu),\nu;\theta_x)\, \mathrm{d}F_\nu \;:\; \boldsymbol{\xi} \in \Xi_I^R\right] . \tag{51}$$

The second term, $D_{\theta_\rho}\boldsymbol{\Sigma}(F_\rho(\theta_\rho))$, will depend on the parametric assumptions made on F_ρ. In the lognormal case, the jacobian $D_{\theta_\rho}\boldsymbol{\Sigma}(F_\rho(\theta_\rho))$ can be calculated using formulas shown below for the quantiles of the distribution and for the lower-truncated conditional expectations. The third term in the matrix above is a zero matrix corresponding to the derivatives with respect to θ_ϕ. Once we have the jacobian of expected flow profits we can compute the jacobian of $\bar{\boldsymbol{V}}_I(\theta_{-\phi})$ by solving the system of linear equations above.[35]

Next, remember that

$$W(\xi_1' \mid \boldsymbol{\xi}, \theta_{-\phi}) = \sum_{\boldsymbol{\xi}_{-1}'} \bar{V}_I(\boldsymbol{\xi}';\theta_{-\phi}) \prod_{k>1} P^\sigma(\xi_k' \mid \boldsymbol{\xi}) .$$

Therefore, we compute $\nabla_\theta W(\xi_1' \mid \boldsymbol{\xi}, \theta_{-\phi})$ by simply taking the corresponding convex combination of the gradients of $\bar{V}_I(\boldsymbol{\xi};\theta_{-\phi})$ terms computed above.[36]

Finally, $\bar{V}^A(\boldsymbol{\xi};\theta_{-\phi})$ is given by

$$\bar{V}^A(\boldsymbol{\xi};\theta_{-\phi}) = \pi(\boldsymbol{\xi}) - \mathbb{E}_\nu[c(\sigma^x(\boldsymbol{\xi},\nu),\nu;\theta_x)] + \sum_{\xi' \in \Xi : \xi' > -\infty} W(\xi' \mid \boldsymbol{\xi}, \theta_{-\phi}) P^A(\xi' \mid \boldsymbol{\xi}) , \tag{52}$$

where $P^A(\xi' \mid \boldsymbol{\xi})$ has been defined in section A.3. We can thus compute the jacobian of $\bar{\boldsymbol{V}}^A$ using the previously computed $\nabla_\theta W(\xi' \mid \boldsymbol{\xi}, \theta_{-\phi})$. This concludes the preliminaries required for the derivations that follow.

A.4.2 The Jacobian of λ^x

Remember that

$$\hat{\lambda}^x(\theta_{-\phi}) := \begin{bmatrix} \hat{\lambda}_1^x(\theta_{-\phi}) & \ldots & \hat{\lambda}_B^x(\theta_{-\phi}) & S_{\zeta^x}(\theta_{-\phi}) \end{bmatrix} .$$

[35] We remind the reader that we compute a decomposition of the matrix $\boldsymbol{I} - \beta \boldsymbol{M}(\boldsymbol{P})$ once and store that in memory, so computing this jacobian is not an expensive operation. Moreover, if $c(x,\nu;\theta_x)$ is linear in θ_x (as in the HvB model we simulate in section 5), the first submatrix of (51) does not depend on θ_x. This reduces the number of linear systems that have to be solved to compute the jacobian of $\bar{\boldsymbol{V}}_I(\theta_{-\phi})$.

[36] The non-linearity of $\bar{V}_I(\theta_{-\phi})$ affects the cost of these computations. Under linearity, these gradients do not depend on the parameters and can be computed only once.

Therefore

$$\underset{(k_x+1)\times d(\theta)}{D_\theta \hat{\lambda}^x(\theta)} = \begin{pmatrix} \underset{k_x \times d(\theta)}{D_\theta \hat{\lambda}^x_{\boldsymbol{\xi}}(\theta)} \\ \underset{1\times d(\theta)}{\nabla_\theta S_{\zeta^x}(\theta)} \end{pmatrix} ,$$

where $\hat{\lambda}^x_{\boldsymbol{\xi}}(\theta)$ is the vector collecting $\hat{\lambda}^x_1$ to $\hat{\lambda}^x_B$ and where we make the functions depend on the entire vector θ even though they only depend on $\theta_{-\phi}$ because we want to obtain the jacobian with respect to the entire vector of parameters; the final $d(\theta_\phi)$ columns of the jacobian are zero.

The object $\hat{\lambda}^x_{\boldsymbol{\xi}}(\theta)$ is an OLS estimate, and thus satisfies

$$(\boldsymbol{X}'\boldsymbol{X})\hat{\lambda}^x_{\boldsymbol{\xi}}(\theta) = \boldsymbol{X}'\sigma^x(\theta)$$

where $\boldsymbol{X}$ is the matrix of features

$$\boldsymbol{X} = \begin{pmatrix} \Psi^x_1(\boldsymbol{\xi}_1) & \dots & \Psi^x_{B^x}(\boldsymbol{\xi}_1) \\ \vdots & \ddots & \vdots \\ \Psi^x_1(\boldsymbol{\xi}_N) & \dots & \Psi^x_{B^x}(\boldsymbol{\xi}_N) \end{pmatrix}$$

and

$$\sigma^x(\theta) = \begin{pmatrix} \sigma^x(\boldsymbol{\xi}_1, \nu_1; \theta_{-\phi}) \\ \vdots \\ \sigma^x(\boldsymbol{\xi}_N, \nu_N; \theta_{-\phi}) \end{pmatrix} .$$

In these equations, $\boldsymbol{\xi}_1$ to $\boldsymbol{\xi}_N$ are the states observed in the data, ν_1 to ν_N are draws from F_ν, and $\sigma^x(\boldsymbol{\xi}_i, \nu_i; \theta_{-\phi})$ was defined in section 3.[37] We therefore have that,

$$(\boldsymbol{X}'\boldsymbol{X})D_\theta\hat{\lambda}^x_{\boldsymbol{\xi}}(\theta) = \boldsymbol{X}'D_\theta\sigma^x(\theta) \tag{53}$$

where

$$D_\theta\sigma^x(\theta) = \begin{pmatrix} \nabla_\theta\sigma^x(\boldsymbol{\xi}_1, \nu_1; \theta_{-\phi}) \\ \vdots \\ \nabla_\theta\sigma^x(\boldsymbol{\xi}_N, \nu_N; \theta_{-\phi}) \end{pmatrix} .$$

By the Implicit Function Theorem, the gradients in the matrix above are given by

$$\nabla_\theta\sigma^x(\boldsymbol{\xi}, \nu; \theta_{-\phi}) = -\left[\frac{\partial f}{\partial x}(x^*, \boldsymbol{\xi}, \nu; \theta_{-\phi})\right]^{-1} \nabla_\theta f(x^*, \boldsymbol{\xi}, \nu; \theta_{-\phi})$$

[37] We draw the ν_i shocks once and store them in memory so that all parameter values use the same ν_i shocks.

where $x^* = \sigma^x(\boldsymbol{\xi}, \nu; \theta_{-\phi})$ and $f(\cdot)$ is the investment first-order condition, namely

$$f(x, \boldsymbol{\xi}, \nu; \theta_{-\phi}) := -\frac{\partial c(x, \nu; \theta_x)}{\partial x} + \beta \sum_{\{\xi' \in \Xi \,:\, \xi' > -\infty\}} W(\xi' | \boldsymbol{\xi}, \theta_{-\phi}) \frac{\partial P(\xi' | \xi, x)}{\partial x} ,$$

The derivative $\partial f / \partial x$ is immediate from the previous equation. The gradient $\nabla_\theta f$ follows from the previous equation and the calculation of $\nabla_\theta W$ in section A.4.1. We can then solve for $D_\theta \hat{\lambda}^x_{\boldsymbol{\xi}}$ from equation 53.

Next, we need $\nabla_\theta S_{\zeta^x}(\theta)$. We have

$$S_{\zeta^x}(\theta) = \left\{ \frac{1}{n - B^x} \sum_{i=1}^{n} [\sigma^x(\boldsymbol{\xi}, \nu; \theta_{-\phi}) - \boldsymbol{\Psi}^x(\boldsymbol{\xi}_i)' \hat{\lambda}^x_{\boldsymbol{\xi}}(\theta)]^2 \right\}^{\frac{1}{2}}$$

where $\boldsymbol{\Psi}^x(\boldsymbol{\xi}_i) := (\Psi^x_1(\boldsymbol{\xi}_i) \ldots \Psi^x_{B^x}(\boldsymbol{\xi}_i))'$. Therefore,

$$\begin{aligned} \nabla_\theta S_{\zeta^x}(\theta) = & \frac{1}{S_{\zeta^x}(\theta)(n - B^x)} \\ & \times \sum_{i=1}^{n} [\sigma^x(\boldsymbol{\xi}, \nu; \theta_{-\phi}) - \boldsymbol{\Psi}^x(\boldsymbol{\xi}_i)' \hat{\lambda}^x_{\boldsymbol{\xi}}(\theta)] \big\{ \nabla_\theta \sigma^x(\boldsymbol{\xi}, \nu; \theta_{-\phi}) - \nabla_\theta [\boldsymbol{\Psi}^x(\boldsymbol{\xi}_i)' \hat{\lambda}^x_{\boldsymbol{\xi}}(\theta)] \big\} \end{aligned}$$

The first gradient in these expressions has already been characterized. The second gradient is

$$\nabla_\theta [\boldsymbol{\Psi}^x(\boldsymbol{\xi}_i)' \hat{\lambda}^x_{\boldsymbol{\xi}}(\theta)] = \boldsymbol{\Psi}^x(\boldsymbol{\xi}_i)' D_\theta \hat{\lambda}^x_{\boldsymbol{\xi}}(\theta)$$

and $D_\theta \hat{\lambda}^x_{\boldsymbol{\xi}}(\theta)$ has just already characterized.

