DUALITY IN DYNAMIC DISCRETE CHOICE MODELS

KHAI X. CHIONG[§], ALFRED GALICHON[†], AND MATT SHUM[♣]

Abstract. Using results from convex analysis, we characterize the identification and estimation of dynamic discrete-choice models based on the random utility framework. We show that the conditional choice proabailities and the choice-specific payoffs in these models are related in the sense of *conjugate duality*. Based on this, we propose a new two-step estimator for these models; interestingly, the first step of our estimator involves solving a linear program which is identical to the classic assignment (two-sided matching) game of Shapley and Shubik (1971). The application of convex-analytic tools to dynamic discrete choice models, and the connection with two-sided matching models, is new in the literature.

1. Introduction

Empirical research utilizing dynamic discrete choice models of economic decision-making has flourished in recent decades, with applications in all areas of applied microeconomics including labor economics, industrial organization, public finance, and health economics. The existing literature on the identification and estimation of these models have recognized a

---

Date: First draft: April 2013. This version: January 2014. Comments welcome.

The authors thank Ben Connault, Thierry Magnac, Emerson Melo, Bob Miller, John Rust, Sorawoot (Tang) Srisuma, Haiqing Xu and audiences at Michigan, Northwestern, NYU, the CEMMAP conference on inference in game-theoretic models (June 2013), UCLA econometrics mini-conference (June 2013), the Boston College Econometrics of Demand Conference (December 2013) and the Toulouse conference on "Recent Advances in Set Identifiation" (December 2013) for helpful comments. Galichon's research has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n°313699 and from FiME, Laboratoire de Finance des Marchés de l'Energie (www.fime-lab.org).

[§]Division of the Humanities and Social Sciences, California Institute of Technology; kchiong@caltech.edu

[†]Department of Economics, Sciences Po; alfred.galichon@sciences-po.fr

[♣]Division of the Humanities and Social Sciences, California Institute of Technology; mshum@caltech.edu.

close link between the conditional choice probabilities (which can be observed and estimated from the data) and the payoffs (or *choice-specific value functions*, which are unobservable to the researcher); indeed, most estimation procedures contain an "inversion" step in which the choice-specific value functions are recovered given the estimated choice probabilities.

This paper has two contributions. First, we explicitly characterize this duality relationship between the choice probabilities and choice-specific payoffs. Specifically, in discrete choice models, the social surplus function (McFadden (1978)) provides us with the mapping from payoffs to the probabilities with which a choice is chosen at each state (conditional choice probabilities). Recognizing that the social surplus function is convex, we utilize the idea that the convex conjugate of the social surplus function gives us the inverse mapping - from choice probabilities to utility indices. More precisely, the subdifferential of the convex conjugate is a correspondence that maps from the observed choice probabilities to an identified set of payoffs. In short, the choice probablities and utility indices are related in the sense of *conjugate duality*. The discovery of this relationship allows us to succinctly characterize the empirical content of discrete choice models, both static and dynamic.

Not only is the convex conjugate of the social surplus function a useful theoretical object; it also provides a new and practical way to "invert" from a given vector of choice probabilities back to the underlying utility indices which generated these probabilities. This is the second contribution of this paper. We show how the conjugate along with its set of subgradients can be efficiently computed by means of linear programming. This linear programming formulation has the structure of an optimal assignment problem similar to the classic Shapley-Shubik (1971) assignment game. This surprising connection enables us to apply insights developed in optimal matchings to discrete choice models.

This paper focuses on the estimation of dynamic discrete-choice models via two-step estimation procedures in which conditional choice probabilities are estimated in the initial stage; this estimation approach was pioneered in Hotz and Miller (HM, 1993) and Hotz, Miller, Sanders, Smith (1994).[1] Our use of tools and concepts from convex analysis to

[1]Subsequent contributions include Aguirregabiria and Mira (2002, 2007), Magnac and Thesmar (2002), Pesendorfer and Schmidt-Dengler (2008), Bajari, et. al. (2009), Arcidiacono and Miller (2011), and Norets and Tang (2013).

study identification and estimation in this dynamic discrete choice setting is novel in the literature. Based on our findings, we propose a new two-step estimator for DDC models. A nice feature of our estimator is that it works for practically any assumed distribution of the utility shocks.[2] Thus, our estimator would make possible the task of evaluating the robustness of estimation to different distributional assumptions.[3]

Section 2 contains our main results regarding duality between choice probabilities and payoffs in discrete choice models. Based on these results, we propose, in Section 3, a two-step estimation approach for these models. We also emphasize here the surprising connection between dynamic discrete-choice and optimal matching models. In Section 4 we discuss computational details for our estimator, focusing on the use of linear programming to compute (approximately) the convex conjugate function from the dynamic discrete-choice model. Monte Carlo experiments (in Section 5) show that our estimator performs well in practice, and we apply the estimator to Rust's (1987) bus engine replacement data (Section 6). Section 7 concludes. The Appendix contains proofs and also a brief primer on results from convex analysis.

2. Basic Model

2.1. **The framework.** In this section we review the basic dynamic discrete-choice setup, as encapsulated in Rust's (1987) seminal paper. The state variable is $x_t \in \mathcal{X}$ which, for convenience, we assume to be finite discrete-valued. Agents choose actions $y_t \in \mathcal{Y}$ from a finite space $\mathcal{Y} = \{0, 1, \ldots, D\}$.

[2]While existing identification results for dynamic discrete choice models allow for quite general specifications of the additive choice-specific utility shocks, many applications of these two-step estimators maintain the restrictive assumption that the utility shocks are distributed i.i.d. type I extreme value, independently of the state variables, leading to choice probabilities which take the multinomial logit form.

[3]We also note that, while they are not the focus in this paper, many applications of dynamic choice models do not utilize HM-type two step estimation procedures, and they allow for quite flexible distributions of the utility shocks, and also for serial correlation in these shocks (examples include Pakes (1986) and Keane and Wolpin (1997)). This literature typically employs simulated method of moments, or simulated maximum likelihood for estimation (see Rust (1994, section 3.3)).

The single-period utility which an agent derives from choosing y_t in period t is

$$\bar{u}\left(y_t, x_t\right) + \epsilon_{y_t}$$

where ϵ_{y_t} denotes the utility shock pertaining to action y_t, which differs across agents. Across agents and time periods, the set of utility shocks $\left\{\epsilon_{y_t}\right\}_{y_t \in \mathcal{Y}}$ is distributed according to a joint distribution function $Q_\epsilon(\cdots; x_t)$ which can depend on the current values of the state variables x_t.

Following Rust (1987), and most of the subsequent papers in this literature, we maintain the following conditional independence assumption (which rules out serially persistent forms of unobserved heterogeneity[4]):

Assumption 1 (Conditional Independence)**.** $(x_t, \vec{\epsilon}_t)$ *is a controlled first-order Markov process, with transition*

$$\begin{aligned} Pr(x_{t+1}, \vec{\epsilon}_{t+1} | y_t, x_t, \vec{\epsilon}_t) =& Pr(\vec{\epsilon}_{t+1} | x_{t+1}, y_t, x_t, \vec{\epsilon}_t) \cdot Pr(x_{t+1} | y_t, x_t, \vec{\epsilon}_t) \\ =& Pr(\vec{\epsilon}_{t+1} | x_{t+1}) \cdot Pr(x_{t+1} | y_t, x_t). \end{aligned}$$

We let $\Pi = (\Pi^1, \ldots, \Pi^D)$ denote the (stationary) Markov transition matrix where $\Pi^y_{ij} = Pr(x_{t+1} = j | y_t = y, x_t = i)$, that is, the ij-th entry of the matrix Π^y denotes the probability that the state transitions from state i to state j when the action taken is y.

The discount rate is β. Agents are dynamic optimizers who solve

$$y_t^* \in \arg\max_{y_t \in \mathcal{Y}} \left\{\bar{u}\left(y_t, x_t\right) + \epsilon_{y_t} + \beta \mathbb{E}\left[\bar{V}\left(x_{t+1}, \epsilon_{t+1}\right) | x_t, y_t\right]\right\}, \tag{1}$$

where, under standard conditions, the value function $\bar{V}$ is recursively defined as

$$\bar{V}\left(x_t, \epsilon_t\right) = \max_{y_t \in \mathcal{Y}} \left\{\bar{u}\left(y_t, x_t\right) + \epsilon_{y_t} + \beta \mathbb{E}\left[\bar{V}\left(x_{t+1}, \epsilon_{t+1}\right) | x_t, y_t\right]\right\}.$$

$V(x_t)$, the ex-ante value function, is defined as:

$$V\left(x_t\right) = \mathbb{E}\left[\bar{V}\left(x_t, \epsilon_t\right) | x_t\right].$$

[4]See Norets (2009), Kasahara and Shimotsu (2009), Arcidiacono and Miller (2011), and Hu and Shum (2012).

The expectation above is conditional on x_t. In the literature, $V(x)$ is called the ex-ante (or integrated) value function, because it measures the continuation value of the dynamic optimization problem before the agent observes his shocks ϵ, so that the optimal action is still stochastic from the agent's point of view.

We now define the *choice-specific value functions* as consisting of two terms: the per-period utility flow and the discounted continuation payoff.

$$w_{y_t}(x_t) \equiv w\left(y_t, x_t\right) = \bar{u}(y_t, x_t) + \beta \mathbb{E}\left[V(x_{t+1})|x_t, y_t)\right],$$

In turn, we can define the *conditional choice probabilities* (CCP's):

$$p(y|x) \equiv Pr(y_t = y|x_t = x) = Pr\{y = \underset{y_t \in \mathcal{Y}}{\operatorname{argmax}}\{w(y_t, x_t = x) + \epsilon_{y_t}\}\}$$

Given these definitions, we proceed to derive the duality between the conditional choice probabilities and the choice-specific value functions.

2.2. **The social surplus function and its convex conjugate.** We start by introducing the indirect expected utility of a decision maker facing the $|\mathcal{Y}|$-dimensional vector of choice-specific values $w.(x)$:

$$\mathcal{G}\left(w.(x)\right) = \mathbb{E}\left[\max_{y \in \mathcal{Y}}\left(w_y(x) + \epsilon_y\right)|x\right]$$

where the expectation is assumed to be finite and is taken over the conditional distribution of the utility shocks, Q_ϵ. This function $\mathcal{G} : \mathbb{R}^{|\mathcal{Y}|} \to \mathbb{R}$, is called the "social surplus function" in McFadden's (1978) random utility framework, and can be interpreted as the expected welfare of a representative agent in the dynamic discrete-choice problem.

Letting Y^* denote the (random) optimal alternative, we get

$$\mathcal{G}\left(w.(x)\right) = \mathbb{E}\left[w_{Y^*}(x) + \epsilon_{Y^*}|x\right] = \sum_{y \in \mathcal{Y}} Pr(Y^* = y|x)\left(w_y(x) + \mathbb{E}[\epsilon_y|Y^* = y, x]\right) \qquad (2)$$

which shows an alternative expression for the social surplus function as a weighted average.

