Identification of Counterfactuals in Dynamic Discrete Choice Models*

Myrto Kalouptsidi[†], Paul T. Scott[‡], Eduardo Souza-Rodrigues[§]
Princeton University and NBER, Toulouse School of Economics, University of Toronto

May 2016

Abstract

Dynamic discrete choice models (DDC) are nonparametrically not identified. However, the non-identification of DDC models does not necessarily imply non-identification of their associated counterfactuals. We provide necessary and sufficient conditions for the identification of counterfactual behavior and welfare for a broad class of counterfactuals. The conditions are simple to check and can be applied to virtually all counterfactuals in the DDC literature. To explore how robust counterfactuals can be to model restrictions in practice, we consider a numerical example of a monopolist's entry problem, as well as an empirical application of agricultural land use. For each case, we provide relevant examples of both identified and non-identified counterfactuals.

KEYWORDS: Identification, Dynamic Discrete Choice, Counterfactual, Land Use

*We are grateful to Victor Aguirregabiria, Peter Arcidiacono, Ivan Canay, Jakub Kastl, Arthur Lewbel, Yao Luo, Thierry Magnac, Ariel Pakes, Junichi Suzuki, Ali Yurukoglu, and seminar participants at various universities for many helpful comments. Adrian Torchiana and Charis Katsiardis 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.

[†]Department of Economics, Princeton University, 315 Fisher Hall, Princeton, NJ 08544-1021, myrto@princeton.edu

[‡]Toulouse School of Economics, 21 allée de Brienne, 31015 Toulouse, France, paul.scott@tse-fr.eu

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

1 Introduction

One of the main reasons applied researchers estimate structural econometric models is to perform counterfactuals, such as policy interventions to the environment under study. Despite a vast literature on the identification of structural models, we know very little about the identification of counterfactuals associated with them. This issue is particularly important when the employed model is not identified, as in the case of dynamic discrete choice (DDC) models. DDC models have proven useful to analyze public policies in a variety of contexts (e.g. labor markets, firm dynamics, health choices). However, non-identification may raise concerns about the robustness of the empirical findings. Typically, there are many different DDC models consistent with an observed dataset, but it may or may not be the case that each of these models predicts the same counterfactual behavior and welfare.[1]

Indeed, are there counterfactuals that are identified even though the model is not? If so, which counterfactuals are identified and which are not? How can researchers select or design identified counterfactuals?

In this paper, we provide a full characterization of the identification of a broad class of counterfactuals in dynamic discrete choice models. Specifically, we provide necessary and sufficient conditions, which are straightforward to check, to determine whether counterfactual behavior and welfare are identified. The results apply to virtually all counterfactuals in the DDC literature, including those involving changes in the choice set and state space, in payoff functions, and in the transition of state variables. One can apply our results on a case-by-case basis to investigate the identification of a particular counterfactual of interest.

To aid practitioners, we explicitly consider a number of important classes of counterfactuals. To give an example, many papers consider hypothetical policy interventions such as subsidies. We show that if the subsidy changes payoffs additively, the counterfactual is identified; in contrast, if the subsidy changes payoffs proportionally, the counterfactual is not identified, except under knife-edge cases.[2] Another common counterfactual eliminates one of the agent's available actions (e.g. remove social security or welfare programs, or remove the ability to experiment with prescription drugs); we show that this is identified.[3] Further, many papers assign the primitives of one group of agents to those of another (e.g. assume preferences of labor market cohorts are equal, or firm entry costs are identical across markets); we show that this counterfactual is generically not identified.

When a counterfactual of interest is not identified, the researcher can add restrictions to the model. These extra restrictions may be insufficient to identify the full model but sufficient to identify the counterfactual. Further, some restrictions may be more plausible than others;

[1] Rust (1994) first showed that DDC models are not identified in standard settings, and Magnac and Thesmar (2002) characterized the degree of underidentification. Heckman and Navarro (2007, p. 342) claim that underidentification "*has fostered the widespread belief that dynamic discrete choice models are identified only by using arbitrary functional form and exclusion restrictions. The entire dynamic discrete choice project thus appears to be without empirical content.*" For recent surveys of the DDC literature, see Aguirregabiria and Mira (2010), Arcidiacono and Ellickson (2011), and Keane, Todd and Wolpin (2011).

[2] In the main text, we explain why many counterfactual changes in payoffs cannot be represented as additive changes. We discuss the few existing results in the literature below.

[3] Sometimes eliminating an action also eliminates states (e.g. that may reflect past actions). We also characterize this case and guide the researcher on how to reallocate the probability mass across the remaining states to obtain identification. Similarly, a counterfactual that adds an action is identified provided the researcher specifies the payoff of the new action appropriately.

thus, researchers would like to know which ones might have an impact on counterfactual behavior. In this paper, we also characterize the minimum number of restrictions on payoffs that are necessary to identify a given counterfactual and guide researchers on the nature of these restrictions.

While we present our main theorem and corollaries above in the context of nonparametric payoffs, we also consider empirically relevant parametric models. Parametric models are popular in practice not only because they reduce the number of parameters to be estimated (helping identifying the model) but also because they allow researchers to extrapolate behavior to states not visited in the data. We formally establish identification of both *payoffs* and *counterfactuals* for a widely used class of parametric models; both results are novel. More important, we show how the nature of the parametric restrictions helps identify a set of counterfactuals of interest. To give an example, a number of papers have implemented a counterfactual that changes the volatility or long-run mean of market states (i.e. it changes state transitions). While such counterfactuals are not identified in a nonparametric setting, we show that most examples of these counterfactuals in the literature are identified given the parametric setting.

Finally, we consider identification of welfare, which is often the ultimate object of interest to policy makers (in terms of both sign and magnitude). We fully characterize this case as well, and show that identification of counterfactual behavior is necessary but not sufficient for the identification of welfare.

To gain some economic intuition and explore how sensitive (or robust) counterfactuals can be to model restrictions in practice, we implement both a numerical exercise and an empirical application. The numerical exercise features a monopoly entry model. To identify such models, researchers must restrict scrap values, entry costs or fixed costs; this is usually accomplished through an exclusion restriction (i.e. these costs do not depend on state variables) and another restriction that fixes some of them to zero. These assumptions, however, are difficult to verify in practice as data on entry costs and scrap values are extremely rare.[4] We implement a number of common counterfactuals (entry cost subsidies, changes in the transition of the demand shocks, across-market comparisons), and show that when a counterfactual is not identified, it can be highly sensitive to assumptions on the scrap values or fixed costs. In some cases, the estimated model predicts changes in behavior and welfare in the wrong direction.

Next, we illustrate 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, and government interventions. We augment Scott's estimation strategy using land resale price data. Similar to Kalouptsidi (2014*a*), 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.[5]

We consider two counterfactuals within the land use model. The first, a long-run land use elasticity, measures the sensitivity of land use to a persistent change in crop returns (Scott (2013)). The second features an increase in the cost of replanting crops and resembles a fertilizer tax. While the long-run elasticity is identified, the fertilizer tax is not. Thus,

[4] Kalouptsidi (2014*a*) using some external information of entry costs and scrap values (in this case, new ship prices and demolition prices) shows that in the shipping industry the latter vary dramatically over states.