A.4.3 The Jacobian of λ^E

We have

$$\hat{\lambda}^E(\theta) = (X_E' X_E)^{-1} X_E' [1 - F_\rho(\hat{\bar{V}}^A(\boldsymbol{\xi}_i; \theta_{-\phi}) - \pi(\boldsymbol{\xi}_i); \theta_\rho)]_{i=1}^M$$

where X_E denotes the design matrix for the exit auxiliary model, i indexes observations, and M denotes the sample size (i.e., the number of observations in which the firm is active). Therefore,

$$D_\theta \hat{\lambda}^E(\theta) = (X_E' X_E)^{-1} X_E' [\nabla_\theta \mathbb{P}(\text{Exit} \mid \boldsymbol{\xi}_i; \theta)]_{i=1}^M$$

where $\mathbb{P}(\text{Exit} \mid \boldsymbol{\xi}_i; \theta) := 1 - F_\rho\left(\hat{\hat{V}}^A(\boldsymbol{\xi}_i; \theta_{-\phi}) - \pi(\boldsymbol{\xi}); \theta_\rho\right)$. Now observe that

$$\begin{aligned}\nabla_\theta \mathbb{P}(\text{Exit} \mid \boldsymbol{\xi}; \theta) = &-f_\rho(\hat{\hat{V}}^A(\boldsymbol{\xi}; \theta_{-\phi}) - \pi(\boldsymbol{\xi}); \theta_\rho)\nabla_\theta \hat{\hat{V}}^A(\boldsymbol{\xi}; \theta_{-\phi}) \\ &- \nabla_\theta F_\rho(x; \theta_\rho)\Big|_{x=\hat{\hat{V}}^A(\boldsymbol{\xi};\theta_{-\phi})-\pi(\boldsymbol{\xi})},\end{aligned}$$

where $f_\rho = F'_\rho$ is the density of the scrap value distribution and the second term is the gradient of the cumulative distribution function F_ρ evaluated at $\hat{\hat{V}}^A(\boldsymbol{\xi}; \theta_{-\phi}) - \pi(\boldsymbol{\xi})$. We adopt that notation to make clear that the gradient is with respect to parameters only, and not the gradient of the composite function $F_\rho(\hat{\hat{V}}^A(\boldsymbol{\xi}; \theta_{-\phi}) - \pi(\boldsymbol{\xi}); \theta_\rho)$. The terms f_ρ and $\nabla_\theta F_\rho$ are primitives. The term $\nabla_\theta \hat{\hat{V}}^A(\boldsymbol{\xi}; \theta_{-\phi})$ was characterized in section A.4.1.

The Jacobian of λ^N We have

$$\lambda^N(\theta) = (X'_N X_N)^{-1} X'_N \left[F_\phi\left(\hat{\hat{V}}^A_E(\boldsymbol{\xi}_{-1}; \boldsymbol{\theta}_{-\phi}); \boldsymbol{\theta}_\phi\right)\right]_{i=1}^M$$

where M denotes the sample size and X_N is the design matrix of the auxiliary entry model. Note that λ^N is the only auxiliary parameter that does depend on θ_ϕ. Therefore,

$$D_\theta \lambda^N(\theta) = (X'_N X_N)^{-1} X'_N [\nabla_\theta^T \mathbb{P}(\text{Entry}|\boldsymbol{\xi}_{-1}; \theta)]_{i=1}^M\,, \tag{54}$$

where $\mathbb{P}(\text{Entry}|\boldsymbol{\xi}_{-1}; \theta) := F_\phi(\hat{\hat{V}}^A_E(\boldsymbol{\xi}_{-1}; \boldsymbol{\theta}_{-\phi}); \boldsymbol{\theta}_\phi)$. Similarly to the exit case, we have

$$\begin{aligned}\nabla_\theta \mathbb{P}(\text{Entry} \mid \boldsymbol{\xi}_{-1}; \theta) = &f_\phi(\hat{\hat{V}}^A_E(\boldsymbol{\xi}_{-1}; \boldsymbol{\theta}_{-\phi}); \boldsymbol{\theta}_\phi)\nabla_\theta \hat{\hat{V}}^A_E(\boldsymbol{\xi}_{-1}; \theta_{-\phi}) \\ &+ \nabla_\theta F_\phi(x; \boldsymbol{\theta}_\phi)\Big|_{x=\hat{\hat{V}}^A_E(\boldsymbol{\xi}_{-1};\boldsymbol{\theta}_{-\phi})}.\end{aligned}$$

The terms f_ϕ and ∇_θ are primitives. The term $\nabla_\theta \hat{\hat{V}}^A_E(\boldsymbol{\xi}_{-1}; \theta_{-\phi})$ was characterized in section A.4.1.

Appendix B Monte Carlo: Additional Results

This appendix presents estimate histograms for each BBL estimator separately, as well as run time and memory allocation for the different steps of our simulation averaged over 500 runs.

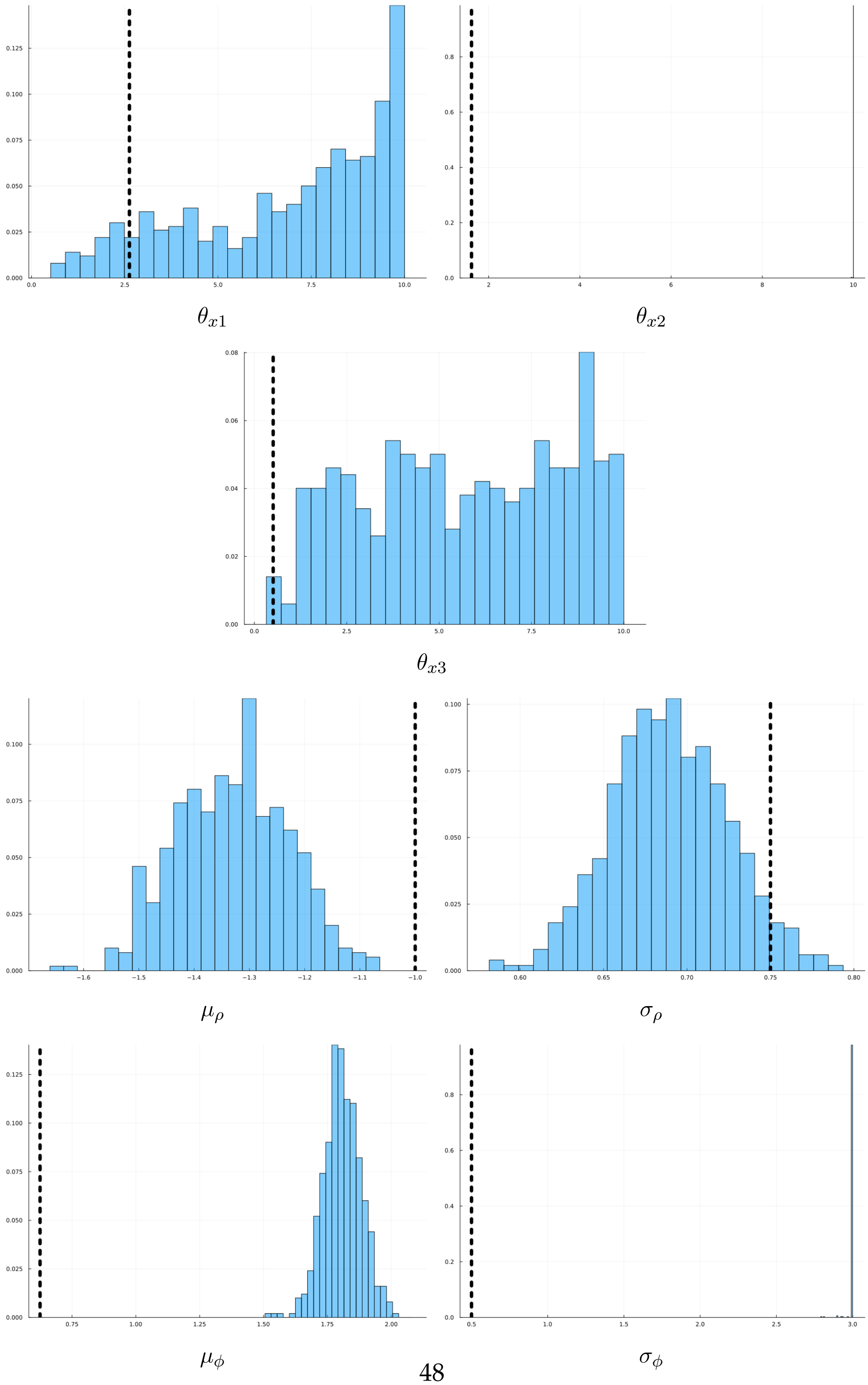
A caveat to comparing estimator performance based on Table 3 is that we tune estimator termination conditions to ensure all of the estimators are guaranteed comparable time per simulation. As a result, there should be – by construction – no clear winner in average time.

