PARTIAL IDENTIFICATION AND INFERENCE FOR DYNAMIC MODELS AND COUNTERFACTUALS

By

Myrto Kalouptsidi, Yuichi Kitamura, Lucas Lima, and Eduardo Souza-Rodrigues

February 2020

COWLES FOUNDATION DISCUSSION PAPER NO. 2221

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Partial Identification and Inference for Dynamic Models and Counterfactuals*

Myrto Kalouptsidi, Yuichi Kitamura, Lucas Lima, and Eduardo Souza-Rodrigues†

February, 2020

Abstract

We provide a general framework for investigating partial identification of structural dynamic discrete choice models and their counterfactuals, along with uniformly valid inference procedures. In doing so, we derive sharp bounds for the model parameters, counterfactual behavior, and low-dimensional outcomes of interest, such as the average welfare effects of hypothetical policy interventions. We characterize the properties of the sets analytically and show that when the target outcome of interest is a scalar, its identified set is an interval whose endpoints can be calculated by solving well-behaved constrained optimization problems via standard algorithms. We obtain a uniformly valid inference procedure by an appropriate application of subsampling. To illustrate the performance and computational feasibility of the method, we consider both a Monte Carlo study of firm entry/exit, and an empirical model of export decisions applied to plant-level data from Colombian manufacturing industries. In these applications, we demonstrate how the identified sets shrink as we incorporate alternative model restrictions, providing intuition regarding the source and strength of identification.

KEYWORDS: Dynamic Discrete Choice, Counterfactual, Partial Identification, Subsampling, Uniform Inference, Structural Model.

*We would like to thank Victor Aguirregabiria, Phil Haile, Bo Honoré, Arthur Lewbel, Rob McMillan, Ismael Mourifié, Salvador Navarro, Ariel Pakes, John Rust, Elie Tamer, Xun Tang, Petra Todd, and Yuanyuan Wan for helpful discussions, and workshop participants at Cornell University, Ohio University, Saint Louis FED, UCLA, University of Toronto, the ASSA Meetings 2019, the 3rd Conference on Structural Dynamic Models, the Counterfactuals with Economic Restrictions Conference at Western Ontario, the 5th Banff Empirical Microeconomics Workshop, the 2019 Interactions Conference: Bringing Together Econometrics and Applied Microeconomics at the University of Chicago, and the First IO Workshop at Javeriana University for additional comments. Financial support from the University of Toronto Mississauga is gratefully acknowledged. All remaining errors are our own.

†Affiliations: Myrto Kalouptsidi, Harvard University, CEPR and NBER; Yuichi Kitamura, Yale University and Cowles Foundation for Research in Economics; Lucas Lima, Harvard University; Eduardo Souza-Rodrigues, University of Toronto.

1 Introduction

Structural models have been used to answer a wide range of counterfactual questions in various fields of economics, including industrial organization, labor, public finance, and trade. For problems involving dynamic tradeoffs, the class of structural dynamic discrete choice (DDC) models has arguably been the most commonly used in applied work; see Aguirregabiria and Mira (2010), Keane, Todd, and Wolpin (2011), and Low and Meghir (2017) for surveys of the literature. In such models, forward-looking agents choose among discrete actions in order to maximize their expected discounted stream of payoffs given a finite state space. In the canonical setting proposed by Rust (1987), the flow payoffs are allowed to depend freely on the current action and state, and are additively separable to an unobservable i.i.d. shock whose distribution is (typically) assumed known by the researcher. This class of models can be estimated using data on individual choices and state variables (Rust, 1987; Hotz and Miller, 1993; Aguirregabiria and Mira, 2002).

Despite being widely used by practitioners, Rust (1994) and Magnac and Thesmar (2002) have shown that, in this class of models, a continuum of payoff functions can rationalize observed choice behavior. That is a fundamental identification problem as different flow payoffs that are equally compatible with the data can generate different behavioral responses in a counterfactual environment. While applied researchers have typically addressed this problem by imposing restrictions that select among observationally equivalent models, economic theory does not always offer guidance as to the correct assumptions necessary to identify the true model. Given that such difficulties can threaten the credibility of structural estimation, a recent literature has started to investigate this problem, and has shown that only a narrow class of counterfactual experiments results in counterfactual behavior that is point identified (i.e., invariant to identifying restrictions imposed on the model); see Aguirregabiria and Suzuki (2014), Norets and Tang (2014), and Kalouptsidi, Scott, and Souza-Rodrigues (2019). Still, these important recent results leave open a critical question: How much can be learned about counterfactual outcomes of interest under minimal economic assumptions for a large (and empirically relevant) class of counterfactual experiments?

The main contribution of our paper is the development of a new framework to address this question. The framework is tractable and involves minimal economic assumptions that may not suffice to point-identify the model parameters, giving rise to partial identification analysis. We show how to characterize and compute sharp bounds – that is, bounds that exhaust all implications of the model and data – for counterfactual outcomes of interest, along with a uniformly valid inference procedure. We focus on bounds for low-dimensional counterfactual objects that are relevant to researchers' main conclusions and are amenable to economic interpretation, such as the change in average welfare. These objects depend on agents' counterfactual behavior; i.e., the conditional choice probability (CCP) function under the counterfactual environment. Our general procedure is valid for broad classes of counterfactual experiments, model restrictions, and outcomes of interest, as we explain below.

To fix ideas, consider the firm entry/exit problem – our running example. In this application, a firm facing uncertainty about demand shocks and input prices decides in every period whether to enter (exit) a market subject to entry costs (scrap values), with the goal of maximizing its expected discounted stream of payoffs consisting of variable profits minus fixed costs. Typically, to estimate this model, researchers assume the payoff of staying out of the market (the 'outside option') is zero, and also impose that scrap values and/or fixed costs do not depend on state variables and are equal to zero. These assumptions are often referred to as "normalization" assumptions, and they suffice to select among observationally equivalent parameter values. Assuming scrap values or fixed costs are invariant over states however may be a strong restriction for some industries; perhaps more importantly, setting them to have the exact same value as the payoff of the outside option is difficult to justify: economic theory does not provide guidance as to how to set these values, and cost or scrap value data are extremely rare. Further, these assumptions are not always innocuous for important counterfactual questions, as shown in previous research (Aguirregabiria and Suzuki, 2014; Norets and Tang, 2014; Kalouptsidi, Scott, and Souza-Rodrigues, 2019). Given such limitations, we avoid these assumptions and focus directly on the identified set of counterfactual objects (e.g., the welfare impact of a hypothetical entry subsidy) under much milder restrictions (such as that entry costs/scrap values are positive, or that entry is eventually profitable). In the application, we show how the identified sets shrink as we add alternative model restrictions, providing intuition regarding the source and strength of identification.

We start the analysis of dynamic discrete choice models more generally by showing that the sharp identified set for the payoff vector is a convex polyhedron whose dimension depends on the size of the state space and the number of model restrictions that the researcher is willing to impose. Then, we show that for a broad class of counterfactuals involving almost any change in the primitives, the sharp identified set for the counterfactual CCP is a connected manifold with a dimension that can be determined from the data.[1] The set is therefore either empty (which occurs when the model is rejected by the data), or a singleton (implying point-identification), or a continuum. The dimension of the set can be calculated by checking the rank of a specific matrix, which depends on the data, the model restrictions, and the counterfactual transformation, all of which are known by the econometrician. This dimension is typically much smaller than the dimension of the conditional probability simplex, which implies that the identified set is informative. Specific combinations of model restrictions and counterfactual experiments can reduce the dimension of the identified set further, leading to point identification in some cases. To the best of our knowledge, while partial identification and estimation of model parameters in DDC models have been considered previously (see, e.g., Bajari, Benkard, and Levin, 2007; Norets and Tang, 2014; Berry and Compiani, 2019), these are the first analytical results characterizing the identified set of counterfactual behavior.

Given the identified sets for the high-dimensional payoff vector and counterfactual CCPs, we then turn

[1] We consider any change in the primitives except for nonlinear transformations in payoffs (uncommon in practice).

to the low-dimensional outcomes of interest. Here, we show that the sharp identified set is also connected and, under additional mild conditions, compact. This is convenient as in practice it is sufficient to trace the boundary of the set. In addition, when the outcome of interest is a scalar, the identified set becomes a compact interval, in which case it suffices to calculate the lower and upper endpoints. The endpoints can be computed by solving well-behaved constrained minimization and maximization problems. The optimizations can be implemented using standard software (e.g., Knitro), and remain feasible even in high-dimensional cases involving large state spaces or a large number of model parameters. In our experience, standard solvers perform best when the researcher provides the gradient of the objective function; for cases in which computing the gradient is costly, we develop and propose an alternative (stochastic) search procedure (discussed in detail in the Online Appendix). Overall, an attractive feature of this approach is that the researcher can flexibly adjust (i) the set of model restrictions, (ii) the counterfactual experiment, and (iii) the target outcome of interest, all without having to derive additional analytical identification results for each alternative specification.

Our approach leads naturally to an inference procedure for empirical work. We develop an asymptotically uniformly valid inference approach based on subsampling, and construct confidence sets for the true value of the low-dimensional outcome – rather than for the identified set – based on test inversion. We elaborate on our inferential procedure later in the paper, but note here that many existing approaches developed for moment inequalities and other set identified models are not easily amendable to our set-valued counterfactual analysis; see Remark 4 in Section 5 for details. Taken together, these are the first positive results on set-identification and uniformly valid inference procedures for counterfactual outcomes of interest in structural dynamic models. These are the core contributions of our paper.

Finally, we provide evidence that our inference procedure performs well in finite samples based on a Monte Carlo study of firm entry/exit. We then illustrate the policy usefulness of our approach by revisiting the seminal contribution by Das, Roberts, and Tybout (2007) on exporting decisions and subsidies. Based on their plant-level panel data from Colombian manufacturing industries, we explore the identifying power of different model restrictions, discuss the assumptions under which alternative counterfactual subsidies promote large impacts on export revenues per unit cost of subsidy, and generate the same rank of policies as in Das, Roberts, and Tybout (2007) under weaker conditions.

Related Literature. A large body of work studies the identification and estimation of dynamic discrete choice models. Rust (1994) showed that DDC models are not identified nonparametrically, and Magnac and Thesmar (2002) characterized the degree of underidentification. Important advances that followed include (but are not limited to) Heckman and Navarro (2007), Pesendorfer and Schmidt-Dengler (2008), Blevins (2014), Bajari, Chu, Nekipelov, and Park (2016), and Abbring and Daljord (2019). In terms of estimation, Rust (1987) introduced the nested fixed point maximum likelihood estimator in his seminal contribution, and Hotz and Miller (1993) pioneered a computationally convenient two-step estimator

that was then further analyzed by a host of important studies (Hotz, Miller, Sanders, and Smith, 1994; Aguirregabiria and Mira, 2002, 2007; Bajari, Benkard, and Levin, 2007; Pakes, Ostrovsky, and Berry, 2007; Pesendorfer and Schmidt-Dengler, 2008).[2] We build on these literatures on point-identification and estimation, and extend them to partial identification of model parameters and, more importantly, counterfactuals.

Several important papers have considered partial identification and estimation of structural parameters, namely Bajari, Benkard, and Levin (2007), Norets and Tang (2014), Dickstein and Morales (2018), Morales, Sheu, and Zahler (2019), and Berry and Compiani (2019).[3] With the exception of Norets and Tang (2014) (discussed further below), these papers consider classes of models that differ from, and do not necessarily nest, ours. A common issue in this literature concerns the fact that existing inference methods for partially identified models are computationally costly – if not infeasible – when the parameter space is not small (they require repeated inversion of hypothesis testing over the parameter space); thus, prior empirical work has only estimated the most parsimonious specifications. Substantial computational costs may also limit the set of counterfactuals implemented, given that simulations for each parameter value in the identified set are required. In contrast, our approach focuses inference directly on low-dimensional final objects of interest – which are typically nonlinear functions of CCPs and model primitives – thus allowing for a large number of model parameters and richer empirical applications. As such, our approach complements, and can be combined with, the previous contributions.

A small but growing literature investigates the identification of counterfactuals in DDC models. The main contributions in this area are by Aguirregabiria (2010), Aguirregabiria and Suzuki (2014), Norets and Tang (2014), Arcidiacono and Miller (2018), and Kalouptsidi, Scott, and Souza-Rodrigues (2017, 2019).[4] We rely heavily on Kalouptsidi, Scott, and Souza-Rodrigues (2019) (henceforth 'KSS'), which provides the necessary and sufficient conditions for point identification of a broad class of counterfactuals encountered in applied work. The closest paper to ours is by Norets and Tang (2014), who consider binary choice models, relax the assumption that the distribution of the idiosyncratic shocks is known by the econometrician, and obtain partial identification results for structural parameters and for (high-dimensional) counterfactual choice probabilities. They focus on two types of counterfactuals – pre-specified additive changes in payoffs and changes to state transitions – and propose a Bayesian approach to inference, based on Markov Chain

[2] Important early contributions include Miller (1984), Wolpin (1984), and Pakes (1986).

[3] Bajari, Benkard, and Levin (2007) two-step estimator is the first to allow for partially-identified model parameters. Dickstein and Morales (2018) and Morales, Sheu, and Zahler (2019) pioneered the use of Euler-equation-like estimators for DDC models using moment inequalities, requiring minimal distributional assumptions on the error term. Berry and Compiani (2019) allow for serially correlated unobserved states and propose the use of lagged state variables as instrumental variables for (econometrically) endogenous states, for models with both continuous and discrete actions, and obtain partial identification of structural parameters in a discrete choice setting.

[4] Aguirregabiria and Suzuki (2014), Norets and Tang (2014), and Arcidiacono and Miller (2018) have established the identification of two important categories of counterfactuals in different classes of DDC models: counterfactual behavior is identified when flow payoffs change additively by pre-specified amounts; counterfactual behavior is generally not identified when the state transition process changes. Kalouptsidi, Scott, and Souza-Rodrigues (2017) discuss identification of counterfactual best-reply functions and equilibria in dynamic games.

Monte Carlo (MCMC) methods. Compared to Norets and Tang (2014), we assume that the distribution of the unobservables is known, which is common in practice and allows us to characterize the properties of the identified sets analytically, but we consider a broader class of counterfactual experiments. Moreover, our approach for inference is based on subsampling and is guaranteed to be uniformly valid asymptotically, in contrast to Bayesian inference, which is known to be only pointwise valid asymptotically (see, e.g., the discussion in Canay and Shaikh, 2017). Further, differently from Norets and Tang (2014), we focus on confidence sets for low-dimensional outcomes of interest, which are often what researchers are mostly interested in. Developing inference procedures for such objects is not trivial as they involve nonlinear functions of model parameters and counterfactual choice probabilities.

Our inference approach builds on the formulation developed in Kitamura and Stoye (2018), where the implications of economic models are expressed in terms of the minimum value of a quadratic form. The associated quadratic form-based algorithm offers computational advantages, and it also provides a useful framework for asymptotic analysis, especially when asymptotic uniform validity is an important issue. Our model restriction has non-regular features in terms of smoothness, and is thus connected to a large literature initiated by Chernoff's (1954) study of non-regular statistical models, namely, the asymptotic behavior of the minimum distance of a random object to a fixed manifold with possible kinks. In contrast to this literature, we consider the minimum distance to a kinked (i.e., non-regular) and *random* (estimated) and possibly nonconvex set. We avoid standard convexity conditions, even locally, on such objects because they are typically incompatible with our model restrictions.[5] We establish that an appropriate application of subsampling to the quadratic-form-based distance measure yields an asymptotically valid algorithm for inference.[6]

Finally, a recent and increasingly influential line of research emphasizes that (partial) identification of potential effects of policy interventions does not necessarily require identification of all the model parameters. Major contributions outside the class of structural dynamic models include Ichimura and Taber (2000, 2002) and Mogstad, Santos, and Torgovitsky (2018) for selection models; Manski (2007) for static choice models under counterfactual choice sets; Blundell, Browning, and Crawford (2008), Blundell, Kristensen, and Matzkin (2014), Kitamura and Stoye (2019), and Adams (2020) for bounds on counterfactual demand distributions and welfare analysis; Adao, Costinot, and Donaldson (2017) for international trade models; and Bejara (2018) for macroeconomic models. All these approaches, including ours, are consistent with Marschak's (1953) advocacy of solving well-posed economic problems with minimal assumptions. See Heckman (2000, 2010) for excellent discussions of Marschak's approach to identification in structural models.

[5]Note that Kitamura and Stoye (2018) deal with the case where a random vector is projected on a non-smooth but fixed object with some desirable geometric features. They then show that a bootstrap procedure combined with what they call the tightening technique leads to a computationally efficient algorithm with asymptotic uniform validity.

[6]Asymptotic validity of subsampling in nonregular models with more conventional settings, such as standard moment inequality models, have been shown in the literature: see Romano and Shaikh (2008) and Romano and Shaikh (2012).

The rest of the paper is organized as follows: Section 2 sets out the dynamic discrete choice framework; Section 3 presents the partial identification results for the model parameters, then illustrates the identified set under alternative restrictions in the context of a firm entry/exit problem; Section 4 contains our main results regarding the set-identification of counterfactuals; Section 5 establishes uniformly valid confidence sets for target parameters; Section 6 presents the empirical application involving export supply and subsidies; and Section 7 concludes.[7]

2 Dynamic Model

In the model, time is discrete and the time horizon is infinite. Every period t, agent i observes the state s_{it} and chooses an action a_{it} from the finite set $\mathcal{A} = \{0, ..., A\}$ to maximize the expected discounted payoff,

$$\mathbb{E}\left(\sum_{\tau=0}^{\infty} \beta^{\tau} \overline{\pi}\left(a_{it+\tau}, s_{it+\tau}\right) \mid a_{it}, s_{it}\right),$$

where $\overline{\pi}(\cdot)$ is the per period payoff function, and $\beta \in [0, 1)$ is the discount factor. The agent's state s_{it} follows a controlled Markov process. We follow the literature and assume that s_{it} is split into two components, $s_{it} = (x_{it}, \varepsilon_{it})$, where x_{it} is observed by the econometrician and ε_{it} is not. We assume $x_{it} \in \mathcal{X} = \{1, ..., X\}$, $X < \infty$; while $\varepsilon_{it} = (\varepsilon_{0it}, ..., \varepsilon_{Ait})$ is i.i.d. across agents and time, and has joint distribution G that is absolutely continuous with respect to the Lebesgue measure and has full support on $\mathbb{R}^{A+1}$.[8]

The transition distribution function for $(x_{it}, \varepsilon_{it})$ factors as

$$F\left(x_{it+1}, \varepsilon_{it+1} | a_{it}, x_{it}, \varepsilon_{it}\right) = F\left(x_{it+1} | a_{it}, x_{it}\right) G\left(\varepsilon_{it+1}\right),$$

and the payoff function is additively separable in the unobservables,

$$\overline{\pi}\left(a, x_{it}, \varepsilon_{it}\right) = \pi_a\left(x_{it}\right) + \varepsilon_{ait},$$

where $\pi_a(x)$ is a bounded function. Let $V\left(x_{it}, \varepsilon_{it}\right)$ be the expected discounted stream of payoffs under

[7] The Supplemental Material and the Online Appendix complement the main paper. The Supplemental Material contains (a) all proofs of the propositions and theorems presented in the main text; (b) detailed information about our running example (the firm entry/exit problem); (c) our Monte Carlo study; and (d) our replication of Das, Roberts, and Tybout (2007). The Online Appendix presents (e) several useful examples of commonly employed restrictions in applied work (using our notation); (f) the computational algorithm for inference based on subsampling; (g) how to calculate the gradient of the object of interest when it involves long-run average effects; and (h) our proposed stochastic search approach to calculate the lower and upper bounds of the identified set of relevant outcomes, without analytic gradients. The Online Appendix is available on the authors' webpages.

[8] Our results cover static discrete choice models, and can be extended to dynamic models with continuous states, nonstationarity, and that are finite-horizon. Such extensions are however beyond the scope of the paper.

optimal behavior by the agent. By Bellman's principle of optimality,

$$V(x_{it}, \varepsilon_{it}) = \max_{a \in \mathcal{A}} \left\{ \pi_a(x_{it}) + \varepsilon_{ait} + \beta \, \mathbb{E}\left[V(x_{it+1}, \varepsilon_{it+1}) \, | a, x_{it} \right] \right\}.$$

Following the literature, we define the *ex ante value function*, $V(x_{it}) \equiv \int V(x_{it}, \varepsilon_{it}) \, dG(\varepsilon_{it})$, and the *conditional value function*:

$$v_a(x_{it}) \equiv \pi_a(x_{it}) + \beta \, \mathbb{E}\left[V(x_{it+1}) \, | a, x_{it} \right].$$

The ex ante value function takes the expectation of the value function with respect to ε_{it}. The conditional value function is the sum of the agent's current payoff, net of the idiosyncratic shocks ε_{it}, and the expected lifetime payoff given a choice of action a this period and optimal choices from next period onwards.

The *conditional choice probability* (CCP) function is given by

$$p_a(x_{it}) = \int 1\left\{ v_a(x_{it}) + \varepsilon_{ait} \geq v_j(x_{it}) + \varepsilon_{jit}, \text{ for all } j \in \mathcal{A} \right\} dG(\varepsilon_{it}),$$

where $1\{\cdot\}$ is the indicator function. We define the $(A+1) \times 1$ vector of conditional choice probabilities $p(x) = (p_0(x), ..., p_A(x))'$, and the corresponding $(A+1)X \times 1$ vector $p = (p'(1), ..., p'(X))'$, where $'$ denotes transpose.

It is useful to note that for any (a, x) there exists a real-valued function $\psi_a(.)$ derived only from G such that

$$V(x) = v_a(x) + \psi_a\left(p(x)\right). \tag{1}$$

Equation (1) states that the ex ante value function V equals the value obtained by choosing a today and optimally thereafter, v_a, plus a correction term, ψ_a, because choosing action a today is not necessarily optimal. When ε_{it} follows the type I extreme value distribution, $\psi_a(p(x)) = \kappa - \ln p_a(x)$, where κ is the Euler constant.[9]

As we make extensive use of matrix notation below, we define the vectors $\pi_a, v_a, V, \psi_a \in \mathbb{R}^X$, which stack $\pi_a(x)$, $v_a(x)$, $V(x)$, and $\psi_a(p(x))$, for all $x \in \mathcal{X}$. We often use the notation $\psi_a(p)$ to emphasize the dependence of ψ_a on the choice probabilities p. We also define F_a as the transition matrix with (m, n) element equal to $\Pr\left(x_{it+1} = x_n | x_{it} = x_m, a\right)$. The payoff vector $\pi \in \mathbb{R}^{(A+1)X}$ stacks π_a for all $a \in \mathcal{A}$, and, similarly, F stacks (a vectorized version) of F_a for all $a \in \mathcal{A}$.

[9] Equation (1) is shown in Arcidiacono and Miller (2011, Lemma 1). It makes use of the Hotz-Miller inversion (Hotz and Miller, 1993), which, in turn, establishes that the difference of conditional value functions is a known function of the CCPs: $v_a(x) - v_j(x) = \phi_{aj}\left(p(x)\right)$, where $\phi_{aj}(.)$ is again derived only from G. When ε_{it} follows the type I extreme value distribution, $\phi_{aj}(p(x)) = \log p_a(x) - \log p_j(x)$. Chiong, Galichon, and Shum (2016) propose a novel approach that can calculate ψ_a and ϕ_{aj} for a broad set of distributions G (see also Dearing, 2019).

3 Model Restrictions and Identification

In this section, we characterize the identified set of the model parameters, allowing for additional model restrictions that the researcher may be willing to impose, and we illustrate the sets in the context of a firm entry/exit problem.

The primitives of the model are $(\mathcal{A}, \mathcal{X}, \beta, G, F, \pi)$, which generate the endogenous objects $\{p_a, v_a, V : a \in \mathcal{A}\}$. Typically, the researcher has access to a panel data on agents' actions and states, $\{a_{it}, x_{it} : i = 1, ..., N;\ t = 1, ..., T\}$. Under some standard regularity conditions, the researcher can identify and estimate the agents' choice probabilities $p_a(x)$, as well as the transition distribution function F, directly from the data. We therefore take p and F as known for the identification arguments. We also follow the literature and assume the econometrician knows the discount factor β and the distribution of the idiosyncratic shocks G (we discuss these assumptions in Remark 3 below). The main objective here is to identify the payoff function π.

The model is identified if there is a unique payoff that can be inferred from the observed choice probabilities and state transitions. Intuitively, π has $(A+1)X$ parameters, and there are only AX observed CCPs; thus there are X free payoff parameters and X restrictions will need to be imposed to point-identify π (Rust, 1994; Magnac and Thesmar, 2002). We follow KSS to represent the underidentification problem as follows: for all $a \neq J$, where $J \in \mathcal{A}$ is some reference action, π_a can be represented as an affine transformation of π_J:[10]

$$\pi_a = M_a \pi_J + b_a(p), \tag{2}$$

where

$$M_a = (I - \beta F_a)(I - \beta F_J)^{-1}, \tag{3}$$

$$b_a(p) = M_a \psi_J(p) - \psi_a(p), \tag{4}$$

and I is a (comformable) identity matrix. In the logit model, $b_a(p) = \ln p_a - M_a \ln p_J$, where $\ln p_a$ is the $X \times 1$ vector with elements $\ln p_a(x)$. To simplify notation, we omit the dependence of both M_a and $b_a(p)$ on the transition probabilities F.

