Supplement to "Linear IV Regression Estimators for Structural Dynamic Discrete Choice Models

Myrto Kalouptsidi[*], Paul T. Scott[†], Eduardo Souza-Rodrigues[‡]

Harvard University, CEPR and NBER; NYU Stern; University of Toronto

August 2020

This supplemental material consists of the following sections: Section A presents general derivations of Euler equations for both deterministic and stochastic finite dependence. It also extends Proposition 2 in the main text to allow for stochastic sequences of choices (see Proposition A1). Sections B and C present all proofs of the lemmas and propositions presented in the main paper: Section B focuses on the identification results, while Section C shows the proofs of the asymptotic properties of the ECCP estimator. Section D explains the standard CCP approach implemented in the Monte Carlo experiment to estimate the model parameters. Finally, Section E extends the Monte Carlo study presented in Section 6 of the main paper by investigating how the biases in the parameter estimates pass through to biases in countefactuals calculations.

A General ECCP Equation Derivation

In this section, we offer a general derivation of Euler equations in conditional choice probabilities relying first on deterministic finite dependence as defined in Section 3, and then exploring stochastic sequences of choices. We also present Proposition A1, which extends the identification result in Proposition 2 (presented in the main text) to allow for stochastic finite dependence. Recall that finite dependence is not a behavioral assumption (whether based on deterministic or stochastic sequences); it is rather a property that the state transition process may or may not satisfy in the data.

[*]Harvard University, Department of Economics, Littauer Center 124, Cambridge, MA 02138. email: myrto@fas.harvard.edu

[†]Stern School of Business, New York University, Kaufman Management Center, 44 W. 4th St., New York, NY 10012. email: pscott@stern.nyu.edu.

[‡]Department of Economics, University of Toronto, Max Gluskin House, 150 St. George St., Toronto, Ontario M5S 3G7, Canada. email: e.souzarodrigues@utoronto.ca

A.1 ECCP Equation Under Deterministic Finite Dependence

Arcidiacono and Miller (2011) show that the conditional value functions, v_a, can be represented by functions of flow payoffs and conditional choice probabilities for any sequence of future choices, optimal or not. To derive such a representation, begin with an arbitrary initial state ω_{mt}. Consider a sequence of actions from t to $t+\tau$ (where $\tau \geq 1$). Suppose the initial choice at time period t is a, and let j denote another element of the choice set $\mathcal{A}$. Let $a_d \in \mathcal{A}$ denote the $d-$th choice in the sequence following a, and define $j_d \in \mathcal{A}$ similarly.

Recall equation (11) in the main paper, rewritten below for convenience. It stacks vectors across rows of the individual state k and absorbs the aggregate state ω_{mt} into mt subscripts:

$$\pi_{amt} + \beta e^V_{am,t,t+1} = V_{mt} - \beta F^k_{amt} V_{mt+1} - \psi_{amt}.$$

We then substitute for V_{mt+1} using equation (11) again, using a_1 as the action instead of a:

$$\begin{aligned}\pi_{amt} + \beta e^V_{am,t,t+1} &= V_{mt} - \psi_{amt} - \beta F^k_{amt}\left(\pi_{a_1mt+1} + \beta e^V_{a_1,m,t+1,t+2} + \psi_{a_1mt+1}\right) \\ &\quad -\beta^2 F^k_{amt} F^k_{a_1mt+1} V_{mt+2}.\end{aligned}$$

Repeated substitution of V_{mt+d} above leads to:

$$\begin{aligned}\pi_{amt} + \beta e^V_{am,t,t+1} &= V_{mt} - \psi_{amt} \\ &\quad -F^k_{amt}\left[\sum_{d=1}^{\tau} \beta^d \Lambda_{amtd}\left(\pi_{a_dmt+d} + \beta e^V_{a_d,m,t+d,t+d+1} + \psi_{a_dmt+d}\right)\right] \\ &\quad -\beta^{\tau+1} F^k_{amt} \Lambda_{amt,\tau+1} V_{mt+\tau+1}, \end{aligned} \tag{A1}$$

where the matrices Λ_{amtd} are defined recursively:

$$\begin{aligned}\Lambda_{amtd} &= I, & \text{for } d=1, \\ \Lambda_{amtd} &= \Lambda_{amt,d-1}\, F^k_{a_{d-1}mt+d-1}, & \text{for } d \geq 2.\end{aligned}$$

Next, finite dependence allows us to eliminate the $V_{mt+\tau+1}$, resulting in an ECCP equation that forms the basis of our identification arguments. Recall Definition 2 in Section 3: Given τ-period finite dependence, for a pair of actions (a, j), we can construct sequences $(a, a_1, \dots, a_\tau)$ and $(j, j_1, \dots, j_\tau)$ such that[1]

$$F^k_{amt} F^k_{a_1mt+1} \dots F^k_{a_\tau mt+\tau} = F^k_{jmt} F^k_{j_1mt+1} \dots F^k_{j_\tau mt+\tau},$$

i.e.,

$$F^k_{amt} \Lambda_{amt,\tau+1} = F^k_{jmt} \Lambda_{jmt,\tau+1}. \tag{A2}$$

[1]Recall that the terms in the sequences depend on the particular initial pair of actions (a, j) chosen.

We then difference equation (A1) across the two sequences of actions. Because of (A2), the last term cancels, and the result is equation (27).

