Supplement to "Counterfactual Analysis for Structural Dynamic Discrete Choice Models"

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

April 2024

This Supplemental Material consists of the following sections: Section E provides several useful examples of linear restrictions that are commonly employed in applied work. Section F presents detailed calculations in the firm entry/exit problem – our running example. Section G extends our main results to allow for unknown discount factor β and distribution G. Section H shows how to calculate the gradient of the function ϕ when it involves counterfactual average effects based on ergodic distributions of the state variables. And Section I discusses the replication of Das, Roberts, and Tybout (2007).

E Examples of Linear Restrictions

In this section, we provide several useful examples of linear restrictions, $R^{eq}\pi = r^{eq}$ and $R^{iq}\pi \leq r^{iq}$, that are commonly employed in applied work. For ease of exposition, we only consider restrictions on π_J (unless otherwise stated). Recall that $R^{eq} = [R^{eq}_{-J}, R^{eq}_{J}]$ and $R^{iq} = [R^{iq}_{-J}, R^{iq}_{J}]$.

Example 1. (Compact Payoffs) Assume $\delta^l_J \leq \pi_J \leq \delta^u_J$. Then $R^{iq}_{-J} = 0$, $R^{iq}_J = [-I,\ I]'$, $r^{iq} = [-\delta^{l\prime}_J,\ \delta^{u\prime}_J]'$, and the number of inequalities is $m = 2X$.

Example 2. (Exclusion Restriction I) Assume $\pi_J(x_1) = \pi_J(x_2)$. Then, $R^{eq}\pi = r$, with $R^{eq}_{-J} = 0$,

$$R^{eq}_J = \left[\begin{array}{ccccc} 1 & -1 & 0 & \cdots & 0 \end{array}\right],$$

and $r^{eq} = 0$. There is only one equality restriction: $d = 1$.

Example 3. (Exclusion Restriction II) Suppose we split the state space in $x = (k, w)$, where $k \in \mathcal{K} = \{1, ..., K\}$ and $w \in \mathcal{W} = \{1, ..., W\}$, with K, W finite. Assume π_J does not depend on w, i.e.,

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

$\pi_J(k,1) = \pi_J(k,2) = ... = \pi_J(k,W)$ for all k. When $K = 2$, $W = 3$, we obtain $R^{eq}_{-J} = 0$,

$$R^{eq}_J = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix},$$

and $r^{eq} = 0$. The number of linear equalities is now $d = K(W-1) < KW = X$.

Example 4. (Linear-in-Parameters Payoffs) Assume that $\pi = z\gamma$, where z is a known matrix of dimension $(A+1)X \times \eta_\gamma$, γ is a column vector $\eta_\gamma \times 1$, and we assume $(A+1)X \geq \eta_\gamma$. 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: $\gamma = (z'_u z_u)^{-1} z'_u \pi_u$. Substitution in the second equality gives $\pi_l = z_l (z'_u z_u)^{-1} z'_u \pi_u$. Therefore, $[z_l (z'_u z_u)^{-1} z'_u\ ,\ -I]\pi = 0$. I.e.,

$$R^{eq} = \left[z_l \left(z'_u z_u \right)^{-1} z'_u\ ,\ -I \right],$$

and $r^{eq} = 0$. The number of linear equalities is $d = X - \eta_\gamma$.

Example 5. (Monotonicity) Without loss, arrange x in increasing order. Assume π_J increases with x. Then $\pi_J(1) \leq \pi_J(2) \leq ... \leq \pi_J(X)$. In this case, take $m = X - 1$, $r^{iq} = 0$, $R^{iq}_{-J} = 0$, and

$$R^{iq}_J = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix}.$$

Example 6. (Concavity) Arrange x in increasing order, take equidistant points for x. Assume π_J is concave in x. Then $\pi_J(x_{i-1}) - 2\pi_J(x_i) + \pi_J(x_{i+1}) \leq 0$, for all $i = 2, ..., X-1$. In this case, take $m = X - 2$, $r^{iq} = 0$, $R^{iq}_{-J} = 0$, and

$$R^{iq}_J = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix}.$$

Example 7. (Smoothness). Suppose $\pi_J \geq 0$ and assume $\pi_J(x)$ is Lipschitz continuous in x. Then,

$\pi_J(x_i) - \pi_J(x_{i+1}) \leq L|x_i - x_{i+1}|$, for some known constant $L < \infty$, for all x. In this case, $R^{iq}_{-J} = 0$,

$$R^{iq}_J = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix},$$

r^{iq} is the vector with elements $L|x_i - x_{i+1}|$, and $m = X - 2$. Note that we can impose higher order restrictions on the variation of the function π as well. This may be important when we discretize a continuous state space and π is a smooth function of states.

Example 8. (Action-Monotonicity) Take the binary model with actions $\mathcal{A} = \{a, J\}$, and assume that $\pi_a(x) \geq \pi_J(x)$ for some x. Then R^{iq}_a is the vector with -1 at position x and zeros elsewhere. Similarly, R^{iq}_J is the vector with 1 at position x and zero elsewhere. I.e.,

$$R^{iq}\pi = \begin{bmatrix} 0 & \cdots & 0 & -1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \pi_a \\ \pi_J \end{bmatrix} \leq 0,$$

where $r^{iq} = 0$ and $m = 1$.