Table 3: Exercise Timings

Call	Average Time	Average Allocation
Overall		
Solving the Game	41.1s	28.7GiB
First Stage (Shared by All Estimators)	94.2s	43.9GiB
Indirect Inference – Factorization	49.7s	21.0GiB
Indirect Inference – Bootstrapped Weight Matrix	22.1s	20.8GiB
Indirect Inference Second Stage	467s	24.0GiB
Estimation of θ_x and θ_ρ	463s	23.5GiB
Estimation of θ_ϕ	1.05s	421MiB
Bajari Benkard Levin Second Stage: Asymptotic	452s	420GiB
Incumbent Forward Simulation	236s	289GiB
Estimation of θ_x and θ_ρ	145s	52.5GiB
Entrant Forward Simulation	14.5s	13.7GiB
Estimation of θ_ϕ	68.0ms	60.0MiB
Bajari Benkard Levin Second Stage: Multiplicative	480s	428GiB
Incumbent Forward Simulation	249s	291GiB
Estimation of θ_x and θ_ρ	154s	57.9GiB
Entrant Forward Simulation	14.5s	13.7GiB
Estimation of θ_ϕ	64.0ms	59.1MiB
Bajari Benkard Levin Second Stage: Additive	475s	442GiB
Incumbent Forward Simulation	223s	283GiB
Estimation of θ_x and θ_ρ	154s	58.2GiB
Entrant Forward Simulation	14.5s	13.7GiB
Estimation of θ_ϕ	64.3ms	57.9MiB

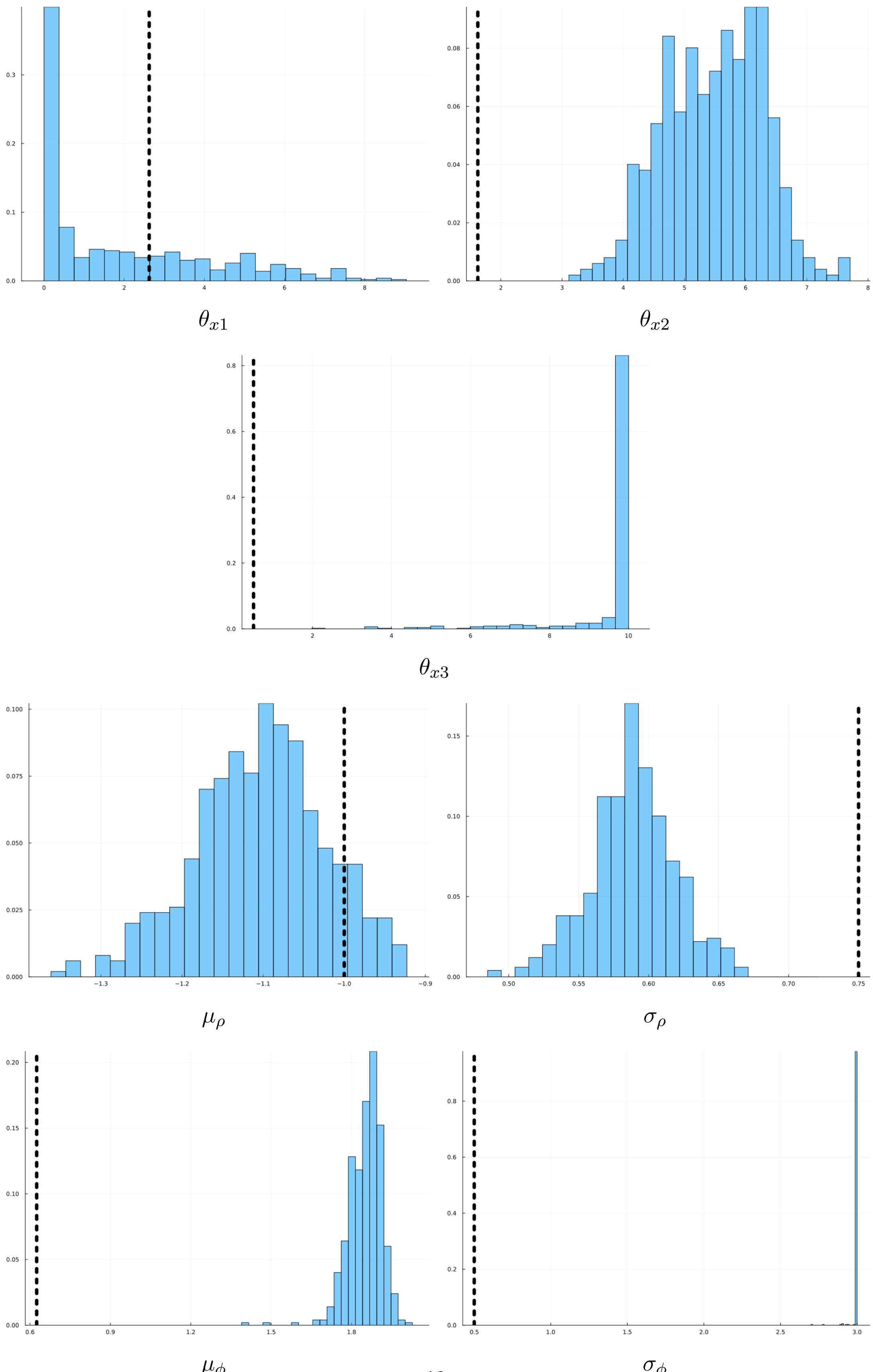
This table reports the computation time and memory allocation associated with each step in the computation of the four estimators discussed above. Note that the memory allocation reported here is not the amount of memory required to compute these estimators, but rather the total size of objects that are written to memory throughout the course of estimation. In “First Stage”, “Factorization” and “Bootstrapped Weight Matrix” are only necessary for the Indirect Inference estimator.

Figure 5: Asymptotic BBL Estimator Parameter Estimates



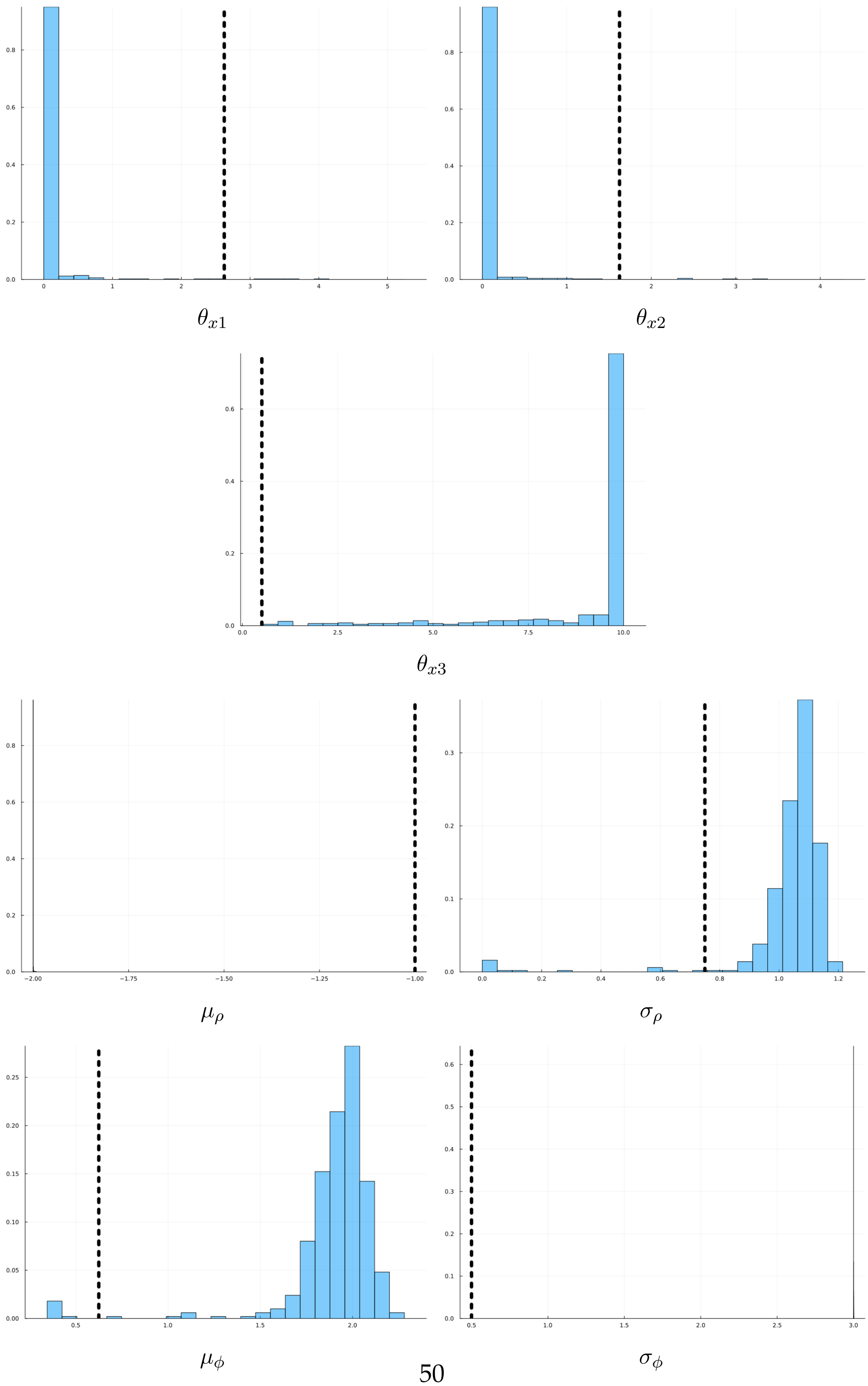
This figure plots the distribution of parameter estimates obtained using the 'Asymptotic' BBL estimator over 500 Monte Carlo replications. The vertical dashed red line indicates the value of the corresponding parameter in the data generating process.

