NBER WORKING PAPER SERIES

IDENTIFICATION OF COUNTERFACTUALS IN DYNAMIC DISCRETE CHOICE MODELS

Myrto Kalouptsidi
Paul T. Scott
Eduardo Souza-Rodrigues

Working Paper 21527
http://www.nber.org/papers/w21527

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2015, Revised February 2020

Previously circulated as "Identification of Counterfactuals and Payoffs in Dynamic Discrete Choice with an Application to Land Use." We are grateful to Victor Aguirregabiria, Ivan Canay, Phil Haile, Bo Honore, Jakub Kastl, Yuichi Kitamura, Arthur Lewbel, Yao Luo, Rob McMillan, Ariel Pakes, Rodrigo Pinto, John Rust, Junichi Suzuki, Elie Tamer, Ali Yurukoglu, and seminar participants at various universities for many helpful comments. We also thank the Editor and the anonymous referees. Charis Katsiardis and Adrian Torchiana provided excellent research assistance. The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme Grant Agreement no. 230589 and Agence Nationale de la Recherche Projet ANR-12-CHEX-0012-01. All remaining errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Myrto Kalouptsidi, Paul T. Scott, and Eduardo Souza-Rodrigues. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Identification of Counterfactuals in Dynamic Discrete Choice Models
Myrto Kalouptsidi, Paul T. Scott, and Eduardo Souza-Rodrigues
NBER Working Paper No. 21527
September 2015, Revised February 2020
JEL No. C5,Q1

ABSTRACT

Dynamic discrete choice (DDC) models are not identified nonparametrically, but the non-identification of models does not necessarily imply the non-identification of counterfactuals. We derive novel results for the identification of counterfactuals in DDC models, such as non- additive changes in payoffs or changes to agents' choice sets. In doing so, we propose a general framework that allows the investigation of the identification of a broad class of counterfactuals (covering virtually any counterfactual encountered in applied work). To illustrate the results, we consider a firm entry/exit problem numerically, as well as an empirical model of agricultural land use. In each case, we provide examples of both identified and non-identified counterfactuals of interest.

Myrto Kalouptsidi
Department of Economics
Harvard University
Littauer 124
Cambridge, MA 02138
and NBER
myrto@fas.harvard.edu

Eduardo Souza-Rodrigues
Department of Economics
University of Toronto
Max Gluskin House,
150 St. George Street, room 324
Toronto, ON, Canada
edusouzarod@gmail.com

Paul T. Scott
NYU Stern School of Business
Kaufman Management Center, 7-77
New York University
New York, NY 10012
ptscott@gmail.com

1 Introduction

Since the seminal contributions of Rust (1994) and Magnac and Thesmar (2002), it is well known that dynamic discrete choice (DDC) models are not identified nonparametrically: several payoff functions can rationalize observed choice behavior. As a result, researchers must impose additional restrictions to estimate DDC models, usually with the goal of performing counterfactuals. When all models consistent with the data generate the same behavioral response in a given counterfactual, then the counterfactual is said to be identified. In some cases, however, different models that are consistent with the data generate different behavioral responses in a counterfactual; then, the counterfactual is not identified. As often there is little guidance as to reasonable restrictions that are necessary to identify the model, one may be concerned about the robustness of the empirical findings.

A recent body of innovative work – Aguirregabiria (2010), Aguirregabiria and Suzuki (2014), Norets and Tang (2014), and Arcidiacono and Miller (2019) – has made valuable progress in this area. These papers have established the identification of two important categories of counterfactuals in different classes of DDC models: counterfactual behavior is identified when flow payoffs change additively by pre-specified amounts; counterfactual behavior is generally not identified when the state transition process changes.

This paper builds on that body of research in three respects. First, we propose a general framework that allows us to consider counterfactuals with non-additive changes in payoffs or with changes to agents' choice sets (in addition to and in combination with the cases considered in previous studies).[1] Examples include assigning the primitives of one group of agents to those of another (e.g., assuming preferences of labor market cohorts are equal, or firm entry costs are identical across markets) and changing payoffs proportionally (e.g., subsidies that reduce firms' entry/sunk costs by some percentage), among others. Second, we investigate how and whether parametric restrictions affect the non-identification of counterfactuals by considering a family of parametric models that encompasses many studies in the literature. Third, we add to existing results that focus on counterfactual behavior by considering the identification of counterfactual welfare, which is often the ultimate object of interest to policy makers.

To that end, we develop a novel approach that allows us to derive the set of necessary and sufficient conditions to identify counterfactual behavior and welfare for a broad class of counterfactuals. We consider counterfactuals that involve almost any change in the primitives, so our results can be used on a case-by-case basis to investigate the identification of particular policy interventions of interest. We first note that Magnac and Thesmar's (2002) underidentification result implies a convenient representation that directly relates counterfactual choice probabilities to a set of "free parameters"; i.e., a subset of the payoff parameters that, if known, deliver all remaining

[1] We consider any (simultaneous) change in all model primitives, with the exception of non-differentiable changes in the payoff function (which are uncommon in practice).

unknown parameters. Based on this representation, we can determine the conditions under which counterfactual behavior and welfare are identified. When such conditions are satisfied, it is not necessary to identify all the individual structural parameters of the model to identify the effects of policy interventions.[2]

Our results imply that the identification of counterfactuals can be verified directly from data on the state transition process. In some cases, identification can be determined without even examining the data. For example, counterfactuals eliminating an action from the choice set result in identified counterfactual behavior; counterfactuals assigning the payoff parameters from one group of agents to another are not identified, except in special cases.[3] While prior studies show that pre-specified additive shifts in flow payoffs result in identified counterfactual behavior, we demonstrate that all other counterfactual transformations of payoff functions are identified only under restrictive conditions that require verification in the data.

Given that some counterfactuals are not identified in the nonparametric setup, it is natural to wonder whether parametric assumptions, which are prevalent in applied work, can help identify specific sets of counterfactuals, even when the model is not fully identified. We find that they do, assuming of course that the parametric restrictions are correctly specified. For instance, a number of papers have implemented a counterfactual that changes the volatility or long-run mean of market states (i.e., a change in the state transition process); while such counterfactuals are not identified in a nonparametric setting, we show that most examples of these counterfactuals in the literature are identified in the parametric setting.

In addition, we consider the identification of welfare, which is often the ultimate object of interest to policy makers (in terms of both sign and magnitude). We find that the identification of counterfactual behavior is necessary but not sufficient for the identification of welfare. We also provide sufficient conditions for welfare identification.

Recognizing that static models are a special case of dynamic models, our framework can be used to understand which counterfactuals are identified in static settings. Our results show that, compared to dynamics, static settings do not require restrictions on state transitions for counterfactual identification as dynamic models do. As a result, a larger set of counterfactuals is identified in static compared to dynamic models.[4]

To gain intuition and explore how sensitive counterfactuals can be to model restrictions in practice, we turn to two applications. The first is a numerical exercise featuring a dynamic firm entry/exit model. To identify this model, the researcher has to restrict scrap values, entry costs or fixed costs; this is usually accomplished by fixing one of them to zero. Such an assumption

[2]See Ichimura and Taber (2000, 2002) for direct estimation of policy impacts in the context of selection models.

[3]Sometimes eliminating an action also eliminates states (e.g. that may reflect past actions). We also characterize this case, providing conditions that the counterfactual state transition process (across the remaining states) must satisfy for identification. Similarly, a counterfactual that adds an action is identified provided that the counterfactual payoffs of the new action are a convex combination of the baseline payoffs.

[4]We also discuss how our results can be extended to dynamic models with continuous choices; see Section 3.4.

is difficult to justify however as cost or scrap value data are extremely rare.[5] The restrictions can affect the parameter estimates and, for non-identified counterfactuals, alter the counterfactual predictions as well. For instance, when fixed costs are set to zero, the estimated profit is high, which provides the incentive to enter and stay in the market. Then, in order to match the observed choices, estimated entry costs and scrap values must be high as well. Although the estimated model with zero fixed costs is observationally equivalent to the true model, when we implement a counterfactual subsidy that reduces entry costs proportionally, the predicted impact on turnover and welfare are exaggerated in the estimated model. Specifically, as the estimated entry costs and scrap values are magnified, it becomes profitable to enter and exit the market repeatedly under the entry subsidy. A similar issue arises when the scrap value is set to zero instead. In this case, the estimated entry costs must be low (or even negative) in order to rationalize observed entry: since the scrap value is zero, entering is not as attractive to a forward-looking agent and so entry costs must be reduced to provide incentives to enter. Applying the same subsidy again results in incorrect counterfactual predictions (which may even go in the wrong direction).

Next, we consider the empirical relevance of our results in the context of US agricultural land use. Following Scott (2013), field owners decide whether to plant crops or not, and face uncertainty (regarding commodity prices, weather shocks, government interventions, etc.) as well as switching costs between land uses (which create dynamic incentives). To estimate the model, Scott (2013) adopts a parametric specification and restricts the value of a subset of the switching cost parameters. As there is little guidance in the literature concerning how to specify the particular values, he sets them to zero. To evaluate the impact of these restrictions on counterfactual analysis, we bring in additional data and augment Scott's estimation strategy using land resale price data. Similar to Kalouptsidi (2014), we treat farmland resale transaction prices as a measure of agents' value functions. The augmented estimator allows us to test Scott's identifying restrictions and reject them.[6]

We then implement two counterfactuals. First, we consider a long-run land use elasticity, which measures the sensitivity of land use to a persistent change in crop returns. This elasticity is an important input to the analysis of several policy interventions, including agricultural subsidies and biofuel mandates (Roberts and Schlenker, 2013; Scott, 2013). The second counterfactual features an increase in the cost of replanting crops and resembles a fertilizer tax (higher fertilizer prices would be a likely consequence of pricing greenhouse gas emissions, as fertilizer production is very fossil-fuel intensive). We show that while the long-run elasticity is identified, the fertilizer tax is not. Thus, a model estimated with our augmented estimator and a model imposing Scott's restrictions both predict the same long-run elasticity, but they predict different responses to the increase in fertilizer taxes (and even responses in different directions).

[5] Using external information on entry costs and scrap values (specifically, new ship prices and demolition prices), Kalouptsidi (2014) shows that the latter vary dramatically over states in the shipping industry.

[6] Relating land resale price data to the model requires another set of assumptions about land markets. See Kalouptsidi (2014) for a full discussion of these restrictions.

Related Literature. Our paper relates to several important prior studies. In addition to Rust's (1994) and Magnac and Thesmar's (2002) seminal contributions, Heckman and Navarro (2007) consider the identification of a semiparametric finite horizon optimal stopping time model that allows for a rich time series dependence on the unobservables. Heckman, Humphries, and Veramendi (2016) then extend the work of Heckman and Navarro (2007) by incorporating both ordered and unordered choice sets, and by decomposing the identification of dynamic treatment effects into direct effects and continuation values. Under a conditional independence assumption on the unobservables, Bajari, Chu, Nekipelov, and Park (2016) study the identification of finite-horizon models with a terminal action, while Abbring and Daljord (2019) investigate the conditions needed to identify the discount factor. Pesendorfer and Schmidt-Dengler (2008) extend Magnac and Thesmar's results to dynamic games.[7]

Regarding the identification of counterfactuals in DDC models, Aguirregabiria (2010) shows identification of counterfactual choice probabilities when the experiment consists of adding a prespecified amount to payoffs in a finite-horizon binary choice model. Aguirregabiria and Suzuki (2014) and Norets and Tang (2014) extend to Aguirregabiria's (2010) result to infinite horizon models. They both provide another important extension by showing nonidentification of behavior under changes in transition probabilities. Arcidiacono and Miller (2019) further extend these results to multinomial choice models for both stationary and nonstationary environments in the presence of long and short panel data.

We focus on infinite horizon multinomial choice models and complement the literature by (a) providing the first full set of necessary and sufficient conditions for identification of counterfactuals involving almost any change in the model primitives, including non-additive changes in payoffs, (b) investigating the identification power of parametric restrictions to the basic framework, and (c) providing identification results for counterfactual welfare. Our results extend to models with permanent unobserved heterogeneity, provided that conditional choice probabilities and transition functions of finitely many unobserved types are identified in a first step, as in Kasahara and Shimotsu (2009) or Hu and Shum (2012). In a companion paper (Kalouptsidi, Scott, and Souza-Rodrigues, 2017), we consider the identification of counterfactual behavior in dynamic games. Recently, Kalouptsidi, Kitamura, Lima, and Souza-Rodrigues (2020) extended our results to investigate partial identification of counterfactual outcomes of interest, along with uniformly valid inference procedures based on subsampling.[8]

[7] Recent work considers the identification of the distribution of the idiosyncratic shocks when agents can make both discrete and continuous choices (Blevins, 2014), or in the presence of continuous states and exclusion restrictions (Chen, 2017), or under linear-in-parameters payoff functions (Buchholz, Shum, and Xu, 2019).

[8] Aguirregabiria and Suzuki (2014) provide results in the context of a monopolist entry/exit problem. In addition to point identification, Norets and Tang (2014) relax the assumption that the distribution of the idiosyncratic shocks is known by the econometrician and, as a consequence, obtain some partial identification results. We do not cover partial identification as in Norets and Tang (2014), and we do not consider nonstationary settings as in Aguirregabiria (2010) and Arcidiacono and Miller (2019).

The paper is organized as follows: Section 2 presents the dynamic discrete choice framework and reconstructs the known results on the nonparametric underidentification of standard DDC models. Section 3 contains our main results relating to the identification of counterfactual behavior and welfare, in both nonparametric and parametric settings. Section 4 discusses our two applications: a numerical firm entry model and an empirical model of agricultural land use. Section 5 concludes. The Appendix contains all mathematical proofs; and the Suplemental Material provides the details of the dataset and the implementation of the empirical application.

2 Dynamic Framework

In each period $t \in \{1, 2, ...\}$, agent i chooses one action a_{it} from the finite set $\mathcal{A} = \{1, ..., A\}$. The per period payoff depends on the state variables $(x_{it}, \varepsilon_{it})$, where x_{it} is observed by the econometrician and ε_{it} is not. We assume $x_{it} \in \mathcal{X} = \{1, ..., X\}$, $X < \infty$, while $\varepsilon_{it} = (\varepsilon_{1it}, ..., \varepsilon_{Ait})$ is i.i.d. across agents and time and has joint distribution G that is absoluetly continuous with respect to the Lebesgue measure and has full support on $\mathbb{R}^A$. The transition distribution function for $(x_{it}, \varepsilon_{it})$ factors as follows:

$$F\left(x_{it+1}, \varepsilon_{it+1} | a_{it}, x_{it}, \varepsilon_{it}\right) = F\left(x_{it+1} | a_{it}, x_{it}\right) G\left(\varepsilon_{it+1}\right).$$

Agents have no private information about future values of x_{it} and ε_{it}. The per period payoff function is given by

$$\pi\left(a, x_{it}, \varepsilon_{it}\right) = \pi_a(x_{it}) + \varepsilon_{ait}.$$

Agent i chooses a sequence of actions to maximize the expected discounted sum of current and future payoffs. Let $V\left(x_{it}, \varepsilon_{it}\right)$ denote the agent's value function. By Bellman's principle of optimality,

$$V\left(x_{it}, \varepsilon_{it}\right) = \max_{a \in \mathcal{A}} \left\{\pi_a(x_{it}) + \varepsilon_{ait} + \beta E\left[V\left(x_{it+1}, \varepsilon_{it+1}\right) | a, x_{it}\right]\right\},$$

where $\beta \in [0, 1)$ is the discount factor. We define both the *ex ante value function* $V\left(x_{it}\right) \equiv \int V\left(x_{it}, \varepsilon_{it}\right) dG\left(\varepsilon_{it}\right)$, and the *conditional value function*

$$v_a\left(x_{it}\right) \equiv \pi_a(x_{it}) + \beta E\left[V\left(x_{it+1}\right) | a, x_{it}\right]. \tag{1}$$

The agent's optimal policy is given by the conditional choice probabilities (CCPs):

$$p_a\left(x_{it}\right) = \int 1\left\{v_a\left(x_{it}\right) + \varepsilon_{ait} \geq v_j\left(x_{it}\right) + \varepsilon_{jit}, \text{ for all } j \in \mathcal{A}\right\} dG\left(\varepsilon_{it}\right),$$

where $1\{\cdot\}$ is the indicator function. We define the vectors $p\left(x\right) = \left[p_1\left(x\right), ..., p_{A-1}\left(x\right)\right]$ and $p = \left[p\left(1\right), ..., p\left(X\right)\right]$.

It is useful to note that for any (a, x) there exists a real-valued function $\psi_a\left(.\right)$ derived only from G such that

$$V(x) = v_a(x) + \psi_a(p(x)). \quad (2)$$

Equation (2) states that the ex ante value function V equals the value obtained by choosing a today and optimally thereafter (v_a) plus a correction term (ψ_a), because choosing action a today is not necessarily optimal. When ε_{it} follows the extreme value distribution, $\psi_a(p(x)) = \kappa - \ln p_a(x)$, where κ is the Euler constant.[9]

As we make extensive use of matrix notation below, we define the vectors $\pi_a, v_a, V, \psi_a \in \mathbb{R}^X$, which stack $\pi_a(x)$, $v_a(x)$, $V(x)$, and $\psi_a(p(x))$, for all $x \in \mathcal{X}$. We often use the notation $\psi_a(p)$ to emphasize the dependence of ψ_a on the choice probabilities p. We also define F_a as the transition matrix with (m,n) element equal to $\Pr(x_{it+1} = x_n | x_{it} = x_m, a)$. The payoff vector $\pi \in \mathbb{R}^{AX}$ stacks π_a for all $a \in \mathcal{A}$, and, similarly, F stacks (a vectorized version) of F_a for all $a \in \mathcal{A}$. While we assume a finite state space, our results can be extended to state variables with continuous support.

Nonparametric Identification of Payoffs. A dynamic discrete choice model consists of the primitives $(\mathcal{A}, \mathcal{X}, \beta, G, F, \pi)$ that generate the endogenous objects $\{p_a, v_a, V,\ a \in \mathcal{A}\}$. Typically, the available data consist of agents' actions at different states, (a_{it}, x_{it}), which implies the CCPs $p_a(x)$ and the transition F are also known. Further, we follow the literature and assume that (β, G) is known as well.[10] The objective is to identify the payoff function π. Intuitively, π has AX parameters, and there are only $(A-1)X$ observed CCPs; thus there are X free payoff parameters and X restrictions will need to be imposed (Rust, 1994; Magnac and Thesmar, 2002).

We represent the underidentification problem as follows. For all $a \neq J$, where $J \in \mathcal{A}$ is some reference action, π_a can be represented as an affine transformation of π_J:

$$\pi_a = A_a \pi_J + b_a(p), \quad (3)$$

where

$$A_a = (I - \beta F_a)(I - \beta F_J)^{-1}, \quad (4)$$

$$b_a(p) = A_a \psi_J(p) - \psi_a(p), \quad (5)$$

[9]Equation (2) is shown in Arcidiacono and Miller (2011, Lemma 1). It makes use of the Hotz-Miller inversion (Hotz and Miller, 1993), which, in turn, establishes that the difference of conditional value functions is a known function of the CCPs: $v_a(x) - v_j(x) = \phi_{aj}(p(x))$, where $\phi_{aj}(.)$ is again derived only from G. When ε_{it} follows the type I extreme value distribution, $\phi_{aj}(p(x)) = \log p_a(x) - \log p_j(x)$. Chiong, Galichon, and Shum (2016) propose a novel approach that can calculate ψ_a and ϕ_{aj} for a broad set of distributions G.

[10]Norets and Tang (2014) have considered the problem of identifying π when G is unknown. Blevins (2014), Chen (2017), and Buchholz, Shum, and Xu (2019) consider identification of G under different model assumptions. Magnac and Thesmar (2002) and Abbring and Daljord (2019) investigate sufficient conditions for identification of the discount factor. It is straightforward to combine assumptions that identify β and G with the results we present in the current paper.

and I is a (comformable) identity matrix.[11] In the logit model, $b_a(p) = \ln p_a - A_a \ln p_J$, where $\ln p_a$ is the $X \times 1$ vector with elements $\ln p_a(x)$. To simplify notation, we omit the dependence of both A_a and $b_a(p)$ on the transition probabilities F.

One can compute A_a and b_a directly from the data (a_{it}, x_{it}) for all $a \neq J$. Equations (3)–(5) therefore explicitly lay out how we might estimate the payoff function if we are willing to fix the payoffs of one action at all states *a priori* (e.g. $\pi_J = 0$). However, this is not the only way to obtain identification: We simply need to add X extra restrictions. Other common possibilities involve reducing the number of payoff function parameters to be estimated using parametric assumptions and/or exclusions restrictions. As long as the extra assumptions add X linearly independent restrictions to the $(A-1)X$ restrictions expressed by (3), π will be uniquely determined. Further, whichever extra restrictions are imposed, they are equivalent to stipulating the payoffs of a reference action; i.e. if π_J^* is the vector of payoffs for the reference action identified by some set of restrictions and (3), then that set of restrictions is equivalent to stipulating $\pi_J = \pi_J^*$ *a priori*.

It is worth noting that in static models, where $\beta = 0$ and/or $F_a = F_J$ for all a, J, equation (3) simplifies to $\pi_a - \pi_J = b_a$, which becomes the standard logistic regression if the shocks ε follow the type I extreme value distribution. As discussed later, all our results apply to static models as well.

For some results, it will be useful to represent (3) for all actions $a \neq J$ as

$$\pi_{-J} = A_{-J}\pi_J + b_{-J}, \tag{6}$$

where π_{-J} stacks π_a for all $a \neq J$; and the matrix A_{-J} and the vector b_{-J} are defined similarly. The underidentification problem is therefore represented by the free parameter π_J.

Remark 1. *(Unobserved Heterogeneity.) In the presence of unobserved heterogeneity, equation (6) holds for each unobserved type. The nature of the underidentification problem is therefore the same after type-specific CCPs and state transitions are identified (e.g., following the strategy proposed by Kasahara and Shimotsu (2009) or Hu and Shum (2012)).*

3 Identification of Counterfactuals

This section presents our main results on the identification of counterfactuals. We begin with a taxonomy of counterfactuals (Subsection 3.1); we then provide the necessary and sufficient con-

[11] To see why, fix the vector $\pi_J \in \mathbb{R}^X$. Then,

$$\pi_a = v_a - \beta F_a V = V - \psi_a - \beta F_a V = (I - \beta F_a) V - \psi_a,$$

where for $a = J$, we have $V = (I - \beta F_J)^{-1} (\pi_J + \psi_J)$. After substituting for V, we obtain the result. As an aside, note that $(I - \beta F_J)$ is invertible because F_J is a stochastic matrix and hence the largest eigenvalue is equal or smaller than one. The eigenvalues of $(I - \beta F_J)$ are given by $1 - \beta\gamma$, where γ are the eigenvalues of F_J. Because $\beta < 1$ and $\gamma \leq 1$, we have $1 - \beta\gamma > 0$.

ditions for identification of counterfactual behavior (Subsection 3.2); next, we investigate several special cases of practical interest (Subsection 3.3); we then provide some intuition for the results, and discuss briefly the implications for static models and continuous choice models (Subsection 3.4); we also analyze the identification of counterfactual behavior under parametric restrictions (Subsection 3.5); finally, we investigate counterfactual welfare (Subsection 3.6).[12]

3.1 Taxonomy of Counterfactuals

A counterfactual is defined by the tuple $\{\widetilde{\mathcal{A}}, \widetilde{\mathcal{X}}, \widetilde{\beta}, \widetilde{G}, h^s, h\}$. The sets $\widetilde{\mathcal{A}} = \{1, ..., \widetilde{A}\}$ and $\widetilde{\mathcal{X}} = \{1, ..., \widetilde{X}\}$ denote the new set of actions and states respectively. The new discount factor is $\widetilde{\beta}$, and the new distribution of the idiosyncratic shocks is $\widetilde{G}$.[13] The function $h^s : \mathbb{R}^{A \times X^2} \to \mathbb{R}^{\widetilde{A} \times \widetilde{X}^2}$ transforms the transition probability F into $\widetilde{F}$. Finally, the function $h : \mathbb{R}^{A \times X} \to \mathbb{R}^{\widetilde{A} \times \widetilde{X}}$ transforms the payoff function π into the counterfactual payoff $\widetilde{\pi}$, so that $\widetilde{\pi} = h(\pi)$, where $h(\pi) \equiv [h_1(\pi), ..., h_{\widetilde{A}}(\pi)]$, with $h_a(\pi) = h_a(\pi_1, ..., \pi_A)$ for each $a \in \widetilde{\mathcal{A}}$. Below, we discuss a number of special cases encountered in applied work.

