Rationalizable Counterfactual Choice Probabilities in Dynamic Binary Choice Processes

Xun TANG[1]

University of Pennsylvania

June 20, 2009

Abstract

We address two issues in nonparametric structural analyses of dynamic binary choice processes (DBCP). First, the DBCP is not testable and decision makers' single-period payoffs (SPP) cannot be identified even when the distribution of unobservable states (USV) is known. Numerical examples show setting SPP from one choice to arbitrary utility levels to identify that from the other can lead to errors in predicting choice probabilities under counterfactual state transitions. We propose two solutions. First, if a data generating process (DGP) has exogenous variations in observable state transitions, the DBCP becomes testable and SPP is identified. Second, exogenous economic restrictions on SPP (such as ranking of states by SPP, or shape restrictions) can be used to recover the identified set of rationalizable counterfactual choice probabilities (RCCP) that are consistent with model restrictions.

The other (more challenging) motivating issue is that when the USV distribution is not known, misspecification of the distribution in structural estimation leads to errors in counterfactual predictions. We introduce a simple algorithm based on linear programming to recover sharp bounds on RCCP. This approach exploits the fact that some stochastic restrictions on USV (such as independence from observable states) and economic restrictions on SPP can be represented (without loss of information for counterfactual analyses) as linear restrictions on SPP and distributional parameters of USV. We use numerical examples to illustrate the algorithm and show sizes of identified sets of RCCP can be quite small relative to the outcome space.

Keywords: Dynamic discrete choice models, counterfactual outcomes, rationalizability, nonparametric and semiparametric identification

JEL Classification: C13, C14, C25

[1]Correspondence to: Xun Tang, Department of Economics, University of Pennsylvania, Philadelphia, PA 19104. Email: xuntang@sas.upenn.edu. Phone: 215-898-7409.

This paper evolves from a chapter of my Ph.D. dissertation at Northwestern University. I thank Victor Aguirregabiria, Aureo De Paula, Philip Haile, Lars Peter Hansen, Charles Manski, Rosa Matzkin, Robert Porter, Elie Tamer, Petra Todd and seminar participants at Northwestern and Penn for helpful discussions. I am especially grateful to Kenneth Wolpin for many insightful comments and detailed discussions that help improve this paper. Any and all errors are my own.

1 Introduction

In a dynamic binary choice process (DBCP), a decision maker's choice of actions each period affects future payoffs through its impact on transitions between current and future states. In each period, the decision maker chooses an action to maximize the present value of expected future payoffs conditional on current states. Structural analyses of DBCP use choice data to estimate underlying model primitives, and use the estimates to infer agents' choice patterns under counterfactual decision environments (e.g. when the transitions between observable states, or the static single-period payoffs (SPP) are changed.) The model has found wide applications in labor economics (e.g. Eckstein and Wolpin (1989), Wolpin (1987), Keane and Wolpin (1997)), industrial organization (e.g. Rust (1987), Pakes, Ostrovsky and Berry (2004), Hendel and Nevo (2006), Aguirregabiria and Mira (2008)) and finance (e.g. Burton and Miller (2006)).

A branch of recent literature on DBCP have studied the estimation of parametric DBCP under increasingly complicated extensions, such as unobserved heterogeneity and serial correlation in the unobserved states. (See Arcidiacono (2003), Brien, Lillard and Stern (2006).) Another branch studied the nonparametric identification of DBCP. Several recent works established that the DBCP model is nonparametrically unidentified in the sense that the SPP cannot be uniquely recovered from the choice patterns observed in data-generating processes (DGP), even when the distribution of unobserved state variables (USV) is known to researchers. (See Magnac and Thesmar (2002), Aguirregabiria (2005), Pesendorfer and Schmidt-Dengler (2007)) Berry and Tamer (2006) showed for the special case of dynamic optimal stopping process (where one of the alternatives is terminal and yields a payoff independent of states) that if USV distribution is known, then the static payoff from the non-terminal choice is nonparametrically identified. Another solution for identifying DBCP model is to introduce an observable outcome variable that can aid the identification. (See Heckman and Navarro (2007), Aguirregabiria (2008).)

This paper contributes to this branch of literature by showing that the counterfactual choice probabilities in the DBCP model can be (informatively) partially identified, despite these strong non-identification results. We first motivate our work by providing two new findings about limitations of nonparametric DBCP models. First, the DBCP model is not testable without restrictions on SPP, even when USV distributions are known to researchers. Second, the practice of "setting the SPP from one of the two choices to an arbitrary constant utility vector in order to identify SPP from the other choice" is not innocuous for certain type of counterfactual predictions if the actual SPP is not independent of states in either choices.[2] We propose two solutions. First, if a data generating process (DGP) has exogenous variations in observable state transitions, the DBCP becomes testable and SPP is identified. To our knowledge, this is the first result that taps into exogenous

[2] When the acutual SPP from one choice is independent of states, then it is innocuous to normalize it to a constant utility vector (provided there are no other exogenous shape restrictions on SPP).

variations in state transitions to identify SPP without involving arbitrary assignment of utility levels to one of the actions.[3] Second, we introduce the concept of rationalizable counterfactual choice probabilities (RCCP). These are defined as the counterfactual choice patterns that would be rationalized, jointly with choices observed in DGP, by primitives that satisfy the nonparametric model restrictions (such as stochastic restrictions on the USV distribution, or shape restrictions on SPP). We show how to exhaust the identifying power of such restrictions to recover sharp bounds (or the identified set) of RCCP efficiently when USV distribution is known.

The second half of the paper is motivated by the need for a dramatic generalization of the idea of partially identifying RCCP when the USV distribution is not known to researchers. Misspecifying USV distribution in structural estimations can lead to errors in counterfactual predictions. That is, model primitives estimated under incorrect parametric assumptions on USV can imply counterfactual outcomes that deviate from true counterfactuals. We introduce a simple, novel algorithm that can recover the sharp bounds on RCCP in the absence of parametric assumptions on either SPP or distributions of USV. Our results should not be interpreted as refuting the use of parametric assumptions in structural estimations of DBCP. Rather, our main objective is to provide a formal characterization of the limits of nonparametric structural analyses of DBCP models, and offer a powerful framework where stochastic restrictions on USV and exogenous shape restrictions on SPP are exploited efficiently to derive sharp bounds on rationalizable counterfactual choice probabilities (RCCP).

Our approach exploits two important features of the DBCP model: First, individuals' dynamic rationality, both in the DGP and counterfactual decision environment, can be formulated as a linear, homogenous system of SPP and nuisance parameters that are functionals of the USV distribution. The choice probabilities, including those observed in DGP and those to be inferred in the counterfactual context, enter the system of linear restrictions through the coefficient matrix. Second, stochastic restrictions on USV distributions (such as independence of USV from observable states) can be equivalently represented as linear inequalities of these nuisance (distributional) parameters, again with choice probabilities in the DGP and counterfactual settings entering the coefficients. Furthermore, in lots of empirical contexts, SPP are often known to satisfy simple linear restrictions, such as ranking of SPP among a subset of observable states. Hence identifying the set of *all* RCCP amounts to collecting choice patterns that would make such a linear system feasible with solutions in SPP and the nuisance parameters. This perspective allows us to use a simple algorithm of linear programming to recover the complete set of RCCP consistent with the DBCP model by checking whether a choice pattern makes such linear systems feasible. We use several numerical, simulated examples to show that the algorithm can yield very informative sets of RCCP in practice.

[3] See the introduction of Section 3 for difference between this paper and some ealrier papers that discussed the identifying power of such exogenous variations.

Our work contributes to the literature on structural analyses of DBCP in two important ways. First, we measure effects of counterfactual policies directly in terms RCCP and does not involve an intermediate step of identifying model structures. Second, our approach only requires nonparametric stochastic restrictions on USV distribution and shape restrictions on SPP. Thus it sheds lights on the limit of what can be learned about counterfactuals when econometricians choose to remain agnostic about the functional form of these model structures. More generally the algorithm applies whenever model restrictions, economic or statistical, can be equivalently represented as systems of linear inequalities.

The rest of this paper is organized as follows. Section 2 reviews the empirical content of the dynamic binary choice model and introduce the two new findings about limitations of a nonparametric DBCP. Section 3 explains the testability of DBCP and the identification of SPP when there is exogenous variation in the transition of states and the USV distribution is known. Section 4 introduces a simple method for finding the identified set of RCCP in the benchmark case where USV distributions are known. Section 5 explains how to recover the identified set of RCCP when USV is only known to be independent of observable states and SPP is known to satisfy exogenous shape restrictions. We illustrate our arguments and algorithms in Section 3,4,5 through several numerical examples. Section 6 concludes. Details of proofs and implementation of the algorithms are included in the appendix.

2 The Empirical Content of Dynamic Binary Choice Processes

In this section, we study the testability and identification of the DBCP model when the distribution of USV is known. We first revisited earlier negative identification results in the literature (Magnac and Thesmar (2002), Pesendorfer and Schmidt-Dengler (2007), Aguirregabiria (2008)) which attributed the source of non-identification of SPP to insufficient ranks in a system of linear equations. Like these previous works, we also focus on discrete supports for observable state variables (OSV) while leaving support of USV unrestricted. We then introduce two new results: (i) the DBCP model is not testable without further restrictions on SPP; and (ii) setting SPP from one of the choices to arbitrary constant utility vectors to identify that from the other can lead to errors in counterfactual choice patterns predicted if the actual SPP in DGP is not independent of observable states. We also illustrate our arguments with a numeric example. Our discussions in this section motivates the two subsequent sections, which address the issues of how to test the model and partially identify the counterfactuals without assigning arbitrary values to $\mathbf{u}_0$.

2.1 Preliminaries

In this section, we specify the model of dynamic binary choice process (DBCP) and define the core concepts such as testability and identification of the model. Consider a single-agent, DBCP in an infinite horizon. The time is discrete and indexed by t. In each period, the decision maker observes states $\mathbf{S}_t = (\mathbf{X}_t, \boldsymbol{\epsilon}_t)$ (where $\epsilon_t \equiv (\epsilon_{1t}, \epsilon_{0t}) \in \mathbb{R}^2$) with support $\Omega_{\mathbf{S}} = \Omega_{\mathbf{X},\boldsymbol{\epsilon}} \subseteq \mathbb{R}^{D+2}$, and chooses j_t from $\mathbf{J} = \{0, 1\}$.[4] The state space $\Omega_{\mathbf{S}}$ is fixed over time. In all periods, the decision maker observes both $\mathbf{X}_t$ and $\boldsymbol{\epsilon}_t$, while econometricians only observe $\mathbf{X}_t$, but not $\boldsymbol{\epsilon}_t$. The return for the decision maker each period is $\upsilon(\mathbf{S}_t, j_t) : \Omega_{\mathbf{S}} \otimes \mathbf{J} \to \mathbb{R}^1$ for all t. Conditional on current states $\mathbf{S}$ and action j, distribution of states in the next period $\mathbf{S}'$ is given by the transition function $H_j(\mathbf{S}'|\mathbf{S}) : \Omega_{\mathbf{S}} \otimes \Omega_{\mathbf{S}} \to [0, 1]$. The decision maker has a constant discount factor $\beta \in (0, 1)$ forever. Both υ and $\{H_j\}_{j=1,2}$ are fixed over time. We drop time subscripts due to the time-homogeneity of υ, $\{H_j\}_{j=1,2}$, $\Omega_{\mathbf{S}}$. The decision maker chooses a deterministic, Markovian decision rule $j(\mathbf{s})$ that maximizes the sum of expected present and future payoffs: $E[\sum_{s=0}^{\infty} \beta^s \upsilon(\mathbf{S}_{t+s}, j_{t+s})|\mathbf{S}_t, j_t]$.[5] We assume β is known to econometricians. In addition, the following restrictions are maintained throughout the paper.

AS (Additive separability) $\upsilon(\mathbf{s}, j) = u_j(\mathbf{x}) + \varepsilon_j$ *for all* $(\mathbf{x}, \boldsymbol{\epsilon}, j)$, *where* $E(\varepsilon_j|\mathbf{x}) = 0$ *for all* $(\mathbf{x}, j)$*;*

CI (Conditional independence) $H_j(\mathbf{s}'|\mathbf{s}) = F_{\boldsymbol{\epsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')G^j(\mathbf{x}'|\mathbf{x})$ $\forall \mathbf{s}, \mathbf{s}' \in \Omega_{\mathbf{S}}$, $j \in \{0, 1\}$, *where* $F_{\boldsymbol{\epsilon}|\mathbf{X}}(.|x)$ *and* $G^j(.|x)$ *are conditional distributions of* ϵ *and* X' *given* $x \in \Omega_{\mathbf{X}}$ *and* $j \in \{0, 1\}$.

DS (Discrete support) The space of observable states is $\Omega_{\mathbf{X}} = \{\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_K\}$, *with* $\mathbf{x}_k \in \mathbb{R}^D$ *for all* $k \in \{1, ., K\}$.

The transitions $\mathbf{G} \equiv [G^1 \; G^0]$ are directly recovered from data of observed states and actions $\{j_t, \mathbf{x}_t\}_{t=0}^{+\infty}$. Let $\Gamma \equiv \{\Omega_{\mathbf{X}}, \beta, \mathbf{G}\}$ denote model primitives, or decision environments, that are directly observable to researchers in a DGP. Note *AS* alone is merely a reparametrization rather than a substantial restriction – the SPP is the expected single-period payoff given $\mathbf{x}$. *CI* requires that persistence between current and future states to be fully captured by dynamics between $\mathbf{x}'$ and $\mathbf{x}$, and current actions affect future states only through $\{G^j\}_{j=0,1}$. Thus the unknown parameters are the decision makers' expected single-period payoffs $\mathbf{u} \equiv [u_0(.) \; u_1(.)]$ (SPP) and the USV distribution $F_{\boldsymbol{\epsilon}|\mathbf{X}}$.[6] We collect some regularity conditions which are necessary only for ensuring the DBCP has

[4] Throughout the paper I use bold letters to denote vectors and matrices.

[5] In general, the optimal policies should be a function of past histories $\mathbf{h}_t = \{\mathbf{s}_j\}_{j=0}^{t}$. However Strauch (1966) showed for any history-dependent poilicy and starting state, there always exists a deterministic, Markovian policy (a policy that depends on the current state only) with the same expeced total discounted payoff. The implication is that for analysis of optimal policies, it suffices to focus on Markovian stationary policies. Throughout the paper we focus on the case where the agent only considers deterministic Markovian policies.

[6] Given our focus on Markovian policies, *CI* implies $\Pr(j_t = 1|\mathbf{x}_t, \mathbf{x}_{t-1}, ., \mathbf{x}_0) = \Pr(j_t = 1|\mathbf{x}_t)$ for all history $(\mathbf{x}_t, \mathbf{x}_{t-1}, ., \mathbf{x}_0)$, which is a testable implication itself.

a well-defined static representation.

REG (Regularity Conditions) (i) For $j \in \{0,1\}$, $u_j \in B(\Omega_{\mathbf{X}})$, *where* $B(\Omega_{\mathbf{X}})$ *is the set of bounded, continuous, real-valued functions on* $\Omega_{\mathbf{X}}$*; (ii) For* $j \in \{0,1\}$, G^j *satisfies the Feller Property;*[7] *(iii) For all* $\mathbf{x} \in \Omega_{\mathbf{X}}$, $j \in \{0,1\}$, $E[\max_{k\in\{0,1\}}\{\varepsilon_{t+1,k}\}|\mathbf{x}_t, j] < \infty$.

To give a general definition of identification, let U, $\mathcal{F}$ denote the parameter space of SPP and USV distributions that satisfy certain generic restrictions (such as parametric specifications of $\mathbf{u}$ or $F_{\boldsymbol{\epsilon}|\mathbf{X}}$, or stochastic restrictions on $F_{\epsilon|\mathbf{X}}$). A DBCP model, and restrictions imposed on its parameters, are fully summarized by $\{\Gamma, U, \mathcal{F}\}$. Let $\mathbf{p} \equiv (p(\mathbf{x}_1), ., p(\mathbf{x}_K))$ denote a generic vector of choice patterns, where $p(\mathbf{x}_k) \equiv \Pr(J = 1|X = \mathbf{x}_k)$. Let $\phi(\mathbf{x}_k; \mathbf{u}, F_{\epsilon|X})$ denote the set of choice probabilities for "the decision maker to chooses alternative 1 conditional on $\mathbf{x}_k$" that is rationalized by a generic pair of parameters $\mathbf{u}, F_{\epsilon|X}$. Later in this section we shall give an implicit definition of ϕ (which is the solution of a nonlinear system given $\mathbf{u}, F_{\epsilon|\mathbf{X}}$, and hence in general may be a correspondence rather than a function).

Definition 1 *A* K*-vector of choice patterns* $\mathbf{p}$ *is rationalized by (or consistent with) a DBCP model* $\{\Gamma, U, \mathcal{F}\}$ *if* $p(\mathbf{x}_k) \in \phi(\mathbf{x}_k; \mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}}, \Gamma)$ *for all* $\mathbf{x}_k \in \Omega_{\mathbf{X}}$ *for some* $(\mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}}) \in U \otimes \mathcal{F}$*. Given a model* $\{\Gamma, U, \mathcal{F}\}$*, the set of rationalizable choice patterns, or testable implications, is the set of all* $\mathbf{p} \in [0,1]^K$ *that are consistent with* $\{\Gamma, U, \mathcal{F}\}$*.*

Definition 2 *Given an observed choice pattern* $\mathbf{p}^*$ *and a model* $\{\Gamma, U, \mathcal{F}\}$ *for the DGP, the identified set of* $(\mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}})$ *is the subset of* $U \otimes \mathcal{F}$ *such that* $p^*(\mathbf{x}_k) \in \phi(\mathbf{x}_k; \mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}}, \Gamma)$ *for all* $\mathbf{x}_k \in \Omega_{\mathbf{X}}$*. The identified set of* $\mathbf{u}$ *in* U *under* $\mathcal{F}$ *is the set of all* $\mathbf{u}$ *in* U *such that* $\exists F_{\epsilon|\mathbf{X}} \in \mathcal{F}$ *with* $p^*(\mathbf{x}_k) \in \phi(\mathbf{x}_k; \mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}}, \Gamma)$ *for all* $\mathbf{x}_k \in \Omega_{\mathbf{X}}$*.*

Note the definition of identification is always relative to the choice patterns observed in DGP. We say a model is testable if its testable implications form a *strict* subset of $[0,1]^K$ (or equivalently, if there exists $\mathbf{p}$ in $[0,1]^K$ such that the corresponding identified set of $(\mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}})$ is empty).