We rely heavily on equation (2). Given the data at hand, one can compute both the $X \times X$ matrix M_a and the $X \times 1$ vector b_a directly for each action $a \neq J$. The payoffs π_a, $a \neq J$, are not identified because the free parameter π_J is unknown. Equation (2) therefore explicitly lays out how we might estimate

[10] To see why, fix the vector $\pi_J \in \mathbb{R}^X$. Then,

$$\pi_a = v_a - \beta F_a V = V - \psi_a - \beta F_a V = (I - \beta F_a) V - \psi_a,$$

where for $a = J$, we have $V = (I - \beta F_J)^{-1}(\pi_J + \psi_J)$. After substituting for V, we obtain the result. As an aside, note that $(I - \beta F_J)$ is invertible because F_J is a stochastic matrix and hence the largest eigenvalue is equal or smaller than one. The eigenvalues of $(I - \beta F_J)$ are given by $1 - \beta\gamma$, where γ are the eigenvalues of F_J. Because $\beta < 1$ and $\gamma \leq 1$, we have $1 - \beta\gamma > 0$.

the payoff function if we are willing to fix the payoffs of one action at all states *a priori* (e.g. $\pi_J = 0$). However, this is not the only way to obtain identification: we simply need to add X extra restrictions. Other common possibilities involve reducing the number of payoff function parameters to be estimated using parametric assumptions and/or exclusions restrictions.

It will be useful to represent (2) for all actions $a \neq J$ at once using two different compact notations. First,

$$\pi_{-J} = M_{-J}\pi_J + b_{-J}, \tag{5}$$

where π_{-J} stacks π_a for all $a \neq J$, and the matrix M_{-J} and vector b_{-J} are defined similarly.[11] Further, define $\mathbf{M} = [I, -M_{-J}]$, and arrange π in the following way: $\pi = [\pi'_{-J}, \pi'_J]'$. Then (5) becomes

$$\mathbf{M}\pi = b_{-J}. \tag{6}$$

Note that identification of π fails because $\mathbf{M}$ is rank-deficient; indeed, $\mathbf{M}$ is an $AX \times (A+1)X$ matrix, and so $rank(\mathbf{M}) = AX < (A+1)X$. In short, equation (6) summarizes all assumptions imposed by the basic dynamic framework: any $\pi \in \mathbb{R}^{(A+1)X}$ satisfying (6) is compatible with the data.[12]

Remark 1. (Static Models) Equation (2) also holds in static models. When agents are myopic ($\beta = 0$) or when choices do not affect the transition of states ($F_a = F_J$, for all $a \neq J$), the matrix M_a becomes the identity matrix, implying that the difference in payoffs $\pi_a - \pi_J$ is identified: it equals $\psi_J(p) - \psi_a(p)$; in the logit model, that differencce equals the log odds ratio of the choice probabilities. All results we present in this paper naturally cover the class of static discrete choice models.

3.1 Model Restrictions

We consider two types of model restrictions. The first is a set of $d \leq X$ linearly independent equalities,

$$R^{eq}\pi = r^{eq}, \tag{7}$$

with $R^{eq} \in \mathbb{R}^{d\times(A+1)X}$, or in block-form, $R^{eq} = [R^{eq}_{-J}, R^{eq}_J]$, where R^{eq}_{-J} defines how π_{-J} enters into the constraints and, similarly, R^{eq}_J for π_J. This formulation is general enough to incorporate several assumptions used in practice. Examples include exclusion restrictions (setting some elements of π equal to each other), prespecifying some π_J (set $R^{eq}_J = I$, $R^{eq}_{-J} = 0$ and r^{eq} accordingly), and parametric assumptions such as $\pi_a(x) = z_a(x)\gamma$, where z_a is some known function of actions and states, and $\gamma \in \Gamma \subset \mathbb{R}^L$ is a parameter vector in the parameter space Γ, with dimension L usually much smaller than

[11] The vectors π_{-J} and b_{-J} have dimension $AX \times 1$, while M_{-J} is an $AX \times X$ matrix.

[12] This model imposes a scale normalization. In general, the payoff function is given by $\overline{\pi}(a, x_{it}, \varepsilon_{it}) = \pi_a(x_{it}) + \sigma\varepsilon_{ait}$, where $\sigma > 0$ is a scale parameter. This means equation (6) is given by $\mathbf{M}(\pi/\sigma) = b_{-J}$. As usual in discrete choice models, when we set $\sigma = 1$ (as we do here), the scale of the payoff is measured relative to the standard deviation of one of the components of ε_{it}.

$(A+1)\,X$.[13]

The second set of restrictions are m linear inequalities:

$$R^{iq}\pi \leq r^{iq}, \tag{8}$$

with $R^{iq} \in \mathbb{R}^{m\times(A+1)X}$, or in block-form, $R^{iq} = [R^{iq}_{-J}, R^{iq}_{J}]$. The inequalities (8) can incorporate shape restrictions, such as monotonicity, concavity, and supermodularity. In Online Appendix E, we explicitly lay out how several examples of assumptions used in applied work can be expressed as (7) or (8).

We assume the restrictions (7) and (8) are not redundant. Equations (6), (7), and (8) summarize therefore all model restrictions.

3.2 Model Identification

The identified set for the payoff function is characterized by all payoffs satisfying all model restrictions. Our first proposition follows (all proofs are in the Supplemental Material).

Proposition 1. *The sharp identified set for the payoff function π is*

$$\Pi^I = \left\{ \pi \in \mathbb{R}^{(A+1)X} : \mathbf{M}\pi = b_{-J},\, R^{eq}\pi = r^{eq},\, R^{iq}\pi \leq r^{iq} \right\}. \tag{9}$$

The identified set Π^I is a convex polyhedron of dimension $X - d$, where $0 \leq d \leq X$.

The identified set is sharp by construction.[14] It is a convex polyhedron because it is the intersection of finitely many closed halfspaces. Note that Π^I can be characterized in practice by linear programming or convex programming methods. In the absence of inequalities (8), the identified set becomes a linear manifold with dimension $X-d$; and it collapses to a singleton (i.e., π is point-identified) when the matrix $\left[\mathbf{M}', R^{eq\prime}\right]'$ is full rank (Magnac and Thesmar, 2002; Pesendorfer and Schmidt-Dengler, 2008).[15]

Before proceeding, remarks regarding unobserved heterogeneity, and the assumption of known β and G are in order.

Remark 2. (Unobserved Heterogeneity.) In the presence of unobserved heterogeneity, equations (6)–(8) hold for each unobserved type. This implies that, after type-specific choice probabilities and transition functions of finitely many unobserved types are identified (e.g., following the strategies proposed by

[13] To see this, note that $\pi = z\gamma$, where z is a known matrix of dimension $(A+1)\,X \times L$, γ is a column vector $L \times 1$, and we assume $(A+1)\,X > L$. Decompose the long $(A+1)\,X$ vector π into an upper part π_u and a lower part π_l, and define z_u and z_l similarly. Then, $\pi_u = z_u\gamma$ and $\pi_l = z_l\gamma$. Suppose the decomposition is such that z_u has full column rank. Then, from the first equality we obtain: $\gamma = \left(z_u'z_u\right)^{-1} z_u'\pi_u$. Substitution in the second equality gives $\pi_l = z_l\left(z_u'z_u\right)^{-1} z_u'\pi_u$. Therefore, $\left[z_l\left(z_u'z_u\right)^{-1} z_u' \;,\; -I\right]\pi = 0$.

[14] A sharp identified set is the smallest set of parameter values that can generate the data.

[15] If we impose a linear-in-parameters restriction on flow payoffs, $\pi = z\gamma$, we can write the identified set for γ in a similar way: $\Gamma^I = \{\, \gamma \in \Gamma : (\mathbf{M}z)\,\gamma = b_{-J},\, (R^{eq}z)\gamma = r^{eq},\, (R^{iq}z)\gamma \leq r^{iq} \,\}$.

Kasahara and Shimotsu (2009) or Hu and Shum (2012)), identified sets given by (9) hold, and can be calculated, for each type.

Remark 3. (Unknown β and G.) Although we assume a known discount factor, it is straightforward to extend our analysis by either making use of the contributions by Magnac and Thesmar (2002) and Abbring and Daljord (2019) to identify β, or by indexing Π^I by β and taking the identified set for π as the union of the sets $\Pi^I(\beta)$'s for all admissible discount factors. Similarly, Blevins (2014), Chen (2017), and Buchholz, Shum, and Xu (2019) consider identification of G under different model assumptions. One can combine their assumptions to identify G and Π^I simultaneously, or take the union of the sets $\Pi^I(G)$ for admissible distributions G as the identified set Π^I; see also Christensen and Connault (2019). Note however that the union of such sets (either $\Pi^I(\beta)$ or $\Pi^I(G)$) is not necessarily a convex polyhedron.

3.3 Example: Firm Entry/Exit Model

Next, we illustrate the payoff identified set in the context of a simple firm entry/exit problem. Suppose a firm i decides whether to enter the market ($a = 1$) or stay out ($a = 0$), so that $\mathcal{A} = \{0, 1\}$. Decompose the state space into $x_{it} = (k_{it}, w_{it})$, where $k_{it} \in \mathcal{K} = \{0, 1\}$ is the lagged decision a_{it-1}, and $w_{it} \in \mathcal{W} = \{1, ..., W\}$ are exogenous shocks determining profits. Assume for convenience (unless otherwise stated) that w_{it} can take two values, low and high: $\mathcal{W} = \{w^l, w^h\}$, with $w^l < w^h$. The size of the state space is therefore $X = KW = 4$, where K and W are the number of values that k and w can take. Transition probabilities decompose as $F(k_{it+1}, w_{it+1}|a_{it}, k_{it}, w_{it}) = F(k_{it+1}|a_{it}, k_{it})F(w_{it+1}|w_{it})$.

Let $\pi_a(k)$ denote the $W \times 1$ vector of payoffs the firm obtains when it chooses action a given k and w, so that $\pi_a = [\pi'_a(0), \pi'_a(1)]'$. We impose the following structure on π:

$$\pi_0 = \begin{bmatrix} \overline{\pi}_0 \\ s \end{bmatrix}, \pi_1 = \begin{bmatrix} vp - fc - ec \\ vp - fc \end{bmatrix}. \tag{10}$$

The payoff the firm obtains when it was out of the market in the previous period and stays out in the current period is the vector $\pi_0(0) = \overline{\pi}_0$ (the value of the outside option); and the payoff when the firm was active and decides to exit is given by the vector of scrap values, $\pi_0(1) = s$. Note that both the outside option and the scrap values can vary with the exogenous state w. The vectors vp, fc, and ec are the variable profits, the fixed costs, and the entry costs, respectively (all of which can vary with w as well). The vector $\pi_1(0) = vp - fc - ec$ measures the profits the firm gets when it enters the market, and $\pi_1(1) = vp - fc$ are the profits when it stays.

In this example, both π_0 and π_1 are 4×1 vectors (and so π has $2X = 8$ elements). To point-identify π we need $X = 4$ restrictions. Typically, researchers identify an entry model by setting $\overline{\pi}_0 = 0$ and further setting either $s = 0$ or assuming $vp - fc$ is known (e.g., by assuming variable profits vp can be recovered "offline," using price and quantity data, and setting $fc = 0$). When $\overline{\pi}_0 = s = 0$, then $\pi_0 = 0$, and point

identification of π follows directly from (2); it is essentially a restriction on a reference action. When instead $\overline{\pi}_0 = 0$ and $vp - fc$ is known, we identify the remaining elements of π by combining (2) and (7).

Assuming the outside option equals the scrap value or the fixed costs (and all are equal to zero) may be difficult to justify in practice, as cost or scrap value data are extremely rare (Kalouptsidi, 2014). When the researcher is not willing to impose such restrictions, π is not point-identified. Yet, the payoff function can be set-identified under weaker conditions. Consider, for instance, the following set of assumptions:

1. $\overline{\pi}_0 = 0, \quad fc \geq 0, \quad ec \geq 0$, and vp is known.

2. $\pi_1(1, w^h) \geq \pi_1(1, w^l)$, and $vp - fc \leq ec \leq \frac{\mathbb{E}[vp-fc]}{1-\beta}$, where the expectation is taken over the ergodic distribution of the state variables.

3. s does not depend on w.

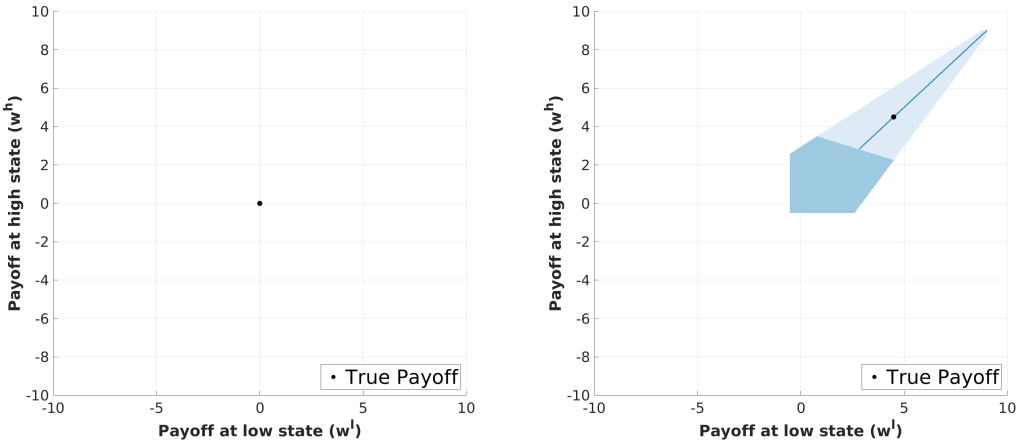
Restriction 1 assumes that the outside option is zero (as usual); fixed costs and entry costs are both positive; and variable profits are known (estimated "offline"). This set of restrictions imposes $d = W = 2$ equality and $m = 4$ inequality constraints. From Proposition 1, the identified set Π^I is a two-dimensional set ($X - d = 2$) in the eight-dimensional space.

Restriction 2 imposes $m = 5$ inequality constraints: profits are increasing in w when the firm is in the market (a monotonicity assumption); entry costs are greater than variable profits minus fixed costs (implying that entry is always costly in the first period of entry); and ec is smaller than the expected present value of future profits when the firm stays forever in the market (meaning that it eventually pays off to enter).

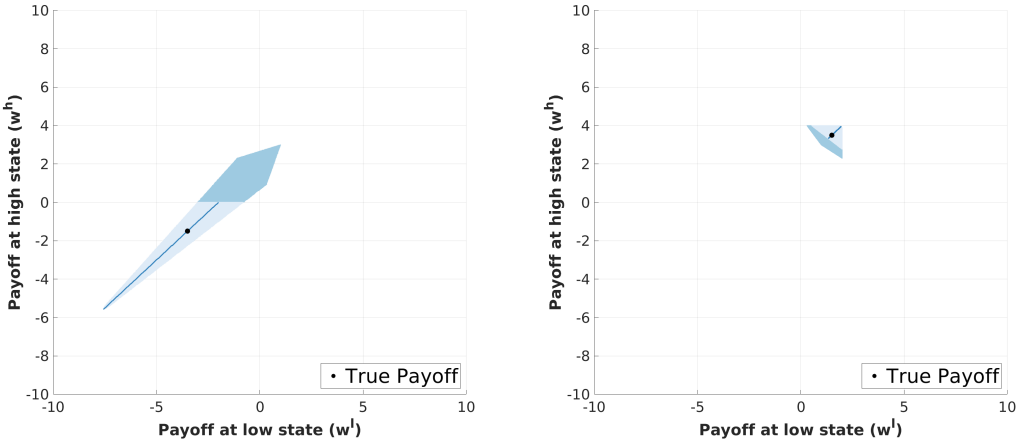
Restriction 3 assumes an exclusion restriction: scrap values are state-invariant. This corresponds to $d = W - 1 = 1$ equality restriction. Note that, by combining Restrictions 1 and 3, we obtain $d = 3$ linear equalities, which makes the identified set Π^I one dimensional. In the Supplemental Material (Section B), we provide explicit characterizations for this example.

Figure 1 presents Π^I for a particular parameter configuration.[16] The larger polyhedron corresponds to Π^I under Restriction 1. The identified set is informative despite the fact that the assumptions imposed are not overly restrictive. To gain intuition regarding the shape of Π^I, consider the set corresponding to scrap values (panel (b)). In this model, equation (6) alone implies that the difference between scrap values and entry costs is point-identified (see Supplemental Material, Section B). As a consequence, the inequality $ec \geq 0$ implies a lower bound on scrap values (for each state w), eliminating from Π^I all values for s below the thresholds indicated in the figure. Similarly, equation (6) implies that the difference between s and the present value of fixed costs is point-identified. In this second case, though, the inequality $fc \geq 0$

[16] We assume scrap values, entry and fixed costs do not depend on w and take the following values: $s = 4.5$, $ec = 5$, and $fc = 0.5$. We also impose $vp(w^l) = 2$ and $vp(w^h) = 4$, so that $\pi_0 = (0, 0, 4.5, 4.5)'$ and $\pi_1 = (-3.5, -1.5, 1.5, 3.5)'$. The discount factor is $\beta = 0.9$, the transition process for w is $Pr(w_{t+1} = w^l | w_t = w^l) = Pr(w_{t+1} = w^h | w_t = w^h) = 0.75$, and the idiosyncratic shocks ε_{it} follow a type 1 extreme value distribution (the scale parameter is set at $\sigma = 1$). Under these assumptions, $\frac{\mathbb{E}[vp-fc]}{1-\beta} = 25$.



(a) Payoff when Stay Out, $\pi_0(0)$ **(b)** Payoff when Exit, $\pi_0(1)$



(c) Payoff when Enter, $\pi_1(0)$ **(d)** Payoff when Stay In, $\pi_1(1)$

Figure 1: Firm Entry/Exit Model: Payoff Identified Set Π^I under Restrictions 1, 2, and 3. The larger polyhedron (including the dark blue areas) correspond to Π^I under Restriction 1. The light blue areas correspond to Π^I under Restrictions 1 and 2. The identified set Π^I under Restrictions 1–3 is represented by the blue lines within the light blue polyhedron. The true π is represented by the black dots.

entails an upper bound on scrap values, eliminating from the identified set all values for s above the diagonal lines shown in the figure. The diagonal lines reflect the fact that equation (6) relates s and the *present value* of fc, so that $fc \geq 0$ leads to restrictions on scrap values *across* states.

Restrictions 1 and 2 together lead to substantial identifying power: Π^I now corresponds to the smaller polyhedron (in light blue), which is substantially smaller in size than the larger polyhedron. Assuming that entry is costly in the first period of entry, $vp - fc \leq ec$, is the main restriction responsible for the reduction in the identified set. This assumption results in another lower bound on s (see panel (b)), but differently from $ec \geq 0$, it involves restrictions on fc and so imposes restrictions on s across states; the other assumptions in Restriction 2 are not as informative in this example; see the Supplemental Material. Interestingly, the payoff function with scrap values that are equal to zero does not belong to Π^I under these two sets of restrictions. As mentioned previously, setting scrap values to zero is a common way to

point-identify π, but, given that $s = 0$ is at odds with Restrictions 1 and 2, such assumption would be rejected by the data.

Finally, Restriction 3 (exclusion restriction on scrap values) also has substantial identifying power as it reduces the dimension of the identified set to one. In the figures, the identified set under Restrictions 1–3 is represented by the blue lines within the light blue polyhedron.

4 Counterfactuals and Outcomes of Interest

The applied literature has implemented several types of counterfactuals that may change one or several of the model's primitives $(\mathcal{A}, \mathcal{X}, \beta, F, G, \pi)$. For instance, a counterfactual may change the action and state spaces (e.g. Gilleskie (1998) restricts access to medical care; Crawford and Shum (2005) do not allow patients to switch medications; Keane and Wolpin (2010) eliminate a welfare program; Keane and Merlo (2010) eliminate job options for politicians; Rosenzweig and Wolpin (1993) add an insurance option for farmers). Some counterfactuals may transform the state transitions (e.g. Collard-Wexler (2013) explores the impact of demand volatility in the ready-mix concrete industry; Hendel and Nevo (2006) study consumers' long-run responsiveness to prices using supermarket scanner data; Kalouptsidi (2014) explores the impact of time to build on industry fluctuations). Other counterfactuals change payoffs through subsidies or taxes (e.g. Keane and Wolpin (1997) consider hypothetical college tuition subsidies; Schiraldi (2011) and Wei and Li (2014), automobile scrap subsidies; Duflo, Hanna, and Ryan (2012), bonus incentives for teachers; Das, Roberts, and Tybout (2007), export subsidies; Lin (2015) and Varela (2018), entry subsidies). Changes in payoffs may also involve a change in the agent's "type" (e.g. Keane and Wolpin (2010) replace the primitives of minorities by those of white women; Eckstein and Lifshitz (2011) substitute the preference/costs parameters of one cohort by those of another; Ryan (2012) replaces firm entry costs post an environmental policy by those before; Dunne, Klimek, Roberts, and Xu (2013) substitute entry costs in some areas by those in others). Finally, a counterfactual may also change the discount factor (e.g., Conlon (2012) studies the evolution of the LCD TV industry when consumers become myopic).

A counterfactual is defined by the tuple $\{\widetilde{\mathcal{A}}, \widetilde{\mathcal{X}}, \widetilde{\beta}, \widetilde{G}, h^s, h\}$. The sets $\widetilde{\mathcal{A}} = \{0, ..., \widetilde{A}\}$ and $\widetilde{\mathcal{X}} = \{1, ..., \widetilde{X}\}$ denote the new set of actions and states respectively. The new discount factor is $\widetilde{\beta}$, and the new distribution of the idiosyncratic shocks is $\widetilde{G}$. The function $h^s : \mathbb{R}^{A \times X^2} \to \mathbb{R}^{\widetilde{A} \times \widetilde{X}^2}$ transforms the transition probability F into $\widetilde{F}$. Finally, the function $h : \mathbb{R}^{AX} \to \mathbb{R}^{\widetilde{A}\widetilde{X}}$ transforms the payoff function π into the counterfactual payoff $\widetilde{\pi}$, so that $\widetilde{\pi} = h(\pi)$. Here, we restrict transformations on payoffs to affine changes $\widetilde{\pi} = \mathcal{H}\pi + g$, where the matrix $\mathcal{H}$ and the vector g are specified by the econometrician. I.e., the payoff $\widetilde{\pi}_a(x)$ at an action-state pair (a, x) is obtained as the sum of a scalar $g_a(x)$ and a linear combination of

all baseline payoffs. It is helpful to write this in a block-matrix equivalent form:

$$\widetilde{\pi} = \begin{bmatrix} \mathcal{H}_{00} & \mathcal{H}_{01} & \cdots & \mathcal{H}_{0A} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{H}_{\widetilde{A}0} & \mathcal{H}_{\widetilde{A}1} & \cdots & \mathcal{H}_{\widetilde{A}A} \end{bmatrix} \begin{bmatrix} \pi_0 \\ \vdots \\ \pi_A \end{bmatrix} + \begin{bmatrix} g_0 \\ \vdots \\ g_{\widetilde{A}} \end{bmatrix}, \tag{11}$$

where the submatrices $\mathcal{H}_{aj}$ have dimension $\widetilde{X} \times X$ for each pair $a \in \widetilde{\mathcal{A}}$ and $j \in \mathcal{A}$. Note that when the counterfactual does not change the set of actions and states (i.e. $\widetilde{\mathcal{A}} = \mathcal{A}$ and $\widetilde{\mathcal{X}} = \mathcal{X}$), $\mathcal{H}$ is a square matrix.

The counterfactual $\{\widetilde{\mathcal{A}}, \widetilde{\mathcal{X}}, \widetilde{\beta}, \widetilde{G}, h^s, h\}$ generates a new set of model primitives $(\widetilde{\mathcal{A}}, \widetilde{\mathcal{X}}, \widetilde{\beta}, \widetilde{G}, \widetilde{F}, \widetilde{\pi})$. The new set of primitives in turn leads to a new optimal behavior, denoted by $\widetilde{p}$ (the counterfactual CCP), and a new lifetime utility, denoted by $\widetilde{V}$ (the counterfactual welfare).

As the state space $\mathcal{X}$ can be large in practice (making both $\widetilde{p}$ and $\widetilde{V}$ high-dimensional vectors), researchers are often interested in low-dimensional objects, such as the average effects of policy interventions. For instance, in the firm entry/exit application, one may be interested in predicting the average effects of an entry subsidy on: (i) how often the firm stays in the market; (ii) prices; (iii) consumer surplus; (iv) the value of the firm; and/or (v) total government expenditures, among others. Denote the low-dimensional counterfactual outcome of interest by $\theta \in \Theta \subset \mathbb{R}^n$, where Θ is the parameter space for θ, and n is much smaller than the size of the state space X (i.e., $n \ll X$). Then, we have

$$\theta = f\left(\widetilde{p}, \pi; p, F\right), \tag{12}$$

where f implicitly incorporates other quantities that may be necessary to calculate θ, such as $\widetilde{\mathcal{A}}$ or $\widetilde{F}$. For instance, take an outcome variable of interest $Y_a(x)$ (e.g., consumer surplus, or the probability of entry), with a corresponding counterfactual given by $\widetilde{Y}_a(x)$. The average treatment effect of the policy intervention on Y is

$$\theta = \mathbb{E}[\widetilde{Y}_a(x)] - \mathbb{E}[Y_a(x)], \tag{13}$$

where $\mathbb{E}[\widetilde{Y}_a(x)]$ integrates over the distribution of actions and states in the counterfactual scenario, while $\mathbb{E}[Y_a(x)]$ integrates over the factual distribution. One may consider the long-run distribution, or may condition on an initial state and estimate short-run effects. (See the Supplemental Material for details.)