Affine Payoff Counterfactuals. In affine payoff counterfactuals, the payoff $\widetilde{\pi}(a, x)$ at an action-state pair (a, x) is obtained as the sum of a scalar $g(a, x)$ and a linear combination of all baseline payoffs, so that:

$$\widetilde{\pi} = \mathcal{H}\pi + g, \tag{7}$$

where $\mathcal{H} \in \mathbb{R}^{\widetilde{A}\widetilde{X} \times AX}$ and g is a $\widetilde{A}\widetilde{X} \times 1$ vector. It is helpful to write this in a block-matrix equivalent form:

$$\widetilde{\pi} = \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1A} \\ \vdots & \vdots & \vdots & \vdots \\ H_{\widetilde{A}1} & H_{\widetilde{A}2} & \cdots & H_{\widetilde{A}A} \end{bmatrix} \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_A \end{bmatrix} + \begin{bmatrix} g_1 \\ \vdots \\ g_{\widetilde{A}} \end{bmatrix}, \tag{8}$$

where the submatrices H_{aj} have dimension $\widetilde{X} \times X$ for each pair $a \in \widetilde{\mathcal{A}}$ and $j \in \mathcal{A}$.

When the counterfactual does not change the set of actions and states (i.e. $\widetilde{\mathcal{A}} = \mathcal{A}$ and $\widetilde{\mathcal{X}} = \mathcal{X}$), $\mathcal{H}$ is a square matrix. When, further, $\widetilde{\pi}_a$ depends solely on π_a, $\mathcal{H}$ is block-diagonal and for all $a \in \mathcal{A}$,

$$\widetilde{\pi}_a = H_{aa}\pi_a + g_a. \tag{9}$$

[12] For dynamic games, we can always treat the problem of solving for an individual player's best response (holding the opponent's strategy fixed) as a single-agent problem. Our identification results can therefore be applied to investigate identification of counterfactual best responses in dynamic games. A full analysis naturally requires strategic considerations and the possibility of multiple equilibria. See Kalouptsidi, Scott, and Souza-Rodrigues (2017) for discussion of how strategic interactions makes the identification of counterfactuals in dynamic games particularly difficult.

[13] As previously mentioned, the discount factor is typically assumed known. We allow however for changes in β for completeness. When β is identified under further restrictions (Magnac and Thesmar, 2002; Abbring and Daljord, 2019), one may be interested in investigating behavior when the discount factor takes different values. For instance, Conlon (2012) studies the evolution of the LCD TV industry when consumers are myopic in the counterfactual experiment (i.e., $\widetilde{\beta} = 0$).

We call these "action diagonal counterfactuals." Below, we contrast three simple special cases of (9) that are common in applications.

Pre-Specified Additive Changes. This counterfactual takes H_{aa} as the identity matrix for all a (i.e. $\mathcal{H} = I$), so that $\widetilde{\pi}_a = \pi_a + g_a$. For instance, Keane and Wolpin (1997) investigate a hypothetical college tuition subsidy. Schiraldi (2011) and Wei and Li (2014) study automobile scrappage subsidies that depend on the car's model and age. Duflo, Hanna, and Ryan (2012) implement optimal bonus incentives for teachers in rural India, where the bonus depends on the number of classes the teachers attend.

"Pre-specified additive changes" have been considered by Aguirregabiria (2010), Aguirregabiria and Suzuki (2014), Norets and Tang (2014), and Arcidiacono and Miller (2019). Note that g is not allowed to depend on π, so the researcher must be able to specify g before estimating the model. Therefore, it is not possible to represent an arbitrary counterfactual $\widetilde{\pi} = h(\pi)$ by an "additive changes" in practice; this would require setting $g = h(\pi) - \pi$. In other words, payoffs must be changed by amounts that can be specified *without estimating the model.*

Proportional Changes. In this case $\mathcal{H}$ is diagonal and $g = 0$. The counterfactual imposes percentage changes on original payoffs, i.e. $\widetilde{\pi}_a(x) = \lambda_a(x)\,\pi_a(x)$. A common example involves entry subsidies represented by percentage changes on entry/sunk costs: for instance, Das, Roberts, and Tybout (2007) study firms' exporting decisions; Varela (2018) studies supermarket entry; Lin (2015) investigates entry and quality investment in the nursing home industry; and Igami (2017) studies innovation in the hard drive industry.[14]

Changes in Types. In this case, the primitives of one type of agents are replaced by those of another, where types can be broadly defined to include markets or regions. For instance, Keane and Wolpin (2010) replace the primitives of minorities by those of white women to investigate the racial-gap in labor markets. Eckstein and Lifshitz (2011) substitute the preference/costs parameters of the 1955's cohort by those of other cohorts to study the evolution of labor market conditions. Ryan (2012) replaces the entry costs post the Clean Air Act Amendment (CAAA) by those before the CAAA in the cement industry. Dunne, Klimek, Roberts, and Xu (2013) substitute entry costs in Health Professional Shortage Areas (HPSA) by those in the non-HPSA for dentists and chiropractors.

To represent such a counterfactual we can explicitly add a time-invariant state, denoted by s, the type, so that the payoff is written $\pi_a(x, s)$. For example, if there are two types, $s \in \{s_1, s_2\}$,

[14] To be precise, many of these applications involve proportional changes in a component of the payoff function, e.g., in fixed or sunk costs rather than in the whole profit function. We discuss this in Section 3.5.

a counterfactual in which the payoff of type s_1 is replaced by that of type s_2 is represented by

$$\begin{bmatrix} \widetilde{\pi}_a(s_1) \\ \widetilde{\pi}_a(s_2) \end{bmatrix} = \begin{bmatrix} 0 & I \\ 0 & I \end{bmatrix} \begin{bmatrix} \pi_a(s_1) \\ \pi_a(s_2) \end{bmatrix}, \tag{10}$$

where $\pi_a(s) \in \mathbb{R}^X$ and $\widetilde{\pi}_a(s) \in \mathbb{R}^{\widetilde{X}}$, for each type s. Note that H_{aa} is not diagonal in this case.

Changes in Choice Sets and State Space. Eliminating an action j leads to $\widetilde{\mathcal{A}} = \mathcal{A} - \{j\}$. In this case, $\widetilde{\pi}$ satisfies (8) with $H_{aa} = I$ and $H_{ak} = 0$ for $a \in \widetilde{\mathcal{A}}$ and $k \in \mathcal{A}$, $a \neq k$. For instance, if $A = 3$ and we drop action $j = 3$, (8) becomes

$$\begin{bmatrix} \widetilde{\pi}_1 \\ \widetilde{\pi}_2 \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix}.$$

Rust and Phelan (1997) eliminate social security in a retirement decision model. Gilleskie (1998) restricts access to medical care in the first days of illness. Crawford and Shum (2005) do not allow patients to switch medications to study the impact of experimentation. Keane and Wolpin (2010) eliminate a welfare program. Keane and Merlo (2010) eliminate the option of private jobs for politicians who leave congress. Note that in some cases, changing the set of actions also changes the set of states (e.g. when $x_t = a_{t-1}$ as in many entry models).

A counterfactual that adds a new action j can also be represented by (8): take $\widetilde{\mathcal{A}} = \mathcal{A} \cup \{j\}$ and let $H_{aa} = I$ and $H_{ak} = 0$ for $a \neq k, j$. Note that, adding an action also requires specifying its payoff $\widetilde{\pi}_j$, the new transition matrix $\widetilde{F}_j$, the (extended) joint distribution of the unobserved shocks $\widetilde{G}$, and possibly new states. Rosenzweig and Wolpin (1993), for example, add an insurance option for farmers in rural India.

Changes in Transitions. Finally, this counterfactual is represented by a function h^s that transforms F to $\widetilde{F}$; it may involve changes in the long-run mean or volatility of market-level variables. This is the second type of counterfactual that has been considered in the literature (Aguirregabiria and Suzuki, 2014; Norets and Tang, 2014; Arcidiacono and Miller, 2019). To give some examples of this counterfactual, Hendel and Nevo (2006) study consumers' long-run responsiveness to prices using supermarket scanner data. Collard-Wexler (2013) explores the effects of demand volatility in the ready-mix concrete industry. Kalouptsidi (2014) investigates the impact of time to build on industry fluctuations for the case of the shipping industry. Chan, Hamilton, and Papageorge (2016) evaluate the value, and the impact on risky behavior, of an HIV treatment breakthrough that affects the likelihood of HIV infection.

3.2 Identification of Counterfactual Behavior: The General Case

We now present our main theorem, which provides a general framework to investigate identification of counterfactual behavior; then, we turn to the special cases and provide some intuition for the results. The starting point is equation (3). This relationship is convenient for two reasons. First, it does not depend on continuation values. Second, the CCP vector generated by the model primitives is the unique vector that satisfies (3).[15]

The counterfactual $\{\widetilde{\mathcal{A}}, \widetilde{\mathcal{X}}, \widetilde{\beta}, \widetilde{G}, h^s, h\}$ determines a new set of primitives $(\widetilde{\mathcal{A}}, \widetilde{\mathcal{X}}, \widetilde{\beta}, \widetilde{G}, \widetilde{F}, \widetilde{\pi})$, with $\widetilde{F} = h^s(F)$ and $\widetilde{\pi} = h(\pi)$, which in turn leads to a new optimal behavior: the counterfactual CCP, denoted by $\widetilde{p}$. The counterfactual counterpart to (3) for any $a \in \widetilde{\mathcal{A}}$, with $a \neq J$, is

$$\widetilde{\pi}_a = \widetilde{A}_a \widetilde{\pi}_J + \widetilde{b}_a(\widetilde{p}), \tag{11}$$

where

$$\begin{aligned} \widetilde{A}_a &= (I - \widetilde{\beta}\widetilde{F}_a)(I - \widetilde{\beta}\widetilde{F}_J)^{-1}, \\ \widetilde{b}_a(\widetilde{p}) &= \widetilde{A}_a \widetilde{\psi}_J(\widetilde{p}) - \widetilde{\psi}_a(\widetilde{p}), \end{aligned}$$

the functions $\widetilde{\psi}_J$ and $\widetilde{\psi}_a$ depend on the new distribution $\widetilde{G}$, and we take without loss of generality a reference action J that belongs to both $\mathcal{A}$ and $\widetilde{\mathcal{A}}$.

It is clear from (3) and (11) that $\widetilde{p}$ is a function of the free parameter π_J. Because the lack of identification of the model is represented by this free parameter, the counterfactual CCP $\widetilde{p}$ is identified if and only if it does not depend on π_J. To determine whether or not this is the case, we apply the implicit function theorem to (11).

Before presenting the general case, we consider a binary choice example to fix ideas. Take $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{\beta} = \beta$, $\widetilde{G} = G$, and assume $\widetilde{\pi}_a$ is action diagonal so that $\widetilde{\pi}_a = h_a(\pi_a)$. Take $J = 2$, and rewrite (11) as

$$h_1(\pi_1) = \widetilde{A}_1 h_2(\pi_2) + \widetilde{b}_1(\widetilde{p}). \tag{12}$$

The implicit function theorem allows us to locally solve (12) with respect to $\widetilde{p}$ provided the matrix

$$\frac{\partial}{\partial \widetilde{p}} \left[h_1(\pi_1) - \widetilde{A}_1 h_2(\pi_2) - \widetilde{b}_1(\widetilde{p}) \right]$$

is invertible. We prove this matrix is indeed invertible in the general case (see Lemma 1 below). Then, it follows from the implicit function theorem that $\widetilde{p}$ does not depend on the free parameter π_2 if and only if

$$\frac{\partial}{\partial \pi_2} \left[h_1(\pi_1) - \widetilde{A}_1 h_2(\pi_2) - \widetilde{b}_1(\widetilde{p}) \right] = 0.$$

[15] Note that a unique CCP vector p is indeed guaranteed from (3): since the Bellman is a contraction mapping, V is unique; from (1) so are v_a and thus so is p.

Because $\pi_1 = A_1\pi_2 + b_1(p)$ from (3), the above equation simplifies to

$$\frac{\partial h_1(\pi_1)}{\partial \pi_1} A_1 - \widetilde{A}_1 \frac{\partial h_2(\pi_2)}{\partial \pi_2} = 0. \tag{13}$$

This equality depends on the (known) counterfactual transformation $\{h^s, h\}$ and on the data, F, through A_1 and $\widetilde{A}_1$. So, in practice, one only needs to verify whether (13) holds for the particular combination $\{h^s, h\}$ of interest.

Next, to facilitate the passage to the general case, rearrange the equality above in matrix form as follows:

$$\begin{bmatrix} \frac{\partial h_1(\pi_1)}{\partial \pi_1} & -\widetilde{A}_1 \frac{\partial h_2(\pi_2)}{\partial \pi_2} \end{bmatrix} \begin{bmatrix} A_1 \\ I \end{bmatrix} = 0$$

or

$$\begin{bmatrix} I & -\widetilde{A}_1 \end{bmatrix} \begin{bmatrix} \frac{\partial h_1(\pi)}{\partial \pi_1} & \frac{\partial h_1(\pi)}{\partial \pi_2} \\ \frac{\partial h_2(\pi)}{\partial \pi_1} & \frac{\partial h_2(\pi)}{\partial \pi_2} \end{bmatrix} \begin{bmatrix} A_1 \\ I \end{bmatrix} = 0,$$

where, in this example, $\frac{\partial h_1(\pi)}{\partial \pi_2} = \frac{\partial h_2(\pi)}{\partial \pi_1} = 0$. Using the property of the Kronecker product $vec(ABC) = (C' \otimes A)vec(B)$, our condition becomes:

$$\left(\begin{bmatrix} A_1' & I \end{bmatrix} \otimes \begin{bmatrix} I & -\widetilde{A}_1 \end{bmatrix} \right) vec\left(\nabla h(\pi)\right) = 0,$$

where $\nabla h(\pi)$ is the matrix with elements $\frac{\partial h_a(\pi)}{\partial \pi_j}$ for $a, j = 1, 2$. So, to identify the counterfactual CCPs, $vec(\nabla h(\pi))$ must lie in the nullspace of a matrix determined by A_1 and $\widetilde{A}_1$.

Now, moving from the binary to the general model, take (11) together with $\widetilde{\pi}_a = h_a(\pi)$ and stack all π_a for $a \neq J$ to obtain:

$$h_{-J}(\pi) = \widetilde{A}_{-J}\, h_J(\pi) + \widetilde{b}_{-J}(\widetilde{p}), \tag{14}$$

where $h_{-J}(\pi)$ stacks $h_a(\pi)$ for all $a \in \widetilde{\mathcal{A}}$ except for J, and the matrix $\widetilde{A}_{-J}$ and vector $\widetilde{b}_{-J}(\widetilde{p})$ are defined similarly. The next lemma guarantees that the implicit function theorem can be applied to (14).

Lemma 1. *The function $\widetilde{b}_{-J}(.)$ is continuously differentiable and its Jacobian is everywhere invertible.*

We state our main theorem below. $vecbr(C)$ rearranges the blocks of matrix C into a block column by stacking the block rows of C; the symbol $\boxtimes$ denotes the block Kronecker product.[16]

[16] The block Kronecker product, $\boxtimes$, of two partitioned matrices B and C is defined by (Koning, Neudecker, and Wansbeek, 1991):

$$B \boxtimes C = \begin{bmatrix} B \otimes C_{11} & \ldots & B \otimes C_{1b} \\ \vdots & \ddots & \vdots \\ B \otimes C_{c1} & \ldots & B \otimes C_{cb} \end{bmatrix}.$$

Theorem 1. *Consider the counterfactual transformation* $\{\widetilde{\mathcal{A}}, \widetilde{\mathcal{X}}, \widetilde{\beta}, \widetilde{G}, h^s, h\}$ *and suppose* h *is differentiable. The counterfactual conditional choice probabilities* $\widetilde{p}$ *are identified if and only if for all* π *satisfying (6),*

$$Q(A_{-J}, \widetilde{A}_{-J}) \times vecbr\left(\nabla h\left(\pi\right)\right) = 0, \tag{15}$$

where

$$Q(A_{-J}, \widetilde{A}_{-J}) = \left[\begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes I, \; - \begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes \widetilde{A}_{-J}\right].$$

The matrix $Q(A_{-J}, \widetilde{A}_{-J})$ *has dimension* $(\widetilde{A}-1)\widetilde{X}X \times (\widetilde{A}\widetilde{X})(AX)$, *while* $vecbr\left(\nabla h\left(\pi\right)\right)$ *has dimension* $(\widetilde{A}\widetilde{X})(AX) \times 1$.

Theorem 1 holds that counterfactual CCPs $\widetilde{p}$ are identified if and only if the Jacobian matrix of h is restricted to lie in the nullspace of a matrix defined by A_{-J} and $\widetilde{A}_{-J}$. So model primitives, data and counterfactual transformations have to interact with each other in a specific way to obtain identification of counterfactual CCPs.[17]

Equation (15) is the minimal set of sufficient conditions that applied researchers need to verify to secure identification of counterfactual behavior. For instance, for "action diagonal" counterfactuals, i.e. $\widetilde{\pi}_a = h_a(\pi_a)$, equation (15) is substantially simplified: $\widetilde{p}$ is identified if and only if for all π satisfying (6) and all $a \in \widetilde{\mathcal{A}}$, $a \neq J$,

$$\frac{\partial h_a}{\partial \pi_a} A_a - \widetilde{A}_a \frac{\partial h_J}{\partial \pi_J} = 0. \tag{16}$$

This implies that it is particularly difficult to identify counterfactual behavior when payoffs change non-linearly, since equation (16) must be satisfied for *all* admissible payoffs π.[18]

It is also worth noting that adding a known vector to $\widetilde{\pi}$ (e.g. $\widetilde{\pi} = h(\pi) + g$) does not affect the Jacobian matrix of h, and so whether $\widetilde{p}$ is identified or not does not depend on vector g. Similarly, changes to the distribution of the idiosyncratic shocks G do not affect whether equation (15) holds and so do not prevent the identification of $\widetilde{p}$.

Remark 2. *(Unobserved Heterogeneity.) Theorem 1 can be extended to incorporate unobserved heterogeneity. Following the discussion in Remark 1, we note that after finitely many type-specific conditional choice probabilities and transition functions are identified (Kasahara and Shimotsu, 2009; Hu and Shum, 2012), equation (15) can be verified for each unobserved type.*

Note that at the entry level, Kronecker rather than ordinary products are employed.

[17] Theorem 1 requires only that h is differentiable – this is a mild restriction typically satisfied in practice. Although the implicit function theorem involves local conditions, equation (15) must be satisfied for *all* payoffs that rationalize observed choice probabilities; i.e. for all π satisfying (6). Note also that the choice of the reference action J does not affect whether or not (15) is satisfied.

[18] One family of counterfactuals that satisfies (16) is a class of periodic functions satisfying $\frac{\partial h(By+c)}{\partial y} = \frac{\partial h(y)}{\partial y}$, for some matrix B and vector c.

Example: Rust's Bus Engine Replacement Problem. Rust (1987) investigates the optimal stopping problem of replacing a bus's engine, trading-off aging and replacement costs. The choice set is $\mathcal{A} = \{\text{replace, keep}\}$; the state variable, x, is the bus mileage which evolves stochastically and is renewed upon replacement; and the payoff function is

$$\pi(a,x) = \begin{cases} -\phi(x) - c(0), & if\ a = \text{replace} \\ -c(x), & if\ a = \text{keep} \end{cases}$$

where $\phi(x)$ is the cost of replacing an engine and $c(x)$ is the operating cost at mileage x. In principle, replacement costs may reflect labor costs of rebuilding an old engine, or the price of a new engine minus scrap values that may depend on the old engine's resale prices. In both cases replacement costs may potentially vary with the mileage on the (old) engine. To identify the model, Rust (1987) adopts an exclusion restriction (i.e. state-invariant replacement costs $\phi(x) = \phi$) and sets operating cost at $x = 0$ to zero (i.e. $c(0) = 0$). This is sufficient to identify payoffs.

In the counterfactual analysis, Rust varies the level of replacement costs and obtains the corresponding (long run) replacement choice probabilities, or a demand curve for engine replacement. One way to represent his counterfactual is to consider counterfactual payoffs as $\widetilde{\pi}(replace, x) = -(1+\lambda)\phi(x) - c(0)$, for various levels of λ, where λ is a parameter capturing the shift in replacement costs. Equivalently,

$$\widetilde{\pi}(replace, x) = \pi(replace, x) + \lambda(\pi(replace, x) - \pi(keep, 0)).$$

Each value of λ corresponds to one point along the demand curve for engine replacement. This representation is appropriate, for instance, if replacement costs depend on labor costs and the counterfactual of interest involves increasing wages. Note that the counterfactual does not affect $\pi(keep, x)$, nor β or F. In Appendix A, we show that Theorem 1 implies that this counterfactual is not identified. Showing this for a particular specification requires only a simple calculation evaluating equation (15).[19]

Interpreted this way, Rust's counterfactual falls within the class of affine payoff transformations, a class of counterfactuals for which equation (15) is simplified. As we show below, we can derive more intuitive conditions for the identification of such counterfactuals.

[19] Formally, $\widetilde{\pi} = \mathcal{H}\pi$, where $\mathcal{H}$ is not block-diagonal:

$$\mathcal{H} = \begin{bmatrix} (1+\lambda)I & [-\lambda 1, 0] \\ 0 & I \end{bmatrix},$$

where I is the identity matrix and 1 is a vector of ones. To evaluate equation (15), consider a simple version of the model in which $a = replace$, $J = keep$, and where the state space is simply $\mathcal{X} = \{new, old\}$ with deterministic transitions. Then $Q(A_{-J}, \widetilde{A}_{-J}) = [A_{replace} \otimes I,\ I \otimes I,\ A'_{replace} \otimes A_{replace},\ I \otimes A_{replace}]$. Finally, with $\beta = .99$ and $\lambda = 0.1$ (representing a 10% increase in replacement costs), we obtain $\|Q(A_{-J}, \widetilde{A}_{-J})vec(\nabla h(\pi))\| = 1.89$, where $\|.\|$ is the matrix 2-norm. Equation (15) is violated, implying the counterfactual is not identified.

3.3 Identification of Counterfactual Behavior: Special Cases

In this section, we discuss several special cases of interest following the taxonomy presented in Section 3.1. Corollary 1 shows how the conditions of Theorem 1 simplify when the payoff transformation $h(\cdot)$ is affine, i.e., $\widetilde{\pi} = \mathcal{H}\pi + g$, or $\widetilde{\pi}_a = \sum_{j\in\mathcal{A}} H_{aj}\pi_j + g_a$. The affine case is prevalent in applied work.

Corollary 1. *("Affine Payoff" Counterfactual) Assume $\widetilde{\pi} = \mathcal{H}\pi + g$.*
(i) The counterfactual CCP $\widetilde{p}$ is identified if and only if for all $a \in \widetilde{\mathcal{A}}$, $a \neq J$,

$$\sum_{l\in\mathcal{A}, l\neq J} \left(H_{al} - \widetilde{A}_a H_{Jl}\right) A_l + H_{aJ} - \widetilde{A}_a H_{JJ} = 0. \tag{17}$$

(ii) Further, if the counterfactual is "action diagonal," $\widetilde{\pi}_a = H_{aa}\pi_a + g_a$, equation (17) becomes, for all $a \in \widetilde{\mathcal{A}}$,

$$H_{aa}A_a - \widetilde{A}_a H_{JJ} = 0. \tag{18}$$

Recalling that $A_a = (I - \beta F_a)(I - \beta F_J)^{-1}$, and noting that the transition matrices F_a can be estimated, it is clear that conditions (17) and (18) can be easily verified from the data. The next set of results make direct use of Corollary 1.