A.2 ECCP Equation Under Stochastic Finite Dependence

Consider now a known mixing sequence of actions from t to $t+\tau$ (where $\tau \geq 1$). Suppose the initial choice at t is action a. For the next period, $t+1$, let $\alpha_{t+1}^{a}(k_{imt+1}, \omega_{mt+1})$ be a vector on $\mathbb{R}^{A+1}$ with elements $\alpha_{lt+1}^{a}(k_{imt+1}, \omega_{mt+1})$, $l \in \mathcal{A}$, such that $\sum_{l=0}^{A} \alpha_{lt+1}^{a}(k_{imt+1}, \omega_{mt+1}) = 1$. Each element $\alpha_{lt+1}^{a}(k_{imt+1}, \omega_{mt+1})$ of the vector α_{t+1}^{a} can be interpreted as the weight given to action $l \in \mathcal{A}$ in period $t+1$ at state (k_{imt}, ω_{mt}) after the initial choice a at t. For deterministic sequences, one element of the vector α_{t+1}^{a} equals one and the others equal zero. For probabilistic sequences, all elements of α_{t+1}^{a} are positive an add up to one. More generally, the mixing may involve negative weights, provided that they sum up to one (Arcidiacono and Miller, 2019). For the other time periods, $d = 2, ..., \tau$, take the sequence of weight choices in a similar way, $\alpha_{t+d}^{a}(k_{imt+d}, \omega_{mt+d})$. In this section, we abuse terminology and use the terms "mixing" and "stochastic" interchangeably.

It is useful to represent the weight choices in matrix notation. Define the diagonal matrix $\alpha_{lmt}^{a} = diag\left\{\alpha_{lt}^{a}\left(k, \omega_{mt}\right); k \in \mathcal{K}\right\}$, for $l \in \mathcal{A}$ and any time period t. In words, α_{lmt}^{a} is a $K \times K$ matrix that collects the individual terms $\alpha_{lt}^{a}(k_{imt}, \omega_{mt})$ for all possible values of k. Note that $\alpha_{lt}^{a}(k_{imt}, \omega_{mt})$ is one element of the vector $\alpha_{t}^{a}(k_{imt}, \omega_{mt})$, as defined in the previous paragraph. Clearly, because the mixing requires $\sum_{l=0}^{A} \alpha_{lt}^{a}(k_{imt}, \omega_{mt}) = 1$, we have that $\sum_{l=0}^{A} \alpha_{lmt}^{a} = I_K$, where I_K is the $K \times K$ identity matrix.

Now, recall equation (11). For any initial choice a, take a mixing $\alpha_{t+1}^{a}(k_{imt+1}, \omega_{mt+1})$ over choices in $\mathcal{A}$ at $t+1$, and replace V_{mt+1} in (11) to get

$$\begin{aligned}\pi_{amt} + \beta e_{am,t,t+1}^{V} &= V_{mt} - \psi_{amt} - \beta F_{amt}^{k}\left[\sum_{l=0}^{A} \alpha_{lmt+1}^{a}\left(\pi_{lmt+1} + \beta e_{lm,t+1,t+2}^{V} + \psi_{lmt+1}\right)\right] \\ &\quad -\beta^{2} F_{amt}^{k}\left[\sum_{l=0}^{A} \alpha_{lmt+1}^{a} F_{lmt+1}^{k}\right] V_{mt+2}.\end{aligned}$$

Next, we follow the steps outlined in Section A.1 of this Appendix. Repeated substitution of V_{mt+d} above leads to:

$$\begin{aligned}\pi_{amt} + \beta e_{am,t,t+1}^{V} &= V_{mt} - \psi_{amt} \\ &\quad -\beta F_{amt}^{k}\left[\sum_{d=1}^{\tau} \beta^{d-1} \Lambda_{am,t,d} \sum_{l=0}^{A} \alpha_{lmt+d}^{a}\left(\pi_{lmt+d} + \beta e_{lm,t+d,t+d+1}^{V} + \psi_{lmt+d}\right)\right] \\ &\quad -\beta^{\tau+1} F_{amt}^{k} \Lambda_{am,t,\tau+1} V_{mt+\tau+1}, \qquad \text{(A3)}\end{aligned}$$

where the observed (estimable) matrices $\Lambda_{am,t,d}$ are defined recursively:

$$\begin{array}{rcll} \Lambda_{am,t,d} & = & I, & \text{for} \quad d=1 \\ \Lambda_{am,t,d} & = & \Lambda_{am,t,d-1}\left[\sum_{l=0}^{A}\alpha_{lmt+d-1}^{a}F_{lmt+d-1}^{k}\right], & \text{for} \quad d\geq 2. \end{array} \tag{A4}$$

As before, we make use of finite dependence to eliminate the $V_{mt+\tau+1}$. First, we extend Definition 2 in Section 3 to stochastic sequences of choices:

Definition 1. (*Stochastic Finite Dependence*) A pair of choices a and j satisfies stochastic τ-period finite dependence if there exist two sequences of mixings starting at a and j such that, for all t,

$$F_{amt}^{k}\Lambda_{am,t,\tau+1} = F_{jmt}^{k}\Lambda_{jm,t,\tau+1}, \tag{A5}$$

where $\Lambda_{am,t,\tau+1}$ is defined in (A4).