[5] Heckman and Navarro (2007) also make use of extra data (labor outcomes such as future earnings) to secure identification in an optimal stopping DDC model.

a model estimated with our augmented estimator and a model imposing Scott's restrictions predict the same long-run elasticity, but predict different responses to the increase in fertilizer taxes (and even responses in different directions).

As an aside, we also consider DDC models with partially observed market states, which are relevant in our land use application. Despite a significant literature on individual level heterogeneity (e.g. Kasahara and Shimotsu (2009), Norets (2009), Arcidiacono and Miller (2011)), little work exists on serially correlated unobserved *market states* in dynamic models. We provide conditions for the identification of payoffs in this context: extra restrictions such as the presence of a renewal or a terminal action are needed.[6]

We close this introduction by relating our results to the literature. As previously mentioned, very few results exist regarding the identification of counterfactuals. Aguirregabiria (2010) showed identification of choice probabilities under additive changes on payoffs. Aguirregabiria and Suzuki (2014) and Norets and Tang (2014) provided an important extension by showing nonidentification of behavior under changes in transition probabilities. Arcidiacono and Miller (2015) extended their results to nonstationary environments. Our paper augments the literature by providing the first full set of necessary and sufficient conditions for identification of both counterfactual behavior and welfare for virtually all counterfactuals in applied work. Indeed, the previous results in the literature are special cases (and straightforward corollaries) of our main theorem. Our paper is the first to characterize important counterfactuals, such as: (i) proportional changes in flow payoffs (e.g. Das, Roberts and Tybout (2007), Lin (2015), Igami (2015)); (ii) replacing preference/costs parameters of one type of agent by parameters of other types (e.g. Keane and Wolpin (2010), Eckstein and Lifshitz (2011), Ryan (2012), Dunne, Klimek, Roberts and Xu (2013)); (iii) changing the choice set available to agents (e.g. Rust and Phelan (1997), Gilleskie (1998), Eckstein and Wolpin (1999), Crawford and Shum (2005), Todd and Wolpin (2006), Gould (2008), Keane and Merlo (2010)); (iv) changing how the flow payoffs respond to some state or outcome variables (e.g. Eckstein and Wolpin (1989), Diermeier, Keane and Merlo (2005), Sweeting (2013)); (v) imposing nonlinear changes on payoffs (e.g. Jeziorski, Krasnokutskaya and Ceccariniz (2015)); (vi) other affine changes in payoffs (e.g. Rust (1987)). Ours is also the first paper to characterize *minimal* model restrictions that are sufficient to identify counterfactuals of interest (specifically, parametric models, as well as linear restrictions on payoffs). It is also the first to consider identification of welfare, which is often the ultimate object of interest to policy makers. In a companion paper (Kalouptsidi, Scott, and Souza-Rodrigues (2016)), we explore the consequences of our identification results in the context of dynamic games.

The paper is organized as follows: Section 2 presents the dynamic discrete choice framework and reconstructs the known results on nonparametric underidentification of standard DDC models. We also present novel results on the identification of parametric models. Section 3 contains our main results on the identification of counterfactuals. Section 4 contains the numerical illustration of our results. Section 5 extends the framework to include resale prices and partially observed market states. Finally, Section 6 presents the empirical exercise, and Section 7 concludes. All proofs are presented in Appendix A. The details of the dataset and estimates of the empirical application are discussed in Appendix B and C.

[6] Our results on unobservable market states are closest to Hu and Shum (2012); we relate our results to theirs in Section 5.2.

2 Modeling Framework

In this section, we lay out the standard empirical framework for dynamic discrete choice models and reconstruct the well-known nonparametric (non)identification results of Rust (1994) and Magnac and Thesmar (2002). Then we provide some novel insights on the identification of parametric payoff functions. We revisit the existing results not only for completeness but also to provide simpler derivations and present a unified treatment of several identification results in the literature. Importantly, our representation of existing results forms the main building block of our counterfactual identification results.

In each period $t \in \{1, 2, ...\}$, agent i chooses one action a_{it} from the finite set $\mathcal{A} = \{1, ..., A\}$. The current 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 $x_{it} \in \mathcal{X} = \{x_1, ..., X\}$, $X < \infty$; while $\varepsilon_{it} = (\varepsilon_{it}(1), ..., \varepsilon_{it}(A))$ is i.i.d. across agents and time and has joint distribution G with continuous support on $\mathbb{R}^A$. The transition distribution function for $(x_{it}, \varepsilon_{it})$ factors as follows:

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

and the current payoff function is given by

$$\pi(a, x_{it}, \varepsilon_{it}) = \pi(a, x_{it}) + \sigma \varepsilon_{it}(a),$$

where $\sigma > 0$ is a scale parameter. Agent i chooses a sequence of actions to maximize the expected discounted payoff:

$$E\left(\sum_{\tau=0}^{\infty} \beta^{\tau} \pi(a_{it+\tau}, x_{it+\tau}, \varepsilon_{it+\tau}) | x_{it}, \varepsilon_{it}\right)$$

where $\beta \in (0, 1)$ is the discount factor. We define $V(x_{it}, \varepsilon_{it})$ as the expected discounted stream of payoffs under optimal behavior. By Bellman's principle of optimality,

$$V(x_{it}, \varepsilon_{it}) = \max_{a \in \mathcal{A}} \{\pi(a, x_{it}) + \sigma \varepsilon_{it}(a) + \beta E[V(x_{it+1}, \varepsilon_{it+1}) | a, x_{it}]\}.$$

Following the literature, we define the *ex ante value function*:

$$V(x_{it}) \equiv \int V(x_{it}, \varepsilon_{it}) dG(\varepsilon_{it}),$$

and the *conditional value function*:

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

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

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

where $1\{\cdot\}$ is the indicator function. We define the vectors $p(x) = [p_1(x), ..., p_{A-1}(x)]$ and $p = [p(x_1), ..., p(X)]$.

The following results provide relations between key objects of the model and are widely used in the literature. We make heavy use of them below.

Lemma 1 *(i) Hotz-Miller inversion: For all $(a,x) \in \mathcal{A} \times \mathcal{X}$ and a given reference action j,*

$$v_a(x) - v_j(x) = \sigma \phi_{aj}(p(x)),$$

where $\phi_{aj}(.)$ are functions mapping the simplex in $\mathbb{R}^A$ onto $\mathbb{R}$ and are derived only from G.

(ii) Arcidiacono-Miller Lemma: For any $(a,x) \in \mathcal{A} \times \mathcal{X}$, there exists a real-valued function $\psi_a(p(x))$ such that

$$V(x) - v_a(x) = \sigma \psi_a(p(x)),$$

where the functions ψ_a are derived only from G. Moreover, $\psi_a(p(x)) = \psi_j(p(x)) - \phi_{aj}(p(x))$, all $a \neq j$.[7]

The second statement is an implication of the Hotz-Miller inversion (Hotz and Miller (1993)) and was formally shown by Arcidiacono and Miller (2011).[8]

2.1 Nonparametric Identification of Payoffs

A DDC model consists of the primitives $(\pi, \sigma, \beta, G, F)$ that generate the endogenous objects $\{p_a, v_a, V,\ a \in \mathcal{A}\}$. The question of interest here is whether we can identify (a subset of) the primitives from the data.