We also introduce the convex conjugate convex conjugate of $\mathcal{G}$, which we denote as $\mathcal{G}^*$:[5]

Definition 1 (Convex Conjugate)**.** *We define $\mathcal{G}^*$, the Legendre-Fenchel conjugate function of $\mathcal{G}$ (a convex function), by*

$$\mathcal{G}^*(p) = \sup_{w \in \mathbb{R}^{\mathcal{Y}}} \left\{ \sum_{y \in \mathcal{Y}} p_y w_y - \mathcal{G}(w) \right\} \tag{3}$$

if p is a probability over the set $\mathcal{Y}$, that is $p_y \geq 0$ and $\sum_{y \in \mathcal{Y}} p_y = 1$, and $\mathcal{G}^*(p) = +\infty$ otherwise. Note that p_y is the y-th component of the $|\mathcal{Y}|$-dimensional vector p.

2.3. **Conjugate duality between choice probabilities and choice-specific value functions.** Because the choice-specific value functions $w_{\cdot}(x)$ and the choice probabilities $p_{\cdot}(x)$ are, respectively, the arguments of the functions $\mathcal{G}$ and its convex conjugate function $\mathcal{G}^*$, we say that $w_{\cdot}(x)$ and $p_{\cdot}(x)$ are related in the sense of *conjugate duality.* Here we explore implications of this duality.

To start, we reiterate a well-known fact that the derivatives of the social surplus function $\mathcal{G}$ at a vector of utilities $\vec{w}(x)$, is the vector of choice probabilities $p(\cdot|x)$:

Proposition 1 (The Williams-Daly-Zachary (WDZ) Theorem)**.**

$$p(\cdot|x) = \nabla \mathcal{G}(w.(x)).$$

This result, which is analogous to Roy's Identity in discrete choice models, is expounded in McFadden (1981) and Rust (1994; Thm. 3.1)). It characterizes the vector of choice probabilities corresponding to optimal behavior in a discrete choice model as the gradient of the social surplus function. (For completeness, we include a proof in the Appendix.)

The WDZ theorem provides a mapping from the choice-specific value functions (which are unobserved by researchers) to the observed choice probabilities $p(\cdot|x)$. The following result, which is the basis of our identification and estimation strategy, provides an "inverse" correspondence from the observed choice probabilities back to the unobserved $w.(x)$, which

[5]Details of convex conjugates are expounded in the Appendix. Convex conjugates are also encountered in production theory. When f is the convex cost function of the firm (decreasing returns to scale in production), then the convex conjugate of the cost function, f^*, is in fact the firm's optimal profit function.

is a necessary step for identification and estimation. It describes one key implication of the conjugate duality between the observed choice probabilities p and the unobserved choice-specific value functions w.

Proposition 2. *The following pair of equivalent statements identify $w_{.}(x)$:*
(i) p is in the subdifferential of $\mathcal{G}$ at w

$$p\left(.|x\right) \in \partial \mathcal{G}\left(w_{.}\left(x\right)\right), \tag{4}$$

(ii) w is in the subdifferential of $\mathcal{G}^*$ at p

$$w_{.}(x) \in \partial \mathcal{G}^*\left(p\left(.|x\right)\right). \tag{5}$$

The definition and properties of the subdifferential of a convex function is provided in Appendix A.[6] Part (i) is, of course, connected to the WDZ theorem above, and hence encapsulates an optimality requirement that the vector of observed choice probabilities p be derived from optimal discrete-choice decision making for some unknown vector w of choice-specific value functions.

Part (ii) of this proposition, which describes the "inverse" mapping from conditional choice probabilities to choice-specific value functions, does not appear to have been exploited in the literature on dynamic discrete choice. It relates to Galichon and Salanié (2012) who use convex analysis to estimate matching games with transferable utilities. It specifically states that the vector of choice-specific value functions can be identified from the corresponding vector of observed choice probabilities p as the subgradient of the convex conjugate function $\mathcal{G}^*(p)$. Eq. (5) is also constructive, and suggests a procedure for computing the choice-specific value functions corresponding to observed choice probabilities; we will fully elaborate this procedure in subsequent sections.[7]

[6] $\mathcal{G}$ is differentiable at w if and only if $\partial\mathcal{G}(w)$ is single-valued. In that case, part (i) of Prop. 2 reduces to $p = \nabla\mathcal{G}(w)$. If, in addition, $\nabla\mathcal{G}$ is one-to-one, then we immediately get $w = \left(\nabla\mathcal{G}\right)^{-1}(p)$, or $\nabla\mathcal{G}^*(p) = \left(\nabla\mathcal{G}\right)^{-1}(p)$, which is the case of the classical Legendre transform. However, as we show below, $\nabla\mathcal{G}(w)$ is not typically one-to-one in discrete choice models, so that the statement in part (ii) of Prop. 2 is more suitable.

[7] Clearly, Proposition 2 also applies to static random utility discrete-choice models, with the $w_{.}(x)$ being interpreted as the utility indices for each of the choices. As such, Eq. (5) relates to results regarding the invertibility of the mapping from utilities to choice probabilities in static discrete choice models (e.g. Berry

It can be shown (cf. Eq. (25) in Appendix A) that $\mathcal{G}(w_.) + \mathcal{G}^*(p_.) = \sum_{y\in\mathcal{Y}} p_y w_y$ if and only if $p_. \in \partial\mathcal{G}(w_.)$. Combining this with Eq. (2), we obtain an alternative expression for the convex conjugate function $\mathcal{G}^*$:

$$\mathcal{G}^*(p(.|x)) = -\sum_y p(y|x)\mathbb{E}[\epsilon_{Y^*}|Y^* = y, x], \tag{6}$$

corresponding to the weighted expectations of the utility shocks ϵ_y conditional on choosing the option y.

Before proceeding, we discuss the example of the logit model, for which the functions and relations above reduce to familiar expressions:

Example 1 (Logit)**.** *When the distribution Q of ϵ obeys an extreme value type I distribution, it follows from Extreme Value theory that $\mathcal{G}$ and $\mathcal{G}^*$ can be obtained in closed form:*[8]

$$\begin{aligned} \mathcal{G}(w) &= \log(\sum_{y\in\mathcal{Y}} \exp(w_y)) + \gamma \\ \mathcal{G}^*(p) &= \sum_{y\in\mathcal{Y}} p_y \log p_y - \gamma, \end{aligned}$$

where $\gamma \approx 0.57$ (Euler's constant). Hence in this case, $\mathcal{G}^*$ is the entropy of distribution p. The subdifferential of $\mathcal{G}^*$ is characterized as follows: $w. \in \partial\mathcal{G}^*(p.)$ if and only if $w_y = \log p_y - K$, for some $K \in \mathbb{R}$.

2.4. **Empirical Content of Dynamic Discrete Choice Model.** To summarize the empirical content of the model, we start with the ex-ante value function V, which solves the following equation

$$V(x) = \sum_{y\in\mathcal{Y}} p(y|x) \left(\bar{u}_y(x) + \mathbb{E}[\epsilon_y|y,x] + \beta \sum_{x'} p(x'|x,y) V(x') \right) \tag{7}$$

found eg. in Pesendorfer and Schmidt-Dengler (2008), where we write $p(x'|x,y) = Pr(x_{t+1} = x'|x_t = x, y_t = y)$. Comparing this equation with the previous equation (7), we see that the

(1994); Haile, Hortacsu, and Kosenok (2008); Berry, Gandhi, and Haile (2013)). Similar results have also arisen in the literature on stochastic learning in games (Hofbauer and Sandholm (2002); Cominetti, Melo and Sorin (2010)).

[8]Relatedly, Arcidiacono and Miller (2011, pp. 1839-1841) discuss computational and analytical solutions for the $\mathcal{G}^*$ function in the generalized extreme value setting.

social surplus and ex-ante value functions coincide:

$$V(x) = \mathcal{G}(w_{.}(x)). \tag{8}$$

From examining the social surplus function $\mathcal{G}$, we see that if $w_{.}(x) \in \partial\mathcal{G}^{*}(p(.|x))$, then it is also true that $w_{.}(x) - K(x) \in \partial\mathcal{G}^{*}(p(.|x))$, where $K(x) \in \mathbb{R}^{|\mathcal{Y}|}$ is vector taking values of $K(x)$ across all $\mathcal{Y}$ components. Indeed, the choice probabilities are only affected by the differences in the levels offered by the various alternatives.

In what follows, we tackle this indeterminacy problem by isolating a particular $w_{.}^{0}(x)$ among those satisfying $w_{.}(x) \in \partial\mathcal{G}^{*}(p(.|x))$, in particular:

$$\mathcal{G}\left(w_{.}^{0}(x)\right) = 0. \tag{9}$$

Theorem 1 below shows that when the distribution of the unobserved heterogeneity ϵ has full support, then Eq. (9) defines $w_{.}^{0}$ uniquely. We stress that this vector $w_{.}^{0}(x)$ need *not* coincide with the "true" vector of choice-specific value functions $w.(x)$ (i.e., that which satisfies $w.(x) \in \partial\mathcal{G}^{*}(p(.|x))$ and Eq. (8)). However, the next theorem shows that for a given vector of CCPs $p(.|x)$, all $w.(x)$ in the subdifferential $\partial\mathcal{G}^{*}(p(.|x))$ are of the form $w_{.}^{0}(x) - K(x)$; hence, the "true" $w.(x)$ will differ from $w_{.}^{0}(x)$ by a constant term $K(x)$. Below, in the second step of the estimation, we show how to point-identify and recover the "true" $w_{.}(x)$ given $w_{.}^{0}(x)$.

Because of this, our choice of $w_{.}^{0}(x)$, as defined in Eq. (9) is without loss of generality; it is not an additional model restriction, but merely a convenient way of *representing* all $w.(x)$ in $\partial\mathcal{G}^{*}(p(.|x))$ with respect to a natural and convenient reference point.[9]

Theorem 1. *Assume the distribution Q_{ϵ} of the utility shocks $(\epsilon_0, \ldots, \epsilon_K)$ is such that $\{\epsilon_k - \epsilon_0\}_k$ has full support on the real line. Let $w.(x)$ be the true choice-specific value function. Then:*

[9]This indeterminacy issue has been resolved in the existing literature on dynamic discrete choice models (eg. Hotz and Miller (1993), Rust (1994), Magnac and Thesmar (2002) by focusing on the *differences* between choice-specific value functions, which is equivalent to setting $w_{y_0}(x)$, the choice-specific value function for a benchmark choice y_0, equal to zero. Compared to this, our choice of $w_{.}^{0}(x)$ satisfying $\mathcal{G}(w_{.}^{0}(x)) = 0$ is more convenient in our context, as it leads to a very simple expression for the constant K (see Theorem 1(iii)).

(i) There exists a unique $w^0_\cdot(x) \in \partial\mathcal{G}^*\left(p\left(.|x\right)\right)$ *such that* $\mathcal{G}\left(w^0_\cdot(x)\right) = 0$*,*

(ii) $w.(x) \in \partial\mathcal{G}^*\left(p\left(.|x\right)\right)$ *if and only if there exists* $K\left(x\right)$ *such that* $w.(x) = w^0_\cdot(x) - K\left(x\right)$*,*

(iii) $K(x) = -V(x)$*.*

The proof of this theorem is in the Appendix. This theorem is our "invertibility" result: for a given vector of choice probabilities $p_\cdot(x)$, it pins down a unique vector of choice-specific value functions $w^0_\cdot(x)$, which differs from the true vector $w_\cdot(x)$ by a constant $K(x)$.[10] Moreover, even when the main condition (the full support of the utility shocks) of Theorem 1 is not satisfied, $w^0_\cdot(x)$ will still be set-identified; Proposition 4 below describes the identified set of $w^0_\cdot(x)$ corresponding to a given vector of choice probabilities $p_\cdot(x)$.