Example 9. (Supermodularity) Take the binary model again. Without loss, arrange x in increasing order. Assume the increasing differences for $x_{i+1} \geq x_i$:

$$\pi_a(x_{i+1}) - \pi_a(x_i) \geq \pi_J(x_{i+1}) - \pi_J(x_i).$$

Then, take

$$R^{iq} = \begin{bmatrix} \underbrace{\begin{matrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 1 & -1 \end{matrix}}_{=R^{iq}_a} & \vdots & \underbrace{\begin{matrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{matrix}}_{=R^{iq}_J} \end{bmatrix},$$

and $r^{iq} = 0$, with $m = X - 1$ inequalities.

F 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 $F_a = F_a^k \otimes F^w$, where F_a^k 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 zero otherwise; and $\otimes$ is the Kronecker product. Specifically,

$$F_0 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \otimes F^w = \begin{bmatrix} F^w & 0 \\ F^w & 0 \end{bmatrix}, \ F_1 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \otimes F^w = \begin{bmatrix} 0 & F^w \\ 0 & F^w \end{bmatrix}. \tag{F1}$$

The payoff vectors are the same as in (23) in the main paper and are rewritten below for convenience,

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

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)

$$\pi_1 = M_1 \pi_0 + b_1(p). \tag{F2}$$

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 (F2) becomes

$$\begin{bmatrix} vp - fc - ec \\ vp - fc \end{bmatrix} = \begin{bmatrix} I + \beta F^w & -\beta F^w \\ \beta F^w & I - \beta F^w \end{bmatrix} \begin{bmatrix} oo \\ s \end{bmatrix} + b_1(p). \tag{F3}$$

Note now that if π_0 is known, namely both the scrap vector s and oo 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:

$$\begin{bmatrix} ec \\ fc \end{bmatrix} = \begin{bmatrix} -I_2 & I_2 \\ 0 & -I_2 \end{bmatrix} \pi_1 - \begin{bmatrix} -I_2 & I_2 \\ 0 & -I_2 \end{bmatrix} \begin{bmatrix} vp \\ vp \end{bmatrix} = \begin{bmatrix} -I_2 & I_2 \\ 0 & -I_2 \end{bmatrix} \pi_1 + \begin{bmatrix} 0 \\ vp \end{bmatrix}. \tag{F4}$$

We next replace π_1 from our main equation to obtain

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

or

$$\begin{bmatrix} ec \\ fc \end{bmatrix} = \begin{bmatrix} s - oo \\ -\beta F^w oo - (I - \beta F^w) s \end{bmatrix} + \begin{bmatrix} b_l(p) - b_u(p) \\ -b_l(p) + vp \end{bmatrix}, \tag{F5}$$

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 $oo = 0$ the above becomes,

$$ec = s + b_l(p) - b_u(p),$$
$$fc = -(I - \beta F^w) s - b_l(p) + vp.$$

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, the first equation above becomes:

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

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: in particular, we can identify the sum of the scrap values and the present value of fixed costs, since we get:

$$s + (I - \beta F^w)^{-1} fc = (I - \beta F^w)^{-1} \left(-b_l(p) + vp\right)$$

and the right-hand-side is known.

Model Restrictions. We now turn to the model restrictions – presented again here for convinience.

1. $oo = 0, \quad fc \geq 0, \quad ec \geq 0$, 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 (F5), $ec \geq 0$ and $fc \geq 0$ translate respectively to:

$$s \geq b_u(p) - b_l(p), \tag{F6}$$

$$(I - \beta F^w) s \leq vp - b_l(p). \tag{F7}$$

Visualizing the set of inequalities (F6) is clear: the positive orthant is shifted to the point $b_u(p) - b_l(p)$.

The hyperplanes under (F7) intersect at a unique point because $(I - \beta F^w)$ is invertible. Suppose $W = 2$, then equation (F7) is written as the following two equations:

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

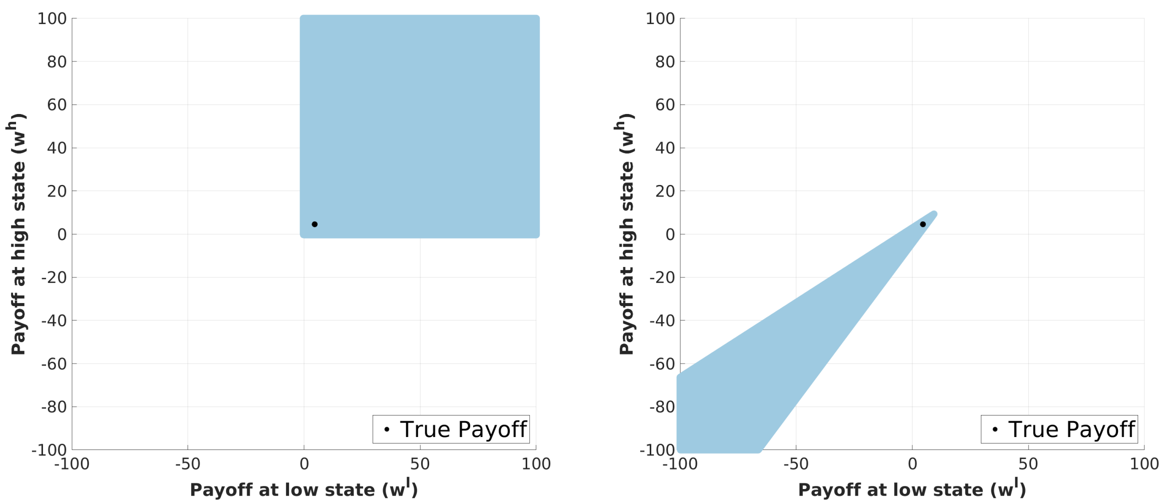
$$-\beta (1 - f_2) s_1 + (1 - \beta f_2) s_2 \leq vp_2 - b_{l2}(p)$$