2.2 Non-testability and non-identification

We start from a benchmark case where the distribution of USV is completely known to the researcher, and show even in this most restrictive case, the model is neither testable nor identified without further restrictions on the form of the parameters. The non-identification result was established by earlier works of Magnac and Thesmar (2002), Aguirregabiria (2005), Pesendorfer and Schmidt-Dengler (2007), but the non-testability result is new. We also offer a new interpretation

[7] $G^j(\mathbf{x}'|\mathbf{x})$ satisfies the *Feller Property* if for each bounded, continuous function $f : \Omega_{\mathbf{X}} \to \mathbb{R}^1$, $\int f(\mathbf{x}')dG^j(\mathbf{x}'|\mathbf{x})$ is also bounded and continuous in $\mathbf{x}$.

of what can be identified nonparametrically from the model with knowledge of the USV distribution: the difference in expected present values between two trivial policies of sticking to one of the alternatives unconditionally forever.[8]

Let $\mathbf{x}_k \equiv (x_{k1}, x_{k2}, ., x_{kD})$ for $k = 1, ., K$ and $p_k \equiv p(\mathbf{x}_k)$and $\mathbf{p} \equiv [p_1, ., p_K]$. Let the K-by-D matrix $\Omega_{\mathbf{X}} \equiv [\mathbf{x}_1\ \mathbf{x}_2\ ..\ \mathbf{x}_K]'$ denote the support of $\mathbf{X}$. Let $\mathbf{G}^j$ denote matrices of transitions when j is chosen, with its (m, n)-th component $G^j_{m,n} \equiv \Pr(\mathbf{x}_n|\mathbf{x}_m, j)$. Let $\Gamma \equiv \{\beta, \mathbf{G}^0, \mathbf{G}^1, \Omega_{\mathbf{X}}\}$ denote structural elements of the model that are known or observable to the researcher. Define

$$\mathbf{G}^j_\infty \equiv \lim_{T\to\infty} \sum_{t=1}^T \beta^t [\mathbf{G}^j]^t$$

where $[\mathbf{G}^j]^t$ denotes the t-th power of $\mathbf{G}^j$. The limit exists since all entries in the matrix $\sum_{t=1}^T \beta^t [\mathbf{G}^j]^t$ are monotone sequences of non-negative numbers smaller than 1. Let $\Delta\epsilon \equiv \epsilon_0 - \epsilon_1$, and for a generic choice pattern $\mathbf{p}$ and specification of the distribution of USV $F_{\Delta\epsilon|X}$, define $\mathbf{Q}(\mathbf{p}; F_{\Delta\epsilon|\mathbf{X}})$ as a K-vector with $Q_k \equiv F^{-1}_{\Delta\epsilon|\mathbf{x}_k}(p_k)$, and $\boldsymbol{\kappa}^0$, $\boldsymbol{\kappa}^1$ be K-vectors which depend on $\mathbf{p}$ and $F_{\Delta\epsilon|\mathbf{X}}$, with k-th coordinates defined as

$$\begin{aligned} \kappa^0_k &\equiv \kappa^0(p_k; F_{\Delta\epsilon|\mathbf{x}_k}) \equiv \int_{-\infty}^{Q_k} (Q_k - s) dF_{\Delta\epsilon|\mathbf{x}_k}(s) \\ \kappa^1_k &\equiv \kappa^1(p_k; F_{\Delta\epsilon|\mathbf{x}_k}) \equiv \int_{Q_k}^{+\infty} [s - Q_k] dF_{\Delta\epsilon|\mathbf{x}_k}(s) \end{aligned} \tag{1}$$

We shall refer to $\kappa^0(.; F_{\Delta\epsilon|\mathbf{X}})$ as the "truncated surplus function" (TSF) related to the unobservable state distribution $F_{\Delta\epsilon|\mathbf{X}}$. Note $\boldsymbol{\kappa}^0(\mathbf{p}; F_{\Delta\epsilon|\mathbf{X}}) - \boldsymbol{\kappa}^1(\mathbf{p}; F_{\Delta\epsilon|\mathbf{X}}) = \mathbf{Q}(\mathbf{p}; F_{\Delta\epsilon|\mathbf{X}})$ for any $(\mathbf{p}, F_{\Delta\epsilon|\mathbf{X}})$, for $E(\Delta\epsilon|\mathbf{x}_k) = 0$. Let $\mathbf{u}_j$ denote a K-vector with its k-th element being the unknown payoff function $u_j(\mathbf{x}_k)$ for $j = 0, 1$. Let $\mathbf{A}_j(\Gamma) \equiv \mathbf{I} + \mathbf{G}^j_\infty$ for $j = 0, 1$ where $\mathbf{I}$ is the K-by-K identity matrix.

Lemma 1 *Suppose a model $\{\Gamma, U, \mathcal{F}\}$ satisfies AS, CI, DS and REG (i)-(iii). Then for a $\mathbf{p}^*$ observed in DGP, the identified set of $(\mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}})$ is the subset in $U \otimes \mathcal{F}$such that:*

$$[\mathbf{A}_1(\Gamma),\ -\mathbf{A}_0(\Gamma)] \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_0 \end{bmatrix} = [\mathbf{A}_1(\Gamma),\ \mathbf{A}_0(\Gamma) - \mathbf{A}_1(\Gamma)] \begin{bmatrix} \mathbf{Q}(\mathbf{p}^*; F_{\Delta\epsilon|\mathbf{X}}) \\ \boldsymbol{\kappa}^0(\mathbf{p}^*; F_{\Delta\epsilon|\mathbf{X}}) \end{bmatrix} \tag{2}$$

The set of solutions to this system of nonlinear equations in $\mathbf{p}^*$ (denoted as $\phi(\mathbf{x}_k; \mathbf{u}, F_{\epsilon|X})$) is the set of all choice probabilities that can be rationalized by a generic pair of parameters $\mathbf{u}, F_{\epsilon|X}$. With knowledge of $F_{\Delta\epsilon|\mathbf{X}}$, we can identify the left-hand side of the equation, which is the difference in present values between two trivial policies of sticking to one of the alternatives unconditionally forever. An equivalent statement of the lemma is that a choice pattern $\mathbf{p}^*$ is consistent with a model $\{\Gamma, U, \mathcal{F}\}$ if and only if it makes the linear system (2) feasible with solutions in $\mathbf{u}$, $\mathbf{Q}$ and $\boldsymbol{\kappa}^0$. The proof is based on the observation that expected (optimal) continuational payoffs conditional

[8] It can be shown that this interpretation is equivalent to the identified feature mentioned in Aguirregabiria (2005).

on current state $\mathbf{x}$ can be decomposed as the sum of a functional of $\mathbf{u}$ and a functional of $F_{\Delta\epsilon|\mathbf{X}}$.[9] An immediate corollary of the lemma is that the scale of $\mathbf{u}$ and $F_{\Delta\epsilon|\mathbf{X}}$ can not be jointly identified in any DGP without additional restrictions. To see this, note for any ($\mathbf{u}$, $F_{\Delta\epsilon|\mathbf{X}}$) consistent with $\{\Gamma, U, \mathcal{F}\}$, their scale transformation ($\tilde{\mathbf{u}}$, $\tilde{F}_{\Delta\epsilon|\mathbf{X}}$) is also consistent with $\{\Gamma, U, \mathcal{F}\}$, where $\tilde{\mathbf{u}}_1 \equiv a\mathbf{u}_1$, $\tilde{\mathbf{u}}_0 \equiv a\mathbf{u}_0$ and $\tilde{F}_{\Delta\epsilon|\mathbf{X}}(t) \equiv F_{\Delta\epsilon|\mathbf{X}}(\frac{t}{a})$ for all t for some constants $a \in \mathbb{R}^1_{++}$. More importantly, note the transformation only involves a positive constant a independent of $(\mathbf{G}^0, \mathbf{G}^1)$ or $\mathbf{u}$. This implies any scale normalization of $F_{\Delta\epsilon|\mathbf{X}}$ must be innocuous for identifying counterfactual choice patterns when either $\mathbf{u}$ or $(\mathbf{G}^0, \mathbf{G}^1)$ are perturbed. Thus, for example, there will be no loss of generality in assuming $F_{\Delta\epsilon|\mathbf{X}}$ is standard normal, if the true USV distribution is known to belong to the normal family. The same argument applies for any parametric location-scale families. (Note there already is a location normalization of $F_{\Delta\epsilon|\mathbf{X}}$ as the assumption AS requires $E(\Delta\epsilon|\mathbf{x}) = 0$.) Next, we use Lemma 1 to derive two (very) negative results about the empirical content of the DBCP model.

Proposition 1 *Consider a model* $\{\Gamma, U, F_{\boldsymbol{\epsilon}|\mathbf{X}}\}$ *that satisfies AS, CI, REG (i)-(iv), and suppose* $F_{\Delta\epsilon|\mathbf{X}}$ *is known. Then (i) any* $\mathbf{p}$ *in* $[0,1]^K$ *is consistent with the model, and (ii) for all* $\mathbf{p} \in [0,1]^K$, $\mathbf{u}$ *is not identified.*

Part (ii) was shown in Magnac and Thesmar (2002), Pesendorfer and Schmidt-Dengler (2007) and Aguirregabiria (2005). The proof of part (i) proceeds by showing the linear system in (2) always has solutions in $\mathbf{u}$ regardless of $\mathbf{p}$ and $F_{\Delta\epsilon|\mathbf{X}}$ in the right-hand side. This implies the model is not testable even when $F_{\Delta\epsilon|\mathbf{X}}$ is known, and any choice pattern observed in data can be rationalized as individuals' optimal behavior given certain structure. This point can be better illustrated by drawing the analogy with static binary choice models (the special case with $\beta = 0$). The static model is not testable without restrictions on $\mathbf{u}$ since for any chosen $F_{\Delta\epsilon|\mathbf{X}}$, all $\mathbf{p} \in [0,1]^K$ can be consistent with the model with appropriately chosen $\mathbf{u}$ such that $\mathbf{u}_1 - \mathbf{u}_0 = [F^{-1}_{\Delta\epsilon|\mathbf{x}_1}(p_1), .. , F^{-1}_{\Delta\epsilon|\mathbf{x}_K}(p_K)]'$. However, the non-testability in the presence of real dynamics (with $\beta > 0$) cannot be taken for granted, for it is not true that a non-homogenous linear system of equations with more unknowns than equations, such as in (2), are always feasible with solutions. To our knowledge this is the first formal proof of non-testability of the DBCP model.

2.3 Linear restrictions and counterfactuals[10]

In practical structural estimation, econometricians sometimes set $\mathbf{u}_0 = 0$ in order to identify $\mathbf{u}_1$. This is often considered a necessary locational normalization, since the linear system in (2) has $2K$ equations and K unknowns. However, in this subsection, we show that, when the real $\mathbf{u}_0$ in the DGP is not independent of states, setting $\mathbf{u}_0 = \mathbf{0}$ amounts to imposing a linear restriction

[9] This idea was initially proposed by Hotz and Miller (1993) in a different form.

[10] I am indebted to Ken Wolpin for detailed comments and discussions that help improve this section. All errors (if any) are my own.

that deviates from the truth, and is not an innocuous normalization. We shall show that setting $\mathbf{u}_0$ to an arbitrary utility vector leads to discrepancies between the predicted and the real choice probabilities in *some (but not all)* counterfactual contexts.

We consider two classes of counterfactual policy changes: (a) perturbing agents' static payoffs each period, or (b) changing transitions between state variables. In both cases, the distribution of USV is left unchanged. Let $\{\tilde{\Gamma}, \Delta\}$ denote the counterfactual policies considered, where $\tilde{\Gamma}$ denotes new transitions of observable states $\mathbf{x}$ and $\Delta \equiv (\Delta_1, \Delta_0)$ denotes changes in SPP. For a given model characterized by $\{\Gamma, U, \mathcal{F}\}$, identifying the set of all rationalizable counterfactual choice probabilities (RCCP) amounts to finding all $\tilde{\mathbf{p}}$ such that an "augmented" linear system consisting of *both* (2) and

$$[\mathbf{A}_1(\tilde{\Gamma}),\ -\mathbf{A}_0(\tilde{\Gamma})] \begin{bmatrix} \mathbf{u}_1 + \Delta_1 \\ \mathbf{u}_0 + \Delta_0 \end{bmatrix} = [\mathbf{A}_1(\tilde{\Gamma}),\ \mathbf{A}_0(\tilde{\Gamma}) - \mathbf{A}_1(\tilde{\Gamma})] \begin{bmatrix} \mathbf{Q}(\tilde{\mathbf{p}}; F_{\Delta\epsilon|\mathbf{X}}) \\ \boldsymbol{\kappa}^0(\tilde{\mathbf{p}}; F_{\Delta\epsilon|\mathbf{X}}) \end{bmatrix} \tag{3}$$

is satisfied jointly for some $\mathbf{u} \in U$ and $F_{\epsilon|\mathbf{X}} \in \mathcal{F}$. (We shall drop $F_{\Delta\epsilon|X}$ from $\mathbf{Q}$ and $\boldsymbol{\kappa}^0$ later for notational ease.) To simplify our exposition, we will discuss the impact of setting "$\mathbf{u}_0 = \mathbf{0}$" on the two types of counterfactuals separately.

Consider the first class of policy changes where SPP is perturbed while Γ remains the same. Let $\mathbf{B}(\Gamma) \equiv [\mathbf{A}_1(\Gamma),\ \mathbf{A}_0(\Gamma) - \mathbf{A}_1(\Gamma)]$. Characterize the DGP by the following linear system

$$\mathbf{A}_1(\Gamma)\mathbf{u}_1 - \mathbf{A}_0(\Gamma)\mathbf{u}_0 = \mathbf{B}(\Gamma)[\mathbf{Q}(\mathbf{p}^*)', \boldsymbol{\kappa}^0(\mathbf{p}^*)']'$$

where $\mathbf{p}^*$ is the choice outcome observed in DGP. We are interested in predicting dynamic rational choice outcomes $\tilde{\mathbf{p}}$ if $\mathbf{u}_j$ are perturbed to $\tilde{\mathbf{u}}_j \equiv \mathbf{u}_j + \boldsymbol{\Delta}_j$ for $j = 1, 0$ respectively. Suppose econometricians normalize $\mathbf{u}_0$ to an arbitrary constant vector $\bar{\mathbf{u}}_0$ and recover $\mathbf{u}_1$ as

$$\bar{\mathbf{u}}_1 = \mathbf{A}_1(\Gamma)^{-1} \left\{ \mathbf{B}(\Gamma)[\mathbf{Q}(\mathbf{p}^*)', \boldsymbol{\kappa}^{\mathbf{0}}(\mathbf{p}^*)']' + \mathbf{A}_0(\Gamma)\bar{\mathbf{u}}_0 \right\} \tag{4}$$

Then identifying counterfactual choice patterns amounts to finding $\tilde{\mathbf{p}}$ such that

$$\mathbf{A}_1(\Gamma)(\bar{\mathbf{u}}_1 + \boldsymbol{\Delta}_1) - \mathbf{A}_0(\Gamma)(\bar{\mathbf{u}}_0 + \boldsymbol{\Delta}_0) = \mathbf{B}(\Gamma)[\mathbf{Q}(\tilde{\mathbf{p}})', \boldsymbol{\kappa}^0(\tilde{\mathbf{p}})']' \Leftrightarrow$$
$$\mathbf{B}(\Gamma)[\mathbf{Q}(\mathbf{p}^*)', \boldsymbol{\kappa}^0(\mathbf{p}^*)']' + \mathbf{A}_1(\Gamma)\boldsymbol{\Delta}_1 - \mathbf{A}_0(\Gamma)\boldsymbol{\Delta}_0 = \mathbf{B}(\Gamma)[\mathbf{Q}(\tilde{\mathbf{p}})', \boldsymbol{\kappa}^0(\tilde{\mathbf{p}})']'$$

The latter equation above is a system of nonlinear equations in $\tilde{\mathbf{p}}$ that does not depend on the choice of $\bar{\mathbf{u}}_0$ given knowledge of Γ, $F_{\epsilon|X}$ and observation of $\mathbf{p}^*$. Thus setting $\mathbf{u}_0 = \bar{\mathbf{u}}_0$ has no impact on the set of rationalizable counterfactual outcomes $\tilde{\mathbf{p}}$ that can be recovered from $\mathbf{p}^*$ observed, provided $F_{\Delta\epsilon|\mathbf{X}}$ is known.

Now consider the second class of counterfactuals labelled as type (b) above, where $\mathbf{A}_j(\Gamma)$ is perturbed to $\mathbf{A}_j(\tilde{\Gamma})$ while $\mathbf{u}_j$ are fixed for both $j = 1, 2$ and $F_{\epsilon|X}$ is known. Again, if we normalize

$\mathbf{u}_0 = \bar{\mathbf{u}}_0$, then $\mathbf{u}_1$ is recovered as in (4) from the DGP. The counterfactual analyses amount to recovering $\tilde{\mathbf{p}}$ such that

$$\begin{aligned}&\tilde{\mathbf{A}}_1\bar{\mathbf{u}}_1 - \tilde{\mathbf{A}}_0\bar{\mathbf{u}}_0 = \mathbf{B}(\tilde{\Gamma})[\mathbf{Q}(\tilde{\mathbf{p}})', \boldsymbol{\kappa}^0(\tilde{\mathbf{p}})']' \Leftrightarrow \\ &\tilde{\mathbf{A}}_1\mathbf{A}_1^{-1}\mathbf{B}(\Gamma)[\mathbf{Q}(\mathbf{p}^*)', \boldsymbol{\kappa}^0(\mathbf{p}^*)']' + \left(\tilde{\mathbf{A}}_1\mathbf{A}_1^{-1}\mathbf{A}_0 - \tilde{\mathbf{A}}_0\right)\bar{\mathbf{u}}_0 = \mathbf{B}(\tilde{\Gamma})[\mathbf{Q}(\tilde{\mathbf{p}})', \boldsymbol{\kappa}^0(\tilde{\mathbf{p}})']'\end{aligned}$$

where $\tilde{\mathbf{A}}_j, \mathbf{A}_j$ are shorthands for $\mathbf{A}_j(\tilde{\Gamma})$ and $\mathbf{A}_j(\Gamma)$ respectively and $\mathbf{B}(\tilde{\Gamma}) \equiv [\tilde{\mathbf{A}}_1, \tilde{\mathbf{A}}_0 - \tilde{\mathbf{A}}_1]$. (For simplicity in exposition, we focus on the case where both $\mathbf{A}_j, \tilde{\mathbf{A}}_j$ are full-rank matrices.) If the perturbation from Γ to $\tilde{\Gamma}$ only involves changing the discount factor from β to $\tilde{\beta}$ while the transition between observable states $\mathbf{G}^j$ are fixed, then $\left(\tilde{\mathbf{A}}_1\mathbf{A}_1^{-1}\mathbf{A}_0 - \tilde{\mathbf{A}}_0\right) = 0$ and setting $\mathbf{u}_0$ to any arbitrary vector $\bar{\mathbf{u}}_0$ is innocuous. On the other hand, if $\mathbf{G}^j$ is perturbed to $\tilde{\mathbf{G}}^j$ in $\tilde{\Gamma}$ while β is fixed, then in general, $\tilde{\mathbf{A}}_1^{-1}\mathbf{A}_1^{-1}\mathbf{A}_0 - \tilde{\mathbf{A}}_0$ is not trivially a zero matrix. As Example 1 below shows, setting $\mathbf{u}_0 = \bar{\mathbf{u}}_0$ may induce discrepancies between the predicted and the true counterfactual outcomes if $\left(\tilde{\mathbf{A}}_1\mathbf{A}_1^{-1}\mathbf{A}_0 - \tilde{\mathbf{A}}_0\right)(\bar{\mathbf{u}}_0 - \mathbf{u}_0^*) \neq \mathbf{0}$ (where $\mathbf{u}_0^*$ denotes the true SPP in the DGP). It can also be shown that in the special case where the actual $\mathbf{u}_0^*$ in the DGP is independent of states, then normalizing $\mathbf{u}_0$ to a constant vector in structural estimation is innocuous.[11] Below we present a numerical example of such a scenario.