Under this condition, $V_{mt+\tau+1}$ is eliminated when we difference equation (A3) across two mixing sequences of actions, starting respectively at a and j. Recalling that $\pi=\overline{\pi}+\xi$, we then obtain the ECCP regression equation:

$$\begin{aligned} &\psi_{jmt}-\psi_{amt}+\beta\sum_{d=1}^{\tau}\sum_{l=0}^{A}\beta^{d-1}\left[F_{jmt}^{k}\Lambda_{jm,t,d}\alpha_{lmt+d}^{j}-F_{amt}^{k}\Lambda_{am,t,d}\alpha_{lmt+d}^{a}\right]\psi_{lmt+d} \\ &=\overline{\pi}_{amt}-\overline{\pi}_{jmt}+\beta\sum_{d=1}^{\tau}\sum_{l=0}^{A}\beta^{d-1}\left[F_{amt}^{k}\Lambda_{am,t,d}\alpha_{lmt+d}^{a}-F_{jmt}^{k}\Lambda_{jm,t,d}\alpha_{lmt+d}^{j}\right]\overline{\pi}_{lmt+d} \\ &+u_{ajmt}, \end{aligned} \tag{A6}$$

where the econometric error term is $u_{ajmt}=\widetilde{\xi}_{ajmt}+\widetilde{e}_{ajmt}^{V}$, with

$$\widetilde{\xi}_{ajmt}=\xi_{amt}-\xi_{jmt}+\beta\sum_{d=1}^{\tau}\sum_{l=0}^{A}\beta^{d-1}\left[F_{amt}^{k}\Lambda_{am,t,d}\alpha_{lmt+d}^{a}-F_{jmt}^{k}\Lambda_{jm,t,d}\alpha_{lmt+d}^{j}\right]\xi_{lmt+d},$$

and

$$\widetilde{e}_{ajmt}^{V}=\beta\left(e_{am,t,t+1}^{V}-e_{jm,t,t+1}^{V}\right)+\beta\sum_{d=1}^{\tau}\sum_{l=0}^{A}\beta^{d}\left[F_{amt}^{k}\Lambda_{am,t,d}\alpha_{lmt+d}^{a}-F_{jmt}^{k}\Lambda_{jm,t,d}\alpha_{lmt+d}^{j}\right]e_{lm,t+d,t+d+1}^{V}.$$

We can now extend Proposition 2 to identify payoff parameters under the assumption of general stochastic finite dependence.

Proposition A1. *Suppose Assumptions 1 and 2 hold. Assume τ-period stochastic finite dependence holds for the agent-level transition process F^{k}, with $\tau<T$. Assume also a linear-in-parameters flow payoff: $\overline{\pi}(a,k,w)=x(a,k,w)\,\theta$, where $\theta\in\mathbb{R}^{P}$ and $x(a,k,w)$ is a known*

$1 \times P$ vector function. Let X_{amt} be a $K \times P$ matrix with elements given by $x(a, k, w_{mt})$, so that $\overline{\pi}_{amt} = X_{amt}\theta$, and define

$$\widetilde{X}_{ajmt} \equiv X_{amt} - X_{jmt} + \beta \sum_{d=1}^{\tau} \sum_{l=0}^{A} \beta^{d-1} \left[F_{amt}^{k} \Lambda_{am,t,d} \alpha_{lmt+d}^{a} - F_{jmt}^{k} \Lambda_{jm,t,d} \alpha_{lmt+d}^{j} \right] X_{lmt+d}. \tag{A7}$$

Denote the $K \times 1$ vector on the left hand side of (A6) by Y_{ajmt}. Stack equation (A6) for all Q feasible combinations of actions $(a, j) \in \mathcal{A}$ to obtain the following equation

$$Y_{mt} = \widetilde{X}_{mt}\theta + u_{mt}, \tag{A8}$$

where the $QK \times 1$ vectors Y_{mt} and u_{mt} stack Y_{ajmt} and u_{ajmt}, respectively, and the $QK \times P$ matrix $\widetilde{X}_{mt}$ stacks $\widetilde{X}_{ajmt}$. Let Z_{mt} be an $L \times QK$ matrix of instrumental variables with $L \geq P$. The parameter θ is identified provided $E[Z_{mt}u_{mt}] = 0$ and $rank(E[Z_{mt}\widetilde{X}_{mt}]) = P$.

The proof of Proposition A1 is identical to the proof of Proposition 2 and is therefore omitted. In fact, Proposition 2 is a special case of Proposition A1, when finite dependence is restricted to satisfy deterministic sequences of actions.

Proposition A1 can be used to extend previous empirical applications exploring stochastic finite dependence (e.g., Ransom, 2019) to incorporate serially correlated unobservable states, measurement error, and endogeneity problems. It can also serve as the basis for identification arguments in future applications featuring all these attributes. Evidently, the same set of issues discussed extensively in Section 4 involving deterministic finite dependence (regarding instrument validity, limitations of and extensions to the ECCP approach) applies here as well.

B Proofs: Identification

B.1 Proof of Proposition 1

Assume single-action τ-period dependence holds for action J. Then, equation (27) simplifies to

$$\begin{aligned} &(\psi_{jmt} - \psi_{amt}) + \left(F_{jmt}^{k} - F_{amt}^{k}\right) \sum_{d=1}^{\tau} \beta^{d} \Lambda_{Jmtd}\, \psi_{Jmt+d} \\ &= \overline{\pi}_{amt} - \overline{\pi}_{jmt} - \left(F_{jmt}^{k} - F_{amt}^{k}\right) \sum_{d=1}^{\tau} \beta^{d} \Lambda_{Jmtd}\, \overline{\pi}_{Jmt+d} + u_{ajmt}, \end{aligned} \tag{B1}$$

where the matrix Λ_{Jmtd} is defined recursively

$$\begin{aligned} \Lambda_{Jmtd} &= I, && \text{for} \quad d = 1 \\ \Lambda_{Jmtd} &= \Lambda_{Jmt,d-1}\, F_{Jmt+d-1}^{k}, && \text{for} \quad d \geq 2, \end{aligned}$$

and the unobservable term is $u_{ajmt} = \widetilde{\xi}_{ajmt} + \widetilde{e}^{V}_{ajmt}$, with

$$\widetilde{\xi}_{ajmt} = (\xi_{amt} - \xi_{jmt}) - (F^{k}_{jmt} - F^{k}_{amt}) \sum_{d=1}^{\tau} \beta^{d} \Lambda_{Jmtd}\, \xi_{Jmt+d}, \tag{B2}$$

$$\widetilde{e}^{V}_{ajmt} = \beta \left(e^{V}_{am,t,t+1} - e^{V}_{jm,t,t+1} \right) - \left(F^{k}_{jmt} - F^{k}_{amt} \right) \sum_{d=1}^{\tau} \beta^{d} \Lambda_{Jmtd}\, e^{V}_{Jm,t+d,t+d+1}. \tag{B3}$$