Figure 6: Multiplicative BBL Estimator Parameter Estimates



This figure plots the distribution of parameter estimates obtained using the 'Multiplicative' BBL estimator over 500 Monte Carlo replications. The vertical dashed red line indicates the value of the corresponding parameter in the data generating process.

Figure 7: Additive BBL Estimator Parameter Estimates



This figure plots the distribution of parameter estimates obtained using the 'Additive' BBL estimator over 500 Monte Carlo replications. The vertical dashed red line indicates the value of the corresponding parameter in the data generating process.

Appendix C Small Dataset Monte Carlo Results

In this section we compare performance of the estimators presented in Section 5 for datasets with sample size in line with Ryan (2012). We consider data for 27 markets recorded over 20 periods. All other simulation parameters are as in Table 1.

We compare the performance of the four estimators presented in Section 5. Table 4 lists true parameter values along with estimate average and standard deviation for the four estimators over 500 simulations. Of the four estimators, the II estimator evidently performs the best, though with more finite sample bias and larger uncertainty than in Section 5. BBL estimators again display substantial bias and large variance. Figures 8 and 9 present the distribution of parameter estimates for each algorithm. II estimates are correctly centered around the true values. As in section 5, the scrap value and entry cost estimates display some skewness. BBL estimators are centered around wrong values, displaying very substantial finite sample bias.

We again relate the performance of the different estimators to the differences in the objective functions that define them. We illustrate this by analysing the shape of each objective function in a neighbourhood of the true parameters for a particular simulated dataset. To do so, we plot slices of the objective function by varying one parameter at a time while holding the others fixed at their true values. To render the shape of different objective functions comparable, we normalise objective values on the parameter grid by dividing each by the objective value at the true parameter. Vertical dashed lines represent true parameter values. Figures 10 and 11 display the results.

Figure 10 shows that, once again, the objective function of the II estimator generally features pronounced local convexity around each true parameter, with minimal distance between the local objective minimum and the objective value at the true parameter. This is reflected in the good performance of the estimator. On the other hand, the objective slice for σ_ρ is a reminder that things can go wrong in a single sample. The picture is very different for the BBL estimators in Figure 11. The objective slices with respect to cost parameter vector θ_x are sometimes flat around the true parameter, sometimes minimized far away from the truth, and sometimes monotonic over the range of values we consider. We also observe monotonicity when considering slices with respect to the entry cost parameters. These slices are consistent with the unsatisfactory summary results in Table 4. Importantly, observe again that Figure 11 shows that differ-

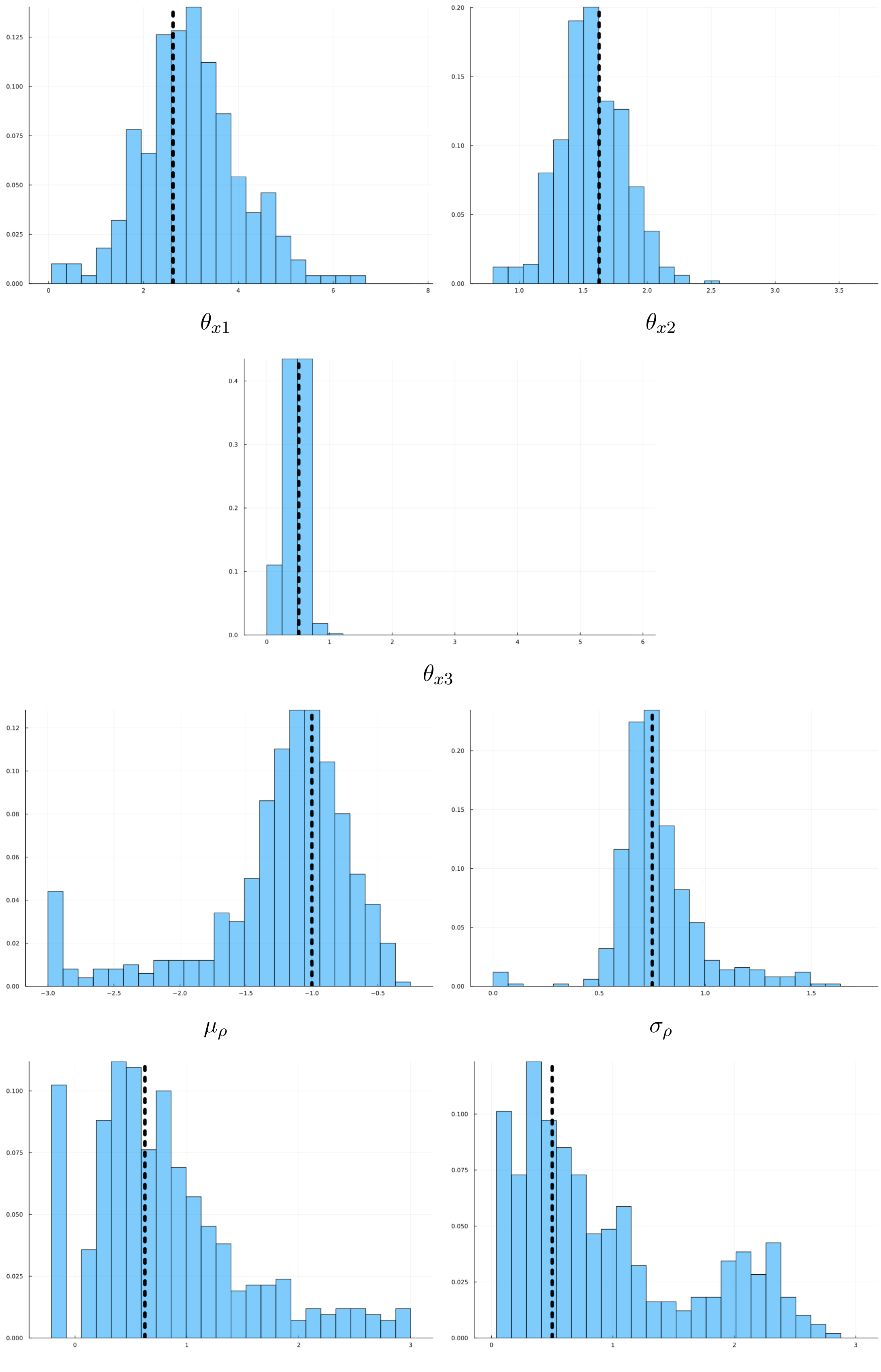
Table 4: Summary of Parameter Estimates

	Value	Indirect Inference	BBL		
			Asymptotic	Multiplicative	Additive
θ_{x1}	2.625	3.016	4.917	5.131	2.426
		1.066	2.624	2.784	2.256
θ_{x2}	1.624	1.57	6.71	5.444	1.666
		0.276	3.003	1.259	1.86
θ_{x3}	0.5096	0.452	5.582	5.521	5.459
		0.299	2.496	2.861	2.811
μ_ρ	-1.0	-1.253	-1.146	-1.175	-1.999
		0.587	0.19	0.187	0.008
σ_ρ	0.75	0.775	0.684	0.574	0.931
		0.21	0.069	0.072	0.314
μ_ϕ	0.625	1.161	1.878	1.646	1.679
		1.03	0.311	0.307	0.526
σ_ϕ	0.5	0.961	2.933	2.927	2.976
		0.762	0.187	0.188	0.107

This table summarizes the results of our Monte Carlo experiment. The first column shows the value of the investment cost, entry cost, and scrap value parameters used in the data generating process. Each subsequent column shows the mean and standard deviation for estimates across 500 Monte Carlo replications. The column labeled "II" shows the results of the estimator we propose in this paper. The columns labeled "Asymptotic", "Multiplicative", and "Additive" display estimates from the three BBL alternatives considered. See text for definitions.