Example 1 *(Impact of arbitrary assignment of* $\mathbf{u}_0$ *on counterfactuals)* Consider a simple example with $K = 3$, $\beta = 0.75$ and define two sets of transitions $\mathbf{G}^j, \tilde{\mathbf{G}}^j$ as:

$$\mathbf{G}^1 \equiv \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \mathbf{G}^0 \equiv \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \tilde{\mathbf{G}}^1 \equiv \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}, \tilde{\mathbf{G}}^0 \equiv \begin{bmatrix} \frac{1}{5} & \frac{1}{5} & \frac{3}{5} \\ \frac{3}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{3}{5} & \frac{1}{5} \end{bmatrix}$$

where $\mathbf{G}^1, \mathbf{G}^0$ are transitions in the DGP, and $\tilde{\mathbf{G}}^1, \tilde{\mathbf{G}}^0$ are counterfactual transitions under which choice outcomes are to be inferred. Suppose $\Delta\epsilon$ is independent of $\mathbf{X}$ and uniformly distributed over $[-1, 1]$. Then the quantile function is $F_{\Delta\epsilon}^{-1}(p) = 2p - 1$ for $p \in (0, 1)$ and $\kappa_k^0(p_k) = p_k^2$ for $p_k \equiv \Pr(j = 1|\mathbf{x}_k)$ and $p_k \in (0, 1)$. Suppose the true SPP are $\mathbf{u}_1^* = [1.3; 1.1; 1.7]$ and $\mathbf{u}_0^* = [1.2, 1.4, 1.0]$. Then the dynamic rationality in both DGP and the counterfactual environments are represented as systems of quadratic equations in choice probabilities. The actual set of rationalizable counterfactual choice probabilities (RCCP) is

$$\Xi(\mathbf{v}^*) \equiv \left\{ \mathbf{p} \in [0, 1]^3 : \underset{\text{3-by-6}}{\mathbf{B}(\tilde{\Gamma})} \big[\underset{\text{6-by-1}}{2\mathbf{p}' - 1 \ , \ \mathbf{p}' Diag(\mathbf{p})} \big]' = \underset{\text{3-by-1}}{\mathbf{v}^*} \right\}$$

[11] Suppose the actual $\mathbf{u}_0^* = \mathbf{1}'c$ where $\mathbf{1}$ is a vector of ones and c is a real number, and $\mathbf{u}_0$ is normalized to $\bar{\mathbf{u}}_0 = (\mathbf{1}'c)\alpha$ for some $\alpha \in \mathbb{R}^1$. Then

$$\left(\tilde{\mathbf{A}}_1\mathbf{A}_1^{-1}\mathbf{A}_0 - \tilde{\mathbf{A}}_0\right)(\bar{\mathbf{u}}_0 - \mathbf{u}_0^*) = 0 \Leftrightarrow \tilde{\mathbf{A}}_1^{-1}\tilde{\mathbf{A}}_0\mathbf{1}'c = \mathbf{A}_1^{-1}\mathbf{A}_0\mathbf{1}'c$$

But the latter must hold by definition of $\mathbf{A}_j = \mathbf{I} + \mathbf{G}_\infty^j = (\mathbf{I} - \beta\mathbf{G}^j)^{-1}$.

where $Diag(\mathbf{p})$ is a 3-by-3 diagonal matrix with the diagonal entries being $\mathbf{p}$ and

$$\mathbf{v}^* \equiv \tilde{\mathbf{A}}_1\mathbf{A}_1^{-1}\mathbf{B}(\Gamma)[\mathbf{Q}(\mathbf{p}^*)', \boldsymbol{\kappa}^0(\mathbf{p}^*)']' + \left(\tilde{\mathbf{A}}_1\mathbf{A}_1^{-1}\mathbf{A}_0 - \tilde{\mathbf{A}}_0\right)\mathbf{u}_0^*$$

and $\mathbf{B}(\Gamma)$, $\mathbf{B}(\tilde{\Gamma})$ are defined as above. We shall see there are discrepancies between the sets of RCCP predicted when the unknowable $\mathbf{u}_0^*$ is replaced with different arbitrary assigned vectors in order to estimate $\mathbf{u}_1$, and that these sets also deviate from the actual set of RCCP $\Xi(\mathbf{v}^*)$.

Suppose econometricians set $\mathbf{u}_0 = \bar{\mathbf{u}}_0 \equiv [0; 0; 0]$ in estimation. Recall $\Gamma \equiv (\beta, \mathbf{G}^1, \mathbf{G}^0)$ and $\tilde{\Gamma} \equiv \{\beta, \tilde{\mathbf{G}}^1, \tilde{\mathbf{G}}^0\}$. Then $\mathbf{u}_1$ is estimated as in (4) and the set of predicted RCCP implied under this particular choice of $\bar{\mathbf{u}}_0$ is $\Xi(\bar{\mathbf{v}})$ where $\bar{\mathbf{v}}$ is similar to $\mathbf{v}^*$, only with $\mathbf{u}_0^*$ replaced by $\bar{\mathbf{u}}_0$. The set includes a rationalizable counterfactual choice pattern: [12]

$$\mathbf{p}^c(\bar{\mathbf{u}}_0) = [0.5859; 0.3566; 0.8075]$$

Now suppose econometricians had specified $\bar{\mathbf{u}}_0' = [0.6; 0.7; 0.5]$. We choose such $\bar{\mathbf{u}}_0'$ deliberately as it is proportional to the truth $\bar{\mathbf{u}}_0$ up to a scale normalization. Then $\bar{\mathbf{u}}_1'$ is estimated as in (4), with $\bar{\mathbf{u}}_0$ replaced by $\bar{\mathbf{u}}_0'$. The set of RCCP implied by this choice of $\bar{\mathbf{u}}_0'$ is $\Xi(\bar{\mathbf{v}}')$ where $\bar{\mathbf{v}}'$ is similar to $\mathbf{v}^*$, only with $\mathbf{u}_0^*$ replaced by $\bar{\mathbf{u}}_0'$. This set includes a rationalizable counterfactual choice pattern:

$$\mathbf{p}^c(\bar{\mathbf{u}}_0') = [0.5554; 0.3644; 0.8302]$$

Substitution of $\mathbf{p}^c(\bar{\mathbf{u}}_0)$ and $\mathbf{p}^c(\bar{\mathbf{u}}_0')$ into the left-hand side of the equation defining the set $\Xi(\mathbf{v}^*)$ verifies neither of the two is in the set of RCCP. Furthermore, similar calculations show $\mathbf{p}^c(\bar{\mathbf{u}}_0) \notin \Xi(\bar{\mathbf{v}}')$ and $\mathbf{p}^c(\bar{\mathbf{u}}_0') \notin \Xi(\bar{\mathbf{v}})$. This is sufficient evidence for discrepancies among the three sets. Thus this numerical example has shown that in general setting $\mathbf{u}_0$ to an arbitrary vector may not be innocuous for analyzing policy effects of perturbations in state transitions, if the true SPP is not independent of states. In fact, as we just showed, even a scale multiplication of the true $\mathbf{u}_0^*$ can lead to errors in counterfactuals implied. (End of Example 1)

3 Exogenous Variations in OSV Transitions

Our discussions in the previous section suggest that assigning arbitrary values to $\mathbf{u}_0$ is not an innocuous solution for the non-identification and non-testability of the DBCP model. In this section, we argue that if the DGP reports sources of exogenous variations in the dynamics between state variables, then this can help identify SPP without further form restrictions on SPP or the USV distribution. In lots of empirical contexts, individuals are observed to make choices under distinct environments, where transitions to future states are different while the expected single period return remains the same. For example, this is true if dynamic decision processes have observable

[12] We use the constrained minimization command in Matlab to solve for the choice probabilities.

heterogeneities that (i) do not enter individuals' single-period payoffs; (ii) are independent of USV conditional on observable states; and (iii) affect the transitions between observable states. Below we define exogenous variation of environments formally, and give examples how they can arise in empirical contexts. Let $\{\Gamma_m\}_{m=1}^{M}$ denote a collection of decision environments with various transitions between observable states but identical discount factors β and observable state spaces $\Omega_{\mathbf{X}}$. That is, $\Gamma_m = \{\beta, \Omega_{\mathbf{X}}, \mathbf{G}_m^0, \mathbf{G}_m^1\}$. Researchers observe the DBCP in all Γ_m and the variations in $\mathbf{G}_m \equiv [\mathbf{G}_m^0, \mathbf{G}_m^1]$ are exogenous in the following sense.

EV (Exogenous variation in environments) The SPP $\mathbf{u}$ *and the USV distribution* $F_{\boldsymbol{\epsilon}|\mathbf{X}}$ *remain the same in all* Γ_m*, for* $m = 1, ., M$*.*

For example, consider the dynamic decisions of engine replacement in Rust (1987). A bus maintenance manager (Mr. Harold Zurcher) in the state of Wisconsin decides how long to operate a bus before replacing its engine with a new one. This problem is represented as a dynamic binary choice process, with the state variable $\mathbf{x}_t$ being the cumulative mileage on a bus since last engine replacement and a decision variable $j_t = 1$ if Mr. Zurcher decides to replace the bus engine and $j_t = 0$ otherwise. Suppose Mr. Zurcher is a cost minimizer and his utility each period depends on labor and costs associated with the engine (maintenance costs if $j_t = 0$ and the costs of a new engine if $j_t = 1$). Assuming that when a bus engine is replaced, it is "as good as new", the transition of $\mathbf{x}_t$ is $G(\mathbf{x}'|\mathbf{x}) = g(\mathbf{x}_{t+1})$ if $j_t = 1$ and $g(\mathbf{x}_{t+1} - \mathbf{x}_t)$ if $j_t = 0$, where $g(.)$ is the density function for distribution of mileages travelled by his bus within one decision period. Now consider similar decisions made by another maintenance manager of a different bus route. She is also a cost minimizer facing the same state variables (labor and engine-related costs on the same market) as Mr. Zurcher in each period, and shares the same utility function per period. However, she observes a different distribution of miles covered each period (denoted $\tilde{g}$), as her bus route serves in different communities than his. The differences might be due to different geographic or demographic features along the routes. To map this example into our framework, the observed heterogeneities in this case are captured by dummy or multinomial variables specifying bus routes.

Fang and Yang (2008) use a related but different assumption of exclusion restrictions where there exists a pair of observable states such that their SPP are the same while transition to future states are different. This allows them to identify the SPP in DBCP with hyperbolic discounting in the special case where the SPP from one of the actions is independent of states and therefore can be innocuously normalized to the zero vector. In comparison, the assumption of exogenous variation in state transitions in this paper is slightly stronger, but allows us to develop two new, stronger results. (i) we are able to derive specific testable implications of the DBCP model (and nonparametric identification of SPP); and (ii) we can identify SPP in a DBCP model where neither of the alternatives yield static payoffs that are independent of observable states. We also specify the rank conditions in terms of $\mathbf{G}^j$ on the transition of observable states that are necessary for attaining these new results.

3.1 Testability and identification of SPP

We first extend the definition of rationalizable choice patterns to the context where DGP reports exogenous variations in observable state transitions. Let $U, \mathcal{F}$ denote generic restrictions on $\mathbf{u}$ and $F_{\Delta\epsilon|\mathbf{X}}$, and $\Gamma_m \equiv \{\beta, \Omega_{\mathbf{X}}, \mathbf{G}^1_m, \mathbf{G}^0_m\}$ denotes the m-th decision environment in the DGP. Let $\phi(\mathbf{u}, F_{\epsilon|X})$ denote the set of choice probabilities rationalized by a generic pair of parameters $(\mathbf{u}, F_{\epsilon|X})$ (which is defined implicitly as the set of solutions to (2) as before).

Definition 3 *An observed choice pattern* $\{\mathbf{p}^*_m\}_{m=1}^M \in \mathbb{R}^{M\times K}$ *is consistent with the model* $\{\Gamma_m, U, \mathcal{F}\}_{m=1}^M$ *if* $\exists \mathbf{u}, F_{\Delta\epsilon|\mathbf{X}}$ *in* $U \otimes \mathcal{F}$ *such that* $p^*_{m,k} \in \phi(\mathbf{x}_k; \mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}}, \{\Gamma_m\}_{m=1}^M)$ *for all k in {1,.,K} and m in {1,.,M}.*

Let $\mathbf{A}^m \equiv [\mathbf{A}_1(\Gamma_m), -\mathbf{A}_0(\Gamma_m)]$ and $\mathbf{B}^m \equiv [\mathbf{A}_1(\Gamma_m), \mathbf{A}_0(\Gamma_m) - \mathbf{A}_1(\Gamma_m)]$. Denote

$$\boldsymbol{\Sigma}_L \equiv \underset{MK\times 2K}{\begin{bmatrix} \mathbf{A}^1 \\ \vdots \\ \mathbf{A}^M \end{bmatrix}} \; ; \; \boldsymbol{\Sigma}_R \equiv \underset{MK\times 2MK}{\begin{bmatrix} \mathbf{B}^1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B}^M \end{bmatrix}} \; ; \; \mathbf{W} \equiv \underset{2MK\times 1}{\begin{bmatrix} \mathbf{Q}(\mathbf{p}_1; F_{\Delta\epsilon|\mathbf{X}}) \\ \boldsymbol{\kappa}^0(\mathbf{p}_1; F_{\Delta\epsilon|\mathbf{X}}) \\ \vdots \\ \mathbf{Q}(\mathbf{p}_M; F_{\Delta\epsilon|\mathbf{X}}) \\ \boldsymbol{\kappa}^0(\mathbf{p}_M; F_{\Delta\epsilon|\mathbf{X}}) \end{bmatrix}}$$

where $\mathbf{p}_m$ is the choice pattern observed under Γ_m. Then stacking the linear restrictions across M decision environments gives us an "augmented" system of MK linear equations with $2K$ unknowns in $\mathbf{u}$. That is

$$\boldsymbol{\Sigma}_L \mathbf{u} = \boldsymbol{\Sigma}_R \mathbf{W}$$

Intuitively, an observed choice pattern $\{\mathbf{p}_m\}_{m=1}^M$ is consistent with the model with multiple environments if and only if it makes the augmented system feasible with solutions in $\mathbf{u}$. Proposition 2 formalizes this idea.

Proposition 2 *Suppose the model* $\{(\Gamma_m)_{m=1}^M, U, F_{\boldsymbol{\epsilon}|\mathbf{X}}\}$ *satisfies AS, CI, DS, REG (i)-(iv) and* $F_{\Delta\epsilon|\mathbf{X}}$ *is known to the researcher. Then (i)* $\{\mathbf{p}_m\}_{m=1}^M$ *is consistent with the model if and only if Rank*$(\boldsymbol{\Sigma}_L)$ = *Rank*$([\boldsymbol{\Sigma}_L, \boldsymbol{\Sigma}_R\mathbf{W}])$*; (ii) Suppose Rank*$(\boldsymbol{\Sigma}_L) = 2K-1$*, then for any* $\{\mathbf{p}_m\}_{m=1}^M$ *consistent with the model,* $[\mathbf{u}_1 \; \mathbf{u}_0]$ *is identified under a locational normalization* $u_{0,k} = c$ *for some* $k \in \{1, ., K\}$*.*

Remark 1: What delivers the testability of the DBCP model under multiple environment is the fact that the highest possible rank of $\boldsymbol{\Sigma}_L$ is $2K-1$. This is true regardless of the number of environments M in the model and the form of transitions in $\{\mathbf{A}^m\}_{m=1}^M$, as the sum of column vectors in $\mathbf{A}_1(\Gamma_m)$ is always equal to the sum of column vectors in $\mathbf{A}_0(\Gamma_m)$ for all m, because both must add up to be a K-vector of constants $\frac{1}{1-\beta}$ by definition. Thus any $\{\mathbf{p}\}_{m=1}^M$ that leads to discrepancies between the rank of Σ_L and $[\boldsymbol{\Sigma}_L, \boldsymbol{\Sigma}_R\mathbf{W}]$ can not be consistent with the DBCP model and this is an easily testable implication. It is important to note that the proposition assumes the knowledge of $F_{\epsilon|X}$.

Remark 2: Part (i) of Proposition 2 suggests a natural approach for testing the rationalizability of the decision maker using choice patterns observed in a DGP with multiple environments. Suppose $Rank(\mathbf{\Sigma}_L) = r \leq 2K - 1$. Let $\Lambda(\mathbf{\Sigma}_L)$ denote the set of all r-by-$2K$ submatrices of $\mathbf{\Sigma}_L$ that have full rank r. For any $\boldsymbol{\lambda} \in \Lambda(\mathbf{\Sigma}_L)$, and let $\boldsymbol{\lambda}^c$ denote the $(M \times K - r)$ row vectors that are in $\mathbf{\Sigma}_L$ but not $\boldsymbol{\lambda}$. Let $\mathbf{m}_{\boldsymbol{\lambda}}$ denote a $(M \times K - r)$-by-r matrix such that $\mathbf{m}_{\boldsymbol{\lambda}}\boldsymbol{\lambda} = \boldsymbol{\lambda}^c$. For any $\boldsymbol{\lambda} \in \Lambda(\mathbf{\Sigma}_L)$, let $\mathbf{\Sigma}_{R,\lambda}$ and $\mathbf{\Sigma}_{R,\lambda^c}$ be a r-by-$2MK$ submatrix and a $(MK - r)$-by-$2MK$ submatrix respectively formed by picking out rows in $\mathbf{\Sigma}_R$ that correspond to the rows in $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}^c$ respectively. Then an algorithm based on enumerating all elements in $\Lambda(\mathbf{\Sigma}_L)$ can be applied to find testable implications of the DBCP model. In the first round, pick any $\boldsymbol{\lambda} \in\Lambda(\mathbf{\Sigma}_L)$, and check whether

$$\left(\mathbf{m}_{\boldsymbol{\lambda}}\mathbf{\Sigma}_{R,\lambda} - \mathbf{\Sigma}_{R,\lambda^c}\right)\mathbf{W} = 0 \tag{5}$$

If the equality holds, then stop and conclude $\{\mathbf{p}_m\}_{m=1}^M$ is consistent with the model. Otherwise proceed to the next round by picking another $\tilde{\boldsymbol{\lambda}} \in\Lambda(\mathbf{\Sigma})$. Then a choice pattern $\{\mathbf{p}_m\}_{m=1}^M$ observed is inconsistent with agents' dynamic rationality *if and only if* the equality fails to hold for *all* $\boldsymbol{\lambda} \in\Lambda(\mathbf{\Sigma}_L)$. We leave the construction of statistical tests for dynamic rationality using these testable implications to future research.) These testable implications do not rely on any restriction on decision makers' static payoffs each period.

Remark 3: The algorithm above can be made more efficient by skipping redundant restrictions as they come up in the iterations. It is easy to see this by considering the computationally most feasible case with $M = 2$, $K = 3$ and $r = Rank(\mathbf{\Sigma}_L) = 5$ (the highest rank possible for $\mathbf{\Sigma}_L$ by construction). In this case, we do not need to enumerate all possible 5-by-6 submatrices in $\mathbf{\Sigma}_L$ that have rank 5. Instead it suffices to check the restriction (5) for any $\boldsymbol{\lambda} \in \Lambda(\mathbf{\Sigma}_L)$ once and there is no need to check it for any other $\tilde{\boldsymbol{\lambda}} \in \Lambda(\mathbf{\Sigma})$. This is because by the Fundamental Theorem of Linear Algebra, the dimension of the null space of the transpose $\mathbf{\Sigma}_L'$ must be one. Hence for any $\boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}} \in \Lambda(\mathbf{\Sigma}_L)$, the row vector $\mathbf{m}_{\boldsymbol{\lambda}}\mathbf{\Sigma}_{R,\lambda} - \mathbf{\Sigma}_{R,\lambda^c}$ must be proportional to $\mathbf{m}_{\tilde{\boldsymbol{\lambda}}}\mathbf{\Sigma}_{R,\tilde{\lambda}} - \mathbf{\Sigma}_{R,\tilde{\lambda}^c}$ (i.e. $\mathbf{m}_{\boldsymbol{\lambda}}\mathbf{\Sigma}_{R,\lambda} - \mathbf{\Sigma}_{R,\lambda^c} = \alpha\left(\mathbf{m}_{\tilde{\boldsymbol{\lambda}}}\mathbf{\Sigma}_{R,\tilde{\lambda}} - \mathbf{\Sigma}_{R,\tilde{\lambda}^c}\right)$ for some $\alpha \neq 0$). Therefore it is redundant to check (5) for more than one $\boldsymbol{\lambda}$ in $\Lambda(\mathbf{\Sigma}_L)$.