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 F1 presents the set of values that s can take for the parameter configuration used in the numerical example presented in Section 5 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 F1.



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

Figure F1: Payoff Identified Set Π^I: Scrap Values under Alternative Restrictions

Remark F1. 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

$$\left\{ (0, s) \in \mathbb{R}^{2W} : \text{such that } s \text{ satisfies equations (F6) and (F7)} \right\}.$$

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

$$\pi_1(0) \leq 0. \tag{F8}$$

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

$f_w^{*\prime}$. Then, the inequality becomes

$$ec \leq \frac{1}{1-\beta}\mathbf{1}f_w^{*\prime}\left(vp - fc\right),$$

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:

$$\pi_1(1) - \pi_1(0) \leq \frac{1}{1-\beta}\mathbf{1}f_w^{*\prime}\pi_1(1)$$

or

$$\left[-I_2,\, I_2 - \frac{1}{1-\beta}\mathbf{1}f_w^{*\prime}\right]\pi_1 \leq 0. \tag{F9}$$

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

$$[0\ 0\ \ 1\ \text{-}1]\,\pi_1 \leq 0. \tag{F10}$$

Now we stack (F8), (F9) and (F10), so that:

$$R_{-J}^{iq}\,\pi_{-J} = R_1^{iq}\,\pi_1 = \begin{bmatrix} I_2 & 0 \\ -I_2 & I_2 - \frac{1}{1-\beta}\mathbf{1}f_w^{*\prime} \\ 0 & [1\ \text{-}1] \end{bmatrix}\pi_1 \leq 0, \tag{F11}$$

and $R_J^{iq} = R_0^{iq} = 0$ and $r^{iq} = 0$. Moreover, multiplying R_1^{iq}, from (F11), with M_1 gives,

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

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 F1 above), which implies

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

or in more detail,

$$
\begin{aligned}
-\beta f_1 s_1 - \beta\left(1-f_1\right) s_2 &\leq b_{u1}(p) \\
-\beta\left(1-f_2\right) s_1 - \beta f_2 s_2 &\leq b_{u2}(p) \\
\left(1-f_{w1}^*\right)\left(s_1-s_2\right) &\leq b_{u1}(p) - b_{l1}(p) + \frac{1}{\beta}\left(f_{w1}^* b_{l1}(p) + \left(1-f_{w1}^*\right) b_{l2}(p)\right) \\
-f_{w1}^*\left(s_1-s_2\right) &\leq b_{u2}(p) - b_{l2}(p) + \frac{1}{\beta}\left(f_{w1}^* b_{l1}(p) + \left(1-f_{w1}^*\right) b_{l2}(p)\right) \\
\left(1-\beta\left(f_1+f_2-1\right)\right)\left(s_1-s_2\right) &\leq b_{l1}(p) + b_{l2}(p),
\end{aligned}
$$

where $f_w^* = [f_{w1}^*, (1-f_{w1}^*)]'$.

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 $f_2 \neq 1 - f_1$.[1] The next two inequalities correspond to condition $ec \leq \mathbb{E}\left[vp - fc\right] / \left(1-\beta\right)$. 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 F1 above, Figure F2 shows the values of s for the parameter configuration presented in Section 5 of the main paper but under Restriction 2. Panel (a) shows the set under condition $vp - fc \leq ec$ (with the two downward sloping lines); panel (b) presents the set under $ec \leq \mathbb{E}\left[vp - fc\right] / \left(1-\beta\right)$ (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.