4.1 Identification of Counterfactual Behavior

We now investigate the identified set for the counterfactual CCP $\widetilde{p}$. To do so, we leverage the counterfactual counterpart to (2) for any action $a \in \widetilde{\mathcal{A}}$, with $a \neq J$. I.e.,

$$\widetilde{\pi}_a = \widetilde{M}_a \widetilde{\pi}_J + \widetilde{b}_a\left(\widetilde{p}\right), \tag{14}$$

where

$$\begin{aligned}\widetilde{M}_a &= (I - \widetilde{\beta}\widetilde{F}_a)\,(I - \widetilde{\beta}\widetilde{F}_J)^{-1}, \\ \widetilde{b}_a\,(\widetilde{p}) &= \widetilde{M}_a\,\widetilde{\psi}_J\,(\widetilde{p}) - \widetilde{\psi}_a\,(\widetilde{p})\,,\end{aligned}$$

the functions $\widetilde{\psi}_J$ and $\widetilde{\psi}_a$ depend on the new distribution $\widetilde{G}$, and, without loss of generality, the reference action J belongs to both $\mathcal{A}$ and $\widetilde{\mathcal{A}}$. As before, we omit the dependence of both $\widetilde{M}_a$ and $\widetilde{b}_a$ on the transition probabilities $\widetilde{F}$ to simplify notation.

By stacking equation (14) for all actions and rearranging it (as done previously for the baseline case), we obtain $\widetilde{\mathbf{M}}\,\widetilde{\pi} = \widetilde{b}_{-J}$, where $\widetilde{\mathbf{M}} = [I, -\widetilde{M}_{-J}]$, I is an identity matrix, and $\widetilde{M}_{-J}$ and $\widetilde{b}_{-J}$ stack $\widetilde{M}_a$ and $\widetilde{b}_a$, for all $a \neq J$, respectively. Next, using the fact that $\widetilde{\pi} = \mathcal{H}\pi + g$, we get

$$(\widetilde{\mathbf{M}}\mathcal{H})\,\pi = \widetilde{b}_{-J}\,(\widetilde{p}) - \widetilde{\mathbf{M}}g. \tag{15}$$

Equation (15) characterizes counterfactual behavior, relating $\widetilde{p}$ and model parameters directly, with no continuation values involved.[17] Importantly, the function $\widetilde{b}_{-J}$ is continuously differentiable with an everywhere invertible Jacobian (see Lemma 1 in KSS). The next proposition follows:

Proposition 2. *The sharp identified set for the counterfactual CCP $\widetilde{p}$ is*

$$\widetilde{\mathbf{P}}^I = \left\{ \begin{array}{c} \widetilde{p} \in \widetilde{\mathbf{P}} : \exists \pi \in \mathbb{R}^{(A+1)X} \text{ such that} \\ \mathbf{M}\pi = b_{-J}(p), \\ R^{eq}\pi = r^{eq},\ \ R^{iq}\pi \leq r^{iq}, \\ (\widetilde{\mathbf{M}}\mathcal{H})\,\pi = \widetilde{b}_{-J}\,(\widetilde{p}) - \widetilde{\mathbf{M}}g \end{array} \right\}, \tag{16}$$

where $\widetilde{\mathbf{P}}$ is the simplex of conditional choice probabilities.

In words, a vector $\widetilde{p}$ lying in the conditional probability simplex $\widetilde{\mathbf{P}}$ belongs to the identified set $\widetilde{\mathbf{P}}^I$ if there exists a payoff π that is compatible with the data (i.e., $\mathbf{M}\pi = b_{-J}$), satisfies the additional model restrictions (i.e., $R^{eq}\pi = r^{eq}$ and $R^{iq}\pi \leq r^{iq}$), and can generate $\widetilde{p}$ in the counterfactual scenario (i.e., $(\widetilde{\mathbf{M}}\mathcal{H})\,\pi = \widetilde{b}_{-J}\,(\widetilde{p}) - \widetilde{\mathbf{M}}g$).[18]

Next, we derive the analytical properties of the identified set $\widetilde{\mathbf{P}}^I$. To that end, we represent the matrix

[17] The CCP vector generated by the model primitives is the unique vector that satisfies (6): since the Bellman is a contraction mapping, V is unique; from the definition of the conditional value function, we conclude that so are v_a and thus so is p (see the argument presented in footnote 10, which leads to equation (6)). The same reasoning applies to $\widetilde{p}$ in (15).

[18] Counterfactuals involving nonlinear transformations on π change the identified set $\widetilde{\mathbf{P}}^I$ defined in (16) by replacing equation (15) by $\widetilde{\mathbf{M}}\,h(\pi) = \widetilde{b}_{-J}\,(\widetilde{p})$. We ignore such counterfactuals because they are uncommon in empirical work (such counterfactuals are considered in KSS); however, extensions to nonlinear transformations are straightforward.

$\mathcal{H}$ in the following way:[19]

$$\mathcal{H} = [\mathcal{H}_{-J}, \mathcal{H}_J] = \begin{bmatrix} \mathcal{H}_{1,-J} & \mathcal{H}_{1J} \\ \mathcal{H}_{2,-J} & \mathcal{H}_{2J} \end{bmatrix}.$$

Proposition 3. *The identified set $\widetilde{\mathbf{P}}^I$ has the following properties:*
(i) It is empty if and only if Π^I is empty.
(ii) It is a connected manifold with boundary, and dimension in the interior given by the rank of the matrix $C(I - P_Q)$, where

$$C = \mathcal{H}_{1,-J} M_{-J} - \widetilde{M}_{-J} \mathcal{H}_{2,-J} M_{-J} + \mathcal{H}_{1J} - \widetilde{M}_{-J} \mathcal{H}_{2J}, \tag{17}$$

and $P_Q = Q^{eq\prime} \left(Q^{eq} Q^{eq\prime}\right)^{-1} Q^{eq}$, with

$$Q^{eq} = R^{eq}_{-J} M_{-J} + R^{eq}_J, \tag{18}$$

and $R^{eq} = [R^{eq}_{-J}, R^{eq}_J]$. Furthermore, $rank(C(I - P_Q)) \leq X - d$.
(iii) It is compact when Π^I is bounded.
(iv) In the absence of equality restrictions (7), the dimension of $\widetilde{\mathbf{P}}^I$ is given by the rank of C.[20]

Intuitively, equation (15) implicitly defines $\widetilde{p}$ as a continuously differentiable function of π. The sharp identified set $\widetilde{\mathbf{P}}^I$ is therefore the image of Π^I under this function. It is clear that $\widetilde{\mathbf{P}}^I$ is empty whenever Π^I is empty (i.e., whenever the model is rejected in the data); $\widetilde{\mathbf{P}}^I$ is connected because Π^I is convex; and $\widetilde{\mathbf{P}}^I$ is compact when Π^I is bounded (recall that Π^I is closed). An implication of the connectedness of the identified set is that a non-empty $\widetilde{\mathbf{P}}^I$ is either a singleton (in which case $\widetilde{p}$ is point-identified) or a continuum.

Proposition 3 also states that $\widetilde{\mathbf{P}}^I$ is a manifold whose dimension is given by the rank of the matrix $C(I - P_Q)$, which is smaller than or equal to $X - d$. The fact that $\widetilde{\mathbf{P}}^I$ cannot have dimension greater than $X - d$ is intuitive: since $\widetilde{\mathbf{P}}^I$ is the image set of a function defined on a $(X - d)$–dimensional polyhedron, $\widetilde{p}$ is specified by at most $X - d$ degrees of freedom rather than by $\widetilde{X}\widetilde{A}$: once $X - d$ elements are specified, the remaining are found from (15). So, whenever $X - d < \widetilde{X}\widetilde{A}$, the dimension of the identified set $\widetilde{\mathbf{P}}^I$ is strictly smaller than the dimension of the conditional probability simplex $\widetilde{\mathbf{P}}$, which implies that the (Lebesgue) measure of $\widetilde{\mathbf{P}}^I$ on $\widetilde{\mathbf{P}}$ is zero. In other words, the identified set $\widetilde{\mathbf{P}}^I$ is informative.

The rank of $C(I - P_Q)$ can be strictly smaller than $X - d$. The exact value depends on (i) the counterfactual transformation (which affects the matrix C, through the elements of $\mathcal{H}$ and $\widetilde{M}_{-J}$, defined by the econometrician), (ii) the model restrictions (which affect P_Q, through Q^{eq}, which in turn depends

[19] Note that $\widetilde{\pi} = \mathcal{H}\pi + g = \mathcal{H}_{-J}\pi_{-J} + \mathcal{H}_J \pi_J + g$. The matrix $\mathcal{H}_{-J}$ has dimension $(\widetilde{A} + 1)\widetilde{X} \times AX$, with the submatrices $\mathcal{H}_{1,-J}$ (with dimension $\widetilde{A}\widetilde{X} \times AX$) and $\mathcal{H}_{2,-J}$ (with dimension $\widetilde{X} \times AX$). Similarly, $\mathcal{H}_J$ is an $(\widetilde{A} + 1)\widetilde{X} \times X$ matrix with the submatrices $\mathcal{H}_{1J}$ (with dimension $\widetilde{A}\widetilde{X} \times X$) and $\mathcal{H}_{2J}$ (with dimension $\widetilde{X} \times X$).

[20] The matrix C has dimension $\widetilde{A}\widetilde{X} \times X$, while Q^{eq} is a $d \times X$ matrix, and both P_Q and $(I - P_Q)$ are $X \times X$ matrices.

on the linear restrictions R^{eq}), and (iii) the data (particularly, on state transitions F, which are part of the matrix M_{-J} – see equation (3) – and possibly part of $\widetilde{M}_{-J}$ through $\widetilde{F} = h^s(F)$). The interaction of these factors can reduce the dimension of the identified set further beyond $X - d$. Of note, once the econometrician establishes the counterfactual of interest and the model restrictions, the rank of $C(I - P_Q)$ can be verified directly from the data.

When $rank(C(I - P_Q)) = 0$, the identified set $\widetilde{\mathbf{P}}^I$ collapses into a singleton – i.e., $\widetilde{p}$ is point-identified. This means that all points $\pi \in \Pi^I$ map onto the same counterfactual CCP. Putting differently, even though the model restrictions do not suffice to point identify the model parameters, they may suffice to identify counterfactual behavior.

4.2 Identification of Counterfactual Outcomes of Interest

We now investigate the identified set of low-dimensional outcomes of interest $\theta \in \Theta \subset \mathbb{R}^n$.

Proposition 4. *The sharp identified set for θ is*

$$\mathbf{\Theta}^I = \left\{ \begin{array}{c} \theta \in \Theta : \exists\, (\widetilde{p}, \pi) \in \widetilde{\mathbf{P}} \times \mathbb{R}^{(A+1)X} \text{ such that} \\ \theta = f\left(\widetilde{p}, \pi;\, p, F\right),\ \mathbf{M}\pi = b_{-J}(p), \\ R^{eq}\pi = r^{eq},\ \ R^{iq}\pi \leq r^{iq}, \\ \left(\widetilde{\mathbf{M}}\mathcal{H}\right)\pi = \widetilde{b}_{-J}\left(\widetilde{p}\right) - \widetilde{\mathbf{M}}g \end{array} \right\}. \tag{19}$$

When f is a continuous function of $(\widetilde{p}, \pi)$, $\mathbf{\Theta}^I$ is a connected set. In addition, if θ is a scalar, then $\mathbf{\Theta}^I$ is an interval. Finally, when Π^I is bounded, $\mathbf{\Theta}^I$ is compact.

Proposition 4 states that a vector θ belongs to $\mathbf{\Theta}^I$ if and only if there exists a payoff π that is compatible with the data (i.e., $\mathbf{M}\pi = b_{-J}$), satisfies the model restrictions (i.e., $R^{eq}\pi = r^{eq}$ and $R^{iq}\pi \leq r^{iq}$), can generate $\widetilde{p}$ in the counterfactual scenario (i.e., $\left(\widetilde{\mathbf{M}}\mathcal{H}\right)\pi = \widetilde{b}_{-J}\left(\widetilde{p}\right) - \widetilde{\mathbf{M}}g$), and the corresponding pair $(\widetilde{p}, \pi)$ can generate θ (i.e., $\theta = f\left(\widetilde{p}, \pi;\, p, F\right)$).

When f is continuous, $\mathbf{\Theta}^I$ is connected because it is the image set of a (composite) continuous function defined on the convex polyhedron Π^I. If the model restrictions make Π^I bounded, $\mathbf{\Theta}^I$ becomes a compact and connected set, which is convenient as it suffices to trace the boundary of $\mathbf{\Theta}^I$ to characterize this set in practice. In addition, when θ is a scalar, $\mathbf{\Theta}^I$ reduces to a compact interval, which is even simpler to characterize: in that case we just need to compute the lower and upper endpoints of the interval $\mathbf{\Theta}^I$.

The upper bound of this interval can be calculated by solving the following constrained maximization problem

$$\theta^U \equiv \max_{(\widetilde{p},\pi) \in \widetilde{\mathbf{P}} \times \mathbb{R}^{(A+1)X}} f\left(\widetilde{p}, \pi;\, p, F\right) \tag{20}$$

subject to

$$\begin{aligned} \mathbf{M}\,\pi &= b_{-J}(p), \\ R^{eq}\,\pi &= r^{eq}, \\ R^{iq}\,\pi &\leq r^{iq}, \\ (\widetilde{\mathbf{M}}\mathcal{H})\,\pi &= \widetilde{b}_{-J}\left(\widetilde{p}\right) - \widetilde{\mathbf{M}}g. \end{aligned} \tag{21}$$

The lower bound of the identified set θ^L is defined similarly (but replacing max by min). For ease of exposition, we focus on the maximization problem hereafter.

The problem (20)–(21) is a nonlinear maximization problem with linear constraints on π and smooth nonlinear constraints on $\widetilde{p}$. When f is differentiable, the optimization can be solved using standard software (e.g., Knitro).

In our experience, standard algorithms are highly efficient in solving (20)–(21) in empirically-relevant high-dimensional problems when the researcher provides the gradient of f. In some cases, however, the gradient of f may be nontrivial to compute; for instance this is the case when the target parameter θ involves counterfactual average effects based on the ergodic distributions of the states, as in equation (13). For such cases, we show in Online Appendix G how to calculate the gradient of f analytically to help the numerical search. In other cases, when numerical gradients are costly to evaluate, standard solvers can be slow to converge. We thus develop a stochastic algorithm that exploits the structure of the problem (20)–(21) and combines the strengths of alternative stochastic search procedures. We discuss and describe our proposed algorithm in Online Appendix H.

4.3 Example: Firm Entry/Exit Model (Continued)

To illustrate the shape and size of the identified sets $\widetilde{\mathbf{P}}^I$ and $\mathbf{\Theta}^I$, we now return to the firm example. Let the baseline CCP be $p = (p_1'(0), p_1'(1))'$, where $p_1(k)$ is the $W \times 1$ vector with the probabilities of being active ($a = 1$) given k and w. Assuming the exogenous shocks can take only two values, low and high, and taking the same parameter values used in the construction of Figure 1, the baseline CCP is given by the vector $p = (0.714, 0.951, 0.804, 0.970)'$: the probability of entry in the low state is $p_1(0, w^l) = 0.714$, while the probability of entry in high state is $p_1(0, w^h) = 0.951$. We observe a higher probability of entry in the high state because higher values of w lead to greater profits and because w follows a persistent Markov process. Similarly, the probability of staying in the market in the low state is $p_1(1, w^l) = 0.804$, while the probability of staying in the high state is $p_1(1, w^h) = 0.970$.

The counterfactual experiment we consider in this example is a subsidy that decreases entry cost by 20%. Formally, $\widetilde{\pi} = \mathcal{H}\pi + g$, with $g = 0$, and $\mathcal{H}$ block-diagonal with the diagonal blocks $\mathcal{H}_{00}$ and $\mathcal{H}_{11}$

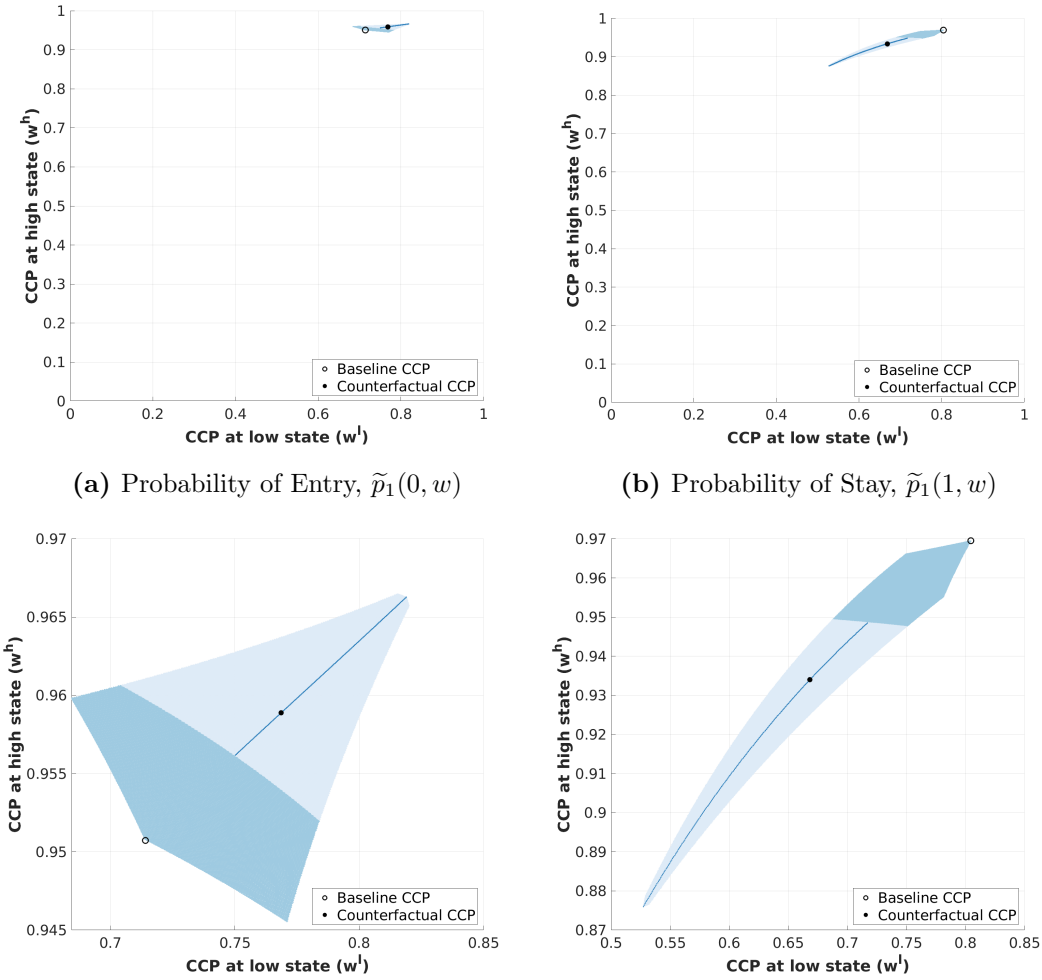
given by

$$\mathcal{H}_{00} = I, \text{ and } \mathcal{H}_{11} = \begin{bmatrix} \tau I & (1-\tau)I \\ 0 & I \end{bmatrix},$$

where $\tau = 0.8$. This implies

$$\widetilde{\pi}_0 = \mathcal{H}_{00}\pi_0 = \pi_0, \text{ and } \widetilde{\pi}_1 = \mathcal{H}_{11}\pi_1 = \begin{bmatrix} vp - fc - \tau \times ec \\ vp - fc \end{bmatrix}.$$

The counterfactual CCP is $\widetilde{p} = (0.768, 0.960, 0.668, 0.934)'$. As expected, the subsidy increases the probability of entry compared to the baseline in both low and high states w. The subsidy also decreases the probability of staying in the market, as it becomes cheaper to re-enter in the future.



(a) Probability of Entry, $\widetilde{p}_1(0, w)$ **(b)** Probability of Stay, $\widetilde{p}_1(1, w)$

(c) Zoom: Probability of Entry, $\widetilde{p}_1(0, w)$ **(d)** Zoom: Probability of Stay, $\widetilde{p}_1(1, w)$

Figure 2: Identified Set for Counterfactual CCPs, $\widetilde{\mathbf{P}}^I$, under Restrictions 1, 2, and 3. The larger sets (including the dark blue areas) correspond to $\widetilde{\mathbf{P}}^I$ under Restriction 1. The light blue areas correspond to $\widetilde{\mathbf{P}}^I$ under Restrictions 1 and 2. The identified set $\widetilde{\mathbf{P}}^I$ under Restrictions 1–3 is represented by the blue lines within the light blue areas. The baseline and counterfactual CCPs, p and $\widetilde{p}$, are represented by the black empty circle and the black full dot, respectively. The bottom panels present the "zoomed-in" versions of the top panels.

We now characterize the identified set $\widetilde{\mathbf{P}}^I$ under Restrictions 1–3. Figure 2 presents the results. Similar to our representation of Π^I in Figure 1, the larger sets (including the dark blue areas) correspond to $\widetilde{\mathbf{P}}^I$ under Restriction 1. The identified set is highly informative: it is a two-dimensional set in a four-dimensional space (recall from Proposition 3 that $\widetilde{\mathbf{P}}^I$ is at most at the same dimension as Π^I), excluding most points in $\widetilde{\mathbf{P}}$ from being possible counterfactual CCPs. Yet, because the baseline CCP p is almost at the boundary of $\widetilde{\mathbf{P}}^I$, it is difficult to rule out in practice the possibility that the entry subsidy has no impact on firm's behavior. Adding Restriction 2 reduces the size of $\widetilde{\mathbf{P}}^I$ substantially (corresponding to the light blue areas in the figure). This is a direct consequence of the smaller set Π^I obtained after imposing Restriction 2 in addition to Restriction 1 (see Figure 1). The baseline CCP does not belong to $\widetilde{\mathbf{P}}^I$ once we add Restriction 2; in fact, the location of p and $\widetilde{\mathbf{P}}^I$ allows us to conclude that the probability of entry increases in the counterfactual and that the probability of staying decreases. In other words, the sign of the treatment effect is identified. The exclusion restriction on scrap values (Restriction 3) has substantial identification power, making $\widetilde{\mathbf{P}}^I$ one-dimensional (because Π^I becomes one-dimensional as well) – see the blue lines in the figure. Note that all identified sets are connected, as expected (Proposition 3), but not necessarily convex.

We now turn to some low-dimensional outcomes θ, in particular, the long-run average impact of the entry subsidy on (i) the probability of staying in the market (labelled θ_P), (ii) consumer surplus (θ_{CS}), and (iii) the value of the firm (θ_V). Table 1 presents the identified sets for each of these outcomes under Restrictions 1–3.[21]

Perhaps surprisingly, the entry subsidy decreases the long-run average probability of the firm staying in the market, by approximately 6.4 percentage points. That is because, while the subsidy induces more entry, it also induces more exit. In the current case, increasing both firm's entry and exit rates results in less time spent in the market in the long run. This in turn reduces the long-run average consumer surplus, and raises the average long-run value of the firm.

As expected, the identified sets are all compact intervals (see Proposition 4), and they all contain the true θ. Under Restriction 1, the upper bound of the identified set for θ_P is a negative number that is very close to zero, leading to the conclusion that the long-run average probability of being active does not increase in the counterfactual. The lower bound implies that the probability of staying active can be reduced by at most 12 percentage points. Similarly, the researcher can conclude that the long-run average consumer surplus does not go up (and decreases by at most \$0.17), while the long-run average

[21] Assuming a (residual) linear inverse demand $P_{it} = w_{it} - \eta Q_{it}$, where P_{it} is the price and Q_{it} is the quantity demanded, and assuming a constant marginal cost mc, the variable profit is given by $vp = (w_{it} - mc)^2/4\eta$. The consumer surplus is $CS = 0$ when the firm is inactive ($a = 0$), and $CS = (w_{it} - mc)^2/8\eta$ when it is active ($a = 1$). Note that consumer surplus is the same in the baseline and counterfactual scenarios; the *average* CS changes in the counterfactual because the firm changes its entry behavior when it receives an entry subsidy. The value of the firm in the baseline is given by the vector $V = (I - \beta F_J)^{-1}\left(\pi_J + \psi_J(p)\right)$, where we take $J = 0$ (see footnote 10), and a similar expression holds for the counterfactual value: $\widetilde{V} = (I - \widetilde{\beta}\widetilde{F}_J)^{-1}(\widetilde{\pi}_J + \widetilde{\psi}_J(\widetilde{p}))$. The *average* firm value (across states) changes in the counterfactual both because the steady state distribution changes, and because the value of the firm is affected by the subsidy in *all* states. See the Supplemental Material, Section B, for explicit formulas for $\theta = (\theta_P, \theta_{CS}, \theta_V)$.