3. Two-Step Estimation procedure

Based upon the derivations in the previous section, we present a two-step estimation procedure. In the first step, we use the results from Theorem 1 to recover the vector of choice-specific value functions $w^0_\cdot(x)$ corresponding to each observed vector of choice probabilities $p_\cdot(x)$. In the second step, we recover the utility flow functions $\bar{u}_y(x)$ given the $w^0_\cdot(x)$ obtained from the first step.

3.1. First step.

In the first step, the goal is to recover the vector of choice-specific value functions $w^0_\cdot(x) \in \partial\mathcal{G}^*(p(\cdot|x))$ corresponding to the vector of observed choice probabilities $p(\cdot|x)$ for each value of x. In doing this, we use Proposition 2 and Theorem 1 above, which show how $w^0_\cdot(x)$ belongs to the subdifferential of the conjugate function $\mathcal{G}^*(p)$. In order to evaluate this function, we use the following proposition, which was derived in Galichon and Salanié (2012, Proposition 2). It characterizes the $\mathcal{G}^*$ function as an optimum of a well-studied mathematical program: the "mass transportation" problem (cf. Villani (2003)).

Proposition 3. *Assume* Q_ϵ *is such that* $\left\{\epsilon_k - \epsilon_0\right\}_k$ *has full support on the real line. Let* $p = \left(p_y\right)_{y\in\mathcal{Y}}$ *be a vector of choice probabilities. Then the function* $\mathcal{G}^*(p)$ *is the value of the*

[10]See Berry (1994), Chiappori and Komunjer (2010), Berry, Gandhi, and Haile (2012), among others, for conditions ensuring the invertibility or "univalence" of demand systems stemming from multinomial choice models, under settings more general than the random utility framework considered here.

mass transportation problem in which the distribution Q_ϵ of utility shocks $\epsilon = \{\epsilon_y\}_{y\in\mathcal{Y}}$ is matched optimally to the distribution of actions y given by the multinomial distribution p, when the cost associated to a match of (ϵ, y) is given by

$$c(y,\epsilon) = -\epsilon_y$$

where ϵ_y is the utility shock from taking the y-th action. That is,

$$\mathcal{G}^*(p) = \sup_{\substack{w,z \\ s.t.\ w(y)+z(\epsilon)\leq c(y,\epsilon)}} \left\{\mathbb{E}_p\left[w(Y)\right] + \mathbb{E}_Q\left[z(\epsilon)\right]\right\} \tag{10}$$

which, by the Monge-Kantorovich duality, coincides with its dual

$$\mathcal{G}^*(p) = \min_{\substack{Y\sim p \\ \epsilon\sim Q_\epsilon}} \mathbb{E}\left[c(Y,\epsilon)\right]. \tag{11}$$

Moreover, $w \in \partial\mathcal{G}^*(p)$ if and only if there exists z such that (w,z) solves (10). Finally, $w^0 \in \partial\mathcal{G}^*(p)$ and $\mathcal{G}(w^0) = 0$ if and only if there exists z such that $\left(w^0, z\right)$ solves (10) and z is such that $\mathbb{E}_Q\left[z(\epsilon)\right] = 0$.

In Eq. (11) above, the minimum is taken across all joint distributions of (Y,ϵ) with marginal distribution equal to, respectively, p and Q_ϵ. It follows from the proposition that the main problem of identification of the choice-specific value functions $w.(x)$ can be recast as a mass transportation problem (Villani (2003)), in which the set of optimizers to Eq. (10) yield vectors of choice-specific value functions $w \in \partial\mathcal{G}^*(p)$.

Moreover, the mass transportation problem can be interpreted as an optimal matching problem. Using a marriage market analogy, consider a setting in which a matched couple consisting of a "man" (with characteristics $y \sim p$) and a "woman" (with characteristics $\epsilon \sim Q_\epsilon$) obtain a joint marital surplus $-c(y,\epsilon) = \epsilon_y$. Accordingly, Eq. (11) is an optimal matching problem in which the joint distribution of characteristics (y,ϵ) of matched couples is chosen to maximum the aggregate marital surplus.

In the case when Q_ϵ is a discrete distribution, the mass transportation problem in the above proposition reduces to a linear-programming problem which coincides with the assignment game of Shapley and Shubik (1971). This connection suggests a convenient way for efficiently computing the $\mathcal{G}^*$ function (along with its subgradient). Specifically, we will show how the dual problem (Eq. (11)) takes the form of a linear programming problem or assignment game, for which some of the associated Lagrange multipliers correspond to the the subgradient $\partial\mathcal{G}^*$, and hence the choice-specific value functions. These computational details are the focus of Section 4 below.

Proof of Proposition 3. This connection between the $\mathcal{G}^*$ function and a matching model follows from manipulation of the variational problem in Equation (2) defining $\mathcal{G}^*$:

$$\begin{aligned}\mathcal{G}^*(p) &= \sup_{w\in\mathbb{R}^{\mathcal{Y}}}\left\{\sum_y p_y w_y - \mathbb{E}_Q\left[\max_{y\in\mathcal{Y}}(w_y+\epsilon_y)\right]\right\} \qquad (12)\\ &= \sup_{w\in\mathbb{R}^{\mathcal{Y}}}\left\{\sum_y p_y w_y + \mathbb{E}_Q\underbrace{\left[\min_{y\in\mathcal{Y}}(-w_y-\epsilon_y)\right]}_{\equiv z(\epsilon)}\right\}.\end{aligned}$$

Defining $c(y,\epsilon)\equiv -\epsilon_y$, one can rewrite the above as

$$\mathcal{G}^*(p) = \sup_{w(y)+z(\epsilon)\leq c(y,\epsilon)}\left\{\mathbb{E}_p[w(Y)] + \mathbb{E}_Q[z(\epsilon)]\right\}. \qquad (13)$$

As is well-known from the results of Monge-Kantorovich (Villani (2003), Thm. 1.3), this is the dual-problem for a mass transportation problem. The corresponding primal problem is

$$\mathcal{G}^*(p) = \min_{\substack{Y\sim p\\ \epsilon\sim\hat{Q}}}\mathbb{E}[c(Y,\epsilon)]$$

which is equivalent to (17)-(19). Comparing Eqs. (12) and (13), we see that the subdifferential $\partial\mathcal{G}^*(p)$ is identified with those elements w such that (w,z), for some z, solves the dual problem (13). □

3.2. **Second step.** From the first step, we obtained $w^0_\cdot(x)$ such that $w_\cdot(x) = w^0_\cdot(x) + V(x)$. Now in the second step, we use the recursive structure of the dynamic model, along with fixing one of the utility flows, to *jointly* pin down the values of $w.(x)$ and $V(x)$. Finally, once $w.(x)$ and $V(x)$ are known, the utility flows can be obtained from $\bar{u}_y(x) = w_y(x) - \beta\mathbb{E}[V(x')|x,y]$

In order to nonparametrically identify $\bar{u}_y(x)$, we need to fix some values of the utility flows. Following Bajari, Chernozhukov, Hong, and Nekipelov (2009), we fix the utility flow corresponding to a benchmark choice y_0 to be constant at zero:[11]

Assumption 2 (Fix utility flow for benchmark choice)**.** $\forall x, \quad \bar{u}(y_0, x) = 0.$

With this assumption, we get

$$0 = w^0_{y_0}(x) + V(x) - \beta \mathbb{E}\left[V\left(x'\right)|x, y = y_0\right]. \tag{14}$$

Let W be the column vector whose general term is $\left(w^0_{y_0}(x)\right)_{x\in\mathcal{X}}$, let V be the column vector whose general term is $\left(V(x)\right)_{x\in\mathcal{X}}$, and let Π^0 be the $|\mathcal{X}| \times |\mathcal{X}|$ matrix whose general term Π^0_{ij} is $Pr\left(x_{t+1} = j | x_t = i, y = y_0\right)$. Equation (14), rewritten in matrix notation, is

$$W = \beta\Pi^0 V - V$$

and for $\beta < 1$, matrix $I - \beta\Pi$ is a diagonally dominant matrix. Hence, it is invertible and Equation (14) becomes

$$V = (\beta\Pi^0 - I)^{-1} W. \tag{15}$$

The right hand side of this equation is uniquely estimated from the data. After obtaining $V(x)$, $\bar{u}_y(x)$ can be nonparametrically identified by

$$\bar{u}_y(x) = w^0_y(x) + V(x) - \beta\mathbb{E}[V(x')|x, y], \tag{16}$$

where $w^0_{\cdot}(x)$ is as in Theorem 1, and V is given by (15).

As a sanity check, one recovers $\bar{u}_{y_0}(.) = W + V - \beta\Pi^0 V = 0$. Also, when $\beta \to 0$, one recovers $\bar{u}_y(x) = w^0_y(x) - w^0_{y_0}(x)$ which is the case in standard static discrete choice.

Eqs. (15) and (16) above, showing how the per-period utility flows can be recovered from the choice-specific value functions via a system of linear equations, echoes similar derivations

[11] In a static discrete-choice setting (i.e. $\beta = 0$), this assumption would be a normalization, and without loss of generality. In a dynamic discrete-choice setting, however, this entails some loss of generality because different values for the utility flows imply different values for the choice-specific value functions, which leads to differences in the optimal choice behavior. Norets and Tang (2013) discuss this issue in greater detail.

in the existing literature (e.g. Aguirregabiria and Mira (2007), Pesendorfer and Schmidt-Dengler (2008), Arcidiacono and Miller (2011, 2013)). Hence, the innovative aspect of our estimator lies not in the second step, but rather in the first step. In the next section, we delve into computational aspects of this first step.

Existing procedures for estimating DDC models typically rely on a small class of distributions for the utility shocks – primarily those in the extreme-value family, as in Example 1 above – because these distributions yield analytical (or near-analytical) formulas for the choice probabilities and $\{\mathbb{E}[\epsilon_y|y,x]\}_y$, the vector of conditional expectation of the utility shocks for the optimal choices, which is required in order to recover the utility flows.[12] Our approach, however, which is based on computing the $\mathcal{G}^*$ function, easily accommodates different choices for Q_ϵ, the (joint) distribution of the utility shocks $(\epsilon_0, \ldots, \epsilon_K)$ conditional on X. Therefore, our findings expand the set of dynamic discrete-choice models suitable for applied work far beyond those with extreme-value distributed utility shocks.[13]

4. Computation Details for First Step

In Section 3.1, we show that the problem of identification in DDC models can be formulated as an optimal transportation problem. In this section, we consider how this may be implemented in practice. In showing how to compute $\mathcal{G}^*$, we exploit the connection, alluded to above, between this function and the assignment game, a model of two-sided matching with transferable utility which have been used to model marriage and housing markets (such as Shapley and Shubik (1971) and Becker (1973)).