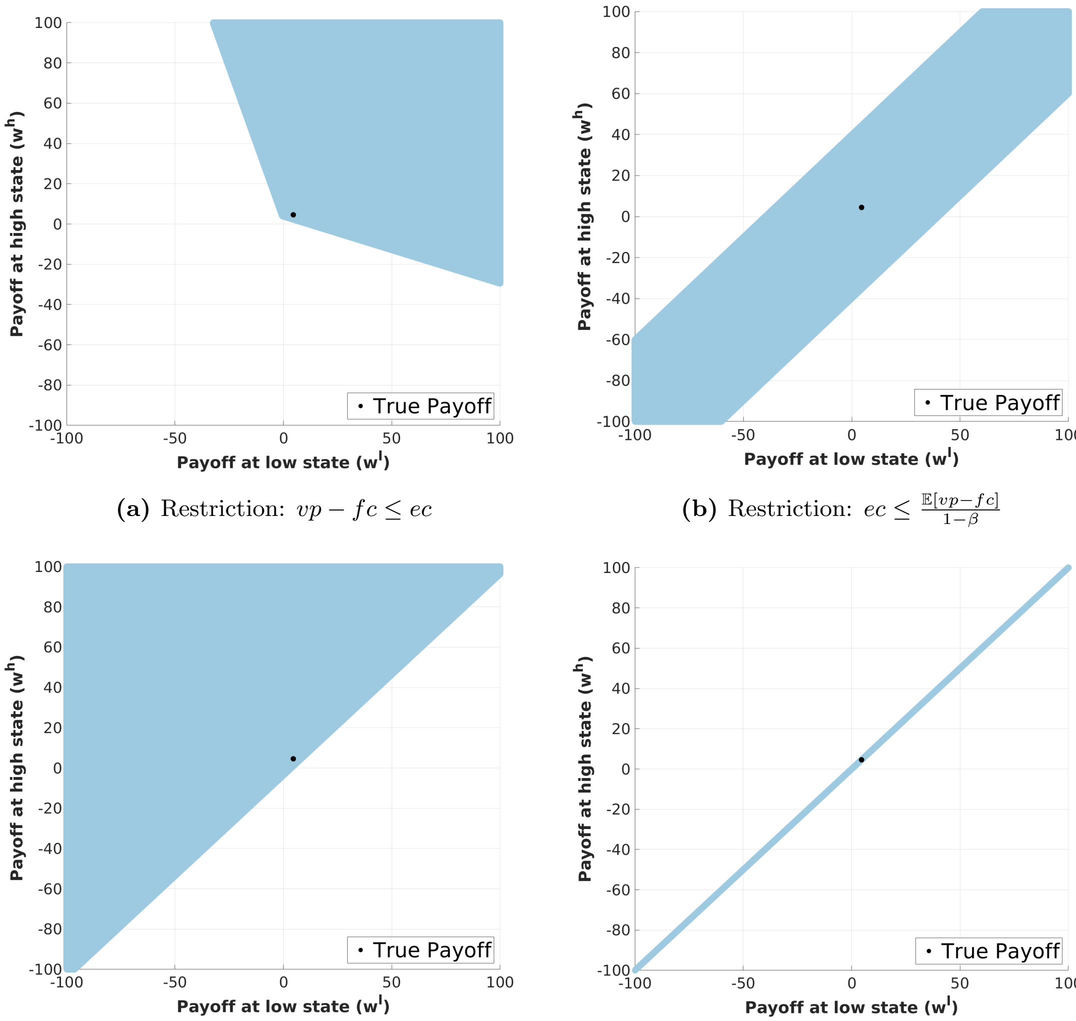
Restriction 3. If $s_1 = s_2 = s$, there is a single free parameter. This clearly results in a single line, presented in panel (d) of Figure F2. Combining Restrictions 1–3 result in the blue line inside the light blue polyhedron in panel (b) of Figure 1.

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

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

It is clear from Proposition 2 that the identified set $\widetilde{\mathbf{P}}^I$ is two-dimensional, regardless of any model restriction, because $\mathcal{H}_{11}$ is diagonalizable with two eigenvalues that are different from one.[2]

[1] If $\det F^w = 0$ then the two constraints collapse to the single constraint: $-\beta f_1 s_1 - \beta\left(1-f_1\right) s_2 \leq \min\left\{b_{u1}(p), b_{u2}(p)\right\}$.
[2] The eigenvalues of $\mathcal{H}_{11}$ are: $\lambda_1 = \lambda_2 = 1$ and $\lambda_3 = \lambda_4 = \tau < 1$.



(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 F2: Payoff Identified Set Π^I: Scrap Values under Alternative Model Restrictions

Indeed, when we calculate $\mathcal{C}_0$, defined in equation (15), we obtain

$$\mathcal{C}_0 = \mathcal{H}_{11} M_1 - M_1 \mathcal{H}_{00} = \begin{bmatrix} \tau I_2 & (1-\tau)\, I_2 \\ 0 & I_2 \end{bmatrix} \begin{bmatrix} I_2 + \beta F^w & -\beta F^w \\ \beta F^w & I_2 - \beta F^w \end{bmatrix} - \begin{bmatrix} I_2 + \beta F^w & -\beta F^w \\ \beta F^w & I_2 - \beta F^w \end{bmatrix}$$
$$= \begin{bmatrix} (\tau - 1)\, I_2 & (1-\tau)\, I_2 \\ 0 & 0 \end{bmatrix}.$$

Clearly, $rank(\mathcal{C}_0) = 2$. We conclude that the restriction $\pi_0(0) = 0$ does not alter the dimension of the identified set of the counterfactual CCP, although it makes (2) – or (F5) – 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_S), 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:[3]

$$\begin{bmatrix} p_0(w) \\ \vdots \\ p_A(w) \end{bmatrix} = \begin{bmatrix} p_0(0, w) & \cdots & p_0(A, w) \\ \vdots & \ddots & \vdots \\ p_A(0, w) & \cdots & p_A(A, w) \end{bmatrix} \begin{bmatrix} p_0(w) \\ \vdots \\ p_A(w) \end{bmatrix}. \tag{F12}$$

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

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

[3]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 $\Pr(a = 1|w) = p_1(0, w)/[1 - p_1(1, w) + p_1(0, w)]$.