Table 1: Sharp Identified Sets for the Long-run Impact of the Entry Subsidy on Outcomes of Interest, $\mathbf{\Theta}^I$

Outcome of Interest	Target parameter	Sharp Identified Sets		
	True	Restriction 1	Restrictions 1–2	Restrictions 1–3
Change in Prob. of Being Active	-0.0638	[-0.1235, -0.0001]	[-0.1235, -0.0419]	[-0.1235, -0.0421]
	True	Restriction 1	Restrictions 1–2	Restrictions 1–3
Change in Consumer Surplus	-0.0875	[-0.1735, -0.0002]	[-0.1735, -0.0571]	[-0.1735, -0.0573]
	True	Restriction 1	Restrictions 1–2	Restrictions 1–3
Change in the Value of the Firm	0.9513	[0.0014, 1.8229]	[0.6375, 1.8229]	[0.6388, 1.8229]

Notes: This table shows the true and the sharp identified sets for the long-run average effect of the 20% entry subsidy on three outcomes of interest in the firm entry/exit problem: the probability of staying active, the consumer surplus, and the value of the firm. The averages are taken with respect to the state variables, using the steady-state distribution. The value of the model parameters and Restrictions 1, 2, and 3 are all specified in Section 3. See the Supplemental Material, Section B, for details.

value of the firm does not go down (and increases at most by \$1.8) in response to the subsidy. These are informative identified sets despite the fact that Restriction 1 is mild.

Adding Restriction 2 makes all identified sets more informative. The upper bound on θ_P is now lower, implying that the average probability of being active is now reduced by a number between 4 and 12 percentage points, which clearly identifies the sign of the impact. The endpoints of the intervals for θ_P and θ_V change similarly. Adding Restriction 3 does not narrow the intervals much further, despite the fact that this restriction has substantial identifying power related to the model parameters π and counterfactual behavior $\widetilde{p}$. That is because, while Restriction 3 reduces the dimension of Π^I and $\widetilde{\mathbf{P}}^I$, it does not eliminate the extreme points of these sets that, in turn, generate the endpoints of $\mathbf{\Theta}^I$. In the Supplemental Material (Section B) we present the three-dimensional identified set $\mathbf{\Theta}^I$.

5 Estimation and Inference

We now present a uniformly valid inference procedure for the main outcomes of interest $\theta \in \Theta \subset \mathbb{R}^n$. More precisely, we are interested in constructing confidence sets (CS's) for the true value of θ (rather than for the identified set $\mathbf{\Theta}^I$). Our approach is similar in spirit to the Hotz and Miller (1993) two-step estimator: we estimate choice probabilities p and transitions of state variables F in the first step, and then we perform inference on θ in the second step.

We assume the econometrician has access to a panel data on agents' actions and states: $\{a_{it}, x_{it} : i = 1, ..., N;\ t = 1, ..., T\}$. We consider asymptotics for the large N and fixed T case, as is typical in microeconometric applications of single-agent models, and assume i.i.d. sampling in the cross-section dimension.[22] Given that actions and states are finite, we consider frequency estimators for both p and F.

[22]If the data is ergodic and an appropriate mixing condition is satisfied then our procedure remains valid when $T \to \infty$

(When states are continuous, one can use kernel estimators for p and F; we leave extensions to continuous states for future research.) Specifically, for all $a \in \mathcal{A}$, and all $x, x' \in \mathcal{X}$,

$$\widehat{p}_{aN}(x) = \frac{\sum_{it} 1\{x_{it} = x, a_{it} = a\}}{\sum_{it} 1\{x_{it} = x\}}, \tag{22}$$

$$\widehat{F}_{aN}(x', x) = \frac{\sum_{it} 1\{x_{it+1} = x', x_{it} = x, a_{it} = a\}}{\sum_{it} 1\{x_{it} = x, a_{it} = a\}}, \tag{23}$$

and the vectors of sample frequencies are denoted by $\widehat{p}_N$ and $\widehat{F}_N$.[23] We collect the terms $\widehat{p}_N$ and $\widehat{F}_N$ into the L–vector $\widehat{\mathfrak{p}}_N = [\widehat{\mathfrak{p}}_{1N}, ..., \widehat{\mathfrak{p}}_{LN}]'$. Similarly, we collect p and F into $\mathfrak{p} = [\mathfrak{p}_1, ..., \mathfrak{p}_L]' := E[\mathbf{e}]$, where $\mathbf{e}$ is a vector of observed indicators. Recall that each matrix $M_a, a \in \mathcal{A}$, is a function of F, which is a subvector of $\mathfrak{p}$, therefore we define $M_a(\mathfrak{p}), a \in \mathcal{A}$, as the value of M_a evaluated at $\mathfrak{p}$ and also define $\mathbf{M}(\mathfrak{p})$ accordingly. We use the same notation for $b_{-J}(\mathfrak{p})$, as well as for $\widetilde{\mathbf{M}}(\mathfrak{p})$, $\widetilde{b}_{-J}(\widetilde{p}, \mathfrak{p})$, and $f(\widetilde{p}, \pi; \mathfrak{p})$ when appropriate.[24]

We construct a confidence set by inverting a test. The test is based on a test statistic $\widehat{J}_N(\theta_0)$ for testing the null $H_0 : \theta = \theta_0$. The nominal level $1 - \alpha$ confidence set for θ is

$$CS = \{\theta \in \Theta : N\widehat{J}_N(\theta) \leq \widehat{c}_{1-\alpha}\}, \tag{24}$$

where $\widehat{c}_{1-\alpha}$ is a data-dependent critical value (discussed below).

To test the null $H_0 : \theta = \theta_0$, we reformulate the problem in the following way. For a fixed value $\theta = \theta_0$, we take the equality constraints on π:

$$R^{eq}\pi = r^{eq}, \ (\widetilde{\mathbf{M}}(\mathfrak{p})\mathcal{H})\pi = \widetilde{b}_{-J}(\widetilde{p}, \mathfrak{p}) - \widetilde{\mathbf{M}}(\mathfrak{p})g, \text{ and } \theta_0 = f(\widetilde{p}, \pi; \mathfrak{p}), \text{ for some } \widetilde{p},$$

and collect them into

$$\mathcal{R}(\theta_0, \pi, \widetilde{p}; \mathfrak{p}) = 0.$$

This leads to the criterion function

$$J(\theta_0) := \min_{\substack{(\widetilde{p}, \pi) \in \widetilde{\mathbf{P}} \times \mathbb{R}^{(A+1)X} : R^{iq}\pi \leq r^{iq}, \\ \mathcal{R}(\theta_0, \pi, \widetilde{p}; \mathfrak{p}) = 0}} [b_{-J}(\mathfrak{p}) - \mathbf{M}(\mathfrak{p})\pi]' \, \Omega \, [b_{-J}(\mathfrak{p}) - \mathbf{M}(\mathfrak{p})\pi] \tag{25}$$

where Ω is a (user-chosen) positive definite weighting matrix. If θ_0 belongs to $\mathbf{\Theta}^I$ then all restrictions are

and N is fixed.

[23] In certain cases some elements of the transition matrix F are degenerate when the corresponding states are known to evolve deterministically; see equation (B1) in the Supplemental Material. We do not estimate these elements, and thus the expressions in (23) are applied only to the rest of the elements of F that need to be estimated.

[24] Note that $\widetilde{\mathbf{M}}$ may also depend on baseline transitions F (and so may have to be estimated in the data). That is because $\widetilde{\mathbf{M}} = [I, -\widetilde{M}_{-J}]$, where $\widetilde{M}_{-J}$ stacks $\widetilde{M}_a$ for all $a \neq J$, with $\widetilde{M}_a = (I - \beta\widetilde{F}_a)(I - \beta\widetilde{F}_J)^{-1}$, and $\widetilde{F} = h^s(F)$. The same applies to $\widetilde{b}_{-J}$, which also depends on $\widetilde{F}$.

satisfied and $J(\theta_0) = 0$, otherwise $J(\theta_0) > 0$. The identified set $\mathbf{\Theta}^I$ can therefore be represented as

$$\mathbf{\Theta}^I = \{\theta \in \Theta : J(\theta) = 0\}, \tag{26}$$

which implies that the null $H_0 : \theta = \theta_0$ is equivalent to $H_0' : J(\theta_0) = 0$.

The empirical counterpart of $J(\theta_0)$ is given by

$$\widehat{J}_N(\theta_0) := \min_{\substack{(\widetilde{p},\pi)\in\widetilde{\mathbf{P}}\times\mathbb{R}^{(A+1)X} : R^{iq}\pi\leq r^{iq}, \\ \mathcal{R}(\theta_0,\pi,\widetilde{p};\widehat{\mathfrak{p}}_N)=0}} [b_{-J}(\widehat{\mathfrak{p}}_N) - \widehat{\mathbf{M}}_N\pi]' \, \widehat{\Omega}_N \, [b_{-J}(\widehat{\mathfrak{p}}_N) - \widehat{\mathbf{M}}_N\pi], \tag{27}$$

where $\widehat{\mathbf{M}}_N = \mathbf{M}(\widehat{\mathfrak{p}}_N)$ and $\widehat{\Omega}_N$ is a consistent estimator for Ω. For the rest of this paper we consider a general specification of Ω so that it can be a (known) continuous function of $\mathfrak{p}$. Denoting the function by $\Omega(\cdot)$, we let $\widehat{\Omega}_N = \Omega(\widehat{\mathfrak{p}}_N)$ in (27).

While a naive bootstrap for $\widehat{J}_N(\theta_0)$ fails to deliver critical values that are asymptotically uniformly valid (see, e.g. Kitamura and Stoye, 2018), subsampling works under weak conditions, as we shall show shortly. Let h_N be the subsample size, with $h_N \to \infty$ as $N \to \infty$. A subsample version of $\widehat{J}_N(\theta_0)$ is

$$\widehat{J}^*_{h_N}(\theta_0) := \min_{\substack{(\widetilde{p},\pi)\in\widetilde{\mathbf{P}}\times\mathbb{R}^{(A+1)X} : R^{iq}\pi\leq r^{iq}, \\ \mathcal{R}\left(\theta_0,\pi,\widetilde{p};\widehat{\mathfrak{p}}^*_{h_N}\right)=0}} [\widehat{b}^*_{-J} - \widehat{\mathbf{M}}^*_{h_N}\pi]'\widehat{\Omega}^*_{h_N}[\widehat{b}^*_{-J} - \widehat{\mathbf{M}}^*_{h_N}\pi], \tag{28}$$

where $\widehat{\mathfrak{p}}^*_{h_N}$ is a subsample estimator of $\mathfrak{p}$, $\widehat{\mathbf{M}}^*_{h_N} = \mathbf{M}(\widehat{\mathfrak{p}}^*_{h_N})$, $\widehat{\Omega}^*_{h_N} = \Omega(\widehat{\mathfrak{p}}^*_{h_N})$ and

$$\widehat{b}^*_{-J} = b_{-J}(\widehat{\mathfrak{p}}^*_{h_N}) - b_{-J}(\widehat{\mathfrak{p}}_N) + \widehat{b}_{-J}(\widehat{\mathfrak{p}}_N),$$

with $\widehat{b}_{-J}(\widehat{\mathfrak{p}}_N)$ being the value of $\widehat{\mathbf{M}}_N\pi$ solving the minimization problem (27). Note that with this definition of $\widehat{b}^*_{-J}$ we implement subsampling with centering.

The testing procedure is simple: We use the empirical distribution of $h_N\widehat{J}^*_{h_N}(\theta_0)$ to obtain the critical value $\widehat{c}_{1-\alpha}$. When the value of the test statistic is smaller than the critical value, $N\widehat{J}_N(\theta_0) \leq \widehat{c}_{1-\alpha}$, we do not reject the null $H_0' : J(\theta_0) = 0$, otherwise we reject it. The $1-\alpha$ confidence set will be the collection of θ_0's for which the tests do not reject the null.

Remark 4. A comment on some approaches that are alternative to ours as outlined above is in order. First, if we treat $(\pi, \widetilde{p})$ as a parameter (while θ_0 is fixed), then our system becomes one of set identified moment *equalities* (composed of equations (6), (12), and (15)), with (inequality) constraints on the parameter space for $(\pi, \widetilde{p})$ (i.e., restrictions (7) and (8)). It is then possible to test the validity of these equalities at each value of $(\pi, \widetilde{p})$. This controls size, but would be extremely conservative; obviously the same can be done to standard moment inequality models but it is not implemented in practice for this reason. Moreover, implementing such a procedure in our context is practically impossible, as the

parameter space for $(\pi, \widetilde{p})$ is too big. Second, we can fix $\widetilde{p}$, but not π, and rewrite the system into a moment inequality form by eliminating π (i.e. solving for other variables). As noted in Kitamura and Stoye (2018), this amounts to transforming, in the language of discrete geometry, a $\mathcal{V}$-representation of a polytope to an $\mathcal{H}$-representation, and it is generally known to be expensive to compute, and impractical even for a moderately sized system. Third, one may try to eliminate both π and $\widetilde{p}$ from the system to get some form of moment inequalities; but this is even harder to implement, especially because of the nonlinear constraints that involve $\widetilde{p}$, and so it is not a practically feasible option either. For example, a recent paper by Kaido, Molinari, and Stoye (2019) is, like ours, concerned with a low dimensional object, though it is not directly applicable as their algorithm requires a moment inequality representation.

Before we state formal assumptions and the asymptotic validity result, it is useful to introduce some notation. First, define the manifold

$$S(\mathfrak{p}, \theta) := \{\mathbf{M}(\mathfrak{p})\pi, \pi \in \mathbb{R}^{(A+1)X} : \mathcal{R}(\theta, \pi, \widetilde{p};\, \mathfrak{p}) = 0, \text{ and } R^{iq}\pi \leq r^{iq} \text{ hold for some } \widetilde{p} \in \widetilde{\mathbf{P}}\}.$$

Note that the minimization problem (25) projects $b_{-J}(\mathfrak{p})$ on $S(\mathfrak{p}, \theta)$ under the weighted norm $\|x\|_\Omega = x'\Omega x$, for $x \in \mathbb{R}^{AX}$. The corresponding value of the objective function $J(\theta)$ in (25) is the squared length of the projection residual vector. Clearly, $\theta \in \mathbf{\Theta}^I$ if and only if the residual vector is zero.

Next, for some positive constants c_1 and c_2, define the set

$$\mathcal{P}_{\theta_0} := \left\{ \begin{array}{c} \mathfrak{p} : \mathfrak{p}_\ell \in (0,1), E\left[\left|\frac{\mathbf{e}_\ell}{\sqrt{\mathfrak{p}_\ell(1-\mathfrak{p}_\ell)}}\right|^{2+c_1}\right] < c_2, 1 \leq \ell \leq L, \\ \exists(\widetilde{p}, \pi) \in \widetilde{\mathbf{P}} \times \mathbb{R}^{(A+1)X} \text{ such that } \mathbf{M}(\mathfrak{p})\pi = b_{-J}(\mathfrak{p}), \\ R^{iq}\pi \leq r^{iq},\ \mathcal{R}(\theta_0, \pi, \widetilde{p}; \mathfrak{p}) = 0, \det(\Omega(\mathfrak{p})) \geq c_1 \end{array} \right\}.$$

This represents the set of permissible data generating processes when the counterfactual value of interest is fixed at θ_0. Note that the first restriction in the definition $\mathcal{P}_{\theta_0}$ is a standard condition imposed to guarantee the Lindeberg condition. The second is the main model restriction. The third and the fourth collect additional constraints on the payoff vector π; the equalities in the fourth restriction include the constraints that arise as we fix the value of the counterfactual θ_0. The final restriction guarantees that the random manifold $S(\widehat{\mathfrak{p}}_N, \theta_0)$ is asymptotically well-behaved.

We impose a weak condition on f and h^s:

Condition 1. *f and h^s are C^1 functions.*

It is useful to impose a mild requirement on $S(\mathfrak{p}, \theta)$ in terms of its local geometric property. To this end, we introduce the notion of tangent cone:

Definition 1. *For a (possibly non-convex) set $A \subset \mathbb{R}^d$, the tangent cone of A at $x \in A$, henceforth*

denoted by $T_A(x)$, is given by

$$T_A(x) := \limsup_{\tau \downarrow 0} \tau^{-1}(A \ominus x),$$

where $\ominus$ denotes the usual Minkowski difference.

See, e.g., Section 6A of Rockafellar and Wets (2009) for a discussion on the role of a tangent cone and other related concepts.

Condition 2. *For every $(\mathfrak{p}, \theta_0)$ such that $\mathfrak{p} \in \mathcal{P}_{\theta_0}$ and $\theta_0 \in \mathbf{\Theta}^I$, the tangent cone $T_{S(\mathfrak{p},\theta_0)}(x)$ of $S(\mathfrak{p}, \theta_0)$ is convex at each $x \in \mathbb{R}^{AX} \in S(\mathfrak{p}, \theta_0)$.*

Then the next theorem follows:

Theorem 1. *Choose h_N such that $h_N \to \infty$ and $h_N/N \to 0$ as $N \to \infty$. Then under Conditions 1 and 2,*

$$\liminf_{N\to\infty} \inf_{\mathfrak{p}\in\mathcal{P}_{\theta_0}} \Pr\{N\widehat{J}_N(\theta_0) \le \widehat{c}_{1-\alpha}\} = 1 - \alpha, \quad \text{for every } \theta_0 \in \mathbf{\Theta}^I,$$

where $\widehat{c}_{1-\alpha}$ is the $1-\alpha$ quantile of $h_N \widehat{J}^*_{h_N}(\theta_0)$, with $0 \le \alpha \le \frac{1}{2}$. The asymptotically uniformly valid $1-\alpha$ confidence set for θ is the collection of θ_0's such that the test does not reject the null $H_0' : J(\theta_0) = 0$.

Our test statistic (27) is the squared minimum distance between the random vector $b_{-J}(\widehat{\mathfrak{p}}_N)$ and the random manifold $S(\widehat{\mathfrak{p}}_N, \theta_0)$. It is therefore crucial to take sampling uncertainty in both objects into account. Also, note that Condition 2 does *not* require that the set $S(\widehat{\mathfrak{p}}_N, \theta_0)$ is convex, even locally. That is, the set does not have to be convex even in a small neighborhood of the true population value b_{-J}, so that there may not exist any positive constant ϵ such that the intersection of the ϵ-neighborhood and $S(\widehat{\mathfrak{p}}_N, \theta_0)$ (or $S(\mathfrak{p}_N, \theta_0)$) is convex. We avoid such standard convexity conditions as they are typically incompatible with our model restrictions, in particular the general equality restrictions $\mathcal{R}(\theta_0, \pi, \widetilde{p}; \mathfrak{p}) = 0$. The above result establishes the asymptotic validity of our procedure, addressing these issues.

Next, we discuss briefly some practical issues when implementing the subsampling procedure. We present further details in Online Appendix F.

Practical Implementation. To simplify, we focus the discussion on the scalar case, $\theta \in \mathbb{R}$. We suggest implementing the subsampling procedure in the following way. First, calculate the lower and upper bounds of the interval $\mathbf{\Theta}^I = [\theta^L, \theta^U]$ by solving the maximization (and minimization) problem (20)–(21) in the full sample; denote them by $\widehat{\theta}^L$ and $\widehat{\theta}^U$. Clearly, $\widehat{J}_N(\theta_0) = 0$ for all $\theta_0 \in [\widehat{\theta}^L, \widehat{\theta}^U]$, so the null hypothesis $H_0' : J(\theta_0) = 0$ will not be rejected for any point in that interval. We therefore start the grid-search at points slightly below $\widehat{\theta}^L$ and slightly above $\widehat{\theta}^U$. Consider the points above $\widehat{\theta}^U$: we start with the point, say, $\theta_0 = \widehat{\theta}^U + 0.01$, and test the null $H_0' : J(\theta_0) = 0$, as described above. If we fail to reject the null, we then move to the next point, say, $\theta_0 = \widehat{\theta}^U + 0.02$ and test the null for that new point. We keep doing so until we reject the null for the first time; we stop the grid-search when we first reject the null because all

points to the right will be rejected by the data as well; we adopt a similar procedure for the lower end $\widehat{\theta}^L$. Changing θ_0 sequentially and incrementally also has the advantage of providing good initial guesses in a series of optimizations: because (27) is a smooth and well-behaved problem, the solution to the latest minimization can be used as an initial value for the next minimization, reducing the total computational costs; the same applies in the critical value calculations.[25] If we reject the null for the first time at the points θ^l and θ^u, then the asymptotically uniformly valid $1-\alpha$ confidence set for the true θ is the interval $[\theta^l, \theta^u]$.

6 Empirical Application

In this section, we illustrate our approach in the context of a dynamic model of export behavior. To that end, we consider the setup of Das, Roberts, and Tybout (2007), henceforth 'DRT', who use plant-level panel data from Colombian manufacturing industries to investigate the impact of export subsidies. As the authors point out, industrial exporters are highly prized in developing countries for generating gains from trade, sustaining production and employment during domestic recessions, and facilitating the absorption of foreign technologies. As a consequence, exporters often receive governmental support. Yet, seemingly similar subsidies may generate different export responses in different industries and time periods, making it difficult for policy makers to know which type of support is optimal. To shed light on these issues, DRT develop a structural dynamic model of firm export decisions and simulate the impact of various subsidies on gains in export revenues per peso of subsidy. Here, we adopt their specification and explore the identifying power of alternative model restrictions.

Data. We consider the knitting mills industry. The dataset is composed of 64 knit fabric producers observed annually during the period 1981–1991; the sample has 704 plant-year observations. Like DRT, we focus on firms that operated continuously in the domestic market, given that they were responsible for most of the exports over this period. The share of exporting firms increased from 12 percent in 1981 to 18 percent by the end of the sample period, possibly a result of the 33% depreciation of Colombia's real exchange rate. This industry also depicts significant turnover: the average probability of re-entry into export markets is 61%. On average, export revenues of exporting firms are approximately 1.4 times the domestic revenues.

Model. DRT assume that export markets are monopolistically competitive; this leads to a specification similar to the firm entry/exit model presented in Sections 3 and 4. In particular, every period t a firm

[25] In addition to the (limited) grid-search and the sequence of optimizations, we can exploit the relation between the optimization problems (20)–(21) and (27) (as well as (28)) to improve the performance of the subsampling further: in our experience it is easier to solve relaxed versions of (20)–(21) to obtain good approximations for $\widehat{J}_N(\theta_0)$ than solving (27) directly. Furthermore, subsampling is amenable to parallelization, which speeds up the procedure. See Online Appendix F for details.

i chooses whether to export or not, $a_{it} \in \mathcal{A} = \{0, 1\}$. The state variables are (i) the lagged decisions ($k_{it} = a_{it-1}$), (ii) exchange rates (e_t), and (iii) demand/supply shocks in export markets (ν_{it}). The exogenous shocks $w_{it} = (e_t, \nu_{it})$ follow (discretized) independent normal-AR(1) processes. The payoff function is given by equation (10) in Section 3. To point identify the model, DRT restrict to zero the payoffs of not exporting (i.e., both the outside value and the scrap value are set to zero). They also impose state-invariant entry and fixed costs, making their model overidentified. We relax these assumptions and instead explore the identifying power of Restrictions 1, 2, and 3 presented in the entry/exit model. In principle, scrap values may differ from zero because they may involve idleness costs (given that exiting is often temporary) or depreciation costs. Similarly, fixed costs and entry costs may depend on the aggregate states, as they involve finding trading partners, setting up distribution networks, maintaining labor and capital abroad, etc. For ease of exposition, we leave the model details to the Supplemental Material.[26]

Counterfactuals and Outcomes of Interest. DRT focus on three counterfactual policies: (i) direct subsidies linked to plants' export revenues, such as a tax rebate that is proportional to foreign sales; (ii) subsidies to the costs of entering into exporting, such as grants for information or technology acquisition for export development; and (iii) payments designed to cover the annual fixed costs of operating in the export market. We follow DRT and consider a 2% export revenue subsidy, a 25% entry cost subsidy, and a 28% fixed cost subsidy.

The main outcome of interest is a benefit–cost ratio based on the average annual gain in total export revenues divided by the average government subsidy expenditures (both averaged over states in the long-run). We denote the ratios for the revenue, fixed costs, and entry costs subsidies by θ_R, θ_F, and θ_E, respectively, and take $\theta = (\theta_R, \theta_F, \theta_E)$ – see the Supplemental Material (Section D) for explicit formulas for θ.

Evaluating ex-ante the impact of different model restrictions on θ is not trivial. Note first that while export revenues are observed in the data, the long-run *average* change in revenues depends on firms' decisions to export given the type of subsidy. This means that all numerators in θ depend on the counterfactual CCPs. Next, note that all denominators in θ equal the long-run average government expenditures, which depend on the fraction of firms exporting in the counterfactual steady-state; i.e., they all depend on $\widetilde{p}$ as well. In addition, θ_F and θ_E depend on the unknown parameters, fc and ec, respectively, since the government expenditures are direct functions of these costs. In the case of the entry cost subsidy, a further complication is that the (subsidized) entry cost is paid only when firms enter, implying that $\widetilde{p}$ affects the direct payments in each state (in addition to affecting the steady-state distribution). In short, θ depends on both $\widetilde{p}$ and π highly nonlinearly.