[12]Cf. Eq. (7). Related papers include Hotz and Miller (1993), Hotz, Miller, Sanders, Smith (1994), Aguirregabiria and Mira (2007), Pesendorfer and Schmidt-Dengler (2008), Arcidiacono and Miller (2011). Norets and Tang (2013) propose another estimation approach for binary dynamic choice models in which the choice probability function is not required to be known.

[13]This remark is also relevant for static discrete choice models. In fact, the random-coefficients multinomial demand model of Berry, Levinsohn, and Pakes (1995) does not have a closed-form expression for the choice probabilities, thus necessitating a simulation-based inversion procedure. In ongoing work (Chiong, Galichon, Shum (2013)), we are exploring the estimation of random-coefficients discrete-choice demand models using our approach.

4.1. **Linear programming approach.** Let $\hat{Q}_\epsilon$ be a discrete approximation to the distribution Q_ϵ. Specifically, consider a S-point approximation to Q, where the support is $\text{Supp}(\hat{Q}_\epsilon) = \{\epsilon^1, \ldots, \epsilon^S\}$. Let $Pr(\hat{Q}_\epsilon = \epsilon^s) = q_s$. The best S-point approximation is such that the support points are equally weighted, $q_s = \frac{1}{S}$, i.e. the best $\hat{Q}$ is a uniform distribution.[14] Therefore, let $\hat{Q}$ be a uniform distribution whose support can be constructed by drawing S points from the distribution Q_ϵ. It is also known that $\hat{Q}_\epsilon$ converges to Q_ϵ uniformly as $S \to \infty$, so that the approximation error from this discretization will vanish when S is large. Under these assumptions, Problem (10)-(11) has a Linear Programming formulation as

$$\max_{\pi \geq 0} \sum_{y,s} \pi_{ys} \epsilon_y^s \tag{17}$$

$$\sum_{s=1}^{S} \pi_{ys} = p_y, \ \forall y \in \mathcal{Y} \tag{18}$$

$$\sum_{y \in \mathcal{Y}} \pi_{ys} = q_s, \ \forall s \in \{1, ..., S\} . \tag{19}$$

For this discretized problem, the set of $w \in \partial \mathcal{G}^* (p)$ is the set of vectors $(w)_y$ of Lagrange multipliers corresponding to constraints (18). To see how we recover w^0, the specific element in $\partial \mathcal{G}^* (p)$ as defined in Theorem 1, we begin with the dual problem

$$\begin{aligned} &\min_{\lambda, z} \sum_{y \in \mathcal{Y}} p_y \lambda_y + \sum_{s=1}^{S} q_s z_s \\ &s.t. \ \lambda_y + z_s \geq \epsilon_y^s \end{aligned} \tag{20}$$

Consider (λ, z) a solution to (20). By duality, λ is a vector of Lagrange multipliers associated to constraint (18), and z, as Lagrange multipliers associated to constraint (19).[15] We have

[14]cf. Kennan (2006).

[15]Because the two linear programs (17) and (20) are dual to each other, the Lagrange multipliers of interest λ_y can be obtained by computing either program. In practice, for the simulations and empirical application below, we computed the primal problem (17).

$\mathcal{G}^*(p) = \sum_{y\in\mathcal{Y}} p_y\lambda_y + \sum_{s=1}^S q_s z_s$, which implies[16] that $\mathcal{G}(\lambda) = -\sum_{s=1}^S q_s z_s$. Also, for any two elements $\lambda, w^0 \in \partial\mathcal{G}^*(p)$, we have $\sum_{y\in\mathcal{Y}} p_y\lambda_y - \mathcal{G}(\lambda) = \sum_{y\in\mathcal{Y}} p_y w_y^0 - \mathcal{G}(w^0)$.

Hence, because w^0 satisfies $\mathcal{G}(w^0) = 0$, we get

$$w_y^0 = \lambda_y - \mathcal{G}(\lambda) = \lambda_y + \sum_{s=1}^S q_s z_s.$$

This quantity converges to the true value of w_y^0 when S is large enough.[17]

4.1.1. *Discretization of Q_ϵ and a second type of indeterminacy issue.* Thus far, we have proposed a procedure for computing $\mathcal{G}^*$ (and the choice-specific value functions w_0) by discretizing the otherwise continuous distribution Q_ϵ. However, as because the support of ϵ is discrete, w_y^0 will generally not be unique.[18] This is due to the non-uniqueness of the solution to the dual of the LP problem in Eq. (17), and corresponds to Shapley and Shubik's (1971) well-known results on the multiplicity of the core in the finite assignment game. Applied to discrete-choice models, it implies that when the support of the utility shocks is finite, the utilities from the discrete-choice model will only be partially identified. In this section, we discuss this partial identification, or indeterminacy, problem further.

Recall that

$$\mathcal{G}^*(p) = \sup_{w(y)+z(e)\le c(y,e)} \left\{\mathbb{E}_p[w(Y)] + \mathbb{E}_Q[z(\epsilon)]\right\} \tag{21}$$

where $c(y,e) = -e_y$. In Proposition 3, this problem was shown to be the dual formulation of an optimal assignment problem.

We call *identified set of payoff vectors*, denoted by $\mathcal{I}(p)$, the set of vectors w such that

$$\Pr\left(w_y + e_y \ge \max_{y'}\{w_{y'} + e_{y'}\}\right) = p(y) \tag{22}$$

and we denote by $\mathcal{I}_0(p)$ the *normalized identified set of payoff vectors*, that is the set of $w \in \mathcal{I}(p)$ such that $\mathcal{G}(w) = 0$. Note that if Q were to have full support, $\mathcal{I}_0(p)$ would

[16] This uses Eq. (25) in Appendix A, which (in our setup) states that $\mathcal{G}^*(p) + \mathcal{G}(\lambda) = p_. \cdot \lambda$, for all Lagrange multiplier vectors $\lambda \in \partial\mathcal{G}^*(p)$.

[17] In Appendix C, we present an alternative approach to computing the $\mathcal{G}^*$ function, based on "power diagrams".

[18] Note that Theorem 1 requires ϵ to have full support.

contain only one element so $\mathcal{I}_0(p) = \{w^0\}$ as in Theorem 1. Instead, when the distribution Q is discrete, the set $\mathcal{I}_0(p)$ contains a multiplicity of vectors w which satisfy (4). One has:

Proposition 4 (Identified set). *The following holds:*

(i) The set $\mathcal{I}(p)$ coincides with the set of w such that there exists z such that (w, z) is a solution to (21). Thus

$$\mathcal{I}(p) = \left\{ w : \exists z, \quad \begin{array}{c} w(y) + z(e) \leq c(y, e) \\ \mathbb{E}_p[w(Y)] + \mathbb{E}_Q[z(\epsilon)] = \mathcal{G}^*(p) \end{array} \right\}.$$

(ii) The set $\mathcal{I}_0(p)$ is determined by the following set of linear inequalities

$$\mathcal{I}_0(p) = \left\{ w : \exists z, \quad \begin{array}{c} w(y) + z(e) \leq c(y, e) \\ \mathbb{E}_p[w(Y)] = \mathcal{G}^*(p) \\ \mathbb{E}_Q[z(\epsilon)] = 0 \end{array} \right\}.$$

This result allows us to easily derive identification bounds using the characterization of the identified set using linear inequalities. Indeed, for each $y \in \mathcal{Y}$, we can obtain upper (resp. lower) bounds on w_y by maximizing (resp. minimizing) w_y subject to the linear inequalities characterizing $\mathcal{I}_0(p)$, which is a linear programming problem.

Furthermore, when the dimensionality of discretization (S) is high, the core typically shrinks to a singleton, and the core collapses to $\{w_0\}$. A detailed discussion of this (along with counterexamples) is provided in Gretsky, Ostroy, and Zame (1999, section 6). In our Monte Carlo experiments below, we provide evidence for the size of this indeterminacy problem under different levels of discretization.

5. Monte Carlo Evidence

In this section, we illustrate our estimation framework using a dynamic model of resource extraction. To illustrate how our method can tractably handle any general distribution of the unobservables, we use a distribution in which shocks to different choices are correlated. We will begin by describing the setup.

At each time t, let $x_t \in \{1, 2, \ldots, 20\}$ be the state variable denoting the size of the resource pool. There are three choices,

$y_t = 0$: The pool of resources is extracted fully. $x_{t+1}|x_t, y_t = 0$ follows a multinomial distribution on $\{1, 2, 3, 4\}$ with parameter $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$. The utility flow is $u(y_t = 0, x_t) = 0.5\sqrt{x_t} - 2 + \epsilon_0$.

$y_t = 1$: The pool of resources is extracted partially. $x_{t+1}|x_t, y_t = 1$ follows a multinomial distribution on $\{\max\{1, x_t - 10\}, \max\{2, x_t - 9\}, \max\{3, x_t - 8\}, \max\{4, x_t - 7\}\}$ with parameter π. The utility flow is $u(y_t = 1, x_t) = 0.4\sqrt{x_t} - 2 + \epsilon_1$.

$y_t = 2$: Agent waits for the pool to grow and does not extract. $x_{t+1}|x_t, y_t = 2$ follows a multinomial distribution on $\{x_t, x_t + 1, x_t + 2, x_t + 3\}$ with parameter π. We fixed the utility flow to be $u(y_t = 2, x_t) = \epsilon_2$.

The joint distribution of the unobserved state variables is given by $(\epsilon_0 - \epsilon_2, \epsilon_1 - \epsilon_2) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$. Other parameters we fix and hold constant for the Monte Carlo study are the discount rate, $\beta = 0.9$ and $\pi = (0.3, 0.35, 0.25, 0.10)$.

5.1. **Asymptotic performance.** As a preliminary check of our estimation procedure, we show that we are able to recover the utility flows using the actual conditional choice probabilities implied by the underlying model. We discretized the distribution of ϵ using $S = 1000$ support points. As is clear from Figure 1, the estimated utility flows (plotted as dots) as a function of states matched the actual utility functions very well.

5.2. **Finite sample performance.** To test the performance of our estimation procedure when there is sampling error in the CCPs, we generate simulated panel data of the following form: $\{y_{it}\ , x_{it} : i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T\}$ where $y_{it} \in \{0, 1, 2\}$ is the dynamically optimal choice at x_{it} after the realization of simulated shocks. We vary the number of cross-section observations N and the number periods T, and for each combination of (N, T), we generate 100 independent datasets.[19]

For each replication or simulated dataset, the root-mean-square error (RMSE) and R^2 are calculated, showing how well the estimated $\bar{u}_y(x)$ fits the true utility function. The averages are reported in Table 2.

[19]In each dataset, we initialized x_{i1} with a random state in $\mathcal{X}$.