Remark 4: To incorporate additional restrictions on SPP (such as linear inequalities due to shape restrictions on $\mathbf{u} \equiv [\mathbf{u}_1, \mathbf{u}_0]$ such as monotonicity or ranking of $\mathbf{x}$ by SPP implied by economic theories), simply append these linear restrictions on $\mathbf{u}$ to the system of MK linear equalities and use the knowledge of $F_{\Delta\epsilon|\mathbf{X}}$ to check whether the $\{\mathbf{p}\}_{m=1}^M$ observed could make the augmented system feasible with solutions in $\mathbf{u}$.

Remark 5: Observing choice patterns under more than two exogenously varying state transitions can help with identification of $\mathbf{u}$ only if they help increase $Rank(\mathbf{\Sigma}_L)$. When $Rank(\mathbf{\Sigma}_L) =$

$r < 2K - 1$, at least $2K - r$ additional linear equality restrictions on $\mathbf{u}$ are needed to uniquely identify SPP. For instance, locational normalizations of $2K - r$ elements in $\mathbf{u}$ will be sufficient for identifying the other payoffs in $\mathbf{u}$. Such normalizations are innocuous for counterfactuals.

Example 2 *(Testable implications under multiple transitions)* Consider the case where the observable state is a scalar variable, and let $K = 3$ and $\beta = 0.8$. Researchers observe decision process under the following two environments $m = 1, 2$:

$$\mathbf{G}_1^1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \mathbf{G}_1^0 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \mathbf{G}_2^1 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}, \mathbf{G}_2^0 = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$$

Straightforward substitution of $\{\mathbf{G}_m\}_{m=1,2}$ into $\mathbf{\Sigma}_L$ shows $rank(\mathbf{\Sigma}_L) = 5$, and Gaussian elimination shows $\{\mathbf{p}_m\}_{m=1,2}$ is consistent with the model if and only if

$$Q_{11} + Q_{12} + Q_{13} = Q_{21} + Q_{22} + Q_{23}$$

where $Q_{mk} \equiv F^{-1}_{\Delta\epsilon|x_k}(p_{mk})$ and $p_{mk} \equiv \Pr(j = 1|\mathbf{x}_k; \Gamma_m)$. Suppose (ϵ_1, ϵ_0) are i.i.d. standard Type-I extreme and jointly independent of $\mathbf{X}$, then $Q(p) = \ln(p) - \ln(1 - p)$ for all $p \in (0, 1)$. Hence $\{\mathbf{p}_m\}_{m=1,2}$ is consistent with the DBCP model if and only if

$$\frac{p_{11}p_{12}p_{13}}{(1 - p_{11})(1 - p_{12})(1 - p_{13})} = \frac{p_{21}p_{22}p_{23}}{(1 - p_{21})(1 - p_{22})(1 - p_{23})} \tag{6}$$

Furthermore, normalizing u_{03} to an arbitrary constant allows us to solve for $\mathbf{u}_1$ and (u_{01}, u_{02}) through backward substitution. Figure 1 plots the set of all choice patterns in $[0, 1]^3$ that satisfies the testable implications. (End of Example 2)

4 Bounding RCCP with Linear Restrictions on SPP

In other situations where the DGP may not report any exogenous variations in the transition of observable states, and the DBCP model is neither testable nor identified due to the insufficient rank issue in (2). Yet it is still possible to extract information from the model structure and restrictions to recover the identified set of rationalizable counterfactual choice probabilities (RCCP). This is defined as the complete set of all counterfactual choice patterns that, jointly with choice probabilities observed in DGP, can be consistent with the model restrictions. The size of this set depends on the transitions in DGP and counterfactual contexts, the form of USV distribution (so far assumed to be known to econometricians), as well as any a priori restrictions on static payoffs. Consider Example 2 above. Suppose instead, econometricians observe $\mathbf{p}_1$ under $\mathbf{G}_1^1, \mathbf{G}_1^0$ in the DGP only, and are interested in learning $\mathbf{p}_2$ under counterfactual $\mathbf{G}_2^1, \mathbf{G}_2^0$ using observables. Then the identified set of RCCP is simply all $\mathbf{p}_2 \in [0, 1]^3$ such that (6) holds for the $\mathbf{p}_1$ observed. In lots of empirical contexts, economic theory may shed lights on some properties of the shape of the static payoffs each period even though their specific functional forms are not known. For example, researchers might have reasons to believe the SPP is monotonically increasing or concave in certain coordinates of state variables, or know how a subset of possible states in Ω_X rank between themselves in terms of SPP. Such restrictions can be equivalently represented as linear restrictions on SPP. Then appending these linear restrictions on SPP to the system of linear equalities (which relate primitives to choice patterns in both DGP and counterfactual environment) helps recover an even smaller identified set of RCCP.

Example 3 *(Recovering RCCP with linear restrictions on SPP and knowledge of $F_{\Delta\epsilon|\mathbf{X}}$)* Let $K = 3$ and β, $\{\mathbf{G}_1^j\}_{j=1,0}$ (transition in DGP), $\{\mathbf{G}_2^j\}_{j=1,0}$ (counterfactual transition) and USV distribution $F_{\Delta\epsilon|\mathbf{X}}$ be defined as in Example 2. Let the true SPP be

$$\mathbf{u}_1 = \begin{bmatrix} \frac{8}{3}\log 2 - \frac{5}{3}\log 3 - \frac{4}{3}\log 5 + 10 \\ \frac{32}{5}\log 2 + \frac{2}{5}\log 3 - \frac{8}{5}\log 5 + 10 \\ \frac{16}{15}\log 2 + \frac{4}{15}\log 3 + \frac{4}{5}\log 5 + 10 \end{bmatrix} \; ; \; \mathbf{u}_0 = \begin{bmatrix} \frac{91}{15}\log 2 - \frac{29}{15}\log 3 - \frac{8}{5}\log 5 + 10 \\ \frac{91}{15}\log 2 - \frac{1}{15}\log 3 - \frac{8}{15}\log 5 + 10 \\ 10 \end{bmatrix}$$

The choice pattern observed in the DGP is $\mathbf{p}_1 = [\frac{1}{3}, \frac{3}{5}, \frac{1}{2}]$ while the true counterfactual pattern is $\mathbf{p}_2 = [\frac{1}{5}, \frac{3}{8}, \frac{5}{6}]$ (which is unknowable to econometricians).[13] It is straightforward to verify the testable implication in (6) is satisfied. An econometrician who analyzes this structural DBCP model can observe $\mathbf{p}_1, \{\mathbf{G}_1^j\}_{j=1,0}$ in the DGP, know β,$F_{\Delta\epsilon|\mathbf{X}}$, and are interested in learning $\mathbf{p}_2$ under the $(\mathbf{G}_2^1, \mathbf{G}_2^0)$ of interests. In addition, he also knows some shape restrictions about how the alternatives compare to each other in terms of SPP for given $\mathbf{x}$'s. That is, he knows $\mathbf{u}$ satisfies

$$u_{11} < u_{01} \; ; \; u_{12} < u_{02} \; ; \; u_{13} > u_{03}$$

where $u_{jk} \equiv u_j(\mathbf{x}_k)$. Then finding the identified set of RCCP under $(\mathbf{G}_2^1, \mathbf{G}_2^0)$ simply amounts to finding the set of all $\tilde{\mathbf{p}}$ such that the following linear system is feasible with solutions in $\mathbf{u}$:

$$\mathbf{A}_1^1\mathbf{u}_1 - \mathbf{A}_0^1\mathbf{u}_0 = \mathbf{A}_1^1\mathbf{Q}(\mathbf{p}^1) + (\mathbf{A}_0^1 - \mathbf{A}_1^1)\boldsymbol{\kappa}^0(\mathbf{p}^1) \tag{7}$$

$$\mathbf{A}_1^2\mathbf{u}_1 - \mathbf{A}_0^2\mathbf{u}_0 = \mathbf{A}_1^2\mathbf{Q}(\tilde{\mathbf{p}}) + (\mathbf{A}_0^2 - \mathbf{A}_1^2)\boldsymbol{\kappa}^0(\tilde{\mathbf{p}}) \tag{8}$$

$$\begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \underset{6\times 1}{\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_0 \end{bmatrix}} > \underset{3\times 1}{\mathbf{0}} \tag{9}$$

where $\mathbf{A}_j^m$ is the shorthand for $\mathbf{A}_j(\Gamma_m)$. First note that setting $u_{0K} = \bar{u}$ for any constant $\bar{u}$ is an innocuous location normalization. To see this, just note $\mathbf{A}_1^m\mathbf{u}_1 - \mathbf{A}_0^m\mathbf{u}_0 = \mathbf{A}_1^m(\mathbf{u}_1 + \mathbf{1}'c) - \mathbf{A}_0^m(\mathbf{u}_0 + \mathbf{1}'c)\,\forall c \in \mathbb{R}^1$ and $m = 1, 2$, where $\mathbf{1}'c$ is a constant vector with all coordinates being c.[14]

[13] While choosing the specifications for the example, we actually work backwards by first choosing $\mathbf{p}_1, \mathbf{p}_2$ that satisfy the equality testable implications and then solve for $\mathbf{u}$ using knowledge of Γ and $F_{\epsilon|X}$.

[14] This is because the column vectors in $\mathbf{A}_j^m$ must add up to be equal to $\mathbf{1}'(\frac{1}{1-\beta})$ for $m = 1, 2$ and $j = 0, 1$.

Hence for any $\mathbf{p}_1, \tilde{\mathbf{p}}$, a vector $(\mathbf{u}_1, \mathbf{u}_0)$ solves the linear system if and only if its locational shift $(\mathbf{u}_1 + \mathbf{1}'c, \mathbf{u}_0 + \mathbf{1}'c)$ is also a solution. The rank of the coefficient matrix formed from (7) and (37) on the left-hand side is 5 (the highest possible). Thus for any fixed $\tilde{\mathbf{p}}$ and after normalizing u_{0K} to 0, the remaining coordinates in $\mathbf{u}_1, \mathbf{u}_0$ can be expressed as nonlinear functions of $\tilde{\mathbf{p}}$ using the linear equalities (7), (8). Substituting these expressions into (9) gives us three nonlinear inequalities involving $\tilde{\mathbf{p}}$. We use a grid-search in the space of $[0,1]^3$ (with grid-width equal to $\frac{1}{50}$) to pick out $\tilde{\mathbf{p}}$ such that these nonlinear inequalities hold jointly. The collection of all such $\tilde{\mathbf{p}}$ is our identified set of RCCP given $\mathbf{p}_1$ observed and the knowledge of $F_{\Delta\epsilon|\mathbf{X}}$. Figure 2.1 shows the set of $\tilde{\mathbf{p}}$ that has the identified set of RCCP under the linear restrictions in (9).

By definition of RCCP, the more restrictions we impose on $\mathbf{u}$, the smaller the size of the identified set of RCCP. To illustrate this point, we introduce additional linear restrictions in our exercise. Suppose in addition to (9), econometricians also know additional shape restrictions that the true SPP satisfies

$$u_{13} - u_{03} > u_{01} - u_{11} > u_{02} - u_{12}$$

Then we can simply augment linear inequalities in (9) with

$$\begin{bmatrix} 1 & 0 & 1 & -1 & 0 & -1 \\ -1 & 1 & 0 & 1 & -1 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_0 \end{bmatrix}}_{6\times 1} > \underset{2\times 1}{\mathbf{0}} \tag{10}$$

Applying the grid-search as before yields a smaller identified set of RCCP shown in Figure 2.2. (End of Example 3)

The algorithm in Example 3 in principle can be used to incorporate any form of exogenously given shape restrictions on $\mathbf{u}$ (not necessarily linear) in the search for identified set of RCCP. There is a subtle issue about the practice of setting u_{0K} to an arbitrary constant vector, as this may not be innocuous under certain restrictions that are different from those in the example. (In our example here, it is innocuous because of the form of coefficient matrices in (9) and (10) only.) Formally, this means there might exists restrictions (such as $\mathbf{Cu} > \mathbf{d}$) such that shifting the location of the solution $\mathbf{u}$ may result in a violation of these restrictions. In such cases, we need to refrain from setting u_{0K} to arbitrary values while trying to recover the identified set of RCCP. On the other hand, if a linear restriction $\mathbf{Cu} > \mathbf{0}$ is such that any location shifts of a solution $\mathbf{u}$ also yields a solution (e.g. the sum of column vectors is equal to a zero vector as it is in Example 3), then any location shifts of the solution $\mathbf{u}$ must be innocuous, and setting u_{0K} to arbitrary constant c is a mere locational normalization that does not change the identified set of RCCP recovered. It is easy to see that exogenous restrictions such as ranking of a subset of the states, or monotonicity or concavity in certain coordinates all satisfy such a requirement for innocuous location normalization.

Finally, note the identified set of RCCP recovered is interesting in its own right, regardless of its actual sizes. This is because such a set reveals the limit of robust structural analyses while remaining agnostic about SPP. Our method also introduces an efficient framework to exhaust the identifying power of a priori, nonparametric restrictions. The next section generalizes this idea of recovering identified sets of RCCP in a more realistic setup where econometricians remain agnostic about the USV distribution.

5 Bounding RCCP with Unknown USV Distributions

So far we have maintained that the distribution of USV is known. We have argued that if the actual USV in DGP follows a location-scale family of distributions (such as normal), then choosing a specific location and scale normalization in structural estimation is innocuous for predicting counterfactuals under different state transitions or SPP. However, in practice, misspecifying USV distribution to have an incorrect parametric form can lead to errors in counterfactual predictions. We propose a novel solution to partially identify rationalizable counterfactual choice patterns (RCCP) while refraining from introducing parametric restrictions on SPP or USV distribution.

Our methodology is based on an observation that decision maker's dynamic rationality, the independence of USV from X, as well as some other exogenous shape restrictions on SPP can be all equivalently expressed as linear restrictions on SPP and finite-dimension parameters in the USV distribution without loss of information for counterfactual analyses. The choice probabilities (both in the DGP and the counterfactuals) enter the linear systems through coefficients. Hence deriving the identified sets of (or "sharp bounds" on) RCCP amounts to finding all counterfactual $\tilde{\mathbf{p}}$ which, combined with $\mathbf{p}$ observed in DGP, would make the linear system feasible with solutions in structural elements (i.e. $\mathbf{u}$ and the nuisance parameters in the USV distribution). Then standard linear programming algorithms can be applied to characterize this identified set of RCCP. We focus on the independence restriction on USV throughout this section.

SI (Statistical independence) $\Delta\epsilon$ is independent from $\mathbf{X}$, and continuously distributed with positive densities on $\mathbb{R}^1$ and $Median(\Delta\epsilon) = 0$.

This assumption essentially requires USV to be exogenous noises superimposed on the process of state transitions, and not to interact with past or current OSV. This assumption is invoked by lots of empirical works in structural estimation.

5.1 Counterfactual changes in state transitions

Let $U, \mathcal{F}$ denote the set of generic restrictions on $\mathbf{u}$ and $F_{\Delta\epsilon|\mathbf{X}}$ that are known to econometricians. Let $\mathbf{p}_1 \in \mathbb{R}^K$ (with the k-th coordinate $p_{1,k} \equiv p_1(\mathbf{x}_k)$) denote choice probabilities observed in a DGP summarized by $\{\Gamma_1, U, \mathcal{F}\}$ with $\Gamma_1 = \{\beta, \mathbf{G}_1^1, \mathbf{G}_1^0\}$. Econometricians are interested in inferring agents' choice patterns $\mathbf{p}_2$ in a counterfactual context $\Gamma_2 \equiv \{\beta, \mathbf{G}_2^1, \mathbf{G}_2^0\}$ where the underlying primitives $\mathbf{u}, F_{\Delta\epsilon|\mathbf{X}}$ are unchanged in U and $\mathcal{F}$. Let ϕ be defined as before.

Definition 4 *The identified set of rationalizable counterfactual choice probabilities (RCCP) under $\{\Gamma_2, U, \mathcal{F}\}$ is the set of all $\mathbf{p}_2 \in [0,1]^K$ such that $\exists(\mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}}) \in U \otimes \mathcal{F}$ with $p_{2,k} \in \phi(\mathbf{x}_k; \mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}}, \Gamma_2)$ and $p_{1,k} \in \phi(\mathbf{x}_k; \mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}}, \Gamma_1)$ for all $\mathbf{x}_k \in \Omega_{\mathbf{X}}$.*

In words, the identified set of RCCP is a collection of all outcomes in the counterfactual context that can be rationalized, jointly with $\mathbf{p}_1$ observed in DGP, by the same structure ($\mathbf{u}$,$F_{\boldsymbol{\Delta\epsilon}|\mathbf{X}}$) satisfying restrictions $U, \mathcal{F}$. In this subsection, let $\mathcal{F}_{SI}$ be the set of $F_{\Delta\epsilon|\mathbf{X}}$ that satisfies *SI*, let $\mathbf{A}_j(\Gamma) \equiv (\mathbf{I} + \mathbf{G}_\infty^j)$. Let U_C denote the set of SPP that satisfy a set of strict linear inequalities $\mathbf{Cu} > \mathbf{0}$ for some known constant matrix $\mathbf{C}$. We start by giving the *sufficient and necessary* conditions for a choice pattern $\mathbf{p}$ to be rationalizable under $U_C, \mathcal{F}_{SI}$ and a generic single environment $\Gamma \equiv (\beta, \mathbf{G}^1, \mathbf{G}^0)$. Let $\bar{\kappa}$ be a positive constant.

Lemma 2 *Suppose AS, SI, DS and REG (i)-(iii) hold. A $\mathbf{p} \in [0,1]^K$ is consistent with $\{\Gamma, U_C, \mathcal{F}_{SI}\}$ if and only if the following linear system has solutions in $\mathbf{Q}$,$\boldsymbol{\kappa}^0$ and $\mathbf{u} \in U_C$:*

$$\mathbf{A}_1(\Gamma)\mathbf{u}_1 - \mathbf{A}_0(\Gamma)\mathbf{u}_0 = \mathbf{A}_1(\Gamma)\mathbf{Q} + [\mathbf{A}_0(\Gamma) - \mathbf{A}_1(\Gamma)]\boldsymbol{\kappa}^0 \ ; \ \mathbf{Cu} > 0 \tag{11}$$

$$Q_l \leq Q_k \Leftrightarrow p_l \leq p_k \text{ and } Q_k \geq 0 \Leftrightarrow p_k \geq \tfrac{1}{2} \ \forall \ l, k \in \{1, ., K\} \tag{12}$$

$$p_{(k-1)}(Q_{(k)} - Q_{(k-1)}) \leq \kappa^0_{(k)} - \kappa^0_{(k-1)} \leq p_{(k)}(Q_{(k)} - Q_{(k-1)}) \ for \ k \geq 2 \tag{13}$$

$$\kappa_k^0 > 0, \ \tfrac{1}{2}Q_k \leq \kappa_k^0 - \bar{\kappa} \leq p_k Q_k, \ \kappa_k^0 = \bar{\kappa} \Leftrightarrow p_k = \tfrac{1}{2} \ \forall k \in \{1, ., K\} \tag{14}$$

where $p_{(k)}$, $Q_{(k)}$, $\kappa^0_{(k)}$ are the k-th smallest element in the K-vectors $\mathbf{p}$, $\mathbf{Q}$ and $\boldsymbol{\kappa}^0$ respectively.