In terms of identification, the benefit-cost ratio of the revenues subsidy θ_R is point-identified. That

[26]The payoff when not exporting (the outside option) may also be different from zero since it includes domestic profits. However, following DRT, we do not explore this possibility given the limitations in the data.

is both because the averages in the numerator and denominator depend on observed revenues (i.e., the integrands are observable), and because $\widetilde{p}$ is identified (since it involves known changes to known quantities, i.e., the identified variable profits; see KSS), implying that the counterfactual steady-state distribution is also point-identified. The other two target objects, θ_F and θ_E, are partially identified both because (i) the counterfactual behavior $\widetilde{p}$ is not point-identified (as the entry subsidy in our toy example in Section 4), and (ii) the denominators in the benefit–cost ratios depend *directly* on model parameters that are partially identified (i.e., on fc and ec, respectively). In sum, both θ_F and θ_E involve ratios of set-identified objects.

Results. We implement our two-step procedure as explained in Sections 4 and 5, and in Online Appendix F.[27] Table 2 presents the benefit–cost ratios under Restrictions 1–3. The revenue subsidy generates an estimated benefit–cost ratio θ_R of approximately 15 pesos of revenue per unit cost. Its impact is statistically significant and economically large, and it is fairly consistent with the estimates in DRT. Because θ_R is point identified, it does not depend on any additional model restriction (other than the basic framework (6)).

We now discuss θ_F and θ_E, which are partially identified. Restriction 1 (i.e., $fc \geq 0$ and $ec \geq 0$) is not sufficiently informative here: the fixed cost subsidies ratio θ_F is between 8 and 30, and the entry cost subsidies ratio θ_E ranges from 4 to 24. These sets are wide because there are still many model parameter values that can rationalize observed behavior. The identified sets overlap and we cannot conclude which policy generates the greatest impact on exports.

Adding Restriction 2 increases the identification power substantially: the ratio for the fixed cost subsidies is now between 11 and 13. This identified set is highly informative and its upper bound is smaller than θ_R, suggesting that the revenue subsidy is more potent than the fixed cost subsidy. Still, there is substantial uncertainty regarding the benefit–cost ratio for the entry cost subsidy: its identified set is now between 7.8 and 17, containing both the estimated θ_R and the identified set of θ_F. Incorporating exclusion restrictions on scrap values (Restriction 3) narrows the identified set for θ_E substantially: the benefit-cost ratio now ranges from 8.9 to 9.4, which is highly informative.[28]

There is a clear ranking of the policies under Restrictions 1–3: revenue subsidies generate the highest export revenues per unit cost, followed by fixed cost subsidies, and then by the entry cost subsidies. That

[27] The transition process for exchange rates is taken from a long-time series as in DRT. Given the small sample size, we discretize the support of each exogenous state, e_t and ν_{it}, in three bins. We estimate CCPs using frequency estimators. To compute confidence intervals, we implement 1000 replications of a standard i.i.d. subsampling, resampling 20 firms over the sample time period, so that the size of each subsample is $h_N = 200 \approx 8 \times \sqrt{NT}$. To minimize the quadratic distances in (27) and (28), we take a diagonal weighting matrix Ω with diagonal elements given by the square-root of the ergodic distribution of the state variable – thus, deviations on more visited states are considered more relevant and receive greater weights. Given that θ_R is known (ex ante) to be point identified, we use the plug-in estimator proposed by Kalouptsidi, Lima, and Souza-Rodrigues (2019) to estimate it, and 1000 standard i.i.d. bootstrap replications at the firm level to construct the confidence intervals for θ_R. To make our results comparable to DRT, we have also estimated the model parameters under their restrictions and obtained very similar results as theirs. See details in Section D of the Supplemental Material.

[28] Of note, the reduction is driven mostly by assuming scrap values do not depend on demand/costs shocks ν_{it}. This (limited) exclusion restriction may be reasonable when scrap values include idleness and depreciation costs incurred abroad, which may depend on exchange rates, but not on, say, demand shocks.

Table 2: Export Revenue/Cost Ratio for Different Subsidies under Alternative Model Restrictions

	Restriction 1	Restrictions 1–2	Restrictions 1–3
		Revenue Subsidies	
Estimated Identified Set	15.13	15.13	15.13
90% Confidence Interval	(11.15, 18.90)	(11.15, 18.90)	(11.15, 18.90)
		Fixed Costs Subsidies	
Estimated Identified Set	[8.41, 30.82]	[11.10, 13.34]	[11.92, 12.60]
90% Confidence Interval	(7.47, 34.98)	(9.65, 14.46)	(9.92, 13.87)
		Entry Costs Subsidies	
Estimated Identified Set	[4.40, 24.04]	[7.85, 17.28]	[8.88, 9.36]
90% Confidence Interval	(3.52, 34.36)	(7.03, 23.49)	(7.34, 14.33)

Notes: This table shows the sharp identified sets for the average gains in total export revenues divided by the average government subsidy expenditures, both averaged over states in the long-run. The top panel shows the gains of a 2% export revenue subsidy; the middle panel, the gains of a 28% fixed cost subsidy; and the bottom panel, the gains of a 25% entry cost subsidy. The (nonsingleton) identified sets are in brackets. The data set is composed of 704 plant-year observations in the Colombian knitting mills industry. The 90% confidence intervals are in parenthesis and were calculated based on 1000 bootstrap replications for the revenue subsidies, and 1000 subsample replications for both fixed and entry costs subsidies (with subsample sizes of 200). Restrictions 1, 2, and 3 are all specified in the main text (Section 3). See Online Appendix D for details.

is exactly the ranking obtained by DRT. The result is intuitive: revenue subsidies affect both volume and entry margins, while fixed costs and entry costs influence only the entry and exit decisions. In addition, fixed cost subsidies do not encourage exit behavior of forward-looking firms, while entry cost subsidies do. Still, notwithstanding these intuitive effects, the ranking seems to hinge on the assumption that scrap values do not depend on state variables.

Of note, the uniformly valid confidence intervals indicate substantial sampling uncertainty, which is not surprising given the size of the data set.

7 Conclusion

In this paper, we study partial identification of model parameters and counterfactual objects in dynamic discrete choice models. We derive analytical properties of the identified sets under alternative model restrictions. We propose computational procedures for estimation and develop an asymptotically uniformly valid inference approach based on subsampling. A Monte Carlo study of firm entry/exit shows the good finite-sample performance of our procedure. Finally, we demonstrate the empirical implications of our results in the study of Das, Roberts, and Tybout (2007) on exporting decisions and subsidies. We leave extensions to identification of optimal policy interventions and to dynamic games for future research.

References

Abbring, J. H., and O. Daljord (2019): "Identifying the Discount Factor in Dynamic Discrete Choice Models," *Quantitative Economics*, p. forthcoming.

Adams, A. (2020): "Mutually Consistent Revealed Preference Demand Predictions," *American Economic Journal: Microeconomics*, 12(1), 42–74.

Adao, R., A. Costinot, and D. Donaldson (2017): "Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade," *American Economic Review*, 107(3), 633–89.

Aguirregabiria, V. (2010): "Another Look at the Identification of Dynamic Discrete Decision Processes: An application to Retirement Behavior," *Journal of Business & Economic Statistics*, 28(2), 201–218.

Aguirregabiria, V., and P. Mira (2002): "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," *Econometrica*, 70(4), 1519–1543.

——— (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75(1), 1–53.

——— (2010): "Dynamic Discrete Choice Structural Models: A Survey," *Journal of Econometrics*, 156(1), 38–67.

Aguirregabiria, V., and J. Suzuki (2014): "Identification and Counterfactuals in Dynamic Models of Market Entry and Exit," *Quantitative Marketing and Economics*, 12(3), 267–304.

Arcidiacono, P., and R. A. Miller (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity," *Econometrica*, 79(6), 1823–1867.

——— (2018): "Identifying Dynamic Discrete Choice Models off Short Panels," *Journal of Econometrics*, p. forthcoming.

Bajari, P., C. L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75(5), 1331–1370.

Bajari, P., C. S. Chu, D. Nekipelov, and M. Park (2016): "Identification and semiparametric estimation of a finite horizon dynamic discrete choice model with a terminating action," *Quantitative Marketing and Economics*, 14(4), 271–323.

Bejara, M. (2018): "Counterfactual Equivalence in Macroeconomics," Discussion paper, MIT.

Berry, S., and G. Compiani (2019): "An Instrumental Variables Approach to Dynamic Models," Discussion paper, Yale University.

Blevins, J. R. (2014): "Nonparametric identification of dynamic decision processes with discrete and continuous choices," *Quantitative Economics*, 5(3), 531–554.

Blundell, R., M. Browning, and I. Crawford (2008): "Best Nonparametric Bounds on Demand Responses," *Econometrica*, (76), 1227–1262.

Blundell, R., D. Kristensen, and R. Matzkin (2014): "Bounding quantile demand functions using revealed preference inequalities," *Journal of Econometrics*, 179(2), 112 – 127.

Buchholz, N., M. Shum, and H. Xu (2019): "Semiparametric Estimation of Dynamic Discrete Choice Models," Discussion paper, Princeton University.

Canay, I. A., and A. M. Shaikh (2017): "Practical and Theoretical Advances in Inference for Partially Identified Models," in *Advances in Economics and Econometrics: Eleventh World Congress*, ed. by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, vol. 2 of *Econometric Society Monographs*, pp. 271–306. Cambridge University Press.

Chen, L.-Y. (2017): "Identification of Discrete Choice Dynamic Programming Models with Nonparametric Distribution of Unobservables," *Econometric Theory*, 33(3), 551–577.

Chernoff, H. (1954): "On the Distribution of the Likelihood Ratio," *The Annals of Mathematical Statistics*, pp. 573–578.

Chiong, K. X., A. Galichon, and M. Shum (2016): "Duality in dynamic discrete-choice models," *Quantitative Economics*, 7(1), 83–115.

Christensen, T., and B. Connault (2019): "Counterfactual Sensitivity and Robustness," Discussion paper, New York University.

Collard-Wexler, A. (2013): "Demand Fluctuations in the Ready-Mix Concrete Industry," *Econometrica*, 81(3), 1003–1037.

Conlon, C. T. (2012): "A Dynamic Model of Costs and Margins in the LCD TV Industry," Discussion paper, New York University.

Crawford, G. S., and M. Shum (2005): "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 73(4), 1137–1173.

Das, S., M. J. Roberts, and J. R. Tybout (2007): "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," *Econometrica*, 75(3), 837–873.

Dearing, A. (2019): "Pseudo-Value Functions and Closed-Form CCP Estimation of Dynamic Discrete Choice Models," Discussion paper, Ohio State University.

Dickstein, M. J., and E. Morales (2018): "What Do Exporters Know?," *The Quarterly Journal of Economics*, 133(4), 1753–1801.

Duflo, E., R. Hanna, and S. P. Ryan (2012): "Incentives work: Getting teachers to come to school," *The American Economic Review*, 102(4), 1241–1278.

Dunne, T., S. D. Klimek, M. J. Roberts, and D. Y. Xu (2013): "Entry, exit, and the determinants of market structure," *The RAND Journal of Economics*, 44(3), 462–487.

Eckstein, Z., and O. Lifshitz (2011): "Dynamic Female Labor Supply," *Econometrica*, 79(6), 1675–1726.

Gilleskie, D. B. (1998): "A Dynamic Stochastic Model of Medical Care Use and Work Absence," *Econometrica*, 66(1), 1–45.

Heckman, J. J. (2000): "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective," *The Quarterly Journal of Economics*, 115(1), 45–97.

——— (2010): "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy," *Journal of Economic Literature*, 48(2), 356–98.

Heckman, J. J., and S. Navarro (2007): "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, 136(2), 341–396.

Hendel, I., and A. Nevo (2006): "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, 74(6), 1637–1673.

Hotz, V. J., and R. A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60(3), 497–529.

Hotz, V. J., R. A. Miller, S. Sanders, and J. Smith (1994): "A Simulation Estimator for Dynamic Models of Discrete Choice," *Review of Economic Studies*, 61(2), 265–89.

Hu, Y., and M. Shum (2012): "Nonparametric identification of dynamic models with unobserved state variables," *Journal of Econometrics*, 171(1), 32–44.

Ichimura, H., and C. Taber (2000): "Direct Estimation of Policy Impacts," *NBER Working Paper 254*.

——— (2002): "Semiparametric Reduced-Form Estimation of Tuition Subsidies," *American Economic Review*, 92(2), 286–292.

Kaido, H., F. Molinari, and J. Stoye (2019): "Confidence intervals for projections of partially identified parameters," *Econometrica*, 87(4), 1397–1432.

Kalouptsidi, M. (2014): "Time to build and fluctuations in bulk shipping," *The American Economic Review*, 104(2), 564–608.

Kalouptsidi, M., L. Lima, and E. Souza-Rodrigues (2019): "On Estimating Counterfactuals Directly in Dynamic Models," Discussion paper, University of Toronto.

Kalouptsidi, M., P. T. Scott, and E. Souza-Rodrigues (2017): "On the Non-identification of Counterfactuals in Dynamic Discrete Games," *International Journal of Industrial Organization*, 50, 362–371.

Kalouptsidi, M., P. T. Scott, and E. A. Souza-Rodrigues (2019): "Identification of Counterfactuals in Dynamic Discrete Choice Models," *Quantitative Economics*, forthcoming.

Kasahara, H., and K. Shimotsu (2009): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica*, 77(1), pp. 135–175.

Keane, M. P., and A. Merlo (2010): "Money, Political Ambition, and the Career Decisions of Politicians," *American Economic Journal: Microeconomics*, 2(3), 186–215.

Keane, M. P., P. E. Todd, and K. I. Wolpin (2011): "The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications," *Handbook of Labor Economics*, 4, 331–461.

Keane, M. P., and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522.

——— (2010): "The Role of Labor and Marriage Markets, Preference Heterogeneity and the Welfare System in the Life Cycle Decisions of Black, Hispanic and White Women," *International Economic Review*, 51(3), 851–892.

Kitamura, Y., and J. Stoye (2018): "Nonparametric Analysis of Random Utility Models," *Econometrica*, 86(6), 1883–1909.

Kitamura, Y., and J. Stoye (2019): "Nonparametric Counterfactuals in Random Utility Models," Discussion paper, Yale University.

Lin, H. (2015): "Quality Choice and Market Structure: A Dynamic Analysis of Nursing Home Oligopolies," *International Economic Review*, 56(4), 1261–1290.

Low, H., and C. Meghir (2017): "The Use of Structural Models in Econometrics," *Journal of Economic Perspectives*, 31(2), 33–58.

Magnac, T., and D. Thesmar (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70(2), 801–816.

Manski, C. F. (2007): "Partial Identification of Counterfactual Choice Probabilities," *International Economic Review*, 48, 1393–1410.

Marschak, J. (1953): "Economic Measurements for Policy and Prediction," in *Cowles Commission Monograph 14: Studies in Econometric Methods*, ed. by W. C. Hood, and T. Koopmans. New York: Wiley.

Miller, R. (1984): "Job matching and occupational choice," *Journal of Political Economy*, 92(6), 1086–1120.

Mogstad, M., A. Santos, and A. Torgovitsky (2018): "Using Instrumental Variables for Inference about Policy Relevant Treatment Parameters," *Econometrica*, 86(5), 1589–1619.

Morales, E., G. Sheu, and A. Zahler (2019): "Extended gravity," *The Review of Economic Studies*, 86(6), 2668–2712.

Norets, A., and X. Tang (2014): "Semiparametric Inference in dynamic binary choice models," *The Review of Economic Studies*, 81(3), 1229–1262.

Pakes, A. (1986): "Patents as options: Some estimates of the value of holding European patent stocks," *Econometrica*, 54(4), 755–784.

Pakes, A., M. Ostrovsky, and S. Berry (2007): "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)," *The RAND Journal of Economics*, 38(2), 373–399.

Pesendorfer, M., and P. Schmidt-Dengler (2008): "Asymptotic Least Squares Estimators for Dynamic Games," *The Review of Economic Studies*, 75(3), 901–928.

Rockafellar, R. T., and R. J.-B. Wets (2009): *Variational analysis*, vol. 317. Springer Science & Business Media.

Romano, J. P., and A. M. Shaikh (2008): "Inference for identifiable parameters in partially identified econometric models," *Journal of Statistical Planning and Inference*, 138(9), 2786–2807.

——— (2012): "On the uniform asymptotic validity of subsampling and the bootstrap," *The Annals of Statistics*, 40(6), 2798–2822.

Rosenzweig, M. R., and K. I. Wolpin (1993): "Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low-Income Countries: Investments in Bullocks in India," *Journal of Political Economy*, 101(2), 223–244.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: an Empirical Model of Harold Zurcher," *Econometrica*, 55(5), 999–1033.

——— (1994): "Structural Estimation of Markov Decision Processes," *Handbook of Econometrics 4*, 4, 3081–3143.

Ryan, S. P. (2012): "The Costs of Environmental Regulation in a Concentrated Industry," *Econometrica*, 80(3), 1019–1061.

Schiraldi, P. (2011): "Automobile replacement: a dynamic structural approach," *The RAND Journal of Economics*, 42(2), 266–291.

Varela, M. J. (2018): "The costs of growth: Accelerated growth and crowd-out in the Mexican supermarket industry," *International Journal of Industrial Organization*, 61, 1–52.

Wei, C., and S. Li (2014): "The Cost of Greening Stimulus: A Dynamic Discrete Choice Analysis of Vehicle Scrappage Programs," Working Papers 2014-12, The George Washington University, Institute for International Economic Policy.

Wolpin, K. (1984): "An estimable dynamic stochastic model of fertility and child mortality," *Journal of Political Economy*, 92(5), 852–874.

Supplement to "Partial Identification and Inference for Dynamic Models and Counterfactuals"

Myrto Kalouptsidi, Yuichi Kitamura, Lucas Lima, and Eduardo Souza-Rodrigues*

February, 2020

This supplemental material consists of the following sections: Section A presents all proofs of the propositions and theorems presented in the main paper. Section B provides detailed information about the firm entry/exit problem – our running example. Section C shows our Monte Carlo study. And Section D discusses our replication of Das, Roberts, and Tybout (2007).

A Proofs

A.1 Proof of Proposition 1

The identified set (9) is sharp by construction because equations (6), (7), and (8) contain all model restrictions. Further, Π^I is a convex polyhedron given that it is the intersection of finitely many closed halfspaces. In the absence of inequalities (8), Π^I is a linear manifold with dimension that equals $X - d$. This implies that the dimension of Π^I under all restrictions also is $X - d$.

A.2 Proof of Proposition 2

The identified set $\widetilde{\mathbf{P}}^I$ defined in (16) is sharp by construction because equations (6), (7), and (8) contain all model restrictions, and equation (15) fully characterizes $\widetilde{p}$ as an (implicit) function of π (see the arguments in footnotes 10 and 17 in the main text, and the proof of Proposition 3).

A.3 Proof of Proposition 3

Clearly, $\widetilde{\mathbf{P}}^I$ is empty whenever Π^I is empty. Assume hereafter that Π^I is non-empty. Recall that the identified set is characterized by the equations (6), (7), (8), and (15). By combining (6) and (7), we get

$$\left(R^{eq}_{-J} M_{-J} + R^{eq}_{J}\right) \pi_J = r^{eq} - R^{eq}_{-J} b_{-J}\left(p\right),$$

*Affiliations: Myrto Kalouptsidi, Harvard University, CEPR and NBER; Yuichi Kitamura, Yale University and Cowles Foundation for Research in Economics; Lucas Lima, Harvard University; Eduardo Souza-Rodrigues, University of Toronto.

which is of the form:

$$Q^{eq}\pi_J = q^{eq}, \tag{A1}$$

where $Q^{eq} = R^{eq}_{-J}M_{-J} + R^{eq}_J$ is a $d \times X$ matrix (defined in equation (18)), and $q^{eq} = r^{eq} - R^{eq}_{-J}b_{-J}(p) \in \mathbb{R}^d$. Equation (A1) incorporates all equality restrictions on π, and expresses them in terms of the "free parameter" $\pi_J \in \mathbb{R}^X$.

The set of solutions to the system (A1) can be represented by

$$\pi_J = Q^{eq\prime}\left(Q^{eq}Q^{eq\prime}\right)^{-1}q^{eq} + (I - P_Q)z, \tag{A2}$$

where $P_Q = Q^{eq\prime}\left(Q^{eq}Q^{eq\prime}\right)^{-1}Q^{eq}$, and the vector $z \in \mathbb{R}^X$ parameterizes the set of solutions. Represent the elements of this set by $\pi_J(z)$. Note that in the absence of the equality restrictions (7), we can just take $\pi_J = z$.[1]

Similarly, combine (6) and (8), to get

$$\left(R^{iq}_{-J}M_{-J} + R^{iq}_J\right)\pi_J \leq r^{iq} - R^{iq}_{-J}b_{-J}(p),$$

which is of the form:

$$Q^{iq}\pi_J \leq q^{iq},$$

where $Q^{iq} = R^{iq}_{-J}M_{-J} + R^{iq}_J$ is an $m \times X$ matrix, and $q^{iq} = r^{iq} - R^{iq}_{-J}b_{-J}(p) \in \mathbb{R}^m$. Substituting π_J in the inequality above by $\pi_J(z)$ defined in (A2) and rearranging, we get the m inequalities defined in terms of $z \in \mathbb{R}^X$:

$$Q^{iq}(I - P_Q)z \leq q^{iq} - Q^{iq}Q^{eq\prime}\left(Q^{eq}Q^{eq\prime}\right)^{-1}q^{eq}. \tag{A3}$$

Define the set

$$\mathcal{Z} = \left\{z \in \mathbb{R}^X : Q^{iq}(I - P_Q)z \leq q^{iq} - Q^{iq}Q^{eq\prime}\left(Q^{eq}Q^{eq\prime}\right)^{-1}q^{eq}\right\}. \tag{A4}$$

Clearly, $\mathcal{Z}$ is a convex polyhedron. By construction, any vector $\pi = [\pi'_{-J}, \pi'_J]'$ such that $\pi_{-J} = M_{-J}\pi_J(z) + b_{-J}$, with $\pi_J(z)$ defined by (A2) for some $z \in \mathcal{Z}$ satisfies (6), (7), and (8). I.e., for any given $z \in \mathcal{Z}$, we can find one π satisfying all model restrictions.

Next, combine (6) and (15) to obtain

$$\underbrace{[I, -\widetilde{M}_{-J}]}_{=\widetilde{\mathbf{M}}}\underbrace{\begin{bmatrix}\mathcal{H}_{1,-J} & \mathcal{H}_{1J}\\ \mathcal{H}_{2,-J} & \mathcal{H}_{2J}\end{bmatrix}}_{=\mathcal{H}}\underbrace{\begin{bmatrix} M_{-J}\pi_J + b_{-J}(p)\\ \pi_J\end{bmatrix}}_{=\pi} = \widetilde{b}_{-J}(\widetilde{p}) - \widetilde{\mathbf{M}}g,$$

[1] If the restrictions (7) suffice to point-identify the model, then Q^{eq} is invertible, $\pi_J = (Q^{eq})^{-1}q^{eq}$, and the remaining π_a, for $a \neq J$, can be recovered from (2). In this case, we can also take $\pi_J = z$.

or,

$$C\pi_J + \left(\mathcal{H}_{1,-J} - M_{-J}\mathcal{H}_{2,-J}\right) b_{-J}(p) = \widetilde{b}_{-J}\left(\widetilde{p}\right) - g_{-J} + \widetilde{M}_{-J}g_J, \tag{A5}$$

where C is the $\widetilde{A}\widetilde{X} \times X$ matrix defined in equation (17).

Noting that $\widetilde{p}$ has to satisfy $\widetilde{X}$ restrictions as it is a collection of conditional probability vectors, let $\widetilde{p}^*$ denote a $\widetilde{A}\widetilde{X}$-vector of independent elements of $\widetilde{p}$, and denote the set of independent elements by $\widetilde{\mathbf{P}}^*$. Substitute (A2) into (A5), and define the function $\mathbf{F} : \mathbb{R}^X \times int(\widetilde{\mathbf{P}}^*) \to \mathbb{R}^{\widetilde{A}\widetilde{X}}$ given by

$$\mathbf{F}\left(z, \widetilde{p}^*\right) = -C\pi_J(z) + \widetilde{b}_{-J}\left(\widetilde{p}^*\right) - \left(\mathcal{H}_{1,-J} - M_{-J}\mathcal{H}_{2,-J}\right) b_{-J}(p) - g_{-J} + \widetilde{M}_{-J}g_J,$$

or, more explicitly,

$$\begin{aligned}
\mathbf{F}\left(z, \widetilde{p}^*\right) &= -C(I - P_Q)z + \widetilde{b}_{-J}\left(\widetilde{p}^*\right) \\
&\quad - \left(\mathcal{H}_{1,-J} - M_{-J}\mathcal{H}_{2,-J}\right) b_{-J}(p) \\
&\quad - g_{-J} + \widetilde{M}_{-J}g_J - CQ^{eq\prime}\left(Q^{eq}Q^{eq\prime}\right)^{-1} q^{eq},
\end{aligned}$$

where $int(\widetilde{\mathbf{P}}^*)$ is the interior of the conditional probability simplex $\widetilde{\mathbf{P}}^*$. Clearly, the model and counterfactual restrictions impose $\mathbf{F}\left(z, \widetilde{p}^*\right) = 0$, for all $z \in \mathcal{Z}$.