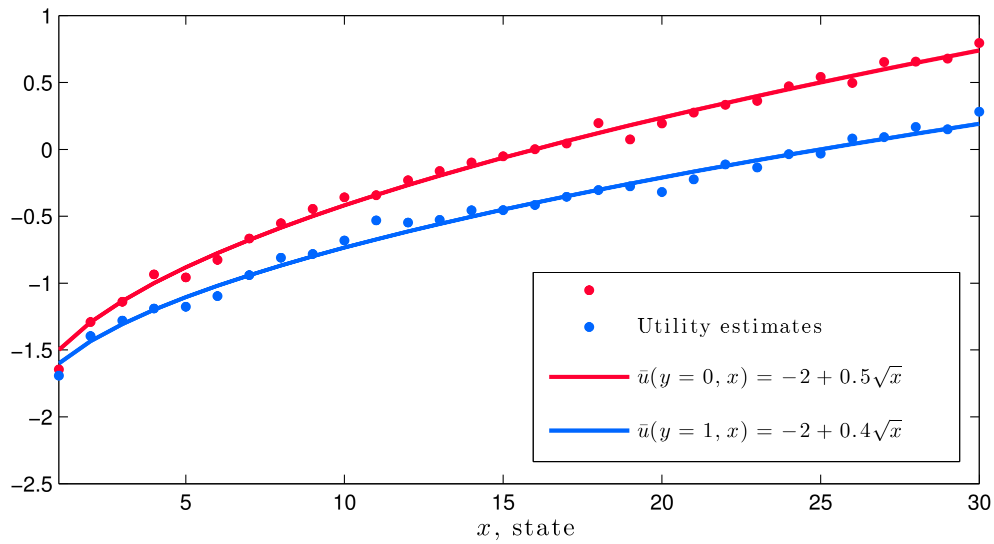


FIGURE 1. Comparison between the estimated and true utility flows.

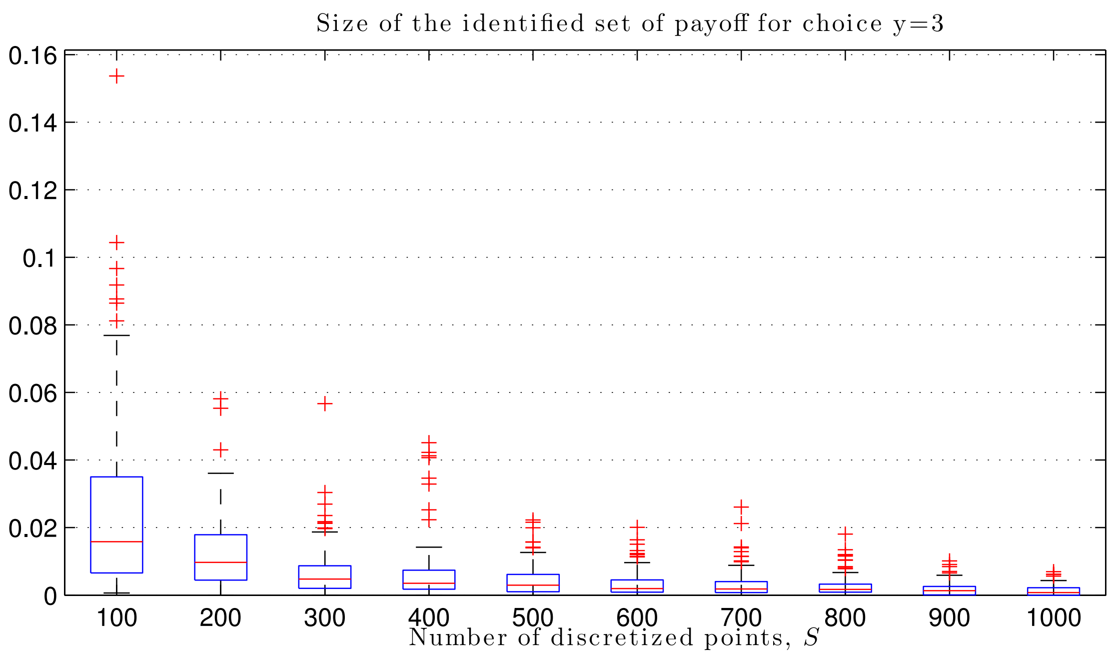
Design	RMSE($y=0$)	RMSE($y=1$)	$R^2(y=0)$	$R^2(y=1)$
$N=100, T=100$	0.5586	0.2435	0.3438	0.7708
$N=100, T=500$	0.1070	0.1389	0.7212	0.9119
$N=100, T=1000$	0.0810	0.1090	0.8553	0.9501
$N=200, T=100$	0.1244	0.1642	0.5773	0.8736
$N=200, T=200$	0.1177	0.1500	0.7044	0.9040
$N=500, T=100$	0.0871	0.1162	0.8109	0.9348
$N=500, T=500$	0.0665	0.0829	0.8899	0.9678
$N=1000, T=100$	0.0718	0.0928	0.8777	0.9647
$N=1000, T=1000$	0.0543	0.0643	0.9322	0.9820

TABLE 1. Average fit across all replications. Standard deviations are reported in the Appendix.

5.3. **Size of the identified set of payoffs.** As mentioned in Section 4.1.1, using a discrete approximation to the distribution of the unobserved state variable introduces a partial identification problem: the identified choice-specific value functions might not be unique.

Using simulations, we next show that the identified set of choice-specific value functions (which we will simply refer to as "payoffs") shrinks to a singleton as S increases, where S is the number of support points in the discrete approximation of Q_ϵ. For S ranging from 100 to 1000 (which is the number of points used in the previous Monte Carlo exercises), we plot in Figure 2, the differences between the largest and smallest choice-specific value function for $y = 3$ across all values of $p \in \Delta^3$ (using the linear programming procedures described in Section 4.1.1).[20]

FIGURE 2. The identified set of payoffs shrinks to a singleton across Δ^3.



For each value of S, we plot the values of the differences $\max_{w \in \partial \mathcal{G}^*(p)} w - \min_{w \in \partial \mathcal{G}^*(p)} w$ across all values of $p \in \Delta^3$. In the boxplot, the central mark is the median, the edges of the box are the 25th and 75th percentiles the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

[20]The analogous plots of the largest and smallest choice-specific value functions for $y = 1$ and $y = 2$ are Figures 5 and 6 in the Appendix.

As is evident, even at small S, the identified payoffs are very close to each other in magnitude. At $S = 1000$, where computation is near-instantaneous, for most of the values in the discretised grid of Δ^3, the core is a singleton; when it is not, the difference in the estimated payoff is less than 0.01. Similar results hold for the choice-specific value functions for choices $y = 1$ and $y = 2$, which are plotted in the Appendix. To sum up, it appears that this indeterminacy issue in the payoffs is not a worrisome problem for even very modest values of S.

6. Empirical Application: Revisiting Harold Zurcher

In this section, we apply our estimation procedure to the bus engine replacement dataset first analyzed in Rust (1987). In each week t, Harold Zurcher (bus depot manager), chooses $y_t \in \{0, 1\}$ after observing the mileage $x_t \in \mathcal{X}$ and the realized shocks ϵ_t. If $y_t = 0$, then he chooses not to replace the bus engine, and $y_t = 1$ means that he chooses to replace the bus engine. The states space is $\mathcal{X} = \{0, 1, \ldots 29\}$, that is, we divided the mileage space into 30 states, each representing a 12,500 increment in mileage since the last engine replacement.[21] Harold Zurcher manages a fleet of 104 identical buses, and we observe the decisions that he made, as well as the corresponding bus mileage at each time period t. The duration between $t + 1$ and t is a quarter of a year, and the dataset spans 10 years. Figures 7 and 8 in the Appendix summarize the frequencies and mileage at which replacements take place in the dataset.

Firstly, we can directly estimate the probability of choosing to replace and not to replace the engine for each state in $\mathcal{X}$. Also directly obtained from the data is the Markov transition probabilities for the observed state variable $x_t \in X$, estimated as:

[21] This grid is coarser compared to Rust's (1987) original analysis of this data, in which he divided the mileage space into increments of 5,000 miles. However, because replacement of engines occurred so infrequently (there were only 61 replacement in the entire ten-year sample period), using such a fine grid size leads to many states that have zero probability of choosing replacement. Our procedure – like all other CCP-based approaches – fail when the vector of conditional choice probability lies on the boundary of the simplex.

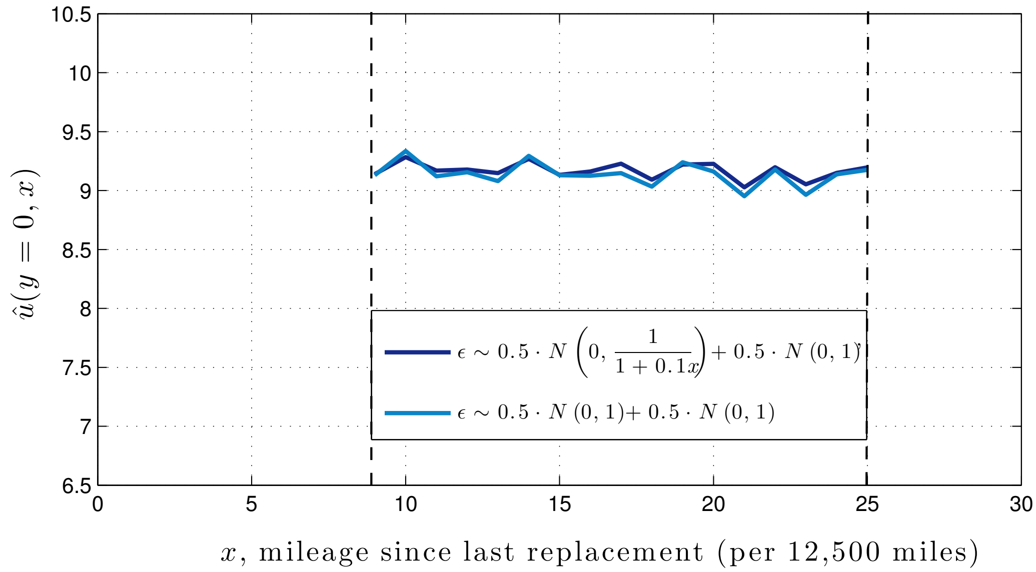


FIGURE 3. Estimates of utility flows $\bar{u}(y=0,x)$, across values of mileage x

$$\hat{\Pr}(x_{t+1}=j|x_t=i,y_t=0)=\begin{cases}0.7405 & \text{if} \quad j=i\\ 0.2595 & \text{if} \quad j=i+1\\ 0 & \text{otherwise}\end{cases}$$

$$\hat{\Pr}(x_{t+1}=j|x_t=i,y_t=1)=\begin{cases}0.7405 & \text{if} \quad j=0\\ 0.2595 & \text{if} \quad j=1\\ 0 & \text{otherwise}\end{cases}$$

For this analysis, we assumed a normal mixture distribution of the error term, specifically, $\epsilon_{t0}-\epsilon_{t1} \sim \frac{1}{2}N(0,1)+\frac{1}{2}N(0,\frac{1}{1+0.1x})$.[22] We chose this mixture distribution in order to allow

[22]In this paper, we restrict attention to the case where the researcher fully knows the distribution of the unobservables $Q_{\bar{\epsilon}}$, so that there are no unknown parameters in these distributions. In principle, the two-step procedure proposed here can be nested inside an additional "outer loop" in which unknown parameters of $Q_{\bar{\epsilon}}$ are considered, but identification and estimation in this case must rely on additional model restrictions in addition to those considered in this paper. We are currently exploring such a model in the context of the simpler static discrete choice setting (Chiong, Galichon and Shum (2014, work in progress)).

the utility shocks to depend on mileage – which accommodates, for instance, operating costs which may be more volatile and unpredictable at different levels of mileage. At the same time, these specifications for the utility shock distribution showcase the flexibility of our procedure in estimating dynamic discrete choice models for any general error distribution. For comparison, we repeat this exercise using an error distribution that is homoskedastic, i.e., its variance does not depend on the state variable x_t. The result appears to be robust to using different distributions of $\epsilon_{t0} - \epsilon_{t1}$. We set the discount rate $\beta = 0.9$.