Denote the dataset by $\{y_{it} : i = 1, ..., N; t = 1, ..., T\}$. A standard dataset includes actions and states: $y_{it} = (a_{it}, x_{it})$. We assume the joint distribution of y_{it}, $\Pr(y)$, is known, which implies the CCPs $p_a(x)$ and the transition distribution function F are also known. Further, we follow the literature and assume that (σ, β, G) is known as well. For the purposes of this section, the objective is to identify the payoff function π.[9]

We base our analysis on the following fundamental relationships between the primitives and the endogenous objects:

$$\pi_a = v_a - \beta F_a V, \quad \text{for } a = 1, ..., A \tag{1}$$

$$v_a - v_j = \sigma \phi_{aj}, \text{ for } a = 1, ..., A,\ a \neq j \tag{2}$$

$$V = v_a + \sigma \psi_a, \quad \text{for } a = 1, ..., A, \tag{3}$$

[7] To derive the Arcidiacono-Miller Lemma from the Hotz-Miller inversion, note that

$$\begin{aligned} V(x) &= \int \max_{j \in \mathcal{A}} \{v_j(x) + \sigma\varepsilon(j)\} \, dG(\varepsilon) \\ &= \int \max_{j \in \mathcal{A}} \{v_j(x) - v_a(x) + \sigma\varepsilon(j)\} \, dG(\varepsilon) + v_a(x) \\ &= \int \max_{j \in \mathcal{A}} \{\sigma\phi_{ja}(p(x)) + \sigma\varepsilon(j)\} \, dG(\varepsilon) + v_a(x) \end{aligned}$$

and take $\psi_a(p(x)) = \int \max_{j \in \mathcal{A}} \{\phi_{ja}(p(x)) + \varepsilon(j)\} \, dG(\varepsilon)$.

[8] When ε_{it} follows the extreme value distribution, then $p_a(x) = \exp(v_a(x)/\sigma) / \left(\sum_{j \in \mathcal{A}} \exp(v_j(x)/\sigma)\right)$; $\phi_{aj}(p(x)) = \log p_a(x) - \log p_j(x)$; and $\psi_a(p(x)) = -\log p_a(x) + \gamma$, where γ is the Euler constant.

[9] A scale normalization is typically adopted (e.g. $\sigma = 1$). However, when we consider data sets involving measures of the value function or variable profits (i.e., cardinal information), scale normalizations are no longer innocent and σ needs to be recovered as well.

where $\pi_a, v_a, V, \phi_{aj}, \psi_a \in \mathbb{R}^X$, with $\pi_a(x) = \pi(a, x)$; F_a is the transition matrix with (m, n) element equal to $\Pr(x_{it+1} = x_n | a, x_{it} = x_m)$. Equation (1) defines the conditional value function; (2) restates the Hotz-Miller lemma; and (3), the Arcidiacono-Miller lemma. Note that using the observed choice probabilities, p, we can compute ϕ_{aj}, as well as ψ_a, for all a.

Equations (1)-(3) form a set of $(3A-1)X$ linear restrictions on $(2A+1)X$ unknowns, (π_a, v_a, V). Because the Arcidiacono-Miller lemma is an implication of the Hotz-Miller inversion, the equations are not all linearly independent. Consider a binary choice problem for illustration. The payoff function involves $2X$ parameters. However, there are only X linearly independent choice probabilities in the data. Thus, there are X free parameters in (1)-(3). If we add X linearly independent restrictions, we can solve for a payoff function that is consistent with the observed data.

Proposition 1 below formalizes this underidentification problem and can be considered a restatement of Proposition 2 in Magnac and Thesmar (2002). We will make heavy use of it, and its compact notation, in deriving our counterfactual identification results in Section 3. All proofs are in Appendix A.

Proposition 1 *Let $J \in \mathcal{A}$ be some reference action. For each $a \neq J$, the payoff function π_a can be represented as an affine transformation of π_J:*

$$\pi_a = A_a \pi_J + b_a, \tag{4}$$

where $A_a = (I - \beta F_a)(I - \beta F_J)^{-1}$ and $b_a = \sigma(A_a \psi_J - \psi_a)$.

Given $\Pr(y)$, one can compute A_a and b_a directly from the data for all $a \neq J$. Proposition 1 therefore explicitly lays 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 (4), 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 (4), then that set of restrictions is equivalent to stipulating $\pi_J = \pi_J^*$ *a priori*.

In the remainder of the paper, it will be useful to represent (4) for all actions $a \neq J$ at once using the compact notation

$$\pi_{-J} = A_{-J} \pi_J + b_{-J} \tag{5}$$

where $\pi_{-J} = \left[\pi_1', ..., \pi_{J-1}', \pi_{J+1}', ..., \pi_A'\right]' \in \mathbb{R}^{(A-1)X}$; $A_{-J} = \left[A_1', ..., A_{J-1}', A_{J+1}', ..., A_A'\right]' \in \mathbb{R}^{(A-1)X \times X}$ and $b_{-J} = \left[b_1', ..., b_{J-1}', b_{J+1}', ..., b_A'\right]' \in \mathbb{R}^{(A-1)X}$, where $'$ denotes matrix transpose. The underidentification problem is therefore represented by the free parameter π_J.

Remark 1 *Magnac and Thesmar (2002) show that the presence of a "terminal action" does not help identify π. In Appendix A we obtain a similar result for "renewal actions", which reset the distribution of states regardless of earlier behavior. However, they do have identifying power in the presence of unobserved market-level states (Section 5.2), and in non-stationary environments (Arcidiacono and Miller (2015)).*

2.2 Identification of Parametric Payoffs

More often than not, applied work relies on parametric restrictions. Economic theory and institutional details often justify parameterizations, while computational constraints favor parsimonious specifications (as the dimension of payoffs $A \times X$ is often large). Moreover, as counterfactuals often involve extrapolations to states not observed in the data, parametric restrictions can be important to perform many counterfactuals of interest. We are the first to formally discuss identification of parametric models; we do so through the lens of equation (4).[10]

Most generally, we could consider an arbitrary parametric payoff function $\pi(a,x) = \pi(a,x;\theta)$ with $\dim(\theta) \lll A \times X$. A parametric model is identified if and only if the Jacobian of equation (4) with respect to θ has full rank. However, we can make more powerful and informative statements when we consider the sorts of parametric restrictions researchers typically impose in practice.

Many examples in the literature decompose the state space into two components, $x = (k,w)$, where $k \in \mathcal{K} = \{1,...,K\}$ are states whose evolution can be affected by individuals' choices, and $w \in \mathcal{W} = \{1,...,W\}$ are states not affected by agents' choices (e.g. market-level states), with K, W finite. Formally, this means the distribution of state transitions is decomposed as follows:

$$F(x'|a,x) = F^k(k'|a,k)\,F^w(w'|w)\,. \tag{6}$$

Then, the transition matrix F_a is written as $F_a = F^w \otimes F_a^k$, where $\otimes$ denotes the Kronecker product. In addition, in such models, it is common to adopt a parametric payoff function with the following form:

$$\pi(a,k,w) = \theta_0(a,k) + R(a,w)'\theta_1(a,k) \tag{7}$$

where $R(a,w)$ is a known function of actions and states w (e.g. observed measure of variable profits or returns) and $\theta_0(a,k)$ is interpreted as a fixed cost component.