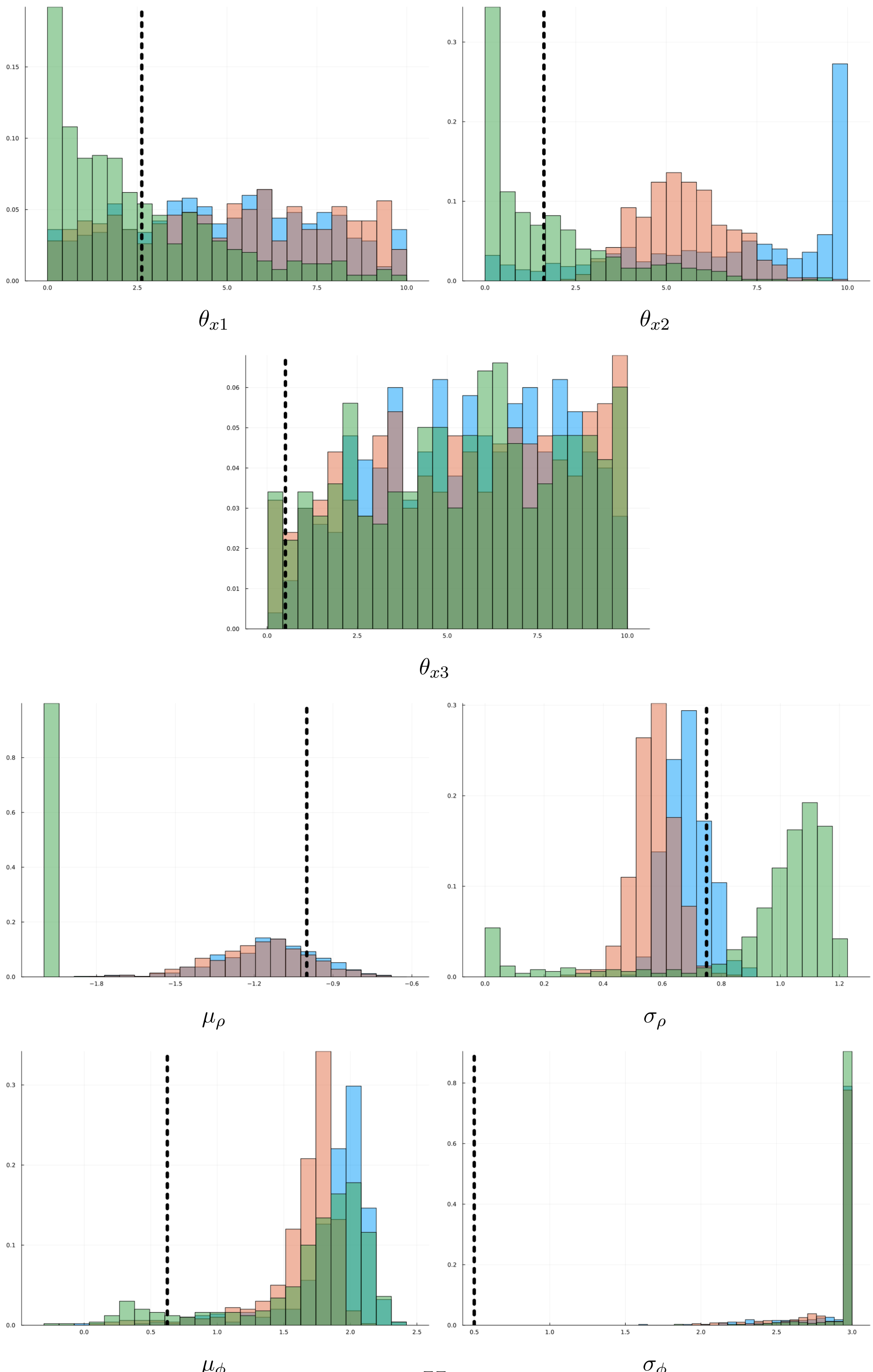
ent choices of deviations in the implementation of BBL will deliver estimators with different performance.

Figure 8: II Estimator Parameter Estimates, Small Dataset



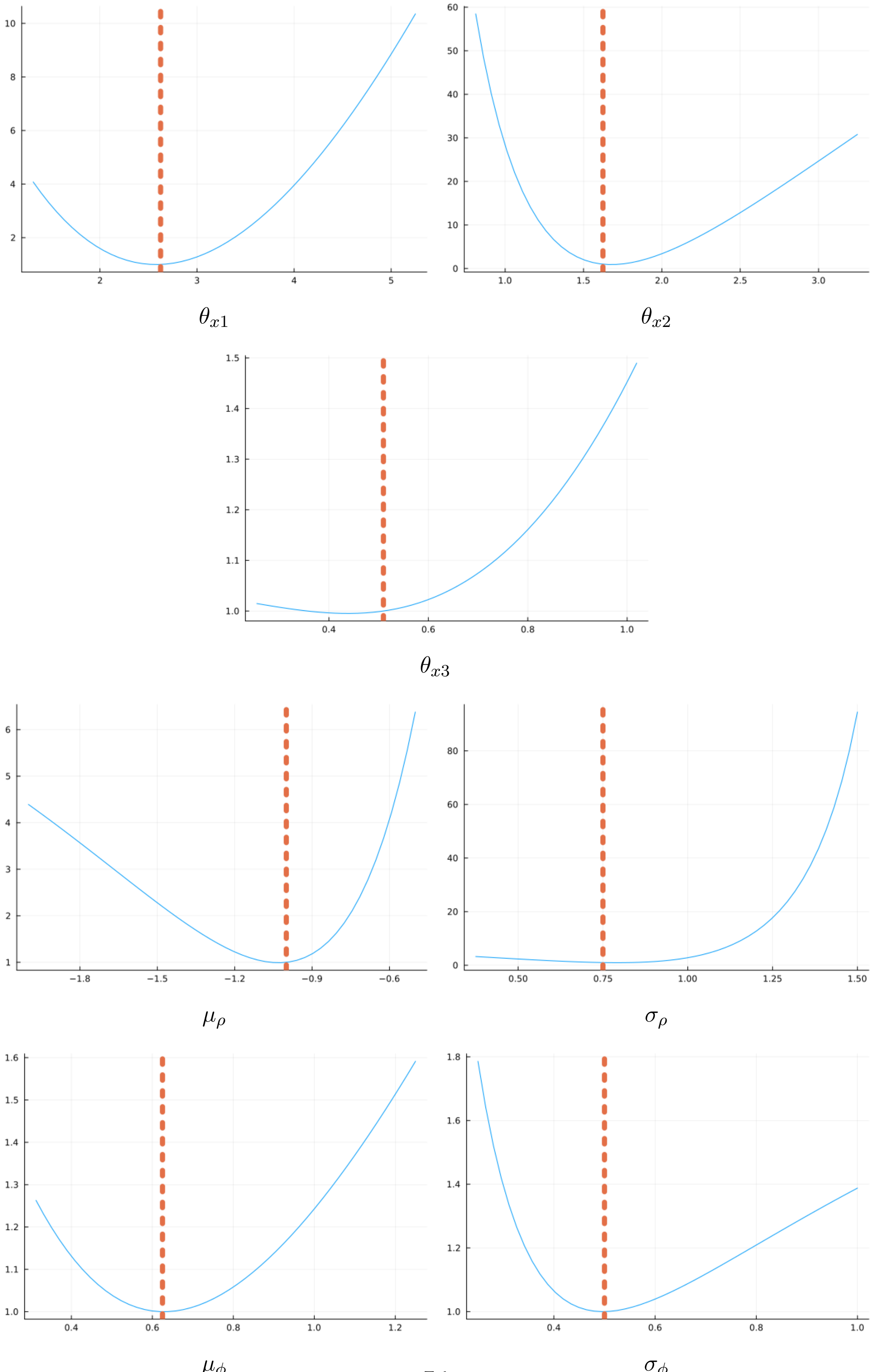
This figure plots the distribution of parameter estimates obtained using the indirect inference estimator defined in Equation (26) over 500 Monte Carlo replications. The vertical dashed red line indicates the value of the corresponding parameter in the data generating process.

