Semi-Parametric Estimation of Counterfactuals in Dynamic Discrete Choice Models

David Hao Zhang*

Harvard University

June 16, 2018

Abstract

I develop a new method for estimating counterfactuals in dynamic discrete choice models, a widely used set of models in economics, without requiring a distributional assumption on utility shocks. Applying my method to the canonical Rust (1987) setting, I find that the typical logit assumption on utility shocks can lead the researcher to conclude that the agent's counterfactual choice probabilities are much more sensitive than what a semi-parametric model would suggest. Therefore, my method may be useful to applied researchers in generating policy counterfactuals that are robust to such distributional assumptions.

*I thank Elie Tamer, Ariel Pakes, Myrto Kalouptsidi, Gaston Illanes, Robin Lee, Nicola Rosaia, Jiacheng Feng, Ashvin Gandhi, Jun Ishii, and Jeffrey Wang, as well as participants at the 2018 International Industrial Organization Conference (IIOC) for helpful comments and discussions. All errors are my own.

1 Introduction and Literature

Dynamic discrete choice models in the style of Rust (1987) are widely used in industrial organization, labor economics, and health economics to model consumer and firm behavior.[1] Yet, commonly used methods for estimating these models such as Rust (1987), Hotz and Miller (1993), and Arcidiacono and Miller (2011) require the researcher to assume a parametric distribution on utility shocks in order to recover the mean utility from choosing an action. Such an assumption is not innocuous in that it impacts the counterfactual predictions of the model. This paper proposes a new methodology for estimating counterfactuals in dynamic discrete choice models without such a distributional assumption, and thus producing counterfactuals that are robust to it.

More philosophically, in the Rust (1987) setting, the mean of engine maintenance costs at various mileages and the distribution of engine maintenance costs around the mean both matter to the agent for when to replace the engine. Different assumptions on how the engine repair costs are distributed can imply different optimal replacement strategies and different responses to policy changes such as taxes or subsidies on engine replacement. Hence, it may not make sense to assume a distribution of engine repair costs but estimate the mean, since both can be considered "primitives" of the model. I apply my proposed methodology to the Rust (1987) setting, and find that the distribution of engine maintenance costs appear somewhat bimodal, leading me to predict significantly attenuated responses to a change in the counterfactual compared to a standard model based on the logit assumption.

Several other papers proposes methodology for the semi-parametric estimation of dynamic discrete choice models, but they lack a computationally tractable way of estimating counterfactuals. Norets and Tang (2014) enumerates a set of necessary and sufficient conditions for a set of conditional choice probabilities (CCPs) to be consistent with a set of model parameters in the discrete state space setting. However, to get the identified set of counterfactuals their method requires the researcher to check whether each possible set of counterfactual choice probabilities is accepted or rejected, which is problematic if the state space is large. More specifically, suppose the state can take 90 possible values, as in Rust (1987), their identification result requires the researcher to check each point of $p^{cf} \in [0,1]^{90}$ for inclusion in the counterfactual, which is intractable through a grid search[2] and difficult to get convergence in MCMC. More recent applications of the dynamic discrete choice model typically involves larger state spaces for which the problem is more severe. Buchholz, Shum, and Xu (2017) shows point identification of the quantile function of utility shocks within a range of CCPs in dynamic binary choice up to an unknown mean and scale factor under given at least one continuous state and some invertibility conditions, but does not give a way of getting counterfactuals without making a further distributional assumption on

[1] See Eckstein and Wolpin (1989), Rust (1994), Pakes (1994), Miller (1999), Aguirregabiria and Mira (2010), Keane, Todd, and Wolpin (2011) for surveys of papers using the dynamic discrete choice model and Shiraldi (2011), Gowrisankaran and Rysman (2012), Lee (2013), Dillon and Stanton (2017), and Agarwal et al. (2018) for some more recent examples.

[2] Even if each state is gridded into 100 points (one for each percentage point), this still involves 100^{90} checks.

the quantile function of utility shocks outside of the observed range of CCPs. Aguirregabiria (2010) shows the non-parametric identification of a class of finite-horizon dynamic binary choice models, but his argument does not apply for infinite horizon problems. Blevins (2014) and Chen (2017) focuses on identification in dynamic choice problems under exclusion restrictions, but does not give a method for getting counterfactuals.

The approach I propose is based on a transformation of the dynamic discrete choice problem from the original action space to a different action space, where the actions are a choice of CCPs and the payoffs incorporate the conditional expectations of the utility shocks given the choice probabilities. Furthermore, I develop a method for obtaining sharp bounds on the required conditional expectations for binary choice problems using either a linear program in the discrete state space case based on Norets and Tang (2014) or a closed-form estimator in the continuous state space case based on Buchholz, Shum, and Xu (2017). For the multivariate choice case where the states are discrete, I also develop non-sharp bounds that can still be computed via a linear program.[3] Therefore, the approaches of Norets and Tang (2014) and Buchholz, Shum, and Xu (2017) are complementary to the one in this paper in that this paper shows how to estimate counterfactuals based on parameter estimates from their procedures in a computationally tractable way.

The proposed method has several drawbacks. First, I sacrifice sharpness on the identified set of counterfactuals in order to gain computational tractability. Nevertheless, I show that the bounds for counterfactuals are still reasonably tight in the Rust (1987) setting, so my method may be useful for applied researchers who want to bound counterfactuals in a way that is robust to distributional assumptions. Second, I grid search over the parameters that characterize the flow payoffs in the discrete state space case, which may be computationally difficult if the flow payoffs are themselves characterized by a large number of parameters. However, we need only to grid over two parameters (the mean and the scale) if at least one continuous state variable is available and the assumptions in Buchholz, Shum, and Xu (2017) are satisfied.

The rest of the paper is organized as follows. Section 2 goes over the model setup and the transformation of the choice problem. Section 3 goes over the estimation method in both the discrete and continuous state case. Section 4 applies the method to real data on engine replacement from Rust (1987). Section 5 concludes.

2 Model setup and main results

I first define a standard dynamic discrete choice model. There exists a set of states $x \in X$ on which a finite set of actions $a \in A$ yields expected payoffs $\pi_a(x) \in \mathbb{R}$ plus an additively separable utility shock

[3] Alternatively, one can use a Norets (2011) approach to develop sharp bounds in an NP-hard procedure.

$\epsilon_a \in \mathbb{R}$ which applies to all non-0th actions.[4]

I then make the following assumptions on the state transitions and $\epsilon = \{\epsilon_a : a \in A\backslash\{0\}\}$.

Assumption 1. *(Markov conditional on observables) State transitions are Markov, being governed by transition probabilities* $f_{x',\epsilon'|x,\epsilon,a}$. *Furthermore, the state transitions depend only on observables, such that* $f_{x',\epsilon'|x,\epsilon,a} = f_{x',\epsilon'|x,a}$.

Assumption 2. *(Independence) In each period, the realized* ϵ_a *is an i.i.d. draw from a distribution with PDF* g_a. *Hence,* $f_{x',\epsilon'|x,a} = f_{x'|x,a}g_a(\epsilon')$.

Assumption 3. *(Full support)* ϵ *has full support, such that* $g(\epsilon) > 0, \forall \epsilon \in \mathbb{R}^{|A|-1}$.

Assumptions 1, 2, and 3 are strong but standard in the literature. In this framework, it would also be possible to relax Assumption 2 by allowing the distribution g_a to depend on the state x. Assumptions 1 and 2 allows us to write the choice problem as:

$$a = \arg\max_{a\in A} \pi_a(x) + \epsilon_a + \beta \int f_{x'|x,a} EV(x')dx' \tag{1}$$

$$EV(x) = E \max_{a\in A} \pi_a(x) + \epsilon_a + \beta \int f_{x'|x,a} EV(x')dx' \tag{2}$$

$$v_a(x) = \pi_a(x) + \beta \int f_{x'|x,a} EV(x')dx' \tag{3}$$

Where $EV(x)$ are the expected value functions and $v_a(x)$ are the choice-specific value functions. Note that I have not yet made any assumptions on whether X is discrete or continuous, and the integral collapses down to matrix multiplication if X is discrete and finite.