For any known (and comformable) function $h(z_{mt})$, multiply both sides of (B1) and take the expectation. We eliminate the error terms $\widetilde{\xi}_{ajmt}$ and $\widetilde{e}^{V}_{ajmt}$ by Assumption 3.(ii)–(iii). Then,

$$E\left[h(z_{mt}) \left((\psi_{jmt} - \psi_{amt}) + \left(F^{k}_{jmt} - F^{k}_{amt} \right) \sum_{d=1}^{\tau} \beta^{d} \Lambda_{Jmtd}\, \psi_{Jmt+d} \right) \right]$$

$$= E\left[h(z_{mt}) \left(\overline{\pi}_{amt} - \overline{\pi}_{jmt} - \left(F^{k}_{jmt} - F^{k}_{amt} \right) \sum_{d=1}^{\tau} \beta^{d} \Lambda_{Jmtd}\, \overline{\pi}_{Jmt+d} \right) \right], \tag{B4}$$

where the expectations are taken over $(z_{mt}, w_{mt}, ..., w_{mt+\tau})$.

The LHS of (B4) can be recovered from the data (using the results of Lemma C1, in Appendix C.4). Then, for any two primitives $\overline{\pi}$ and $\overline{\pi}'$,

$$\begin{aligned} & E\left[h(z_{mt}) \left(\overline{\pi}_{amt} - \overline{\pi}_{jmt} - \left(F^{k}_{jmt} - F^{k}_{amt} \right) \sum_{d=1}^{\tau} \beta^{d} \Lambda_{Jmtd}\, \overline{\pi}_{Jmt+d} \right) \right] \\ = \; & E\left[h(z_{mt}) \left(\overline{\pi}'_{amt} - \overline{\pi}'_{jmt} - \left(F^{k}_{jmt} - F^{k}_{amt} \right) \sum_{d=1}^{\tau} \beta^{d} \Lambda_{Jmtd}\, \overline{\pi}'_{Jmt+d} \right) \right]. \end{aligned}$$

By the completeness condition (Assumption 3.(i)),

$$\overline{\pi}_{amt} - \overline{\pi}_{jmt} - \left(F^{k}_{jmt} - F^{k}_{amt} \right) \sum_{d=1}^{\tau} \beta^{d} \Lambda_{Jmtd}\, \overline{\pi}_{Jmt+d}$$

$$= \overline{\pi}'_{amt} - \overline{\pi}'_{jmt} - \left(F^{k}_{jmt} - F^{k}_{amt} \right) \sum_{d=1}^{\tau} \beta^{d} \Lambda_{Jmtd}\, \overline{\pi}'_{Jmt+d}, \tag{B5}$$

for almost all $(w_{mt}, \ldots, w_{mt+\tau})$. Consider (B5) for $j = J$. Because $\overline{\pi}_{J}(k, w)$ is known for all observed states (k, w), we conclude that $\overline{\pi}_{amt} = \overline{\pi}'_{amt}$ almost surely.

B.2 Proof of Proposition 2

Equation (31) is a linear regression equation, and $E[Z_{mt} u_{ajmt}] = 0$ and $rank(E[Z_{mt} \widetilde{X}_{ajmt}]) = P$ are the standard orthogonality and rank conditions, respectively, for parameter identification.

B.3 Proof of Lemma 1

We omit the subscripts i and m to simplify notation. Suppose Assumption 4 holds.

(i) From the definition of $e^h\left(a,k,\omega_t,\omega_{t+1}^*\right)$,

$$
\begin{aligned}
E\left[e^h\left(a,k,\omega_t,\omega_{t+1}^*\right)|\mathcal{I}_t\right] &= E\left[\sum_{k'} e^h\left(k',\omega_t,\omega_{t+1}^*\right)F^k\left(k'|a,k,w_t\right)|\mathcal{I}_t\right] \\
&= E\left[\sum_{k'}\left(\int_{\omega'} h\left(k',\omega'\right)dF^\omega\left(\omega'|\omega_t\right) - h\left(k',\omega_{t+1}^*\right)\right)F^k\left(k'|a,k,w_t\right)|\mathcal{I}_t\right] \\
&= \sum_{k'}\int_{\omega'} h\left(k',\omega'\right)dF^\omega\left(\omega'|\omega_t\right)F^k\left(k'|a,k,w_t\right) \\
&\quad -\sum_{k'}\int_{\omega_{t+1}^*} h\left(k',\omega_{t+1}^*\right)dF^\omega\left(\omega_{t+1}^*|\omega_t\right)F^k\left(k'|a,k,w_t\right) \\
&= 0.
\end{aligned}
$$

(ii) By the law of iterated expectations,

$$
E\left[e^h\left(a,k,\omega_t,\omega_{t+1}^*\right)|z_t\right] = E\left[E\left[e^h\left(a,k,\omega_t,\omega_{t+1}^*\right)|\mathcal{I}_t\right]|z_t\right] = 0,
$$

where the second equality follows from (i).

Note also that, given that the time-t information set $\mathcal{I}_t$ includes current and past variables, Lemma 1 also implies that $E\left[e^h\left(a,k,\omega_t,\omega_{t+1}^*\right)|z_{t-d}\right] = 0$, for all a,k and any $d \geq 1$. (And in particular, $E\left[e^h\left(a,k,\omega_{t+d},\omega_{t+d+1}^*\right)|z_t\right] = 0$.)