The Jacobian of $\mathbf{F}$ is given by $\nabla\mathbf{F} = \left[\frac{\partial\mathbf{F}}{\partial z}, \frac{\partial\mathbf{F}}{\partial\widetilde{p}^*}\right]$, with

$$\begin{aligned}
\frac{\partial\mathbf{F}}{\partial z} &= -C(I - P_Q), \\
\frac{\partial\mathbf{F}}{\partial\widetilde{p}^*} &= \frac{\partial\widetilde{b}_{-J}}{\partial\widetilde{p}^*}.
\end{aligned}$$

Because $\frac{\partial\widetilde{b}_{-J}}{\partial\widetilde{p}^*}$ is everywhere invertible (see KSS), the implicit function theorem applies. Specifically, for a point $(z^0, \widetilde{p}^{*0}) \in \mathbb{R}^X \times int(\widetilde{\mathbf{P}}^*)$ satisfying $\mathbf{F}\left(z^0, \widetilde{p}^{*0}\right) = 0$, there exist open sets $U \subseteq \mathbb{R}^X$ and $W \subseteq int(\widetilde{\mathbf{P}}^*)$ such that $z^0 \in U$ and $\widetilde{p}^{*0} \in W$, and there exists a continuously differentiable function $\varphi : U \to W$ satisfying $\widetilde{p}^{*0} = \varphi(z^0)$ and that

$$\mathbf{F}\left(z, \varphi\left(z\right)\right) = 0,$$

for all $z \in U$. Furthermore,

$$\frac{\partial\varphi\left(z\right)}{\partial z} = -\left[\frac{\partial\mathbf{F}}{\partial\widetilde{p}^*}\right]^{-1}\frac{\partial\mathbf{F}}{\partial z} = \left[\frac{\partial\widetilde{b}_{-J}}{\partial\widetilde{p}^*}\right]^{-1} C(I - P_Q).$$

The rank of the matrix $\frac{\partial\varphi(z)}{\partial z}$ equals the rank of $C(I - P_Q)$ because $\frac{\partial\widetilde{b}_{-J}}{\partial\widetilde{p}^*}$ is invertible everywhere. Let $rank(C(I - P_Q)) = k$. By the Rank Theorem, the image set of φ is a differentiable k-dimensional manifold in $int(\widetilde{\mathbf{P}}^*)$ (see Theorem 3.5.1 in Krantz and Parks, 2003). Clearly, by restricting z to the convex

polyhedron $\mathcal{Z}$, the image set of φ becomes a k-dimensional manifold with boundary. In the absence of the model restrictions (7), we have $\pi_J = z$ and so the image set of φ becomes a manifold with dimension that equals the rank of C.

We can construct a global function $\bar{\varphi}$ defined on the entire domain $\mathcal{Z}$ based on the local function φ defined above. To do so, we need to show that the constructed $\bar{\varphi}$ is not a set-function on $\mathcal{Z}$. I.e., if for any pair of points $(z^0, \widetilde{p}^{*0})$ and $(z^0, \widetilde{p}^{*1})$ with $z^0 \in \mathcal{Z}$ and $\widetilde{p}^{*0}, \widetilde{p}^{*1} \in int(\widetilde{\mathbf{P}}^*)$, if $\bar{\varphi}(z^0) = \widetilde{p}^{*0}$ and $\bar{\varphi}(z^0) = \widetilde{p}^{*1}$, then we must have $\widetilde{p}^{*0} = \widetilde{p}^{*1}$. Suppose by contradiction that there there exist implicit functions φ^0 and φ^1 defined locally on the neighborhood of the points $(z^0, \widetilde{p}^{*0})$ and $(z^0, \widetilde{p}^{*1})$ such that $\widetilde{p}^{*0} = \varphi^0\left(z^0\right)$ and $\widetilde{p}^{*1} = \varphi^1\left(z^0\right)$, with $\widetilde{p}^{*0} \neq \widetilde{p}^{*1}$. Next, recall that for any point $z^0 \in \mathcal{Z}$, there exists only one vector of payoffs $\pi(z^0) = [\pi'_{-J}(z^0), \pi'_J(z^0)]'$ satisfying all model restrictions: This vector is given by the elements $\pi_{-J}(z^0) = M_{-J}\pi_J(z^0) + b_{-J}$, and $\pi_J(z^0)$ defined by (A2). This leads to the counterfactual payoff $\widetilde{\pi}(z^0)$, which is given by the affine function $\widetilde{\pi}(z^0) = \mathcal{H}\pi(z^0) + g$. Finally, the counterfactual payoff $\widetilde{\pi}(z^0)$ can generate just one conditional choice probability function in the counterfactual scenario (by the uniqueness of the solution of the Bellman equation). We therefore must have $\widetilde{p}^{*0} = \widetilde{p}^{*1}$ (as well as $\varphi^0 = \varphi^1 = \varphi$). The global function $\bar{\varphi}$ equals the local implicit functions everywhere.[2]

We conclude that the identified set $\widetilde{\mathbf{P}}^I$ is the image set of the global function $\bar{\varphi}$, defined on the domain $\mathcal{Z}$. Consequently, $\widetilde{\mathbf{P}}^I$ is a manifold with boundary and with dimension in the interior given by the rank of $C(I - P_Q)$. Further, $\widetilde{\mathbf{P}}^I$ is connected because $\bar{\varphi}$ is a continuous function defined on the convex domain $\mathcal{Z}$. In addition, when Π^I is bounded, so is the closed set $\mathcal{Z}$, which implies that $\bar{\varphi}(\mathcal{Z})$ is compact. Finally, we have $rank(C(I - P_Q)) \leq X - d$ because $rank(C) \leq \min\{\widetilde{A}\widetilde{X}, X\}$ and $rank(I - P_Q) = X - d$ (given that P_Q is symmetric and idempotent).

A.4 Proof of Proposition 4

The identified set $\mathbf{\Theta}^I$ defined in (19) is sharp by construction. We can construct payoff vectors satisfying all model restrictions, denoted by $\pi(z)$, and obtain the counterfactual CCP from the function $\widetilde{p}^* = \bar{\varphi}(z)$, where $\bar{\varphi}$ is continuously differentiable, $z \in \mathcal{Z}$, and $\mathcal{Z}$ is defined in (A4), as explained in the proof of Proposition 3. We have therefore

$$\theta = f(\widetilde{p}, \pi) = f(\bar{\varphi}(z), \pi(z)) = \bar{f}(z),$$

where we omit (p, F) from the notation for simplicity. When the function f is continuous, so is the function $\bar{f}$ because $\bar{\varphi}(z)$ and $\pi(z)$ are both continuous. Clearly, $\mathbf{\Theta}^I$ equals the image set of the function $\bar{f}$ defined on the domain $\mathcal{Z}$. The image set is connected because $\mathcal{Z}$ is convex, and it becomes compact when $\mathcal{Z}$ is compact (which happens when Π^I is bounded, see the proof of Proposition 3). Furthermore,

[2]While different z's can generate the same $\widetilde{p}^*$ (because the function φ is not one-to-one, which is at the heart of the identification problem of dynamic discrete choice models), a single z cannot generate more than one $\widetilde{p}^*$.

when θ is a scalar, the connected set $\mathbf{\Theta}^I$ becomes an interval.

A.5 Proof of Theorem 1

Consider a sequence $\{\mathfrak{p}_N \in \mathcal{P}_{\theta_0}, N \in \mathbb{N}\}$. Recall that p and F are determined by $\mathfrak{p}$. Let $(p_N, F_N) := (p(\mathfrak{p}_N), F(\mathfrak{p}_N))$. In what follows we prove the claim of the theorem for a fixed value $\theta_0 \in \mathbf{\Theta}^I$, and in the course of it we use symbols such as S_N, $\widehat{S}_N$, $\bar{\Pi}_N$, $(\bar{V}_N, V_N, v)$, $(\bar{W}_N, W_N, w)$, B, μ_N, $\binom{\phi}{\psi}$ and Σ (and their appropriate subsample counterparts with an asterisk symbol * in superscript) while omitting their dependence on θ_0 to ease the notational burden in the proof.

Let $S_N := S(\mathfrak{p}_N, \theta_0)$ and $\widehat{S}_N := S(\widehat{\mathfrak{p}}_N, \theta_0)$. Then writing $\|x\|_\Omega^2 := x'\Omega x$ for $x \in \mathbb{R}^{AX}$,

$$\begin{aligned} N\widehat{J}_N(\theta_0) &= \min_{x \in \widehat{S}_N} N\|b_{-J}(\widehat{\mathfrak{p}}_N) - x\|_{\widehat{\Omega}_N}^2 \\ &= \min_{x \in \widehat{S}_N} \|\sqrt{N}[b_{-J}(\widehat{\mathfrak{p}}_N) - b_{-J}(\mathfrak{p}_N)] - \sqrt{N}[x - b_{-J}(\mathfrak{p}_N)]\|_{\widehat{\Omega}_N}^2 \\ &= \min_{\xi \in \sqrt{N}(\widehat{S}_N \ominus b_{-J}(\mathfrak{p}_N))} \|\sqrt{N}[b_{-J}(\widehat{\mathfrak{p}}_N) - b_{-J}(\mathfrak{p}_N)] - \xi\|_{\widehat{\Omega}_N}^2, \end{aligned} \tag{A6}$$

where $\ominus$ denotes the usual Minkowski difference, and for $c \in \mathbb{R}_{++}$ and a set $A \in \mathbb{R}^d$, we let cA denote the set A dilated by the factor c, that is, $\{cx : x \in A\}$.

To show the theorem it suffices to consider sequences $\mathfrak{p}_N, N \in \mathbb{N}$ such that

(i) $\inf_{x \in \mathrm{bdy}(S_N)} \|b_{-J}(\mathfrak{p}_N) - x\|_\Omega = O(1/\sqrt{N})$, where $\mathrm{bdy}(S_N)$ is the boundary of S_N, and

(ii) Each sequence $\{\mathfrak{p}_N, N = 1, 2, ...\}$ converges.

Suppose $\mathfrak{p}_N, N \in \mathbb{N}$ satisfies (i) and (ii). The restrictions imposed on $\mathcal{P}_{\theta_0}$ guarantee that along the sequence $\mathfrak{p}_N$ it holds that

$$\sqrt{N}[b_{-J}(\widehat{\mathfrak{p}}_N) - b_{-J}(\mathfrak{p}_N)] \xrightarrow{d} \phi,$$

where ϕ is a zero mean Gaussian vector. In what follows we also use the following notation: for finite sets $V, W \subseteq \mathbb{R}^d$ we let $\mathrm{conv}(V)$ and $\mathrm{cone}(W)$ denote the convex hull of V and the cone spanned by W, respectively; then the Minkowski sum $\mathrm{conv}(V) \oplus \mathrm{cone}(W)$ is a polyhedron. We approximate the last term in equation (A6) following Chernoff (1954). Under Conditions 1 and 2 we have:

$$N\widehat{J}_N(\theta_0) \overset{d}{=} \min_{\xi \in \bar{\Pi}_N} \|\phi - \xi\|_\Omega^2 + o_p(1), \tag{A7}$$

where $\bar{\Pi}_N = \mathrm{conv}(\bar{V}_N) \oplus \mathrm{cone}(\bar{W}_N)$ is a random polyhedron, with $\bar{V}_N = V_N + v$, $V_N \in \mathbb{R}^{AX \times m}$, $\bar{W}_N = W_N + w$, $W_N \in \mathbb{R}^{AX \times n}$, and v and w are $\mathbb{R}^{AX \times m}$-valued and $\mathbb{R}^{AX \times n}$-valued zero-mean Gaussian random matrices, respectively, for some $m, n \in \mathbb{N}$. Note that the estimation uncertainty in $\widehat{S}_N$ makes the polyhedron $\bar{\Pi}_N$ that appears in the asymptotic approximation (A7) random. Also define a (deterministic) sequence of polyhedra $\Pi_N = \mathrm{conv}(V_N) \oplus \mathrm{cone}(W_N)$. By the representation theorem for polyhedra (see,

for example, Theorem 1.2 in Ziegler (2012)) we can write

$$\Pi_N = \{\xi : B\xi \leq \mu_N\} \text{ for some } B \in \mathbb{R}^{\ell \times AX},$$

where $\mu_N \geq 0$ for all N and $\mu_N = O(1)$.

Recalling that each transition matrix $F_a, a \in \mathcal{A}$ depends on $\mathfrak{p}_N$ (as so does F), write

$$\det(M_a(\mathfrak{p}_N)) = \det\left((1 - \beta F_a(\mathfrak{p}_N))(1 - \beta F_J(\mathfrak{p}_N))^{-1}\right) = \frac{\det(I - \beta F_a(\mathfrak{p}_N))}{\det(I - \beta F_J(\mathfrak{p}_N))}.$$

Let $\{\lambda_a^i(\mathfrak{p}_N)\}$ and $\{\lambda_J^i(\mathfrak{p}_N)\}$ be the eigenvalues of $F_a(\mathfrak{p}_N)$ and $F_J(\mathfrak{p}_N)$, then

$$\begin{aligned} \det(M_a(\mathfrak{p}_N)) &= \frac{\det(\beta^{-1} I - F_a(\mathfrak{p}_N))}{\det(\beta^{-1} I - F_J(\mathfrak{p}_N))} \\ &= \frac{\prod_{i=1}^X (\beta^{-1} - \lambda_a^i(\mathfrak{p}_N))}{\prod_{i=1}^X (\beta^{-1} - \lambda_J^i(\mathfrak{p}_N))} \\ &> c \qquad \text{for every } a \in \mathcal{A} \text{ and every } N \in \mathbb{N} \end{aligned} \tag{A8}$$

holds for some $c > 0$ that does not depend on N as β is fixed in the unit interval $(0, 1)$ and $\{\lambda_a^i(\mathfrak{p}_N)\}$ and $\{\lambda_J^i(\mathfrak{p}_N)\}$ are inside the unit circle for every N.

Note that the approximation (A7) holds for any sequence $\{V_N', W_N'\}_{N \in \mathbb{N}}$ such that $V_N' = V_N + o(1)$ and $W_N' = W_N + o(1)$, and with Condition 1 and (A8) we can choose $\{V_N, W_N\}_{N \in \mathbb{N}}$ such that the matrix B above does not depend on N. Then we have an alternative representation for the random polyhedron $\bar{\Pi}_N$ as well: for some positive definite matrix Σ it holds that

$$\bar{\Pi}_N = \{\xi : B\xi \leq \mu_N + \psi\},$$

where the vector $\binom{\phi}{\psi} \sim \mathrm{N}(0, \Sigma)$. In sum, we have

$$N\widehat{J}_N(\theta_0) \overset{d}{=} \min_{\xi : B\xi \leq \mu_N + \psi} \|\phi - \xi\|_\Omega^2 + o_p(1). \tag{A9}$$

Next we turn to the subsample statistic $\widehat{J}_{h_N}^*(\theta_0)$. To show the uniform validity of subsampling we can instead analyze the asymptotic behavior of the statistic $\widehat{J}_{h_N}$, the $\widehat{J}$-statistic calculated from a random sample of size h_N, drawn according to $\mathfrak{p}_N$ (Romano and Shaikh, 2012). That is, we now study the limiting behavior of the CDF $G_{h_N}(x, \mathfrak{p}_N), N = 1, 2, ...$, where $G_\ell(x, \mathfrak{p}) := \Pr_{\mathfrak{p}}\{\ell \widehat{J}_\ell(\theta_0) \leq x\}$ for $\ell \in \mathbb{N}$. Then proceeding as before, along the sequence $\mathfrak{p}_N$ we have

$$h_N \widehat{J}_{h_N}(\theta_0) \overset{d}{=} \min_{\xi \in \bar{\Pi}_{h_N, N}^*} \|\phi^* - \xi\|_\Omega^2 + o_p(1) \tag{A10}$$

where $\bar{\Pi}^*_{h_N,N} = \text{conv}(\bar{V}^*_{h_N,N}) \oplus \text{cone}(\bar{W}^*_{h_N,N})$, with $\bar{V}^*_{h_N,N} = V^*_{h_N,N} + v^*$, $V^*_{h_N,N} \in \mathbb{R}^{AX\times m}$, $\bar{W}^*_{h_N,N} = W^*_{h_N,N} + w^*$, $W^*_{h_N,N} \in \mathbb{R}^{AX\times m}$, and ϕ^*, v^* and w^* are zero-mean Gaussian random elements taking values in $\mathbb{R}^{AX}$, $\mathbb{R}^{AX\times m}$ and $\mathbb{R}^{AX\times n}$ with $(\phi^*, v^*, w^*) \overset{d}{=} (\phi, v, w)$. Define $\Pi^*_{h_N,N} = \text{conv}(V^*_{h_N,N}) \oplus \text{cone}(W^*_{h_N,N})$ and observe that it has a half-space based representation $\Pi^*_{h_N,N} = \{\xi : B\xi \leq \sqrt{\frac{h_N}{N}}\mu_N\}$. We now have

$$\bar{\Pi}^*_N = \left\{\xi : B\xi \leq \sqrt{\frac{h_N}{N}}\mu_N + \psi^*\right\}.$$

Recall that $\mu_N = \text{O}(1)$, and moreover, we have $\binom{\phi^*}{\psi^*} \sim \text{N}(0, \Sigma)$. Therefore

$$h_N \widehat{J}_{h_N}(\theta_0) \xrightarrow{d} \min_{\xi: B\xi \leq \psi} \|\phi - \xi\|^2_\Omega. \tag{A11}$$

In sum, for every sequence $\mathfrak{p}_N, N \in \mathbb{N}$ satisfying conditions (i) and (ii) above, by (A9) and (A11) and noting $\mu_N \geq 0$ for every N, we have

$$\limsup_{N\to\infty} \sup_x (G_{h_N}(x, \mathfrak{p}_N) - G_N(x, \mathfrak{p}_N)) \leq 0.$$

We can now invoke Theorem 2.1 in Romano and Shaikh (2012) to conclude.

B Firm Dynamic Entry/Exit Model

We now provide explicit formulas for the main equations and outcomes of interest presented in the paper in the context of the firm entry/exit model. By revisiting the numerical example shown in the main text we focus on the role that each individual model restriction plays in shaping the payoff identified set Π^I.

In the example, the transition matrix of the state variables $x = (k, w)$ becomes $F_a = F^k_a \otimes F^w$, where F^k_a is the 2×2 transition matrix for k, with (l, j) elements $\Pr[k_{it+1} = j | a_{it} = l, k_{it}]$ that equal one when $j = l$, and equal zero otherwise; and $\otimes$ is the Kronecker product. Specifically,

$$F_0 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \otimes F^w = \begin{bmatrix} F^w & 0 \\ F^w & 0 \end{bmatrix}, \; F_1 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \otimes F^w = \begin{bmatrix} 0 & F^w \\ 0 & F^w \end{bmatrix}. \tag{B1}$$

The payoff vectors are the same as in (10) in the main paper and are rewritten below for convenience,

$$\pi_0 = \begin{bmatrix} \overline{\pi}_0 \\ s \end{bmatrix}, \pi_1 = \begin{bmatrix} vp - fc - ec \\ vp - fc \end{bmatrix}.$$

The vector of CCPs is composed of $p_a(k, w)$. To simplify notation, we let $p_a(k)$ be a vector of dimension W (i.e., we fix k and run over w) so that $p = \left(p'_0(0), p'_1(0), p'_0(1), p'_1(1)\right)'$.

Consider the main equality constraint resulting from the DDC framework and take $J = 0$ (i.e., equation

(2) presented in the main text)

$$\pi_1 = M_1\pi_0 + b_1(p). \tag{B2}$$

This equation indicates that $X = KW = 2W$ parameters need to be specified for point-identification. Thus, if π_0 is known, then π_1 is recovered. Indeed, let us first compute M_1, defined in (3). Here, we have

$$M_1 = \begin{bmatrix} I & -\beta F^w \\ 0 & I - \beta F^w \end{bmatrix} \begin{bmatrix} I - \beta F^w & 0 \\ -\beta F^w & I \end{bmatrix}^{-1},$$

where the inverse in the above expression is easily verified to be

$$\begin{bmatrix} (I - \beta F^w)^{-1} & 0 \\ (I - \beta F^w)^{-1} \beta F^w & I \end{bmatrix}$$

and therefore,

$$M_1 = \begin{bmatrix} I + \beta F^w & -\beta F^w \\ \beta F^w & I - \beta F^w \end{bmatrix}.$$

Next, note that in the logit model, $b_1(p) = M_1\psi_0(p) - \psi_1(p)$ becomes (see equation (4)):

$$b_1(p) = \begin{bmatrix} \ln p_1(0) \\ \ln p_1(1) \end{bmatrix} - \begin{bmatrix} I + \beta F^w & -\beta F^w \\ \beta F^w & I - \beta F^w \end{bmatrix} \begin{bmatrix} \ln p_0(0) \\ \ln p_0(1) \end{bmatrix},$$

given that $\psi_a(p(x)) = \kappa - \ln p_a(x)$, where κ is the Euler constant. Thus equation (B2) becomes

$$\begin{bmatrix} vp - fc - ec \\ vp - fc \end{bmatrix} = \begin{bmatrix} I + \beta F^w & -\beta F^w \\ \beta F^w & I - \beta F^w \end{bmatrix} \begin{bmatrix} \bar{\pi}_0 \\ s \end{bmatrix} + b_1(p). \tag{B3}$$

Note now that if π_0 is known, namely both the scrap vector s and $\bar{\pi}_0$ are given, they suffice to identify π_1, but they do not suffice to separate the $3W$ parameters, vp, fc, and ec. Suppose in addition that vp is known. Then, we rewrite π_1 separating the unknowns ec and fc:

$$\pi_1 = \begin{bmatrix} -I_2 & -I_2 \\ 0 & -I_2 \end{bmatrix} \begin{bmatrix} ec \\ fc \end{bmatrix} + \begin{bmatrix} vp \\ vp \end{bmatrix},$$

where I_2 is the 2×2 identity matrix.

We want to find an explicit relation between ec, fc, and s. First, we invert the equation above to obtain the unknowns ec and fc:

$$\begin{bmatrix} ec \\ fc \end{bmatrix} = \begin{bmatrix} -I_2 & I_2 \\ 0 & -I_2 \end{bmatrix} \pi_1 - \begin{bmatrix} -I_2 & I_2 \\ 0 & -I_2 \end{bmatrix} \begin{bmatrix} vp \\ vp \end{bmatrix} = \begin{bmatrix} -I_2 & I_2 \\ 0 & -I_2 \end{bmatrix} \pi_1 + \begin{bmatrix} 0 \\ vp \end{bmatrix} \tag{B4}$$

We next replace π_1 from our main equation to obtain

$$\begin{bmatrix} ec \\ fc \end{bmatrix} = \begin{bmatrix} -I_2 & I_2 \\ 0 & -I_2 \end{bmatrix} \left(\begin{bmatrix} I + \beta F^w & -\beta F^w \\ \beta F^w & I - \beta F^w \end{bmatrix} \begin{bmatrix} \bar{\pi}_0 \\ s \end{bmatrix} + b_1(p) \right) + \begin{bmatrix} 0 \\ vp \end{bmatrix}$$

or

$$\begin{bmatrix} ec \\ fc \end{bmatrix} = \begin{bmatrix} s - \bar{\pi}_0 \\ -\beta F^w \bar{\pi}_0 - (I - \beta F^w) s \end{bmatrix} + \begin{bmatrix} b_l(p) - b_u(p) \\ -b_l(p) + vp \end{bmatrix}, \tag{B5}$$

where the vectors $b_u(p)$ and $b_l(p)$ constitute the upper and lower parts of $b_1(p)$, that is $b_1(p) = [b_u'(p), b_l'(p)]'$.

In particular, if $\bar{\pi}_0 = 0$ the above becomes,

$$\begin{aligned} ec &= s + b_l(p) - b_u(p), \\ fc &= -(I - \beta F^w) s - b_l(p) + vp. \end{aligned}$$

Clearly, given any one of the three parameters ec, fc, s, the remaining two are uniquely determined.

These equations have an interesting interpretation. In the case of logit shocks, and assuming that $\bar{\pi}_0 = 0$, the first equation above becomes:

$$s - ec = \ln \frac{p_1(0)}{p_0(0)} - \ln \frac{p_1(1)}{p_0(1)}.$$

The difference between the scrap values and the entry cost is identified; the difference is given by the contrast between the odds of the probability of entry ($p_1(0)/p_0(0)$) and the odds of the probability of staying in the market ($p_1(1)/p_0(1)$). Intuitively, in the data, the larger the probability of entry relative to the probability of staying, the smaller the entry cost relative to the scrap value. (A similar interpretation relating scrap values and fixed costs holds for the second equation above.)

Model Restrictions. We now turn to the model restrictions. For ease of exposition, we focus on the restrictions presented in the main paper:

1. $\bar{\pi}_0 = 0, \quad fc \geq 0, \quad ec \geq 0$, and vp is known.
2. $vp - fc \leq ec \leq \frac{\mathbb{E}[vp - fc]}{1-\beta}$, and $\pi_1(1, w^h) \geq \pi_1(1, w^l)$.
3. s does not depend on w.

Restriction 1. Under equation (B5), $ec \geq 0$ and $fc \geq 0$ translate respectively to:

$$s \geq b_u(p) - b_l(p), \tag{B6}$$

$$(I - \beta F^w)\, s \leq vp - b_l\,(p)\,. \tag{B7}$$

Visualizing the set of inequalities (B6) is clear: the positive orthant is shifted to the point $b_u\,(p) - b_l\,(p)$.