Assumption 3 then enables us to use the Hotz and Miller (1993) inversion. Gven a set of choice-specific value functions $v(x) = \{v_a(x), a \in A\}$, we can define a mapping a mapping $\mathcal{F} : \mathbb{R}^{|A|-1} \to (0,1)^{|A|-1}$ such that $p(x) = \mathcal{F}(v(x))$, where $p_0 = 1 - \sum_{a=1}^{|A|} p_a$ is excluded. Then, the Hotz and Miller (1993) inversion implies that $\mathcal{F}$ is invertible, so that $\mathcal{F}^{-1}$ exists.

Furthermore, I define $\xi_a(x)$ as the expected value of ϵ_a conditional on selection into action a at state x, multiplied by the probability of that happening:

$$\xi_a(v(x)) = \Pr(v_a(x) + \epsilon_a \geq v_{a'}(x) + \epsilon_{a'}, \forall a' \in A) E(\epsilon_a | v_a(x) + \epsilon_a \geq v_{a'}(x) + \epsilon_{a'}, \forall a' \in A) \tag{4}$$

Intuitively, $\xi_a(v(x))$ is term that describes selection on the utility shocks: an action is chosen only when the utility shock for that action is high relative to the mean utilities given the state and the utility shock for other actions.

Then, I obtain the following lemma.

[4] Shocks on the 0th action can be normalized to 0 by subtracting it from shocks on other actions.

Lemma 1. *Under Assumptions 1-3, the following maximization problems are equivalent.*

$$EV(x) = E \max_{a \in A} v_a(x) + \epsilon_a = \max_{\{p(x)\}} \sum_{a \in A} p_a(x) v_a(x) + \xi_a(\mathcal{F}^{-1}(p(x))) \tag{5}$$

$$= \max_{\{p(x)\}} \sum_{a \in A} p_a(x) \left(\pi_a(x) + \beta \int f_{x'|x,a} EV(x') dx' \right) + \xi_a(\mathcal{F}^{-1}(p(x))) \tag{6}$$

Proof. See Appendix A.1 □

Lemma 1 implies that the expected value function $EV(x)$ of the stochastic choice problem can be re-cast into a choice problem where $p(x)$ are taken as actions instead. In other words, given the flow payoffs π, the transition probabilities $f_{x'|x,a}$, and a selection function ξ, one can compute the expected value function $EV(x|\pi, f_{x'|x,a\in A}, \xi)$.

Now, we consider bounding counterfactual CCPs given a set of counterfactual primitives $\tilde{\pi}, \tilde{f}_{x'|x,a}, \tilde{\xi}$. These counterfactual primitives can be functions of the original primitives $\pi, f_{x'|x,a}, \xi$. Note that in this paper, we take the counterfactuals as given and focus on estimating it, and refer readers to Norets and Tang (2014) and Kalouptsidi, Scott, and Souza-Rodrigues (2017) for identification of counterfactuals under normalization.

The rest of the results in this section makes use of Lemma 1 along with $\tilde{\pi}, \tilde{f}_{x'|x,a}$ and bounds on $\tilde{\xi}$ and on the CDF of differences $\epsilon_a - \epsilon_{a'}$, $G_{a,a'}$ to get bounds on counterfactuals. I will show in Section 3 how one can obtain the estimates of these payoffs, transitions, and bounds using methods based on either Norets and Tang (2014) or Buchholz, Shum, and Xu (2017). For now, I take these as given, an assumption expressed in the following Assumption 4.

Assumption 4. *Suppose that $\tilde{\pi}, \tilde{f}_{x'|x,a}$ is known, and that there exists bounds $\underline{\xi}(v), \overline{\xi}(v)$ and $\underline{G}_{a,a'}(v_a - v_{a'}), \overline{G}_{a,a'}(v_a - v_{a'})$ such that $\underline{\xi}(v) \leq \tilde{\xi}(v) \leq \overline{\xi}(v)$ and $\underline{G}_{a,a'}(v_a - v_{a'}) \leq G_{a,a'}(v_a - v_{a'}) \leq \overline{G}_{a,a'}(v_a - v_{a'})$, $\forall v \in \mathbb{R}^{|A|-1}$ and $v_a, v'_a \in \mathbb{R}$.*

The general strategy that Lemma 1 enables, given Assumption 4, is to first obtain bounds on differences in choice specific value functions in the counterfactual $\tilde{\delta}_{a,a'}$, which is defined based on the choice specific value functions $\tilde{v}_a, \tilde{v}_{a'}$ as:

$$\tilde{\delta}_{a,a'}(x) = \tilde{v}_a(x) - \tilde{v}_{a'}(x) \tag{7}$$

$$= \tilde{\pi}_{a,a'}(x) - \tilde{\pi}_{a'}(x) + \beta \int \left(\tilde{f}_{x'|x,a} - \tilde{f}_{x'|x,a'} \right) EV(x'|\tilde{\xi}, \tilde{\pi}, \tilde{f}_{x'|x,a\in A}) dx' \tag{8}$$

Based on this definition, one way to construct bounds on the differences in choice-specific value functions is:

$$\underline{\delta}_{a,a'}(x) = \tilde{\pi}_a(x) - \tilde{\pi}_{a'}(x) + \min_{\hat{\xi}\in[\underline{\xi},\overline{\xi}]} \beta \int \left(\tilde{f}_{x'|x,a} - \tilde{f}_{x'|x,a'}\right) EV(x'|\hat{\xi},\tilde{\pi},\tilde{f}_{x'|x,a\in A})dx' \leq \tilde{\delta}_a \tag{9}$$

$$\overline{\delta}_{a,a'}(x) = \tilde{\pi}_a(x) - \tilde{\pi}_{a'}(x) + \max_{\hat{\xi}\in[\underline{\xi},\overline{\xi}]} \beta \int \left(\tilde{f}_{x'|x,a} - \tilde{f}_{x'|x,a'}\right) EV(x'|\hat{\xi},\tilde{\pi},\tilde{f}_{x'|x,a\in A})dx' \geq \tilde{\delta}_a \tag{10}$$

While elegant, these quantities cannot be computed easily because the function to be minimized or maximized is not guaranteed to be convex. The next results in this section, in particular Corollaries 1-4, are all about getting more easily computable bunds on $\underline{\delta}_a, \overline{\delta}_a$. Once we have this, we can be translate them into bounds on counterfactual CCPs:

$$\prod_{a'\in A\setminus\{a\}} 1 - \overline{G}_{a,a'}(\overline{\delta}_{a,a'}) \leq \tilde{p}_a(x) \leq \prod_{a'\in A\setminus\{a\}} 1 - \underline{G}_{a,a'}(\underline{\delta}_{a,a'}) \tag{11}$$

Empirically, I find that uncertainly in the CDF of the unobserved variable (that is, the gap $\overline{G}_{a,a'} - \underline{G}_{a,a'}$) in Equation 11 does not contribute much to uncertainty in bounds in the Rust (1987) case, because the bounds on the CDF is very tight, and in fact we can get point identification of G within the observed range of CCPs if the assumptions in Buchholz, Shum, and Xu (2017) are satisfied. The main contributor to the width of the bounds in counterfactuals is the difference between $\overline{\delta}_a$ and $\underline{\delta}_a$, which is why I focus the rest of the section on them.

Inspecting Lemma 1 while noting that $\tilde{f}_{x'|x,a} > 0$ immediately implies the following corollary.