Changes in Payoffs. We now consider counterfactuals that only change agents' payoff functions, holding fixed the remaining primitives. As already mentioned, previous work has shown that one particular case here yields identified counterfactual behavior: "pre-specified additive changes," which are of the form $\widetilde{\pi}(a,x) = \pi(a,x) + g(a,x)$ (Aguirregabiria, 2010; Aguirregabiria and Suzuki, 2014). Norets and Tang (2014) also proved identification when $\widetilde{\pi} = \lambda\pi + g$, where λ is a scalar. The identification of $\widetilde{p}$ for this class of counterfactuals is an immediate implication of Corollary 1. Note that in this case, we have $H_{aa} = \lambda I$ for all a, and so equation (18) is clearly satisfied.

Following the taxonomy, we now consider "proportional changes" counterfactuals. For this class, recall that we take $\widetilde{\pi}_a(x) = \lambda_a(x)\pi_a(x)$, or, compactly, $\widetilde{\pi}_a = H_{aa}\pi_a$, with H_{aa} diagonal.

Proposition 1. *("Proportional Changes") Consider $\widetilde{\pi}_a = H_{aa}\pi_a$ with H_{aa} diagonal. Assume $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{\beta} = \beta$, $\widetilde{G} = G$, and $\widetilde{F} = F$. Then, to identify $\widetilde{p}$ it is necessary that $H_{aa} = H_{jj}$, for all $a, j \in \mathcal{A}$.*

Assume the matrices H_{aa} are identical for all a, and denote them by H. Suppose further that H has d distinct diagonal elements $\lambda_1, ..., \lambda_d$, each occurring $n_1, ..., n_d$ times so that H can be written as $H = diag(\lambda_1 I_{n_1}, ..., \lambda_d I_{n_d})$. The following statements are equivalent:
(i) $\widetilde{p}$ is identified.
(ii) A_a is block diagonal with diagonal blocks, $(A_a)_i$, of comformable sizes $n_1, ..., n_d$.
(iii) Let $(F_a)_{ij}$ be the $n_i \times n_j$ submatrix of F_a that comforms with $H = diag(\lambda_1 I_{n_1}, ..., \lambda_d I_{n_d})$. For

all $a \in \mathcal{A}$ and $i \neq j$, the block partitions of F_a and F_J satisfy

$$(F_a)_{ij} = (A_a)_i \, (F_J)_{ij} \, , \tag{19}$$

where $(A_a)_i \equiv (I - \beta(F_a)_{ii})(I - \beta(F_J)_{ii})^{-1}$.

The necessary conditions to identify $\widetilde{p}$ in the case of "proportional changes" are restrictive. For one, if we change the payoff of action a in state x by $\lambda(x)$, $\widetilde{\pi}(a, x) = \lambda(x) \pi(a, x)$, then it is necessary to change the payoff of any other action a in state x by exactly the same proportion $\lambda(x)$. Furthermore, identification requires special conditions on the A_a matrices (part ii), which are equivalent to special conditions on the transition process F (part iii). In particular, an implication of the proposition is that, if all diagonal elements of H are pairwise distinct, then identification of $\widetilde{p}$ requires $F_a = F_J$ for all $a \in \mathcal{A}$. This condition however will not be satisfied in any dynamic model of interest.[20]

Another set of payoff counterfactuals is the "changes in types." We prove the following proposition for two types and two actions for notational simplicity; the extension to multiple types and actions is straightforward.

Proposition 2. *("Changes in Types") Suppose the payoff is $\pi_a(x, s)$, where s is a time-invariant state (type) that takes two values, $s \in \{s_1, s_2\}$, and that there are two actions $\mathcal{A} = \{a, J\}$. Suppose also that $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{\beta} = \beta$, $\widetilde{G} = G$, and $\widetilde{F} = F$.*
(i) If the counterfactual replaces the payoff of type s_1 by that of s_2 for one of the actions, then $\widetilde{p}$ is not identified.
(ii) If the counterfactual replaces the payoff s_1 by that of s_2 for all actions, then $\widetilde{p}$ is identified if and only if $(I - \beta F_a^{s_1})(I - \beta F_J^{s_1})^{-1} = (I - \beta F_a^{s_2})(I - \beta F_J^{s_2})^{-1}$, where F_a^s is the transition matrix corresponding to type s.

Proposition 2 results in nonidentification if payoffs of a subset of actions are replaced; and requires strong restrictions on transition probabilities if payoffs at all actions are replaced (much like the "proportional changes" case). It is worth noting that if the types have the same transitions ($F_a^{s_1} = F_a^{s_2}$), the condition is satisfied and the counterfactual is identified.

Proposition 1 and 2 express specific and verifiable restrictions on the transition process that must hold for the identification of counterfactual behavior. A natural question then is whether some transformations of payoffs – and which – can be said to be identified for *any* transition process. In other words, when can we say a counterfactual transformation of payoffs is identified without having to estimate or specify the specific transition process? In such a case, the researcher can establish identification ex ante, regardless of the data at hand. Our next result holds that only a limited set of payoff transformations can be said to be identified without verifying the restrictions on the transition process.

[20]In general, if a diagonal element of H is unique, i.e. $n_i = 1$ for some i, identification requires that the i-row of F_a and F_J are identical. That is, conditional on state x_i, the transition probabilities do not depend on the action.

Proposition 3. *Assume* $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{\beta} = \beta$, $\widetilde{G} = G$, $\widetilde{F} = F$, *and* $\widetilde{\pi} = \mathcal{H}\pi + g$. *Counterfactual behavior* $\widetilde{p}$ *is identified without restrictions on the transition process* F *(i.e., counterfactual behavior is identified for every transition* F*) if and only if*
(i) $\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2$, *where* $\mathcal{H}_1 = \lambda I$, λ *is a scalar, and* $\mathcal{H}_2$ *has identical rows;*
(ii) or, equivalently, the transformation of payoffs can be expressed in the following form for all $a \in \mathcal{A}$ *and* $x \in \mathcal{X}$*:*

$$\widetilde{\pi}_a(x) = \lambda \pi_a(x) + L(\pi) + g_a(x), \tag{20}$$

where $L(\cdot)$ *is the scalar-valued function given by* $L(\pi) = \sum_{j \in \mathcal{A}} \sum_{x \in \mathcal{X}} \rho_{jx} \pi_j(x)$, *and the vector* $[\rho_{11}, ..., \rho_{AX}]$ *corresponds to one row of* $\mathcal{H}_2$.[21]

Equation (20) shows that only the following changes are identified regardless of the transition process:

1. Pre-specified additive changes $g_a(x)$ which may depend arbitrarily on actions and states but does not depend on the baseline payoff function.

2. Multiplication of baseline payoffs by a scalar λ, which does not depend on actions or states. For $\lambda > 0$, this resembles a *change in the scale of the payoff function.*[22]

3. Addition of a scalar-valued function $L(\pi)$, which does not depend on actions or states. This corresponds to a *change in the level of the payoff function.*[23]

Proposition 3 states that the only meaningful counterfactual transformations of payoffs that can be said to be identified with no restrictions on the state transition process are (1) pre-specified additive changes and (2) changes in the level and scale of the payoff function. As described above, different versions of the "if" direction of Proposition 3 have been proved in the literature. Our result shows that *any other* counterfactual transformations of payoffs are identified only under restrictions on the state transition processes that require verification in the data. Theorem 1 provides the conditions the transition process must satisfy most generally; Corollary 1 and Propositions 1 and 2 provide simpler conditions for certain classes of payoff transformations.

Changes in Choice Sets and State Space. Following the taxonomy, we now consider a counterfactual that adds an option to agent's choice set. This counterfactual naturally requires pre-specifying $\widetilde{\pi}_j$ and $\widetilde{F}_j$ for the new choice j (as well as the joint distribution of the idiosyncratic shocks $\widetilde{G}$).

[21]To connect parts (i) and (ii), note that $\widetilde{\pi} = \mathcal{H}\pi + g = \mathcal{H}_1\pi + \mathcal{H}_2\pi + g$, and that $\mathcal{H}_2\pi$ is a constant vector because $\mathcal{H}_2$ has identical rows. So, $\mathcal{H}_2\pi$ corresponds to $L(\pi)\,1$, where 1 is a vector of ones.

[22]Strictly speaking, multiplication of π by a positive scalar is *not* equivalent to a scale normalization of the whole utility function, $\pi + \varepsilon$, because the distribution of the idiosyncratic shocks ε is fixed. More formally, $\widetilde{\pi} = \lambda\pi$ for $\lambda > 0$ is equivalent to multiplying the variance of the idiosyncratic shocks by λ^{-2}.

[23]Note that a shift in the level of payoffs by the same number L for every action and state does not affect agents' incentives.

Proposition 4. *("Add an Action" Counterfactual) Suppose* $\widetilde{\mathcal{A}} = \mathcal{A} \cup \{j\}$*, where* j *is the new action. Assume* $\widetilde{\mathcal{X}} = \mathcal{X}$*,* $\widetilde{\beta} = \beta$*,* $\widetilde{F} = F$*,* $\widetilde{\pi}_a = \pi_a$ *for all* $a \in \mathcal{A}$*, and*

$$\widetilde{\pi}_j = \sum_{a \in \mathcal{A}} H_{ja} \pi_a + g_j.$$

Let 1 *be an* $X \times 1$ *vector of ones. Then* $\widetilde{p}$ *is identified if and only if* $\sum_{a \in \mathcal{A}} H_{ja} 1 = 1$*, and*

$$\widetilde{F}_j = \sum_{a \in \mathcal{A}} H_{ja} F_a + \beta^{-1} \left(I - \sum_{a \in \mathcal{A}} H_{ja} \right).$$

In words, to obtain identification it is necessary that the payoff of the new action j is a "convex combination" of existing payoffs, and the new transition matrix is an "affine" combination of existing transitions. This is reminiscent of predicting consumers' choices when a new good is introduced in a static differentiated product demand framework. Predicting the demand for the new good requires that the attributes of the new good are a combination of the attributes of existing goods in the market. The same applies in the dynamic context; here we additionally need restrictions on the transitions in order to predict behavior when a new choice is available.

Consider next a counterfactual that eliminates one action (extensions to eliminating more actions are straightforward).

Proposition 5. *("Eliminate an Action" Counterfactual) Suppose* $\widetilde{\mathcal{A}} = \mathcal{A} - \{j\}$*, where* j *is the action to be eliminated. If* $\widetilde{\mathcal{X}} = \mathcal{X}$*,* $\widetilde{\beta} = \beta$*,* $\widetilde{F}_a = F_a$*, and* $\widetilde{\pi}_a = \pi_a$ *for all* $a \in \widetilde{\mathcal{A}}$*, then* $\widetilde{p}$ *is identified.*

Here, the key to identification is that transitions do not change.[24] However, elimination of an action often implies elimination of some states as well (e.g. when $x_t = a_{t-1}$ as in many entry models), which necessarily changes transitions. In that case, identification depends on how the probability mass is reallocated from $\mathcal{X}$ into the remainder set of states $\widetilde{\mathcal{X}}$. Lemma A2 in Appendix A provides the necessary and sufficient conditions for identification in this case. Below, we consider a special case that is common in applied work. Decompose the state variables as $x = (k, w)$, where $k_t = a_{t-1}$, and w is an exogenous state (i.e., its evolution does not depend on choices a). Formally, $F_a = F^w \otimes F_a^k$, where F^w is the transition matrix for w, F_a^k is the transition for k, and $\otimes$ denotes the Kronecker product. The firm entry/exit problem presented in Section 4 satisfies these restrictions. The counterfactual CCP is indeed identified in this case.

Proposition 6. *("Eliminate an Action and States" Counterfactual) Suppose* $\widetilde{\mathcal{A}} = \mathcal{A} - \{j\}$*, where* j *is the action to be eliminated. Without loss of generality, let the set of states be* $\widetilde{\mathcal{X}} = \{1, ..., \overline{x}\}$ *and* $\mathcal{X} = \{1, ..., \overline{x}, \overline{x}+1, ..., X\}$*. Assume* $\widetilde{\pi}_a = H_{aa} \pi_a$ *with* $H_{aa} = [I_{\overline{x}}, 0]$ *for all* $a \in \widetilde{\mathcal{A}}$*, where* $I_{\overline{x}}$ *is the* $\overline{x} \times \overline{x}$ *identity matrix. Suppose* $x = (w, k)$ *with transition matrix* $F_a = F^w \otimes F_a^k$ *and* $k_t = a_{t-1}$*. Then, the counterfactual CCP* $\widetilde{p}$ *is identified.*

[24] Note that eliminating action j is not equivalent to a pre-specified additive change with $g_j = -\infty$ because the Blackwell sufficient conditions for a contraction are not satisfied in the corresponding Bellman equation.

Changes in Transitions. For completeness, we mention briefly another existing result, first proven by Aguirregabiria and Suzuki (2014) and Norets and Tang (2014), regarding counterfactuals that only change the state transitions. Specifically, when the only primitive that changes in the counterfactual is the transition process, $\widetilde{F} \neq F$, then counterfactual behavior is identified only under restrictive conditions on these transitions. In our notation, the necessary and sufficient condition is $A_a = \widetilde{A}_a$, for all $a \in \widetilde{\mathcal{A}}$, $a \neq J$.[25] In Section 3.5, we discuss how (correctly specified) parametric assumptions on the payoff function can lead to less restrictive requirements on transitions for the identification of this class of counterfactuals.

3.4 Some Intuition

Intuitively, the reason payoffs are not identified in a DDC model is twofold. First, although we can identify the difference in continuation values $v_a - v_J$ from the observed choice probabilities (Hotz and Miller, 1993), the discrete choice nature of the data does not allow us to separate v_a from v_J. (This feature is shared by static discrete choice models as well.) Second, we also cannot separate the two components of $v_a(x)$ nonparametrically, i.e. $\pi(a, x)$ and $E[V(x')|a, x]$, since they both depend on the same arguments.

To obtain some intuition for why some counterfactuals are identified while others are not, take a simple example of a binary choice, $\mathcal{A} = \{a, J\}$ and consider the counterfactual $\widetilde{\pi}_a = H_{aa}\pi_a + g_a$, for all a, with $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{\beta} = \beta$, and $\widetilde{G} = G$. Next, note that by rearranging our main equation (3), we obtain

$$(I - \beta F_a)^{-1} \pi_a - (I - \beta F_J)^{-1} \pi_J = (I - \beta F_a)^{-1} b_a(p).$$

The left-hand side is the difference of the expected discounted present value obtained by always choosing a versus always choosing J. This difference is known, as the right-hand side is known.

For the counterfactual scenario, rearrange the equivalent counterfactual equation (11) as above. Assuming the counterfactual is identified (i.e., $H_{aa}A_a = \widetilde{A}_a H_{JJ}$), it is easy to show that

$$H_{aa}(I - \beta F_a)\left[(I - \beta F_a)^{-1} \pi_a - (I - \beta F_J)^{-1} \pi_J\right] + (g_a - \widetilde{A}_a g_J) = \widetilde{b}_a(\widetilde{p}).$$

In words, the counterfactual CCP $\widetilde{p}$ depends on π *only through* the difference in present values: $(I - \beta F_a)^{-1} \pi_a - (I - \beta F_J)^{-1} \pi_J$. All other terms of the left-hand side, as well as the function $\widetilde{b}_a$, are known. So, to calculate $\widetilde{p}$, it is not necessary to identify the payoff function π.

To make the argument more transparent, consider the "additive changes" counterfactual. The left-hand side now becomes the sum of two terms: $(I - \beta F_a)^{-1} \pi_a - (I - \beta F_J)^{-1} \pi_J$ and

[25] That is an implication of equation (18) in Corollary 1. Note that identification of a counterfactual that only changes the discount factor, $\widetilde{\beta} \neq \beta$, requires the same condition: $A_a = \widetilde{A}_a$, for all $a \neq J$, but with $\widetilde{A}_a = (I - \widetilde{\beta} F_a)(I - \widetilde{\beta} F_J)^{-1}$.

$(I-\beta F_a)^{-1} g_a - (I-\beta F_J)^{-1} g_J$. Both terms are known and thus the change in the choice probabilities is also known. There is no need to re-optimize agents' dynamic behavior in the counterfactual scenario. This is possible only because the counterfactual difference $(I-\widetilde{\beta}\widetilde{F}_a)^{-1}\widetilde{\pi}_a - (I-\widetilde{\beta}\widetilde{F}_J)^{-1}\widetilde{\pi}_J$ is a *known* function of the *observed* difference $(I-\beta F_a)^{-1}\pi_a - (I-\beta F_J)^{-1}\pi_J$.

When the equality $H_{aa}A_a = \widetilde{A}_a H_{JJ}$ is not satisfied, the counterfactual difference in continuation values depends *directly* on π. It is no longer sufficient to know the observed differences $(I-\beta F_a)^{-1}\pi_a - (I-\beta F_J)^{-1}\pi_J$ and therefore we cannot identify the counterfactual behavior. At least not without additional restrictions.

Before investigating the role of additional restrictions in the next section, we discuss briefly some implications for static discrete choice models and dynamic models with continuous choices.

Remark 3. *(Static vs Dynamic Models.) Our framework can be used to understand which counterfactuals are identified in a static setting, and how that compares to dynamic problems. To be precise, a static model can be characterized by either agents being myopic (i.e., $\beta = 0$), or by state transitions not affected by agents' choices (i.e., $F_a = F_J$, for all a), or both. In such cases, it is clear from the discussion above that counterfactual behavior is identified when it depends on π only through the difference $\pi_a - \pi_J$. This is intuitive as static models can only recover differences in flow payoffs from the data. Counterfactual behavior is not identified when it depends directly on π, i.e., when payoff* levels *matter.*

An alternative way to see this is to note that when $\beta = 0$ or $F_a = F_J$ hold for all a, both in the baseline and in the counterfactual scenarios, then $A_a = \widetilde{A}_a = I$, for all a. This implies that condition (18) in Corollary 1 simplifies to $H_{aa} = H_{JJ}$, for all a. This in turn means that the "additive changes," the "eliminating actions," and the "changes in transitions" counterfactuals are all identified: equation (18) is satisfied trivially in all these cases (since $H_{aa} = I$ for all a). On the other hand, "proportional changes" and "adding actions" are counterfactuals that are not trivially identified. Much like in dynamic models, "proportional changes" are not identified when we multiply different payoffs by different amounts; "adding actions" is not identified when the payoff of the new action is not a convex combination of the payoffs of existing choices. That is because payoff levels matter for both cases. In contrast to dynamic models though, while these two counterfactuals require restrictions in how payoffs are changed, they do not demand restrictions on state transitions for identification.[26] In sum, not all counterfactuals of interest are identified in static models, but they demand fewer restrictions for identification when compared to dynamic models and so a larger number of cases are identified in static models.

Remark 4. *(Discrete vs Continuous Choices.) While we do not investigate formally the identification of counterfactuals for dynamic models with continuous choices, our approach suggests that the same issues apply to that class of models. Intuitively, just like in a discrete choice setting,*

[26] In static models, "proportional changes" satisfy condition (ii) in Proposition 1 trivially, and "adding actions" does not require the new state transition to be a combination of existing state transitions (see Proposition 4).

where the econometrician can only recover differences in payoffs, in a continuous choice setting, we can only recover the derivative of payoffs. Consistent with this, Blevins (2014) has shown that nonparametric identification of dynamic models with continuous actions and states requires location restrictions on flow payoffs (in the same manner as in the class of nonseparable models studied by Matzkin (2003)). When cardinal properties of the payoff function matter for the counterfactual, location restrictions necessary to identify the model may prevent the identification of counterfactual behavior, even when actions and states are continuous.[27]

3.5 Identification of Counterfactual Behavior Under Parametric Restrictions

More often than not, applied work relies on parametric restrictions. We thus consider identification of counterfactuals under a specific parametric model that encompasses many applied models in the literature.

We decompose the state space into two components, $x = (k, w)$, where $k \in \mathcal{K} = \{1, ..., K\}$ are controlled states (i.e. their evolution is affected by agents' choices), and $w \in \mathcal{W} = \{1, ..., W\}$ are exogenous (e.g. market-level states), with K and W finite. Therefore,

$$F\left(x'|a, x\right) = F^{k}\left(k'|a, k\right) F^{w}\left(w'|w\right). \tag{21}$$

and the transition matrix F_a is written as $F_a = F^w \otimes F_a^k$, where $\otimes$ denotes the Kronecker product. In addition, we assume the following parametric payoff is true:

$$\pi(a, k, w) = \theta_0(a, k) + Z(a, w)'\theta_1(a, k), \tag{22}$$

where $Z\left(a, w\right)$ is a known function of actions and states w (e.g. observed measures of variable profits or returns) and $\theta_0\left(a, k\right)$ is interpreted as a fixed cost component. For instance, in the firm entry/exit problem considered in Section 4.1, $Z(a, w)'\theta_1(a, k)$ represents variable profits which may be either directly observed or a flexible function of observables such as market size and input prices, while $\theta_0(a, k)$ denotes entry/exit/fixed cost depending on the action and state.

Proposition 7 provides sufficient conditions for the identification of this parametric model. For notational simplicity, we focus on a binary choice with $\mathcal{A} = \{a, J\}$ and assume $Z\left(a, w\right)$ is scalar. The proposition also holds in the more general case of $F^w\left(w'|w, a\right)$ and multivariate $Z\left(a, w\right)$.

[27] Furthermore, Blevins (2014) shows in his Theorem 2 that π is nonparametrically identified only in the region in which unobservable shocks ε are compatible with optimal choices. The fact that π is not identified nonparametrically for all possible realizations of ε limits possible extrapolations and poses additional difficulties to counterfactual identification. Identifying π everywhere requires further restrictions on payoffs (see, e.g., his Theorem 3).

Proposition 7. *Assume (21) and (22) hold. Let*

$$\begin{aligned} D_a &= (I - \beta F_a)^{-1}, \\ Z_a &= [Z_a(1)I_k, ..., Z_a(W)I_k]', \end{aligned}$$

and similarly for D_J and Z_J. I_k is the identity matrix of size K and $e'_w = [0, 0, ..., I_k, 0, ...0]$ with I_k in the w position. Suppose $W \geq 3$ and there exist $w, \widetilde{w}, \overline{w}$ such that the matrix

$$\begin{bmatrix} (e'_w - e'_{\widetilde{w}}) D_a Z_a & (e'_{\widetilde{w}} - e'_w) D_J Z_J \\ (e'_w - e'_{\overline{w}}) D_a Z_a & (e'_{\overline{w}} - e'_w) D_J Z_J \end{bmatrix} \tag{23}$$

is invertible. Then the true parameters $[\theta_1(a,k), \theta_1(J,k)]$ are identified, but $[\theta_0(a,k), \theta_0(J,k)]$ are not identified.

Intuitively, the "slope" coefficients θ_1 are identified provided there is "sufficient variation" in w (guaranteed by the invertibility of matrix (23)). This requires w to significantly change the conditional expected values of Z_a and Z_J (naturally, it is necessary that $Z_a \neq Z_J$).[28] The "intercept" parameters θ_0, however, are not identified. To identify this vector, we have to add K linearly independent restrictions to the model, much as we have to impose X linearly independent restrictions in the nonparametric setting.

Counterfactuals. In addition to the counterfactuals discussed in Section 3.1, one may be interested in changes in either θ_0, i.e.:

$$\widetilde{\pi}(a,k,w) = h_0[\theta_0(a,k)] + Z(a,w)'\theta_1(a,k); \tag{24}$$

or in $Z'\theta_1$, i.e.:

$$\widetilde{\pi}(a,k,w) = \theta_0(a,k) + h_1\left[Z(a,w)'\theta_1(a,k)\right]. \tag{25}$$

These counterfactuals allow for changes in how the flow payoff responds to some state (k,w), or to the outcome variable Z.

We show that transformations in $Z'\theta_1$ result in identified counterfactuals, while transformations in $\theta_0(a,k)$ may not. Indeed, since $\theta_1(a,k)$ is identified, a counterfactual that changes $Z'\theta_1$ resembles a "pre-specified additive change," which is an identified counterfactual (see Section 3.3). In contrast, since $\theta_0(a,k)$ is not identified, one needs to follow our analysis of counterfactuals for nonparametric payoffs to establish whether a particular counterfactual is identified or not.