The inner sum in the first term equals $\widetilde{p}_1(w)$ due to (F12). 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:

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

In the special case in which $x = (k, w)$, and k is the lagged action and w are exogenous shocks, we compute the consumer surplus for each action and state, $S(a,k,w)$, 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 $S(a,k,w) = 0$ when the firm is inactive ($a = 0$), and $S(a,k,w) = (w - mc)^2/8\eta$ when it is active ($a = 1$). So,

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

Note that, in our example, 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* S 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

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

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

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

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

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

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 F3 presents the identified set for θ based on the parameter configuration of the firm entry/exit model in Section 5. 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.

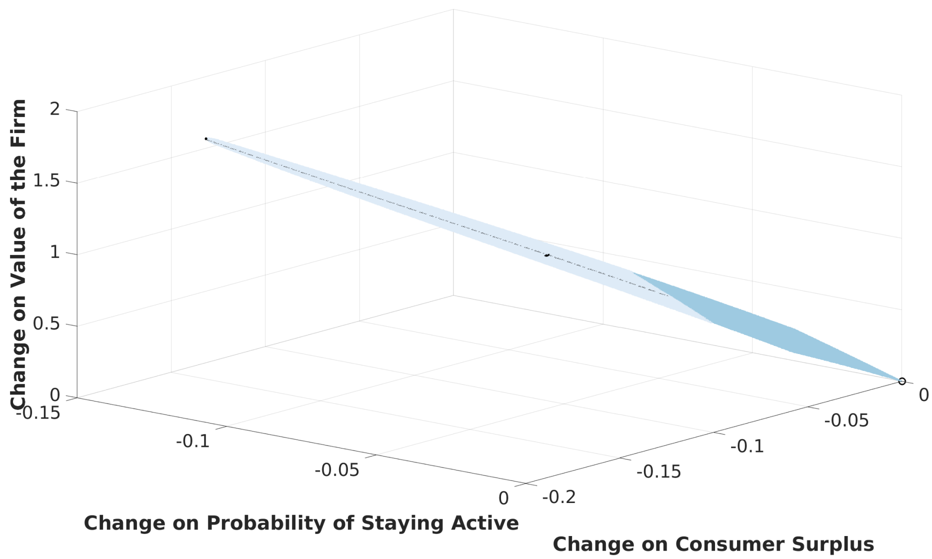


Figure F3: Identified Set $\mathbf{\Theta}^I$ under Restrictions 1–3

G Extension: Unknown β and G

We now extend our results to allow for unknown discount factor β and distribution G. First, we characterize the identified sets for $\widetilde{p}$ and θ. Then, we discuss practical implementation and its computational challenges.[4]

Formally, we assume that $\beta \in \mathcal{B} \subset [0, 1)$, and that the distribution of the idiosyncratic shocks ε_{it} belongs to a set of parametric distributions that are absolutely continuous with respect to the Lebesgue measure and have full support on $\mathbb{R}^{A+1}$. We denote that distribution by $G(\varepsilon_{it}; \lambda)$, where $\lambda \in \Lambda \subset \mathbb{R}^q$. The ex-ante value function, denoted here by $V(x; \lambda)$, is adjusted accordingly. The next proposition follows:

Proposition 1. *Assume that $\beta \in \mathcal{B} \subset [0, 1)$, where $\mathcal{B}$ is a compact interval, and that $G(\varepsilon_{it}; \lambda)$, with $\lambda \in \Lambda \subset \mathbb{R}^q$, where Λ is compact and convex. Consider the counterfactual $\{\widetilde{\mathcal{A}}, \widetilde{\mathcal{X}}, \widetilde{\beta}, \widetilde{G}, h^s, h\}$, where $\widetilde{\beta}$ and $\widetilde{G}$ are continuous functions of β and G, respectively, and assume that $V(x; \lambda)$ is continuously differentiable with respect to λ.[5] Then, the sharp identified set for the counterfactual CCP $\widetilde{p}$ is a connected manifold.*

[4] Magnac and Thesmar (2002) and Abbring and Daljord (2020) discuss identification of β; Blevins (2014), Chen (2017), and Buchholz, Shum, and Xu (2020) consider identification of G in binary choice models under different model assumptions.

[5] Formally, we need that, for any measurable and $\|\cdot\|_\infty$-bounded function $\bar{v}_a(x)$ on $\mathcal{A} \times \mathcal{X}$, the function $\bar{V}(\bar{v}(x); \lambda) := \int_\varepsilon \max_{a \in \mathcal{A}}\{\bar{v}_a(x) + \varepsilon_a\} \mathrm{d}G(\varepsilon; \lambda)$ is continuously differentiable with respect to λ. See Theorem 4.1 of Rust (1988).

Moreover, if ϕ is a continuous function of $(\widetilde{p}, \pi, \beta, \lambda)$, then the sharp identified set for θ is a connected set (and it is also compact when Π^I is bounded).