Corollary 1. *Let* $\underline{EV} = EV(x'|\underline{\xi},\tilde{\pi},\tilde{f}_{x'|x,a})$ *and* $\overline{EV} = EV(x'|\overline{\xi},\tilde{\pi},\tilde{f}_{x'|x,a})$*. Then,*

$$\underline{\delta}_{a,a'}(x) = \tilde{\pi}_a(x) - \tilde{\pi}_{a'}(x) + \min_{\hat{EV}\in[\underline{EV},\overline{EV}]} \beta \int \left(\tilde{f}_{x'|x,a} - \tilde{f}_{x'|x,a'}\right) \hat{EV}(x')dx' \leq \tilde{\delta}_a \tag{12}$$

$$\overline{\delta}_{a,a'}(x) = \tilde{\pi}_a(x) - \tilde{\pi}_{a'}(x) + \max_{\hat{EV}\in[\underline{EV},\overline{EV}]} \beta \int \left(\tilde{f}_{x'|x,a} - \tilde{f}_{x'|x,a'}\right) \hat{EV}(x')dx' \geq \tilde{\delta}_a \tag{13}$$

Proof. Via noting EV in Lemma 1 is monotonic in ξ. □

The bounds in Corollary 1 are easily computable: they only involve doing two value function iterations to get $\underline{EV}, \overline{EV}$, and computing the minimum or maximum of the integral simply involves setting the $\hat{EV}$ to either its minimum or maximum depending on the sign of $\tilde{f}_{x'|x,a} - \tilde{f}_{x'|x,a'}$. Hence, Corollary 1 combined with Equation 11 already yields a very efficient methodology for computing bounds on counterfactuals. This approach yields fairly tight bounds if the discount factor β is relatively low, such as $\beta = .9$ in the Rust (1987) case, as shown in Figure 5. It should also work well if the action is terminal, based on Corollary 2. Otherwise, the bounds be loose, the main reason being that $\hat{EV} \in [\underline{EV}, \overline{EV}]$ is a rather loose condition that may not result in a $\hat{EV}$ that is consistent with any particular $\hat{\xi} \in [\underline{\xi}, \overline{\xi}]$. The following Lemma 2 develops an approach to address this problem.

Lemma 2. *Allow* $\hat{\xi}_a$ *to vary by state, so that* $\hat{\xi}_a(x) \in [\underline{\xi}, \overline{\xi}], \forall x \in X$*. Let* u *be a vector that is equal* *to 1 for state x and zero for every other state. Then, setting $\hat{\xi}_a(x) = \underline{\xi}_a, \hat{\xi}_{a'}(x) = \overline{\xi}_{a'}$ minimizes (and $\hat{\xi}_a(x) = \overline{\xi}_a, \hat{\xi}_{a'}(x) = \underline{\xi}_{a'}$ maximizes) $\hat{\delta}_{a,a'}(x')$ if:*

$$\min_{p \in \mathcal{P}} \int_{x''} (\tilde{f}_{x''|x',a} - \tilde{f}_{x''|x',a'}) ev(p, u(x), \beta) dx'' \geq 0 \tag{14}$$

Where $\mathcal{P}$ is the space of counterfactual CCPs which may be $[0,1]^{|A|-1}$, and ev is the expected value of the dynamic choice problem with payoffs u. The converse is also true, where setting $\hat{\xi}_a(x) = \overline{\xi}_a, \hat{\xi}_{a'}(x) = \underline{\xi}_{a'}$ minimizes (and $\hat{\xi}_a(x) = \underline{\xi}_a, \hat{\xi}_{a'}(x) = \overline{\xi}_{a'}$ maximizes) $\hat{\delta}_{a,a'}(x')$ if:

$$\max_{p \in \mathcal{P}} \int_{x''} (\tilde{f}_{x''|x',a} - \tilde{f}_{x''|x',a'}) ev(p, u(x), \beta) dx'' \leq 0 \tag{15}$$

The intuition behind Lemma 2 is that it checks whether for any set of possible counterfactual CCPs, state x might be more important to the agent when taking action a as opposed to action a' in state x'. While not easily checkable by itself (because the condition is again non-convex), Lemma 2 allows us to develop several corollaries which are useful in computation. The first Corollary 2 concerns actions which are terminal for some action a in state x such the problem ends afterwards.

Corollary 2. *Suppose an action a' is terminal in state $x' \in X$, so that $\tilde{f}_{x''|x',a'} = 0$. Then, setting $\hat{\xi}_a(x) = \underline{\xi}_a, \hat{\xi}_{a'}(x) = \overline{\xi}_{a'}$ minimizes (and $\hat{\xi}_a(x) = \overline{\xi}_a, \hat{\xi}_{a'}(x) = \underline{\xi}_{a'}$ maximizes) $\hat{\delta}_{a,a'}(x')$ for all $x \in X$.*

Proof. By noticing that $ev(p, u, \beta) \geq 0$ so Equation 16 is always satisfied. □

Therefore, Corollary 2 implies that for terminal actions, the approach in Corollary 1 already gives the tightest possible bounds on $\tilde{\delta}_{a,a'}$ based on Lemma 2, and no further improvements can be made without exploiting more structure in the bounds on ξ, G.

The next Corollary 3 helps determine how to select $\xi_a(x)$ to bound $\tilde{\delta}_{a,a'}(x')$ in the case where a state cannot be reached except through some other states. In particular, Corollary 3 can be applied to the Rust (1987) case by noticing that an engine cannot lose mileage except through replacement, so it pins down the set of $\xi_a(x)$ to be applied for all mileages that comes before the current mileage.

Corollary 3. *Suppose a state $x \in X$ cannot be reached at any point in the future from $f_{x''|x',a'}$ except through the transition probabilities $f_{x''|x',a}$. Then, setting $\hat{\xi}_a(x) = \underline{\xi}_a, \hat{\xi}_{a'}(x) = \overline{\xi}_{a'}$ minimizes (and $\hat{\xi}_a(x) = \overline{\xi}_a, \hat{\xi}_{a'}(x) = \underline{\xi}_{a'}$ maximizes) $\hat{\delta}_{a,a'}(x')$.*

Proof. By noticing that in this case, the value of $\int_{x''} \tilde{f}_{x''|x',a} ev(p, u(x), \beta) dx''$ will be weakly greater than the value of $\int_{x''} \tilde{f}_{x''|x',a'} ev(p, u(x), \beta) dx''$ regardless of p. □

Finally, we can break down the components of Lemma 2 into separate dynamic programming problems, which leads to the following Corollary 4 which applies to all problems but is more effective when there are already some bounds on the counterfactual CCPs $\tilde{p}$.

Corollary 4. *Let u be a vector that is equal to 1 for state x and zero for every other state. Then, setting $\hat{\xi}_a(x) = \underline{\xi}_a, \hat{\xi}_{a'}(x) = \overline{\xi}_{a'}$ minimizes (and $\hat{\xi}_a(x) = \overline{\xi}_a, \hat{\xi}_{a'}(x) = \underline{\xi}_{a'}$ maximizes) $\hat{\delta}_{a,a'}(x')$ if:*

$$\min_{p\in\mathcal{P}} \int_{x''} \tilde{f}_{x''|x',a} ev(p, u(x), \beta) dx'' - \max_{p\in\mathcal{P}} \int_{x''} \tilde{f}_{x''|x',a'} ev(p, u(x), \beta) dx'' \geq 0 \tag{16}$$

Where $\mathcal{P}$ is the space of counterfactual CCPs which may be $[0,1]^{|A|-1}$, and ev is the expected value of the dynamic choice problem with payoffs u. The converse is also true, where setting $\hat{\xi}_a(x) = \overline{\xi}_a, \hat{\xi}_{a'}(x) = \underline{\xi}_{a'}$ minimizes (and $\hat{\xi}_a(x) = \underline{\xi}_a, \hat{\xi}_{a'}(x) = \overline{\xi}_{a'}$ maximizes) $\hat{\delta}_{a,a'}(x')$ if:

$$\max_{p\in\mathcal{P}} \int_{x''} \tilde{f}_{x''|x',a} \tilde{f}_{x''|x',a'}) ev(p, u(x), \beta) dx'' - \min_{p\in\mathcal{P}} \int_{x''} \tilde{f}_{x''|x',a'}) ev(p, u(x), \beta) dx'' \leq 0 \tag{17}$$

Proof. By breaking Lemma 2 into separate dynamic programming problems. □

For non-terminal actions, Corollaries 3 and 4 yields tighter bounds on $\tilde{\delta}_a$ compared to Corollary 1, but the bounds they yield are more difficult to compute, though still tractable. In particular, it requires computing the EV $2|A||X|$ times, or 360 times in the Rust (1987) case, for each accepted set of parameters, which finishes in 1-2 minutes for each accepted set of parameters on a Macbook Pro.