(iii) Next, fix a and k, and simplify notation further by defining $e^h\left(a,k_t,\omega_t,\omega_{t+1}^*\right) = e_{t+1}^h$. Note that not only current and past states (k,ω) belong to the information set available to agents $\mathcal{I}_t$, but also past prediction errors. I.e., $\left\{e_t^h, e_{t-1}^h, ..., e_1^h\right\} \in \mathcal{I}_t$. We can then let $z_t = e_{t-d}^h$ for $d \geq 1$ and use result (ii) above to establish that $E\left[e_{t-d}^h e_t^h\right] = 0$. Thus, expectational errors are serially uncorrelated.

C Proofs: Estimation and Inference

C.1 Proof of Lemma 2

Given that $\{a_{imt}, k_{imt} : i = 1, ..., N\}$ are i.i.d. conditional on ω_{mt}, the first part of the Lemma (the almost sure convergence) follows by an immediate application of the Law of Large Numbers for exchangeable random variables (see Hall and Heyde (1980), p. 202, (7.1)).

The second part is obtained in three steps. First, Horvath and Yandell (1988) presents a Law of Iterated Logarithm (LIL) applied to both kernel and nearest neighbor estimators for conditional probabilities (see their Corollary 5.1). The i.i.d. sample in Horvath and Yandell (1988) can be replaced by the assumption that the sample is i.i.d. conditional on the common shocks following the arguments in Souza-Rodrigues (2016).[2] The LIL then holds for almost all ω_{mt}. Finally, it is straightforward to adapt the kernel regression results to simple frequency estimators (i.e., use simple indicator functions as kernel functions).

C.2 Proof of Proposition 3

Recall that $g_{mt}(\theta) = h(z_{mt})\, u_{mt}(\theta, \delta_{mt})$. Define the following functions:

$$\widetilde{g}_M(\theta) = \frac{1}{M(T-\tau)} \sum_{m=1,t=1}^{M,(T-\tau)} g_{mt}(\theta), \tag{C1}$$

and

$$\widetilde{Q}_M(\theta) = \widetilde{g}_M(\theta)' \mathbf{W}_M \widetilde{g}_M(\theta). \tag{C2}$$

The criterion function $\widetilde{Q}_M(\theta)$ is similar to $\widehat{Q}_M(\theta)$ but makes use of δ_{mt} instead of the estimator $\widehat{\delta}_{mt}$. I.e., $\widetilde{Q}_M(\theta)$ is an unfeasible GMM criterion function, while $\widehat{Q}_M(\theta)$ is feasible. The unfeasible estimator $\widetilde{\theta}_M$ (approximately) minimizes $\widetilde{Q}_M(\theta)$ over Θ.

A straightforward application of Theorem 2.6 in Newey and McFadden (1994) proves that the unfeasible estimator $\widetilde{\theta}_M$ is a consistent estimator of θ_0. To show that the feasible estimator $\widehat{\theta}_M$ is consistent as well, it suffices to show that $\widehat{Q}_M(\theta)$ converges in probability to $\widetilde{Q}_M(\theta)$ uniformly over Θ. To do so, define the difference $v_{mt} = \widehat{g}_{mt}(\theta) - g_{mt}(\theta)$, and

$$v_M(\theta) = \frac{1}{M(T-\tau)} \sum_{m=1,t=1}^{M,(T-\tau)} v_{mt}(\theta).$$

Then,

$$\begin{aligned} \widehat{Q}_M(\theta) &= [\widetilde{g}_M(\theta) + v_M(\theta)]' \mathbf{W}_M [\widetilde{g}_M(\theta) + v_M(\theta)] \\ &= \widetilde{Q}_M(\theta) + v_M(\theta)' \mathbf{W}_M v_M(\theta) + 2\widetilde{g}_M(\theta)' \mathbf{W}_M v_M(\theta). \end{aligned}$$

Given Condition 2(ii), it suffices to show that both $\widetilde{g}_M(\theta)$ and $v_M(\theta)$ converge to zero in probability uniformly over Θ.

By Conditions 2(i),(iii),(v), and (vi), $\widetilde{g}_M(\theta)$ satisfies the uniform Weak Law of Large Numbers, and therefore converges in probability to zero uniformly over Θ as $M \to \infty$. Now consider $v_M(\theta)$. Note that

$$v_{mt} = h(z_{mt}) \left(u_{mt}(\theta, \widehat{\delta}_{mt}) - u_{mt}(\theta, \delta_{mt}) \right),$$

[2] Souza-Rodrigues (2016) establishes the asymptotic properties of the kernel regression estimator for cross-sectional data in the presence of common shocks.

and take a mean-value expansion of $u_{mt}(\theta, \widehat{\delta}_{mt})$ about δ_{mt}:

$$u_{mt}(\theta, \widehat{\delta}_{mt}) - u_{mt}(\theta, \delta_{mt}) = \nabla_{\delta} u_{mt}(\theta, \delta^*_{mt}) \left(\widehat{\delta}_{mt} - \delta_{mt} \right),$$

where δ^*_{mt} lies between $\widehat{\delta}_{mt}$ and δ_{mt}. Next, note that

$$\begin{aligned} E\left[\sup_{\theta\in\Theta} \| v_M(\theta) \| \right] &\leq \frac{1}{M(T-\tau)} \sum_{m=1,t=1}^{M,(T-\tau)} E\left[\sup_{\theta\in\Theta} \| h(z_{mt}) \nabla_{\delta} u_{mt}(\theta, \delta^*_{mt}) \| \left\| \widehat{\delta}_{mt} - \delta_{mt} \right\| \right] \\ &\leq \frac{B}{M(T-\tau)} \sum_{m=1,t=1}^{M,(T-\tau)} E\left[\left\| \widehat{\delta}_{mt} - \delta_{mt} \right\|^2 \right] \end{aligned} \tag{C3}$$