To illustrate, we consider affine action-diagonal counterfactuals as an example. In particular, let:

$$\widetilde{\theta}_0(a) = H_0(a)\theta_0(a) \tag{26}$$

[28] Note that $e'_w D_a Z_a$ is the expected discounted value of Z when action a is always chosen conditional on observing state w today.

for $a = 1, ..., J$, $\theta_0(a)$ is obtained by stacking $\theta_0(a, k)$ for all k, and $H_0(a)$ is a $K \times K$ matrix. Extending to a more general (differentiable) function of θ_0 is straightforward.

Proposition 8. *(Parametric Model) Assume $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{\beta} = \beta$, and that the conditions of Proposition 7 hold.*

(i) $\widetilde{\mathrm{p}}$ is identified when the counterfactual only changes the term $Z(a, w)'\theta_1(a, k)$ of $\pi(a, k, w)$ as in (25).

(ii) $\widetilde{\mathrm{p}}$ is identified under the affine action diagonal counterfactual (26) if and only if for all $a \neq J$

$$H_0(a)A_a^k - A_a^k H_0(J) = 0$$

where $A_a^k = (I - \beta F_a^k)(I - \beta F_J^k)^{-1}$.

(iii) $\widetilde{\mathrm{p}}$ is identified when the counterfactual only changes the transition F_a^k if and only if $A_a^k = \widetilde{A}_a^k$, $a \neq J$, where $\widetilde{A}_a^k = (I - \beta \widetilde{F}_a^k)(I - \beta \widetilde{F}_J^k)^{-1}$.

(iv) $\widetilde{\mathrm{p}}$ is identified when the counterfactual changes the transition F^w.

In a nonparametric setting, changes in the transition process generically result in non-identified counterfactual behavior (in the sense that the necessary conditions are bound to be restrictive). However, Proposition 8 shows that the intuition from the nonparametric setting does not necessarily carry over to parametric models. When a counterfactual changes the transition process for state variables that are part of the identified component of the payoff function, counterfactual behavior is identified. For instance, the response to a change in the volatility of demand shocks in the firm entry/exit example is identified. Even though Aguirregabiria and Suzuki (2014) and Norets and Tang (2014) have explored changes in transitions in the nonparametric context, most implementations of these counterfactuals in practice are done in the parametric context (Hendel and Nevo, 2006; Collard-Wexler, 2013) and so based on our results, are in fact identified if the parametric model is correctly specified.

3.6 Identification of Counterfactual Welfare

In this section, we discuss the identification of counterfactual welfare and provide the minimal set of sufficient conditions for identification. For simplicity, we only consider affine action-diagonal counterfactuals; i.e. $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{\beta} = \beta$, and $\widetilde{\pi}_a = H_{aa}\pi_a + g_a$, all a. Extensions to more general cases are straightforward, but at the cost of substantially more cumbersome notation. The feature of interest here is the value function difference $\Delta V = \widetilde{V} - V$, where $\widetilde{V}$ is the counterfactual value function.

Proposition 9. *(Welfare) Assume $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{\beta} = \beta$, and $\widetilde{\pi}_a = H_{aa}\pi_a + g_a$, for all a. The welfare difference ΔV is identified if, for all $a \neq J$,*

$$H_{aa}A_a - \widetilde{A}_a H_{JJ} = 0,$$

and

$$H_{JJ} = (I - \beta \widetilde{F}_J)(I - \beta F_J)^{-1}. \tag{27}$$

Proposition 9 shows that identification of $\widetilde{p}$ (which is implied by the proposition's first condition) is not sufficient to identify ΔV; we also need (27). The second condition is satisfied, for instance, when the counterfactual transformation does not affect option J: $H_{JJ} = I$ and $\widetilde{F}_J = F_J$. For "proportional changes" counterfactuals the two conditions are satisfied only when all matrices H_{aa} equal the identity matrix; i.e. $\widetilde{\pi} = \pi$, which is equivalent to saying that ΔV is not identified. On a positive note, an immediate implication of Proposition 9 is that the welfare impact of "additive changes" is identified. Therefore, "additive changes" are robust to nonidentification of the model primitives: both $\widetilde{p}$ and ΔV are identified.[29]

Finally, the next corollary considers identification of ΔV for the parametric model of Section 3.5. As expected, identification is guaranteed when counterfactuals change $Z'\theta_1$ and/or F^w.

Corollary 2. *(Welfare, Parametric Model) Assume the conditions of Proposition 7 hold. Suppose* $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{\beta} = \beta$, *and* $\widetilde{\theta}_0(a) = H_0(a)\theta_0(a)$. *The welfare difference* ΔV *is identified if, for all* $a \neq J$,

$$H_0(a)A_a^k - \widetilde{A}_a^k H_0(J) = 0,$$

and

$$H_0(J) = (I - \beta \widetilde{F}_J^k)(I - \beta F_J^k)^{-1}.$$

Furthermore, if $H_0(a) = I$ *and* $\widetilde{F}_a^k = F_a^k$ *for all* a, *then* ΔV *is identified for any counterfactual transformation on* $Z(a, w)'\theta_1(a, k)$ *and* F^w.

4 Applied Examples

4.1 Numerical Example: Firm Entry and Exit Problem

This section illustrates some of our theoretical results using a firm entry/exit problem; the model adopts the parameterization of Section 3.5 and has been commonly used in the literature (Das, Roberts, and Tybout, 2007; Aguirregabiria and Suzuki, 2014; Lin, 2015; Igami, 2017; Varela, 2018). Consider a firm deciding between two actions: whether to be active ($a = 1$) or inactive ($a = 0$) in a market, so that $\mathcal{A} = \{0, 1\}$. Let the state variables be $x = (k, w)$, where k is the lagged chosen action, and w is an exogenous shock determining variable operating profits, $\overline{\pi}(w)$. The flow payoff is

$$\pi(a, k, w) = \begin{cases} k \times sv & if\ a = 0 \text{ (inactive)} \\ k(\overline{\pi}(w) - fc) - (1 - k)\, ec, & if\ a = 1 \text{ (active)} \end{cases}$$

[29] Proposition 9 is an immediate consequence of Lemma A5 in the Appendix, which provides the full set of necessary and sufficient conditions to identify ΔV.

where sv is the scrap value, fc is the fixed cost, and ec is the entry cost.[30] Note that the payoff when $a = 0$ and $k = 0$ (i.e. when the firm was and remains inactive) is set equal to zero.

We assume that w follows a first-order Markov process, and can take three values: high, medium or low, $w \in \{w^H, w^M, w^L\}$. Variable profits $\overline{\pi}(w)$ are determined by static profit maximization. We assume the econometrician knows (or estimates): (a) the true CCP, $\Pr(active|k, w)$; (b) the transition $\Pr(w_{t+1}|w_t)$; and (c) the variable profits $\overline{\pi}(w)$, which can be recovered "offline," using price and quantity data.[31]

First, we solve the true model, obtain the baseline CCP, and recover π. Typically, researchers identify the model by setting either $sv = 0$ or $fc = 0$. As previously discussed, there is little guidance to justify these assumptions because cost or scrap value data are extremely rare (e.g., see Kalouptsidi, 2014). We estimate the model twice to compare the different sets of restrictions. Under the first restriction that scrap values are equal to zero ($sv = 0$), identification of π follows directly from equation (3), $\pi_a = A_a \pi_J + b_a(p)$: in this case, note that the payoff π_J for $J = inactive$ is equal to zero for all states (k, w) and thus (3) directly delivers π_a for $a = active$. Under the alternative restriction that fixed costs are equal to zero ($fc = 0$), we can recover the remaining elements of π by adding $\pi(active, 1, w) = \overline{\pi}(w)$ to the system described by (3).

Table 1 presents the true and the two estimated payoff functions. Note that under the first restriction ($sv = 0$) the estimated entry costs have the wrong sign. This is because if there is no scrap value gained upon exiting, entering the market becomes less attractive and entry costs must become low (in fact, negative) to explain the observed entry patterns. Under the second restriction ($fc = 0$), both the estimated entry costs and scrap values are considerably larger than their true values. To see why, consider an active firm ($k = 1$). Fixing $fc = 0$ implies higher profits when active, which gives incentives to stay more often in the market. Therefore, the estimated scrap value must increase in order to provide incentives to exit and match the observed exit rate. Similarly, when the firm is out ($k = 0$), increasing profits when active provides incentives to enter. Entry costs must then increase to compensate for this incentive and explain the observed entry rate.

Next, given the recovered payoffs, we implement four counterfactuals and compare the true

[30] The model falls within the parametric framework of Section 3.5, with $\theta_0(0,0) = 0$, $\theta_0(1,0) = -ec$, $\theta_0(0,1) = sv$, and $\theta_0(1,1) = -fc$. Variable profits, $\overline{\pi}(w)$, are assumed known/estimated outside of the dynamic problem using price and quantity data, one may then take $Z(a, w) = \overline{\pi}(w)$ and $\theta_1(a, k) = 1$. Alternatively, one might assume a reduced form profit function $\overline{\pi}(w) = w'\gamma$, in which $Z(a, w) = w$ and $\theta_1(a, k) = \gamma$, with γ identified under sufficient variation on w.

[31] We assume the firm faces the (residual) inverse demand curve $P_t = w_t - \eta Q_t$ and has constant marginal cost c, so that $\overline{\pi}(w_t; \eta, c) = (w_t - c)^2 / 4\eta$. The idiosyncratic shocks ε_{it} follow a type 1 extreme value distribution. We ignore sampling variation for simplicity and set: $c = 11, \eta = 1.5, w = (20, 17, 12), \beta = 0.95, fc = 5.5, sv = 10, ec = 9$, while the transition matrix for w is

$$F(w_{t+1}|w_t) = \begin{bmatrix} 0.4 & 0.35 & 0.25 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}.$$

Table 1: Numerical Example – True vs. Estimated Profits

States: (k, w)	**True Profit**	**Estimated Profit** *scrap value = 0*	**Estimated Profit** *fixed cost = 0*
$a =$ inactive			
$\pi(a, k = 0, w_H) = 0$	0	0	0
$\pi(a, k = 0, w_M) = 0$	0	0	0
$\pi(a, k = 0, w_L) = 0$	0	0	0
$\pi(a, k = 1, w_H) = sv$	10	0	120
$\pi(a, k = 1, w_M) = sv$	10	0	120
$\pi(a, k = 1, w_L) = sv$	10	0	120
$a =$ active			
$\pi(a, k = 0, w_H) = -ec$	–9	0.5	–113.5
$\pi(a, k = 0, w_M) = -ec$	–9	0.5	–113.5
$\pi(a, k = 0, w_L) = -ec$	–9	0.5	–113.5
$\pi(a, k = 1, w_H) = \overline{\pi}(w_H) - fc$	8	7.5	13.5
$\pi(a, k = 1, w_M) = \overline{\pi}(w_M) - fc$	0.5	0	6
$\pi(a, k = 1, w_L) = \overline{\pi}(w_L) - fc$	–5.33	–5.83	0.167

and the inferred counterfactual CCPs and welfare. In the first two, the government provides subsidies to encourage entry. Counterfactual 1 is an additive subsidy that reduces entry costs: $\widetilde{\pi}(active, 0, w) = \pi(active, 0, w) + g_1$. Counterfactual 2 is a proportional subsidy: $\widetilde{\pi}(active, 0, w) = H_1\pi(active, 0, w)$.[32] As shown in Section 3, while the counterfactual CCPs and welfare are identified in the first case, they are not identified in the second case. Specifically, counterfactual behavior corresponding to "additive changes" (the first case) depends on the difference in present values of always choosing $a = 0$ versus $a = 1$ (i.e., $(I - \beta F_1)^{-1}\pi_1 - (I - \beta F_0)^{-1}\pi_0$) *plus* the present value of the subsidies (i.e., $(I - \beta F_1)^{-1}g_1$). The difference in present values is itself identified from the data, and the present value of the subsidies is known by the researcher (since g_1 is pre-specified), implying identification of the counterfactual CCPs (see Section 3.4). In contrast, "proportional changes" (the second case) require knowledge of baseline payoffs in levels, except in special cases unlikely to be satisfied in practice. Indeed, the current example does not satisfy these special conditions and hence counterfactual CCP and welfare are not identified here (see Propositions 1 and 9).

Table 2 presents the results from counterfactuals 1 and 2 for the true model and the two estimated models. In both counterfactuals, the *true* counterfactual probability of entering increases because of the subsidy; and the probability of staying in the market decreases because it is cheaper

[32] We choose the additive and proportional subsides so that the true counterfactual CCP and welfare are the same. As $\pi(active, 0, w) = -ec$, and the true $ec = 9$, we set $g_1 = 0.9$ and $H_1 = 0.9$, so that in both cases the true counterfactual entry cost becomes $\widetilde{\pi}(active, 0, w) = -8.1$.

Table 2: Counterfactuals 1 and 2 – Additive and Proportional Entry Subsidies

States: (k, w)	**Baseline**	**True CF**	**Estimated CF** *scrap value = 0*	**Estimated CF** *fixed cost = 0*
			CF1: $\widetilde{\pi}_0 = \pi_0, \widetilde{\pi}_1 = \pi_1 + g_1$	
CCP: $\Pr(active \vert x)$				
$(k = 0, w_H)$	93.61%	94.95%	94.95%	94.95%
$(k = 0, w_M)$	87.48%	90.27%	90.27%	90.27%
$(k = 0, w_L)$	72.99%	80.33%	80.33%	80.33%
$(k = 1, w_H)$	99.99%	99.99%	99.99%	99.99%
$(k = 1, w_M)$	80.91%	69.59%	69.59%	69.59%
$(k = 1, w_L)$	0.48%	0.29%	0.29%	0.29%
Welfare: $\widetilde{V} - V$				
$(k = 0, w_H)$	-	5.420	5.420	5.420
$(k = 0, w_M)$	-	5.445	5.445	5.445
$(k = 0, w_L)$	-	5.539	5.539	5.539
$(k = 1, w_H)$	-	4.535	4.535	4.535
$(k = 1, w_M)$	-	4.727	4.727	4.727
$(k = 1, w_L)$	-	5.219	5.219	5.219
			CF2: $\widetilde{\pi}_0 = \pi_0, \widetilde{\pi}_1 = H_1\pi_1$	
CCP: $\Pr(active \vert x)$				
$(k = 0, w_H)$	93.61%	94.95%	93.53%	99.87%
$(k = 0, d_M)$	87.48%	90.27%	87.31%	99.84%
$(k = 0, d_L)$	72.99%	80.33%	72.53%	99.81%
$(k = 1, d_H)$	99.99%	99.99%	99.99%	90.59%
$(k = 1, w_M)$	80.91%	69.59%	81.44%	0.44%
$(k = 1, w_L)$	0.48%	0.29%	0.49%	0.00%
Welfare: $\widetilde{V} - V$				
$(k = 0, w_H)$	-	5.420	–0.289	88.255
$(k = 0, w_M)$	-	5.445	–0.290	88.829
$(k = 0, w_L)$	-	5.539	–0.295	89.756
$(k = 1, w_H)$	-	4.535	–0.239	77.068
$(k = 1, w_M)$	-	4.727	–0.248	82.836
$(k = 1, w_L)$	-	5.219	–0.278	84.802

to re-enter in the future. So, the firm enters and exits more often in the true counterfactual. The entry subsidies also increase the value of the firm in all states.

In counterfactual 1 (additive subsidy), as expected, the counterfactual CCPs and welfare are identical in the true model and under both estimated models. Thus, when the counterfactual is identified, it does not matter what restrictions the researcher chooses when estimating the model. In contrast, counterfactual 2 (proportional subsidy) results in very different outcomes under the two restrictions. When we set $sv = 0$, the changes in the CCPs are all in the wrong direction: while the true entry probability increases relative to the baseline, the predicted counterfactual entry probability decreases. Similarly, the counterfactual exit probability decreases in the true model, while it increases in the estimated model. Welfare also has the wrong sign in all states. This is a direct consequence of the fact that the estimated entry cost under this restriction has the wrong sign: in the true model, multiplying $\pi(active, 0, w)$ by H_1 represents a subsidy, but in the estimated model, it becomes a tax. This illustrates the importance of the identifying restrictions in driving conclusions, especially when the researcher does not know the sign of the true parameter. When instead we restrict $fc = 0$, since the estimated entry costs and scrap values are magnified, it is profitable to enter and exit the market repeatedly when the entry cost is reduced in the counterfactual. Predicted turnover and welfare are therefore excessive.

Counterfactual 3 changes the transition process $\Pr(w_{t+1}|w_t)$. As discussed previously, changes in the transition process generically result in non-identified counterfactual behavior. However, counterfactual behavior and welfare are identified in the present case due to the parametric restrictions (Proposition 8 and Corollary 2). To see why, recall that when counterfactual behavior depends on baseline payoffs only through the (identified) difference in present values of always choosing $a = 0$ versus $a = 1$, then the counterfactual is identified. Here, the new CCP depends on this baseline difference in present values *plus* the change in present values of variable profits $\bar{\pi}(w)$ that results from changing $\Pr(w_{t+1}|w_t)$. This second quantity is known not just because $\bar{\pi}(w)$ is known, but because the evolution of the exogenous states w is known in both the baseline and the counterfactual scenarios. Counterfactual behavior (and also welfare) are therefore identified, and the top panel of Table 3 confirms the results.[33]

Finally, counterfactual 4 implements a "change in types" experiment. In particular, we add a second market with different parameter values: market 2 is more profitable than market 1 both through lower entry costs and higher variable profits. We identify the parameters for market 2 as before and perform a counterfactual that substitutes the entry cost of market 1 by the estimated entry cost of market 2.[34]

[33] We set $\widetilde{\Pr}(w'|w) = 1/3$, for all (w', w). Aguirregabiria and Suzuki (2014) also implement a change in transitions in a similar model. But they consider a change in F_a^k, i.e., a change in the transition of states that enter the nonidentified part of payoffs. As expected, their counterfactual is not identified. Similar to our counterfactual 2 under the restriction $sv = 0$, they obtained counterfactual predictions in the wrong direction.

[34] For market 2, we set: $c_2 = 9$, $\eta_2 = 1.7$, $w_2 = (18, 15, 11)$, $fc_2 = 3$, $sv_2 = 8$, $ec_2 = 6$. The discount factor and transition matrix in market 2 is the same as in market 1. The estimated profit under the first restriction ($sv_2 = 0$)

Table 3: Counterfactuals 3 and 4 – Change in $F(w_{t+1}/w_t)$ and Change Markets' Entry Costs

States: (k,w)	**Baseline**	**True CF**	**Estimated CF** *scrap value = 0*	**Estimated CF** *fixed cost = 0*
			CF3: $\widetilde{\pi}_0=\pi_0,\widetilde{\pi}_1=\pi_1,\widetilde{F}^w\neq F^w$	
CCP: $\Pr(active\|x)$				
$(k=0,w_H)$	93.61%	86.97%	86.97%	86.97%
$(k=0,w_M)$	87.48%	86.97%	86.97%	86.97%
$(k=0,w_L)$	72.99%	86.97%	86.97%	86.97%
$(k=1,w_H)$	99.99%	99.99%	99.99%	99.99%
$(k=1,w_M)$	80.91%	80.19%	80.19%	80.19%
$(k=1,w_L)$	0.48%	1.17%	1.17%	1.17%
Welfare: $\widetilde{V}-V$				
$(k=0,w_H)$	-	0.542	0.542	0.542
$(k=0,w_M)$	-	1.347	1.347	1.347
$(k=0,w_L)$	-	2.530	2.530	2.530
$(k=1,w_H)$	-	0.468	0.468	0.468
$(k=1,w_M)$	-	1.350	1.350	1.350
$(k=1,w_L)$	-	1.808	1.808	1.808
			CF4: $\widetilde{\pi}_0^1=\pi_0^2,\widetilde{\pi}_1^1=\pi_1^1$	
CCP: $\Pr(active\|x)$				
$(k=0,w_H)$	93.61%	97.28%	95.22%	100.00%
$(k=0,w_M)$	87.48%	95.08%	90.83%	100.00%
$(k=0,w_L)$	72.99%	91.44%	81.74%	100.00%
$(k=1,w_H)$	99.99%	99.95%	99.99%	0%
$(k=1,w_M)$	80.91%	36.86%	66.67%	0%
$(k=1,w_L)$	0.48%	0.09%	0.02%	0%
Welfare: $\widetilde{V}-V$				
$(k=0,w_H)$	-	19.778	6.684	482.861
$(k=0,w_M)$	-	19.883	6.715	483.667
$(k=0,w_L)$	-	20.198	6.831	484.849
$(k=1,w_H)$	-	16.816	5.602	449.773
$(k=1,w_M)$	-	17.752	5.846	457.934
$(k=1,w_L)$	-	19.044	6.438	459.999

The bottom panel of Table 3 presents the results. Similar to counterfactual 2, turnover increases in the true counterfactual compared to the baseline; and again, the two identifying restrictions generate very different outcomes. This is expected given Proposition 2. Under the first restriction ($sv = 0$), counterfactual CCPs and welfare are all in the right direction, even though the estimated entry costs have the wrong sign in both markets. This happens because replacing the entry cost of market 1 by that of market 2 amounts to an increase in entry costs in the restricted model. However, even though the CCP moves in the right direction, the magnitude is bound to be wrong and turnover under this restriction is not as large as the true counterfactual turnover. Under the second identifying restriction ($fc = 0$), turnover and welfare are again exaggerated, to the point that counterfactual choice probabilities are (numerically close to) either zero or one.

4.2 Empirical Example: Agricultural Land Use Model

In this section, we explore the impact of identifying restrictions on counterfactuals using actual data on agricultural land use. We estimate a dynamic model of farmers' planting choices and perform two counterfactuals of interest: the long-run land use elasticity and a fertilizer tax. We emphasize the impact of identifying restrictions on counterfactuals and relegate the details of the estimation methodology to the Supplemental Material (Section C).

Empirical Model. Each year, field owners decide whether to plant crops or not; i.e. $\mathcal{A} = \{c, nc\}$, where c stands for "crops" and nc stands for "no crops" (e.g. pasture, hay, non-managed land). Fields are indexed by i and counties are indexed by m. We partition the state x_{imt} into:

1. time-invariant field and county characteristics, s_{im}, e.g. slope, soil composition;
2. number of years since field was last in crops, $k_{imt} \in \mathcal{K} = \{0, 1, ..., K\}$; and
3. aggregate state, w_{mt} (e.g. input and output prices, government policies).

Per period payoffs are specified as in (22) so that

$$\pi(a, k, s, w) = \theta_0 \left(a, k, s\right) + \theta_1 Z \left(a, w\right),$$

where $\theta_0 \left(a, k, s\right)$ captures switching costs between land uses and $Z \left(a, w\right)$ are observable measures of returns. The dependence of θ_0 on k is what creates dynamic incentives for landowners. The action of "no crops" leaves the land idle, slowly reverting it to natural vegetation, rough terrain, etc. The farmer needs to clear the land in order to convert to crop and start planting. The costs of switching to crop may be rising as the terrain gets rougher. At the same time, however, there may

is $ec_2 = 1.6, \pi^2 \left(active, 1, w\right) = (8.52, 1.89, -2.82)$; and under the second restriction ($fc_2 = 0$) is: $ec_2 = -63, sv_2 = 68, \pi^2 \left(active, 1, w\right) = (11.91, 5.29, 0.59)$.

be benefits to switching, e.g. planting crops may be more profitable after the land is left fallow for a year. In summary, we expect $\theta_0(a, k, s)$ to differ across actions and states.

The transition of state variables follows the decomposition (21), $F(k', w'|a, x) = F^k(k'|a, k) F^w(w'|w)$, so that farmers do not affect the evolution of the aggregate state w; this implies that farmers are small (price takers) and that there are no externalities across fields. The transition rule of k is

$$k'(a, k) = \begin{cases} 0 & if\ a = c \text{ (crops)} \\ \min\{k+1, K\}, & if\ a = nc \text{ (no crops)} \end{cases}$$

so that if "no crop" is chosen, the field state since last crop increases by one, up to K, while if "crop" is chosen, the field state is reset to zero. Planting crops is therefore a "renewal" action. We return to the market state w below. Finally, note that the type s is time-invariant.