Proposition 2 provides sufficient conditions for the identification of parametric models with the above form. For notational simplicity, we focus on binary choice with $\mathcal{A} = \{a, J\}$ and assume $R(a,w)$ is scalar. The proposition also holds in the more general case of $F^w(w'|w,a)$ and multivariate $R(a,w)$.

Proposition 2 *Assume (6) and (7) hold. Let*

$$\begin{aligned} D_a &= \left[I - \beta\left(F^w \otimes F_a^k\right)\right]^{-1} \\ R_a &= [R_a(w_1)I_k, ..., R_a(W)I_k]' \end{aligned}$$

and similarly for D_J and R_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

$$\begin{bmatrix} \left(e_w' - e_{\widetilde{w}}'\right) D_a R_a & \left(e_{\widetilde{w}}' - e_w'\right) D_J R_J \\ \left(e_w' - e_{\overline{w}}'\right) D_a R_a & \left(e_{\overline{w}}' - e_w'\right) D_J R_J \end{bmatrix} \tag{8}$$

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

[10] Although Magnac and Thesmar (2002) and Pesendorfer and Schmidt-Dengler (2008) pointed out that parametric assumptions may aid identification by reducing the number of parameters, they do not formally investigate the conditions under which the parameters are identified.

To see why Proposition 2 holds, note first that the term $e'_w D_a R_a$ is the expected discounted present value of R_a given w when the agent always chooses action a. Existence of the inverse of (8) requires w to significantly change the conditional expected values of R_a and R_J (evidently, it is necessary that $R_a \neq R_J$). In addition, the additive separability on payoffs implies

$$(I - \beta F_a)^{-1} \pi_a(\theta) = D_a \theta_0(a, k) + D_a R_a(w)' \theta_1(a, k).$$

We can therefore exploit variation in w (ensured by the existence of the inverse of (8)) to eliminate $\theta_0(a, k)$ in (5) and identify $\theta_1(a, k)$. Because there is no variation in observables that allows us to separate $D_a \theta_0(a, k)$ from $D_J \theta_0(J, k)$, we cannot identify $[\theta_0(a, k), \theta_0(J, k)]$. 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.

2.3 Examples

We present three illustrative examples to showcase the role of identifying restrictions. We return to them later to discuss examples of identified and non-identified counterfactuals.

Example 1: Rust's Bus Engine Replacement Problem. In Rust (1987), the agent faces the optimal stopping problem of replacing a bus's engine, trading-off aging and replacement costs. He has two actions: to replace or keep the engine, $\mathcal{A} = \{\text{replace, keep}\}$. The state variable, x, is the bus mileage which evolves stochastically and is renewed upon replacement. The payoff function is

$$\pi(a, x, \theta) = \begin{cases} -\phi(x, \gamma_1) - c(0, \gamma_2), & if\ a = 0 \text{ (replace)} \\ -c(x, \gamma_2), & if\ a = 1 \text{ (keep)} \end{cases}$$

where $\phi(x, \gamma_1)$ is the cost of replacing an engine (net of scrap value); and $c(x, \gamma_2)$ is the operating cost at mileage x. To identify the model, Rust (1987) adopts an exclusion restriction (state-invariant replacement cost $\phi(x, \gamma_1) = \phi$) and a normalization (operating cost at $x = 0$ is zero, i.e. $c(0, \gamma_2) = 0$). This is sufficient to identify payoffs.

Example 2: Monopolist Entry/Exit Problem. Consider a monopolist deciding whether to be active or exit from a market, so that $\mathcal{A} = \{\text{active, inactive}\}$. Let $x = (k, w)$ with $k_{it} = a_{it-1}$, and w a vector of market characteristics relevant for the firm's variable profits $\overline{\pi}$ (e.g. demand shifters, input prices). The firm's flow payoff is

$$\pi(a, k, w) = \begin{cases} k\phi^s & if\ a = 0 \text{ (inactive)} \\ k(\overline{\pi}(w) - fc) - (1 - k)\phi^e, & if\ a = 1 \text{ (active)} \end{cases} \quad (9)$$

where ϕ^s is the scrap value, fc is the fixed cost, and ϕ^e is the entry cost. Note that this model already imposes that (i) if the firm is inactive at $t-1$ and decides to not enter at t, it receives zero and (ii) $\{fc, \phi^s, \phi^e\}$ are invariant over w. The first restriction seems natural, but the second is more questionable *a priori*. It is also difficult to test as data on entry costs and scrap values are extremely rare (Kalouptsidi (2014*a*)). With the restrictions, the model falls within the parametric framework described above, with $\theta_0(0, 0) = 0$, $\theta_0(1, 0) = -\phi^e$, $\theta_0(0, 1) = \phi^s$, and $\theta_0(1, 1) = -fc$. If variable profits, $\overline{\pi}(w)$, are estimated outside of the dynamic problem using price and quantity data, one may take $R(a,w) = \overline{\pi}(w)$ and $\theta_1(a,k) = 1$; otherwise, one may assume a reduced form profit function $\overline{\pi}(w) = w'\gamma_1$, in which case $R(a,w) = w$ and $\theta_1(a,k) = \gamma_1$, with γ_1 identified under sufficient variation on w. To complete identification we need to add K linearly independent restrictions. Since $K = 2$ here, and we already imposed $\theta_0(0,0) = 0$, we have to restrict one of ϕ^e, ϕ^s or fc. Most commonly, researchers set either $\phi^s = 0$ or $fc = 0$.

Example 3: Agricultural Land Use Model. This example closely follows Scott (2013). Each year, field owners decide whether to plant crops or not; i.e. $\mathcal{A} = \{\text{crop, no crop}\}$ where "no crop" includes pasture, hay, non-managed land, etc. Let k denote the number of years since the field was last in crops, and let w be a vector of aggregate states (e.g. input and output prices, government policies). Per period payoffs are specified as in (7); here, $\theta_0(a,k)$ captures switching costs between land uses and $R(a,w)$ are observable measures of returns. The slope $\theta_1(a,k)$ is identified provided there is sufficient variation on $R(a,w)$. Switching costs between land uses, $\theta_0(a,k)$, on the other hand are not identified; Scott (2013) restricts $\theta_0(nocrop, k) = 0$ for all k to estimate the model. As is common in applied work, there is little guidance to specify the particular values that $\theta_0(a,k)$ should take.

3 Identification of Counterfactuals

This section presents our main results on the identification of counterfactuals. We begin with a taxonomy of counterfactuals. We then provide the necessary and sufficient conditions to identify counterfactual behavior. Next, we investigate several special cases of practical interest, including parametric models, partially restricted (but still not fully identified) models, and identification of welfare.