where the second inequality follows from the Cauchy–Schwarz inequality and Condition 2(vii). Because $\left\| \widehat{\delta}_{mt} - \delta_{mt} \right\| \xrightarrow{p} 0$, as $N \to \infty$, by Lemma 2, we have that $E\left[\left\| \widehat{\delta}_{mt} - \delta_{mt} \right\|^2 \right] = o(1)$, and, so, the right-hand-side of (C3) converges to zero as $N \to \infty$ for all M and T. We conclude that

$$\sup_{\theta\in\Theta} \left\| \widehat{Q}_M(\theta) - \widetilde{Q}_M(\theta) \right\| \xrightarrow{p} 0, \ \ as \ \ (M, N) \to \infty.$$

C.3 Proof of Proposition 4.

By standard arguments (see Theorem 3.2 in Newey and McFadden (1994), the unfeasible estimator $\widetilde{\theta}_M$ satisfies

$$\widetilde{\theta}_M - \theta_0 = -\left[\mathbf{G}'\mathbf{W}\mathbf{G}\right]^{-1} \mathbf{G}'\mathbf{W} g(\theta_0) + o_p\left(1/\sqrt{M}\right), \tag{C4}$$

and is asymptotically normal,

$$\sqrt{M}\left(\widetilde{\theta}_M - \theta_0\right) \xrightarrow{p} N(0, \mathbf{V}),$$

under Conditions 3(i)-(iv). The asymptotic distribution of the feasible estimator $\widehat{\theta}_M$ is the same as the asymptotic distribution of the unfeasible $\widetilde{\theta}_M$ provided

$$\left\| \widehat{\theta}_M - \widetilde{\theta}_M \right\| = o_p\left(\frac{1}{\sqrt{M}} \right).$$

From (C4), it is clear that

$$\widetilde{\theta}_M - \widehat{\theta}_M = \left[\mathbf{G}'\mathbf{W}\mathbf{G}\right]^{-1} \mathbf{G}'\mathbf{W} v_M(\theta_0) + o_p\left(1/\sqrt{M}\right).$$

So,

$$\left\| \widehat{\theta}_M - \widetilde{\theta}_M \right\| \leq \left\| \left[\mathbf{G}'\mathbf{W}\mathbf{G}\right]^{-1} \right\| \|\mathbf{G}\| \|\mathbf{W}\| \|v_M(\theta_0)\| + o_p\left(1/\sqrt{M}\right).$$

Note that

$$E\left[\left\|v_M\left(\theta_0\right)\right\|\right] \leq \frac{B}{M\left(T-\tau\right)} \sum_{m=1,t=1}^{M,(T-\tau)} \left(E\left[\left\|\widehat{\delta}_{mt}-\delta_{mt}\right\|^2\right]\right)^{1/2}$$

by Condition 2(vii). Because $E\left[\left\|\widehat{\delta}_{mt}-\delta_{mt}\right\|^2\right] = O\left(\frac{\log\log N}{N}\right)$, by Lemma 2, we have that $\|v_M(\theta_0)\| = O_p\left(\sqrt{\frac{\log\log N}{N}}\right)$, which implies

$$\sqrt{M}\left\|\widehat{\theta}_M - \widetilde{\theta}_M\right\| = O_p\left(\sqrt{\frac{M\log\log N}{N}}\right) = o_p\left(1\right),$$

provided $\frac{M\log\log N}{N} \to 0$.

C.4 Additional Result

The next lemma provides a result that is used in Proposition 1. Proposition 1 claims that, for a known function f of $\delta_{mt}^{\tau} = (\delta_{mt}, ..., \delta_{mt+\tau})$, quantities of the type $E\left[h\left(z_{mt}\right) f\left(\delta_{mt}^{\tau}\right)\right]$ can be recovered from the data. (More specifically, $f\left(\delta_{mt}^{\tau}\right)$ in the proof of Proposition 1 corresponds to the term in parenthesis on the LHS of equation (B4).)

Lemma C1. *Suppose the vector (w_{mt}, z_{mt}) is i.i.d. across markets m. Assume*

$$E\left[\left\|h\left(z_{mt}\right)\nabla_{\delta} f\left(\delta_{mt}^{\tau}\right)\right\|^2\right] \leq C < \infty.$$

Then

$$\frac{1}{M}\sum_{m=1}^{M} h\left(z_{mt}\right) f(\widehat{\delta}_{mt}^{\tau}) \xrightarrow{p} E\left[h\left(z_{mt}\right) f(\delta_{mt}^{\tau})\right],$$

as $(M, N) \to \infty$.[3]

Proof. First, note that

$$\frac{1}{M}\sum_{m=1}^{M} h\left(z_{mt}\right) f(\widehat{\delta}_{mt}^{\tau}) = \frac{1}{M}\sum_{m=1}^{M} h\left(z_{mt}\right) f\left(\delta_{mt}^{\tau}\right) + \frac{1}{M}\sum_{m=1}^{M} h\left(z_{mt}\right)\left[f(\widehat{\delta}_{mt}^{\tau}) - f\left(\delta_{mt}^{\tau}\right)\right].$$