It is important to note that these linear inequalities in the lemma are *not only necessary but also sufficient* for a choice pattern $\mathbf{p}$ to be rationalizable by a distribution of USV that is only restricted to be independent of $\mathbf{X}$. The necessity follows from the characterization of rationalizability in Lemma 1 and that when the USV is independent of X, observable states only affect the distributional parameters in this characterization (i.e. $\mathbf{Q}$ and $\boldsymbol{\kappa}^0$) through $\mathbf{p}$. The intuition for sufficiency is that, though $F_{\Delta\epsilon}$ is an infinite-dimensional parameter under SI, it only affects individuals' decisions through a pair of appropriately chosen, finite-dimensional (K-) vector of nuisance parameters (i.e. quantiles and truncated surplus functions). The existence of solutions in $\mathbf{u}$, $\mathbf{Q}$, $\boldsymbol{\kappa}^0$ in the linear system can be used to construct a rationalizing USV distribution through interpolation.

The result in Lemma 2 is only for a single decision environment, but can be extended easily to recover the identified set of RCCP, which is an exercise involving two decision environments. By definition of counterfactual analyses, the structure $\mathbf{u}$ and $F_{\Delta\epsilon}$ are fixed both in the DGP and the counterfactual context. Thus we can stack the linear systems from the DGP and the counterfactual context together and recover all counterfactual choice patterns that can be rationalized, jointly with $\mathbf{p}_1$ observed, in the augmented system with solutions in $\mathbf{u}, \mathbf{Q}, \boldsymbol{\kappa}^0$. The proposition below formalizes this idea. Let $\mathbf{A}(\Gamma_m) \equiv [\mathbf{A}_1(\Gamma_m), -\mathbf{A}_0(\Gamma_m)]$, $\mathbf{B}(\Gamma_m) \equiv [\mathbf{A}_1(\Gamma_m),\ \mathbf{A}_0(\Gamma_m) - \mathbf{A}_1(\Gamma_m)]$, $Q_{m,l}$, $p_{m,l}$ are the l-th coordinate of $\mathbf{Q}_m, \mathbf{p}_m$ for $m = 1, 2$, and $p_{(k)}$, $Q_{(k)}$, $\kappa_{(k)}$ are the k-th smallest elements in the $2K$ vectors $[\mathbf{p}_1, \mathbf{p}_2]$, $[\mathbf{Q}_1, \mathbf{Q}_2]$, $[\boldsymbol{\kappa}_1^0, \boldsymbol{\kappa}_2^0]$ respectively. Let $\bar{\kappa}$ be an arbitrary constant.

Proposition 3 *Suppose AS, CI, DS, REG (i)-(iv) hold. Let $\mathbf{p}_1$ be choice probabilities observed in Γ_1 under restrictions $U_C, \mathcal{F}_{SI}$. The identified set of rationalizable counterfactual choice probabilities under Γ_2 is the set of $\mathbf{p}_2$ such that the following system has solutions in $\mathbf{u}_1, \mathbf{u}_0$ and $\{\mathbf{Q}_m, \boldsymbol{\kappa}_m^0\}_{m=1}^2$ (where $\mathbf{u}_1, \mathbf{u}_0, \mathbf{Q}_m, \boldsymbol{\kappa}_m^0$ are all K-vectors):*

$$\underbrace{\begin{bmatrix} \mathbf{A}(\Gamma_1) \\ \mathbf{A}(\Gamma_2) \end{bmatrix}}_{2K\text{-by-}2K} \underbrace{\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_0 \end{bmatrix}}_{2K\text{-by-}1} = \underbrace{\begin{bmatrix} \mathbf{B}(\Gamma_1) & \mathbf{0} \\ \mathbf{0} & \mathbf{B}(\Gamma_2) \end{bmatrix}}_{2K\text{-by-}4K} \underbrace{\begin{bmatrix} \mathbf{Q}_1' & \boldsymbol{\kappa}_1^{0\prime} & \mathbf{Q}_2' & \boldsymbol{\kappa}_2^{0\prime} \end{bmatrix}'}_{4K\text{-by-}1} ;\ \mathbf{Cu} > \mathbf{0} \tag{15}$$

$$Q_{m,l} \le Q_{n,k} \Leftrightarrow p_{m,l} \le p_{n,k} \text{ and } Q_{m,k} \ge 0 \Leftrightarrow p_{m,k} \ge 1/2\ \forall l,k \in \{1,.,K\}, m,n \in \{1,2\} \tag{16}$$

$$p_{(k-1)}(Q_{(k)} - Q_{(k-1)}) \le \kappa_{(k)}^0 - \kappa_{(k-1)}^0 \le p_{(k)}(Q_{(k)} - Q_{(k-1)}) \text{ for } 2 \le k \le 2K, \tag{17}$$

$$\tfrac{1}{2}Q_{m,k} \le \kappa_{m,k}^0 - \bar{\kappa} \le p_{m,k}Q_{m,k} \text{ and } \kappa_{m,k}^0 = \bar{\kappa} \Leftrightarrow p_{m,k} = \tfrac{1}{2} \text{ and } \kappa_k^0 > 0 \text{ for } 2 \le k \le 2K \tag{18}$$

Note there are $6K$ unknowns including $(\mathbf{u}_1, \mathbf{u}_0, \mathbf{Q}_1, \boldsymbol{\kappa}_1^0, \mathbf{Q}_2, \boldsymbol{\kappa}_2^0)$ in the linear system. In addition to the shape restrictions $\mathbf{Cu} > \mathbf{0}$, there are $2K$ equalities in (15) as well as $2K$ effective inequality constraints in (16) (which arises from the ordering of $\mathbf{Q}_1, \mathbf{Q}_2$ and $\mathbf{0}$), and $4K+1$ effective inequalities in (17) and (18) together (which arise from the upper and lower bounds on differences between adjacent elements from the ordering of $\boldsymbol{\kappa}_1^0, \boldsymbol{\kappa}_2^0, \bar{\kappa}$ and the nonnegativity constraint $\kappa_{(1)}^0 > 0$). There might exist a set of $\mathbf{p}_2$ that can make the linear system infeasible, depending on the $\mathbf{p}_1$ observed in the DGP as well as environments Γ_1, Γ_2 and shape restrictions $\mathbf{C}$. The identified scope $\mathbf{p}_2$ can be recovered through any linear programming algorithm that checks feasibility of systems of linear inequalities. (We will give more details about the implementation of the algorithm in the examples below.)

Some remarks: (1) The proposition can be easily extended to accommodate any generic linear restrictions on $\mathbf{u}$ under SI. Such restrictions may include linearity, monotonicity or concavity of $\mathbf{u}$ in $\mathbf{x}$. (2) Note the maximum possible rank for $[\mathbf{A}(\Gamma_1)', \mathbf{A}(\Gamma_2)']'$ is $2K-1$ by construction. Some locational normalization on $\mathbf{u}$ (such as $\mathbf{u}_{0K} = 0$) can be applied innocuously to simplify the algorithm in the search of identified scope of RCCP as long as $\mathbf{Cu} > \mathbf{0}$ implies $\mathbf{C}(\mathbf{u} + \mathbf{1}'c) > \mathbf{0}$ (which is the case in examples below). (3) The proposition can also be used to test whether a profile of observed choice probabilities $\{\mathbf{p}_1, \mathbf{p}_2\}$ are rationalized in a multiple-environment DGP given by (Γ_1, Γ_2) under U_C and $\mathcal{F}_{SI}$.

Example 4.1 *(Identified set of RCCP when the true USV distribution is unknown and Extreme Type I)* Let's consider exactly the same specification as in Example 3, except that now econometricians do not know USV are i.i.d. extreme type I. Instead, they only know (a) the difference $\Delta\epsilon$ is independent of X and has zero median; and (b) the true SPP satisfies the shape restrictions in (9) and (10). We choose to stick to the previous specifications, because it is easier to compare the results below with earlier results when $F_{\Delta\epsilon|\mathbf{X}}$ is known. As in Example 3, the actual choice probabilities observed under the DGP Γ_1 is $\mathbf{p}_1 = [\frac{1}{3}, \frac{3}{5}, \frac{1}{2}]$. If the truth were to be known, the true counterfactual choice outcome under Γ_2 would be $\mathbf{p}_2 = [\frac{1}{5}, \frac{3}{8}, \frac{5}{6}]$. Econometricians are interested in finding out all counterfactual outcomes $\mathbf{p}_2$ under Γ_2 that could be consistent with the model restrictions, while remaining agnostic about the parametric form of the USV distribution. We use a grid-search in the space of $[0,1]^3$ (with grid-width being $\frac{1}{20}$) to pick out $\tilde{\mathbf{p}}$ that makes the augmented linear systems from DGP Γ_1 and counterfactual environment Γ_2 in Proposition 3 jointly feasible (given $\mathbf{p}_1$ observed). By Proposition 3, the complete set of all such $\tilde{\mathbf{p}}$ forms the identified set of RCCP. (We include in the appendix the details in implementing the algorithm pointwise for a generic element in $[0,1]^3$.) The result is shown in Figure 3.1. (End of Example 4.1)

The identified set of RCCP recovered under (9), (10) and the knowledge of the i.i.d. extreme type one USV in Example 3 is a subset of that recovered in Example 4.1 above. The difference in sizes of the two sets is a graphic illustration of the additional information about counterfactuals derived from knowing exactly the distribution of USV. To illustrate the generality of our algorithm, we carry out a similar exercise in Example 4.2 below. Compared with Example 4.1, we experiment with a different transition matrix between observable states so that linear equalities due to dynamic rationality in (15) takes a more complicated form.

Example 4.2 Let $\beta = 0.8$ but consider a different specification of the transition of observable states

$$\mathbf{G}_1^1 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \mathbf{G}_1^0 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \mathbf{G}_2^1 = \begin{bmatrix} \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{bmatrix}, \mathbf{G}_2^0 = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

and let USV be distributed as extreme type I. Using results in Proposition 2, we can show the counterfactual outcome must necessarily satisfy

$$Q_{11} + \tfrac{1}{10}Q_{12} + \tfrac{1}{2}Q_{13} + \tfrac{2}{5}\kappa_{12}^0 - \tfrac{2}{3}\kappa_{13}^0 = Q_{21} + \tfrac{1}{10}Q_{22} + \tfrac{1}{2}Q_{23} + \tfrac{2}{5}\kappa_{22}^0 - \tfrac{2}{3}\kappa_{23}^0 \tag{19}$$

where $Q_{mk} = \ln p_{mk} - \ln(1-p_{mk})$, $\kappa_{mk}^0 \equiv -\ln(1-p_{mk})$ for $m = 1,2$ and $k = 1,2,3$. Furthermore suppose the choice patterns observed in DGP ($\mathbf{p}_1$) and the true counterfactual ($\mathbf{p}_2$) are :

$$\mathbf{p}_1 = (\tfrac{1}{3}, \tfrac{3}{5}, \tfrac{1}{2})\ ;\ \mathbf{p}_2 = \left(\tfrac{3^{1/5}2^{2/5}5^{2/5}}{3^{1/5}2^{2/5}5^{2/5}+6}, \tfrac{1}{3}, \tfrac{3}{4}\right)$$

which satisfy (19).[15] Then backward calculations using the first restriction in (15) shows the true

[15] While setting up the example, we actually choose $\mathbf{p}_1, p_{22}, p_{23}$ first and use (19) to solve for p_{21}.

SPP satisfies the following restrictions:

$$\mathbf{C} \equiv \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 & -1 & -1 \\ -1 & 0 & -1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_0 \end{bmatrix} > 0$$

Suppose econometricians do not know the true $\mathbf{u}$ but know it satisfies these linear restrictions. Then a new linear system similar to (15)-(18) can be constructed with $\mathbf{A}(\Gamma_m)$ and $\mathbf{C}$ defined accordingly in (15), and the identified set of RCCP is the set of all $\mathbf{p}_2$ that makes the system feasible. We implement the same algorithm for checking the feasibility of such linear systems, and Figure 3.2 gives a scatterplot of the recovered set of RCCP. (End of Example 4.2)

5.2 Perturbation in single-period payoffs

In practice, policy makers often implement changes with known effects on individuals' payoffs per period while holding the transition between states fixed. For instance, when payoffs for individuals are measured in monetary terms, policy makers can modify static payoffs $\mathbf{u}$ by introducing lump-sum subsidies (which leads to additive perturbations), or by levying taxes proportional to individual payoffs (which leads to multiplicative perturbations). We argue that a slight modification of the algorithm in the previous section can help recover the identified set of RCCP under this class of policy changes. Let $U, \mathcal{F}$ denote generic restrictions on $\mathbf{u}$ and $F_{\Delta\epsilon|\mathbf{X}}$.

Definition 5 *Suppose $\mathbf{p}_1$ is observed in the model $\{\Gamma, U, \mathcal{F}\}$. The identified set of rationalizable counterfactual choice probabilities under payoff perturbations (denoted $\rho(\mathbf{u})$) is the set of $\mathbf{p}_2 \in [0,1]^K$ such that $\exists(\mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}}) \in U \otimes \mathcal{F}$ with $p_{2,k} \in \phi(\mathbf{x}_k; \rho(\mathbf{u}), F_{\boldsymbol{\epsilon}|\mathbf{X}}, \Gamma)$ and $p_{1,k} \in \phi(\mathbf{x}_k; \mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}}, \Gamma)$ for all $\mathbf{x}_k \in \Omega_{\mathbf{X}}$.*

Suppose a counterfactual policy perturbs SPP. The size of the percentage changes in each possible state is known to econometricians even though the true $\mathbf{u}$ in the DGP is not. These changes are summarized as

$$\tilde{\mathbf{u}}_1 = \mathbf{D}_1 \mathbf{u}_1 \ ; \ \tilde{\mathbf{u}}_0 = \mathbf{D}_0 \mathbf{u}_0$$

where $\mathbf{D}_j$ is a K-by-K diagonal matrix with its (k,k)-th entry being the gross percentage change in $\mathbf{u}$. There may be additional linear restrictions on $\mathbf{u}$ such as in Proposition 3. Proposition 4 below suggests a slight modification of the algorithm in Proposition 3 can be used to recover the identified set of RCCP in this case. The proof is similar to Proposition 3 and omitted for brevity.

Proposition 4 *Suppose AS, CI, DS, REG (i)-(iv) hold. Let $\mathbf{p}_1$ be choice probabilities observed under Γ and restrictions $U_C, \mathcal{F}_{SI}$. The identified scope of counterfactual outcomes under perturbations $\mathbf{D}_1, \mathbf{D}_0$ (with $\mathbf{u}$ and $F_{\Delta\epsilon}$ fixed) is the set of $\mathbf{p}_2$ such that a linear system that consists of*

(16), (17), (18) and the following linear equations have solutions in $\mathbf{u}$ *and* $\{\mathbf{Q}_m, \boldsymbol{\kappa}_m^0\}_{m=1}^2$ *(where* $\mathbf{Q}_m, \boldsymbol{\kappa}_m^0$ *are both* K*-vectors):*

$$\underbrace{\begin{bmatrix} \mathbf{A}_1(\Gamma), -\mathbf{A}_0(\Gamma) \\ \mathbf{A}_1(\Gamma)\mathbf{D}_1, -\mathbf{A}_0(\Gamma)\mathbf{D}_0 \end{bmatrix}}_{2K\text{-by-}2K} \underbrace{\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_0 \end{bmatrix}}_{2K\text{-by-}1} = \underbrace{\begin{bmatrix} \mathbf{B}(\Gamma) & \mathbf{0} \\ \mathbf{0} & \mathbf{B}(\Gamma) \end{bmatrix}}_{2K\text{-by-}4K} \underbrace{\begin{bmatrix} \mathbf{Q}_1' & \boldsymbol{\kappa}_1^{0\prime} & \mathbf{Q}_2' & \boldsymbol{\kappa}_2^{0\prime} \end{bmatrix}'}_{4K\text{-by-}1} \; ; \; \mathbf{Cu} > \mathbf{0} \tag{20}$$

Example 5 *(Multiplicative perturbations of SPP)* Let $\beta = 0.8$ and the transition matrix $\mathbf{G}^1, \mathbf{G}^0$ be defined in the same way as $\mathbf{G}_2^1, \mathbf{G}_2^0$ in Example 2. Let USV be i.i.d. extreme type I. Suppose $\mathbf{p}$ is observed under $\beta, \mathbf{G}^1, \mathbf{G}^0$ in DGP and we are interested in predicting counterfactual choice outcomes $\tilde{\mathbf{p}}$ when $\beta, \mathbf{G}^1, \mathbf{G}^0$ and $F_{\Delta\epsilon}$ are fixed while $\mathbf{u}$ is perturbed to $\tilde{\mathbf{u}}$ in the following way:

$$\tilde{u}_{11} = \tfrac{8}{10}u_{11} \ , \ \tilde{u}_{12} = \tfrac{9}{10}u_{12} \ , \ \tilde{u}_{13} = \tfrac{11}{10}u_{13} \ ; \ \tilde{u}_{01} = \tfrac{11}{10}u_{01} \ , \ \tilde{u}_{02} = u_{02} \ , \ \tilde{u}_{03} = \tfrac{9}{10}u_{03}$$

Suppose the true probabilities in DGP and counterfactual contexts are $\mathbf{p}_1 = [\frac{1}{2}, \frac{3}{4}, \frac{1}{5}]$ and $\mathbf{p}_2 = [\frac{2}{3}, \frac{2}{5}, \frac{1}{6}]$ with the true SPP satisfies:[16]

$$\mathbf{C} \equiv \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 & 1 & -1 \\ -1 & 0 & -1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_0 \end{bmatrix} > 0$$

We then recover the identified set of RCCP under the multiplicative counterfactual changes by checking the linear system of (20), (16), (17) and (18). Figure 4 suggests the identified set of RCCP in this case is much larger relative to Example 4.1 and Example 4.2. This is not surprising, as the rank of the coefficient matrix on the left-hand side has full rank at $2K$. Thus, compared with the other two examples, there is one fewer equality restrictions resulting from dynamic rationality alone. (End of Example 5)

Policy-makers may also introduce lump-sum (rather than percentage) changes on SPP (i.e. $\tilde{\mathbf{u}} = \mathbf{u} + \boldsymbol{\Delta}$) while keeping $\mathbf{G}^1, \mathbf{G}^0$ fixed. Aguirregabiria (2005) suggests if both the changes in SPP ($\boldsymbol{\Delta}$) and the USV distribution are known, then the identified set of RCCP would be recovered directly as the (not necessarily unique) solutions to a system of nonlinear equations

$$\mathbf{B}(\Gamma) \begin{bmatrix} \mathbf{Q}(\tilde{\mathbf{p}}) \\ \boldsymbol{\kappa}^0(\tilde{\mathbf{p}}) \end{bmatrix} = \mathbf{B}(\Gamma) \begin{bmatrix} \mathbf{Q}(\mathbf{p}) \\ \boldsymbol{\kappa}^0(\mathbf{p}) \end{bmatrix} + \mathbf{A}(\Gamma) \begin{bmatrix} \boldsymbol{\Delta}_1 \\ \boldsymbol{\Delta}_0 \end{bmatrix}$$

where $\Gamma = \{\beta, \mathbf{G}^1, \mathbf{G}^0\}$, $\mathbf{A}(\Gamma) = [\mathbf{A}_1(\Gamma), -\mathbf{A}_0(\Gamma)]$, $\mathbf{B}(\Gamma) = [\mathbf{A}_1(\Gamma), \mathbf{A}_0(\Gamma) - \mathbf{A}_1(\Gamma)]$, $\mathbf{p}$ is the choice pattern observed in DGP, and $\tilde{\mathbf{p}}$ is the RCCP to be solved for. We argue that when USV

[16] While specifying this example, we actually choose $\mathbf{p}_1$ and $\mathbf{p}_2$ first and then calculate $\mathbf{u}$ backwards using (20) and knowledge of $F_{\Delta\epsilon}$.

distribution is not known, a slight modification of the algorithm above can be applied to recover the sharp bounds on RCCP. Moreover, an attractive feature of our approach is that it provides a convenient way for incorporating exogenous economic restrictions on SPP that take the form $\mathbf{Cu} > \mathbf{0}$ while recovering the identified set of RCCP. Specifically, we can just replace restrictions in (20) with

$$\begin{bmatrix} \mathbf{A}_1(\Gamma)\mathbf{u}_1 - \mathbf{A}_0(\Gamma)\mathbf{u}_0 \\ \mathbf{A}_1(\Gamma)(\mathbf{u}_1 + \Delta_1) - \mathbf{A}_0(\Gamma)(\mathbf{u}_0 + \Delta_0) \end{bmatrix} = \begin{bmatrix} \mathbf{B}(\Gamma) & \mathbf{0} \\ \mathbf{0} & \mathbf{B}(\Gamma) \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1' & \boldsymbol{\kappa}_1^{0\prime} & \mathbf{Q}_2' & \boldsymbol{\kappa}_2^{0\prime} \end{bmatrix}' \; ; \; \mathbf{Cu} > \mathbf{0} \tag{21}$$

and collect all $\mathbf{p}_2$ such that (21), (16), (17) and (18) are feasible with solutions in $\mathbf{u}$ and $\{\mathbf{Q}_m, \boldsymbol{\kappa}_m^0\}_{m=1}^2$.