Proof. To extend the argument from Proposition 1 we need to prove two results: first, that the function $\bar{\varphi}(\kappa|\beta, \lambda)$ is jointly continuous on κ and (β, λ), where $\bar{\varphi}(\kappa|\beta, \lambda)$ corresponds to the function $\bar{\varphi}(\kappa)$ defined in Proposition 1 given the parameters β and λ. Second, that the set $\bigcup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} \mathcal{K}|_{\beta,\lambda}$ is connected, where $\mathcal{K}|_{\beta,\lambda}$ denotes the set $\mathcal{K}$ in Proposition 1 given β and λ. If that is the case, then the image set of $\bigcup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} \mathcal{K}|_{\beta,\lambda} \times \{\beta, \lambda\}$ under the function $\bar{\varphi}$ is a connected set.

Joint continuity of $\bar{\varphi}(\kappa|\beta, \lambda)$ follows because β affects the Bellman contraction mapping continuously provided it is bounded away from 1, and because of Theorem 4.1 in Rust (1988), as V is continuously differentiable with respect to λ and a discrete state-space implies his assumptions (A11) and (A12).[6]

The same conditions imply that $b_{-J}(p|\beta, \lambda)$ are continuous in (β, λ). Therefore, if $\mathcal{B} \times \Lambda$ is connected, then $\bigcup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} \mathcal{K}|_{\beta,\lambda}$ is connected as well because $\mathcal{K}|_{\beta,\lambda}$ depends on $b_{-J}(p|\beta, \lambda)$ continuously. It follows that the set $\bigcup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} \mathcal{K}|_{\beta,\lambda} \times \{\beta, \lambda\}$ is connected as the cartesian product of connected sets is connected. ∎

Naturally, the identified sets for $\widetilde{p}$ and θ in Proposition 1 are wider, and the computational costs involved are greater (as discussed below) than when β and G are known. This extension also requires choosing the compact intervals to which β and λ belong a priori, which can be difficult in practice – though we note that in some cases some prior information exist and can help; e.g., β might be known to be greater than or equal to 0.9 and smaller than 1. Nonetheless these difficulties, this extension has the benefit of preserving the smoothness (and so the computational feasibility) of our optimization while allowing for a set of flexible distributions for ε_{it}.

Implementation. We now briefly discuss some practical issues regarding the computation of the identified sets. To find the bounds for a scalar θ in practice, we need to solve the (augmented) constrained optimization problem:

$$\theta^U \equiv \max_{(\widetilde{p},\pi,\beta,\lambda)\in\widetilde{\mathbf{P}}\times\mathbb{R}^{(A+1)X}\times\mathcal{B}\times\Lambda} \phi\left(\widetilde{p}, \pi, \beta, \lambda;\, p, F\right) \tag{G13}$$

subject to

$$\begin{aligned}
\mathbf{M}(F, \beta)\,\pi \quad &= \quad b_{-J}(p, F; \beta, \lambda), \\
R^{eq}\,\pi \quad &= \quad r^{eq}, \\
R^{iq}\,\pi \quad &\leq \quad r^{iq}, \\
(\widetilde{\mathbf{M}}(F, \beta)\mathcal{H})\,\pi \quad &= \quad \widetilde{b}_{-J}\left(\widetilde{p}, F; \beta, \lambda\right) - \widetilde{\mathbf{M}}(F, \beta) g.
\end{aligned} \tag{G14}$$

[6]The other assumptions in Rust's (1988) Theorem 4.1 are already satisfied by our model or hold vacuously. See the discussion in Rust (1988), especially page 1015, and Norets (2010) for a more general result.

As before, the lower bound is obtained similarly, by replacing max by min. (Note that we subsumed in the notation for $\widetilde{\mathbf{M}}(F,\beta)$ and $\widetilde{b}_{-J}\left(\widetilde{p},F;\beta,\lambda\right)$ the fact that $\widetilde{\beta}$ and $\widetilde{G}$ are continuous functions of β and G.)

The optimization (G13)–(G14) is well-behaved and can be solved using standard software, as before. Yet, two additional computational challenges arise, besides the fact that the search now must be performed on a larger space. First, the discount factor β enters the constraints through the matrix $\mathbf{M}(F,\beta)$, which stacks M_a for all $a \neq J$, which in turn require calculating inverse matrices for any given β; see equation (3). This challenge can be avoided when action J is a renewal or terminal action (or under general forms of finite dependence), because M_a simplifies and does not involve an inverse matrix.[7] (The same argument holds for the matrix $\widetilde{\mathbf{M}}(F,\beta)$.) Second, for flexible distributions $G(\varepsilon;\lambda)$ for which there is no closed-form expression for ψ_a, one may need to use the linear program algorithm proposed by Chiong, Galichon, and Shum (2016) in order to calculate ψ_a (and obtain the corresponding b_{-J} and $\widetilde{b}_{-J}$) for each given λ.[8] The additional computational costs reflect the difficulties in calculating counterfactuals in DDC models more generally when β and G are unknown. Norets (2011) characterizes the identified set for $\widetilde{p}$ in dynamic multinomial choice models when G is completely nonparametric, but the identified set is infeasible to compute in practice. We, in contrast, maintain the parametric assumption, allowing for a set of flexible distributions while preserving the smoothness – and the computational feasibility – of the augmented constrained optimization.