The first term on the right-hand-side converges in probability to $E\left[h\left(z_{mt}\right) f\left(\delta_m\right)\right]$ as $M \to \infty$ by the Weak Law of Large Numbers. Applying a mean-value expansion on the second term, we get

$$\frac{1}{M}\sum_{m=1}^{M} h\left(z_{mt}\right)\left[f(\widehat{\delta}_{mt}^{\tau}) - f\left(\delta_{mt}^{\tau}\right)\right] = \frac{1}{M}\sum_{m=1}^{M} h\left(z_{mt}\right)\nabla_{\delta} f\left(\delta_{mt}^{\tau *}\right)\left[\widehat{\delta}_{mt}^{\tau} - \delta_{mt}^{\tau}\right]$$

[3]The same result applies if (w_{mt}, z_{mt}) is stationary and ergodic, if we average the term $[h\left(z_{mt}\right) f(\widehat{\delta}_{mt}^{\tau})]$ over $T-\tau$ time periods, and if we take $(T, N) \to \infty$.

where $\delta_{mt}^{\tau*}$ lies between $\widehat{\delta}_{mt}^{\tau}$ and δ_{mt}^{τ}. Next, note that

$$\begin{aligned} E\left[\left\|h\left(z_{mt}\right)\nabla_{\delta}f(\delta_{mt}^{\tau*})\left[\widehat{\delta}_{mt}^{\tau}-\delta_{mt}^{\tau}\right]\right\|\right] &\leq \left(E\left[\left\|h\left(z_{mt}\right)\nabla_{\delta}f\left(\delta_{mt}^{\tau*}\right)\right\|^{2}\right]E\left[\left\|\widehat{\delta}_{mt}^{\tau}-\delta_{mt}^{\tau}\right\|^{2}\right]\right)^{1/2} \\ &\leq C\left(E\left[\left\|\widehat{\delta}_{mt}^{\tau}-\delta_{mt}^{\tau}\right\|^{2}\right]\right)^{1/2}, \end{aligned}$$

where the first inequality follows from the Cauchy–Schwarz inequality, and the second inequality from the regularity condition $E\left[\left\|h\left(z_{mt}\right)\nabla_{\delta}f\left(\delta_{mt}^{\tau}\right)\right\|^{2}\right]\leq C<\infty$. By Lemma 2, $E\left[\left\|\widehat{\delta}_{mt}^{\tau}-\delta_{mt}^{\tau}\right\|^{2}\right]$ converges to zero as $N\to\infty$, which implies

$$\frac{1}{M}\sum_{m=1}^{M}h\left(z_{mt}\right)\left[f(\widehat{\delta}_{mt}^{\tau})-f\left(\delta_{mt}^{\tau}\right)\right]\xrightarrow{p}0,\ \text{ as } N\to\infty, \text{ for all } M.$$

∎

D The Standard CCP Estimator

Here we explain the standard CCP approach implemented in the Monte Carlo experiment to estimate the model parameters. By "standard," we mean involving a full specification of how all state variables evolve, and not relying on Euler equations. Following Hotz and Miller (1993), this CCP approach avoids the computational burden of solving the dynamic problem within the estimation algorithm associated with Rust's (1987) nested fixed point approach.

The estimation here follows Section 2 of Kalouptsidi, Scott, and Souza-Rodrigues (2019) and we refer readers to it for details. Estimation begins by estimating choice probabilities conditional on individual states and the modeled exogenous state variable, i.e., $p\left(k,w\right)$. Let F_b represent the stochastic matrix for observable state variables (k,w) conditional on buying the product, and let F_{nb} represent the stochastic matrix when the action is not buying the product. Kalouptsidi, Scott, and Souza-Rodrigues (2019) shows that

$$\pi_b = A\pi_{nb} + \boldsymbol{b},$$

where $A=\left(I-\beta F_b\right)\left(I-\beta F_{nb}\right)^{-1}$ and $\boldsymbol{b}=A\psi_{nb}-\psi_b$, where ψ_a stacks $\psi_a\left(p\left(k,w\right)\right)$ across all values of (k,w).

We estimate the payoff parameters θ using a Minimum Distance estimator, i.e., by minimizing the L2 norm of

$$\pi_b\left(\theta\right)-A\pi_{nb}\left(\theta\right)-\boldsymbol{b}.$$

Given the parameterization, this is achieved by a linear regression of the vector $\boldsymbol{b}$ on the matrix

$$\left[\begin{array}{cc} (\mathbf{1} - A\boldsymbol{k}) & , \boldsymbol{w} \end{array}\right],$$

where $\mathbf{1}$ is a vector of ones, $\boldsymbol{k}$ is a dummy vector equal to one in states where the good is owned, and $\boldsymbol{w}$ is the vector of prices.[4]

E Monte Carlo: Counterfactual

In the Monte Carlo presented in Section 6, we consider only the estimation of the parameters of agents' utility function. Typically, applied researchers are also interested in the outcomes of policy simulations or counterfactuals. In this section, we extend the Monte Carlo experiment in Section 6 to study counterfactual simulations. Specifically, we consider how the biases in the parameter estimates pass through to biases in countefactuals. Before doing so, we must consider the question of how to do counterfactuals within the ECCP framework.

Much of the ECCP approach's appeal comes from the fact that it takes seriously the possibility that the econometrician might be facing important measurement issues; e.g., some market-level state variables might not be observed, and/or it might be difficult to specify how they evolve. However, when doing counterfactuals, researchers typically solve for a new equilibrium of the model, which normally involves fully specifying all the relevant state variables and how they evolve. Thus, *prima facie*, ECCP estimation seems to be at odds with doing counterfactual simulations.