3.1 Taxonomy of Counterfactuals

A counterfactual is defined by the tuple $\left\{\widetilde{\mathcal{A}}, \widetilde{\mathcal{X}}, h, h^s\right\}$. The sets $\widetilde{\mathcal{A}} = \left\{1, ..., \widetilde{A}\right\}$ and $\widetilde{\mathcal{X}} = \left\{x_1, ..., \widetilde{X}\right\}$ denote the new set of actions and states respectively. 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 $\widetilde{\pi}$; counterfactual payoffs are given by $\widetilde{\pi} = h(\pi)$, where $h(\pi) \equiv [h_1(\pi), ..., h_{\widetilde{A}}(\pi)]$, with $h_a(\pi) = h_a(\pi_1, ..., \pi_A)$ for each $a \in \widetilde{\mathcal{A}}$. We allow h to be any differentiable function. Finally, 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 $\widetilde{F}$, representing for instance changes in the long-run mean or volatility of some market-level variable (e.g. Hendel and Nevo (2007); Collard-Wexler (2013)), or a change in the transition process for an agent-specific state variable like health status (e.g. Chan, Hamilton, and Papageorge (2015)).[11]

[11] Note that 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 multiplicity of equilibria. For a discussion of identification of counterfactuals in dynamic games, see Kalouptsidi, Scott, and Souza-Rodrigues (2016).

A simple family of counterfactuals are the *affine payoff counterfactuals*:

$$\widetilde{\pi} = \mathcal{H}\pi + g, \tag{10}$$

where $\mathcal{H} \in \mathbb{R}^{\widetilde{A}\widetilde{X} \times AX}$ and g is a $\widetilde{A}\widetilde{X} \times 1$ vector. In words, the payoff $\widetilde{\pi}(a, x)$ at a certain action-state pair (a, x) is obtained as the sum of a scalar $g(a, x)$ and a linear combination of all baseline payoffs at all actions and states. It is helpful to write this in a block matrix equivalent form that emphasizes actions at the block level:

$$\widetilde{\pi} = \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1A} \\ \vdots & \vdots & \vdots & \vdots \\ H_{\widetilde{A}1} & H_{\widetilde{A}2} & \cdots & H_{\widetilde{A}A} \end{bmatrix} \pi + g \tag{11}$$

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

Counterfactuals encountered in applied work are usually simpler than the above general setup. When the set of actions and states do not change, we have $\widetilde{\mathcal{A}} = \mathcal{A}$ and $\widetilde{\mathcal{X}} = \mathcal{X}$, which implies $\mathcal{H}$ is a square matrix. When $\widetilde{\pi}_a$ depends solely on π_a, $\mathcal{H}$ is block-diagonal and we can write for all $a \in \mathcal{A}$

$$\widetilde{\pi}_a = H_a \pi_a + g_a, \tag{12}$$

we call these "action diagonal counterfactuals."

We contrast two simple special cases of (12) that are common in applied work. The first case, which we call "additive transfers," takes H_a as the identity matrix for all a (i.e. $\mathcal{H} = I$), so that $\widetilde{\pi}_a = \pi_a + g_a$. Note that "additive transfers" are flexible to the extent that g can vary across actions and states; yet g is not allowed to depend on π, so that the researcher must be able to specify g before estimating the model (e.g. Keane and Wolpin (1997), Schiraldi (2011), Duflo, Hanna and Ryan (2012) and Li and Wei (2014)).[12]

The second special case of affine counterfactuals, which we call "proportional changes" counterfactual, sets $\mathcal{H}$ to be diagonal and $g = 0$. It considers counterfactuals that impose percentage changes on original payoffs, with changes that may differ across actions and states; i.e. $\widetilde{\pi}_a(x) = \lambda_a(x)\,\pi_a(x)$ (e.g. Das, Roberts and Tybout (2007), Varela (2013), Lin (2015), and Igami (2015)).[13] In contrast to "additive transfers", the researcher does not have to specify the difference $\widetilde{\pi} - \pi$ before estimating the model.[14]

Another family of affine counterfactuals replaces the primitives of one observable type of agents by those of another, where types can be broadly defined to include markets or regions (e.g. Keane and Wolpin (2010), Eckstein and Lifshitz (2011), Ryan (2012), and Dunne,

[12] Keane and Wolpin (1997) investigate college tuition subsidies; Schiraldi (2011) and Li and Wei (2014) study automobile scrappage subsidies; and Duflo, Hanna and Ryan (2012) implement optimal bonus incentives for teachers in rural India.

[13] Das, Roberts and Tybout (2007) study firms' exporting decisions; Varela (2013) studies supermarkets' entry decisions; Lin (2015) investigates entry and quality investment in the nursing home industry; and Igami (2015) studies innovation in the hard drive industry. They all implement percentage changes subsidies on entry/sunk costs.

[14] Note that because π is unknown/not identified, in practice it is not possible to represent a "proportional changes" counterfactual by an "additive transfers" (by taking $g = (\mathcal{H} - I)\,\pi$).

Klimek, Roberts and Xu (2013)).[15] To represent such a counterfactual we can explicitly consider time-invariant states s (i.e. observable types), so that the payoff is $\pi_a(x, s)$. For instance, if there are two types, $s \in \{s_1, s_2\}$, a counterfactual in which the payoff of type s_1 is replaced by the payoff of type s_2 is represented by

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

Note that H_a is not diagonal in this case. We call this a "change in types" counterfactual.

Finally, we consider counterfactuals that change the set of actions available to agents. Eliminating an action leads to $\widetilde{\mathcal{A}} = \mathcal{A} - \{j\}$, where j is the action to be eliminated. Then $\widetilde{\pi}$ satisfies (11) 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 $j = 3$, (11) becomes

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

In some cases, changing the set of actions also changes the set of states (e.g. when $x_t = a_{t-1}$ as in our monopolist example) and we allow for this possibility as well. A counterfactual that only adds a new action can also be represented by (11): take $\widetilde{\mathcal{A}} = \mathcal{A} \cup \{j\}$, where j is the new action 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 unobservables G, and possibly new states (e.g. Rust and Phelan (1997), Gilleskie (1998), Keane and Wolpin (2010), Crawford and Shum (2005), Keane and Merlo (2010), and Rosenzweig and Wolpin (1993)).[16]

For the parametric model (6)-(7) of Section 2.2, in addition to the general transformations, one may be interested in changes in either θ_0 or in $R'\theta_1$. The first type of counterfactual is represented by the function

$$\widetilde{\pi}(a, k, w) = h_0 [\theta_0(a, k)] + R'(a, w)\,\theta_1(a, k); \tag{13}$$

while the second type is represented by

$$\widetilde{\pi}(a, k, w) = \theta_0(a, k) + h_1 [R'(a, w)\,\theta_1(a, k)]. \tag{14}$$

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

Below, we use our three examples to illustrate some of these types of counterfactuals.

[15] 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 after-change (Clean Air Act Amendment) entry cost by the before-change entry cost in the cement industry. Dunne, Klimek, Roberts and Xu (2013) replace entry costs in HPSA (Health Professional Shortage Areas) by those in the non-HPSA focusing on dentists and chiropractors.