The hyperplanes under (B7) intersect at a unique point because $(I - \beta F^w)$ is invertible. Suppose $W = 2$, then equation (B7) is written as the following two equations:

$$(1 - \beta f_1)\, s_1 - \beta\,(1 - f_1)\, s_2 \leq vp_1 - b_{l1}\,(p)$$

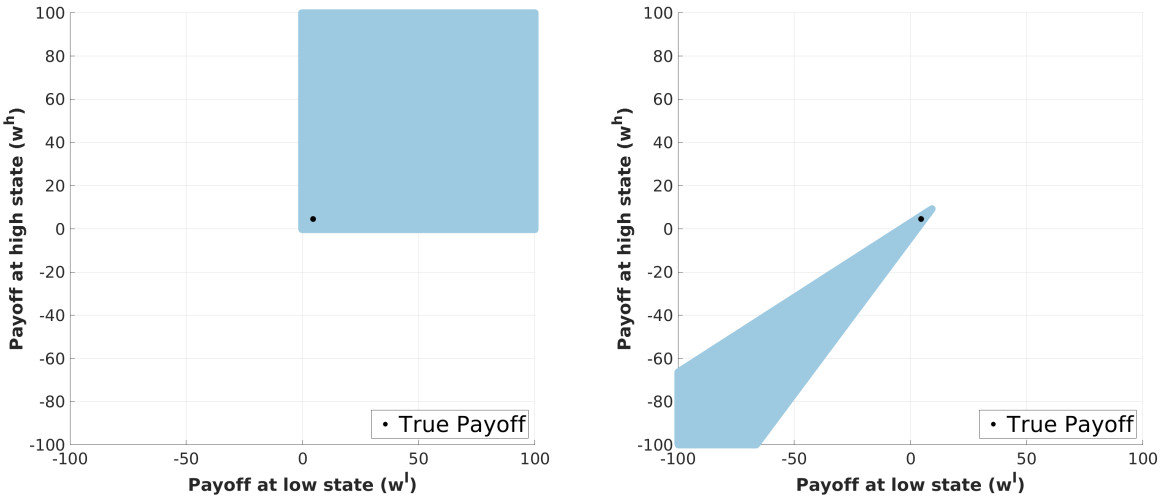
$$-\beta\,(1 - f_2)\, s_1 + (1 - \beta f_2)\, s_2 \leq vp_2 - b_{l2}\,(p)$$

where

$$F^w = \begin{bmatrix} f_1 & 1 - f_1 \\ 1 - f_2 & f_2 \end{bmatrix},$$

$s = [s_1, s_2]'$, and similarly for the vectors vp and $b_l(p)$. Both lines in the inequalities above have positive slope and are thus increasing.

Figure B1 presents the set of values that s can take for the parameter configuration used in the numerical example presented in Section 3 of the main paper. In the left panel, we present the set implied by $ec \geq 0$; on the right panel, the set implied by $fc \geq 0$. In both panels, the horizontal axis represents scrap values when the shock is low, w^l, and the vertical axis, scrap values when the shock is high, w^h. (For ease of exposition, we limit the values in the figures to be between -100 and 100.) The true s is represented by the black dots. Clearly, the larger polygon presented in panel (b) of Figure 1 in the main text combines all restrictions presented separately in Figure B1.



(a) Restriction: $ec \geq 0$ **(b)** Restriction: $fc \geq 0$

Figure B1: Payoff Identified Set Π^I: Scrap Values under Alternative Restrictions

Remark B1. In summary, given the reference action $J = 0$, the polytope

$$\Pi_J^I = \left\{ \pi_J \in \mathbb{R}^X : (R_{-J}^{eq} M_{-J} + R_J^{eq})\pi_J = r^{eq} - R_{-J}^{eq} b_{-J},\ (R_{-J}^{iq} M_{-J} + R_J^{iq})\pi_J \leq r^{iq} - R_{-J}^{iq} b_{-J} \right\}$$

is given by the $W-$dimensional polyhedral set

$$\left\{ (0, s) \in \mathbb{R}^{2W} : \text{such that } s \text{ satisfies equations (B6) and (B7)} \right\}.$$

Restriction 2. We first express the three sets of inequalities of Restriction 2 in terms of the payoffs π_0 and π_1. Condition $vp - fc \leq ec$ becomes

$$\pi_1(0) \leq 0. \tag{B8}$$

Next, we focus on $ec \leq \mathbb{E}\left[vp - fc\right] / (1 - \beta)$. Let q denote the stationary distribution of F^w, i.e. $q' F^w = q'$. Then, the inequality becomes

$$ec \leq \frac{1}{1-\beta} \mathbf{1} q' \left(vp - fc\right),$$

where $\mathbf{1}$ is a $W \times 1$ vector of ones. From the definition of π_1 we have that $ec = \pi_1(1) - \pi_1(0)$ and $vp - fc = \pi_1(1)$. Therefore, we get:

$$\pi_1(1) - \pi_1(0) \leq \frac{1}{1-\beta} \mathbf{1} q' \pi_1(1)$$

or

$$\left[-I_2,\ I_2 \frac{1}{1-\beta} \mathbf{1} q' \right] \pi_1 \leq 0. \tag{B9}$$

Finally, monotonicity in $\pi_1(1)$ means

$$[0\ 0\ \ 1\ \text{-}1]\, \pi_1 \leq 0. \tag{B10}$$

Now we stack (B8), (B9) and (B10), so that:

$$R_{-J}^{iq} \pi_{-J} = R_1^{iq} \pi_1 = \begin{bmatrix} I_2 & 0 \\ -I_2 & I_2 - \frac{1}{1-\beta} \mathbf{1} q' \\ 0 & [1\ \text{-}1] \end{bmatrix} \pi_1 \leq 0, \tag{B11}$$

and $R_J^{iq} = R_0^{iq} = 0$ and $r^{iq} = 0$. Moreover, multiplying R_1^{iq}, from (B11), with M_1 gives,

$$R_1^{iq} M_1 = \begin{bmatrix} I + \beta F^w & -\beta F^w \\ -\left(I_2 + \frac{\beta}{1-\beta} \mathbf{1} q' \right) & I_2 - \mathbf{1} q' \\ \beta\, [1\ \text{-}1]\, F^w & [1\ \text{-}1]\, (I - \beta F^w) \end{bmatrix}.$$

The scrap values are confined by the inequalities $(R_1^{iq}M_1 + R_0^{iq})\pi_0 \leq r^{iq} - R_1^{iq}b_1$ (see Remark B1 above), which implies

$$\begin{aligned} -\beta F^w s &\leq -b_u(p) \\ \left(I_2 - \mathbf{1}q'\right) s &\leq b_u(p) - b_l(p) + \frac{1}{1-\beta}\mathbf{1}q' b_l(p) \\ [1\ \text{-}1]\,(I - \beta F^w)\, s &\leq b_{l2}(p) - b_{l1}(p), \end{aligned}$$

or in more detail,

$$\begin{aligned} -\beta f_1 s_1 - \beta\,(1-f_1)\, s_2 &\leq b_{u1}(p) \\ -\beta\,(1-f_2)\, s_1 - \beta f_2 s_2 &\leq b_{u2}(p) \\ (1-q_1)\,(s_1 - s_2) &\leq b_{u1}(p) - b_{l1}(p) + \frac{1}{\beta}\left(q_1 b_{l1}(p) + (1-q_1)\, b_{l2}(p)\right) \\ -q_1\,(s_1 - s_2) &\leq b_{u2}(p) - b_{l2}(p) + \frac{1}{\beta}\left(q_1 b_{l1}(p) + (1-q_1)\, b_{l2}(p)\right) \\ \left(1 - \beta\,(f_1 + f_2 - 1)\right)(s_1 - s_2) &\leq b_{l1}(p) + b_{l2}(p), \end{aligned}$$

where $q = [q_1, 1-q_1]'$.

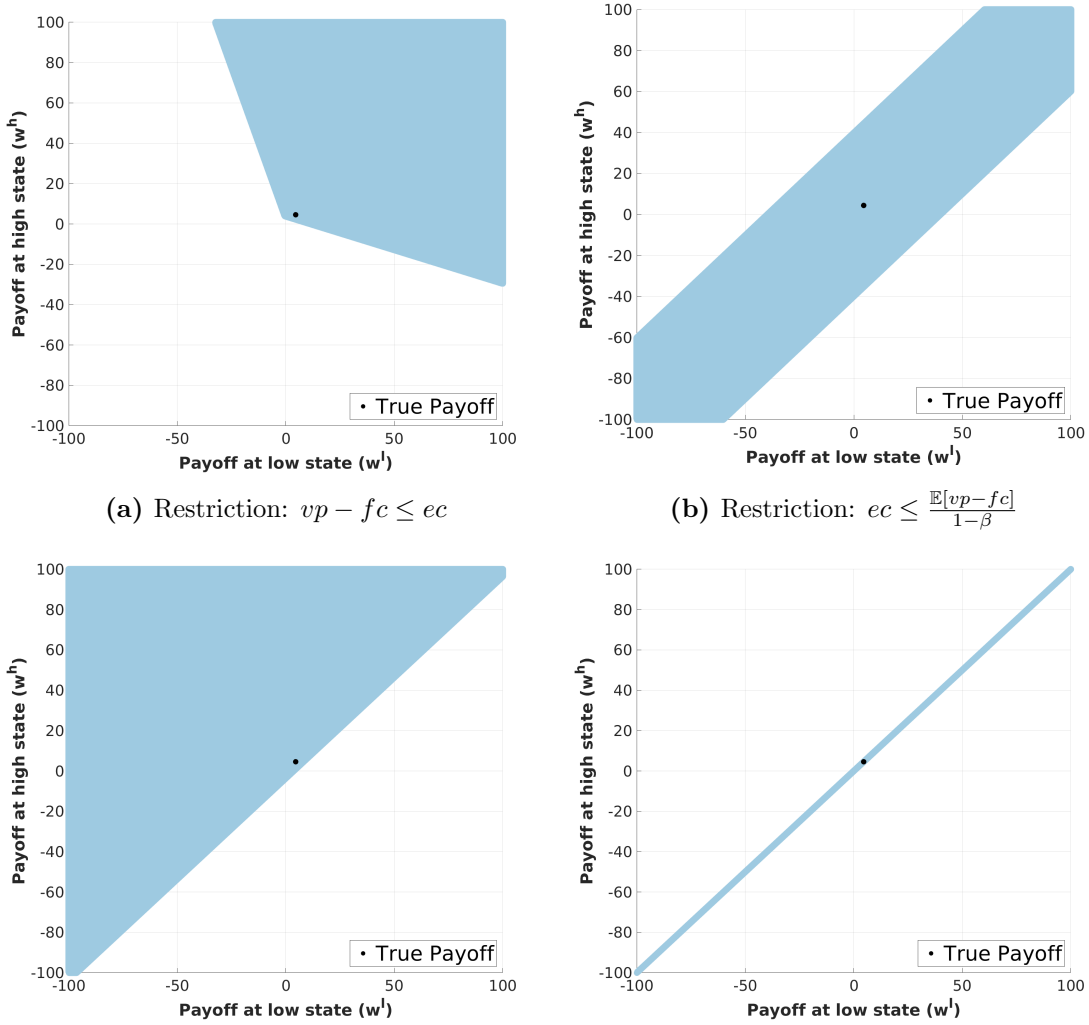
The first two inequalities correspond to the restriction $vp - fc \leq ec$. They imply lower bounds on scrap values. Note that these first two lines have negative slope and hence are decreasing. They have a unique intersection if $\det F^w \neq 0$ or $f_2 \neq 1 - f_1$.[3] The next two inequalities correspond to condition $ec \leq \mathbb{E}\,[vp - fc]\,/\,(1-\beta)$. They define a box constraining the difference $s_1 - s_2$. And the monotonicity in $\pi_1(1)$ assumption implies the fifth inequality above. That line has positive slope and so any point above that line satisfies the restriction.

Like Figure B1 above, Figure B2 shows the values of s for the parameter configuration presented in Section 3 of the main paper but under Restriction 2. Panel (a) shows the set under condition $vp - fc \leq ec$ (with the two downward slope lines); panel (b) presents the set under $ec \leq \mathbb{E}\,[vp - fc]\,/\,(1-\beta)$ (with $s_1 - s_2$ constrained in a box); and panel (c) shows the set under the monotonicity condition. Their intersection result in the light blue polygon presented in panel (b) of Figure 1 in the main text.

Restriction 3. If $s_1 = s_2 = s$, there is a single free parameter. This clearly results in a single line, presented in panel (d) of Figure B2. Combining Restrictions 1–3 result in the blue line inside the light blue polyhedron in panel (b) of Figure 1.

[3] If $\det F^w = 0$ then the two constraints collapse to the single constraint:

$$-\beta f_1 s_1 - \beta\,(1-f_1)\, s_2 \leq \min\left\{b_{u1}(p), b_{u2}(p)\right\}.$$



(a) Restriction: $vp - fc \leq ec$

(b) Restriction: $ec \leq \frac{\mathbb{E}[vp-fc]}{1-\beta}$

(c) Restriction: $\pi_1(1, w^h) \geq \pi_1(1, w^l)$

(d) Restriction: s does not depend on w

Figure B2: Payoff Identified Set Π^I: Scrap Values under Alternative Model Restrictions

Counterfactuals. In the firm example, we consider a counterfactual experiment that decreases entry cost by 20%, and holds everything else the same as in the baseline. This means we take $g = 0$ and $\mathcal{H}$ block-diagonal with diagonal blocks given by $\mathcal{H}_{00} = I$ and

$$\mathcal{H}_{11} = \begin{bmatrix} \tau I_2 & (1-\tau) I_2 \\ 0 & I_2 \end{bmatrix}.$$

Combining equations (5) and (15), we obtain

$$\widetilde{b}_1(\widetilde{p}) = C\pi_J + (\mathcal{H}_{11} - M_1\mathcal{H}_{21}) b_1(p) - g_1 + \widetilde{M}_1 g_0,$$

where C is defined in equation (17), and $g = [g_0', g_1']'$. Since $g = 0$ and $\mathcal{H}_{21} = 0$, the above becomes:

$$\widetilde{b}_1(\widetilde{p}) = C\pi_J + \mathcal{H}_{11} b_1(p). \tag{B12}$$

We next calculate C:

$$C = \mathcal{H}_{11}M_1 - M_1\mathcal{H}_{00} = \begin{bmatrix} \tau I_2 & (1-\tau) I_2 \\ 0 & I_2 \end{bmatrix} \begin{bmatrix} I_2 + \beta F^w & -\beta F^w \\ \beta F^w & I_2 - \beta F^w \end{bmatrix} - \begin{bmatrix} I_2 + \beta F^w & -\beta F^w \\ \beta F^w & I_2 - \beta F^w \end{bmatrix}$$
$$= \begin{bmatrix} (\tau - 1) I_2 & (1-\tau) I_2 \\ 0 & 0 \end{bmatrix}.$$

Clearly, $rank(C) = 2$. We thus conclude that even in the absence of any restrictions (e.g. $\pi_0(0) = \overline{\pi}_0 = 0$), the counterfactual CCPs live in a 2-dimensional manifold. (See Proposition 3.)

To see whether the restriction $\pi_0(0) = 0$ reduces the dimension of the identified set for the counterfactual CCPs, we need to verify the rank of $C(I - P_Q)$, where $P_Q = Q^{eq'}(Q^{eq}Q^{eq'})Q^{eq}$ (Proposition 3). Note that this restriction means that $Q^{eq} = [I_2\ 0]$ (see equation (18) in the main text defining the matrix Q^{eq}). But $Q^{eq}Q^{eq'} = [I_2\ 0] \begin{bmatrix} I_2 \\ 0 \end{bmatrix} = I_2$. Thus $P_Q = \begin{bmatrix} I_2 \\ 0 \end{bmatrix} [I_2\ 0] = \begin{bmatrix} I_2 & 0 \\ 0 & 0 \end{bmatrix}$ and $I - P_Q = \begin{bmatrix} 0 & 0 \\ 0 & I_2 \end{bmatrix}$. It follows that

$$C\left(I - P_Q\right) = \begin{bmatrix} 0 & (1-\tau) I_2 \\ 0 & 0 \end{bmatrix},$$

and $rank(C(I - P_Q)) = 2$. The added restriction does not alter the dimension of the counterfactual CCP, although it makes equation (2) (or (B5)) simpler.

Counterfactual Outcomes of Interest. In our example, we consider the long-run average impact of the entry subsidy τ on (i) the probability of staying in the market (labelled θ_P), (ii) the consumer surplus (θ_{CS}), and (iii) the value of the firm (θ_V).

Probability of Being Active. The long-run average effect on the probability of being active is given by

$$\theta_P = \mathbb{E}[\widetilde{p}_1(x)] - \mathbb{E}[p_1(x)],$$

where the expectations are taken with respect to the ergodic distributions of the state variables x in the counterfactual and baseline scenarios. Specifically,

$$\theta_P = \sum_{x \in \widetilde{\mathcal{X}}} \widetilde{p}_1(x)\, \widetilde{f}^*(x) - \sum_{x \in \mathcal{X}} p_1(x)\, f^*(x),$$

where $\widetilde{f}^{*}(x)$ is the ergodic distribution of the (endogenous) Markovian process

$$\widetilde{F}(x'|x) = \sum_{a \in \widetilde{\mathcal{A}}} \widetilde{F}(x'|x,a)\, \widetilde{p}_a(x),$$

and a similar expression holds for the baseline ergodic distribution $f^{*}(x)$.

When $x = (k, w) \in \mathcal{K} \times \mathcal{W}$, and k is the lagged action, the expression for θ_P simplifies. First, note that the probability of choosing action a at time period t conditioned on the exogenous states w is given by

$$\Pr(a_{it} = a|w_{it}) = \sum_{k \in \mathcal{K}} \Pr(a_{it} = a|k_{it} = k, w_{it}) \Pr(k_{it} = k|w_{it}),$$

which implies

$$\Pr(a_{it} = a|w_{it}) = \sum_{k \in \mathcal{K}} p_a(k, w_{it}) \Pr(a_{it-1} = k|w_{it}).$$

Define $p_a(w) \equiv \Pr(a_{it} = a|w)$. The steady state condition implies that the vector $[p_0(w), ..., p_A(w)]'$ satisfies the fixed-point:[4]

$$\begin{bmatrix} p_0(w) \\ \vdots \\ p_A(w) \end{bmatrix} = \begin{bmatrix} p_0(0,w) & \cdots & p_0(A,w) \\ \vdots & \ddots & \vdots \\ p_A(0,w) & \cdots & p_A(A,w) \end{bmatrix} \begin{bmatrix} p_0(w) \\ \vdots \\ p_A(w) \end{bmatrix}. \tag{B13}$$

Let $\widetilde{f}^{*}_{W}$ and f^{*}_{W} be the steady-state distributions of the exogenous variables in the counterfactual and baseline scenarios, respectively. Then

$$\theta_P = \mathbb{E}[\widetilde{p}_1(k,w)] - \mathbb{E}[p_1(k,w)] = \sum_{k,w} \widetilde{p}_1(k,w)\, \widetilde{f}^{*}(k|w) \widetilde{f}^{*}_{W}(w) - \sum_{k,w} p_1(k,w)\, f^{*}(k|w) f^{*}_{W}(w).$$

The inner sum in the first term equals $\widetilde{p}_1(w)$ due to (B13). A similar remark holds for the inner sum of the second term which becomes $p_1(w)$. Thus

$$\theta_P = \sum_{w \in \widetilde{\mathcal{W}}} \widetilde{p}_1(w)\, \widetilde{f}^{*}_{W}(w) - \sum_{w \in \mathcal{W}} p_1(w)\, f^{*}_{W}(w).$$

Consumer Surplus. The long-run average change on the consumer surplus is:

$$\theta_{CS} \;=\; \sum_{a \in \widetilde{\mathcal{A}}, x \in \widetilde{\mathcal{X}}} \widetilde{CS}(a,x)\, \widetilde{p}_a(x) \widetilde{f}^{*}(x) - \sum_{a \in \mathcal{A}, x \in \mathcal{X}} CS(a,x)\, p_a(x) f^{*}(x).$$

In the special case in which $x = (k, w)$, and k is the lagged action and w are exogenous shocks, we

[4]For instance, in the binary choice model, we have $\Pr(a = 1|w) = p_1(0,w)(1 - \Pr(a = 1|w)) + p_1(1,w)\Pr(a = 1|w)$, which implies $\Pr(a = 1|w) = p_1(0,w)/[1 - p_1(1,w) + p_1(0,w)]$.

compute the consumer surplus for each action and state, $CS(a,k,w)$, by assuming a (residual) linear inverse demand $P = w - \eta Q$, where P is the price and Q is the quantity demanded, and assuming a constant marginal cost mc. These imply that $CS(a,k,w) = 0$ when the firm is inactive ($a = 0$), and $CS(a,k,w) = (w - mc)^2/8\eta$ when it is active ($a = 1$). So,

$$\begin{aligned}\theta_{CS} &= \mathbb{E}[\widetilde{CS}(a,k,w) \times 1\{a=1\}] - \mathbb{E}[CS(a,k,w) \times 1\{a=1\}] \\ &= \sum_{w \in \widetilde{\mathcal{W}}} CS\left(w\right) \widetilde{p}_1(w) \widetilde{f}^*_W\left(w\right) - \sum_{w \in \mathcal{W}} CS(w) p_1(w) f^*_W\left(w\right).\end{aligned}$$

Note that the consumer surplus function is the same in the baseline and counterfactual scenarios (and so is the distribution of the exogenous states, $\widetilde{f}^*_W = f^*_W$). The *average* CS changes in the counterfactual because the firm changes its entry behavior when it receives an entry subsidy.

Value of the Firm. The value of the firm in the baseline is given by the $X \times 1$ vector

$$V = \left(I - \beta F_J\right)^{-1} \left(\pi_J + \psi_J(p)\right),$$

where we take $J = 0$ (see footnote 10 in the main text). A similar expression holds for the counterfactual value: $\widetilde{V} = (I - \widetilde{\beta}\widetilde{F}_J)^{-1}(\widetilde{\pi}_J + \widetilde{\psi}_J(\widetilde{p}))$. The long-run average change in the value of the firm is given by

$$\theta_V = \sum_{x \in \widetilde{\mathcal{X}}} \widetilde{V}\left(x\right) \widetilde{f}^*\left(x\right) - \sum_{x \in \mathcal{X}} V\left(x\right) f^*\left(x\right).$$

As before, let $\widetilde{f}^*$ and f^* denote the vector of steady-state distributions, then

$$\begin{aligned}\theta_V &= \widetilde{f}^{*\prime} \times (I - \widetilde{\beta}\widetilde{F}_J)^{-1}(\widetilde{\pi}_J + \widetilde{\psi}_J(\widetilde{p})) \\ &\quad - f^{*\prime} \times \left(I - \beta F_J\right)^{-1} \left(\pi_J + \psi_J(p)\right).\end{aligned}$$

The *average* firm value (across states) changes in the counterfactual both because the steady state distribution changes, and because the value of the firm is affected by the subsidy in *all* states.

Figure B3 presents the identified set for θ based on the parameter configuration of the firm entry/exit model in Section 3. As before, the larger set (including the dark blue area) depicts $\mathbf{\Theta}^I$ under Restriction 1, while the smaller set (in light blue) shows the identified set under Restrictions 1–2, and the blue line shows $\mathbf{\Theta}^I$ under Restrictions 1–3. The true θ is represented by the black dot.

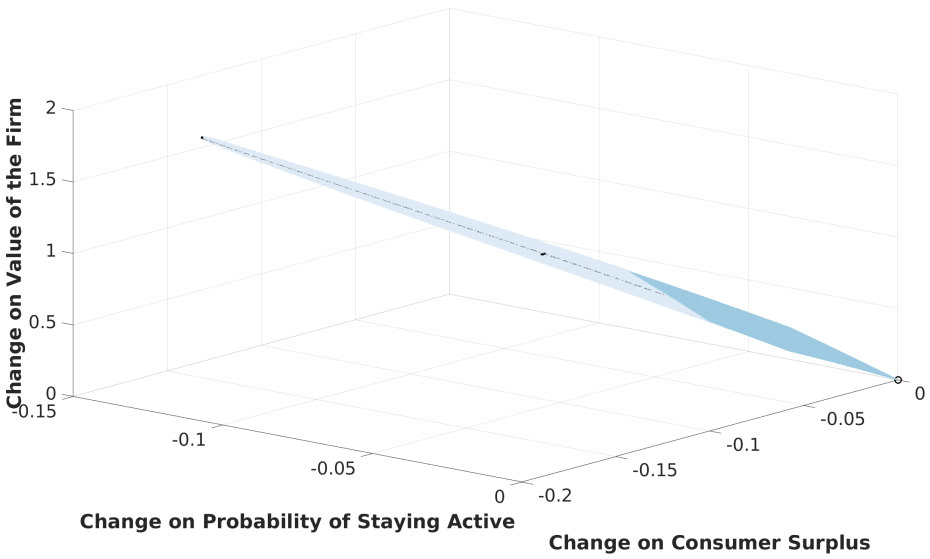


Figure B3: Identified Set $\boldsymbol{\Theta}^I$ under Restrictions 1–3

C A Monte Carlo Study

In this section, we present a Monte Carlo study to illustrate the finite-sample performance of our inference procedure. We start with the setup, and then we show the results.