Figure 9: BBL Estimator Parameter Estimates, Small Dataset



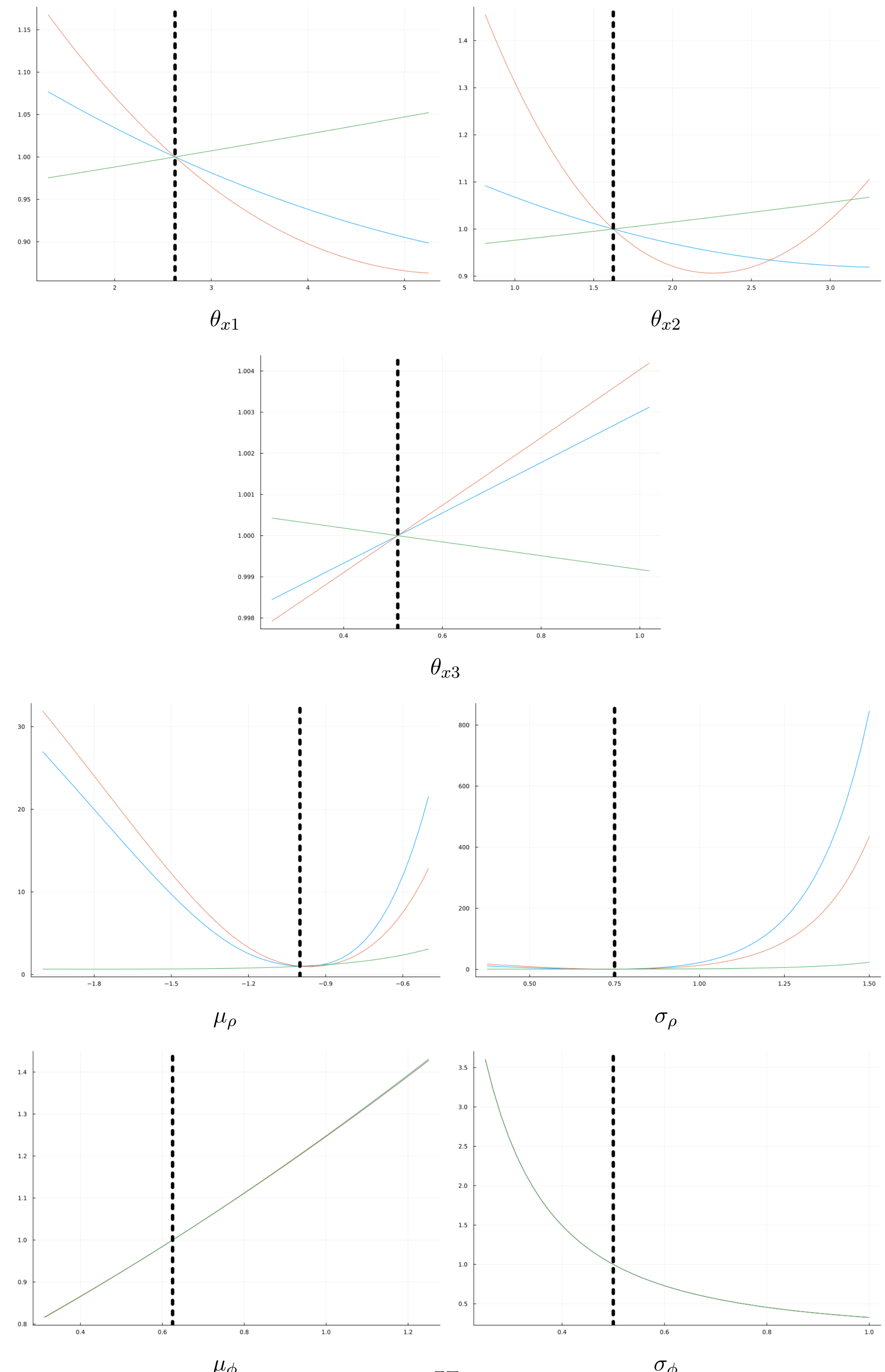
This figure plots the distribution of parameter estimates obtained using all BBL estimators over 500 Monte Carlo replications. 'Asymptotic' BBL in blue, 'Multiplicative' in red, 'Additive' in green. The vertical dashed red line indicates the value of the corresponding parameter in the data generating process.

Figure 10: II Estimator Objective Slices, Small Dataset



This figure plots the value of the objective function of the indirect inference estimator varying one parameter at a time while holding the other parameters fixed at their true values. The vertical dashed line indicates the value of the corresponding parameter in the data generating process.

Figure 11: BBL Estimator Objective Slices, Small Dataset



This figure plots the value of the objective function of the BBL estimators varying one parameter at a time while holding the other parameters fixed at their true values. 'Asymptotic' BBL in blue, 'Multiplicative' in red, 'Additive' in green. The vertical dashed line indicates the value of the corresponding parameter in the data generating process.

Appendix D Monte Carlo Results for a Model without Entry and Exit

In this Appendix we present Monte Carlo evidence on the performance of estimators discussed in Sections 3 and 4 for a model without entry and exit. The model is the same as the one presented in Section 2, aside for the fact that we assume the number of firms in each market is fixed over time (and this is known to firms).

We present results for two estimators that use the recursive equilibrium conditions. In addition to the II estimator discussed at length above, we also consider a nonlinear least squares (NLLS) estimator, as defined in equation (27).[38] One main reason to analyse estimator performance under this simpler setup is that the $g(\boldsymbol{\xi}, \sigma'; \sigma, \boldsymbol{\theta})$ terms entering the BBL objective function (37) are linear in parameters. This is important not only because it further reduces the computational burden of the estimator, but also because it allows us to consider the possibility that the BBL objective is not uniquely minimized without substantial complication.[39]

Indeed, when the incumbent value function is linear in parameters, BBL inequalities define a polyhedron that can be readily computed using existing software. If the polyhedron is found to be empty (i.e. there exists no parameter vector that satisfies all BBL inequalities), it is possible (but not guaranteed) that minimization of the BBL objective will return a single parameter vector. If however the polyhedron is found not to be empty, all parameter vectors lying within the polyhedron will set the BBL objective to zero.[40] For the purpose of this section we therefore first check whether the inequalities used in estimation characterize a non-empty polyhedron. If they do, we set estimate parameters by projection; otherwise, we minimise the BBL objective. Furthermore, because now each simulation can return a set rather than point estimates, we compute average lower and upper bounds across estimated sets instead of an average

[38] Other estimators could be considered. For instance, one could minimize the distance between estimated and predicted conditional expectations of investment. This would be the continuous-control analog of the Pesendorfer and Schmidt-Dengler (2008) estimator. In results not reported here we find that this estimator performs similarly to the NLLS estimator.

[39] See e.g. Aguirregabiria, Collard-Wexler, and Ryan (2021): "[...] in most applications of the BBL method, the relatively small set of alternative CCPs selected by the researcher does not provide enough moment inequalities to achieve point identification such that the BBL method provides set estimation of the structural parameters."

[40] When the incumbent value function is *not* linear in parameters, as in the case of the model with entry and exit, BBL inequalities still define a potentially non-empty set, but they do not define a polyhedron. Computation of the set estimate is then more cumbersome.

of point estimates. Also, we follow Manski and Tamer (2002) and compute, for each parameter, the shortest interval that covers 95% of the estimated sets.[41] While these steps may be of independent interest to empiricists using the BBL estimator, their purpose in our case is to clarify whether the poor performance of the BBL estimator in our Monte Carlo (both in Section 5 and in this Appendix) is limited identifying power resulting from choosing too few inequalities. This is not the case: across the 100 simulations presented in this Appendix, we always find the polyhedron defined by the BBL inequalities to be empty. We treat this result as evidence that an insufficient number of inequalities leading to a large set-estimate for the structural parameters does not explain the poor performance of BBL under our model and simulated data set, and we do not conduct the same analysis for the model with entry and exit in Section 5.[42]

Table 5: Summary of Parameter Estimates

	Value	Indirect Inference	BBL		NLLS
			Asymptotic	Multiplicative	
θ_{x1}	2.625	2.616	[0.0, 0.0]	[0.0, 0.0]	3.004
		(1.404, 3.621)	(0.0, 0.0)	(0.0, 0.0)	(1.561, 4.68)
θ_{x2}	1.624	1.594	[131.546, 131.546]	[3.762, 3.762]	1.524
		(1.017, 2.197)	(118.08, 145.552)	(2.567, 5.574)	(0.842, 2.167)
θ_{x3}	0.5096	0.498	[18.156, 18.156]	[6.0, 6.0]	1.234
		(0.393, 0.578)	(0.0, 35.586)	(0.0, 23.826)	(0.0, 3.519)

This table summarizes the results of the Monte Carlo experiment for the model without entry and exit. The first column shows the value of the parameters of the investment cost function in the data generating process. Each subsequent column shows the mean and standard error of estimates across Monte Carlo replications. The column labeled "II" shows the results of the Indirect Inference estimator. The column labeled "Asymptotic" shows the results of the Asymptotic BBL estimator. The column labeled "Multiplicative" shows the results of the Multiplicative BBL estimator. The column labeled "NLLS" shows the results of the estimator defined in Equation (27).

Table 5 lists true parameter values along with estimate average and standard deviation for the four estimators over 100 simulations. Of the four estimators, the II estimator evidently performs the best, though with more finite sample bias and larger uncertainty than in Section 5. The NLLS estimator and,

[41] We treat point estimates as sets with coinciding upper and lower bounds.

[42] Even if it were the case that our BBL implementations and number of inequalities led to large set-estimates, that would of course not invalidate the main message of this paper that the estimator based on recursive equilibrium conditions has superior performance relative to BBL, as we compare the two estimators while holding their computational cost fixed.

in particular, BBL estimators display substantial bias and larger standard deviation than the II estimator. Figures 12, 13, 14, and 15 present the distribution of parameter estimates for each algorithm; vertical dotted lines represent true parameter values. II estimates are correctly centered around the true values and seem close to Gaussian. Additive and Multiplicative BBL estimators are centered around wrong values, displaying a bias that can be orders of magnitude larger than the parameter to be estimated. The NLLS estimator performs substantially better than the two BBL implementations, but also significantly underperforms the II estimator.