Example 6 *(Additive perturbations of SPP)* Let $\beta, \mathbf{G}^0, \mathbf{G}^1$ and $F_{\Delta\epsilon}$ be specified as in Example 5 where $K = 3$. Let the SPP and its counterfactual changes be

$$\mathbf{u}_1 = \begin{bmatrix} 10 + \log\left(2^{-\frac{23}{31}} 5^{\frac{6}{31}}\right) \\ 10 + \log\left(2^{\frac{14}{31}} 3^1 6^{5-\frac{5}{31}}\right) \\ 10 + \log\left(2^{-\frac{53}{31}} 5^{-\frac{1}{31}}\right) \end{bmatrix} \; ; \; \boldsymbol{\Delta}_1 = \begin{bmatrix} \frac{26}{31} + \log\left(2^{\frac{54}{31}} 3^{\frac{4}{31}} 5^{-\frac{17}{32}} 6^{\frac{6}{31}}\right) \\ \frac{61}{31} + \log\left(2^{\frac{17}{31}} 3^{-\frac{55}{31}} 5^{\frac{9}{31}} 6^{-\frac{5}{31}}\right) \\ \frac{37}{31} + \log\left(2^{\frac{53}{31}} 5^{-\frac{23}{31}} 6^{-\frac{1}{31}} 3^{-\frac{11}{31}}\right) \end{bmatrix}$$

and $\mathbf{u}_0 = [10; 10; 10]$, $\Delta_0 = [1; 2; 1]$.[17] Under these specifications, the choice outcomes in the DGP and under counterfactual changes are respectively $\mathbf{p}_1 = [\frac{1}{2}, \frac{3}{4}, \frac{1}{5}]$, $\mathbf{p}_2 = [\frac{2}{3}, \frac{2}{5}, \frac{1}{6}]$. Suppose an econometrician observes $\mathbf{p}_1$, knows β, G^0, G^1 and knows that $\mathbf{u}_0$ is independent of states. He does not know the form of $\mathbf{u}_1$ or the utility level in $\mathbf{u}_0$, or the parametric form of the USV distribution.[18] The econometrician is interested in predicting $\mathbf{p}_2$ under counterfactual changes Δ_1, Δ_0 while only knowing $\Delta\epsilon$ is independent of X and choose to normalize $\mathbf{u}_0$ to the zero vector in structural estimations.

We apply the algorithm under two scenarios. First, the econometrician are not aware of any additional shape restrictions on $\mathbf{u}_1, \mathbf{u}_0$. In this case, after imposing the innocuous normalization $\mathbf{u}_0 = 0$, recovering the identified set of RCCP is equivalent to checking the existence of solutions in $\{\mathbf{Q}, \boldsymbol{\kappa}^0\}_{m=1,2}$ in the linear system

$$\mathbf{B}(\Gamma) \begin{bmatrix} \mathbf{Q}_2' & \boldsymbol{\kappa}_2^{0\prime} \end{bmatrix}' - \mathbf{B}(\Gamma)] \begin{bmatrix} \mathbf{Q}_1' & \boldsymbol{\kappa}_1^{0\prime} \end{bmatrix}' = \mathbf{A}_1(\Gamma)\Delta_1 - \mathbf{A}_0(\Gamma)\Delta_0 \tag{22}$$

joint with (16), (17) and (18). (We do not need to check the first K equalities in (21) characterizing the DGP because there are no restrictions on $\mathbf{u}$ and $rank(\mathbf{A}_1(\Gamma)) = 5$ implies the equalities always have solutions in $\mathbf{u}$ regardless of the right-hand side.) The identified set of RCCP is plotted in Figure 5.1, which covers the actual RCCP as a subset.

[17] While designing the example, we first specify $\mathbf{u}_0$ and Δ_0 as well as $\mathbf{p}_1$ and $\mathbf{p}_2$ that we want the model to generate. Then we solve for $\mathbf{u}_1$ and Δ_1 using the closed form of $\mathbf{Q}$ and $\boldsymbol{\kappa}^0$ under the Extreme Type I specification of the USV distribution.

[18] Note this is different from all previous examples and makes the normalization $\mathbf{u}_0 = 0$ innocuous.

Now consider a second scenario where the econometrician still does not know the form of USV distribution or SPP, but know the true SPP in DGP satisfies the following shape restrictions:

$$u_{11} < u_{10}, u_{12} > u_{02}, u_{13} < u_{03} \tag{23}$$

$$u_{03} - u_{13} > u_{12} - u_{02} > u_{01} - u_{11} \tag{24}$$

Then the econometrician can incorporate these additional information into the linear system of inequalities by appending the following linear inequality to the system (22), (16), (17) and (18):

$$\mathbf{C}_1 \mathbf{A}_1(\Gamma)^{-1} \mathbf{B}(\Gamma) \left[\begin{array}{cc} \mathbf{Q}_1' & \boldsymbol{\kappa}_1^{0\prime} \end{array} \right]' > 0 \tag{25}$$

where $\mathbf{C}_1$ is the first $K = 3$ columns of the 5-by-6 matrix of coefficients $\mathbf{C}$ derived from exogenous shape restrictions in (23) and (24). The identified set of RCCP in this case is plotted in Figure 5.2. It is slightly smaller than that recovered without the shape restrictions in Figure 5.1. (End of Example 6)

6 Conclusions

We have proposed new approaches to address two main challenges in nonparametric structural DBCP models in this paper. Within the benchmark framework where unobservable state distribution is assumed to be known, we first show exogenous variations in state transitions can help derive testable implications of the dynamic binary choice process and identify single-period payoffs even when both actions yield static payoffs that are not independent of observable states. Then we argue nonparametric shape restrictions on SPP that derive exogenously from economic restrictions can be exploited to find sharp bounds on the rationalizable counterfactual choice probabilities that are consistent with the model restrictions. More interestingly, we generalize the approach of partial identification of RCCP to the more challenging case where USV distribution is not known. We propose simple algorithms to recover the sharp bounds on RCCP, and use numerical examples and simulations to show the algorithm is feasible and the resultant bounds on RCCP can be very informative.

There are several interesting directions for future research. First, search for analytical properties on the identified sets of RCCP that may help reduce the computational intensity of the algorithm. Second, extend the algorithm to more general cases with multinomial choices or continuous states. A third direction is to construct test statistics or estimators based on our identification results, and show their asymptotic properties.

7 Appendix

7.1 Proofs of main identification results

Proof of Proposition 1. *Proof of (i) :* Given any environment $\Gamma \equiv \{\beta, \mathbf{G}^0, \mathbf{G}^1, \mathbf{S}\}$ (β is known) with $F_{\Delta\epsilon|\mathbf{X}}$ known, (2) is a generally non-homogenous linear system of K equations in $2K$ unknowns $(\mathbf{u}_1, \mathbf{u}_0)$ for any $\mathbf{p}$ in $[0,1]^K$. Denote the augmented matrix of coefficients of (2) as

$$\tilde{\mathbf{A}}(\mathbf{p}; \Gamma, F_{\Delta\epsilon|\mathbf{X}}) \equiv \left[\begin{array}{cc} \mathbf{A}(\Gamma), & \mathbf{B}(\Gamma) \left[\begin{array}{c} \mathbf{Q}(\mathbf{p}; F_{\Delta\epsilon|\mathbf{X}}) \\ \boldsymbol{\kappa}^0(\mathbf{p}; F_{\Delta\epsilon|\mathbf{X}}) \end{array} \right] \end{array} \right]$$

where $\mathbf{A}(\Gamma) \equiv [\mathbf{A}_1(\Gamma), -\mathbf{A}_0(\Gamma)]$ and $\mathbf{B}(\Gamma) \equiv [\mathbf{A}_1(\Gamma),\ \mathbf{A}_0(\Gamma) - \mathbf{A}_1(\Gamma)]$. The system has solutions in $\mathbf{u}$ if and only if $Rank(\mathbf{A}) = Rank(\tilde{\mathbf{A}})$. Suppose $\mathbf{A}$ has full rank K, then $Rank(\mathbf{A}) = Rank(\tilde{\mathbf{A}}) = K$ regardless of the last column in $\tilde{\mathbf{A}}$ where $\mathbf{p}$ enters. Suppose $Rank(\mathbf{A}) < K$. Then by construction of $\mathbf{A}, \mathbf{B}$, any set of basic matrix operations that reduce $\mathbf{A}$ to its row echelon form also reduce $\mathbf{B}$ to its own row echelon form, with rows of zeros in $\mathbf{B}$ exactly matching the rows of zeros in $\mathbf{A}$. Hence $Rank(\mathbf{A})$ $=$$Rank(\tilde{\mathbf{A}})$ also holds even if $Rank(\mathbf{A}) < K$. *Proof of (ii) :* follows immediately from the fact that the number of unknowns in $\mathbf{u}$ is greater than the number of equations in the linear system.

Proof of Proposition 2. Part (i) follows immediately from the fact that the augmented system with MK linear equations has solutions in $2K$ unknown parameters in $\mathbf{u}$ if and only if the rank of $\boldsymbol{\Sigma}$ is equal to the rank of $\tilde{\boldsymbol{\Sigma}}(\{\mathbf{p}_m\}_{m=1}^M)$. Part (ii) follows immediately from the fact that the reduced row echelon form of every non-zero matrix is unique.

Proof of Lemma 2. *(Necessity)* Suppose $\mathbf{p}$ is consistent with $\{\Gamma, U, \mathcal{F}_{SI}\}$. By the definition of rationalizability and *Lemma 1*, there exists $F_{\boldsymbol{\epsilon}|\mathbf{X}} \in \mathcal{F}_{SI}$ such that (11) is satisfied with some $\mathbf{u}$ in U_C, as well as $F_{\Delta\epsilon}$ in $\mathcal{F}_{SI}$ such that

$$\mathbf{Q}\text{=}\ \mathbf{Q}(\mathbf{p}; F_{\boldsymbol{\epsilon}|\mathbf{X}}) \equiv [F^{-1}_{\Delta\epsilon|\mathbf{x}_1}(p_1), ., F^{-1}_{\Delta\epsilon|\mathbf{x}_K}(p_K)]$$

and $\boldsymbol{\kappa}^0$= $\boldsymbol{\kappa}^0(\mathbf{p}; F_{\boldsymbol{\epsilon}|\mathbf{X}})$, where the k-th coordinate is defined as

$$\kappa^0_k \equiv \kappa^0(p(\mathbf{x}_k)) \equiv \int_{-\infty}^{F^{-1}_{\Delta\epsilon|\mathbf{x}_k}(p_k)} F^{-1}_{\Delta\epsilon|\mathbf{x}_k}(p_k) - s dF_{\Delta\epsilon|\mathbf{x}_k}(s)$$

Then independence of ϵ of $\mathbf{X}$ implies $F^{-1}_{\Delta\epsilon}(p_l) \geq F^{-1}_{\Delta\epsilon}(p_k)$ if and only if $p_l \geq p_k$. That $Med(\Delta\epsilon) = 0$ implies $F^{-1}_{\Delta\epsilon}(p_k) \geq 0$ if and only if $p_k \geq 1/2$. That $\kappa^0_k > 0$ $\forall k$ follows by definition of truncated surplus functions. Note for any pair $\mathbf{x}$, $\mathbf{x}' \in \Omega_{\mathbf{X}}$, κ_0 can be written as

$$\kappa_0(p(\mathbf{x}')) = \kappa_0(p(\mathbf{x})) + \int_{q(\mathbf{x})}^{q(\mathbf{x}')} F_{\Delta\epsilon}(s) ds$$

where $q(\mathbf{x}) \equiv F_{\Delta\epsilon}^{-1}(p(\mathbf{x}))$. Thus it is clear from the equation above that the difference between any pair $\kappa^0_{(k-1)}, \kappa^0_{(k)}$ (the $k-1$ and k-th smallest among the K coordinates in $\boldsymbol{\kappa}^0$) has to be bounded between $p_{(k-1)}[F_{\Delta\epsilon}^{-1}(p_{(k)}) - F_{\Delta\epsilon}^{-1}(p_{(k-1)})]$ and $p_{(k)}[F_{\Delta\epsilon}^{-1}(p_{(k)}) - F_{\Delta\epsilon}^{-1}(p_{(k-1)})]$. Also $\frac{1}{2}Q_{(k)} \leq \kappa^0_{(k)} - \kappa^* \leq p_{(k)}Q_{(k)}$ must hold for all k, where κ^* is the true truncated surplus function evaluated at the median $\kappa^0(\frac{1}{2}; F_{\Delta\epsilon})$. But then we can simply change the scale of $\mathbf{u}, \mathbf{Q}, \boldsymbol{\kappa}^0$ defined above by multiplying all elements in $\mathbf{u}, \mathbf{Q}, \boldsymbol{\kappa}^0$ with the constant $\bar{\kappa}/\kappa^*$. This gives solutions that satisfy the linear system (11), (12), (13) and (14). *(Sufficiency)* We need to show if a $\hat{\mathbf{p}}$ makes the system (11), (12), (13) and (14) feasible with some solutions $\hat{\mathbf{u}}$ and $(\hat{\mathbf{Q}}, \hat{\boldsymbol{\kappa}}^0)$, then we can construct a USV distribution $\hat{F}_{\Delta\epsilon}$ independent of states X such that (a) its quantiles and truncated surplus functions are equal to $\hat{\mathbf{Q}}, \hat{\boldsymbol{\kappa}}^0$ and (b) $\hat{\mathbf{p}}$ is the dynamic rational choice probabilities given $\hat{\mathbf{u}}, \hat{F}_{\epsilon|X}$. Such a distribution is conveniently constructed as follows. First, define the $\hat{p}_k$-th quantile as $\hat{F}_{\Delta\epsilon}^{-1}(\hat{p}_k) = \hat{Q}_k$. This will fix K quantiles of the distribution $\hat{F}_{\Delta\epsilon}$. Then interpolating the distribution $\hat{F}_{\Delta\epsilon}$ between these K quantiles so that the truncated surplus functions $\kappa^0(\hat{p}_k; \hat{F}_{\Delta\epsilon}) = \hat{\kappa}^0_k$. This is always possible precisely because the vector $\hat{\boldsymbol{\kappa}}^0$, as part of the solution of the linear system in the lemma satisfies (13), (14) by construction. Therefore, the $\hat{F}_{\Delta\epsilon}$ constructed in this way belongs to $\mathcal{F}_{SI}$ and have conformable quantiles and truncated surplus functions. Also, when combined with $\hat{\mathbf{u}} \in U_C$, $\hat{F}_{\Delta\epsilon}$ generates the choice pattern $\hat{\mathbf{p}}$ observed because (11) is satisfied by construction.

Proof of Proposition 3. (Necessity) Suppose $\mathbf{p}_2$ is such that $\exists \mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}}$ in $U_C \otimes \mathcal{F}_{SI}$ that satisfies:

$$
\begin{aligned}
p_{2,k} &\in \phi(\mathbf{x}_k; \mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}}, \Gamma_2) \\
p_{1,k} &\in \phi(\mathbf{x}_k; \mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}}, \Gamma_1)
\end{aligned}
$$

for all $x_k \in \Omega_{\mathbf{X}}$. First note (15) is necessary for $\mathbf{p}_1, \mathbf{p}_2$ to be observed under Γ_1, Γ_2. Furthermore, by independence of (ϵ, X) and similar arguments for the necessity in Lemma 2, there must be $(\mathbf{Q}_1, \boldsymbol{\kappa}^0_1, \mathbf{Q}_2, \boldsymbol{\kappa}^0_2)$ that satisfies (16), (17) and (18) (Sufficiency) Suppose $\mathbf{p}_2$ is such that the system (15), (16), (17) and (18) is satisfied with some $\mathbf{u} \in U_C$ and $(\mathbf{Q}_1, \boldsymbol{\kappa}^0_1, \mathbf{Q}_2, \boldsymbol{\kappa}^0_2)$. We can construct a distribution of disturbance $F_{\Delta\epsilon}$ whose quantiles and truncated expectations are conformable with $(\mathbf{Q}_1, \boldsymbol{\kappa}^0_1, \mathbf{Q}_2, \boldsymbol{\kappa}^0_2)$ just as in Lemma 2. Then it follows that there always exists a $2K$ vector $\mathbf{u}$ which, along with $F_{\Delta\epsilon}$ constructed above, can rationalizes $(\mathbf{p}_1, \mathbf{p}_2)$ under (Γ_1, Γ_2) respectively.

7.2 Proof of Lemma 1

To prove Lemma 1, we need to first prove the results in *Lemma A1* and *Lemma A2*. We adopt the sup norm on the space of $\mathbb{R}^2$-valued functions $||\mathbf{u}||_\infty \equiv \sup_{j\in\{0,1\},\mathbf{x}\in\Omega_{\mathbf{X}}} |u_j(\mathbf{x})|$.

Lemma A1 *(Rust 1994) Under AS, CI and REG (i)-(iii), the value function of the dynamic binary decision process has a static representation:*

$$
j(\mathbf{s}) = \arg\max_{j\in\{0,1\}} \delta_j(\mathbf{x}; \mathbf{u}, F_{\boldsymbol{\epsilon}|\mathbf{X}}, \Gamma) + \varepsilon_j
$$

where $\boldsymbol{\delta}(\mathbf{x};\mathbf{u},F_{\boldsymbol{\epsilon}|\mathbf{X}},\Gamma)\equiv[\delta_0(\mathbf{x})\ \delta_1(\mathbf{x})]'$ *solves the fixed point equation* $T\circ\boldsymbol{\delta}(\mathbf{x})\equiv[T_1(\mathbf{x};\boldsymbol{\delta})\ T_0(\mathbf{x};\boldsymbol{\delta})]$*, where*

$$T_j(\mathbf{x};\boldsymbol{\delta})\equiv u_j(\mathbf{x})+\beta\int \max\nolimits_{k\in\{0,1\}}\{\delta_k(\mathbf{x}')+\varepsilon_k'\}dF_{\boldsymbol{\epsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x}) \tag{26}$$

To prove *Lemma A1*, we need some preliminary results collected by *Lemma A1.1, Lemma A1.2* and *Lemma A1.3* below. Let $\mathbf{B}^2(\Omega_{\mathbf{X}})$ denote the space of bounded, continuous $\mathbb{R}^2$-valued functions, i.e. $\mathbf{B}^2(\Omega_{\mathbf{X}})=B(\Omega_{\mathbf{X}})\otimes B(\Omega_{\mathbf{X}})$, where $B(\Omega_{\mathbf{X}})$ is the space of bounded, real-valued functions defined on $\mathbf{X}$.Define the norm on $\mathbf{B}^2(\Omega_{\mathbf{X}})$ as $||\mathbf{f}(\mathbf{x})||=\sup_{j\in\{0,1\},\mathbf{x}\in\Omega_{\mathbf{X}}}|f_j(\mathbf{x})|$.