Having discussed how to obtain bounds on counterfactuals conditional on knowing $\tilde{\pi}, \tilde{f}_{x'|x,a}$ as well as bounds on $\tilde{\xi}$ and $G_{a,a'}$, in the following Section 3 we will discuss how to obtain those primitives in the setups of either Norets and Tang (2014) or Buchholz, Shum, and Xu (2017).

3 Obtaining the model primitives

3.1 The discrete states case

When states are discrete, we can without loss of generality use the approach of Norets and Tang (2014) to obtain the required primitives using the following normalization Assumption 6.

Assumption 5. *(Discrete) X is discrete and takes finitely many values. Hence, the transition probabilities $f_{x'|x,a}$ can be written in matrix form as F^a.*

Assumption 6. *(Normalizations) For all $a \in A$, $E(\epsilon_a) = 0$. Furthermore, $\epsilon_0 = 0$, and for some p_1^c, $E(\epsilon_1|\epsilon_1 \geq G_1^{-1}(p_1^c)) = \hat{c}$, where $\hat{c} > 0$ is a constant picked by the researcher.*

A convenient choice for $p_1^c, \hat{c}$ is the median of the observed CCPs $p_1^c = \text{med}(p_1)$, and the corresponding logit conditional expectation $\hat{c} = -p_1^c \log(p_1^c) - (1-p_1^c)\log(1-p_1^c)$. I include a proof in Appendix A.2 that such scale normalizations can be done without affecting choice probabilities.

To extend the Norets and Tang (2014) approach to the multinomial case and to generate bounds on the CDF of the differences in the error terms $\epsilon_a - \epsilon_{a'} \sim G_{a,a'}$, I define the conditional expectation of this difference given that it is greater than its quantile $Q_{a,a'}$ at probability q:

$$e_{a,a'}(q) = \int_{Q_{a,a'}(q)}^{\infty} \epsilon_{a,a'} dG_{a,a'} \tag{18}$$

Then, the following linear program generates the require bounds on the primitives in Assumption 4. These bounds are sharp in the binary case. In the multinomial case, the bounds are not sharp but can still be computed in a linear program.

Lemma 3. *Suppose Assumptions 1-3 and Assumptions 5-6 holds. Order the CCPs by magnitude, so that* $p_a^1 \leq p_a^2 \leq \ldots \leq p_a^{|X|}$*, and call* $Q_{a,a'}(p_a^j) = Q_{a,a'}^j$*,* $e_{a,a'}(p_a^j) = e_{a,a'}^j$*. Given* $q \in [0,1]$ *where we want to bound the CDF of the unobservables. First, suppose* $p_a^1 \leq q \leq p_a^{|X|}$*, we can then define* $l = \arg\max_{t:p_a^t<=q} p_a^t$*,* $n = \arg\min_{t:p_a^t>q} p_a^t$ *as the last and next CCP in the data, and the following constitutes bounds on* $e_{a,a'}(q), Q_{a,a'}(q)$*:*

$$\frac{e_{a,a'}^n(q-p_a^l) + e_{a,a'}^l(p_a^n - q)}{p_a^n - p_a^l} \leq e_{a,a'}(q) \leq \min\left(e_{a,a'}^l - Q_{a,a'}^l(q-p_a^l), e_a^n + Q_{a,a'}^n(p_a^n - q)\right) \tag{19}$$

$$\max\left(\frac{-e_{a,a'}(q) + e_{a,a'}^l}{q - p_a^l}, Q_{a,a'}^l\right) \leq Q_{a,a'}(q) \leq \min\left(\frac{e_{a,a'}(q) - e_{a,a'}^n}{p_a^n - q}, Q_{a,a'}^n\right) \tag{20}$$

Whereas if $q < p_a^1$*, then the bounds are the same as above with* $e_{a,a'}^l = 0, p_a^l = 0, Q_{a,a'}^l \in (-\infty, \infty)$*, and if* $q > p_a^{|X|}$*, the bounds are the same as above except* $e_{a,a'}^n = 0, p_a^n = 1, Q_{a,a'}^n \in (-\infty,\infty)$*.*

For a set of flow payoffs π_a*, the parameters in the bounds in turn satisfy the following:*

1. $e_{a,a'}^j, Q_{a,a'}(p_a^j)$*s satisfy:*

$$-\frac{e_{a,a'}^1}{p_a^1} < Q_{a,a'}^1 < \ldots < Q_{a,a'}^{i-1} < \frac{e_{a,a'}^{i-1} - e_{a,a'}^i}{p_a^i - p_a^{i-1}} < Q_{a,a'}^i < \ldots < Q_{a,a'}^{|X|} < \frac{e_{a,a'}^{|X|}}{1 - p_a^{|X|}} \tag{21}$$

2. $e_{a,a'}, Q_{a,a'}$ *relate to the* $\delta_{a,a'}^j$ *and* $\xi_{a,a'}^j$ *in the following way:*

$$Q_{a,a'}(p_{a'}^j) \leq \delta_{a,a'}^j \leq Q_{a,a'}(1-p_a^j) \tag{22}$$

$$e_{a,a'}(p_{a'}^j)\frac{p_a^j}{1-p_{a'}^j} \leq \xi_{a,a'}^j \leq e_{a,a'}(1-p_a^j) \tag{23}$$

3. *The* $\delta_{a,a'}^j, \xi_{a,a'}^j$ *satisfy a relationship implied by the Bellman equation:*

$$\delta_{a,a'} = \pi_a - \pi_{a'} + \beta(F^a - F^{a'})(I - \beta F^{a'})^{-1}\left[\pi_{a'} + \sum_{a'' \in A\setminus\{a'\}} (diag(p_{a''})\delta_{a'',a'} - \xi_{a'',a'})\right] \tag{24}$$

If no such $e_{a,a'}^j, Q_{a,a'}^j, \delta_{a,a'}^j, \xi_{a,a'}^j$ *satisfying the constraint exists, the set of flow payoffs* π_a *rejected. These bounds are sharp in the binary case where* $a \in \{0,1\}$*.*

Proof. Proof follows from extending Norets and Tang (2014) Lemma 1 to allow for any $q \in [0,1]$ to be bounded, and then applying it to the restriction imposed on δ, ξ in Norets (2011) after extending it to apply to differences in actions a, a'. The bounds in condition 2 are due to the independence Assumption 6, and collapse to equality in the binary case of $a \in \{0, 1\}$. $\square$