Data. First, we employ high-resolution annual land use data in the United States from the Cropland Data Layer (CDL) database. We then merge the CDL with an extensive dataset of land transactions obtained from DataQuick (which includes information on price, acreage, field address and other characteristics). Then, we incorporate detailed data from NASA's Shuttle Radar Topography Mission database (with fine topographical information on altitude, slope and aspect); the Global Agro-Ecological Zones dataset (with information on soil categories and on protected land); and various public databases on agricultural production and costs from the USDA. The final dataset goes from 2010 to 2013 for 515 counties and from 2008 to 2013 for 132 counties.

Further details about the construction of the dataset, as well as some summary statistics, are presented in the Supplemental Material, Section B. Here we only emphasize that land use exhibits substantial persistence. The average proportion of cropland in the sample is 15%; the probability of keeping the land in crop is about 85%, while the probability of switching to crops after two years as non-crop is quite small: 1.6%. Finally, the proportion of fields that switch back to crops after one year as "no crop" ranges from 27% to 43% on average depending on the year, which suggests some farmers enjoy benefits from leaving land fallow for a year.

Estimation. Our estimation strategy mostly follows Scott (2013); we augment this land use model to allow for unobserved market states by the econometrician. In other words, we allow for the aggregate state variable w to have an unobserved component. This may be important as the econometrician may not be able to capture the entire information set of the land owner (commodity prices, government policy, etc.). See the Supplemental Material (Section C) for details.[35]

The parameters of interest are θ_1 and $\theta_0(a, k, s)$, for all a, k, s. The slope θ_1 is identified

[35] In a previous version of this paper (see NBER Working Paper 21527) we discuss identification of DDC models with unobserved states. This material is now part of Kalouptsidi, Scott, and Souza-Rodrigues (2018), where we provide the details of a general setup with unobserved market states and characterize the identification of payoffs.

provided there is sufficient variation in $Z(a, w)$; switching costs between land uses, $\theta_0(a, k, s)$, on the other hand are not identified (Proposition 7). Thus, additional restrictions are required to obtain these costs parameters. The sensitivity of certain counterfactuals to identifying restrictions on payoffs calls out for some means to assess the accuracy of these restrictions. To do so, we present and compare two estimators.

First, we estimate the model using the observed data on farmers' actions and states, following Scott (2013). We call this the "CCP estimator." Naturally, in this case we need some identifying restrictions, and as in Scott (2013), we impose $\theta_0(nocrop, k, s) = 0$ for all k and s. These K restrictions (for each field type s) suffice to identify the remaining switching cost parameters. Specifically, farmers' payoffs for $a = nocrop$ are known under these restrictions, since $\pi(nocrop, k, s, w) = \theta_1 Z(nocrop, w)$. Identification of $\theta_0(crop, k, s)$ now follows directly from (3): given that $\pi(nocrop, k, w)$ is known for all states (k, w), equation (3) directly delivers $\pi(crop, k, w)$, implying that $\theta_0(crop, k, s)$ is identified. However, as is common in applied work, there is little guidance to specify the particular values that $\theta_0(nocrop, k, s)$ should take. To evaluate the impact of these restrictions in this real-data setup, we bring in additional data, namely, the land resale prices.

Our second estimator makes use of resale prices to avoid the restrictions $\theta_0(nocrop, k, s) = 0$. We call it the "V-CCP estimator." We assume that the resale prices provide direct information on the value function V.[36] Clearly, if V is known, so are the payoffs: if V is known at all states, we can recover the conditional value functions v_a from equation (2), i.e. $v_a = V - \psi_a$. Then, π_a can be retrieved from the definition of v_a, (1), and since θ_1 is identified, we can obtain θ_0. As explained in the Supplemental Material, our estimator is designed so that the *only* role of the resale price data is to avoid the identifying restrictions $\theta_0(nocrop, k, s) = 0$ for all k, s. By construction of the estimator, θ_1 is the same as that of the CCP estimator.

Section C of the Supplemental Material explains both estimators in detail. Here, we only emphasize that the CCP estimator imposes restrictions on $\theta_0(a, k, s)$ for identification, while the "V-CCP" estimator replaces these *a priori* restrictions with more data-driven restrictions.

Parameter Estimates. Table 4 presents the estimated parameters using the CCP and V-CCP estimators. For brevity we only present the average of the ratio $\theta_0(a, k, s)/\theta_1$ across field types s, where we divide by θ_1 so that the parameters can be interpreted in dollars per acre; $\overline{\theta}_0(a, k)$ denotes the average of $\theta_0(a, k, s)$ across s. We set $K = 2$ due to data limitations and because after

[36] There are numerous ways to model resale markets, and different models may imply different mappings between transaction prices and agents' value function. Here, we essentially consider the simplest possible setting: in a world with a large number of homogeneous agents, a resale transaction price must equal the value of the asset. A similar approach is adopted in Kalouptsidi (2014, 2018). To address concerns that transacted fields may be selected, we compare the transacted fields (in DataQuick) to all US fields (in the CDL) in Table B3 of the Supplemental Material. Overall, the two sets of fields look similar. We also explore whether land use changes upon resale and find no such evidence (see Table C2 in the Supplemental Material).

Table 4: Empirical Results

Estimator:	CCP	V-CCP
$\bar{\theta}_0\,(crop,0)\,/\theta_1$	-721.93 (-1350,-542)	-1228.9 (-2700,-804)
$\bar{\theta}_0\,(crop,1)\,/\theta_1$	-2584.4 (-5500,-1740)	-1119.4 (-4020,-284)
$\bar{\theta}_0\,(crop,2)\,/\theta_1$	-5070.8 (-11060,-3340)	-4530.4 (-10037,-2940)
$\bar{\theta}_0\,(nocrop,0)\,/\theta_1$	0	-2380.3 (-4050,-1900)
$\bar{\theta}_0\,(nocrop,1)\,/\theta_1$	0	470.05 (-777,829)
$\bar{\theta}_0\,(nocrop,2)\,/\theta_1$	0	-454.58 (-1240,-229)
θ_1^{-1}	734.08 (358,1110)	734.08 (358,1110)

θ_0 values are means across all fields in the sample, divided by θ_1 so that their units are in dollars. 95% confidence intervals in parentheses. Note that θ_1 is proportional to standard deviation of idiosycratic shocks, when payoff function is measured in dollars.

2 years out of crops there are very few conversions back to crops in the data.[37]

The mean switching cost parameters from the CCP estimator are all negative and increase in magnitude with k. One may interpret this as follows: when $k = 0$, crops were planted in the previous year. According to the estimates, preparing the land to replant crops costs on average \$722/acre. When $k = 1$, the land was not used for crops in the previous year. In this case, it costs more to plant crops than when $k = 0$. Conversion costs when $k = 2$ are even larger. Of course this interpretation hinges on the assumption that $\theta_0\,(nocrop,k,s) = 0$ for all k,s. As is typical in switching cost models, estimated switching costs are large in order to explain the observed persistence in choices; unobserved heterogeneity – which is beyond the scope of this paper – can alleviate this issue (see Scott, 2013).

The estimated parameters of the V-CCP estimator do not impose $\theta_0\,(nocrop,k,s) = 0$. When $k = 0$, switching out of crops is now expensive on average (not zero anymore). In fact we test the

[37] We weight observations as in Scott (2013) and cluster standard errors by year. We construct the confidence intervals for $\overline{\theta}_0\,(a,k)\,/\theta_1$ by sampling from the estimated asymptotic distribution of $(\widehat{\theta}_0,\widehat{\theta}_1)$. The details of the first stage estimator are in the Supplemental Material.

joint hypothesis $\overline{\theta}_0\,(nocrop, k) = 0$, for all k, and we reject it. This is reasonable because the "no crops" option incorporates, in addition to fallow land, pasture, hay, and other land uses. While staying out of crops for one year may be the result of the decision to leave land fallow, staying out of crops for longer periods reflects other land usages (since land will likely not stay idle forever) with their associated preparation costs. Furthermore, the estimated value of $\theta_0\,(crop, k, s)$ is also affected when we drop the restriction. Indeed, the absolute value of the estimated $\overline{\theta}_0\,(crop, 0)$ is now larger than the absolute value of $\overline{\theta}_0\,(crop, 1)$. This reflects the benefits of leaving land fallow for one year (i.e. smaller replanting costs). This potential benefit is not apparent when we restrict $\theta_0\,(nocrop, k, s)$. Given that the probability of planting crops after one year of fallow is lower than the probability of planting crops after crops in the data (in most counties), in order to rationalize the choice probabilities, the restricted model (imposing $\theta_0\,(nocrop, k, s) = 0$) must assign higher costs to crops after fallow than after crops. We view this as an appealing feature of the V-CCP model – it is arguably not plausible that leaving land out of crops for one year would increase the costs of planting crops in the following year dramatically.[38]

Counterfactuals. We implement two counterfactuals: the long-run elasticity (LRE) of land use and an increase in the costs of replanting crops.

The LRE measures the long-run sensitivity of land use to an (exogenous) change in crop returns, $Z(c, w)$. As previously mentioned, the LRE is an important input to evaluate several policy interventions, including agricultural subsidies and biofuel mandates (Roberts and Schlenker, 2013; Scott, 2013). We compare the share of cropland in the steady-state obtained when $Z(c, w)$ is held fixed at their average recent levels and when $Z(c, w)$ is held fixed at 10% higher levels. The LRE is defined as the arc elasticity between the total acreage in the two steady states.[39]

As shown in Table 5, the CCP and V-CCP estimators give *exactly* the same LRE. This is no coincidence. By Proposition 7, the slope parameter θ_1 is identified and by Proposition 8(i), a counterfactual that changes only the identified part of payoffs is also identified. Intuitively, because θ_1 is identified, a counterfactual that changes $\theta_1 Z(c, w)$ by 10% resembles a "pre-specified additive change," which is an identified counterfactual. Therefore, the LRE is not affected by identifying restrictions on θ_0, and the *only* difference between the CCP estimator and the V-CCP estimator is that the latter relies on land values to identify the profit function while the former relies on *a priori* restrictions. Since the counterfactual of interest is identified, it does not matter which restrictions are imposed when estimating the model parameters.

[38] One could also argue that it is not plausible that staying out of crops for only two years would lead to dramatically higher costs of planting crops. However, as mentioned previously, we observe very few fields in the data with field state $k = 2$ which have not been out of crops for longer than two years; i.e., fields which have been out of crops for at least two years have typically been out of crops for a long time.

[39] See Scott (2013) for a formal definition and further discussion. The LREs estimated here are higher than those found in Scott (2013) (although not significantly so). We find that this is largely due to our different sample combined with the absence of unobserved heterogeneity: when Scott's estimation strategy is applied to our sample of counties ignoring unobserved heterogeneity, LREs are very similar to those presented here.

The second counterfactual increases the crop replanting costs as

$$\widetilde{\theta}_0 (crop, 0, s) = \theta_0 (crop, 0, s) + \lambda (\theta_0 (crop, 1, s) - \theta_0 (crop, 0, s)).$$

The difference $\theta_0 (crop, 1, s) - \theta_0 (crop, 0, s)$ captures the benefits of leaving land out of crops for a year. One such benefit is to allow soil nutrient levels to recover, reducing the need for fertilizer inputs. When it is difficult to measure the fertilizer saved by leaving land fallow, one can use the switching cost parameters to implement a counterfactual that resembles a fertilizer tax. A motivation for this type of counterfactual is that higher fertilizer prices would be a likely consequence of pricing greenhouse gas emissions, as fertilizer production is very fossil-fuel intensive. Here we impose $\lambda = 0.1$. So, this exercise changes the costs of replanting crops in a way that reflects 10% of the benefits of leaving land out of crops for one year.

In terms of identification, the counterfactual choice probabilities here depend on the baseline model parameters in levels, not just on payoff differences that can be recovered directly from the data, as shown in Section 3.4. As such, one should expect this counterfactual to be sensitive to model identifying restrictions. Indeed, by Proposition 8(ii), the counterfactual choice probability is not identified here.[40]

As shown in Table 5, the identifying restrictions do matter when it comes to this counterfactual. The CCP estimator leads to a 32% increase in cropland, while the V-CCP estimator predicts a decrease in cropland, as expected. In other words, the CCP estimator errs in predicting not just the magnitude, but also the sign of the change in crop acreage. The reason behind this is that the CCP estimator cannot capture the benefits from leaving land fallow (on average) and thus interprets this counterfactual as a subsidy rather than a tax.

To summarize, when we only change the identifying restrictions (i.e. moving from the CCP to the V-CCP estimator), the LRE does not change, as it involves only a transformation of the identified component of the profit function. However, the land use pattern in the second counterfactual, which involves a transformation of the non-identified part of payoffs, is substantially altered when we modify the identifying restrictions.

5 Conclusion

This paper studies the identification of counterfactuals in dynamic discrete choice models. We provide the set of necessary and sufficient conditions that determine whether counterfactual behavior

[40]Formally, $\theta_0 (a)$ is a 3×1 vector (omitting s in the notation to simplify), and we take $\widetilde{\theta}_0 (a) = H_0 (a) \theta_0 (a)$, with $H_0 (nocrop) = I$, and

$$H_0 (crop) = \begin{bmatrix} 1-\lambda & \lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

These matrices do not satisfy the identification condition in Proposition 8(ii).

Table 5: Policy Counterfactuals

Estimator:	CCP	V-CCP
Long-run elasticity	0.57	0.57
Fertilizer tax	0.32	-0.16

Fertilizer tax statistic is percentage change in long-run cropland.

Long-run elasticity is a 10 percent arc elasticity.

and welfare are identified for a broad class of counterfactuals of interest, including non-additive changes in payoffs or changes to agents' choice sets. We also investigate the identification power of parametric restrictions. For a large class of interventions, the identification conditions are straightforward to verify in practice.

We investigate relevant counterfactuals in two applied examples (a firm's entry/exit decisions and a farmer's land use decisions). The results call for caution while leaving room for optimism: although counterfactual behavior and welfare can be sensitive to identifying restrictions imposed on the model, there exists important classes of counterfactuals that are robust to such restrictions.

References

Abbring, J. H., and O. Daljord (2019): "Identifying the Discount Factor in Dynamic Discrete Choice Models," *Quantitative Economics*, forthcoming. 1, 10, 13

Aguirregabiria, V. (2010): "Another Look at the Identification of Dynamic Discrete Decision Processes: An application to Retirement Behavior," *Journal of Business & Economic Statistics*, 28(2), 201–218. 1, 1, 8, 3.1, 3.3

Aguirregabiria, V., and J. Suzuki (2014): "Identification and Counterfactuals in Dynamic Models of Market Entry and Exit," *Quantitative Marketing and Economics*, 12(3), 267–304. 1, 1, 8, 3.1, 3.1, 3.3, 3.3, 3.5, 4.1, 33

Arcidiacono, P., and R. A. Miller (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity," *Econometrica*, 79(6), 1823–1867. 9

——— (2019): "Identifying Dynamic Discrete Choice Models off Short Panels," *Journal of Econometrics*, forthcoming. 1, 1, 8, 3.1, 3.1

Bajari, P., C. S. Chu, D. Nekipelov, and M. Park (2016): "Identification and semiparametric estimation of a finite horizon dynamic discrete choice model with a terminating action," *Quantitative Marketing and Economics*, 14(4), 271–323. 1

Blevins, J. R. (2014): "Nonparametric identification of dynamic decision processes with discrete and continuous choices," *Quantitative Economics*, 5(3), 531–554. 7, 10, 4, 27

Buchholz, N., M. Shum, and H. Xu (2019): "Semiparametric Estimation of Dynamic Discrete Choice Models," *Journal of Econometrics*, forthcoming. 7, 10

Chan, T. Y., B. H. Hamilton, and N. W. Papageorge (2016): "Health, Risky Behavior and the Value of Medical Innovation for Infectious Diseases," *Review of Economic Studies*, 83(4), 1465–1510. 3.1

Chen, L.-Y. (2017): "Identification of Discrete Choice Dynamic Programming Models with Nonparametric Distribution of Unobservables," *Econometric Theory*, 33(3), 551–577. 7, 10

Chiong, K. X., A. Galichon, and M. Shum (2016): "Duality in dynamic discrete-choice models," *Quantitative Economics*, 7(1), 83–115. 9

Collard-Wexler, A. (2013): "Demand Fluctuations in the Ready-Mix Concrete Industry," *Econometrica*, 81(3), 1003–1037. 3.1, 3.5

Conlon, C. T. (2012): "A Dynamic Model of Costs and Margins in the LCD TV Industry," Discussion paper, New York University. 13

Crawford, G. S., and M. Shum (2005): "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 73(4), 1137–1173. 3.1

Das, S., M. J. Roberts, and J. R. Tybout (2007): "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," *Econometrica*, 75(3), 837–873. 3.1, 4.1

Duflo, E., R. Hanna, and S. P. Ryan (2012): "Incentives work: Getting teachers to come to school," *The American Economic Review*, 102(4), 1241–1278. 3.1

Dunne, T., S. D. Klimek, M. J. Roberts, and D. Y. Xu (2013): "Entry, exit, and the determinants of market structure," *The RAND Journal of Economics*, 44(3), 462–487. 3.1

Eckstein, Z., and O. Lifshitz (2011): "Dynamic Female Labor Supply," *Econometrica*, 79(6), 1675–1726. 3.1

Gilleskie, D. B. (1998): "A Dynamic Stochastic Model of Medical Care Use and Work Absence," *Econometrica*, 66(1), 1–45. 3.1

Heckman, J. J., J. E. Humphries, and G. Veramendi (2016): "Dynamic treatment effects," *Journal of Econometrics*, 191(2), 276 – 292, Innovations in Measurement in Economics and Econometrics. 1

Heckman, J. J., and S. Navarro (2007): "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, 136(2), 341–396. 1

Hendel, I., and A. Nevo (2006): "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, 74(6), 1637–1673. 3.1, 3.5

Hotz, V. J., and R. A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60(3), 497–529. 9, 3.4, A.1.1

Hu, Y., and M. Shum (2012): "Nonparametric identification of dynamic models with unobserved state variables," *Journal of Econometrics*, 171(1), 32–44. 1, 1, 2

Ichimura, H., and C. Taber (2000): "Direct Estimation of Policy Impacts," Discussion Paper No 254, NBER. 2

——— (2002): "Semiparametric Reduced-Form Estimation of Tuition Subsidies," *American Economic Review*, (92), 286–292. 2

Igami, M. (2017): "Estimating the Innovator's Dilemma: Structural Analysis of Creative Destruction in the Hard Disk Drive Industry, 1981–1998," *Journal of Political Economy*, 125(3), 798–847. 3.1, 4.1

Kalouptsidi, M. (2014): "Time to build and fluctuations in bulk shipping," *The American Economic Review*, 104(2), 564–608. 1, 5, 6, 3.1, 4.1, 36

——— (2018): "Detection and Impact of Industrial Subsidies: The Case of Chinese Shipbuilding," *Review of Economics Studies*, 85(2), 1111–1158. 36

Kalouptsidi, M., Y. Kitamura, L. Lima, and E. Souza-Rodrigues (2020): "Partial Identification and Inference for Dynamic Models and Counterfactuals," Discussion Paper No 2221, Cowles Foundation. 1

Kalouptsidi, M., P. T. Scott, and E. Souza-Rodrigues (2017): "On the Non-identification of Counterfactuals in Dynamic Discrete Games," *International Journal of Industrial Organization*, 50, 362–371. 1, 12

Kalouptsidi, M., P. T. Scott, and E. A. Souza-Rodrigues (2018): "Linear IV Regression Estimators for Structural Dynamic Discrete Choice Models," *NBER Working Paper 25134*. 35

Kasahara, H., and K. Shimotsu (2009): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica*, 77(1), pp. 135–175. 1, 1, 2

Keane, M. P., and A. Merlo (2010): "Money, Political Ambition, and the Career Decisions of Politicians," *American Economic Journal: Microeconomics*, 2(3), 186–215. 3.1

Keane, M. P., and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522. 3.1

——— (2010): "The Role of Labor and Marriage Markets, Preference Heterogeneity and the Welfare System in the Life Cycle Decisions of Black, Hispanic and White Women," *International Economic Review*, 51(3), 851–892. 3.1, 3.1

Koning, R. H., H. Neudecker, and T. Wansbeek (1991): "Block Kronecker Products and the vecb Operator," *Linear Algebra and Its Applications*, pp. 165–184. 16

Lin, H. (2015): "Quality Choice and Market Structure: A Dynamic Analysis of Nursing Home Oligopolies," *International Economic Review*, 56(4), 1261–1290. 3.1, 4.1

Magnac, T., and D. Thesmar (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70(2), 801–816. 1, 1, 2, 10, 13

Matzkin, R. L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71(5), 1339–1375. 4

Norets, A., and X. Tang (2014): "Semiparametric Inference in dynamic binary choice models," *The Review of Economic Studies*, 81(3), 1229–1262. 1, 1, 8, 10, 3.1, 3.1, 3.3, 3.3, 3.5

Pesendorfer, M., and P. Schmidt-Dengler (2008): "Asymptotic Least Squares Estimators for Dynamic Games," *The Review of Economic Studies*, 75(3), 901–928. 1

Roberts, M. J., and W. Schlenker (2013): "Identifying Supply and Demand Elasticities of Agricultural Commodities: Implications for the US Ethanol Mandate," *American Economic Review*, 103(6), 2265–95. 1, 4.2

Rosenzweig, M. R., and K. I. Wolpin (1993): "Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low-Income Countries: Investments in Bullocks in India," *Journal of Political Economy*, 101(2), 223–244. 3.1

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: an Empirical Model of Harold Zurcher," *Econometrica*, 55(5), 999–1033. 3.2

——— (1994): "Structural Estimation of Markov Decision Processes," *Handbook of Econometrics 4*, 4, 3081–3143. 1, 1, 2

Rust, J., and C. Phelan (1997): "How social security and medicare affect retirement behavior in a world of incomplete markets," *Econometrica: Journal of the Econometric Society*, pp. 781–831. 3.1

Ryan, S. P. (2012): "The Costs of Environmental Regulation in a Concentrated Industry," *Econometrica*, 80(3), 1019–1061. 3.1

Schiraldi, P. (2011): "Automobile replacement: a dynamic structural approach," *The RAND Journal of Economics*, 42(2), 266–291. 3.1

Scott, P. T. (2013): "Dynamic Discrete Choice Estimation of Agricultural Land Use," *Working Paper*. 1, 4.2, 4.2, 37, 4.2, 39

Varela, M. J. (2018): "The costs of growth: Accelerated growth and crowd-out in the Mexican supermarket industry," *International Journal of Industrial Organization*, 61, 1 – 52. 3.1, 4.1

Wei, C., and S. Li (2014): "The Cost of Greening Stimulus: A Dynamic Discrete Choice Analysis of Vehicle Scrappage Programs," Working Papers 2014-12, The George Washington University, Institute for International Economic Policy. 3.1

A Appendix: Proofs

Appendix A summarizes the proofs of the claims given in the main body of the paper. The subsections here correspond to the main paper's sections.

A.1 Identification of Counterfactual Behavior: The General Case

A.1.1 Proof of Lemma 1

To prove Lemma 1, we make use of Lemma A1 below. Assume without loss of generality that $J = A$. Define

$$\frac{\partial \phi_{-J}}{\partial p} = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \cdots & \Phi_{1,A-1} \\ \Phi_{21} & \Phi_{22} & \cdots & \Phi_{2,A-1} \\ \vdots & \vdots & \vdots & \vdots \\ \Phi_{A-1,1} & \Phi_{A-1,2} & ... & \Phi_{A-1,A-1} \end{bmatrix} \equiv \Phi,$$

where Φ_{ij} are the $X \times X$ matrices with elements $\frac{\partial \phi_{iJ}(p(x))}{\partial p_j(x')}$, with $x, x' \in \mathcal{X}$ for each $i, j = 1, ..., A-1$. Note that each Φ_{ij} is diagonal because $\frac{\partial \phi_{iJ}(p(x))}{\partial p_j(x')} = 0$ when $x \neq x'$.

Next, define the diagonal matrices, P_a, with diagonal element i, $p_a(x_i)$ for $a = 1, ..., A-1$; and let $P = [P_1, P_2, ..., P_{A-1}]$.