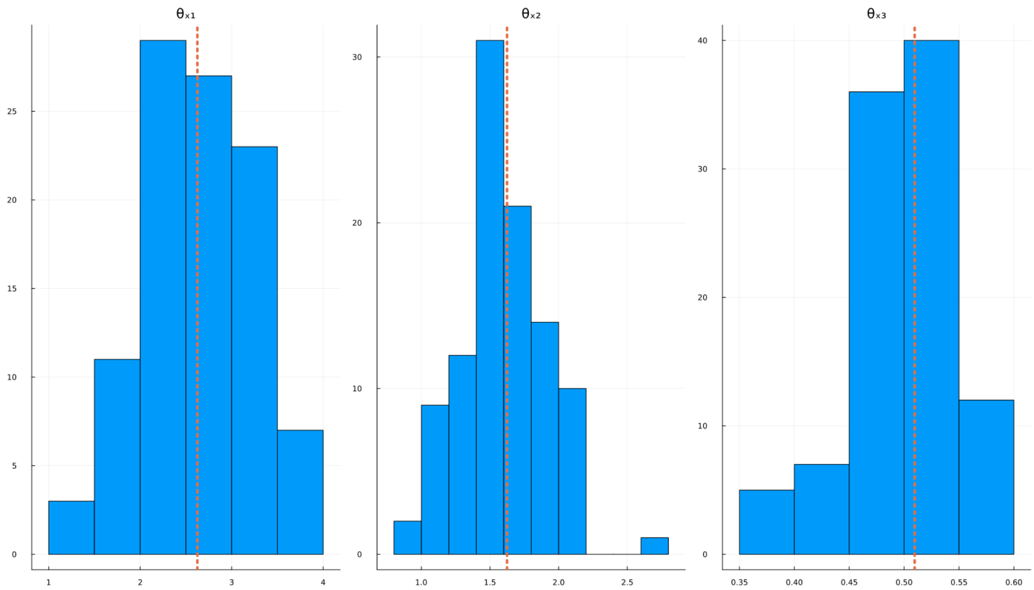
As in Section 5, we plot slices of the objective function by varying one parameter at a time while holding the others fixed at their true values. To render the shape of different objective functions comparable, we normalise objective values on the parameter grid by dividing each by the objective value at the true parameter. Figures 16 to 19 display the results. Vertical lines represent true parameter values.

Figure 16 shows that the objective function of the II estimator features pronounced local convexity around each true parameter, with minimal distance between the local objective minimum and the objective value at the true parameter. This is reflected in the good performance of the estimator, both in terms of short 95% confidence intervals and small bias.

The picture is very different for the HvB and BBL estimators. They are approximately flat around the true parameter, and zooming in reveals that they display no convexity at all for the considered grid – they are almost linear around the correct parameter values. This is consistent with the poor performance displayed in Figures 13 and 14.

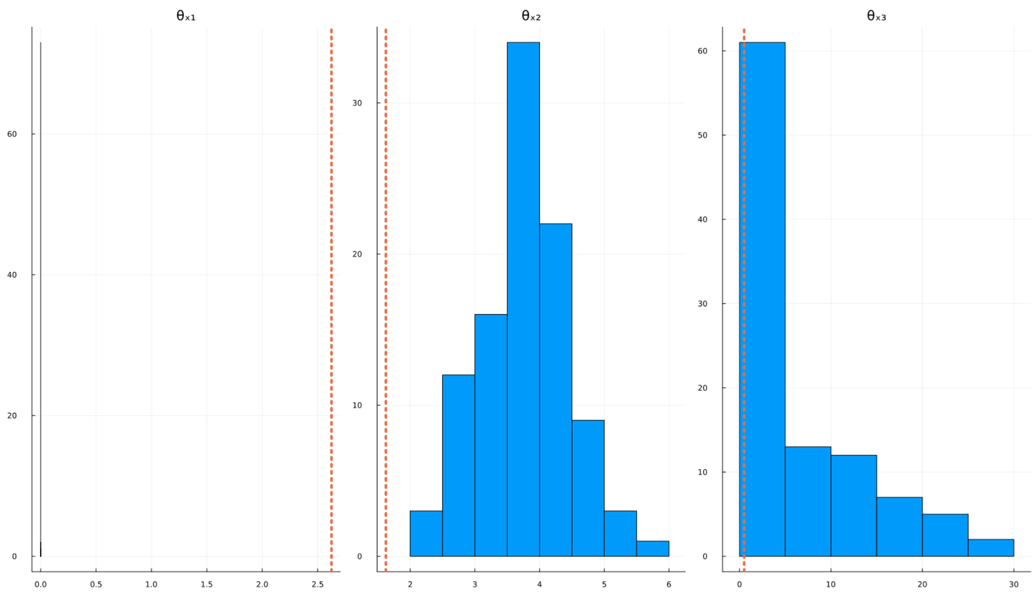
Finally, objective plots for the NLLS estimator are informative about its imprecision. Objective function minima are not far from the minima at the true parameter value (low bias) for θ_{x1}, θ_{x2}, but the NLLS objective function is much less convex around local minima than the II objective function. On the other hand, the objective function is not centered around the true parameter value for θ_{x3}.

Figure 12: II Estimator Parameter Estimates



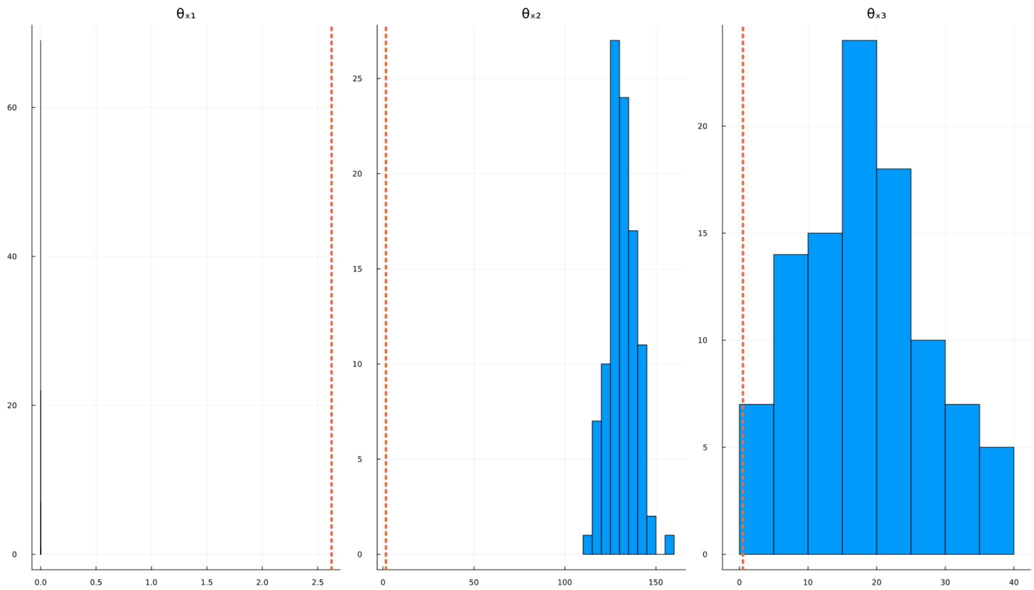
This figure plots the distribution of parameter estimates obtained using the II estimator described in Equation (26) over 100 Monte Carlo replications. The vertical dashed red line indicates the value of the corresponding parameter in the data generating process.

Figure 13: Multiplicative BBL Estimator Parameter Estimates



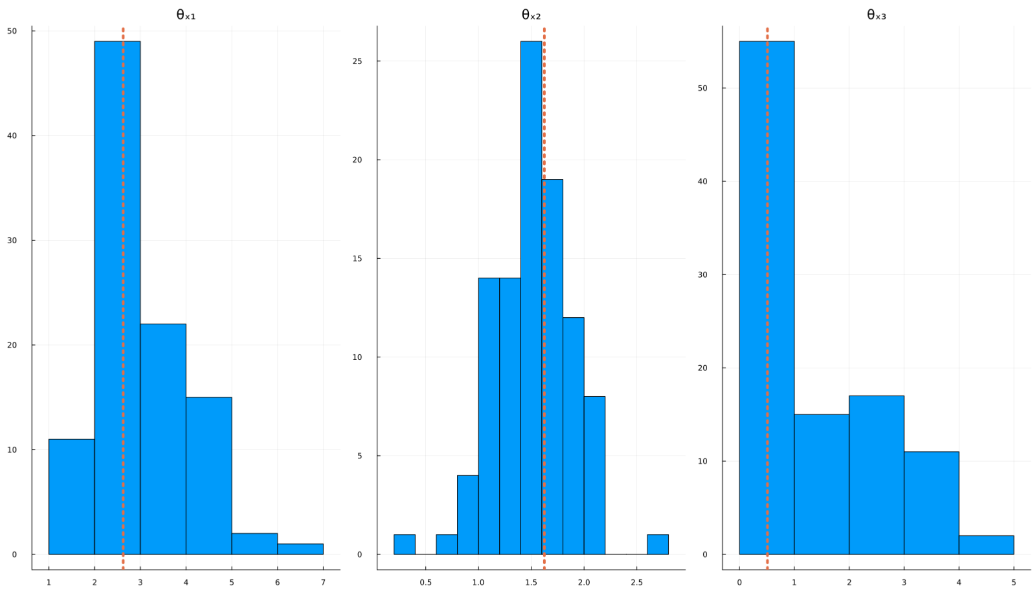
This figure plots the distribution of parameter estimates obtained using the Multiplicative BBL estimator over 100 Monte Carlo replications. The vertical dashed red line indicates the value of the corresponding parameter in the data generating process.