To non-parametrically estimate $\bar{u}(y = 0, x)$, we fixed $\bar{u}(y = 1, x)$ to 0 for all $x \in X$. Hence, our estimates of $\bar{u}(y = 0, x)$ should be interpreted as the magnitude of operating costs[23] relative to replacement costs[24], with positive values implying that replacement costs exceed operating costs. The estimated utility flow from choosing $y = 0$ (don't replace) relative to $y = 1$ (replace engine) are plotted in Figure 3. We only present estimates for mileage within the range $x \in [9, 25]$, because within this range, the CCPs are in the interior of the probability simplex (cf. footnote 22 and Figure 8 in appendix).

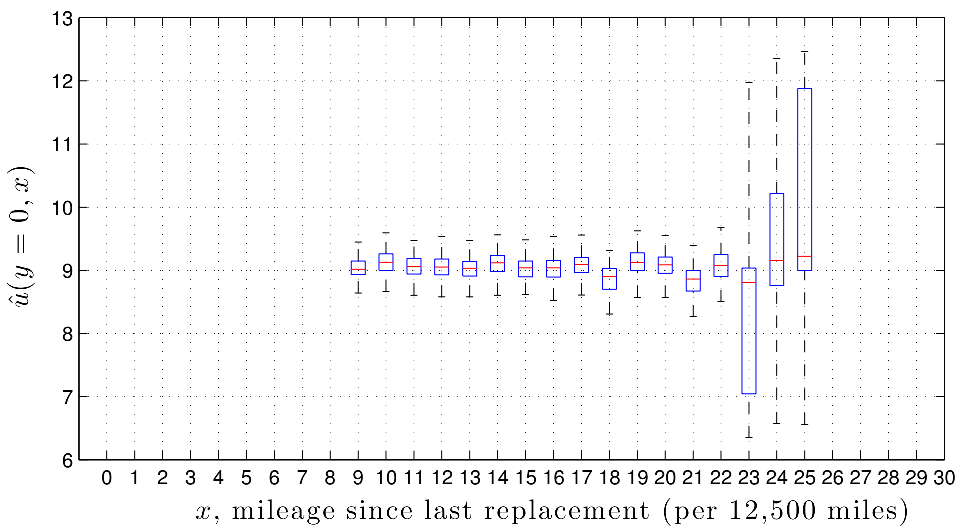
Within this range, the estimated utility function does not vary much with increasing mileages, i.e. it has slope that is not significantly different from zero. The recovered utilities fall within the narrow band of 9 and 9.5, which implies that on average the replacement cost is much higher than the maintenance cost, by a magnitude of 18 to 19 times the variance of the utility shocks. It is somewhat surprising that our results suggest that when the mileage goes beyond the cutoff point of 100,000 miles, Harold Zurcher perceived the operating costs to be inelastic with respect to accumulated mileage. It is worth noting that Rust (1987) mentioned: "According to Zurcher, monthly maintenance costs increase very slowly as a function of accumulated mileage."

To get an idea for the effect of sampling error on our estimates, we bootstrapped our estimation procedure. For each of 100 resamples, we randomly drew 80 buses with replacement from the dataset, and re-estimated the utility flows $\bar{u}(y = 0, x)$ using our procedure. The results are plotted in Figure 4. The evidence suggests that we are able to obtain fairly tight

[23] Operating costs include maintenance, fuel, insurance costs, plus Zurcher's estimate of the costs of lost ridership and goodwill due to unexpected breakdowns.

[24] To be pedantic, this also includes the operating cost at $x = 0$.

Figure 4. Boostrapped estimates of utility flows $\bar{u}(y = 0, x)$



We plot the values of the bootstrapped resampled estimates of $\bar{u}(y = 0, x)$. In each boxplot, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the 5th and 95th percentiles.

cost estimates for states where there is at least one replacement, i.e. for $x \geq 9$ ($x \geq 112,500$ miles), and for states that are reached often enough; i.e. for $x \leq 22$ ($x \leq 275,000$ miles).

7. Conclusions

In this paper, we have shown how results from convex analysis can be fruitfully applied to study identification in dynamic discrete choice models; modulo the use of these tools, a large class of dynamic discrete choice problems with quite general utility shocks becomes no more difficult to compute and estimate than the Logit model encountered in most empirical applications. This has allowed us to provide a natural and holistic framework encompassing the papers of Rust (1987), Hotz and Miller (1993), and Magnac and Thesmar (2002). While the identification results in this paper are comparable to other results in the literature, the approach we take, based on the convexity of the social surplus function $\mathcal{G}$ and the resulting duality between choice probabilities and choice-specific value functions, appears new. Far more than providing a mere reformulation, this approach is powerful, and has significant implications in several dimensions:

First, by drawing the (surprising) connection between the computation of the $\mathcal{G}^*$ function and the computation of optimal matchings in the classical assignment game, we can apply the powerful tools developed to compute optimal matchings to dynamic discrete-choice models.[25] Moreover, by reformulating the problem as an optimal matching problem, all existence and uniqueness results are inherited from the theory of optimal transportation. For instance, the uniqueness of a systematic utility rationalizing the consumer's choices follows from the uniqueness of a potential in the Monge-Kantorovich theorem.

We believe the present paper opens a more flexible way to deal with discrete choice models. While identification is exact for a fixed structure of the unobserved heterogeneity, one may wish to parameterize the distribution of the utility shocks and do inference on that parameter. The results and methods developed in this paper may also extend to dynamic discrete games, with the utility shocks reinterpreted as players' private information (see, e.g. Aguirregabiria and Mira (2007) or Pesendorfer and Schmidt-Dengler (2008)). However, we leave these directions for future exploration.

References

[1] V. Aguirregabiria and P. Mira. Swapping the nested fixed point algorithm: A class of esitmators for discrete Markov decision models. *Econometrica*, 70:1519-1543, 2002.

[2] V. Aguirregabiria and P. Mira. Sequential estimation of dynamic discrete games. *Econometrica*, 75:1–53, 2007.

[3] P. Arcidiacono and R. Miller. Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity. *Econometrica*, 79: 1823-1867, 2011.

[4] P. Arcidiacono and R. Miller. Identifying Dynamic Discrete Choice Models off Short Panels. Working paper, 2013.

[5] F. Aurenhammer. Power diagrams: properties, algorithms, and applications. *SIAM Journal on Computing*, 16(1):78-96, 1987.

[25] While the present paper has used standard Linear Programming algorithms such as the Simplex algorithm, other, more powerful matching algorithm such as the Hungarian algorithm may be efficiently put to use when the dimensionality of the problem grows.

[6] C. Aliprantis and K. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer-Verlag, 2006.

[7] P. Bajari, V. Chernozhukov, H. Hong, and D. Nekipelov. Nonparametric and semiparametric analysis of a dynamic game model. Preprint, 2009.

[8] S. Berry, A. Gandhi, and P. Haile. Connected Substitutes and Invertibility of Demand. *Econometrica* 81: 2087-2111, 2013.

[9] S. Berry. Estimating Discrete-Choice models of Production Differentiation. *RAND Journal of Economics*, 25:242-262, 1994.

[10] S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica*, 63:841–890, July 1995.

[11] P. Chiappori and I. Komunjer. On the Nonparametric Identification of Multiple Choice Models. Working paper, 2010.

[12] K. Chiong, A. Galichon, and M. Shum. Simulation and Partial Identification in Random Coefficient Discrete Choice Demand Models. Work in progress, 2014.

[13] R. Cominetti, E. Melo, and S. Sorin. A payoff-based learning procedure and its application to traffic games. *Games and Economic Behavior*, 70:71-83, 2010.

[14] A. Galichon and B. Salanié. Cupid's invisible hand: Social surplus and identification in matching models. Preprint, 2012.

[15] N. Gretsky, J. Ostroy, and W. Zame. Perfect Competition in the Continuous Assignment Model. *Journal of Economic Theory*, Vol. 85, pp. 60-118, 1999.

[16] P. Haile, A. Hortacsu, and G. Kosenok. On the Empirical Content of Quantal Response Models. *American Economic Review*, 98:180-200, 2008.

[17] J. Hofbauer and W. Sandholm. On the Global Convergence of Stochastic Fictitious Play. *Econometrica*, 70: 2265-2294, 2002.

[18] J. Hotz and R. Miller. Conditional choice probabilties and the estimation of dynamic models. *Review of Economic Studies*, 60:497–529, 1993.

[19] J. Hotz, R. Miller, S. Sanders, and J. Smith. A Simulation Estimator for Dynamic Models of Discrete Choice. *Review of Economic Studies*, 61:265-289, 1994.

[20] Y. Hu and M. Shum. Nonparametric Identification of Dynamic Models with Unobserved Heterogeneity. *Journal of Econometrics*, 171: 32-44, 2012.

[21] H. Kasahara and K. Shimotsu. Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choice. *Econometrica*, 77: 135–175, 2009.

[22] M. Keane and K. Wolpin. The career decisions of young men. *Journal of Political Economy*, 105: 473–522, 1997.

[23] J. Kennan. A Note on Discrete Approximations of Continuous Distributions. Mimeo, University of Wisconsin at Madison, 2006.

[24] T. Magnac and D. Thesmar. Identifying dynamic discrete decision processes. *Econometrica*, 70:801–816, 2002.

[25] D. McFadden. Modelling the choice of residential location. In A. Karlquist et. al., editor, *Spatial Interaction Theory and Residential Location*. North Holland Pub. Co., 1978.

[26] D. McFadden. Economic Models of Probabilistic Choice. In C. Manski and D. McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, 1981.

[27] A. Norets. Inference in dynamic discrete choice models with serially correlated unobserved state variables. *Econometrica*, 77: 1665-1682, 2009.

[28] A. Norets and S. Takahashi. On the Surjectivity of the Mapping Between Utilities and Choice Probabilities. *Quantitative Economics* 4.1 (2013): 149-155.

[29] A. Norets and X. Tang. Semiparametric Inference in Dynamic Binary Choice Models. Preprint, Princeton University, 2013.

[30] A. Pakes. Patents as options: some esitmates of the value of holding European patent stocks. *Econometrica*, 54:1027-1057, 1986.

[31] M. Pesendorfer and P. Schmidt-Dengler. Asymptotic least squares estimators for dynamic games. *Review of Economic Studies*, 75:901–928, 2008.

[32] R. Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[33] J. Rust. Structural Estimation of Markov Decision Processes. *Handbook of Econometrics*, Volume 4 (ed. R. Engle and D. McFadden). North-Holland, 1994.

[34] J. Rust. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55:999–1033, 1987.

[35] L. Shapley and M. Shubik. The assignment game I: The core. *International Journal of Game Theory*, 1(1):111–130, 1971.

[36] C. Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics, Vol. 58. American Mathematical Society, 2003.

Appendix A. Background results

A.1. **Convex analysis for Discrete-choice Models.** Here, we give a brief review of the main notions and results used in the paper. We keep an informal style and do not give proofs, but we refer to Rockafellar (1970) for an extensive treatment of the subject.