Lemma A1. *The Arcidiacono-Miller function $\psi_J(p)$ is continuously differentiable with derivative:*

$$\frac{\partial \psi_J}{\partial p} = P\Phi.$$

Proof. Recall that

$$\psi_J(p(x)) = \int \max_{k\in\mathcal{A}} \{\phi_{kJ}(p(x)) + \varepsilon_k\} \, dG(\varepsilon).$$

Because $\phi_{jJ}(p(x))$ is a continuously differentiable function, as shown by Hotz and Miller (1993), so is $\psi_J(p(x))$. For $x \neq x'$, $\frac{\partial \psi_J(p(x))}{\partial p_a(x')} = 0$ for all a, because $\frac{\partial \phi_{kJ}(p(x))}{\partial p_a(x')} = 0$ for all k. For $x = x'$, apply the Chain Rule and obtain

$$\begin{aligned}
\frac{\partial \psi_J(p(x))}{\partial p_a(x)} &= \int \frac{\partial}{\partial p_a(x)} \left[\max_{k\in\mathcal{A}} \{\phi_{kJ}(p(x)) + \varepsilon_k\}\right] dG(\varepsilon) \\
&= \sum_{j=1}^{J-1} \int 1\left\{j = \arg\max_{k\in\mathcal{A}} \{\phi_{kJ}(p(x)) + \varepsilon_k\}\right\} dG(\varepsilon) \frac{\partial \phi_{jJ}(p(x))}{\partial p_a(x)} \\
&= \sum_{j=1}^{J-1} p_j(x) \frac{\partial \phi_{jJ}(p(x))}{\partial p_a(x)}
\end{aligned}$$

Note that

$$\frac{\partial \psi_J}{\partial p} = [\Psi_1, ..., \Psi_{J-1}]$$

where Ψ_a is the $X \times X$ diagonal matrix with elements $\frac{\partial \psi_J(p(x))}{\partial p_a(x)}$, $x \in \mathcal{X}$, for $a = 1, ...J-1$. Hence,

$$\frac{\partial \psi_J}{\partial p} = [P_1, P_2, ..., P_{J-1}] \, \Phi.$$

∎

To simplify notation, consider the function b_{-J}, instead of $\widetilde{b}_{-J}$. Recall the definition of $b_{-J}(p): \mathbb{R}^{(A-1)X} \to \mathbb{R}^{(A-1)X}$ in Section 2. Because $\psi_a = \psi_J - \phi_{aJ}$, we have

$$b_{-J}(p) = \begin{bmatrix} A_1 - I \\ \vdots \\ A_{J-1} - I \end{bmatrix} \psi_J(p) + \phi_{-J}(p) = \mathbf{A}\psi_J(p) + \phi_{-J}(p),$$

where $\mathbf{A}$ has dimension $(A-1)X \times X$ and $\psi_J(p)$ is a column vector with entries $\psi_J(p(x))$, $x \in \mathcal{X}$, and $\phi_{-J}(p)$ is an $(A-1)X$-valued function with elements $\phi_{aJ}(p(x))$. Because both functions $\psi_J(p)$ and $\phi_{-J}(p)$ are differentiable, by Lemma A1 we have

$$\frac{\partial b_{-J}}{\partial p} = \mathbf{A}\frac{\partial \psi_J}{\partial p} + \frac{\partial \phi_{-J}}{\partial p} = [\mathbf{A}P + I] \, \Phi$$

Note that, by the Hotz-Miller inversion (Hotz and Miller, 1993), all block-matrices Φ_{ij} of Φ are invertible. Further, the blocks are all linearly independent, so Φ is invertible as well. Thus $\left[\frac{\partial b_{-J}(p)}{\partial p}\right]$ will be invertible if $[\mathbf{A}P + I]$ is. Using the identity $det(I + AB) = det(I + BA)$ and the property $\sum_a P_a = I$, we obtain

$$det\left(\mathbf{A}P + I\right) = det\left(I + \sum_{a=1}^{J-1} P_a\left(A_a - I\right)\right) = det\left(P_J + \sum_{a=1}^{J-1} P_a A_a\right)$$

But $A_a = (I - \beta F_a)(I - \beta F_J)^{-1}$ and therefore

$$\begin{aligned} det\left(\mathbf{A}P + I\right) &= det\left(P_J + \sum_{a=1}^{J-1} P_a (I - \beta F_a)(I - \beta F_J)^{-1}\right) \\ &= det\left(P_J (I - \beta F_J) + \sum_{a=1}^{J-1} P_a (I - \beta F_a)\right) det\left((I - \beta F_J)^{-1}\right) \\ &= det\left(\sum_{a=1}^{J} P_a (I - \beta F_a)\right) det\left((I - \beta F_J)^{-1}\right) \\ &= det\left(I - \beta \sum_{a=1}^{J} P_a F_a\right) det\left((I - \beta F_J)^{-1}\right) \end{aligned}$$

Note that $\sum_{a=1}^{J} P_a F_a$ is a stochastic matrix, since all its elements are non-negative and

$$\left(\sum_{a=1}^{J} P_a F_a\right) 1 = \sum_{a=1}^{J} P_a 1 = \left(\sum_{a=1}^{J} P_a\right) 1 = 1,$$

where 1 is a $X \times 1$ vector of ones. Thus, $det\left(I - \beta \sum_{a=1}^{J} P_a F_a\right)$ is nonzero and $det\left(\mathbf{A}P + I\right) \neq 0$.

A.1.2 Proof of Theorem 1

Assume without loss of generality that action $J = A$ belongs to both sets $\mathcal{A}$ and $\widetilde{\mathcal{A}}$. The implicit function theorem allows us to locally solve (14) with respect to $\widetilde{p}$ provided the matrix

$$\frac{\partial}{\partial \widetilde{p}} \left[h_{-J}\left(\pi\right) - \widetilde{A}_{-J} h_J(\pi) - \widetilde{b}_{-J}\left(\widetilde{p}\right)\right] = -\frac{\partial}{\partial \widetilde{p}} \widetilde{b}_{-J}\left(\widetilde{p}\right)$$

is invertible; this is proved in Lemma 1.[41]

The vector $\widetilde{p}$ does not depend on the free parameter π_J if and only if

$$\frac{\partial}{\partial \pi_J} \left[h_a\left(\pi_1, \pi_2 ..., \pi_J\right) - \widetilde{A}_a h_J\left(\pi_1, \pi_2 ..., \pi_J\right) - \widetilde{b}_a\left(\widetilde{p}\right)\right] = 0$$

[41] Because $\pi \in \mathbb{R}^{A \times X}$, the set of payoffs that satisfies (6) is an open linear manifold. Therefore, for any point π in this manifold, there exists a neighborhood for which the implicit function theorem is valid.

for all $a \in \widetilde{\mathcal{A}}$, with $a \neq J$, and all π satisfying (6). But, the above yields

$$\sum_{l \in \mathcal{A}, l \neq J} \frac{\partial h_a}{\partial \pi_l} \frac{\partial \pi_l}{\partial \pi_J} + \frac{\partial h_a}{\partial \pi_J} = \widetilde{A}_a \left(\sum_{l \in \mathcal{A}, l \neq J} \frac{\partial h_J}{\partial \pi_l} \frac{\partial \pi_l}{\partial \pi_J} + \frac{\partial h_J}{\partial \pi_J} \right)$$

where, for each $a \in \widetilde{\mathcal{A}}$ and $l \in \mathcal{A}$, the matrix $\left[\frac{\partial h_a}{\partial \pi_l}\right]$ has dimension $\widetilde{X} \times X$; while $\widetilde{A}_a$ is an $\widetilde{X} \times \widetilde{X}$ matrix. Using (3),

$$\sum_{l \in \mathcal{A}, l \neq J} \frac{\partial h_a}{\partial \pi_l} A_l + \frac{\partial h_a}{\partial \pi_J} = \widetilde{A}_a \left(\sum_{l \in \mathcal{A}, l \neq J} \frac{\partial h_J}{\partial \pi_l} A_l + \frac{\partial h_J}{\partial \pi_J} \right) \tag{A1}$$

or,

$$\begin{bmatrix} \frac{\partial h_a}{\partial \pi_1} & \frac{\partial h_a}{\partial \pi_2} & \dots & \frac{\partial h_a}{\partial \pi_J} \end{bmatrix} \begin{bmatrix} A_{-J} \\ I \end{bmatrix} = \widetilde{A}_a \begin{bmatrix} \frac{\partial h_J}{\partial \pi_1} & \frac{\partial h_J}{\partial \pi_2} & \dots & \frac{\partial h_J}{\partial \pi_J} \end{bmatrix} \begin{bmatrix} A_{-J} \\ I \end{bmatrix}$$

For $a \in \widetilde{\mathcal{A}}$, define the $\widetilde{X} \times AX$ matrix (recall $J = A$)

$$\nabla h_a(\pi) = \begin{bmatrix} \frac{\partial h_a}{\partial \pi_1} & \frac{\partial h_a}{\partial \pi_2} & \dots & \frac{\partial h_a}{\partial \pi_J} \end{bmatrix}.$$

Then, stacking the above expressions for all $a \in \widetilde{\mathcal{A}}$, with $a \neq J$, we obtain

$$\nabla h_{-J}(\pi) \begin{bmatrix} A_{-J} \\ I \end{bmatrix} = \widetilde{A}_{-J} \nabla h_J(\pi) \begin{bmatrix} A_{-J} \\ I \end{bmatrix}.$$

Now apply the property $\mathit{vecbr}(BCA') = (A \boxtimes B)\, \mathit{vecbr}(C)$ to obtain:

$$\begin{aligned} \left(\begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes I\right) \mathit{vecbr}(\nabla h_{-J}(\pi)) - \left(\begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes \widetilde{A}_{-J}\right) \mathit{vecbr}(\nabla h_J(\pi)) &= 0 \\ \underbrace{\begin{bmatrix} \begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes I, & -\begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes \widetilde{A}_{-J} \end{bmatrix}}_{(\widetilde{A}-1)\widetilde{X}X \times (\widetilde{A}\widetilde{X})(AX)} \underbrace{\begin{bmatrix} \mathit{vecbr}(\nabla h_{-J}(\pi)) \\ \mathit{vecbr}(\nabla h_J(\pi)) \end{bmatrix}}_{(\widetilde{A}\widetilde{X})(AX) \times 1} &= 0, \end{aligned}$$

which is (15). Note that $\begin{bmatrix} A'_{-J} & I \end{bmatrix}$ is an $X \times AX$ matrix, while $\left(\begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes I\right)$ is an $(\widetilde{A}-1)\widetilde{X}X \times (\widetilde{A}-1)A\widetilde{X}X$ matrix. Similarly, $\left(\begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes \widetilde{A}_{-J}\right)$ is an $(\widetilde{A}-1)\widetilde{X}X \times A\widetilde{X}X$ matrix, and $\widetilde{A}_{-J}$ is an $(\widetilde{A}-1)\widetilde{X} \times \widetilde{X}$ matrix.[42]

[42] With abuse of notation, the identity matrix in $\begin{bmatrix} A'_{-J} & I \end{bmatrix}$ is an $X \times X$ matrix, while the identity matrix after $\boxtimes$ in $\left(\begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes I\right)$ is $(\widetilde{A}-1)\widetilde{X} \times (\widetilde{A}-1)\widetilde{X}$.

A.1.3 Proof of the Bus Engine Replacement Example in Section 3.2

Let $a = 1$ if *replace*, and $a = 2$ if *keep*. Then

$$\widetilde{\pi} = \begin{bmatrix} (1+\lambda)\,I & -\lambda\,[1,0] \\ 0 & I \end{bmatrix} \pi,$$

where 1 is a vector of ones and 0 is a matrix with zeros. Let $J = 2$. By equation (17) in Corollary 1, $\widetilde{p}$ is identified if and only if

$$\left(H_{11} - \widetilde{A}_1 H_{21}\right) A_1 + H_{12} - \widetilde{A}_1 H_{22} = 0$$

or $(1+\lambda)\,IA_1 - \lambda\,[1,0] - A_1 = 0$, which implies $\lambda A_1 = \lambda\,[1,0]$. This implies A_1 is non-invertible, which is a contradiction.

A.2 Identification of Counterfactual Behavior: Special Cases

A.2.1 Proof of Corollary 1

Because $\frac{\partial h_a}{\partial \pi_l} = H_{al}$, equation (A1) in the proof of Theorem 1 becomes

$$\sum_{l \neq J} H_{al} A_l + H_{aJ} = \widetilde{A}_a \left(\sum_{l \neq J} H_{Jl} A_l + H_{JJ} \right).$$

In the "action diagonal" case, $H_{al} = H_{Jl} = 0$ for all $a, J \neq l$, and the condition simplifies to

$$H_{aa} A_a = \widetilde{A}_a H_{JJ}.$$

A.2.2 Proof of Proposition 1

Equation (18) implies $H_{aa} = A_a H_{JJ} A_a^{-1}$, for all $a \neq J$. So all H_{aa} must be similar. Diagonal similar matrices are equal to each other, which implies $H_{aa} = H_{jj} \equiv H$, for all a and j.

Let A_a be partitioned comformably with H:

$$A_a = \begin{bmatrix} (A_a)_{11} & \dots & (A_a)_{1d} \\ \vdots & \vdots & \vdots \\ (A_a)_{d1} & \dots & (A_a)_{dd} \end{bmatrix}$$

Then, $HA_a - A_a H = 0$ implies that for all $i \neq j$, $(\lambda_i - \lambda_j)\,(A_a)_{ij} = 0$, and since $\lambda_i \neq \lambda_j$, it must be $(A_a)_{ij} = 0$ and A_a is block-diagonal. This proves the equivalence of statements (i) and (ii).

We next prove that statement (ii) implies statement (iii). Suppose A_a is block-diagonal. Then,

$A_a(I - \beta F_J) = (I - \beta F_a)$, or

$$I - A_a = \beta(F_a - A_a F_J) \tag{A2}$$

The left-hand side is block diagonal and its (i, j) block is equal to zero. Therefore,

$$0 = (F_a)_{ij} - \sum_k (A_a)_{ik} (F_J)_{kj}$$

Since A_a is block-diagonal, $(A_a)_{ik} = 0$ for $i \neq k$ and thus $(F_a)_{ij} = (A_a)_{ii} (F_J)_{ij}$. Moreover, if we equate the diagonal blocks in (A2), we have:

$$I - (A_a)_{ii} = \beta\left((F_a)_{ii} - (A_a)_{ii} (F_J)_{ii}\right)$$

and $(A_a)_{ii} (F_J)_{ii} = (A_a F_J)_{ii}$, since A_a is block-diagonal. Rearranging, we establish the claim.

Finally, we show the reverse. Consider the first block-row of A_a, $a_1 = [(A_a)_{11} \ (A_a)_{12} \ \ldots \ (A_a)_{1d}]$. Let $e_1 = [I \ 0 \ \ldots 0]$. Then $a_1 = e_1 A_a = e_1(I - \beta F_a)(I - \beta F_J)^{-1}$. But $e_1(I - \beta F_a) = [I - \beta(F_a)_{11} \ \ldots I - \beta(F_a)_{1d}]$ and statement (iii) implies that

$$e_1(I - \beta F_a) = [(A_a)_{11} (I - \beta(F_J)_{11}) \ \ldots (A_a)_{11} (I - \beta(F_J)_{1d})] = (A_a)_{11} e_1 (I - \beta F_J)$$

Therefore,

$$a_1 = e_1 A_a = (A_a)_{11} e_1 (I - \beta F_J)(I - \beta F_J)^{-1} = (A_a)_{11} e_1 = [(A_a)_{11} \ 0 \ \ldots 0]$$

We conclude that $(A_a)_{1j} = 0$ for $j \neq 1$. The same argument applied to all block-rows shows that A_a is block diagonal with block entries given by $(A_a)_i$.

A.2.3 Proof of Proposition 2

Suppose the counterfactual replaces the payoff of type s_1 by that of s_2 for action J only. Then: $H_{aa} = I$ and

$$H_{JJ} = \begin{bmatrix} 0 & I \\ 0 & I \end{bmatrix} \tag{A3}$$

From Corollary 1 for affine counterfactuals, identification requires that, $H_{aa} A_a = A_a H_{JJ}$. We partition A_a comformably with H_{JJ}, i.e.

$$A_a = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

The identification condition then leads to $A_{11} = A_{21} = 0$ with A_{12}, A_{22} arbitrary.

But in the case of time-invariant types, A_a is block diagonal, as it is not possible to transit from one type to the other (i.e. $A_{ii} = (I - \beta F_a^{s_i})(I - \beta F_J^{s_i})^{-1}$, for $i = 1, 2$ and $A_{12} = A_{21} = 0$). Invertibility of A_a implies that the diagonal blocks are invertible as well, and therefore, the counterfactual is not identified.

Next, suppose the counterfactual replaces the payoff of type s_1 by that of s_2 for all actions. Then, $H_{aa} = H_{JJ}$ and both are given by (A3).

Therefore identification requires $HA = AH$, or, $A_{21} = 0$ and $A_{11} + A_{12} = A_{22}$, $A_{21} + A_{22} = A_{22}$. Given the form of A_a, $A_{21} = 0$ is automatically satisfied, while $A_{12} = 0$ implies that identification requires $A_{22} = A_{11}$ and A_{22} arbitrary.

A.2.4 Proof of Proposition 3

The equivalence of statements (i) and (ii) and the sufficiency part are obvious. Next we prove necessity. Assume $\widetilde{\pi} = \mathcal{H}\pi + g$. We prove the statement in three steps. First we show that all off-diagonal submatrices $H_{aj}, a \neq j$, must have identical rows. Second, we show that, when $J \geq 3$, the off-diagonal blocks in column a of $\mathcal{H}$ must be identical to each other; i.e., $H_{ja} = H_{la} = \overline{H}_a$, for any combination of $j \neq l \neq a$. Finally, we show that all diagonal blocks must be of the form $H_{aa} = \lambda I + \overline{H}_a$ for any choice a, for some scalar λ.

(a) By Corollary 1, equation (17) must hold for all $a \neq J$ and for any arbitrary process F. Take an action $a \neq J$ and post-multiply (17) by $(I - \beta F_J)$. We get

$$\sum_{l \in \mathcal{A}, l \neq J} \left(H_{al} - A_a H_{Jl}\right)\left(I - \beta F_l\right) + H_{aJ}\left(I - \beta F_J\right) - A_a H_{JJ}\left(I - \beta F_J\right) = 0.$$

Take $F_J = I$ (this is allowable), then $A_a = \left(I - \beta F_a\right)\left(1 - \beta\right)^{-1}$ for all $l \neq J$, and

$$\sum_{l \in \mathcal{A}, l \neq J} \left[\left(H_{al} - (1-\beta)^{-1}\left(I - \beta F_a\right) H_{Jl}\right)\left(I - \beta F_l\right)\right] + (1-\beta) H_{aJ} - \left(I - \beta F_a\right) H_{JJ} = 0.$$

Rearranging, we get

$$\begin{aligned}&\sum_{l \in \mathcal{A}, l \neq J} \left[H_{al} - (1-\beta)^{-1} H_{Jl} + (1-\beta) H_{aJ} - H_{JJ}\right] + \left[(1-\beta)^{-1} \beta H_{Jl} - \beta H_{al}\right] F_l \\ &+ F_a \left[(1-\beta)^{-1} \beta H_{Jl} + \beta H_{JJ}\right] - F_a \left[(1-\beta)^{-1} \beta^2 H_{Jl}\right] F_l = 0.\end{aligned}$$

This equals

$$\mathbf{A} + \mathbf{B} F_a + F_a \mathbf{C} + F_a \mathbf{D} F_a = 0, \tag{A4}$$

where

$$\begin{aligned}
\mathbf{A} &= \sum_{l\in\mathcal{A},l\neq J}\left[H_{al}-(1-\beta)^{-1}H_{Jl}+(1-\beta)H_{aJ}-H_{JJ}\right]+\sum_{l\in\mathcal{A},l\neq a,J}\left[(1-\beta)^{-1}\beta H_{Jl}-\beta H_{al}\right]F_l\\
\mathbf{B} &= (1-\beta)^{-1}\beta H_{Ja}-\beta H_{aa}\\
\mathbf{C} &= \sum_{l\in\mathcal{A},l\neq J}\left[(1-\beta)^{-1}\beta H_{Jl}+\beta H_{JJ}\right]-\sum_{l\in\mathcal{A},l\neq a,J}\left[(1-\beta)^{-1}\beta^2 H_{Jl}\right]F_l\\
\mathbf{D} &= -(1-\beta)^{-1}\beta^2 H_{Ja}
\end{aligned}$$

with $F_a \geq 0$ and $F_a 1 = 1$, where 1 is a vector of ones. Fix F_J and F_l for $l \neq a$. The left hand side of equation (A4) can be viewed as a quadratic function in F_a. If this identity is satisfied for all $F_a \geq 0$, then all derivatives of the quadratic function with respect to F_a must be equal to zero.

Let the columns of F_a be

$$F_a = \left[f_1 \quad f_2 \quad ... \quad f_{n-1} \quad 1-\sum_i^{n-1} f_i\right]$$

where n is the number of columns (note that $n = X$). We first take the second derivative, and so we focus on the term $F_a\mathbf{D}F_a$. For $j \neq n$, the (i, j) entry is:

$$(F_a\mathbf{D}F_a)_{i,j} = \sum_{l,k} f_{ik}d_{lk}f_{kj} = \sum_{l,k} d_{lk}f_{ik}f_{kj}$$

Isolate the last entry and substitute in:

$$\begin{aligned}
(F_a\mathbf{D}F_a)_{i,j} &= \sum_{l\neq n,k} d_{lk}f_{ik}f_{kj}+\sum_k d_{lk}f_{in}f_{kj} = \sum_{l\neq n,k} d_{lk}f_{ik}f_{kj}+\left(1-\sum_m^{n-1} f_m\right)\sum_k d_{nk}f_{kj}\\
&= \sum_{l\neq n,k} f_{ik}f_{kj}\left(d_{lk}-d_{nk}\right)
\end{aligned}$$

and therefore we must have

$$d_{lk} = d_{nk}, \text{ for all } k, l \neq n$$

Consider now $j = n$. Then,

$$\begin{aligned}
(F_a\mathbf{D}F_a)_{i,n} &= \sum_{l,k} f_{ik}d_{lk}f_{kn} = \sum_{l\neq n,k} f_{ik}d_{lk}\left(1-\sum_{m=1}^{n-1} f_{km}\right)+\sum_k\left(1-\sum_{l=1}^{n-1} f_{il}\right)d_{lk}\left(1-\sum_{m=1}^{n-1} f_{km}\right)\\
&= \sum_{k.l.m} f_{il}f_{km}\left(d_{nk}-d_{lk}\right)
\end{aligned}$$

which already holds. We conclude that $\mathbf{D}$ has identical rows, which implies H_{Ja} also has identical rows. Because this argument holds for any $a \neq J$, and because the choice of J is arbitrary, each off-diagonal submatrix H_{aj} must have identical rows for *any* pair of actions a and j.

The following facts will be useful below. First note that $A_a 1 = 1$ for any a, since

$$\begin{aligned} A_a 1 &= (I-\beta F_a)(I-\beta F_J)^{-1} 1 = (I-\beta F_a) \sum_{n=0}^{\infty} \beta^n F_J^n 1 = \frac{1}{1-\beta}(I-\beta F_a) 1 \\ &= \frac{1}{1-\beta}(1-\beta F_a 1) = \frac{1}{1-\beta}(1-\beta) 1 = 1. \end{aligned}$$

Given that, take any two actions, j and l, and let $H_{jl} = [\rho_{1jl}1, ..., \rho_{Xjl}1]$. Then,

$$A_a H_{jl} = A_a [\rho_{1jl}1, ..., \rho_{Xjl}1] = [\rho_{1jl} A_a 1, ..., \rho_{Xjl} A_a 1] = H_{jl}.$$

(b) Next, consider the case $J \geq 3$, and take $j \neq a, J$. Return to equation (17). Rearrange it and isolate the terms involving j:

$$\begin{aligned} (H_{aj} - A_a H_{Jj})(I-\beta F_j) &= A_a H_{JJ}(I-\beta F_J) - H_{aJ}(I-\beta F_J) \\ &\quad - \sum_{l \in \mathcal{A}, l \neq j, J} (H_{al} - A_a H_{Jl})(I-\beta F_l). \end{aligned}$$

Fix F_J and F_l, for all $l \neq j, J$, and view this as a function of F_j. The right-hand-side does not depend on F_j, and the term $(H_{aj} - A_a H_{Jj})$ on the left-hand-side is fixed. We need the equality to hold for any F_j. Because $(I - \beta F_j)$ is full rank, the only way this equality can be satisfied for all choices of F_j is for

$$H_{aj} - A_a H_{Jj} = 0.$$

We have shown that $A_a H_{jl} = H_{jl}$ for any pair of actions j and l. We therefore obtain

$$H_{aj} = H_{Jj}.$$

The argument holds for any combination of $j \neq a \neq J$. Take the block-column j of $\mathcal{H}$, then all off-diagonal blocks in column j are identical to each other (in addition to having identical rows each). Denote the off-diagonal matrices in the block-column j by $\overline{H}_j$. I.e., $\overline{H}_j = H_{aj}$, for all pairs $a \neq j$.