[16] 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 medication after choosing one drug treatment (i.e., they do not allow for experimentation). Keane and Wolpin (2010) eliminate a welfare program. Keane and Merlo (2010) eliminate the option of private jobs for politicians who leave congress. Rosenzweig and Wolpin (1993) add an insurance option to farmers' choice sets.

Example 1: Rust's Bus Engine Replacement Problem. A policy maker might be interested in maintaining low-mileage buses to reduce breakdowns. To this end, she could offer replacement subsidies. Such subsidies may take the form of "additive transfers" $\widetilde{\pi}(replace, x) = \pi(replace, x) + g(replace, x)$. Alternatively, recall that $\pi(replace, x) = -\phi(x) - c(0)$, where $\phi(x)$ is replacement costs (net of scrap values), and $c(x)$ is operating costs. Then, the policy maker can set $\widetilde{\pi}(replace, x) = -(1 + \lambda(x))\phi(x) - c(0)$, or

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

Such subsidies could depend on the mileage of the bus, but would not affect $\pi(keep, x)$ nor F.[17]

Example 2: Monopolist Entry/Exit Problem. Subsidizing entry costs can increase the frequency at which a monopolist's product is offered in the market and so may increase social welfare. The policy maker needs to estimate how much more often the firm is active in the market, and compare the welfare gains to consumers and the monopolist with the subsidy costs. In our model, a subsidy that decreases entry costs by, say, 10%, can be represented by a proportional change on $\theta_0(a, k)$ for $a = 1$, $k = 0$. Another counterfactual of interest may be how changes in variable profits affect investment decisions. In the monopolist example, this translates into changes in $\overline{\pi}(w)$ (e.g. changes in market size), which can be represented by changes in the identified part of the profit function, $R'\theta_1$. A third possibility is a counterfactual that changes the volatility of demand shocks to investigate the impact of uncertainty on investment. This counterfactual takes $\widetilde{\pi} = \pi$, $\widetilde{F}^k = F^k$ but $\widetilde{F}^w \neq F^w$, where the variance of $\widetilde{F}^w$ is larger than the variance of F^w.

Example 3: Dynamic Land Use Model. Land use is a critical component in the evaluation of several agricultural policies, including agricultural subsidies and biofuel mandates (Roberts and Schlenker (2013), Scott (2013)). Some policies may be reflected in transformations on returns R, while others on conversion costs θ_0. In Section 6.3 we consider a long-run land use elasticity of crop returns and a change in the costs of replanting crops.

3.2 Identification of Counterfactual CCP: The General Case

We now present our main identification theorem. The starting point is equation (4). This relationship is useful for two reasons. First, it does not involve non-primitive objects such as continuation values. Second, the CCP vector generated by the primitives $(\pi, \sigma, \beta, G, F)$ is the unique vector that satisfies (4).[18]

[17] In the second case, we have $\widetilde{\pi} = \mathcal{H}\pi$, where $\mathcal{H}$ is not block-diagonal. More specifically,

$$\begin{bmatrix} \widetilde{\pi}(replace, 0) \\ \vdots \\ \widetilde{\pi}(replace, X) \\ \widetilde{\pi}(keep, 0) \\ \vdots \\ \widetilde{\pi}(keep, X) \end{bmatrix} = \begin{bmatrix} (1+\lambda(0)) & 0 & \cdots & 0 & -\lambda(0) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (1+\lambda(X)) & -\lambda(X) & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \pi(replace, 0) \\ \vdots \\ \pi(replace, X) \\ \pi(keep, 0) \\ \vdots \\ \pi(keep, X) \end{bmatrix}.$$

[18] Note that a unique CPP vector p is indeed guaranteed from (4): since the Bellman is a contraction mapping, V is unique; from (1) so are v_a and thus so is p.

The counterfactual counterpart to (4) for any action $a \in \widetilde{\mathcal{A}}$, with $a \neq J$, is

$$\widetilde{\pi}_a = \widetilde{A}_a \widetilde{\pi}_J + \widetilde{b}_a\left(\widetilde{p}\right), \tag{15}$$

where $\widetilde{A}_a = \left(I - \beta \widetilde{F}_a\right)\left(I - \beta \widetilde{F}_J\right)^{-1}$; $\widetilde{b}_a\left(\widetilde{p}\right) = \sigma\left(\widetilde{A}_a \psi_J\left(\widetilde{p}\right) - \psi_a\left(\widetilde{p}\right)\right)$; $\widetilde{p}$ is the counterfactual CCP; and we take without loss a reference action J that belongs to both $\mathcal{A}$ and $\widetilde{\mathcal{A}}$.

It is clear 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 (15).

Before presenting the general case, we consider a binary choice example to fix ideas. Take $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, and assume $\widetilde{\pi}_a$ is action diagonal so that $\widetilde{\pi}_a = h_a\left(\pi_a\right)$. Take $J = 2$, and rewrite (15) as

$$h_1\left(\pi_1\right) = \widetilde{A}_1 h_2\left(\pi_2\right) + \widetilde{b}_1\left(\widetilde{p}\right). \tag{16}$$

The implicit function theorem allows us to locally solve (16) with respect to $\widetilde{p}$ provided the matrix

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

is invertible. We prove this matrix is indeed invertible in the general case (see Lemma 2 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

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

Because $\pi_1 = A_1\pi_2 + b_1(p)$ from (4), the above equation simplifies to

$$\frac{\partial h_1\left(\pi_1\right)}{\partial \pi_1} A_1 = \widetilde{A}_1 \frac{\partial h_2\left(\pi_2\right)}{\partial \pi_2}. \tag{17}$$

The counterfactual CCP does not depend on π_2 if and only if this equality holds. The equality depends on the (known) counterfactual transformation $\{h, h^s\}$ and on the data, F, through A_1 and $\widetilde{A}_1$. So, in practice, one only needs to verify whether (17) holds for the particular combination $\{h, h^s\}$ of interest.

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

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

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:

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

where $\nabla h(\pi)$ is the matrix with elements $\frac{\partial h_a(\pi)}{\partial \pi_j}$ for $a, j = 1, 2$. So, to satisfy (17) (and for counterfactual CCPs to be identified), $vec(\nabla h(\pi))$ must lie in the nullspace of a matrix determined by A_1 and $\widetilde{A}_1$.

Moving from the binary choice case to the general case, take (15) together with $\widetilde{\pi}_a = h_a(\pi)$ and stack all payoff vectors for $a \neq J$ to obtain:

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

where $h_{-J}(\pi)$ stacks $h_a(\pi)$ for all $a \in \widetilde{\mathcal{A}}$ except for J and matrix $\widetilde{A}_{-J}$ and vector $\widetilde{b}_{-J}(\widetilde{p})$ are defined similarly.

Lemma 2 *The function $b_{-J}(.)$ is continuously differentiable and its Jacobian is everywhere invertible.*

Lemma 2 implies that the function $\widetilde{p}(\pi_J)$ inherits the differentiability of $h(.)$ and $\widetilde{b}_{-J}(.)$ and that we can apply the implicit function theorem to (15).