The constraints in Lemma 3 are all linear, and Lemma 3 can therefore be solved by a linear program for any position on the CDF q. In particular, it is straightforward to code, with the first constraint being equivalent to the following matrix, and with rest of the constraints are already in linear form.[5]

$$\begin{bmatrix} 1 & 0 & \dots & 0 & \frac{1}{p_a^1} & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & -\frac{1}{p_a^2 - p_a^1} & \frac{1}{p_a^2 - p_a^1} & \dots & 0 \\ \dots & & & & & & & \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & \frac{1}{p_a^{|X|} - p_a^{|X|-1}} \\ -1 & 0 & \dots & 0 & \frac{1}{p_a^2 - p_a^1} & -\frac{1}{p_a^2 - p_a^1} & \dots & 0 \\ 0 & -1 & \dots & 0 & 0 & \frac{1}{p_a^3 - p_a^2} & \dots & 0 \\ \dots & & & & & & & \\ 0 & 0 & \dots & -1 & 0 & 0 & \dots & \frac{1}{1 - p_a^{|X|}} \end{bmatrix} \begin{bmatrix} \delta_{a,a'}^1 \\ \delta_{a,a'}^2 \\ \dots \\ \delta_{a,a'}^{|X|} \\ e_{a,a'}^1 \\ e_{a,a'}^2 \\ \dots \\ e_{a,a'}^{|X|} \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \tag{25}$$

Linear programming is efficient in general, but it can still be costly to run it tens of thousands of times just to restrict the CDF for a grid of points $q \in [0, 1]$. We can further improve on computational efficiency by noticing that the conditions 1, 2, and 3 in Lemma 3 do not relate to the particular q chosen but rather on the structure of $e_{a,a'}^j, Q_{a,a'}^j$. So, sacrificing some tightness, we can instead use a linear program to bound $e_{a,a'}^j, Q_{a,a'}^j$ for a subset of states, and then bound $e_{a,a'}(q), Q_{a,a'}(q)$ as piecewise linear functions of these bounds using Equation 19 which would be valid for all $q \in [0, 1]$ with no error from gridding. This leads to a reduction in the number of linear programming evaluations to a maximum of $|A||X|$, or 182 in the Rust (1987) setup. Furthermore, it keeps the bounds on $\xi_{a,a'}(q)$ piecewise-linear for the binary case, which is useful for computation of counterfactuals using Lemma 1.[6]

3.2 The continuous states, binary action case

I now restrict my attention to a dynamic binary problem where at least one state is continuous, following the setup of Buchholz, Shum, and Xu (2017). Then, the problem can be represented as:

$$u(x, 0) = \beta_0 W_0(\mathbf{x}) \tag{26}$$

$$u(x, 1) = \beta_1 W_1(\mathbf{x}) + \epsilon_1' \tag{27}$$

Suppose the following assumptions hold:

[5] We find Matlab's linprog command, with its dual-simplex method, to be fairly efficient for solving this problem.

[6] Though not necessary for tractability since the computational aspects of this paper was completed without using this simplification

Assumption 7. *For all* $s \geq 1$, $E(||W_d^s|||X) < \infty$ *a.s., where* s *indicates the next period values.*

Assumption 8. $p_1(X)$ *is continuously distributed with support on a closed interval,* $[\underline{p}_1, \overline{p}_1] \subseteq [0, 1]$.

Then, let:

$$m(X) = \phi(X) - \sum_{s=1}^{\infty} \beta^s \left\{ E\left[\int_{\underline{p}}^{p(X^s)} \mathcal{B}(\tau)d\tau|X, Y = 1\right] - E\left[\int_{\underline{p}}^{p(X^s)} \mathcal{B}(\tau)d\tau|X, Y = 0\right]\right\} \tag{28}$$

with$\phi, \mathcal{B}$ being defined the same as in Buchholz, Shum, and Xu (2017). And assume:

Assumption 9. $m(X)$ *is continuously distributed with a joint probability density function* f_m*. The matrix* $E(m(X)m(X)^T)$ *is invertible.*

The second half of the assumption is testable, and impose a parameter normalization:

Assumption 10. $||\beta|| = 1$

Then, Buchholz, Shum, and Xu (2017) says we can identify β and $Q_1(q), q \in [\underline{p}, \overline{p}]$. Note, however, that we do not know anything about the mean of ϵ_1' or its impact on the value function. Therefore, it is not possible to use this knowledge to get counterfactuals. To get counterfactuals, we impose an unknown mean and scale transformation on the problem, such that for (μ, λ), $\lambda > 0$:

$$u(x, 0) = \lambda\beta_0 W_0(\mathbf{x}) \tag{29}$$

$$u(x, 1) = \lambda\beta_1 W_1(\mathbf{x}) + \mu + \lambda\left(\epsilon_1' - \frac{1}{\lambda}\mu\right) \tag{30}$$

Such that, defining $\epsilon_1 = \epsilon_1' - \frac{1}{\lambda}\mu$, we now have an unobservable that is subject to the following normalizations:

$$E(\epsilon_1) = 0 \tag{31}$$

$$\xi_1(Q_1(\overline{p})) = \xi_1^c. \tag{32}$$

This is now a two-parameter problem, where (λ, μ) are the parameters. The following Lemma 4 defines how we can get the required primitives in Assumption 4.

Lemma 4. *Suppose Assumptions 1-3 and Assumptions 7-10 holds. Then, the accepted range of* λ, μ *is defined by the following triangle:*

$$\lambda Q(\underline{p}) - \mu > -\frac{\xi_0}{\underline{p}} \tag{33}$$

$$\lambda Q(\overline{p}) - \mu + \lambda\int_{\underline{p}}^{\overline{p}} Q(p)dp - \mu(\overline{p} - \underline{p}) < \xi_0 \tag{34}$$

$$\lambda > 0 \tag{35}$$

ξ_1 is bounded by the following:

$$\xi_1(q) = \xi_1^c + \int_q^{\overline{p}} Q_1(p)dp, q \in [\underline{p}, \overline{p}] \tag{36}$$

$$\frac{\xi_1(\underline{q})q}{\underline{p}} \leq \xi_1(q) \leq \xi_1(\underline{q}) + Q_1(\underline{p})(\underline{p} - q), q < \underline{p} \tag{37}$$

$$\frac{\xi_1^c(1-q)}{1-\overline{p}} \leq \xi_1(q) \leq \xi_1^c - Q_1(\overline{p})(q - \overline{p}), q > \overline{p} \tag{38}$$

And Q_1 is bounded by the following outside of the identified range:

$$\frac{-\xi_1(q)}{q} \leq Q_1(q) \leq \min\left(\frac{\xi_1(q)}{1-q}, Q_1(\overline{p})\right), q < \overline{p} \tag{39}$$

$$\max\left(\frac{-\xi_1(q) + \xi_1^c}{q - \overline{p}}, Q_1(\overline{p})\right) \leq Q_1(q) \leq \frac{\xi_1(q)}{1-q}, q > \overline{p} \tag{40}$$

Proof. An extension of Norets and Tang (2014) Lemma 1 to allow for continuous states that, substituted and rearranged. □

Based on Lemma 4, one can then grid the two-parameter accepted triangle of (λ, μ) for the primitives, where for each accepted parameter the results in Section 2 leads to bounds on counterfactuals.