(c) Now, return to the case $J \geq 2$. We investigate the block-diagonal terms of $\mathcal{H}$. Again, take (17) for $a \neq J$,

$$(H_{aa} - A_a H_{Ja}) A_a + \sum_{l \in \mathcal{A}, l \neq a, J} (H_{al} - A_a H_{Jl}) A_l + H_{aJ} - A_a H_{JJ} = 0.$$

Given that all off-diagonal terms H_{al}, for all pairs $a \neq l$, must satisfy $H_{al} - A_a H_{Jl} = 0$, equation (17) simplifies to

$$(H_{aa} - A_a H_{Ja}) A_a + H_{aJ} - A_a H_{JJ} = 0.$$

In addition, for all pairs $a \neq l$, we have that $H_{al} = \overline{H}_l$, which implies

$$\left(H_{aa} - \overline{H}_a\right) A_a - A_a \left(H_{JJ} - \overline{H}_J\right) = 0.$$

Furthermore, if we take $F_a = F_J$ (this is allowable), we get

$$\left(H_{aa} - \overline{H}_a\right) = \left(H_{JJ} - \overline{H}_J\right).$$

If we take $F_J = I$ instead, we get

$$\left(H_{aa} - \overline{H}_a\right) (I - \beta F_a) - (I - \beta F_a) \left(H_{JJ} - \overline{H}_J\right) = 0.$$

Rearranging, we obtain

$$\left(H_{aa} - \overline{H}_a\right) F_a = F_a \left(H_{aa} - \overline{H}_a\right).$$

So, $\left(H_{aa} - \overline{H}_a\right)$ and F_a must commute, where F_a is an arbitrary (stochastic) matrix. This implies $\left(H_{aa} - \overline{H}_a\right)$ must be of the form

$$\left(H_{aa} - \overline{H}_a\right) = \lambda I,$$

for all a, where λ is a constant. Finally, note that, for $\mathcal{H}$ with diagonal blocks $H_{aa} = \lambda I + \overline{H}_a$, and off-diagonal blocks $H_{aj} = \overline{H}_j$, we obtain

$$\widetilde{\pi} = \mathcal{H}\pi + g = \lambda\pi + \left[\sum_{j \in \mathcal{A}} \overline{H}_j \pi_j\right] 1 + g$$

where all rows of $\sum_{j \in \mathcal{A}} \overline{H}_j \pi_j$ are identical, and therefore all elements of the vector $[\sum_{j \in \mathcal{A}} \overline{H}_j \pi_j]1$ are the same.

A.2.5 Proof of Proposition 4

Suppose $\mathcal{A} = \{1, 2, ..., A\}$. Without loss of generality, take the reference action to be $J = 1$ and suppose action $j = A + 1$ is new, so that $\widetilde{\mathcal{A}} = \{1, 2, ..., A + 1\}$. Assume $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{F}_a = F_a$, and $\widetilde{\pi} = \mathcal{H}\pi + g$, with $\widetilde{\pi}_a = \pi_a$ for all $a \in \mathcal{A}$, and

$$\widetilde{\pi}_j = \sum_{a=1}^{A} H_{ja} \pi_a + g_j.$$

The identification condition (17) becomes $A_a = \widetilde{A}_a$, for $a = 2, ..., A$, and

$$H_{j1} + \sum_{a=2}^{A} H_{ja} A_a = \widetilde{A}_j, \tag{A5}$$

for $j = A + 1$, since $H_{al} = 0$ and $H_{aa} = I$ for all $a, l \neq j$. The first set of restrictions are satisfied, since transitions are unaffected. Now, post-multiply (A5) by $(I - \beta F_1) = (I - \beta \widetilde{F}_1)$ to obtain:

$$H_{j1}(I - \beta F_1) + \sum_{a=2}^{A} H_{ja}(I - \beta F_a) = I - \beta \widetilde{F}_j$$

or

$$\widetilde{F}_j = \sum_{a=1}^{A} H_{ja} F_a + \beta^{-1} \left(I - \sum_{a=1}^{A} H_{ja} \right)$$

Since transitions are stochastic matrices, we have that $\widetilde{F}_j 1 = 1$, so that

$$1 = \sum_{a=1}^{A} H_{ja} 1 + \beta^{-1} \left(1 - \sum_{a=1}^{A} H_{ja} 1 \right)$$

or $\sum_{a=1}^{A} H_{ja} 1 = 1$.

A.2.6 Proof of Proposition 5

If $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{F}_a = F_a$, and $\widetilde{\pi}_a = \pi_a$ for all $a \in \widetilde{\mathcal{A}}$, then $H_{aa} = I$ and $H_{ak} = 0$ for $a \in \widetilde{\mathcal{A}}$ and $k \in \mathcal{A}$, $a \neq k$, and so (17) becomes $A_a = \widetilde{A}_a$ for all $a \in \widetilde{\mathcal{A}}$, which is satisfied because $\widetilde{F}_a = F_a$ for all a.

A.2.7 Proof of Proposition 6

The proof relies on the following lemma:

Lemma A2. *Set the reference action to be $J = 1$ and suppose action A is eliminated. Suppose further that the first m states are maintained and the remaining $X - m$ are eliminated. The counterfactual is specified by:*

$$\widetilde{\pi}_a = \begin{bmatrix} I_m & 0 \end{bmatrix} \pi_a \tag{A6}$$

We partition the transition matrix as follows:

$$F_a = \begin{bmatrix} \widehat{F}_a & f_a \\ g_a & q_a \end{bmatrix} \tag{A7}$$

where $\widehat{F}_a$ is the $m \times m$ top left submatrix of F_a, corresponding to the maintained states; f_a has dimension $m \times (X - m)$; g_a is $(X - m) \times m$; and q_a is $(X - m) \times (X - m)$. Counterfactual transitions adjust the maintained states as follows,

$$\widetilde{F}_a = \widehat{F}_a + f_a r \tag{A8}$$

where r is a $(X - m) \times m$ matrix such that $r1 = 1$, to secure that $\widetilde{F}_a$ is a stochastic matrix. The *counterfactual is identified if and only if*

$$(I-\beta\widehat{F}_a)^{-1}f_a = (I-\beta\widehat{F}_1)^{-1}f_1 \tag{A9}$$

or $f_a = \hat{A}_a f_1$, *where* $\hat{A}_a = (I-\beta\widehat{F}_a)(I-\beta\widehat{F}_1)^{-1}$.

Proof. The identification condition is $H_{aa}A_a = \tilde{A}_a H_{aa}$ or $H_{aa}(I-\beta F_a) = \tilde{A}_a H_{aa}(I-\beta F_1)$. Combining (A6) and (A7), we obtain:

$$\begin{bmatrix} I-\beta\hat{F}_a & -\beta f_a \end{bmatrix} = \begin{bmatrix} \tilde{A}_a\left(I-\beta\hat{F}_1\right) & -\beta\tilde{A}_a f_1 \end{bmatrix}$$

or

$$I-\beta\hat{F}_a = \tilde{A}_a\left(I-\beta\hat{F}_1\right) \tag{A10}$$

and

$$f_a = \tilde{A}_a f_1$$

We show that (A10) is redundant when (A8) holds. Indeed, (A10) is written $I-\tilde{A}_a = \beta(\hat{F}_a - \tilde{A}_a\hat{F}_1)$, while by definition, $I-\tilde{A}_a = \beta(\tilde{F}_a - \tilde{A}_a\tilde{F}_1)$. Thus, $\hat{F}_a - \tilde{A}_a\hat{F}_1 = \tilde{F}_a - \tilde{A}_a\tilde{F}_1$ and using (A8), $\hat{F}_a - \tilde{A}_a\hat{F}_1 = \widehat{F}_a + f_a r - \tilde{A}_a(\widehat{F}_a + f_a r)$, or $f_a r = \tilde{A}_a f_1 r$, which holds because of (A10). ∎

Next, we return to Proposition 6. Assume $x=(k,w)$, then

$$F_a = F_a^k \otimes F^w = \begin{bmatrix} f_{11}^a F^w & f_{12}^a F^w & \ldots & f_{1K}^a F^w \\ \vdots & \vdots & \ldots & \vdots \\ f_{K1}^a F^w & f_{K2}^a F^w & \ldots & f_{KK}^a F^w \end{bmatrix}$$

where $f_{ij}^a = \Pr(k'=j|k=i,a)$ are the elements of F_a^k. Because $k_t = a_{t-1}$, F_a is a matrix with zeros except in the $a-th$ block-column. The $a-th$ block-column is a block-vector with blocks F^w. If action $a=A$ is eliminated from $\mathcal{A}=\{1,2,...,A\}$, then for all $a\neq A$, we have $f_a = 0$, where f_a is defined in (A7). Because $f_J = 0$ as well, condition (A9) in Lemma A2 is trivially satisfied.

A.3 Identification of Counterfactual Behavior Under Parametric Restrictions

A.3.1 Proof of Proposition 7

We make use of two lemmas. Lemma A3 provides sufficient conditions for the identification of parametric models with linear-in-parameters payoff functions.

Lemma A3. *If* $\pi_a(x)$ *satisfies*

$$\pi_a(x;\theta) = \overline{\pi}_a(x)\,\theta, \tag{A11}$$

where θ is a finite dimensional parameter. The parameter θ is identified provided rank $[\overline{\pi}_{-J} - A_{-J}\overline{\pi}_J] = \dim(\theta)$, *where* $\overline{\pi}_{-J} = [\overline{\pi}_1, ..., \overline{\pi}_{J-1}, \overline{\pi}_{J+1}, ..., \overline{\pi}_A]$.

Proof. Equation (A11) implies $\pi_{-J} = \overline{\pi}_{-J}\theta$. Then (6) becomes $[\overline{\pi}_{-J} - A_{-J}\overline{\pi}_J]\,\theta = b_{-J}$. So, if the rank of the matrix $[\overline{\pi}_{-J} - A_{-J}\overline{\pi}_J]$ equals $\dim(\theta)$, then θ is uniquely determined. ∎

Next, Lemma A4 provides results that are used to prove Proposition 7.

Lemma A4. *Let $D_a = [I - \beta(F^w \otimes F^k_a)]^{-1}$, where I is the identity matrix of size $KW \times KW$. Let I_k be the identity matrix of size K, and $\mathbf{1}$ be the block vector $\mathbf{1} = \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix}$ of size $KW \times K$. Finally, let $A^k_a = \left(I_k - \beta F^k_a\right)\left(I_k - \beta F^k_J\right)^{-1}$. The following properties hold:*

(i) $D_a^{-1}\mathbf{1} = \left(I - \beta(F^w \otimes F^k_a)\right)\mathbf{1} = \mathbf{1}(I_k - \beta F^k_a)$.

(ii) $D_a\mathbf{1} = \left(I - \beta(F^w \otimes F^k_a)\right)^{-1}\mathbf{1} = \mathbf{1}(I_k - \beta F^k_a)^{-1}$.

(iii) $A_a\mathbf{1} = \mathbf{1}A^k_a$.

Statements (ii) and (iii) state that the sum of block entries on each block row of D_a and A_a is constant for all block rows.

Proof. (i) Since F^w is a stochastic matrix, its rows sum to 1: $\sum_j f^w_{ij} = 1$, where f^w_{ij} is the (i,j) element of F^w. By the definition of the Kronecker product,

$$(F^w \otimes F^k_a)\mathbf{1} = \begin{bmatrix} f^w_{11}F^k_a & f^w_{12}F^k_a & ... & f^w_{1W}F^k_a \\ \vdots & \vdots & ... & \vdots \\ f^w_{W1}F^k_a & f^w_{W2}F^k_a & ... & f^w_{WW}F^k_a \end{bmatrix} \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} = \begin{bmatrix} \left(\sum_j f^w_{1j}\right) F^k_a \\ \vdots \\ \left(\sum_j f^w_{Wj}\right) F^k_a \end{bmatrix} = \mathbf{1}F^k_a$$

Thus, $(I - \beta(F^w \otimes F^k_a))\mathbf{1} = \mathbf{1}(I_k - \beta F^k_a)$.

(ii) Let n be a non-negative integer. Then, $(F^w)^n$ is a stochastic matrix with rows summing to 1. Therefore,

$$(F^w \otimes F^k_a)^n = (F^w)^n \otimes (F^k_a)^n$$

and following the proof of (i), we obtain $(F^w \otimes F^k_a)^n\mathbf{1} = \mathbf{1}(F^k_a)^n$. Now,

$$D_a\mathbf{1} = \sum_{n=0}^{\infty} \beta^n (F^w \otimes F^k_a)^n \mathbf{1} = \mathbf{1}\sum_{n=0}^{\infty} \beta^n (F^k_a)^n = \mathbf{1}(I_k - \beta F^k_a)^{-1}.$$

(iii) The proof is a direct consequence of (i) and (ii). Indeed,

$$A_a\mathbf{1} = (I - \beta(F^w \otimes F^k_a))D_J\mathbf{1} = (I - \beta(F^w \otimes F^k_a))\mathbf{1}(I_k - \beta F^k_J)^{-1} = \mathbf{1}(I_k - \beta F^k_a)(I_k - \beta F^k_J)^{-1} = \mathbf{1}A^k_a.$$

∎

We now prove Proposition 7; we focus on the binary choice $\{a, J\}$ for notational simplicity, but the general case is obtained in the same fashion. Let θ be the vector of $4K$ unknown parameters (e.g. $\theta_0^a = [\theta_0(a,1), ..., \theta_0(a,K)]'$),

$$\theta = \begin{bmatrix} \theta_0^a \\ \theta_0^J \\ \theta_1^a \\ \theta_1^J \end{bmatrix}.$$

The parametric form of interest is linear in the parameters; stacking the payoffs for a given w and all k we have:

$$\pi_a(w) = [I_k, 0_k, Z_a(w)I_k, 0_k]\,\theta$$

and

$$\pi_J(w) = [0_k, I_k, 0_k, Z_J(w)\,I_k]\,\theta$$

Collecting $\pi_a(w)$ for all w, we get $\pi_a = \overline{\pi}_a\theta$, where

$$\overline{\pi}_a = \begin{bmatrix} I_k & 0_k & Z_a(1)I_k & 0_k \\ \vdots & \vdots & \vdots & \vdots \\ I_k & 0_k & Z_a(W)I_k & 0_k \end{bmatrix} \tag{A12}$$

and similarly for π_J. In Lemma A3, we showed that identification hinges on the matrix $(\overline{\pi}_a - A_a\overline{\pi}_J)$. This matrix equals:

$$\overline{\pi}_a - A_a\overline{\pi}_J = \begin{bmatrix} 1, & -A_a1, & Z_a, & -A_aZ_J \end{bmatrix} \tag{A13}$$

where $Z_a = [Z_a(1)I_k, ..., Z_a(W)I_k]'$ (the same for Z_J).

It follows from Lemma A4 that the first two block columns of (A13) consist of identical blocks each (the first block column has elements I_k, and the second, $-A_a^k$). As a consequence, the respective block parameters θ_0^a, θ_0^J, are not identified unless extra restrictions are imposed.[43] The remaining parameters, θ_1^a, θ_1^J, are identified as follows.

Consider $(\overline{\pi}_a - A_a\overline{\pi}_J)\,\theta = b_a$, or using (A13):

$$1\theta_0^a - 1A_a^k\theta_0^J + Z_a\theta_1^a - \left[I - \beta\left(F^w \otimes F_a^k\right)\right]\left[I - \beta\left(F^w \otimes F_J^k\right)\right]^{-1} Z_J\theta_1^J = b_a.$$

Left-multiplying both sides by $D_a = \left[I - \beta\left(F^w \otimes F_a^k\right)\right]^{-1}$ and using Lemma A4, we obtain:

$$1\left(I_k - \beta F_a^k\right)^{-1}\theta_0^a - 1\left(I_k - \beta F_J^k\right)^{-1}\theta_0^J + D_aZ_a\theta_1^a - D_JZ_J\theta_1^J = D_ab_a.$$

[43] In the multiple choice one block column is a linear combination of the remaining $(J-1)$ corresponding to θ_0; therefore we need to fix θ_0^J for one action J to identify θ_0^{-J}.

Take the w block row of the above:

$$\left(I_k - \beta F_a^k\right)^{-1} \theta_0^a - \left(I_k - \beta F_J^k\right)^{-1} \theta_0^J + e_w' D_a Z_a \theta_1^a - e_w' D_J Z_J \theta_1^J = e_w' D_a b_a \tag{A14}$$

where $e_w' = [0, 0, ..., I_k, 0, ...0]$ with I_k in the w position. Since $W \geq 3$, take two other distinct block rows corresponding to $\widetilde{w}, \overline{w}$ and difference both from the above to obtain the parameter θ_1^a, θ_1^J.

A.3.2 Proof of Proposition 8

(i) Consider the counterfactual payoff $\widetilde{\pi}(a,k,w) = \theta_0(a,k) + h_1 \left[Z'(a,w)\theta_1(a,k)\right]$. Since the term $Z'(a,w)\theta_1(a,k)$ is known for all (a,k,w), we can write this as an "additive changes" as follows: $\widetilde{\pi}(a,k,w) = \pi(a,k,w) + g$, where $g = h_1 \left[Z'(a,w)\theta_1(a,k)\right] - Z'(a,w)\theta_1(a,k)$ is known.

(ii) Consider the counterfactual

$$\widetilde{\pi}(a,w) = H_0\left(a\right)\theta_0\left(a\right) + Z'\left(a,w\right)\theta_1\left(a\right)$$

for $a = 1, ..., J$, where we stack $\theta_0(a,k)$ and $\theta_1\left(a,k\right)$ for all k and $H_0(a)$ is a $K \times K$ matrix. From the proof of Proposition 7, equation (A14), we know that for any w, the w block row of (3) is

$$\left(I_k - \beta F_a^k\right)^{-1} \theta_0^a - \left(I_k - \beta F_J^k\right)^{-1} \theta_0^J + e_w' D_a Z_a \theta_1^a - e_w' D_J Z_J \theta_1^J = e_w' D_a b_a\left(p\right).$$

The corresponding w block row for the counterfactual scenario is

$$\left(I_k - \beta F_a^k\right)^{-1} H_0\left(a\right) \theta_0^a - \left(I_k - \beta F_J^k\right)^{-1} H_0\left(J\right) \theta_0^J + e_w' D_a Z_a \theta_1^a - e_w' D_J Z_J \theta_1^J = e_w' D_a b_a\left(\widetilde{p}\right).$$

Lack of identification of θ_0 is represented by the free parameter θ_0^J. Using (A14), we prove the claim.

(iii) From item (ii) above, it is clear that when F_a^k changes, $\widetilde{p}$ is identified if and only if for all $a \neq J$, $A_a^k = \widetilde{A}_a^k$.

(iv) When $\widetilde{F}^w \neq F^w$ and $\widetilde{F}_a^k = F_a^k$, the equality $A_a^k = \widetilde{A}_a^k$ trivially holds.

A.4 Identification of Counterfactual Welfare

A.4.1 Proof of Proposition 9

Proposition 9 is a direct consequence of Lemma A5 below.

Lemma A5. *Assume $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{\beta} = \beta$, and let $h_a\left(\pi_a\right) = H_{aa}\pi_a + g_a$, all a. Let $C = H_{-J}A_{-J} - \widetilde{A}_{-J}H_{JJ}$ and $D = \left(I - \beta\widetilde{F}_J\right)\left(I - \beta F_J\right)^{-1} - H_{JJ}$. Then ΔV is identified if and only if*

$$\widetilde{P}\left[C - \widetilde{\mathbf{A}}D\right] = D. \tag{A15}$$

where the matrices $\widetilde{\mathbf{A}}$ and $\widetilde{P}$ are defined as in Lemma 1 (but based on $\widetilde{A}_a$ and $\widetilde{p}$).

Proof. We know that

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

and similarly for $\widetilde{V}$

$$\widetilde{V} = \left(I - \beta \widetilde{F}_J\right)^{-1} (h_J (\pi_J) + \psi_J (\widetilde{p})) .$$

Then,

$$\frac{\partial \Delta V}{\partial \pi_J} = \left(I - \beta \widetilde{F}_J\right)^{-1} \left(H_{JJ} + \frac{\partial \psi_J (\widetilde{p})}{\partial \widetilde{p}} \frac{\partial \widetilde{p}}{\partial \pi_J}\right) - (I - \beta F_J)^{-1} .$$

Therefore, $\frac{\partial \Delta V}{\partial \pi_J} = 0$ if and only if

$$\frac{\partial \psi_J (\widetilde{p})}{\partial \widetilde{p}} \frac{\partial \widetilde{p}}{\partial \pi_J} = D \tag{A16}$$

From Lemma A1, we know that

$$\frac{\partial \psi_J}{\partial \widetilde{p}} = \widetilde{P}\widetilde{\Phi},$$

where $\widetilde{P}$ and $\widetilde{\Phi}$ are the counterfactual counterpart of P and Φ defined in Lemma A1. By the Implicit Function Theorem, we know that

$$\frac{\partial \widetilde{p}}{\partial \pi_J} = \left[\frac{\partial \widetilde{b}_{-J} (\widetilde{p})}{\partial \widetilde{p}}\right]^{-1} \left(H_{-J} A_{-J} - \widetilde{A}_{-J} H_{JJ}\right).$$

and, by Lemma 1,

$$\left[\frac{\partial \widetilde{b}_{-J} (\widetilde{p})}{\partial \widetilde{p}}\right]^{-1} = \widetilde{\Phi}^{-1} \left(\widetilde{\mathbf{A}}\widetilde{P} + I\right)^{-1},$$

Thus (A16) becomes:

$$\widetilde{P} \left(\widetilde{\mathbf{A}}\widetilde{P} + I\right)^{-1} C = D \tag{A17}$$

Note that[44]

$$\left(\widetilde{\mathbf{A}}\widetilde{P} + I\right)^{-1} = I - \widetilde{\mathbf{A}} \left(I + \widetilde{P}\widetilde{\mathbf{A}}\right)^{-1} \widetilde{P}.$$

Define $M = (I + \widetilde{P}\widetilde{\mathbf{A}})$. Then,

$$\widetilde{P}(\widetilde{\mathbf{A}}\widetilde{P} + I)^{-1} = \widetilde{P} - \widetilde{P}\widetilde{\mathbf{A}}M^{-1}\widetilde{P} = \widetilde{P} - (M - I) M^{-1}\widetilde{P} = M^{-1}\widetilde{P}$$

Then, (A17) becomes $M^{-1}\widetilde{P}C = D$, or $\widetilde{P}C = MD = (I + \widetilde{P}\widetilde{\mathbf{A}})D$, or $\widetilde{P}(C - \widetilde{\mathbf{A}}D) = D$, which is (A15). ∎

[44]The equality makes use of the identity $(I - BA)^{-1} = I + B (I - AB)^{-1} A$.

A.4.2 Proof of Corollary 2

Lack of identification of θ_0 is represented by the free parameter θ_0^J. So, applying the same argument as in Lemma A5, but differentiating ΔV with respect to θ_0^J, we prove the claim.

Supplement to "Identification of Counterfactuals in Dynamic Discrete Choice Models

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

Harvard University, CEPR and NBER, NYU Stern School of Business, University of Toronto

February 2020

This supplemental material consists of the following sections: Section B presents the data sources, explains the construction of the variables used in the empirical application, and shows some summary statistics. Section C discusses the implementation of the empirical exercise based on a dynamic model of farmers land use decisions.