A counterfactual is a function of the model parameters, and sometimes that function does not depend (or depends only minimally) on the presence of unobservable variables or on the precise specification of how state variables evolve. Therefore, the modeling issues that motivate the ECCP approach need not undermine the use of parameter estimates for counterfactual analysis. De Groote and Verboven (2019) provide a clear example. They use an ECCP estimator to estimate the rate of time discounting of Belgian households in deciding whether to install solar photovoltaic systems (the ECCP estimator allows them to flexibly include demand shocks and avoid specifying a process for how government policy evolved). They find that households' estimated discount rate is considerably lower than the interest rate that the Belgian government can borrow at. As they argue, this disparity means that it would be more cost effective for the government to support solar PV installations with up-front payments, rather than the ongoing payments that the government actually used. This conclusion follows intuitively from the disparity in discount rates and plausibly is not affected in an important way by how government policy and unobservable states evolve. The conclusion, however, may be highly sensitive to biases in the *estimation* of the discount factor. In

[4] Note that one can estimate the model parameters either by minimizing the distance between $\boldsymbol{b}$ and $\pi_b(\theta) - A\pi_{nb}(\theta)$, or by minimizing the distance between the (nonparametrically) estimated CCP, p, and the CCP generated by the model, $p(\theta)$. See Pesendorfer and Schmidt-Dengler (2008).

other words, the estimation of a mis-specified model may crucially affect policy recommendations.

In what follows, we show that some counterfactuals – specifically, long-run demand elasticities from our durable good demand model – are robust to the omission of unobservable state variables that are present in the data generating process. Or more specifically, long-run demand elasticities are well approximated by a model in which we set ξ at its long-run mean (i.e., $\xi = 0$). Furthermore, we show that the biases that result from leaving the unobservable shocks out of the counterfactual simulations can be smaller than the biases that result from using an estimation approach that is not robust to their presence.

We perform both a *real* and *feasible* counterfactuals for the durable good demand model. Our *real* counterfactuals take the parameter estimates from various estimations presented in Section 6 and plug them into a counterfactual that uses the true data generating process (notably including the true law of motion for the unobservable demand shock ξ_{mt}). That is, the real counterfactuals rely on our understanding of an unobservable that an econometrician who was not simulating the data would not have access to. Our *feasible* counterfactuals, in contrast, simulate a simple model that an econometrician could easily implement: a model that sets $\xi = 0$.[5]

The counterfactual we consider is an increase in the mean price level (formally, we increase γ_0 by .01; see Table 1 in Section 6), and we calculate the long-run change in the demand level. That is, we calculate the unconditional probability of purchase $Pr\,(a = b)$ in the steady state after solving the consumer's dynamic problem. We present this counterfactual in the form of a *long-run demand elasticity*, i.e., the ratio of the percentage change in the probability of purchase to the percentage change in the long-run price.

Table E1 shows the counterfactuals from the ECCP (OLS and IV) and standard CCP estimators based on the parameter estimates from the above simulations with $M = 160$ and $T = 160$. A first observation is that the real and feasible counterfactuals at the true parameters differ by a factor of about 10%. Second, consistent with the biases in the underlying parameter estimates, we find that the ECCP IV estimates yield very little bias in the counterfactuals relative to the true values while the other estimators result in substantially biased counterfactuals. Furthermore, the biases in the long-run elasticities from the OLS and standard CCP estimators (whether we consider the real or feasible versions) are larger than the gap between the real and feasible estimators.

Evidently, counterfactuals are not always robust to setting ξ at its unconditional mean. The broader point we make in this section is that robustness to the presence of unobserved shocks can be assessed through a procedure similar to what we do here. That is, when researchers are concerned about the presence of unobservables, they might adopt a robust estimation approach that delivers consistent estimates of important parameters despite the unobservables. Then, when it comes to counterfactual simulations, they can perform the simulation in several ways to assess

[5]To solve the feasible counterfactual, we need to specify how the exogenous state variable w_{mt} evolves. We consider the residual from the pricing equation (39) as the econometrician can measure it. I.e., $w = \gamma_0 + \gamma_1 z + \nu$, where $\nu = \gamma_2 \xi + \varepsilon^w$. So, we calculate the true evolution of ν given the underlying processes and assume the econometrician is able to estimate it.

whether and how the results of interest might be sensitive to the presence of unobservables and how they evolve.

Table E1: Sample size, structure and bias

	True value		ECCP OLS	ECCP IV	Standard CCP
Real LRE	-1.106	Mean Estimate	60.15	-1.104	0.01471
		Relative Bias	-5540%	-0.1561%	-101.3%
		SD	16.62	0.04227	0.02545
		RMSE	63.48	0.04231	1.121
Feasible LRE	-1.022	Mean Estimate	-1.187e4	-1.064	0.03888
		Relative Bias	1.162e6%	4.114%	-103.8%
		SD	1.382e6	0.1184	0.06774
		RMSE	1.382e6	0.1256	1.063

Notes: 5000 replications with sample structure $M = T = 160$. SD is the standard deviation across replications. RMSE is root-mean squared error. Relative Bias is bias as percentage of the true parameter.

References

Arcidiacono, P., and R. A. Miller (2019): "Nonstationary Dynamic Models with Finite Dependence," *Quantitative Economics*, 10, 853–890.

De Groote, O., and F. Verboven (2019): "Subsidies and Time Discounting in New Technology Adoption: Evidence from Solar Photovoltaic Systems," *American Economic Review*, 109(6), 2137–72.

Hall, P., and C. C. Heyde (1980): *Martingale limit theory and its application*. Academic press.

Horvath, L., and B. S. Yandell (1988): "Asymptotics of conditional empirical processes," *Journal of Multivariate Analysis*, 26(2), 184–206.

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.

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

Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, 4, 2111–2245.

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

Ransom, T. (2019): "Labor Market Frictions and Moving Costs of the Employed and Unemployed," Discussion Paper 12139, IZA Institute of Labor Economics.

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

Souza-Rodrigues, E. A. (2016): "Nonparametric Regression with Common Shocks," *Econometrics*, 4(3), 36.