Next, we state our main theorem, which makes use of the following notation: $vecbr(C)$ 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.[19]

Theorem 1 *Consider the counterfactual transformation $\left\{\widetilde{\mathcal{A}}, \widetilde{\mathcal{X}}, h, h^s\right\}$ and suppose h is differentiable. The counterfactual conditional choice probabilities $\widetilde{p}$ are identified if and only if for all π satisfying* (5),

$$Q\left(A, \widetilde{A}\right) vecbr(\nabla h(\pi)) = 0 \tag{18}$$

where

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

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

Theorem 1 holds that counterfactual choice probabilities $\widetilde{p}$ 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 $\widetilde{A}_{-J}$, which in turn are determined by the transition probabilities F and $\widetilde{F}$.[20] So model primitives, data and counterfactual transformations have to interact with each other in a specific way to obtain identification of counterfactual CCPs. The only requirement in the theorem is that h must be differentiable – this is a mild restriction typically satisfied in practice. Restricting π to satisfy (5) means the theorem applies to all payoffs that can rationalize observed choice probabilities.

---

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

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

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

[20]The choice of the reference action J does not affect whether or not (18) is satisfied. If it is satisfied for one choice of J, it will be satisfied for any choice of J.

Equation (18) is the minimal set of sufficient conditions that applied researchers need to verify to secure identification of counterfactual behavior. Note that (18) is computable from F, $\widetilde{F}$ and h.

Example 1: Rust's Bus Engine Replacement Problem. Rust (1987) investigates the demand for replacement investment. He varies the level of replacement costs and obtains the corresponding (long run) replacement choice probabilities. This counterfactual can be represented by $\widetilde{\pi}\left(replace, x\right) = -\left(1+\lambda\right)\phi\left(x\right) - c\left(0\right)$, for various levels of λ, or

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

as in Example 1 of Section 3.1. Each λ corresponds to one point along the replacement demand curve, and can be understood as a not "action-diagonal" counterfactual. In Appendix A, we show that Theorem 1 implies that Rust's counterfactual is not identified. Showing this for a particular specification requires only a simple calculation evaluating equation (18).[21]

When, in contrast to Rust's counterfactual, the counterfactual payoffs to an action are a function only of the baseline payoffs of *that* action, we say that the counterfactual is *action diagonal*, and the conditions for the identification of counterfactual behavior become substantially simpler:

Corollary 1 *("Nonlinear Action Diagonal" Counterfactual) In "action diagonal" counterfactuals, $\widetilde{\pi}_a = h_a\left(\pi_a\right)$, $\widetilde{p}$ is identified if and only if for all π satisfying* (5) *and all $a \in \widetilde{\mathcal{A}}$, $a \neq J$,*

$$\frac{\partial h_a}{\partial \pi_a} A_a = \widetilde{A}_a \frac{\partial h_J}{\partial \pi_J}. \tag{19}$$

Corollary 1 implies that nonlinear counterfactuals are not identified in general. For counterfactual behavior to be identified, equation (19) must be satisfied for *all* of the various payoff functions π which are compatible with the observed CCP data. However, given that the counterfactual is nonlinear, the values of the derivatives $\frac{\partial h_a}{\partial \pi_a}$ and $\frac{\partial h_J}{\partial \pi_J}$ will be different for different values of π. The matrices A_a and $\widetilde{A}_a$ have full rank, so changing the value of these derivatives will change the values of the respective sides of equation (19). Thus, if the equation is satisfied for some value of π, it will typically not be satisfied for other values of π. Consequently, the only nonlinear counterfactuals which could be identified would be special cases in which the nonlinearities in the counterfactual are such that the derivatives $\frac{\partial h_a}{\partial \pi_a}$ and $\frac{\partial h_J}{\partial \pi_J}$ change in the same way when π_a and π_J are changed according to equation (5).

[21] Consider a simple version of Rust's model in which $a = replace$ and $J = keep$ and where the state space is simply $X = \{new, old\}$ with deterministic transitions:

$$F_{replace} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \qquad F_{keep} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

Then $A_{-J} = A_{replace} = \left(I - \beta F_{replace}\right)\left(I - \beta F_{keep}\right)^{-1}$,

$$Q\left(A, \widetilde{A}\right) = \left[A_{replace} \otimes I,\, I \otimes I,\, A'_{replace} \otimes A_{replace},\, I \otimes A_{replace}\right],$$

and $\nabla h\left(\pi\right)$ is defined in footnote 17. Finally, with $\beta = .99$ and $\lambda\left(x\right) = 0.1$ for all x (representing a 10% increase in replacement costs), we have $\left\|Q\left(A, \widetilde{A}\right) vec\left(\nabla h\left(\pi\right)\right)\right\| = 1.89$, where $\|.\|$ is the matrix 2-norm. Equation (18) is violated, implying the counterfactual is not identified.

3.3 Identification of Counterfactual CCP: Affine Transformations

In the common case of affine counterfactuals, $\widetilde{\pi} = \mathcal{H}\pi + g$, Theorem 1 and Corollary 1 simplify to:

Corollary 2 *(Affine Counterfactual) Assume $\widetilde{\pi} = \mathcal{H}\pi + g$. The counterfactual CCP $\widetilde{p}$ is identified if and only if for all π satisfying* (5) *and all $a \in \widetilde{\mathcal{A}}$, $a \neq J$,*

$$\sum_{l \in \mathcal{A}, l \neq J} \left(H_{al} - \widetilde{A}_a H_{Jl} \right) A_l + H_{aJ} - \widetilde{A}_a H_{JJ} = 0. \tag{20}$$

When $\widetilde{\pi}_a$ is "action diagonal," $\widetilde{\pi}_a = H_a \pi_a + g_a$, the above condition becomes

$$H_a A_a = \widetilde{A}_a H_J. \tag{21}$$

Corollary 2 provides a simple-to-verify set of minimal sufficient conditions to identify $\widetilde{p}$. Indeed, it is possible to test if $\widetilde{p}$ is identified by testing (20) given data on F.[22]

We now turn to some further special cases of interest. First, consider the "additive transfers" counterfactual: $\widetilde{\pi} = \pi + g$. It is obvious from (21) that $\widetilde{p}$ is identified in this case:

Corollary 3 *("Additive Transfers" Counterfactual) When $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{F} = F$, and $\widetilde{\pi} = \pi + g$, the counterfactual CCP $\widetilde{p}$ is identified.*

This result was first shown by Aguirregabiria (2010).[23] Here, we note that, more generally, adding a known vector to $\widetilde{\pi}$ does not affect the Jacobian matrix of h, even for nonlinear functions, and so whether $\widetilde{p}$ is identified or not does not depend on this vector.

The next corollary is also immediate:

Corollary 4 *("Action Diagonal" Counterfactual) Suppose $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{F} = F$, and $\widetilde{\pi}_a = H_a \pi_a$ for all $a \in \mathcal{A}$. Then:*

(i) To identify the counterfactual CCP, it is necessary that H_a, all a, are similar matrices.

(ii) Further, if H_a is diagonal for all a, then to identify $\widetilde{p}$ it is necessary that $H_a = H$, all a, and that $H = A_a H A_a^{-1}$.

Corollary 4 places a strong restriction on H_a: they have to be similar matrices. This must be the case whether or not H_a are diagonal. For instance, a "change in types" counterfactual that replaces preference parameters of one type of agents by those of another type (in which case H_a are not diagonal), must satisfy Corollary 4(i) to identify $\widetilde{p}$.