B Data and Summary Statistics

Table B1 lists our data sources. All are publicly available for download save DataQuick's land values. Our main sample is based on a sub-grid of the Cropland Data Layer (CDL), a high-resolution (30-56m) annual land-use data that covers the entire contiguous United States since 2008. We took a 840m sub-grid of the CDL within those counties appearing in our DataQuick database.[1] DataQuick collects transaction data from about 85% of US counties and reports the associated price, acreage, parties involved, field address and other characteristics. The coordinates of the centroids of transacted parcels in the DataQuick database are known. To assign transacted parcels a land use, we associate a parcel with the nearest point in the CDL grid.

A total of 91,198 farms were transacted between 2008 to 2013 based on DataQuick. However, we dropped non-standard transactions and outliers from the data. First, because we are interested in the agricultural value of land (not residential value), we only consider transactions of parcels for

*Department of Economics, Harvard University, Littauer Center, Cambridge, MA 02138, myrto@fas.harvard.edu

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

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

[1] The 840m grid scale was chosen for two reasons. First, it provides comprehensive coverage (i.e., most large agricultural fields are sampled) without providing too many repeated points within any given parcel. Second, the CDL data changed from a 56m to a 30m grid, and the 840 grid size allows us to match points across years where the grid size changed while matching centers of pixels within 1m of each other. The CDL features crop-level land cover information. See Scott (2013) for how "crops" and "non-crops" are defined.

Table B1: Data Sources

Dataset	Description	Source
Cropland Data Layer	Land cover	http://nassgeodata.gmu.edu/CropScape/
DataQuick	Real estate transactions, assessments	DataQuick
US Counties	County boundaries	http://www.census.gov/cgi-bin/geo/shapefiles2010/layers.cgi
GAEZ Database	Protected land, soil type	http://www.gaez.iiasa.ac.at/
SRTM	Topographical – altitude and slope	http://dds.cr.usgs.gov/srtm/
NASS Quick Stats	Yields, prices, pasture rental rates	http://www.nass.usda.gov/Quick_{}{}{}{}{}{}{}Stats/
ERS	Operating costs	http://www.ers.usda.gov/data-products/commodity-costs-and-returns.aspx
NOOA Urban Centers	Urban center locations and populations	%{}{}{}{}{} http://www.nws.noaa.gov/geodata/catalog/national/html/urban.

which the municipal assessment assigned zero value to buildings and structures. Additionally, we drop transactions featuring multi-parcels, transactions between family members, properties held in trust, and properties owned by companies. Finally, we drop transactions with extreme prices: those with value per acre greater than \$50,000, total transaction price greater than \$10,000,000, or total transaction price less than \$60; these are considered measurement error. After applying the selection criteria, there remained $24,643$ observations (transactions).

To obtain a rich set of field characteristics, we use soil categories from the Global Agro-Ecological Zones database and information on protected land from the World Database on Protected Areas. Protected land was dropped from all analyses. The NASA's Shuttle Radar Topography Mission (SRTM) database provides detailed topographical information. The raw data consist of high-resolution (approx. 30m) altitudes, from which we calculated slope and aspect, all important determinants of how land is used. Characteristics such as slopes and soil categories are assigned to fields/parcels using nearest neighbor interpolation.

To derive a measure of nearby developed property values, we find the five restaurants nearest to a field, and we average their appraised property values. For each field, we also compute the distance to the nearest urban center with a population of at least 100,000. Location of urban centers were obtained from the National Oceanic and Atmospheric Administration (NOAA).

Finally, we use various public databases on agricultural production and costs from the USDA. The final dataset goes from 2010 to 2013 for 515 counties and from 2008 to 2013 for 132 counties. Crop returns are based on information on yields, prices received, and operating expenditures; non-crop returns are based on more sparse information on pasture land rental rates (see Scott (2013)).

Table B2 presents some summary statistics. Table B3 compares the transacted fields (in DataQuick) to all US fields (in the CDL). Overall, the two sets of fields look similar. In particular, the probability of keeping (switching to) crops is very similar across the two datasets.

Table B2: Summary Statistics

Statistics	Mean	Std Dev	Min	Max
In Cropland	0.147	0.354	0	1
Switch to Crops	0.0162	0.126	0	1
Keep Crops	0.849	0.358	0	1
Crop Returns ($)	228	112	43	701
Slope (grade)	0.049	0.063	0	0.702
Altitude (m)	371	497	−6	3514
Distance to Urban Center (km)	79.8	63.7	1.22	362
Nearest commercial land value ($/acre)	159000	792000	738	73369656
Land value ($/acre)	7940	9720	5.23	50000

A slope of 1 refers to a perfect incline and a slope of 0 refers to perfectly horizontal land.

Table B3: Dataquick vs CDL Data – Time Invariant Characteristics

Mean by dataset	DataQuick	CDL
In Cropland	0.147	0.136
Switch to Crops	0.0162	0.0123
Keep Crops	0.849	0.824
Crop Returns ($)	228	241
Slope (grade)	0.049	0.078
Altitude (m)	371	688
Distance to Urban Center (km)	79.8	103
Nearest commercial land value ($/acre)	159000	168000

C Dynamic Land Use Model and Estimation

C.1 Model with Unobserved States

As mentioned in the main text, we augment the empirical model by allowing for unobserved market states, following Scott (2013). The per period payoff becomes:

$$\pi\left(a, k_{imt}, w_{mt}, s_{im}, \varepsilon_{imt}\right) = \theta_0\left(a, k_{imt}, s_{im}\right) + \theta_1 Z\left(a, w_{mt}\right) + \xi\left(a, k_{imt}, w_{mt}, s_{im}\right) + \varepsilon_{imt} \tag{C1}$$

where $\xi\left(a, k, w, s\right)$ captures unobservable variation in returns, and the idiosyncratic shock ε_{it} has a logistic distribution. (Without loss of generality, $\xi\left(a, k, w, s\right)$ is mean-zero for all $\left(a, k, w, s\right)$.) We construct returns $Z^a_{mt} \equiv Z\left(a, w_{mt}\right)$ using county-year information (expected prices and realized yields for major US crops, as well as USDA cost estimates) as in Scott (2013).[2] As described below, identification requires exclusion restrictions on $\xi\left(a, k, w, s\right)$ (see also Kalouptsidi, Scott, and Souza-Rodrigues, 2018).

C.2 Payoff Parameter Estimation

Throughout this section, we use t-subscripts in place of explicitly writing the aggregate state variable w_{mt}. We also omit the subscripts i (fields) and m (counties) to simplify notation. The derivation relies on two crucial assumptions: (a) agents are small; i.e., changing the action of any agent at time t does not affect the distribution of w_{t+1}, and (b) agents have rational expectations.

Here, we consider two estimators for the payoff function. Let $p^c_t\left(k, s\right)$ denote the probability of choosing action "crops" at time period t given state k for a field of type s, and let σ be the scale parameter of the logit shocks (we discuss this further below). We begin with Scott's (2013) linear estimating equation for a dynamic model with logit errors; we refer the interested reader to Scott (2013) (see also Kalouptsidi, Scott, and Souza-Rodrigues, 2018) for the derivation of the following equation:

$$Y_t\left(k, s\right) = \tilde{\theta}_0\left(k, s\right) + \theta_1\left(Z_t\left(c, s\right) - Z_t\left(nc, s\right)\right) + \tilde{\xi}_{k,s,t} + \tilde{e}_{k,s,t} \tag{C2}$$

[2] We refer the interested reader to Scott (2013) for details of constructing the measure of observed returns Z. Due to data limitations, we restrict Z to depend only on $\left(a, w_{mt}\right)$. One important difference from Scott (2013) is that we have field level observable characteristics s_{im} and they affect land use switching costs.

where

$$
\begin{aligned}
Y_t(k,s) &\equiv \ln\left(\frac{p_t^c(k,s)}{1-p_t^c(k,s)}\right)+\beta\ln\left(\frac{p_{t+1}^c(0,s)}{p_{t+1}^c(k'(nc,k),s)}\right)\\
\tilde{\theta}_0(k,s) &\equiv (\theta_0(c,k,s)-\theta_0(nc,k,s))/\sigma\\
&\quad+\beta(\theta_0(c,0,s)-\theta_0(c,k'(nc,k),s))/\sigma\\
\theta_1 &\equiv 1/\sigma\\
\tilde{\xi}_{k,s,t} &\equiv \xi_t(c,k,s)-\xi_t(nc,k,s)\\
&\quad+\beta(\xi_{t+1}(c,0,s)-\xi_{t+1}(c,k'(nc,k),s))\\
\tilde{e}_{k,s,t} &\equiv \beta(E_t[V_t(0,s)]-V_t(0,s))\\
&\quad-\beta(E_t[V_{t+1}(k'(nc,k),s)]-V_{t+1}(k'(nc,k),s)).
\end{aligned}
$$

Ultimately, this is a linear equation that can be used to estimate the parameters of the payoff function with no need to solve the agent's dynamic optimization problem.

On the left hand side of equation (C2), we have a dependent variable which is a function of conditional choice probabilities (which are estimated in a first stage, described below in Section C.3) and the discount factor (which is imputed; we assume it equals 0.95).

On the right hand side of (C2), the intercept term $\tilde{\theta}_0$ is a combination of intercepts of the payoff function θ_0. We discuss the identification of θ_0 in more detail below, for this is essentially where the two estimators differ.

The error term has two components, $\tilde{\xi}$ and $\tilde{e}$. The term $\tilde{\xi}$ is a function of ξ, representing unobservable variation in returns, while $\tilde{e}$ is a function of expectational error terms. Because Z and ξ may be correlated, we follow Scott (2013) and implement an instrumental variable estimator. To do so, we need exclusion restrictions of the form

$$E\left[\nu_{k,s,t}\left(\tilde{\xi}_{k,s,t}+\tilde{e}_{k,s,t}\right)\right]=0, \tag{C3}$$

where $\nu_{k,s,t}$ is a vector of instrumental variables. Given that agents have rational expectations, $\tilde{e}_{k,s,t}$ is uncorrelated with any function of variables in the time-t information set by construction. For this reason, $E[\nu_{k,s,t}\tilde{e}_{k,s,t}]=0$ holds for any $\nu_{k,s,t}$ in the time-t information set and the question of whether equation (C3) is valid becomes a question of whether $E[\nu_{k,s,t}\tilde{\xi}_{k,s,t}]=0$. Such a restriction is a substantive assumption as exclusion restrictions for instrumental variables typically are.[3]

We take first-differences for each field and field state, implicitly allowing for $\tilde{\xi}_{k,s,t}$ to have fixed effects for s and k (interacted).[4] After taking first differences, the instruments we use are: a

[3]If we were willing to assume that $E[(Z^c_{s,t}-Z^{nc}_{s,t})\tilde{\xi}_{k,s,t}]=0$, then we could estimate equation (C2) using ordinary least squares.

[4]Note that we predict CCPs for each field state k, not just for the field state actually observed on the field, so we can take these first differences for each k regardless of the actual path of k for the field.

constant term, caloric yields, and the lagged value of $Z^{c}_{s,t} - Z^{nc}_{s,t}$.[5] The moment restrictions are used to estimate θ_1. We then form estimates of $\tilde{\theta}_0(k,s)$ by averaging over the residuals for each (k,s) pair.

Up to this point, our two estimators coincide; i.e., our two estimators agree on the estimates of θ_1 and $\tilde{\theta}_0(k,s)$. The estimators differ when it comes from mapping the estimates of $\tilde{\theta}_0(k,s)$ to estimates of $\theta_0(\cdot,k,s)$. Notice that for each type s, equation (C2) includes one intercept parameter $\tilde{\theta}_0(k,s)$ for each field state k. However, the original payoff function involves two intercept parameters ($\theta_0(c,k,s)$ and $\theta_0(nc,k,s)$) for each (s,k) combination. Hence, the need for restrictions for the identification of the model (and our claim in Section 3.5 that θ_0 is not identified without restrictions).

Our first estimator (the CCP estimator) imposes the following restrictions on θ_0:

$$\forall k,s: \quad \theta_0(nc,k,s) = 0. \tag{C4}$$

After imposing (C4), we can solve for $\theta_0(c,k,s)$ from our $\tilde{\theta}_0(k,s)$ estimates, recalling that

$$\tilde{\theta}_0(k,s) \equiv (\theta_0(c,k,s) - \theta_0(nc,k,s))/\sigma + \beta(\theta_0(c,0,s) - \theta_0(c,k'(nc,k),s))/\sigma, \tag{C5}$$

noting that equations (C4) and (C5) represent six linearly independent equations in six unknowns for each (k,s) pair. (And noting that the scale parameter σ is identified given that $\theta_1 \equiv 1/\sigma$.)

Our second estimator (the V-CCP estimator) does not impose equation (C4), and instead uses additional information in resale prices. In order to relate observed resale prices to farmer's payoff and value functions, we need a model of transaction prices. We assume that resale prices measure farmer's ex-ante value functions; i.e.,

$$\ln p^{RS}_t = \ln\left(\tilde{V}_t(k,s)\right) + \eta_t, \tag{C6}$$

where p^{RS}_t is the resale price of a field, η_t is measurement error, and we will explain the reason for the tilde on the value function below. Using resale prices as signals of the value function can be justified by assuming that there is a competitive market for buying farms – see Kalouptsidi (2014) for further discussion of this assumption in the context of bulk shipping.

We estimate a flexible model of how resale prices depend on (k,s,t), much like Kalouptsidi (2014) (see Section C.4 for details about the implementation). Fitted values from this regression can be used as estimates of the value function, but an important caveat is that we must consider the scale of the utility function when interpreting the estimates. In econometric discrete choice models, we typically impose a scale normalization on the model that sets the variance of the idiosyncratic shocks equal to a convenient number (e.g., unity for a probit model of $\pi^2/6$ for a logit model). In our parametric land use model, the coefficient on returns, θ_1, reflects this normalization: the parameter

[5] See Scott (2013) for the measurement of caloric yields.

θ_1 can be understood as the scalar we need to multiply by to convert the units from dollars to utils. When we estimate a hedonic model of the value function, the value function is measured in dollars. Therefore, to convert from the estimated value function to the scale-normalized value function we should multiply by θ_1:

$$V_t(k, s) = \theta_1 \tilde{V}_t(k_{it}, s_{it}).$$

A relationship between value functions and the payoff function can be derived as follows:

$$\begin{aligned} V_t(k, s) &= v_t(c, k, s) + \psi_c(p_t^c(k, s)) \\ &= \pi_t(c, k, s) + \beta E_t[V_{t+1}(k'(c, k), s)] + \psi_c(p_t^c(k, s)) \\ &= \pi_t(c, k, s) + \beta V_{t+1}(k'(c, k), s) + e_{k,s,t} + \psi_c(p_t^c(k, s)) \end{aligned}$$

where

$$e_{k,s,t} \equiv \beta(E_t[V_t(k'(c, k), s)] - V(k'(c, k), s)).$$

Ultimately, we can write the payoff function as a function of conditional choice probabilities (estimated in a first stage), value functions (estimated using retail prices in a first stage), and an expectational error term (mean zero):

$$\pi_t(c, k, s) = V_t(k, s) - \beta V_{t+1}(k'(c, k), s) - \psi_c(p_t^c(k, s)) - e_{k,s,t}. \tag{C7}$$

Recalling that the measured version of the value function needs to be converted from dollars to utils to be on the same scale as the normalized payoff function, we have

$$\pi_t(c, k, s) = \theta_1 \left(\tilde{V}_t(k, s) - \beta \tilde{V}_{t+1}(k'(c, k), s)\right) - \psi_c(p_t^c(k, s)) - e_{k,s,t}. \tag{C8}$$

Noting that an estimate of θ_1 can be obtained from the CCP estimator, we can then obtain estimates of payoffs using equation (C8), simply by plugging in the estimated values of θ_1, $\tilde{V}$ and p.[6] More to the point, we can obtain estimates of the intercept parameters:

$$\theta_0(c, k, s) = -\theta_1 Z_t(c, s) + \theta_1 \left(\tilde{V}_t(k, s) - \beta \tilde{V}_{t+1}(k'(c, k), s)\right) - \psi_c(p_t^c(k, s)) - e_{k,s,t}. \tag{C9}$$

The V-CCP estimator uses equation (C9) to estimate $\theta_0(c, k, s)$ by averaging the right-hand-side of (C9) over time. Finally, the estimates of $\theta_0(nc, k, s)$ are then recovered from equation (C5).

Note that we could alternatively estimate $\theta_0(nc, k, s)$ from an equation like (C9), but using non-

[6]Recall that estimating θ_1 with the CCP estimator does not require any identifying restrictions on θ_0. Consider equation (C2), a regression equation that allows us to estimate θ_1 and $\tilde{\theta}_0$. The identifying restrictions are only needed if we want to map from $\tilde{\theta}_0$ to θ_0.

crops as the action instead of crops. Thus, we have over-identifying restrictions. As the primary reason we consider the V-CCP estimator is to replace the *a priori* identifying restrictions in the CCP estimator with a more data-driven approach, we only take as much information as we need from the resale prices to fully identify the payoff function. If we were to use more information from the resale prices, then the two estimators might not agree on the value of $\tilde{\theta}_0(k,s)$, an object that is identified from CCP data without imposing identifying restrictions. Our two estimators only differ when it comes to parameters that cannot be identified from CCP data without restrictions. Thus, by comparing these two estimators, we isolate the impact of identifying restrictions.

C.3 Conditional Choice Probabilities

We estimate conditional choice probabilities using a semiparametric logit model. The model is fully flexible over field states and year, but smooth across counties. In particular, we maximize the following log likelihood function:

$$\max_{\theta_{ckt}} \sum_{m' \in S_m} \sum_{i \in I_{m'}} w_{m,m'} I\left[k_{imt} = k\right] \left\{ \begin{array}{c} I\left[a_{imt} = c\right] \log\left(p_{mt}\left(c, k, s_{im}; \theta_{ckt}\right)\right) \\ +I\left[a_{imt} = nc\right] \log\left(1 - p_{mt}\left(c, k, s_{im}; \theta_{ckt}\right)\right) \end{array} \right\}$$

where S_m is the set of counties in the same US state as m, I_m is the set of fields in county m, $w_{m,m'}$ is the inverse squared distance between counties m and m', and $I[.]$ is the indicator function. The conditional choice probability is parameterized as follows:

$$p_{mt}\left(c, k, s_{im}; \theta_{ckt}\right) = \frac{\exp\left(s_{im}'\theta_{ckt}\right)}{1 + \exp\left(s_{im}'\theta_{ckt}\right)}.$$

Note that without fields' observable characteristics, this regression would amount to taking frequency estimates for each county, field state, and year, with some smoothing across counties. Including covariates allows for within-county field heterogeneity. The final specification for the conditional choice probabilities only uses $slope_{im}$ among regressors because it proved to be the most powerful predictor of agricultural land use decisions (after conditioning on county and field state).

The set of counties in S_m only includes counties which also appear in the DataQuick database. For some states, the database includes a small number of counties, so in these cases we group two or three states together. For example, only one county in North Dakota appears in our sample, and it is a county on the eastern border of North Dakota, so we combine North Dakota and Minnesota. Thus, for each county m in North Dakota or Minnesota, S_m represents all counties in both states in our sample.[7]

[7] In particular, we form a number of groups for such cases: Delaware and Maryland; North Dakota and Minnesota; Idaho and Montana; Arkansas, Louisiana, and Mississippi; Kentucky and Ohio; Illinois, Indiana, and Wisconsin; Nebraska and Iowa; Oregon and Washington; Colorado and Wyoming.

Table C1: Land Resale Price Regression

VARIABLES	(1) log(land value)
log(distance to urban center)	-0.471***
	(0.0297)
commercial land value	0.102***
	(0.00930)
slope	-1.654***
	(0.160)
alt	-0.000226**
	(9.00e-05)
Observations	24,643
R-squared	0.318

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Ommitted: soil, county, year, and field state dummies

as well as interactions with returns.

For the sake of precision, rather than only estimating CCPs using the CDL sample that was merged with resale data, we used the full 840m sub-grid of fields from the CDL (848,384 fields) for the CCP estimation. We then predicted CCPs and estimated payoff functions using fields that were merged with the resale data.

C.4 Resale Price Regression

Next, we discuss how we estimate the value function from resale prices. We view that our resale market assumptions are not overly restrictive in the context of rural land which features a large number of small agents. The land resale market is arguably thick, with a large number of transactions taking place every year.[8] Moreover, we are able to control for a rich set of field characteristics. Finally, we did not find evidence of selection on land use changes upon resale, as discussed below.

As our transaction data is much more sparse than our choice data, we adopt a more restrictive (parametric) form for modeling land values. We estimate the following regression equation:

$$\ln p_{it}^{RS} = X_{it}'\theta_V + \eta_{it}, \tag{C10}$$

where p_{it}^{RS} is a transaction price (in dollars per acre), and X_{it} is a vector of characteristics for the corresponding field. The covariates X_{it} include all variables in Table B2 (i.e. k, slope, altitude, distance to urban centers, nearby commercial values). They also include year dummies, returns interacted with year dummies, field state dummies interacted with year dummies, and county

[8]Comparing DataQuick with the CDL data we see that 1.4–2% of fields are resold every year. Moreover, the USDA reports that in Wisconsin there are approximately 100 thousand acres transacted every year (about 1000 transactions) out of 14.5 million acres of farmland (seemingly information on other states is not available).

Table C2: Land use and transactions

VARIABLES	(1) incrops2010	(2) incrops2011	(3) incrops2012	(4) incrops2013
soldin2009	0.000647			
	(0.00604)			
soldin2010	0.000116	0.00364		
	(0.00326)	(0.00334)		
soldin2011	-0.00117	0.000629	-0.00159	
	(0.00316)	(0.00324)	(0.00330)	
soldin2012		-0.000620	-0.00472	0.00411
		(0.00306)	(0.00313)	(0.00265)
soldin2013			-0.00962***	-0.000445
			(0.00306)	(0.00256)
Observations	23,492	23,492	23,492	23,492
R-squared	0.666	0.698	0.717	0.757

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Linear probability model. Omitted covariates include current returns, field state, US state, slope, local commercial land value, distance to nearest urban center, and interactions.

dummies.

Table C1 presents the estimated coefficients. Although not shown in the table, the estimated coefficients of k are significant and have the expected signs (the large number of interactions makes it difficult to add them all in the table). This is important for the second stage estimation, as k is the main state variable included in the switching cost parameters $\theta_0(a, k)$.

Note that, because field acreage is available only in the DataQuick dataset, when merging with the CDL and remaining datasets we lose this information. This implies, for example, that acreage cannot be a covariate in the choice probabilities. For this reason, we choose a specification for the value function that regresses price per acre on covariates. The value of our R^2 in our regression is a direct consequence of this fact. When we use total land prices as the dependent variable and include acres on the covariates we obtain R^2 as high as 0.8.

Finally, we briefly discuss the possibility of selection on transacted fields. As shown previously in Table B3 of Section B, the characteristics of the transacted fields (in DataQuick) look similar to all US fields (in the CDL). Furthermore, we investigate whether land use changes upon resale. Using a linear probability model we find no such evidence (see Table C2). We regress the land use decision on dummy variables for whether the field was sold in the current, previous, or following year as well as various control variables. In regressions within each cross section, ten of the eleven coefficients on the land transaction dummy variables are statistically insignificant, and the estimated effect on the probability of crops is always less than 1% (see Table C2). We have tried alternative specifications such as modifying the definition of the year to span the planting year rather than calendar year, and yet we have found no evidence indicating that there is an important connection between land transactions and land use decisions.

References

Kalouptsidi, M. (2014): "Time to build and fluctuations in bulk shipping," *The American Economic Review*, 104(2), 564–608. C.2

Kalouptsidi, M., P. T. Scott, and E. A. Souza-Rodrigues (2018): "Linear IV Regression Estimators for Structural Dynamic Discrete Choice Models," *NBER Working Paper 25134*. C.1, C.2

Scott, P. T. (2013): "Dynamic Discrete Choice Estimation of Agricultural Land Use," *Working Paper*. 1, B, C.1, C.1, C.2, 2, C.2, 5