4 Application to Rust (1987) Data

In this section I apply my methodology to Rust (1987) data. The Rust (1987) model is as follows.

$$\pi_0(x) = -RC \tag{41}$$

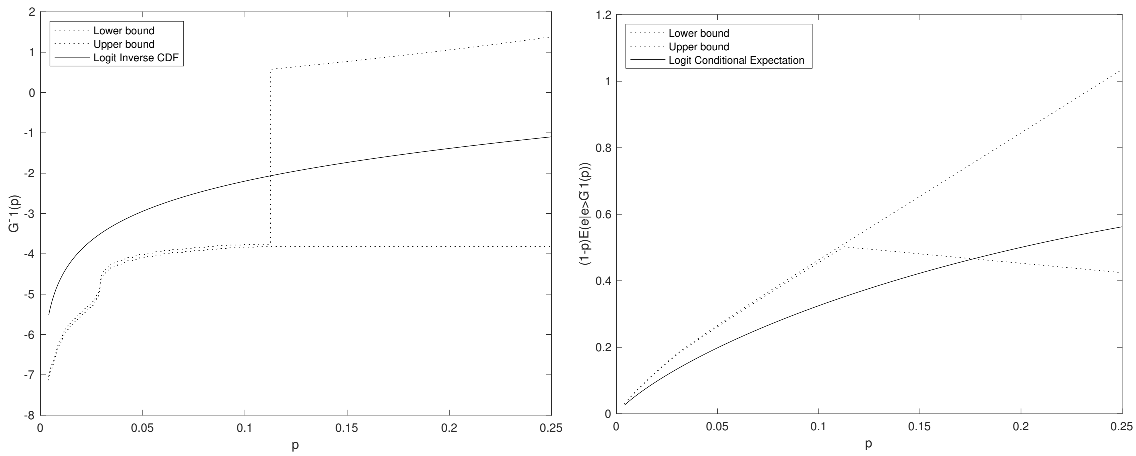
$$\pi_1(x) = -\theta_1 x \tag{42}$$

And the following transition probabilities:

$$F_0(x_{t+1}|x_t, \theta) = \begin{cases} \theta_{20}, x_{t+1} = 0 \\ \theta_{21}, x_{t+1} = 1 \\ 1 - \theta_{20} - \theta_{21}, x_{t+1} = 2 \end{cases} \qquad F_1(x_{t+1}|x_t, \theta) = \begin{cases} \theta_{20}, x_{t+1} = x_t \\ \theta_{21}, x_{t+1} = x_t + 1 \\ 1 - \theta_{20} - \theta_{21}, x_{t+1} = x_t + 2 \end{cases} \tag{43}$$

Where x_t is engine mileage divided into 90 buckets. I show the bounds on the CDF in Figure 1, which are extremely tight within the range of the observed CCPs, but are loose otherwise. This makes sense since even though we have discrete states, the states are tight enough together that we should approach the point identification within the range of observed CCPs in Buchholz, Shum, and Xu (2017).

Figure 1: Bounds on the distribution of the error term based on Lemma 3 with Rust (1987) data, given a random set of accepted flow payoffs π.



(a) Bounds on the Q_1.

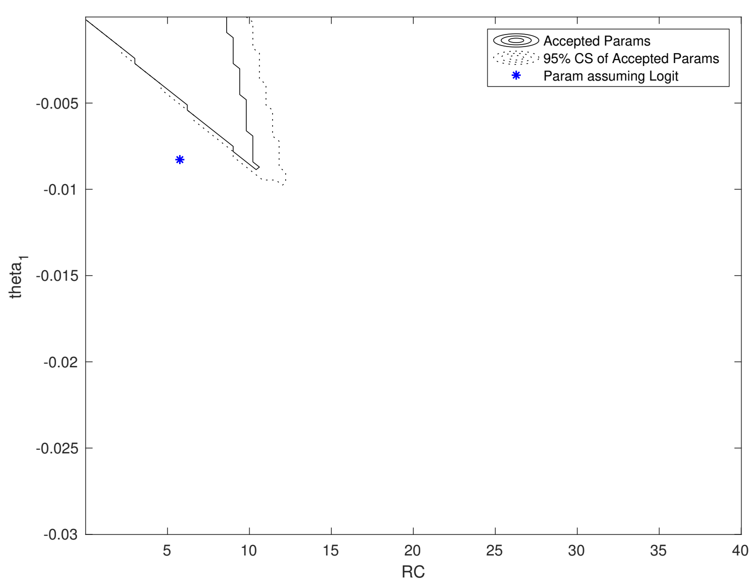
(b) Bounds on ξ_1

We then compute a credible set by bootstrapping to get an estimate of the probability of a parameter being in the identified set Δ_I given data X:

$$\mathbb{C}^{\theta}_{1-\alpha}(X) = \left\{ \hat{\theta} : P\left(\hat{\theta} \in \Delta_I | X \right) \geq \alpha \right\} \tag{44}$$

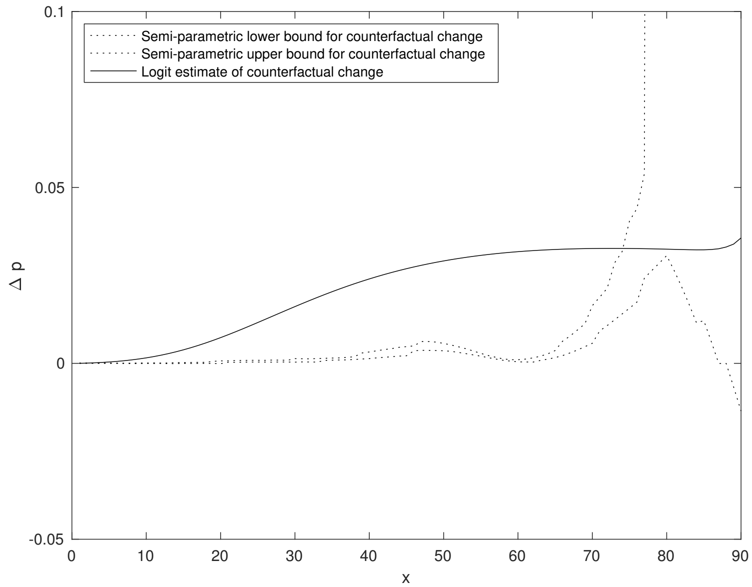
Where the distribution of X is approximated via 500 bootstraps over the set of buses in the data. The accepted range of parameters is plotted in Figure 2.

Figure 2: Accepted parameters using the Rust data



Now, I consider a counterfactual $\tilde{\theta}_{20} = .9, \tilde{\theta}_{21} = .05$. This represents a significant decrease in the usage of the engine. As shown in Figure 3, the logit model predicts that the agent will compensate for the decrease in engine usage by replacing the engine more often, especially when the mileage is high. This comes from a change in the option value of replacement due to the error shocks. Intuitively, the logit result does not make much sense: why should a decrease in engine usage lead to higher replacement probabilities? The semi-parametric model shows that replacement probabilities are likely unchanged, which is perhaps more sensible.

Figure 3: Counterfactual Change in Choice Probabilities using Rust Data



I also show the demand for engines in Table 1 approximated by the amount of time the engine is left running without replacement ("expected lifespan"). After the engine becomes used less often, the logit model predicts a much smaller change in the expected lifespan (from a baseline of $\hat{T} = 89.3$) than the semi-parametric model, because it predicts that the agent will increase choice probabilities to compensate.

Table 1: Counterfactual change in the expected lifespan of a new engine

	CF Change, $\Delta T = \tilde{T} - T$
Semi-Parametric Estimate	[230.6, 248.1]
Logit Estimate	70.9

5 Conclusion

In this paper I have described a tractable method for the semi-parametric estimation of dynamic discrete choice problems. My method can also be used to bound the distribution of unobserved shocks in dynamic games with private shocks, but counterfactual estimation in dynamic games is more difficult because we would have to find the set of all equilibria that can be supported by the identified range of error term distributions. Therefore, bounding the equilibria to dynamic games given bounds on CDFs of the error shocks may be a useful direction for future research.

References

Agarwal, Nikhil, Itai Ashlagi, Michael Rees, Paulo Somaini, and Daniel Waldinger. 2018. "An Empirical Framework for Sequential Assignment: The Allocation of Deceased Donor Kidneys." *Working paper.*

Aguirregabiria, Victor. 2010. "Another Look at the Identification of Dynamic Discrete Decision Processes: An Application to Retirement Behavior." *Journal of Business & Economic Statistics* 28: 201–208.