Lemma A1.1 $(B^2(\Omega_{\mathbf{X}}),||.||)$ *is a complete normed vector space.*

Proof. Standard and omitted for brevity.

Lemma A1.2 *Suppose the operator* $T: B^2(\Omega_{\mathbf{X}})\to B^2(\Omega_{\mathbf{X}})$ *satisfies (a)* $\forall f,g\in B^2(\Omega_{\mathbf{X}})$, $f(x)\le g(x)$ *for all* $x\in\Omega_{\mathbf{X}}$ *implies* $(T\circ f)(x)\le(T\circ g)(x)$ *for all* $x\in\Omega_{\mathbf{X}}$ *(where the inequality is component-wise in* $\mathbb{R}^2$*); (b)* $\exists\beta\in(0,1).s.t.T\circ(f(x)+a1_2)\le T\circ f(x)+\beta a$, $\forall f\in B^2(\Omega_{\mathbf{X}})$, $a\ge 0$, $x\in\Omega_{\mathbf{X}}$ *(where* $1_2\equiv[1\ 1]'$*). Then* T *is an contraction mapping with modulus* β.

Proof. We need to show that $\forall\mathbf{f},\mathbf{g}\in\mathbf{B}^2(\Omega_{\mathbf{X}}),||T\circ\mathbf{f}-T\circ\mathbf{g}||\le\beta||\mathbf{f}-\mathbf{g}||$.Note:

$$\begin{aligned}\mathbf{f}&\le \mathbf{g}+||\mathbf{f}-\mathbf{g}||\mathbf{1}_2\\ &\Longrightarrow T\circ\mathbf{f}\le T\circ(\mathbf{g}+||\mathbf{f}-\mathbf{g}||\mathbf{1}_2)\le T\circ\mathbf{g}+\beta||\mathbf{f}-\mathbf{g}||\mathbf{1}_2\\ &\Longrightarrow T\circ\mathbf{f}-T\circ\mathbf{g}\le\beta||\mathbf{f}-\mathbf{g}||\mathbf{1}_2\end{aligned}$$

Likewise by interchanging the role of $\mathbf{f}$ and bg, we have $T\circ\mathbf{g}-T\circ\mathbf{f}\le\beta||\mathbf{f}-\mathbf{g}||\mathbf{1}_2$. Combining the two inequalities proves $||T\circ\mathbf{f}-T\circ\mathbf{g}||\le\beta||\mathbf{f}-\mathbf{g}||$.

Lemma A1.3 *(Contraction Mapping) Define the operator* $T\circ f(x)\equiv[T_1(x;f)\ T_0(x;f)]$*, where*

$$T_j(\mathbf{x})\equiv u_j(\mathbf{x})+\beta\int\max\nolimits_{k\in\{0,1\}}\{f_k(\mathbf{x}')+\varepsilon_k'\}dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})$$

Under REG, T *is a contraction mapping that maps from* $B^2(\Omega_{\mathbf{X}})$ *into* $B^2(\Omega_{\mathbf{X}})$.

Proof. Note $\max_{k\in\{0,1\}}\{f_k(\mathbf{x})\}$ is bounded since $\mathbf{f}\in\mathbf{B}^2(\Omega_{\mathbf{X}})$. Also:

$$\begin{aligned}&\int\max\nolimits_{k\in\{0,1\}}\{f_k(x')+\varepsilon_k'\}dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})\\ \le&\int\max\nolimits_{k\in\{0,1\}}\{f_k(x')\}dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})+\int\max\nolimits_{k\in\{0,1\}}\{\varepsilon_k'\}dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})\end{aligned}$$

Both terms as well as $\mathbf{u}(\mathbf{x})$ are bounded and continuous under *REG*. Hence $T\circ(\mathbf{f}(\mathbf{x}))$ is bounded and continuous. Suppose $\mathbf{f},\mathbf{g}\in\mathbf{B}^2(\Omega_{\mathbf{X}})$, and $\mathbf{f}(\mathbf{x})\le\mathbf{g}(\mathbf{x})$ for all $\mathbf{x}\in\Omega_{\mathbf{X}}$. Then

$$\begin{aligned}T_j(\mathbf{x};\mathbf{f})&= u_j(\mathbf{x})+\beta\int\max\nolimits_{k\in\{0,1\}}\{f_k(\mathbf{x}')+\varepsilon_k'\}dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})\\ &\le u_j(\mathbf{x})+\beta\int\max\nolimits_{k\in\{0,1\}}\{g_k(\mathbf{x}')+\varepsilon_k'\}dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})=T_j(\mathbf{x};\mathbf{g})\end{aligned}$$

And:

$$T_j(\mathbf{x};\mathbf{f}+a\mathbf{1}_2) = u_j(\mathbf{x})+\beta\int \max_{k\in\{0,1\}}\{f_k(\mathbf{x}')+a+\varepsilon_k'\}dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})$$
$$= u_j(\mathbf{x})+\beta\int \max_{k\in\{0,1\}}\{f_k(x')+\varepsilon_k'\}dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})+\beta a$$

By *Lemma A2*, the operator T is a contraction mapping.

Now we are ready to prove *Lemma A1*.

Proof of Lemma A1. By definition, the Bellman Equation is:

$$V(\mathbf{s}) = \max_{j\in\{0,1\}} \upsilon(\mathbf{s},j)+\beta\int V(\mathbf{s}')dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})$$

Under *AS* and *CI*, $V(\mathbf{s}) = \max_{j\in\{0,1\}}\{\delta_j(\mathbf{x})+\varepsilon_j\}$, where

$$\delta_j(\mathbf{x}) \equiv u_j(\mathbf{x})+\beta\int V(\mathbf{x}',\boldsymbol{\varepsilon}')dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})$$

Substitute expression for $V(\mathbf{s})$ into the definition of $\delta_j(\mathbf{x})$ for $j\in\{0,1\}$,

$$\delta_j(\mathbf{x}) = u_j(\mathbf{x})+\beta\int \max_{j\in\{0,1\}}\{\delta_k(\mathbf{x}')+\varepsilon_k'\}dF_{\boldsymbol{\varepsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})$$

It follows from *Lemma A3* that under *REG*, the operator is well-defined for any $\{u,\beta,F_{\boldsymbol{\varepsilon}|\mathbf{X}}\}$, and that a fixed point $\delta(\mathbf{x})$ exists.

As a result of *Lemma A1*, conditional choice probabilities have a static representation:

$$p(\mathbf{x};\mathbf{u},F_{\boldsymbol{\epsilon}|\mathbf{X}},\Gamma)=F_{\Delta\epsilon|\mathbf{X}}[\Delta\delta(\mathbf{x};\mathbf{u},F_{\boldsymbol{\epsilon}|\mathbf{X}},\Gamma)|\mathbf{x}] \tag{27}$$

where $\Delta\epsilon\equiv\epsilon_0-\epsilon_1$ and $\Delta\delta(\mathbf{x})\equiv\delta_1(\mathbf{x})-\delta_0(\mathbf{x})$, where $\delta_j(\mathbf{x}_t)$ is the expected return from choosing j in the current period conditional on $\mathbf{x}_t$. In fact, the conditional independence restriction can be weakened to *A2'*: $H_j(.|\mathbf{s}) = H_j(.|\mathbf{x}), \forall j,\mathbf{s}$ and the representation result is still valid. Now we are ready to prove *Lemma A2*.

Lemma A2 *Suppose a model $\{\Gamma,U,\mathcal{F}\}$ satisfies restrictions AS, CI and REG (i)-(iv). For a given vector of observed choice probabilities $\mathbf{p}^*$, the joint identification region of $(\mathbf{u},F_{\boldsymbol{\epsilon}|\mathbf{X}})$ is*

$$\Theta_I\equiv\{(\mathbf{u},F_{\boldsymbol{\epsilon}|\mathbf{X}})\in U\otimes\mathcal{F}:\Delta\omega(\mathbf{x}_k;\mathbf{u}) = F^{-1}_{\Delta\varepsilon|\mathbf{X}}(p^*(\mathbf{x}_k)|\mathbf{x}_k)-\Delta\xi(\mathbf{x}_k;F_{\Delta\epsilon|\mathbf{X}},p^*)\ \forall\mathbf{x}_k\in\Omega_{\mathbf{X}}\} \tag{28}$$

where $\Delta\omega(\mathbf{x};\mathbf{u})\equiv\omega_1(\mathbf{x})-\omega_0(\mathbf{x})$, $\Delta\xi(\mathbf{x};F_{\Delta\epsilon|\mathbf{X}},p^*)\equiv\xi_1(\mathbf{x})-\xi_0(\mathbf{x})$; $\omega_j(\mathbf{x})$ and $\xi_j(\mathbf{x})$ are unique fixed points of following operators:

$$T_\omega\circ(\omega_j(\mathbf{x})) \equiv u_j(\mathbf{x})+\beta\int\omega_j(\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x}) \tag{29}$$

$$T_\xi\circ(\xi_j(\mathbf{x})) \equiv \beta\int\kappa_j(\mathbf{x}';p^*,F_{\Delta\varepsilon|\mathbf{X}})+\xi_j(\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x}) \tag{30}$$

with κ_d defined as

$$\begin{aligned}\kappa_0(\mathbf{x};p^*,F_{\Delta\varepsilon|\mathbf{X}}) &\equiv \int_{-\infty}^{q(\mathbf{x})}[q(\mathbf{x})-s]dF_{\Delta\epsilon|\mathbf{X}}(s|\mathbf{x})\\ \kappa_1(\mathbf{x};p^*,F_{\Delta\varepsilon|\mathbf{X}}) &\equiv \int_{q(\mathbf{x})}^{+\infty}[s-q(\mathbf{x})]dF_{\Delta\epsilon|\mathbf{X}}(s|\mathbf{x})\end{aligned}$$

where $q(\mathbf{x})\equiv F^{-1}_{\Delta\varepsilon|\mathbf{X}}(p^*(\mathbf{x})|\mathbf{x})$.

Proof of Lemma A2. We need to show that a generic pair $(\mathbf{u},F_{\boldsymbol{\epsilon}|\mathbf{X}})$ can generate the same observed choice probabilities $p(\mathbf{x})$ if and only if it satisfies conditions in the proposition. *(Sufficiency)* Suppose $\mathbf{u}$,$F_{\boldsymbol{\epsilon}|\mathbf{X}}$ satisfies the conditions in the proposition. Then for $j=0,1$,

$$\delta_j(\mathbf{x};u_j,F_{\Delta\epsilon|\mathbf{X}},p)\equiv\omega_j(\mathbf{x};u_j)+\xi_j(\mathbf{x};F_{\Delta\epsilon|\mathbf{X}},p)$$

is the unique fixed point for the following operator:

$$T_j\circ\delta_j(\mathbf{x};u_j,F_{\Delta\epsilon|\mathbf{X}},p)=u_j(\mathbf{x})+\beta\int\delta_j(\mathbf{x}';u_j,F_{\Delta\epsilon|\mathbf{X}},p)+\kappa_j(\mathbf{x}';p,F_{\Delta\epsilon|\mathbf{X}})dG_j(\mathbf{x}'|\mathbf{x})$$

By our supposition in the proposition, for all $\mathbf{x}\in\Omega_{\mathbf{X}}$,

$$\Delta\delta(\mathbf{x};\mathbf{u},F_{\Delta\epsilon|\mathbf{X}},p)=\Delta\omega(\mathbf{x};\mathbf{u})+\Delta\xi(\mathbf{x};F_{\Delta\epsilon|\mathbf{X}},p)=F^{-1}_{\Delta\epsilon|\mathbf{X}}(p(\mathbf{x})|\mathbf{x})$$

Substitution implies

$$\begin{aligned}\kappa_0(\mathbf{x};p,F_{\Delta\epsilon|\mathbf{X}}) &= \int\max\{\Delta\delta(\mathbf{x};\mathbf{u},F_{\Delta\varepsilon|\mathbf{X}},p)-s,0\}dF_{\Delta\epsilon|\mathbf{X}}(s|\mathbf{x})\\ \kappa_1(\mathbf{x};p,F_{\Delta\epsilon|\mathbf{X}}) &= \int\max\{s-\Delta\delta(\mathbf{x};\mathbf{u},F_{\Delta\epsilon|\mathbf{X}},p),0\}dF_{\Delta\epsilon|\mathbf{X}}(s|\mathbf{x})\end{aligned}$$

Since $E(\epsilon_j|\mathbf{x})=0$ for all $\mathbf{x}\in\Omega_{\mathbf{X}}$, we have for $j=1,0$ and $\mathbf{x}\in\Omega_{\mathbf{X}}$

$$\begin{aligned}&\int\delta_j(\mathbf{x}';u_j,F_{\Delta\varepsilon|\mathbf{X}},p)+\kappa_j(\mathbf{x}';p,F_{\Delta\varepsilon|\mathbf{X}})dG_j(\mathbf{x}'|\mathbf{x})\\ =&\int\max_{k\in\{0,1\}}\{\delta_k(\mathbf{x}';u_k,F_{\Delta\epsilon|\mathbf{X}},p)+\varepsilon'_k\}dF_{\boldsymbol{\epsilon}|\mathbf{X}}(\boldsymbol{\epsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x}).\end{aligned}$$

Therefore $\boldsymbol{\delta}(\mathbf{x};\mathbf{u},F_{\Delta\epsilon|\mathbf{X}},p)\equiv[\delta_1(\mathbf{x};u_1,F_{\Delta\varepsilon|\mathbf{X}},p)\ \delta_0(\mathbf{x};u_0,F_{\Delta\varepsilon|\mathbf{X}},p)]'$ is the unique fixed point of the operator $T\circ\boldsymbol{\psi}(\mathbf{x})\equiv[T_1(\mathbf{x};\boldsymbol{\psi})\ T_0(\mathbf{x};\boldsymbol{\psi})]$, where

$$T_j(\mathbf{x};\boldsymbol{\psi})\equiv u_j(\mathbf{x})+\beta\int\max_{k\in\{0,1\}}\{\psi_k(\mathbf{x}')+\varepsilon'_k\}dF_{\boldsymbol{\epsilon}|\mathbf{X}}(\boldsymbol{\varepsilon}'|\mathbf{x}')dG_j(\mathbf{x}'|\mathbf{x})$$

Then the proof of sufficiency is completed by noting that by construction, $\Delta\delta(\mathbf{x})=F^{-1}_{\Delta\epsilon|\mathbf{X}}(p(\mathbf{x})|\mathbf{x})$ for all $\mathbf{x}\in\Omega_{\mathbf{X}}$. *(Necessity)* Now suppose $(\mathbf{u},F_{\boldsymbol{\epsilon}|\mathbf{X}})$ generates $p(\mathbf{x})$. This requires $\Delta\delta(\mathbf{x};\mathbf{u},F_{\boldsymbol{\epsilon}|\mathbf{X}})=F^{-1}_{\Delta\epsilon|\mathbf{X}}(p(\mathbf{x})|\mathbf{x})$ for all $\mathbf{x}\in\Omega_{\mathbf{X}}$, where $\boldsymbol{\delta}(\mathbf{x};\mathbf{u},F_{\boldsymbol{\epsilon}|\mathbf{X}})\equiv[\delta_1(\mathbf{x})\ \delta_0(\mathbf{x})]'$ is the unique fixed point of the operator T. Recursive substitution of $\boldsymbol{\delta}(\mathbf{x})$ into the definition of T suggests $\delta_j(\mathbf{x};\mathbf{u},F_{\boldsymbol{\epsilon}|\mathbf{X}}) = \omega_j(\mathbf{x};u_j)+\xi_j(\mathbf{x};p,F_{\Delta\epsilon|\mathbf{X}})$ for $j=0,1$. (See Aguirregabiria (2008) for more details.) It follows immediately $F^{-1}_{\Delta\epsilon|\mathbf{X}}(p(\mathbf{x})|\mathbf{x}) = \Delta\omega(\mathbf{x};\mathbf{u})+\Delta\xi(\mathbf{x};F_{\Delta\epsilon|\mathbf{X}})$ for all $\mathbf{x}\in\Omega_{\mathbf{X}}$.

The restriction of finite state spaces is not essential for the result. By finiteness of $\Omega_{\mathbf{X}}$ and recursive substitution of ω_j and ξ_j in (28), the identification region can be characterized by the linear system:where $\mathbf{I}$ is the K-by-K identity matrix, and $\mathbf{A}$ and $\mathbf{B}$ are K-by-$2K$ matrices defined as

$$
\begin{aligned}
\mathbf{A}(\Gamma) &\equiv [(\mathbf{I}+\mathbf{G}^1_\infty),\ -(\mathbf{I}+\mathbf{G}^0_\infty)] \\
\mathbf{B}(\Gamma) &\equiv [\mathbf{I}+(\mathbf{I}+\mathbf{G}^1_\infty)\beta\mathbf{G}^1,\ (\mathbf{I}+\mathbf{G}^0_\infty)\beta\mathbf{G}^0-(\mathbf{I}+\mathbf{G}^1_\infty)\beta\mathbf{G}^1] \\
&= [(\mathbf{I}+\mathbf{G}^1_\infty),\ (\mathbf{I}+\mathbf{G}^0_\infty)-(\mathbf{I}+\mathbf{G}^1_\infty)]
\end{aligned}
$$

When $\mathbf{I}-\beta\mathbf{G}^j$ has full rank, $\mathbf{I}+\mathbf{G}^j_\infty$ can be calculated easily as $(\mathbf{I}-\beta\mathbf{G}^j)^{-1}$. This completes the proof of Lemma 1.

7.3 Details in Implementing the Algorithm for Recovering RCCP

Below we describe detailed steps in implementing the algorithm used to recover the identified set of RCCP in Example 4.1. (Implementation of the algorithm in Examples 4.2, and Examples 5,6 all follow almost identical steps.) We show how to check whether a candidate vector of choice probabilities $\mathbf{p}_2$ is consistent with the model restrictions and the counterfactual environment Γ_2, given $\mathbf{p}_1$ is observed in the DGP Γ_1.