Figure 14: Additive BBL Estimator Parameter Estimates



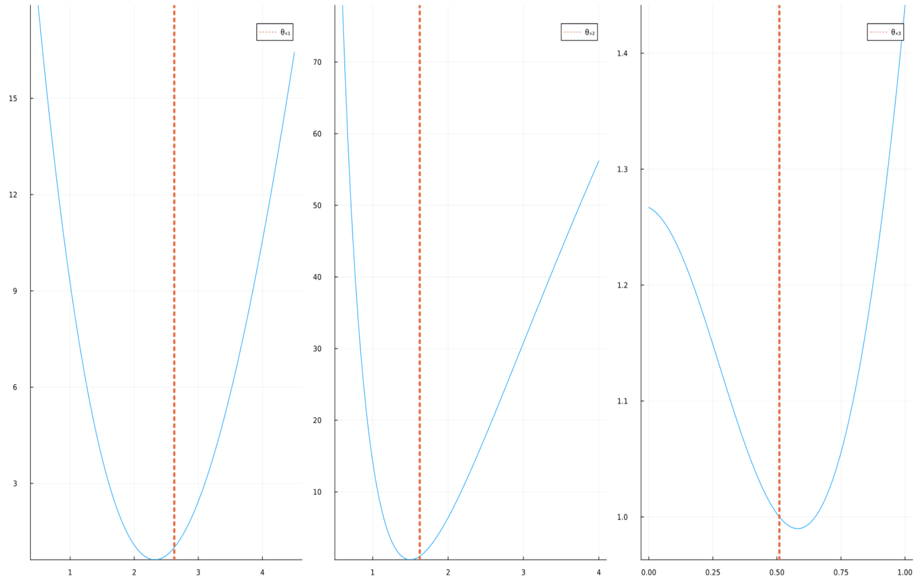
This figure plots the distribution of parameter estimates obtained using the Additive BBL estimator over 100 Monte Carlo replications. The vertical dashed red line indicates the value of the corresponding parameter in the data generating process.

Figure 15: NLLS Estimator Parameter Estimates



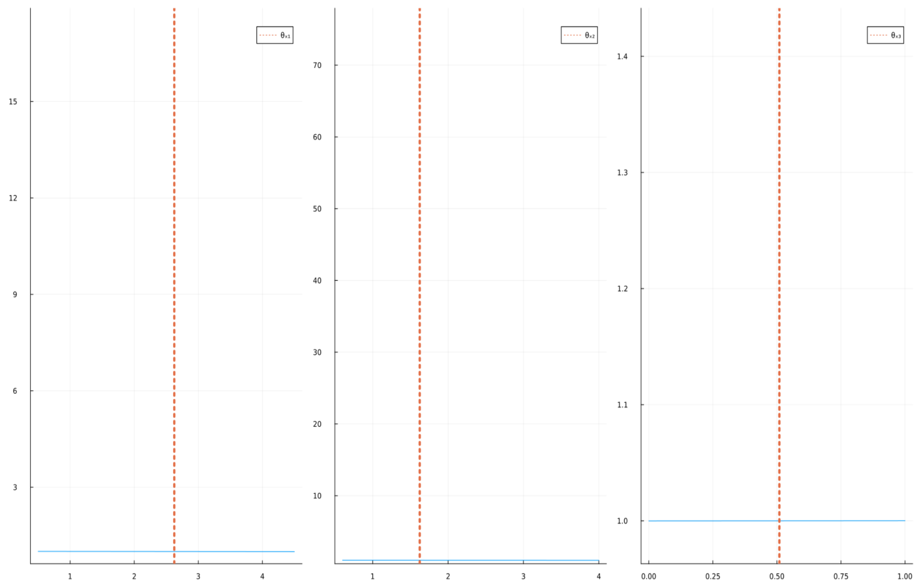
This figure plots the distribution of parameter estimates obtained using the NLLS estimator defined in Equation (27) over 100 Monte Carlo replications. The vertical dashed red line indicates the value of the corresponding parameter in the data generating process.

Figure 16: II Estimator Objective Slices



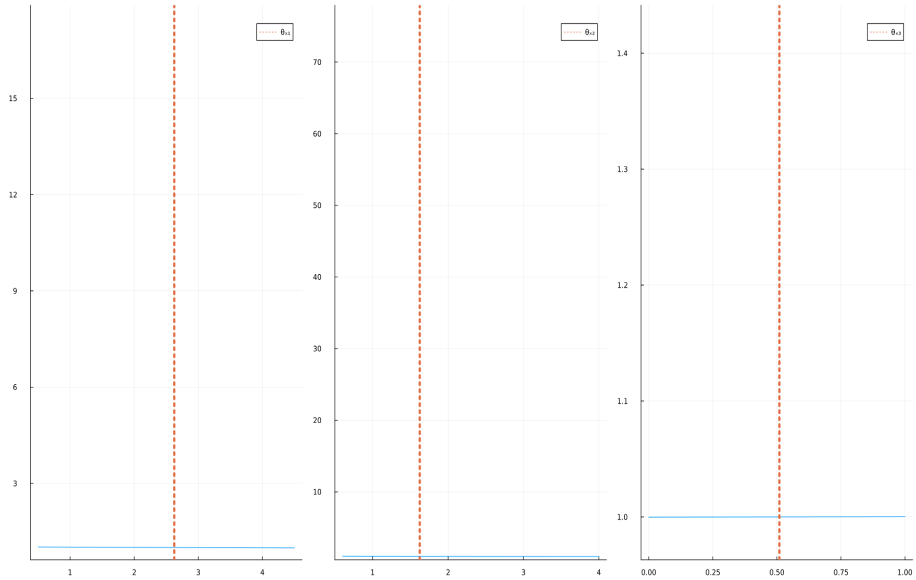
This figure plots the value of the II estimator objective function varying one parameter at a time while holding the other two parameters fixed at their true values. Values are scaled by the value of the objective at the true parameters.

Figure 17: Asymptotic BBL Estimator Objective Slices



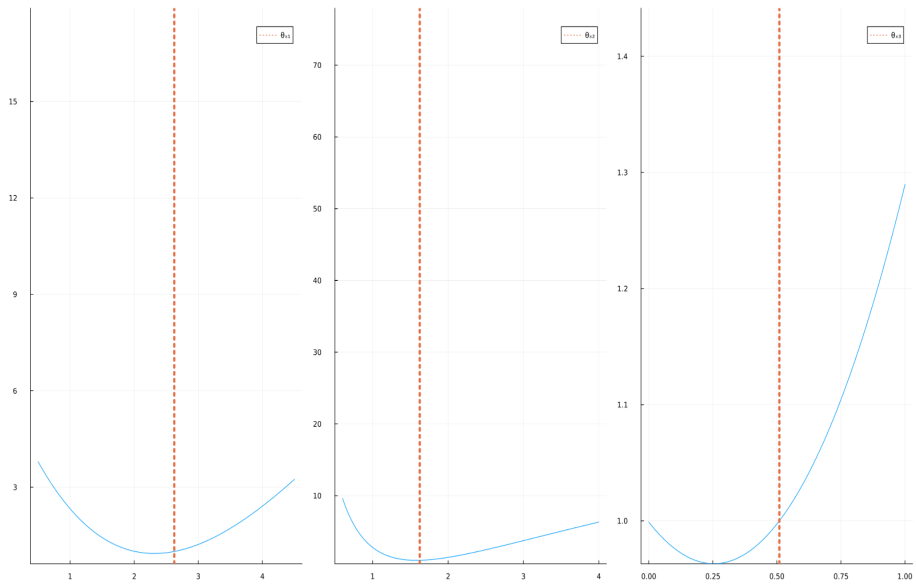
This figure plots the value of the objective function of the Asymptotic BBL estimator varying one parameter at a time while holding the other two parameters fixed at their true values. Values are scaled by the value of the objective at the true parameters.

Figure 18: Multiplicative BBL Estimator Objective Slices



This figure plots the value of the objective function of the Multiplicative BBL estimator varying one parameter at a time while holding the other two parameters fixed at their true values. Values are scaled by the value of the objective at the true parameters.

Figure 19: NLLS Estimator Objective Slices



This figure plots the value of the objective function of the estimator defined in Equation (27) varying one parameter at a time while holding the other two parameters fixed at their true values. Values are scaled by the value of the objective at the true parameters.