H Gradient of ϕ involving Ergodic Distribution

In this section, we show how to calculate the gradient of the function ϕ when it involves counterfactual average effects based on ergodic distributions of the state variables.

Omit for convenience (p,F) from the notation, and assume the function ϕ is given by

$$\phi(\widetilde{p},\pi) = \sum_{x\in\widetilde{\mathcal{X}}} \widetilde{Y}(x;\pi)\widetilde{f}^{*}(x) - \sum_{x\in\mathcal{X}} Y(x;\pi) f^{*}(x),$$

where $\widetilde{Y}(x;\pi)$ and $Y(x;\pi)$ are outcome variables of interest in the counterfactual and baseline scenarios and that may depend on baseline payoffs π (e.g., consumer surplus, or the firm value); and $\widetilde{f}^{*}(x)$ and $f^{*}(x)$ are the counterfactual and baseline steady-state distributions of the state variables (which in turn depend on state transitions and choice probabilities, as shown below). Note that all target parameters θ presented in our firm entry/exit example (discussed in details in Appendix F) and in our Monte Carlo study (Appendix D) are of this type; in the empirical application of Section 7 in the main text, we take the ratio of objects of this type.

[7] When action J is either a renewal or a terminal action, then for all $a,\ j \in \mathcal{A}$, $F_aF_J = F_jF_J$, which implies that $M_a = I + \beta\left(F_J - F_a\right)$, for all $a \in \mathcal{A}$. The matrix M_a is linear in β under more general forms of finite dependence.

[8] This clearly adds non-negligible computational costs. Not only does it require solving linear programming multiple times to maximize/minimize θ, given that we would need to calculate $\widetilde{\psi}$ for various $\widetilde{p}$ in an inner loop, but it is also nontrivial to obtain the best search directions to speed up convergence of the algorithm.

The term $\widetilde{f}^*(x)$ is the ergodic distribution of the (endogenous) Markovian process for the state variables

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

(A similar expression holds for $f^*(x)$.) In matrix notation, we have

$$\phi(\widetilde{p},\pi) = \widetilde{\mathbf{Y}}'\,\widetilde{f}^* - \mathbf{Y}'\,f^*,$$

where $\widetilde{\mathbf{Y}}$ and $\mathbf{Y}$ are vectors of the outcome variables in the counterfactual and baseline; and $\widetilde{f}^*$ is the vector of the ergodic distribution satisfying the steady-state condition

$$\widetilde{f}^{*\prime} = \widetilde{f}^{*\prime}\,\widetilde{\mathbf{F}}, \tag{H15}$$

where

$$\widetilde{\mathbf{F}} = \sum_{a\in\widetilde{\mathcal{A}}} \widetilde{P}_a\widetilde{F}_a, \tag{H16}$$

and $\widetilde{P}_a$ is a diagonal matrix with $\widetilde{p}_a$ in its diagonal, and $\widetilde{F}_a$ is the counterfactual transition matrix conditional on the choice a. (Again, a similar expression holds for f^*.) Importantly, the ergodic distribution $\widetilde{f}^*$ depends directly on $\widetilde{p}$ (through equations (H15)–(H16)), and indirectly on the baseline payoff π, since $\widetilde{p}$ depends on π through equation (13) presented in the main text.

We want to know the derivative of ϕ with respect to π and $\widetilde{p}$, holding all other arguments of ϕ constant (e.g., the baseline CCP p and the state transitions F). We focus on the derivative with respect to π; obtaining the derivative with respect to $\widetilde{p}$ is similar but simpler, and is therefore omitted. Clearly, we have

$$\frac{\partial\phi}{\partial\pi'} = \left(\widetilde{f}^{*\prime}\,\frac{\partial\widetilde{\mathbf{Y}}}{\partial\pi'}\right) - \left(f^{*\prime}\,\frac{\partial\mathbf{Y}}{\partial\pi'}\right) + \left(\widetilde{\mathbf{Y}}'\,\frac{\partial\widetilde{f}^*}{\partial\pi'}\right).$$

The derivatives $\frac{\partial\widetilde{\mathbf{Y}}}{\partial\pi'}$ and $\frac{\partial\mathbf{Y}}{\partial\pi'}$ depend on the specific outcome of interest. Here, we focus on the third term of the right-hand-side, $\frac{\partial\widetilde{f}^*}{\partial\pi'}$. By the chain rule, we have

$$\frac{\partial\widetilde{f}^*}{\partial\pi'} = \frac{\partial\widetilde{f}^*}{\partial\widetilde{p}'}\frac{\partial\widetilde{p}}{\partial\pi'}.$$

By equation (13), we know that

$$\frac{\partial\widetilde{p}}{\partial\pi'} = \left(\frac{\partial\widetilde{b}_{-J}}{\partial\widetilde{p}'}\right)^{-1}\widetilde{\mathbf{M}}\mathcal{H}.$$

See Appendix B.2 for a discussion of how one can take advantage of the structure of $\widetilde{b}_{-J}$ to reduce the cost of inverting $\left(\frac{\partial\widetilde{b}_{-J}}{\partial\widetilde{p}'}\right)$ in practice.