C.1 Setup

We extend the firm entry/exit problem presented in Sections 3 and 4 of the main text, allowing now for a larger state space. Specifically, we assume the presence of three exogenous states, $w_t = (w_{1t}, w_{2t}, w_{3t})$, reflecting demand and supply shocks. The exogenous states are independent to each other, and each follows a discrete-AR(1) process with W support points (obtained by discretizing latent normally-distributed AR(1) processes). The (residual) inverse demand function is linear, $P_t = \bar{w} + w_{1t} + w_{2t} - \eta Q_t$, where P_t is the price of the product, Q_t is the quantity demanded, $\bar{w}$ is the intercept, w_{1t} and w_{2t} are demand shocks, and η is the slope. We assume constant marginal costs mc_t (i.e., mc_t does not depend on Q_t), and let the supply shocks w_{3t} affect marginal costs. To simplify, we just take $mc_t = w_{3t}$. Variable profits are then $vp_t = (\bar{w} + w_{1t} + w_{2t} - mc_t)^2/4\eta$. The idiosyncratic shocks ε follow the type 1 extreme value distribution. The model parameters are presented in Table C1.

The counterfactual we consider is the same as in the example in Section 4: a subsidy that reduces entry costs by 20%. The target parameter θ is the long-run average probability of staying in the market given the subsidy, where the long-run average is based on the ergodic distribution of the state variables; the specific formula for θ is provided in Section B (but note that here we do not take the difference between the counterfactual and the baseline average probabilities).

In order to analyze the sensitivity of the target parameter θ to alternative model restrictions, we follow

Table C1: Parameters of the Monte Carlo Data Generating Process

Demand Function:	$\bar{w}$	6.8	$w_{1t} \sim$ Normal AR(1):	ρ_{01}	0
	η	4		ρ_{11}	0.75
				σ_1^2	0.02
Payoff Parameters:	s	4.5	$w_{2t} \sim$ Normal AR(1):	ρ_{02}	0
	ec	5		ρ_{12}	0.75
	fc	0.5		σ_2^2	0.025
Scale parameter:	σ	1	$w_{3t} \sim$ Normal AR(1):	ρ_{03}	0
				ρ_{13}	0.75
Discount Factor:	β	0.9		σ_3^2	0.03

the example again and impose the three sets of restrictions:

1. $\overline{\pi}_0 = 0, \quad fc \geq 0, \quad ec \geq 0$, and vp is known.

2. $\pi_1(1, w^h) \geq \pi_1(1, w^l)$, and $vp - fc \leq ec \leq \frac{\mathbb{E}[vp-fc]}{1-\beta}$, where the expectation is taken over the ergodic distribution of the state variables.

3. s does not depend on w.[5]

We generate 1000 Monte Carlo replications for each of the following sample sizes: the small sample, with $N = 100$ firms on separated (independent) markets and $T = 5$ time periods, and the large sample, with $N = 1000$ firms and $T = 15$ time periods. For the first sample period, the value of the state variables are drawn from their steady-state distributions. Given that each exogenous state variable w_{jt} can take W values, the dimension of the state space is $X = 2 \times W^3$. We consider three sizes for the state space: $X = 16$, 54 and 250, which correspond to $W = 2$, 3 and 5. The choices of the state space were dictated by the sample size, not by computational constraints, given that the method makes use of a frequency, or a nonparametric estimator for the CCP in the first stage. (As discussed in Online Appendix H, it is feasible to solve the optimization problem (20)–(21) for state spaces that are larger in size.)

In each sample, we estimate the lower and upper bounds for the target parameter, θ^L and θ^U, by solving the minimization and maximization problems (20)–(21). We estimate CCPs using frequency estimators, and we use the true transition matrix F, both in calculating test statistics and critical values. (The results do not change significantly when we estimate transition probabilities as well.) We solve the problem (20)–(21) using the Knitro MATLAB function. We provide initial values for π by solving the

[5]When we impose Restriction 3, we replace the inequalities defined in Restriction 1 by their average versions. This does not affect the identified set, but it improves the finite-sample behavior of the estimators when the sample size is small and the state space is large.

following quadratic programming problem (with the GUROBI solver, using its MATLAB interface):

$$\min_{\substack{\pi\in\mathbb{R}^{(A+1)X}:\\ R^{eq}\pi=r^{eq},R^{iq}\pi\leq r^{iq}}} [b_{-J}(\widehat{\mathfrak{p}}_N) - \widehat{\mathbf{M}}_N\,\pi]'\,\widehat{\Omega}_N\,[b_{-J}(\widehat{\mathfrak{p}}_N) - \widehat{\mathbf{M}}_N\,\pi].$$

We specify $\widehat{\Omega}_N$ to be a diagonal matrix with diagonal terms given by the square-root of the ergodic distribution of the exogenous state variables, implied by the transition process F^w. We opt for this weighting matrix so that deviations on more visited states receive greater weights and are, therefore, considered more relevant. This is the weighting matrix we use to compute $\widehat{J}_N(\theta_0)$.

We approximate the value of $\widehat{J}_N(\theta_0)$ for any fixed θ_0 in practice by solving a relaxed version of the optimization problem (20)–(21). We do so because when f is costly to evaluate (as it is in the present case), it is difficult to solve directly the minimization problem (27) (and (28)), as it requires searching over $(\widetilde{p},\pi)$ to minimize $J_N(\theta_0)$ when the constraint $\theta_0 = f(\widetilde{p},\pi;p,F)$ must be satisfied exactly for a fixed θ_0. Putting differently, finding particular values for $(\widetilde{p},\pi)$ that satisfy $\theta_0 = f(\widetilde{p},\pi;p,F)$ can be computationally costly, while solving relaxed versions of the well-behaved problem (20)–(21) is substantially simpler. To be specific, we solve the problem (F3)–(F4) for several values of ϵ, as explained in Online Appendix F.[6]

We calculate 90% confidence sets for θ using the procedure described in Section 5 of the main text and in Online Appendix F. For each sample, we generate 1000 replicated samples with size that is approximately $h_N \approx 8 \times \sqrt{NT}$. Specifically, we implement a standard i.i.d. subsampling, resampling firms over the full time period: For the small sample we draw 36 firms randomly, and for the large sample, we draw 65 firms. The computations were run on the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University.

C.2 Monte Carlo Results

We now discuss the results of the Monte Carlo simulations. In the baseline scenario, the long-run average probability that the firm stays in the market is 90.5%, while the long-run average probability of being active reduces to 83.3% in the counterfactual scenario (so that $\theta = 0.833$). The impact of the entry subsidy is to reduce the long-run average by 7.2 percentage points. Similar to the example presented in the main paper, the entry subsidy increases the exit rate of forward-looking firms, which translates into firms staying less often in the market in the steady state. This result is invariant to the alternative discretizations of the state space, since the discretizations are performed on the same underlying AR(1) processes.

Table C2 presents the Monte Carlo results. The top, middle, and bottom panels show the results for the alternative state spaces: small ($X = 16$), medium-sized ($X = 54$), and large ($X = 250$), respectively. In each panel, the top subpanel presents the results for the small sample ($N = 100, T = 5$), and the bottom

[6] We let ϵ range from 0 to 1 in an equally spaced grid with 50 points.

subpanel, for the large sample ($N = 1000, T = 15$). In each subpanel, we show for each alternative set of Restrictions 1–3, (i) the populational (true) identified set, (ii) the average estimates of the lower and upper bounds, θ^L and θ^U, (iii) the average bias of the estimated bounds, (iv) the average endpoints and the average length of the 90% confidence sets, (v) the coverage probability of the confidence sets, and (vi) the average time taken to estimate θ^L and θ^U (in seconds), and the average time taken to compute the confidence intervals (in minutes).

The identified sets under the alternative Restrictions 1–3 are all compact intervals containing the true θ (Proposition 4), and vary slightly with the size of the state space. Restriction 1 alone is highly informative: the counterfactual long-run average probability of being active is between 75% and 90.5%. It does however include the baseline probability (at the upper end of the interval). Adding Restriction 2 reduces the upper bound to 87.8%, which suffices to identify the sign of the impact of the subsidy. And adding Restriction 3 pushes the upper bound further down to 86.8%.

In all cases, the estimated lower and upper bounds of the identified sets appear to be consistent, with smaller biases in larger samples. The coverage probabilities of the confidence sets converge to the nominal level 90%, as expected (Theorem 1). And the confidence sets' average lengths are wider (though not substantially) than the length of the true identified sets, for all sample sizes and state spaces. E.g., in the small state space case and large sample, the average length of the confidence set is 0.1782 under Restriction 1, while the length of the (true) identified set is just 0.1536; and in the large state space and small sample, the average length of the confidence set under the same restriction is 0.25.

Naturally, the finite sample performance of our inference procedure depends on both the state space and the sample size. In the larger state space cases, we obtain slightly greater average biases for the point estimates. These are expected: larger state spaces imply less (effective) degrees of freedom, as the number of model parameters increases with the state space. (Recall that π is an $(A+1)X$ vector.)

In terms of the computer time required to solve the minimization and maximization problems (20)–(21), it takes approximately 0.03 seconds to solve both optimization problems under Restrictions 1 and 1–2, and that time is reduced to just 0.01 seconds under Restrictions 1–3, in the small state space case. Subsampling is computationally intensive but feasible: for the same state space, the average time required to run it varies from two minutes under Restriction 1 to one minute under Restrictions 1–3.

As expected, it takes longer to solve (20)–(21) when the state space is larger. E.g., under Restriction 1, it takes approximately 0.3 seconds on average in the medium-sized state space case ($X = 54$), and approximately 6 seconds on average in the large state space case ($X = 250$). It also takes longer to run the subsampling procedure: between 7 and 28 minutes on average in the medium-sized state space, and between 150 and 580 minutes on average in the large state space, depending on the sample size and the restrictions imposed. It is important to stress, however, that the average computer time here is based on a sequential implementation of subsampling, which does not take advantage of parallelization.

Table C2: Monte Carlo Results

Target Parameter: θ = Long-run Average Probability of Being Active			
Small State Space: $X = 16$			
$T = 5, N = 100$	Restrictions 1	Restrictions 1–2	Restrictions 1–3
True Identified Set	[0.7500, 0.9036]	[0.7500, 0.8763]	[0.7500, 0.8662]
Average Estimated Bounds	[0.7583, 0.9036]	[0.7579, 0.8727]	[0.7580, 0.8651]
Average Bias	[0.0083, 0.0000]	[0.0079, -0.0036]	[0.0080, -0.0011]
Confidence Sets: Average Endpoints	[0.6729, 0.9214]	[0.6734, 0.8951]	[0.6757, 0.8870]
Confidence Sets: Average Length	0.2485	0.2217	0.2113
Coverage Probability (90% nominal)	0.9060	0.9010	0.9050
Time Estimation (sec)	0.04	0.05	0.01
Time Inference (min)	2	2	1
$T = 15, N = 1000$	Restrictions 1	Restrictions 1–2	Restrictions 1–3
True Identified Set	[0.7500, 0.9036]	[0.7500, 0.8763]	[0.7500, 0.8662]
Average Estimated Bounds	[0.7507, 0.9036]	[0.7507, 0.8761]	[0.7507, 0.8661]
Average Bias	[0.0007, -0.0000]	[0.0007, -0.0002]	[0.0007, -0.0001]
Confidence Sets: Average Endpoints	[0.7296, 0.9079]	[0.7296, 0.8806]	[0.7297, 0.8713]
Confidence Sets: Average Length	0.1782	0.1510	0.1417
Coverage Probability (90% nominal)	0.9090	0.9010	0.9040
Time Estimation (sec)	0.04	0.04	0.01
Time Inference (min)	2	2	0.7
Medium-sized State Space: $X = 54$			
$T = 5, N = 100$	Restriction 1	Restrictions 1–2	Restrictions 1–3
True Identified Set	[0.7503, 0.9057]	[0.7503, 0.8784]	[0.7503, 0.8682]
Average Estimated Bounds	[0.7591, 0.9036]	[0.7581, 0.8710]	[0.7586, 0.8641]
Average Bias	[0.0089, -0.0021]	[0.0078, -0.0074]	[0.0083, -0.0041]
Confidence Sets: Average Endpoints	[0.6656, 0.9235]	[0.6589, 0.9042]	[0.6628, 0.8932]
Confidence Sets: Average Length	0.2579	0.2453	0.2304
Coverage Probability (90% nominal)	0.8940	0.9050	0.8910
Time Estimation (sec)	0.34	0.41	0.03
Time Inference (min)	28	22	12
$T = 15, N = 1000$	Restriction 1	Restrictions 1–2	Restrictions 1–3
True Identified Set	[0.7503, 0.9057]	[0.7503, 0.8784]	[0.7503, 0.8682]
Average Estimated Bounds	[0.7509, 0.9056]	[0.7509, 0.8782]	[0.7509, 0.8680]
Average Bias	[0.0006, -0.0001]	[0.0006, -0.0002]	[0.0006, -0.0002]
Confidence Sets: Average Endpoints	[0.7292, 0.9101]	[0.7290, 0.8831]	[0.7290, 0.8748]
Confidence Sets: Average Length	0.1809	0.1541	0.1459
Coverage Probability (90% nominal)	0.9020	0.9070	0.8990
Time Estimation (sec)	0.30	0.27	0.03
Time Inference (min)	24	17	7
Large State Space: $X = 250$			
$T = 5, N = 100$	Restriction 1	Restrictions 1–2	Restrictions 1–3
True Identified Set	[0.7504, 0.9060]	[0.7504, 0.8787]	[0.7503, 0.8685]
Average Estimated Bounds	[0.7612, 0.9027]	[0.7605, 0.8701]	[0.7593, 0.8638]
Average Bias	[0.0108, -0.0033]	[0.0102, -0.0086]	[0.0090, -0.0047]
Confidence Sets: Average Endpoints	[0.6678, 0.9253]	[0.6602, 0.9096]	[0.6621, 0.8979]
Confidence Sets: Average Length	0.2575	0.2494	0.2358
Coverage Probability (90% nominal)	0.8960	0.9090	0.9080
Time Estimation (sec)	7	8	0.7
Time Inference (min)	578	477	252
$T = 15, N = 1000$	Restrictions 1	Restrictions 1–2	Restrictions 1–3
True Identified Set	[0.7504, 0.9060]	[0.7504, 0.8787]	[0.7503, 0.8685]
Average Estimated Bounds	[0.7532, 0.9064]	[0.7532, 0.8790]	[0.7510, 0.8685]
Average Bias	[0.0028, 0.0004]	[0.0028, 0.0003]	[0.0007, 0.0000]
onfidence Sets: Average Endpoints	[0.7321, 0.9106]	[0.7287, 0.8845]	[0.7288, 0.8757]
Confidence Sets: Average Length	0.1786	0.1558	0.1469
Coverage Probability (90% nominal)	0.9070	0.9000	0.9020
Time Estimation (sec)	6	6	0.6
Time Inference (min)	505	457	150

Note: T = number of periods, N = number of markets, X = number of states.

D Replication of Das, Roberts, and Tybout (2007)

We now present briefly our replication of Das, Roberts, and Tybout (2007), as well as the details of our counterfactual exercise.

Parameter Estimates. As explained in the main text, every period t a firm i chooses whether to export or not, $a_{it} \in \mathcal{A} = \{0, 1\}$, after observing the state variables k_{it} (the lagged decision), e_t (the exchange rate), ν_{it} (the demand/supply shocks in export markets), and the logit shocks ε_{it}. Both states k_{it} and e_t are observed by the econometrician, while ν_{it} can be recovered from data on export revenues, as explained below.

The payoff function is given by equation (10) in Section 3. DRT specify the (log of) variable profits as

$$\ln vp_{it} = \psi_0 + \psi_1 z_i + \psi_2 e_t + \nu_{it},$$

where z_i is a dummy variable indicating whether the firm is large or not (based on domestic sales in year 0). They also assume the profit shocks ν_{it} equal the sum of two independent AR(1) processes (so that ν_{it} follows an ARMA(2,1) process). We instead assume ν_{it} is AR(1); the results are not sensitive to this simplification.

We estimate the parameters of vp "offline." Following DRT, we impose monopolistic competition in export markets; it yields a simple expression for vp in terms of export revenues: $vp_{it} = \eta_i^{-1} R_{it}^f$, where $\eta_i > 1$ is a firm-specific foreign demand elasticity, and R_{it}^f are export revenues.[7] This relationship is useful because R_{it}^f is observed in the data while vp_{it} is not. That implies the regression equation

$$\ln R_{it}^f = \ln \eta_i + \psi_0 + \psi_1 z_i + \psi_2 e_t + \nu_{it}, \tag{D1}$$

which can be used for estimation. Although ψ_2 can be estimated directly by differencing the fixed-effects out in (D1), we still need to estimate the demand elasticities η_i to recover the state variable ν_{it}. To deal with the incidental parameters $\{\eta_i\}_{i=1}^N$, DRT assume monopolistic competition in domestic markets and impose that the ratio of foreign demand elasticities to domestic demand elasticities is constant for all producers and equals $(1 + \upsilon)$. Then, by exploiting the markup equation in both domestic and foreign markets, they obtain

$$1 - \frac{C_{it}}{R_{it}} = \eta_i^{-1} \left(1 + \upsilon \, \frac{R_{it}^d}{R_{it}} \right) + \xi_{it}, \tag{D2}$$

where C_{it} and R_{it} are total costs and total revenues (from both domestic and foreign markets), R_{it}^d are domestic revenues, and ξ_{it} is an error term that accommodates noise in this relationship. Based on data on costs and revenues, we estimate $\{\eta_i\}_{i=1}^N$ and υ applying a Nonlinear Least Squares estimator to equation

[7] The standard markup equation implied by profit maximization under monopolistic competition is $R_{it}^f(1 - \eta_i^{-1}) = C_{it}^f$, where C_{it}^f is the variable cost of exporting.

(D2). Then, given all estimated η_i's, we regress $\ln R_{it}^f - \ln \eta_i$ on z_i and e_t to estimate ψ_0, ψ_1, and ψ_2 in equation (D1) using Ordinary Least Squares. The parameters of the ν_{it} process are estimated using the Maximum Likelihood estimator applied to the residuals of that regression. Following DRT, we assume the exchange rate e_t follows an AR(1) process and take the values estimated by Ocampo and Villar (1995) based on a longer time-series, 1968–1992. After the parameters of the profit function, vp, and of the state transitions, ν_{it} and e_t, are estimated we move to the estimation of the dynamic parameters (namely, s, ec, and fc).

To estimate the dynamic parameters, we discretize the state space and estimate CCPs using frequency estimators. Given the small sample size, we discretize the support of each exogenous state in three bins, and ignore firms' types (z_i). Because ν_{it} is observed only when the firm is exporting, we assume that every time a firm decides to start exporting, it draws a value from ν_{it}'s ergodic distribution. (This implies that when the firm is not exporting, the only exogenous state is e_t.) Like DRT, we set the discount factor to 0.9. Finally, we estimate the dynamic parameters, as well as the scale parameter σ, by searching the values that best fit the dynamic equation (6), $\mathbf{M}\pi = \sigma b_{-J}$ (i.e, we use a Minimum Distance estimator). Here, we impose DRT's identification assumptions: scrap values are equal to zero, and fixed and entry costs do not depend on states.

Table D1: Model Parameter Estimates

Profit Function Parameters	(1)	(2)	Dynamic Parameters	(1)
ψ_0 (intercept)	-10.89	-9.03	ec (entry cost)	127.45
	(-20.46, -1.30)	(-19.09, 1.03)		(37.88, 239.34)
ψ_1 (large domestic size)	1.45	-		
	(0.76, 2.15)	-		
ψ_2 (exchange rate coefficient)	3.79	3.61	fc (fixed cost)	7.08
	(1.76, 5.81)	(1.48, 5.76)		(2.03, 10.83)
λ_{AR} (AR root)	0.797	0.823		
	(0.785, 0.807)	(0.818, 0.834)		
σ_{AR} (AR unconditional std)	1.12	1.12	σ (scale parameter)	26.28
	(1.10, 1.14)	(1.10, 1.15)		(7.94, 48.07)

Table D1 presents our results, with 90% confidence intervals in parentheses. Although our point estimates are not identical to DRT's estimates (as expected, given the small adjustments that we made), they all lie in the range estimated by them (see column 4 of their Table 1, on page 851).[8]

Inference on Counterfactuals. We implement our inference procedure for $\theta = (\theta_R, \theta_F, \theta_E)$ in the following way: In the first step, we estimate (i) the state transitions, (ii) the variable profits as specified by DRT

[8]DRT do not implement a two-step approach as we do here. Instead, they estimate all model parameters simultaneously by maximizing the likelihood function using a Bayesian MCMC estimator. Another difference is that they assume normally distributed idiosyncratic shocks ε_{it}, while we assume a logit model. To make the scale parameters comparable, we need to multiply our estimated σ by $\frac{\pi}{\sqrt{6}}$. This is approximately 33.7, which is close to their estimates.

(but omitting z_i), and (iii) the conditional choice probabilities – all of them as explained above. In the second step, we estimate the identified sets for each element of θ under alternative model restrictions by solving the optimization problems (20)–(21). (To make our results comparable to DRT, we fix the scale parameter σ at the estimated value presented in Table D1.) We then calculate the corresponding confidence intervals as explained in Sections 4 and 5 of the main text and in Online Appendix F. We implement 1000 replications of a standard i.i.d. subsampling, resampling 20 firms over the sample time period, so that the size of each subsample is $h_N = 200 \approx 8 \times \sqrt{NT}$. To calculate the test statistic used in the subsampling, $\widehat{J}_N(\theta_0)$, we minimize the quadratic distances in (27) and (28), as explained in Online Appendix F, and we take a diagonal weighting matrix Ω with diagonal terms given by the square-root of the ergodic distribution of the state variable – in this way, deviations on more visited states are considered more relevant and receive greater weights. Given that the benefit-cost ratio of the revenues subsidy θ_R is known (ex ante) to be point identified, we use the plug-in estimator proposed by Kalouptsidi, Lima, and Souza-Rodrigues (2019) to estimate it, and 1000 standard i.i.d. bootstrap replications at the firm level to construct the confidence intervals for θ_R.

The exact formula for each element of θ follows. Let $\widetilde{f}^*$ and f^* be vectors with the ergodic distributions of the state variables in the counterfactual and in the baseline scenarios, respectively, arranged first by k_{it} and then by e_t and ν_{it}. (We abuse notation and use the same $\widetilde{f}^*$ for different counterfactuals.) The first counterfactual is a 2% revenue subsidy; the benefit-cost ratio is given by

$$\theta_R = \frac{(\widetilde{f}^* - f^*)' \times \mathbf{R}^f}{\widetilde{f}^{*\prime} \times 0.02 \times \mathbf{R}^f},$$

where $\mathbf{R}^f$ is the vector of export revenues ranging over the states $x_{it} = (k_{it}, e_t, \nu_{it})$; i.e.,

$$\mathbf{R}^f = \begin{bmatrix} 0 \\ R^f \end{bmatrix},$$

where the zero vector at the top indicates that the firm is not exporting in the steady-state, $k = 0$, and R^f are the export revenues ranging over e_t and ν_{it} when $k = 1$, according to equation (D1). (To simplify, we set η_i at its estimated median.)

The second counterfactual is a fixed cost subsidy of 28% (which approximately matches the 2 million pesos that DRT consider under their full set of restrictions). The benefit-cost ratio is now

$$\theta_R = \frac{(\widetilde{f}^* - f^*)' \times \mathbf{R}^f}{\widetilde{f}^{*\prime} \times 0.28 \times \begin{bmatrix} 0 \\ fc \end{bmatrix}},$$

where, as in the revenue subsidy, the vector in the denominator has a zero at the top indicating that firms

are not exporting in the steady-state when $k = 0$.

Finally, the third counterfactual is an entry cost subsidy of 25%. The benefit-cost ratio here is

$$\theta_R = \frac{(\widetilde{f}^* - f^*)' \times \mathbf{R}^f}{\widetilde{f}^{*\prime} \times 0.25 \times \begin{bmatrix} ec \circ \widetilde{p}_1 \\ 0 \end{bmatrix}},$$

where $\circ$ is the Hadamard (i.e., element-wise) multiplication, and $\widetilde{p}_1$ is the counterfactual entry probability vector. Note that the multiplication $ec \circ \widetilde{p}_1$ in the denominator reflects the fact that subsidies are paid only when the firm enters (which happens with probability $\widetilde{p}_1$).

When solving the optimization problems (20)–(21) for each element of $\theta = (\theta_R, \theta_F, \theta_E)$, we provide the numerical algorithm the gradients of θ based on the derivations presented in the Online Appendix G.

References

Chernoff, H. (1954): "On the Distribution of the Likelihood Ratio," *The Annals of Mathematical Statistics*, pp. 573–578.

Das, S., M. J. Roberts, and J. R. Tybout (2007): "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," *Econometrica*, 75(3), 837–873.

Kalouptsidi, M., L. Lima, and E. Souza-Rodrigues (2019): "On Estimating Counterfactuals Directly in Dynamic Models," Discussion paper, University of Toronto.

Krantz, S. G., and H. R. Parks (2003): *The Implicit Function Theorem: History, Theory, and Applications*. Birkäuser Basel, 1 edn.

Ocampo, J. A., and L. Villar (1995): *Colombian Manufactured Exports, 1967–91* pp. 54–98. New York: Routledge.

Romano, J. P., and A. M. Shaikh (2012): "On the uniform asymptotic validity of subsampling and the bootstrap," *The Annals of Statistics*, 40(6), 2798–2822.

Ziegler, G. M. (2012): *Lectures on Polytopes*, vol. 152. Springer Science & Business Media.