Let $u \in \mathbb{R}^{|\mathcal{Y}|}$ be a vector of utility indices. For utility shocks $\{\epsilon_y\}_{y\in\mathcal{Y}}$ distributed according to a joint distribution function Q_ϵ, we define the social surplus function as

$$\mathcal{G}(u) = \mathbb{E}[\max_y \{u_y + \epsilon_y\}], \tag{23}$$

where u_y is the y-th component of u. If $\mathbb{E}(\epsilon_y)$ exists and is finite, then the function $\mathcal{G}$ is a proper convex function that is continuous everywhere. Moreover assuming that Q_ϵ is sufficiently well-behaved (for instance, if it has a density with respect to the Lebesgue measure), $\mathcal{G}$ is differentiable everywhere.

Define the *Legendre-Fenchel conjugate*, or *convex conjugate* of $\mathcal{G}$ as $\mathcal{G}^*(p) = \sup_{u\in\mathbb{R}^{|\mathcal{Y}|}}\{p \cdot u - \mathcal{G}(u)\}$. Clearly, $\mathcal{G}^*$ is a convex function as it is the supremum of affine functions. Note that the inequality

$$\mathcal{G}(u) + \mathcal{G}^*(p) \geq p \cdot u \tag{24}$$

holds in general. The domain of $\mathcal{G}^*$ consists of $p \in \mathbb{R}^{|\mathcal{Y}|}$ for which the supremum is finite. In the case when $\mathcal{G}$ is defined by (23), it follows from Norets and Takahashi (2013) that the domain of $\mathcal{G}^*$ contains the simplex $\Delta^{|\mathcal{Y}|}$, which is the set of $p \in \mathbb{R}^{|\mathcal{Y}|}$ such that $p_y \geq 0$ and $\sum_{y\in\mathcal{Y}} p_y = 1$. This means that our convex conjugate function is always well-defined.

The *subgradient* $\partial\mathcal{G}\,(u)$ of $\mathcal{G}$ at u is the set of $p \in \mathbb{R}^{|\mathcal{Y}|}$ such that

$$p \cdot u - \mathcal{G}(u) \geq p \cdot u' - \mathcal{G}(u')$$

holds for all $u' \in \mathbb{R}^{|\mathcal{Y}|}$. Hence $\partial\mathcal{G}$ is a set-valued function or correspondence. $\partial\mathcal{G}\,(u)$ is a singleton if and only if $\mathcal{G}(u)$ is differentiable at u; in this case, $\partial\mathcal{G}\,(u) = \nabla\mathcal{G}\,(u)$.

One sees that $p \in \partial\mathcal{G}\,(u)$ if and only if $p \cdot u - \mathcal{G}(u) = \mathcal{G}^*(p)$, that is if equality is reached in inequality (24):

$$\mathcal{G}(u) + \mathcal{G}^*(p) = p \cdot u. \tag{25}$$

By symmetry in (25), one sees that $p \in \partial\mathcal{G}\,(u)$ if and only if $u \in \partial\mathcal{G}^*(p)$. In particular, when both $\mathcal{G}$ and $\mathcal{G}^*$ are differentiable, then $\nabla\mathcal{G}^* = \nabla\mathcal{G}^{-1}$.

Appendix B. Proofs

Proof of Proposition 1. Consider the y-th component, corresponding to $\frac{\partial \mathcal{G}(w.)}{\partial w_y}$:

$$\frac{\partial \mathcal{G}(w.)}{\partial w_y} = \frac{\partial}{\partial w_y} \int \max_y [w_y + \epsilon_y] dQ_\epsilon \tag{26}$$

$$= \int \frac{\partial}{\partial w_y} \max_y [w_y + \epsilon_y] dQ_\epsilon \tag{27}$$

$$= \int \mathbb{1}(w_y + \epsilon_y \geq w_{y'} + \epsilon_{y'}), \forall y' \neq y) dQ_\epsilon = p(y). \tag{28}$$

(We have suppressed the dependence of x for convenience.) $\square$

Proof of Proposition 2. This follows directly from Fenchel's inequality (see Rockafellar (1970), Theorem 23.5). $\square$

Proof of Theorem 1. In this proof we shall drop x from the notation for the sake of clarity.

For a vector w we shall denote $Y(w, \varepsilon)$ be the value of y which maximizes $w_y + \varepsilon_y$.

Because ϵ has full support, the choice probabilities p will lie strictly in the interior of the simplex $\Delta^{|\mathcal{Y}|}$. Let $\tilde{w} \in \partial \mathcal{G}^* (p)$, and let $w_y = \tilde{w}_y - \mathcal{G}(\tilde{w})$. One has $\mathcal{G}(w) = 0$, and an immediate calculation shows that $\partial \mathcal{G}(w) = p$. Let us now show that w is unique. Consider w and w' such that $\mathcal{G}(w) = \mathcal{G}(w') = 0$, and $p \in \partial \mathcal{G}(w)$ and $p \in \partial \mathcal{G}(w')$. Assume $w \neq w'$ to get a contradiction; then there exist two distinct y_0 and y_1 such that $w_{y_0} - w_{y_1} \neq w'_{y_0} - w'_{y_1}$; without loss of generality one may assume

$$w_{y_0} - w_{y_1} > w'_{y_0} - w'_{y_1}.$$

Let S be the set of ε's such that

$$w_{y_0} - w_{y_1} > \varepsilon_{y_1} - \varepsilon_{y_0} > w'_{y_0} - w'_{y_1}$$

$$w_{y_0} + \varepsilon_{y_0} > \max_{y \neq y_0, y_1} w_y + \varepsilon_y$$

$$w'_{y_1} + \varepsilon_{y_1} > \max_{y \neq y_0, y_1} w'_y + \varepsilon_y$$

Because ε has full support, S has positive probability.

Let $\bar{w} = \frac{w+w'}{2}$. Because $p \in \partial\mathcal{G}(w)$ and $p \in \partial\mathcal{G}(w')$, one has $\mathcal{G}(\bar{w}) = 0$, thus

$$\begin{aligned} 0 &= \mathbb{E}\left[\bar{w}_{Y(\bar{w},\varepsilon)} + \varepsilon_{Y(\bar{w},\varepsilon)}\right] = \frac{1}{2}\mathbb{E}\left[w_{Y(\bar{w},\varepsilon)} + \varepsilon_{Y(\bar{w},\varepsilon)}\right] + \frac{1}{2}\mathbb{E}\left[w'_{Y(\bar{w},\varepsilon)} + \varepsilon_{Y(\bar{w},\varepsilon)}\right] \\ &\leq \frac{1}{2}\mathbb{E}\left[w_{Y(w,\varepsilon)} + \varepsilon_{Y(w,\varepsilon)}\right] + \frac{1}{2}\mathbb{E}\left[w'_{Y(w',\varepsilon)} + \varepsilon_{Y(w',\varepsilon)}\right] \\ &= \frac{1}{2}\left(\mathcal{G}(w) + \mathcal{G}\left(w'\right)\right) = 0 \end{aligned}$$

Hence equality holds term by term, and

$$\begin{aligned} w_{Y(w,\varepsilon)} + \varepsilon_{Y(w,\varepsilon)} &= w_{Y(\bar{w},\varepsilon)} + \varepsilon_{Y(\bar{w},\varepsilon)} \\ w'_{Y(w',\varepsilon)} + \varepsilon_{Y(w',\varepsilon)} &= w'_{Y(\bar{w},\varepsilon)} + \varepsilon_{Y(\bar{w},\varepsilon)} \end{aligned}$$

For $\varepsilon \in S$, $Y(w,\varepsilon) = Y(\bar{w},\varepsilon) = y_0$ and $Y(w',\varepsilon) = Y(\bar{w},\varepsilon) = y_1$, and we get the desired contradiction.

Hence $w = w'$, and the uniqueness of w follows.

Finally, to prove (iii), we will exploit the structure of the dynamic optimization problem in order to determine the value of K

From $\mathcal{G}\left(w^0_{\cdot}(x)\right) = 0$ and $w_{\cdot}(x) = w^0_{\cdot}(x) - K(x)$, it follows that $\mathcal{G}(w_{\cdot}(x)) = -K(x)$. But we also know that $V(x) = \mathcal{G}(w_{\cdot}(x))$, therefore we get $K(x) = -V(x)$, so that

$$w_{\cdot}(x) = w^0_{\cdot}(x) + V(x).$$

□

Proof of Proposition 4. (i) follows from Proposition 3 and the fact that if $w(y) + z(e) \leq c(y,e)$, then $\mathbb{E}_p[w(Y)] + \mathbb{E}_Q[z(\epsilon)] = \mathcal{G}^*(p)$ if and only if (w,z) is a solution to the dual problem.

(ii) follows from the fact that $-z(\epsilon) = \sup_y\{w(y) - c(y,\epsilon)\} = \sup_y\{w(y) + \epsilon_y\}$, thus $\mathbb{E}_Q[z(\epsilon)] = 0$ is equivalent to $\mathbb{E}_Q\left[\sup_y\{w(y) + \epsilon_y\}\right] = 0$, that is $\mathcal{G}(w) = 0$. □

Appendix C. Power cells approach for computing $\mathcal{G}^*$ function

Second, we can give geometric insights for the locus of ϵ which lead to the choice of some given y. For this, we need to reinterpret the utility shock ϵ_y as a scalar product in a higher dimensional space – a classical trick. For $y \in \mathcal{Y}$, let $\iota_y \in \{0,1\}^{\mathcal{Y}}$ the vector such that $(\iota_y)_{y'} = 1\,(y = y')$. Introduce $S_{\mathcal{Y}} = \{\iota_y : y \in \mathcal{Y}\}$, which is nothing else than the canonical basis of $\mathbb{R}^{\mathcal{Y}}$. Denoting $\cdot$ the scalar product in $\mathbb{R}^{\mathcal{Y}}$, one has $\epsilon_y = \epsilon \cdot \iota_y$, and letting P be the distribution over $S_{\mathcal{Y}}$ which gives probability p_y to point ι_y, problem (11) rewrites as

$$\mathcal{G}^*\left(p\right) = -\max_{\substack{Z \sim P \\ \epsilon \sim Q_\epsilon}} \mathbb{E}\left[\epsilon \cdot Z\right].$$

Hence, $-\mathcal{G}^*\left(p\right)$ is the value of a Monge-Kantorovich problem with a quadratic surplus. This problem is very well studied, and by Brenier's theorem, there exists a convex map $V : \mathbb{R}^{\mathcal{Y}} \to \mathbb{R}$ such that the optimal coupling (Z, ϵ) is such that $Z \in \partial V\left(\epsilon\right)$. As a result, Y^* is defined in (1) is related to ϵ by

$$\iota_{Y^*} \in \partial V^w\left(\epsilon\right) \tag{29}$$

where V is a convex, piecewise linear function given by $V^w\left(\epsilon\right) = \max_{y \in \mathcal{Y}}\left\{\epsilon \cdot \iota_y - w_y\right\}$. Because V^w is a convex function, it is (Lebesgue-) almost everywhere differentiable, so if the distribution of ϵ is absolutely continuous, then $\nabla V^w\left(\epsilon\right)$ exists almost surely, and (29) rewrites as $\iota_{Y^*} = \nabla V^w\left(\epsilon\right)$.