Step 1: Construct an augmented matrix from coefficients of the linear equations in (15)

$$
\left[\begin{array}{c|cc}
\mathbf{A}(\Gamma_1) & \mathbf{B}(\Gamma_1) & \mathbf{0} \\
\mathbf{A}(\Gamma_2) & \mathbf{0} & \mathbf{B}(\Gamma_2)
\end{array}\right] \tag{31}
$$

where the four matrices $\{A(\Gamma_m),B(\Gamma_m)\}_{m=1,2}$ are calculated from β, $\{\mathbf{G}^1_m,\mathbf{G}^0_m\}_{m=1,2}$. Find the reduced row echelon form of the matrix in (31) through Gaussian eliminations:

$$
\left[\begin{array}{cccccc|c}
1 & 0 & 0 & 0 & 0 & -1 & \\
0 & 1 & 0 & 0 & 0 & -1 & \mathbf{R}^I_1(\Gamma_1,\Gamma_2) \\
0 & 0 & 1 & 0 & 0 & -1 & r\text{-by-}4K \\
0 & 0 & 0 & 1 & 0 & -1 & \\
0 & 0 & 0 & 0 & 1 & -1 & \mathbf{R}^E_1(\Gamma_1,\Gamma_2) \\
0 & 0 & 0 & 0 & 0 & 0 & (2K-r)\text{-by-}4K
\end{array}\right]
$$

where $K = 3$ and $r = rank(\Sigma_L) = 5$ in Example 4.1, and

$$\mathbf{R}_1^I(\Gamma_1,\Gamma_2) \equiv$$
$$\begin{bmatrix} 0 & -3 & -4/3 & -4/3 & 0 & 4/3 & 1 & 3 & 4/3 & 1/3 & 0 & -1/3 \\ 0 & -2 & -18/5 & 0 & -8/5 & 8/5 & 0 & 3 & 18/5 & 0 & 3/5 & -3/5 \\ 0 & 4/15 & -1/3 & 4/15 & -4/15 & 0 & 0 & -4/15 & 4/3 & -4/15 & 4/15 & 0 \\ 0 & -10/3 & -29/15 & -4/3 & -4/15 & 8/5 & 0 & 10/3 & 29/15 & 1/3 & 4/15 & -3/5 \\ 0 & -7/5 & -10/3 & 4/15 & -8/5 & 4/3 & 0 & 7/5 & 10/3 & -4/15 & 3/5 & -1/3 \end{bmatrix}$$
$$\mathbf{R}_1^E(\Gamma_1,\Gamma_2) \equiv \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 \end{bmatrix}$$

(We shall drop the notation Γ_1, Γ_2 for notational ease.) We distinguish equalities from strict inequalities among the linear restrictions, and this is crucial for our algorithm for reasons that will become clear later. Let $\boldsymbol{\theta} \equiv \begin{bmatrix} \mathbf{Q}_1' & \boldsymbol{\kappa}_1^{0\prime} & \mathbf{Q}_2' & \boldsymbol{\kappa}_2^{0\prime} \end{bmatrix}'$ denote the 12-by-1 vector of unknown distributional parameters (i.e. $Q_{m,k} \equiv F_{\Delta\epsilon}^{-1}(p_{m,k})$, $\kappa_{m,k}^0 \equiv \kappa^0(p_{m,k}; F_{\Delta\epsilon}) \equiv \int_{-\infty}^{Q_{m,k}} (Q_{m,k} - s) dF_{\Delta\epsilon}(s)$ for $m = 1, 2$ and $k = 1, 2, 3$). The pick any real number d (e.g. $d = 0$ for simpler algebra) and set $u_{0K} = d$ for a locational normalization. Thus $\mathbf{u}$ can be expressed as

$$\mathbf{u} = [\mathbf{R}_1^I; \mathbf{R}_1^E]\boldsymbol{\theta} + \mathbf{1}'c$$

where $\mathbf{1}'$ is a conformable column vector of 1. Then substitute this into exogenous shape restrictions $\mathbf{Cu} > \mathbf{0}$, and reformulate these restrictions as a system of linear inequalities on $\boldsymbol{\theta}$,

$$\mathbf{C}[\mathbf{R}_1^I; \mathbf{R}_1^E]\boldsymbol{\theta} > \mathbf{d}_1^I \equiv -\mathbf{C}\mathbf{1}'d$$

Then the statement that "$\exists$ $\mathbf{u}, \boldsymbol{\theta}$ that solve (15)-(18)" is equivalent to the statement that "$\exists$ $\boldsymbol{\theta}$ that satisfy (16), (17), (18) and

$$\mathbf{C}[\mathbf{R}_1^I; \mathbf{R}_1^E]\boldsymbol{\theta} > \mathbf{d}_1^I \quad (32)$$
$$\mathbf{R}_1^E\boldsymbol{\theta} = \mathbf{0} \quad (33)$$

", where the strict linear inequalities capture the shape restrictions in C and the equality restrictions capture the testable implications on $\mathbf{p}_1, \mathbf{p}_2$ under (Γ_1, Γ_2) jointly. The fact that $\boldsymbol{\kappa}_1^0, \boldsymbol{\kappa}_2^0$ do not enter the testable implications in (33) is a coincidence solely due to the form of $\mathbf{G}_m^1, \mathbf{G}_m^0$ chosen.

Step 2: For the pair of $(\mathbf{p}_1, \mathbf{p}_2)$ considered, formulate restrictions in (16), (17) and (18)) in the form of linear inequalities and equalities

$$\mathbf{R}_2^I(\mathbf{p}_1, \mathbf{p}_2)\boldsymbol{\theta} \;>\; \mathbf{d}_2^I \quad (34)$$
$$\mathbf{R}_2^E(\mathbf{p}_1, \mathbf{p}_2)\boldsymbol{\theta} \;=\; \mathbf{d}_2^E \quad (35)$$

where the matrices of coefficients in the inequality and equality restrictions (denoted $\mathbf{R}_2^I$ and $\mathbf{R}_2^E$ respectively) are completely determined once $(\mathbf{p}_1, \mathbf{p}_2)$ are given. For example, the choice probabilities observed under Γ_1 is $\mathbf{p}_1 = [1/3, 3/5, 1/2]$ and suppose the hypothetical counterfactual choice outcomes under Γ_2 considered is $\mathbf{p}_2 = [1/5, 1/4, 3/4]$. Then the first three rows in $\mathbf{R}_2^I(\mathbf{p}_1, \mathbf{p}_2)$ that involves comparing $Q_{11}, Q_{12}, \kappa_{11}^0, \kappa_{12}^0$ is

$$\begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -3/5 & 3/5 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & -1/3 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

with the corresponding first three coordinates in the right-hand side vector $\mathbf{d}_2^I$ is a 3-by-1 zero vector. Likewise,

$$\mathbf{R}_2^E(\mathbf{p}_1, \mathbf{p}_2) = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

with $\mathbf{d}_2^E = [0, 1]$ (where we have used the normalization $\kappa^0(0.5) = 1$).

Step 3: It remains to check the feasibility of the linear system

$$\mathbf{C}[\mathbf{R}_1^I; \mathbf{R}_1^E]\boldsymbol{\theta} > \mathbf{d}_1^I \ ; \ \mathbf{R}_2^I\boldsymbol{\theta} > \mathbf{d}_2^I \tag{36}$$

$$\mathbf{R}_1^E\boldsymbol{\theta} = \mathbf{0} \ ; \ \mathbf{R}_2^E\boldsymbol{\theta} = \mathbf{d}_2^E \tag{37}$$

where the decision environment (Γ_1, Γ_2) enters $\mathbf{R}_1^I, \mathbf{R}_1^E$ and $(\mathbf{p}_1, \mathbf{p}_2)$ enters $\mathbf{R}_2^I, \mathbf{R}_2^E$. We then eliminate the equality constraints by substituting out a subvector of $\boldsymbol{\theta}$ using the reduced row echelon form of the equality restrictions in (37). This leads to an equivalent representation of the linear system:

$$\tilde{\mathbf{R}}\tilde{\boldsymbol{\theta}} > \tilde{\mathbf{d}} \tag{38}$$

where $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^{\tilde{K}}$ is a subvector of $\boldsymbol{\theta} \in \mathbb{R}^{4K}$ after substitution of (37) into (36), and $\tilde{\mathbf{R}}$ is the corresponding new matrix of coefficients.

Step 4: We check the feasibility of the linear system (38) by solving the following linear programming problem:

$$\min_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^{\tilde{K}},\ t \in \mathbb{R}^1} \quad t \qquad s.t. \ -\tilde{\mathbf{R}}\tilde{\boldsymbol{\theta}} + \tilde{\mathbf{d}} \leq \mathbf{1}'t \tag{39}$$

where $\mathbf{1}'$ is a conformable column vector of 1. If $\tilde{\mathbf{R}}\tilde{\boldsymbol{\theta}} > \tilde{\mathbf{d}}$ holds for some $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^{4K}$, then the solution to (39) must be strictly negative. Otherwise, if for all $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^{\tilde{K}}$ at least one of the strict inequalities in (38) fail to hold, then the solution of (39) must converge to some number $t^* \geq 0$. We use the LINPROG command in MatLab to solve this linear programming problem and exclude the $\mathbf{p}_2$ considered from the identified scope of counterfactual outcomes under Γ_2 if the solution converges to $t^* \geq 0$.

8 Figures in the Examples

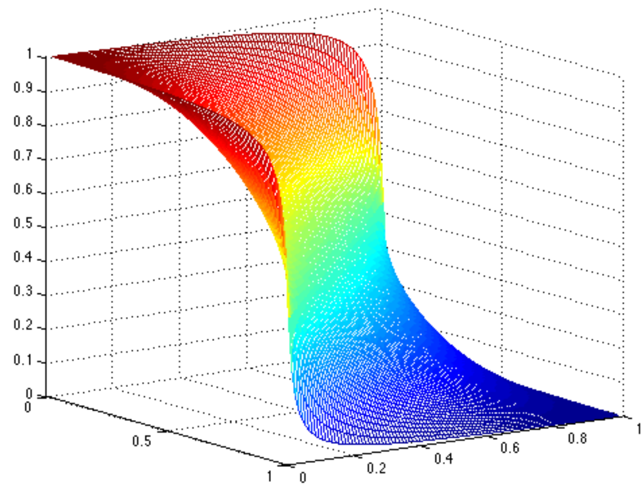

Figure 1: Testable Implications in Example 2

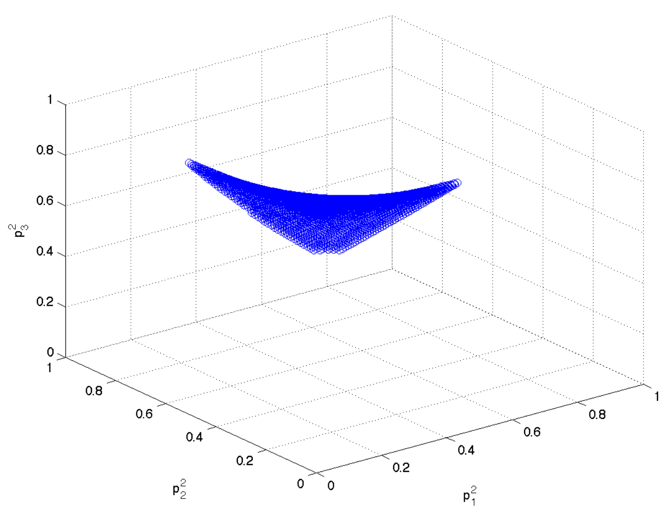

Figure 2.1: RCCP in Example 3 under shape restrictions in (9)

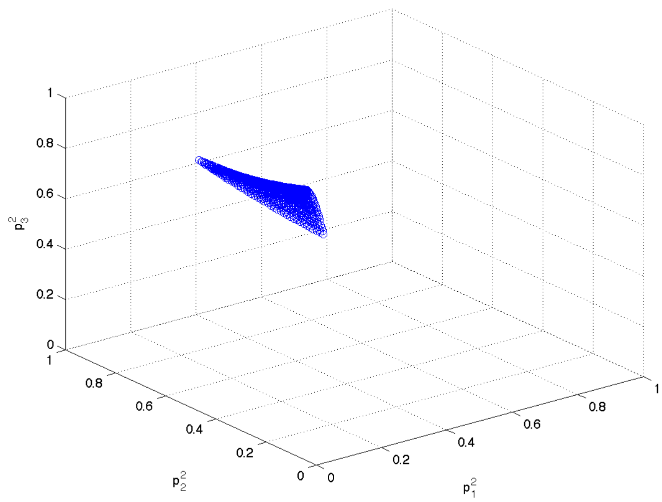

Figure 2.2: RCCP in Example 3 under shape restrictions in (9) and (10)

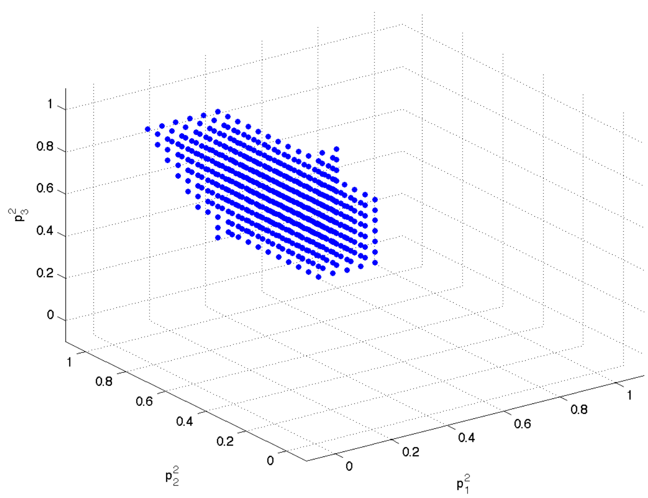

Figure 3.1: RCCP in Example 4.1

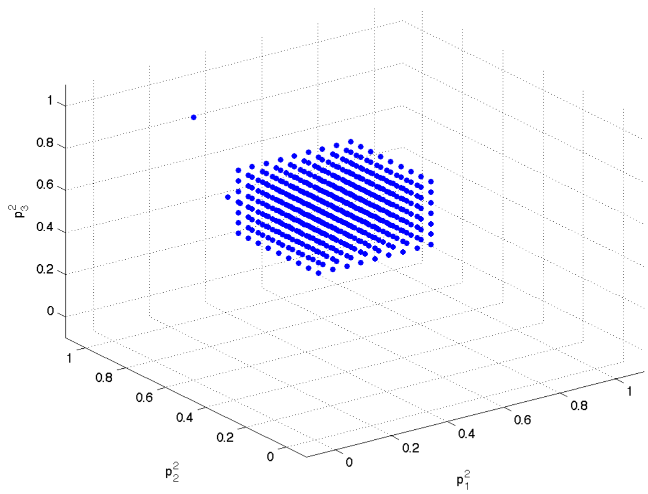

Figure 3.2: RCCP in Example 4.2

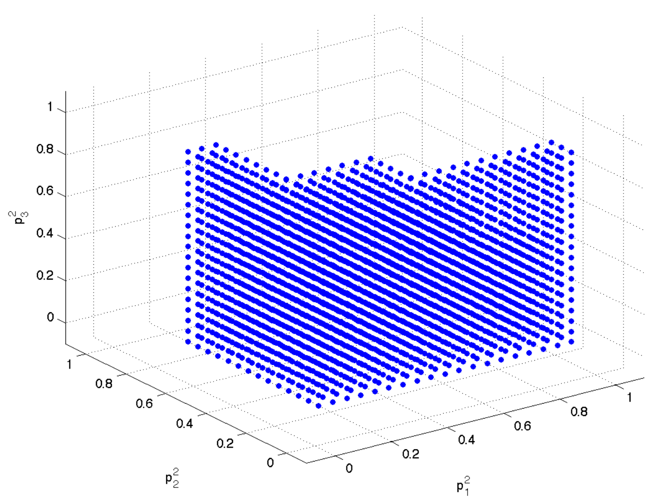

Figure 4: RCCP in Example 5

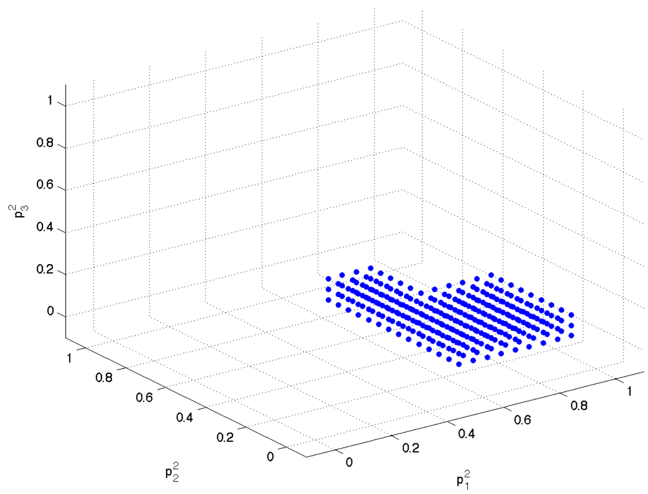

Figure 5.1: RCCP in Example 6 with no shape restriction on SPP

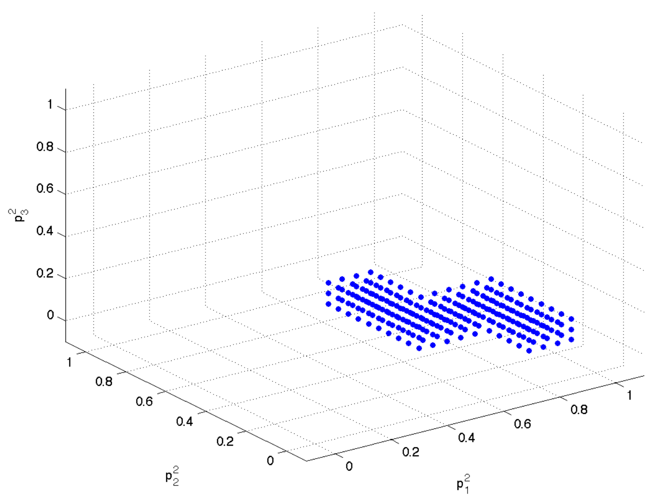

Figure 5.2: RCCP in Example 6 with shape restrictions on SPP

References

[1] Aguirregabiria, V., "Another look at the identification of dynamic discrete decision processes: With an application to retirement behavior", Journal of Business and Economic Statistics, 2008

[2] Aguirregabiria, V., "Nonparametric Identification of Behavioral Response to Counterfactual Interventions of Dynamic Discrete Choice Process", Economics Letters, 2005

[3] Aguirregabiria, V. and P. Mira, "Dynamic Discrete Choice Structural Models: A Survey", working paper, University of Toronto July 2008

[4] Arcidiacono, P. and J. Jones, "Finite mixture distributions, sequential likelihood, and the EM Algorithm", Econometrica 71, 933-946, 2003

[5] Berry, S. and E. Tamer, "Identification in Models of Oligopoly Entry," working paper, April 2006.

[6] Brien, M, L. Lillard, and S. Stern, "Cohabitation, Marriage, and Divorce in a Model of Match Quality", International Economic Review 47, 451-494, 2006

[7] Eckstein, Z., and K. Wolpin, "The specification and estimation of dynamic stochastic discrete choice models", Journal of Human Resources 24, 562-598, 1989

[8] Fang, H. and Y. Wang, "", manuscript, Duke University, 2008

[9] Heckman, J. and S. Navarro, "Dynamic Discrete Choice and Dynamic Treatment Effects", Journal of Econometrics 136, 341-396, 2007

[10] Hendel, I. and A. Nevo, "Measuring the Implications of Sales and Consumer Stockpiling Behavior", Econometrica 2006

[11] Keane, M. and K. Wolpin, "The career decisions of young men", Journal of Political Economy 105, 473-522, 1997

[12] Magnac, T. and D. Thesmar, "Identifying dynamic discrete decision processes", Econometrica, 70, p801-816, 2002

[13] Rust, J., "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher", Econometrica, 55, p999-1033, 1987

[14] Rust, J., "Structural estimation of Markov decision processes", Handbook of econometrics Volume 4, North-Holland, 1994

[15] Pakes, A., M. Ostrovsky, and S. Berry, "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)", Rand Journal of Economics 38, 373-399, 2007

[16] Pesendorfer, M. and Schmidt-Dengler, "Asymptotic Least Squares Estimators for Dynamic Games", Review of Economic Studies, 2008.

[17] Wolpin, K., "Estimating a Structural Search Model: The Transition from Schooling to Work", Econometrica 55, 801-818, 1987