We now derive the remaining term $\frac{\partial \widetilde{f}^*}{\partial \widetilde{p}'}$. Let $x, x', \tilde{x}$ be arbitrary states and $\widetilde{a} \neq J$. Then (H15) pointwise becomes

$$\widetilde{f}^*\left(x'\right) = \sum_x \widetilde{f}^*\left(x\right) \sum_a \widetilde{p}_a\left(x\right) \widetilde{F}\left(x'|x, a\right).$$

Therefore,

$$\frac{\partial \widetilde{f}^*\left(x'\right)}{\partial \widetilde{p}_{\widetilde{a}}\left(\tilde{x}\right)} = \sum_x \frac{\partial \widetilde{f}^*\left(x\right)}{\partial \widetilde{p}_{\widetilde{a}}\left(\tilde{x}\right)} \widetilde{F}\left(x'|x\right) + \widetilde{f}^*\left(\tilde{x}\right)\left[\widetilde{F}\left(x'|\tilde{x}, \widetilde{a}\right) - \widetilde{F}\left(x'|\tilde{x}, J\right)\right].$$

This is written compactly in matrix as,

$$\frac{\partial \widetilde{f}^*}{\partial \widetilde{p}_a'} = -(\widetilde{\mathbf{F}}' - I)^+ (\widetilde{F}_a' - \widetilde{F}_J')\, \widetilde{\mathbf{f}}^*, \tag{H17}$$

where $(\widetilde{\mathbf{F}}' - I)^+$ is the pseudo-inverse of $(\widetilde{\mathbf{F}}' - I)$, and $\widetilde{\mathbf{f}}^*$ is a diagonal matrix with $\widetilde{f}^*$ in its diagonal.

I 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 directly observed by the econometrician, while ν_{it} can be recovered from data on export revenues, as explained below. We take $w_{it} = (e_t, \nu_{it})$.

The payoff function is given by equation (23) in Section 5. 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.[9] This relationship is useful

[9]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.

because R^f_{it} is observed in the data while vp_{it} is not. That implies the regression equation

$$\ln R^f_{it} = \ln \eta_i + \psi_0 + \psi_1\, z_i + \psi_2\, e_t + \nu_{it}, \tag{I18}$$

which can be used for estimation. Although ψ_2 can be estimated directly by differencing the fixed-effects out in (I18), 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

$$1 - \frac{C_{it}}{R_{it}} = \eta_i^{-1}\left(1 + \upsilon\, \frac{R^d_{it}}{R_{it}}\right) + \xi_{it}, \tag{I19}$$

where C_{it} and R_{it} are total costs and total revenues (from both domestic and foreign markets), R^d_{it} 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 (I19). Then, given all estimated η_i's, we regress $\ln R^f_{it} - \ln \eta_i$ on z_i and e_t to estimate ψ_0, ψ_1, and ψ_2 in equation (I18) 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 (7), $\mathbf{M}\pi = \sigma b_{-J}$ (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 I1 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).[10]

[10] 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

Table I1: 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)

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 (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 (21)–(22). To make our results comparable to DRT, we fix the scale parameter σ at the estimated value presented in Table I1. We then calculate the corresponding confidence intervals as explained in Section 6 of the main text and in Appendix C. We implement 1000 replications of a standard i.i.d. subsampling, resampling 16 firms over the sample time period, so that the size of each subsample is $h_N = 16 \approx N^{2/3}$. To minimize the quadratic distances in (27) and (28), 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. The grid set for the approximations in the main sample is an equally spaced grid with $K = 50$ points and a range from $\epsilon_0 = 0$ to $\epsilon_{max} = 0.1$; for the subsamples, we have the same number of equally-distant points but over $\epsilon_0 = 0$ to $\epsilon_{max} = 1$. We randomly generated a set of 25 initial values for the optimizations as discussed in Setion C. 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 (2024) 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 slightly and use the same $\widetilde{f}^*$ for different counterfactuals.

multiply our estimated σ by $\frac{\pi}{\sqrt{6}}$. This is approximately 33.7, which is close to their estimates.

The first counterfactual is a 2% revenue subsidy; the benefit-cost ratio is given by

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

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 (I18); 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

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

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

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

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 $\widetilde{p}_1$). Importantly, Condition 1, required for our inference procedure (Theorem 1), is satisfied in this application, as discussed in Appendix A.

When solving the optimization problems (21)–(22) 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 Appendix H.

References

Abbring, J. H., and O. Daljord (2020): "Identifying the Discount Factor in Dynamic Discrete Choice Models," *Quantitative Economics*, 11(2), 471–501.

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

Buchholz, N., M. Shum, and H. Xu (2020): "Semiparametric Estimation of Dynamic Discrete Choice Models," Discussion paper, Princeton University.

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

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

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 (2024): "On Estimating Counterfactuals Directly in Dynamic Models," Discussion paper, University of Toronto.

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

Norets, A. (2010): "Continuity and Differentiability of Expected Value Functions in Dynamic Discrete Choice Models," *Quantitative Economics*, 1(2), 305–322.

——— (2011): "Semiparametric Identification of Dynamic Multinomial Choice Models," Discussion paper, Brown University.

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

Rust, J. (1988): "Maximum likelihood Estimation of Discrete Dontrol Processes," *SIAM Journal on Control and Optimization*, 26(5), 1006–1024.