Aguirregabiria, Victor, and Pedro Mira. 2010. "Dynamic Discrete Choice Structural Models: A Survey." *Journal of Econometrics* 156.

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

Blevins, Jason R. 2014. "Nonparametric Identification of Dynamic Decision Processes with Discrete and Continuous Choices." *Quantitative Economics* 5: 531–554.

Buchholz, Nicholas, Matthew Shum, and Haiqing Xu. 2017. "A Closed-Form Estimator for Dynamic Discrete-Choice Models: Assessing Taxicab Drivers' Dynamic Labor Supply." *Working Paper.*

Chen, Le-Yu. 2017. "Identification of Discrete Choice Dynamic Programming Models with Nonparametric Distribution of Unobservables." *Econometric Theory* 33: 551–577.

Dillon, Eleanor W., and Christopher T. Stanton. 2017. "Self-Employment Dynamics and the Returns to Entrepreneurship." *Working paper.*

Eckstein, Zvi, and Kenneth I. Wolpin. 1989. "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey." *The Journal of Human Resources.*

Gowrisankaran, Gautam, and Marc Rysman. 2012. "Dynamics of Consumer Demand for New Durable Goods." *Journal of Political Economy* 120: 1173–1219.

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

Kalouptsidi, Myrto, Paul Scott, and Eduardo Souza-Rodrigues. 2017. "Identification of Counterfactuals in Dynamic Discrete Choice Models." *Working paper.*

Keane, Michael P., Petra E. Todd, and Kenneth I. Wolpin. 2011. "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications." *Handbook of Labor Economics* 4a.

Lee, Robin S. 2013. "Vertical Integration and Exclusivity in Platform and Two-Sided Markets." *American Economic Review* 103(7): 2960–3000.

Miller, Robert A. 1999. "Estimating Models of Dynamic Optimization with Microeconomic Data." In *Handbook of Applied Econometrics Volume 2: Microeconomics*, eds. M. Hashem Pesaran and Peter Schmidt. Blackwell Publishers Ltd.

Norets, Andriy. 2011. "Semiparametric Identification of Dynamic Multinomial Choice Models." *Working paper.*

Norets, Andriy, and Xun Tang. 2014. "Semiparametric Inference in Dynamic Binary Choice Models." *Review of Economic Studies* 81(3): 1229–1262.

Pakes, Ariel. 1994. "Dynamic Structural Models: Problems and Prospects. Mixed Continuous Discrete Controls and Market Interactions." In *Advances in Econometrics: Proceedings of the 6th World Congress of the Econometric Society*, eds. J.J. Laffont and C. Sims. The Econometric Society.

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

Rust, John. 1994. "Structural estimation of markov decision processes." In *Handbook of Econometrics, Volume IV*, eds. R.F. Engle and D.L. McFadden. Elsevier Science B.V.

Shiraldi, Pasquale. 2011. "Automobile replacement: a dynamic structural approach." *The RAND Journal of Economics* 42: 266–291.

A Proofs

A.1 Proof of Lemma 1

Let $\mathbb{1}$ be an indicator function. Then,

$$E \max_{a \in A} \delta_a(x) + \epsilon_a = E \sum_{a \in A} \mathbb{1}(\delta_a(x) + \epsilon_a \geq \delta_{a'}(x) + \epsilon_{a'}, \forall a')(\delta_a(x) + \epsilon_a) \tag{45}$$

$$= \max_{\{q \in \mathbb{R}^{|A|-1}\}} E \sum_{a \in A} \mathbb{1}(q_a + \epsilon_a \geq q_{a'} + \epsilon_{a'}, \forall a')(\delta_a(x) + \epsilon_a) \tag{46}$$

$$= \max_{\{q \in \mathbb{R}^{|A|-1}\}} \sum_{a \in A} \Pr(q_a + \epsilon_a \geq q_{a'} + \epsilon_{a'}, \forall a')\delta_a(x) + \xi_a(q) \tag{47}$$

The second line follows from the fact that first, we know that the choice of $q = \delta(x)$ would lead the agent to always select the action with the highest payoff and hence no other choice of q would lead to a higher expected payoff, and $q = \delta(x)$ is a possible solution. The third line then follows from the linearity of expectations.

Finally, we invoke the Hotz and Miller (1993) inversion. Suppose that, for some $p'(x)$:

$$\sum_{a \in A} p'_a(x)\delta_a + \xi_a(\mathcal{F}^{-1}(p'(x))) > \max_{\{q \in \mathbb{R}^{|A|-1}\}} \sum_{a \in A} \Pr(q_a + \epsilon_a \geq q_{a'} + \epsilon_{a'}, \forall a')\delta_a(x) + \xi_a(q) \tag{48}$$

Then, there exists $\{q'\} = \mathcal{F}^{-1}(p'(x))$ such that:

$$\sum_{a \in A} p'_a(x)\delta_a + \xi_a(\mathcal{F}^{-1}\{p'(x)\}) = \sum_{a \in A} P(q'_a + \epsilon_a \geq q'_{a'} + \epsilon_{a'}, \forall a')\delta_a + \xi_a(q') \tag{49}$$

$$> \max_{\{q \in \mathbb{R}^{|A|-1}\}} \sum_{a \in A} \Pr(q_a + \epsilon_a \geq q_{a'} + \epsilon_{a'}, \forall a')\delta_a(x) + \xi_a(q) \tag{50}$$

Which is a contradiction. The opposite direction is also a contradiction by using the direct map $\mathcal{F}$ instead of the inverse map. More specifically, suppose that, for some q'':

$$\max_{p(x)} \sum_{a \in A} p_a(x)\delta_a + \xi_a(\mathcal{F}^{-1}(p(x))) < \sum_{a \in A} \Pr(q''_a + \epsilon_a \geq q''_{a'} + \epsilon_{a'}, \forall a')\delta_a(x) + \xi_a(q'') \tag{51}$$

Then, there exists $p'' = \mathcal{F}(q'')$ such that:

$$\sum_{a \in A} \Pr(q''_a + \epsilon_a \geq q''_{a'} + \epsilon_{a'}, \forall a')\delta_a(x) + \xi_a(q'') = \sum_{a \in A} p''_a \delta_a(x) + \xi_a(\mathcal{F}^{-1}(p'')) \tag{52}$$

$$> \max_{p(x)} \sum_{a \in A} p_a(x)\delta_a + \xi_a(\mathcal{F}^{-1}(p(x))) \tag{53}$$

Which is again a contradiction as required.

A.2 A proof that scale normalizations do not affect choice probabilities

A proof that a scale normalization can be done without affecting choice probabilities.

Proof. Suppose for some constant $k > 0$ we have $\pi^*(x, a) = k\pi(x, a), \epsilon^*_a = k\epsilon_a$. Then, by Equation **??**, we have the expected value function of the normalized setting being:

$$EV^*(x) = E \max_{a \in A} \{\pi^*(x, a) + \epsilon^*_a + \beta F(x, a) EV^*\} \tag{54}$$

$$= E \max_{a \in A} \{k\pi(x, a) + k\epsilon_a + \beta F(x, a) EV^*\} \tag{55}$$

Then, we know that $EV^*(x) = kEV(x)$ is a solution to this system. Then, the choice probabilities are:

$$p^*_a(x) = \Pr(\pi^*(x, a) + \epsilon^*_a + \beta F(x, a) EV^* \geq \pi^*(x, a') + \epsilon^*_{a'} + \beta F(x, a') EV^*, \forall a') \tag{56}$$

$$= \Pr(k\pi(x, a) + k\epsilon_a + \beta k F(x, a) EV \geq k\pi(x, a') + k\epsilon_{a'} + \beta k F(x, a') EV, \forall a') \tag{57}$$