Define C_y^w as the set of ϵ which lead to the choice of y, that is

$$C_y^w = \left\{\epsilon \in \mathbb{R}^{\mathcal{Y}} : \iota_y \in \partial V^w\left(\epsilon\right)\right\}.$$

C_y^w are closed convex polytopes which are called *Power Diagrams* in combinatorial geometry, see Aurenhammer (1987). The probability of choice of y is hence $Q\left(C_y^w\right)$, the mass assigned by distribution Q to set C_y^w. Routines in combinatorial geometry provide the computation of the area of C_y^w. Note that

$$Q\left(C_y^w\right) = \int_{C_y^w} dQ$$

which can also be approximated using simulation techniques.

Once $Q\left(C_y^w\right)$ is computed, we can use the following result to obtain w^y:

Theorem 2. *When Q_ϵ has full support, Problem (10)-(11) reformulates as*

$$\mathcal{G}^*(p) = -\min_{w \in \mathbb{R}^{\mathcal{Y}}} \sum_{y \in \mathcal{Y}} p_y w_y + Q\left(C_y^w\right). \tag{30}$$

Problem (30) is a convex optimization problem and can be solved using a gradient descent of the form

$$w_y^{t+1} = w_y^{t+1} + \delta\left(w_y - Q\left(C_y^{w^t}\right)\right).$$

Appendix D. Additional Figures

Design	RMSE($y=0$)	RMSE($y=1$)	$R^2(y=0)$	$R^2(y=1)$
$N=100, T=100$	0.5586 (3.7134)	0.2435 (0.1155)	0.3438 (0.7298)	0.7708 (0.2073)
$N=100, T=500$	0.1070 (0.0541)	0.1389 (0.0638)	0.7212 (0.2788)	0.9119 (0.0820)
$N=100, T=1000$	0.0810 (0.0376)	0.1090 (0.0425)	0.8553 (0.1285)	0.9501 (0.0352)
$N=200, T=100$	0.1244 (0.0594)	0.1642 (0.0628)	0.5773 (0.6875)	0.8736 (0.1112)
$N=200, T=200$	0.1177 (0.0736)	0.1500 (0.0816)	0.7044 (0.2813)	0.9040 (0.0842)
$N=500, T=100$	0.0871 (0.0375)	0.1162 (0.0430)	0.8109 (0.2468)	0.9348 (0.0650)
$N=500, T=500$	0.0665 (0.0261)	0.0829 (0.0290)	0.8899 (0.1601)	0.9678 (0.0374)
$N=1000, T=100$	0.0718 (0.0340)	0.0928 (0.0344)	0.8777 (0.1320)	0.9647 (0.0314)
$N=1000, T=1000$	0.0543 (0.0176)	0.0643 (0.0162)	0.9322 (0.0577)	0.9820 (0.0101)

Table 2

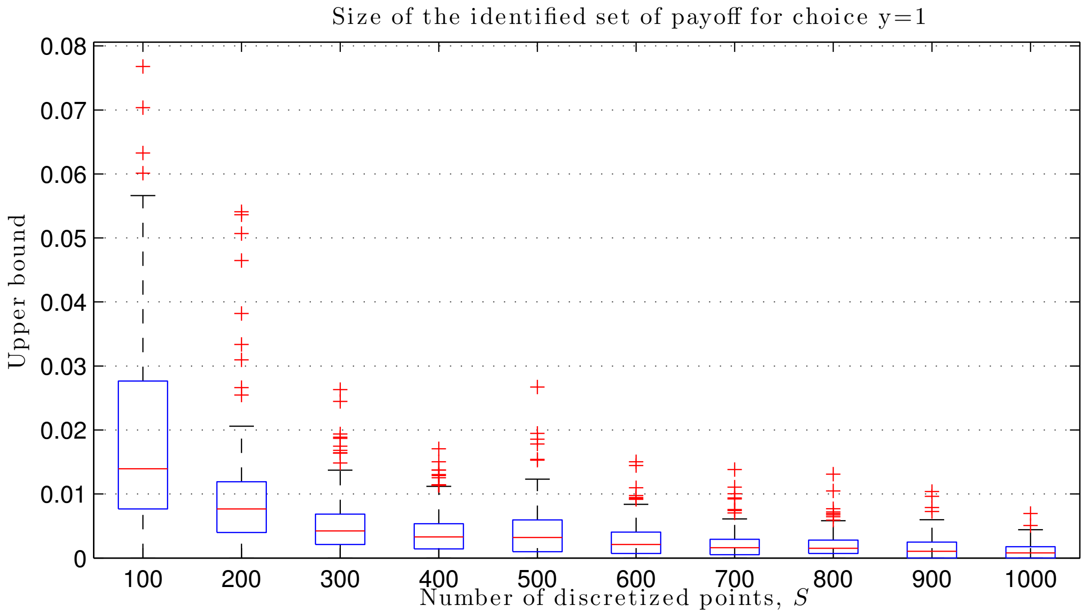


FIGURE 5. For each value of S, we plot the values of the differences $\max_{w\in\partial\mathcal{G}^*(p)} w_1 - \min_{w\in\partial\mathcal{G}^*(p)} w_1$ across all values of $p \in \Delta^3$. In the box-plot, the central mark is the median, the edges of the box are the 25th and 75th percentiles the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

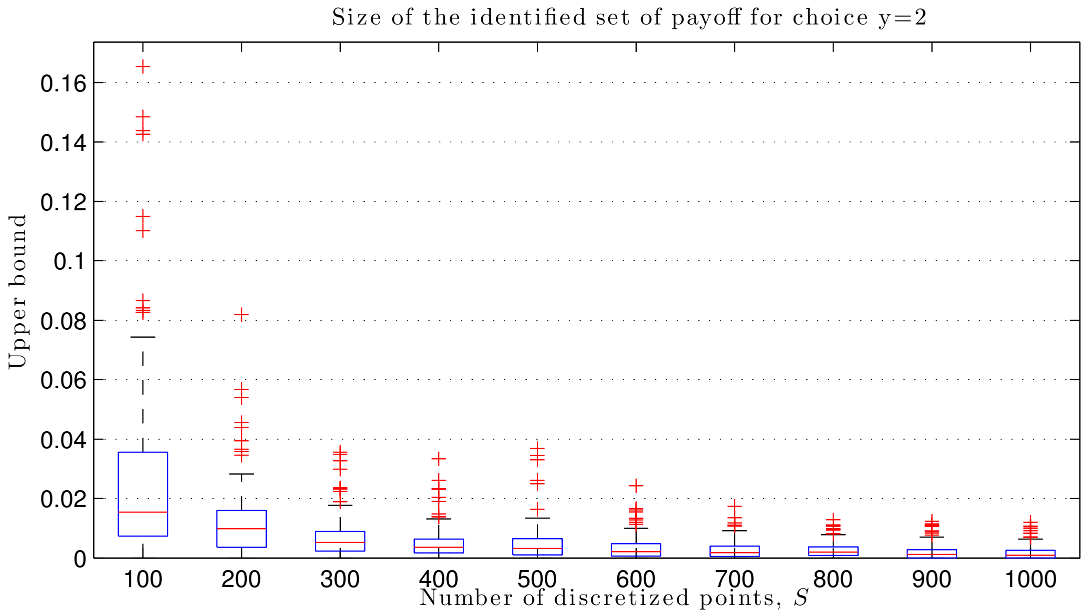


FIGURE 6. For each value of S, we plot the values of the differences $\max_{w\in\partial\mathcal{G}^*(p)} w_2 - \min_{w\in\partial\mathcal{G}^*(p)} w_2$ across all values of $p \in \Delta^3$. In the box-plot, the central mark is the median, the edges of the box are the 25th and 75th percentiles the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

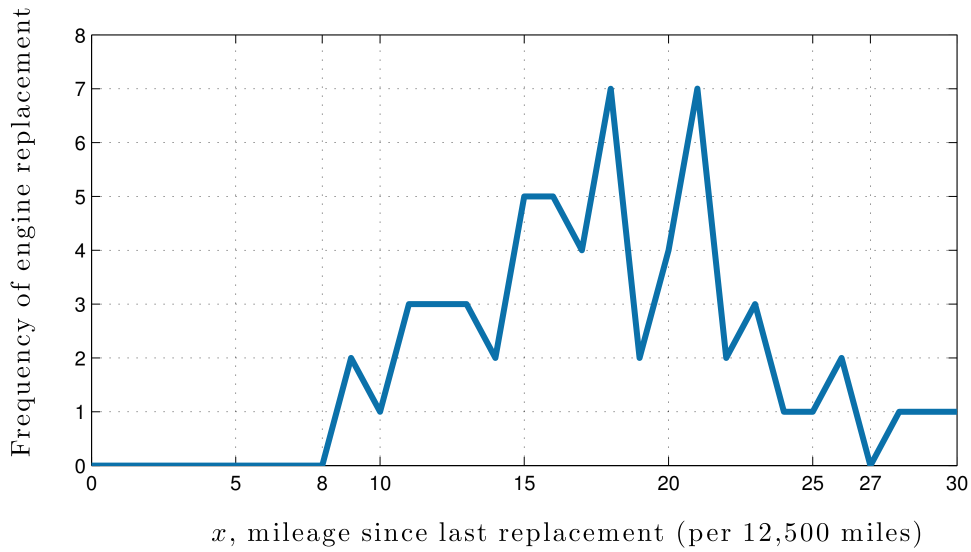


Figure 7

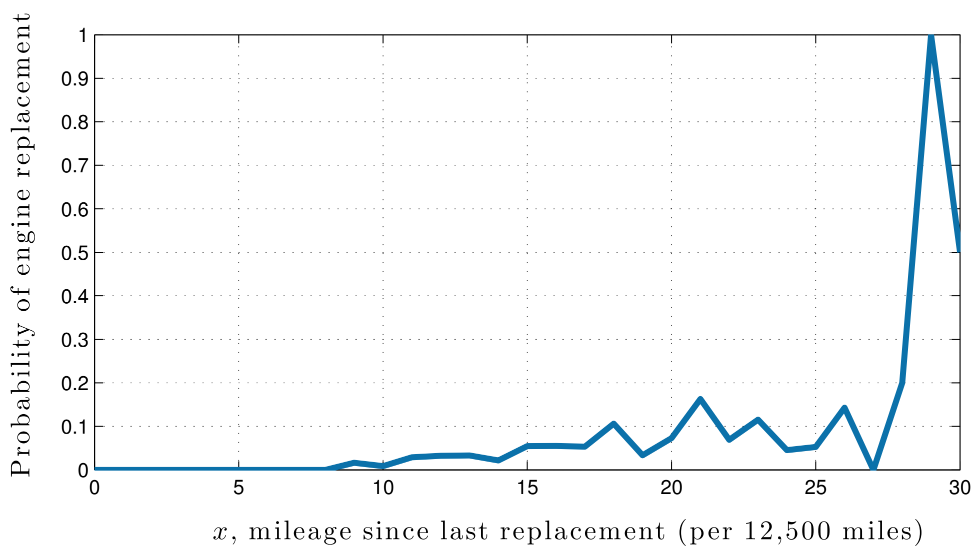


Figure 8