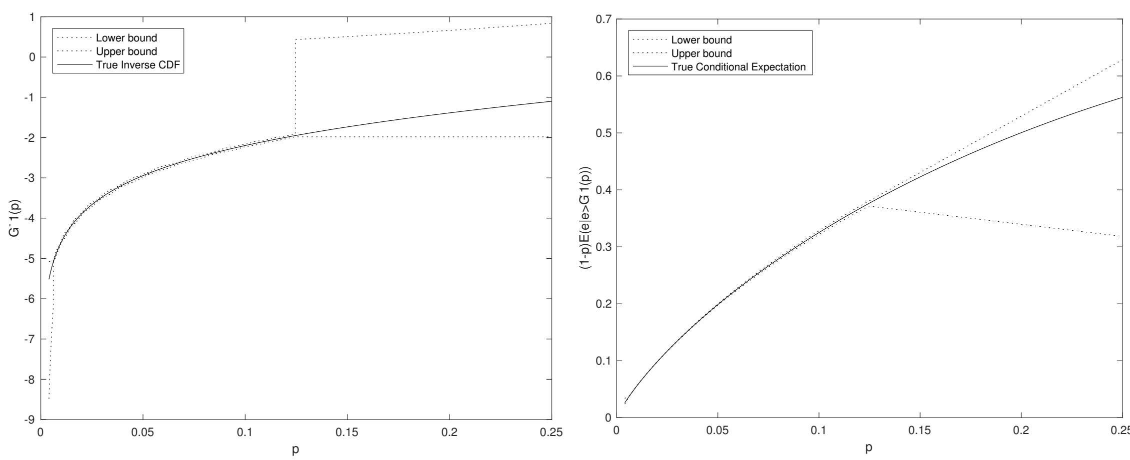
$$= \Pr(\pi(x, a) + \epsilon_a + \beta F(x, a) EV \geq \pi(x, a') + \epsilon_{a'} + \beta F(x, a') EV, \forall a') = p_a(x) \tag{58}$$

Hence, the scaled payoffs yields the same choice probabilities as the original payoffs. □

B Simulation results

In this section I check my method using simulated data from a logit distribution. The following Figure 4 shows that it works extremely well.

Figure 4: Bounds on the distribution of the error term based on Lemma 3 using simulated data, given true flow payoffs π.

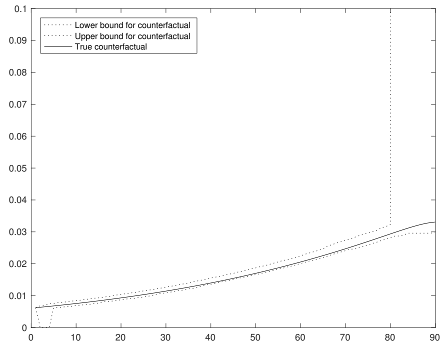


(a) Bounds on the inverse CDF.

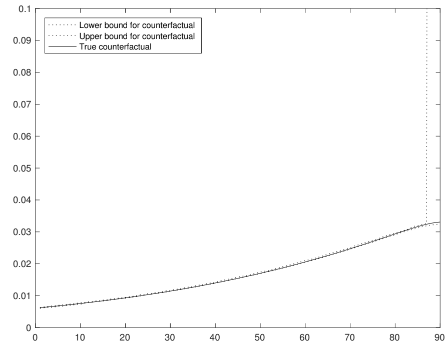
(b) Bounds on conditional expectations

In the following Figure 5 I illustrate the effectiveness of bounds from Corollary 1 and Corollaries 3 and 4 with two iterations, for the counterfactual $\tilde{\theta}_{20} = .6, \tilde{\theta}_{21} = .3$. This is a "change in state transitions" that is not identified without a normalizing assumption on $\pi_1(1)$ according to Norets and Tang (2014) and Kalouptsidi, Scott, and Souza-Rodrigues (2017). Nevertheless, it was used in Norets and Tang (2014) because the parameters in Rust (1987) is not well identified so counterfactuals on parameters are also not well identified. In settings where the parameters are better identified, for example in a model where a wider range of CCPs are observed leading to less flexibility in the CDFs of the error terms, we may be better able to explore counterfactuals that are based on the parameter estimates.

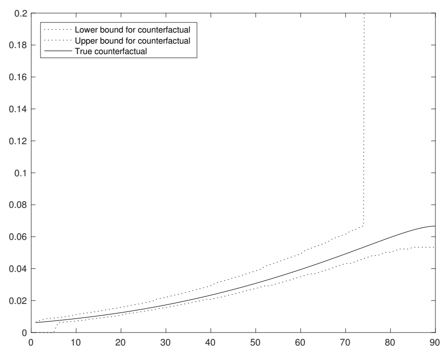
Figure 5: Bounds on the counterfactual given true counterfactual flow payoffs $\tilde{\pi}_a$ using simulated data



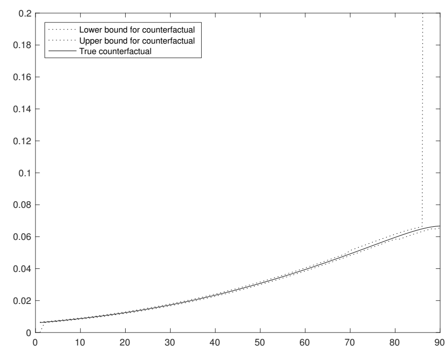
(a) $\beta = .9$, using Corollary 1.



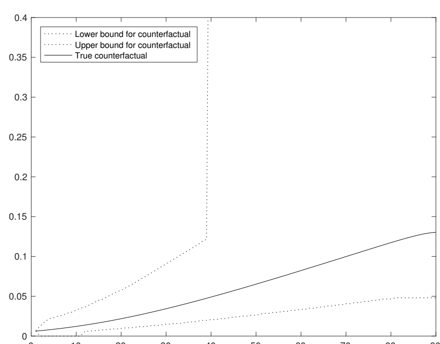
(b) $\beta = .9$, using Corollaries 3 and 4.



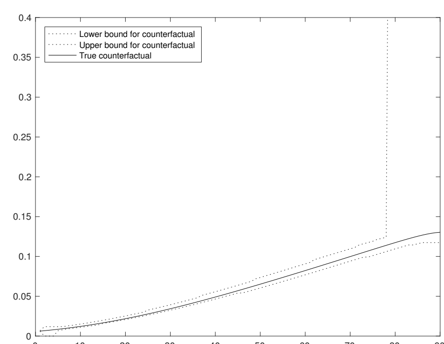
(c) $\beta = .95$, using Corollary 1.



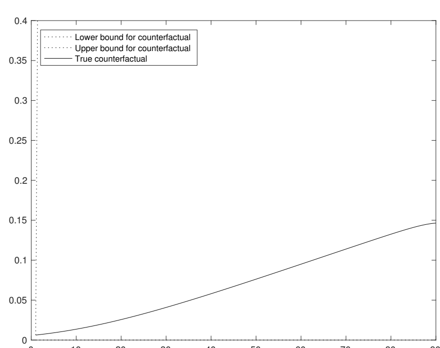
(d) $\beta = .95$, using Corollaries 3 and 4.



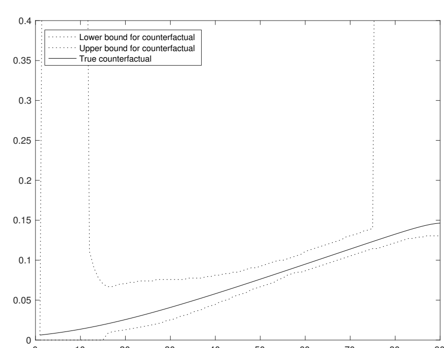
(e) $\beta = .99$, using Corollary 1.



(f) $\beta = .99$, using Corollaries 3 and 4.



(g) $\beta = .999$, using Corollary 1.



(h) $\beta = .999$, using Corollaries 3 and 4.