For the "proportional changes" counterfactual, we have diagonal H_a and $g = 0$. In this case, Corollary 4(ii) states that if we change the payoff of action a in state x by $\lambda(x)$,

[22] The matrices H_{al} are stipulated by the counterfactual; the matrices A_a and $\widetilde{A}_a$ depend on transition probabilities F_a and $\widetilde{F}_a$; and $\widetilde{F} = h^s(F)$ is known for any known function h^s and any F. We can therefore take the test statistic to be $T = \left\| \sum_{l \neq J} (H_{al} - \widetilde{A}_a H_{Jl}) A_l + H_{aJ} - \widetilde{A}_a H_{JJ} \right\|$, where $\|.\|$ is a matrix norm, and test if T is sufficiently close to zero. Its asymptotic distribution can be obtained from the asymptotic distribution of the estimator $\widehat{F}$ combined with the delta method. We leave the investigation of the asymptotic properties of such a test for future research.

[23] Aguirregabiria (2010) proved this result for finite-horizon models. Aguirregabiria and Suzuki (2014) and Norets and Tang (2014) showed the result extends to infinite-horizon models.

$\widetilde{\pi}(a,x) = \lambda(x)\pi(a,x)$, then we also need to change the payoff of any other action a in state x by exactly the same proportion $\lambda(x)$. Otherwise, the counterfactual CCP is not identified. Furthermore, restrictions on the transition matrices F_a are also necessary, as they must satisfy the equality $H = A_a H A_a^{-1}$.

Corollary 5 *("Proportional Changes" Counterfactual) Suppose $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{F} = F$, and $\widetilde{\pi}_a = H\pi_a$ for all $a \in \mathcal{A}$, with H diagonal.*

(i) If $H = \lambda I$, then $\widetilde{p}$ is identified.

(ii) If H has pairwise distinct elements, to identify $\widetilde{p}$ it is necessary that the transition probabilities are action invariant.

(iii) If H has a simple eigenvalue, a necessary condition to identify $\widetilde{p}$ is that the row of each F_a corresponding to the simple eigenvalue has to be invariant across a.

Corollary 5 shows that the "proportional changes" counterfactual is identified *only if* proper restrictions on the transition matrices F_a hold. The degree of freedom of these restrictions is controlled by the number of distinct elements of the diagonal matrix H. For example, if all elements of H are pairwise distinct, $\widetilde{p}$ is not identified unless F_a are the same for all actions. Action-invariant transition probabilities, however, effectively imply that the agent is not facing a dynamic problem.

Another consequence of Corollary 4 is that if a counterfactual leaves π_J and F_J unaffected (i.e. $\widetilde{\pi}_J = \pi_J$ and $\widetilde{F}_J = F_J$), but imposes $\widetilde{\pi}_a = H_a\pi_a$, with H_a diagonal for some action $a \neq J$, then, $\widetilde{p}$ is identified only if $H_a = I$. In words, if a counterfactual does not alter one action, but changes π_a proportionally for some other action a, it is not identified.[24]

3.4 Changes in Transitions and Sets of Actions and States

We now consider counterfactuals that change only transition probabilities, $\widetilde{F} \neq F$, or the set of actions and states, $\widetilde{\mathcal{A}} \neq \mathcal{A}$, $\widetilde{\mathcal{X}} \neq \mathcal{X}$.

Corollary 6 *("Change in Transition" Counterfactual) Suppose $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = X$, and $\widetilde{\pi}_a = \pi_a$ for all $a \in \mathcal{A}$, but $\widetilde{F} \neq F$. Then $\widetilde{p}$ is identified if and only if $A_a = \widetilde{A}_a$, for all $a \in \widetilde{\mathcal{A}}$, $a \neq J$.*

This result is also documented by Aguirregabiria and Suzuki (2014), Norets and Tang (2014) and Arcidiacono and Miller (2015). It is an immediate consequence of Corollary 2. This result says that counterfactuals in which the transition matrices change are typically not identified except in knife-edge cases. Here we add another non-identification result with a knife-edge qualification: when both transitions and payoffs change, the two must change together in a way that satisfies (19) or else counterfactual behavior is not identified.

Next, we consider counterfactuals that only change the set of actions (and possibly of states too). First we present counterfactuals that eliminate one action; then we turn to counterfactuals that add a new action. Extensions to eliminating or adding more than one action are straightforward.

[24] To see why, recall that identification requires $H_a A_a = \widetilde{A}_a H_J$ or, since $H_J = I$, $H_a = (I - \beta\widetilde{F}_a)(I - \beta F_a)^{-1}$. For any two stochastic matrices, $(I - \beta\widetilde{F}_a)(I - \beta F_a)^{-1}1 = 1$, where 1 is a vector of ones. Let λ be the vector of diagonal elements of H_a. Then, $H_a 1 = \lambda = 1$. Therefore, $H_a = I$.

Corollary 7 *("Eliminate an Action" Counterfactual) Suppose $\widetilde{\mathcal{A}} = \mathcal{A} - \{j\}$, where j is the action to be eliminated. If $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{F}_a = F_a$, and $\widetilde{\pi}_a = \pi_a$ for all $a \in \widetilde{\mathcal{A}}$, then $\widetilde{p}$ is identified.*

In other words if a counterfactual just eliminates an action, it is identified. Here, it is key that transitions do not change; however, elimination of an action often implies elimination of some states (e.g. the monopolist entry/exit example) 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 $\widetilde{\mathcal{X}}$. In Appendix A, we provide the necessary and sufficient conditions for identification in this case (see Lemma 8). Below, we consider the special case in which $x = (k, w)$, with transition probabilities given by (6) and $k_t = a_{t-1}$ (similar to the model presented in Section 2.2, but with no parametric assumptions on π). The counterfactual CCP is indeed identified in this case.

Proposition 3 *("Eliminate an Action and States" Counterfactual) Suppose $\widetilde{\mathcal{A}} = \mathcal{A} - \{j\}$, where j is the action to be eliminated. Without loss, let the set of states be $\widetilde{\mathcal{X}} = \{1, ..., \overline{x}\}$ and $\mathcal{X} = \{1, ..., \overline{x}, \overline{x}+1, ..., X\}$. Assume $\widetilde{\pi}_a = H_a \pi_a$ with $H_a = [I_{\overline{x}}, 0]$ for all $a \in \widetilde{\mathcal{A}}$. 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.*

Finally, we consider a counterfactual that only adds an action to $\mathcal{A}$. This case requires prespecifying $\widetilde{\pi}_j$ and $\widetilde{F}_j$ for the new action.

Corollary 8 *("Add an Action" Counterfactual) Suppose $\widetilde{\mathcal{A}} = \mathcal{A} \cup \{j\}$, where j is the new action. Assume $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{F}_a = F_a$, $\widetilde{\pi}_a = \pi_a$ for all $a \in \mathcal{A}$ and*

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

Then $\widetilde{p}$ is identified if and only if

$$\sum_{a \in \mathcal{A}} H_{ja} 1 = 1,$$

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

where 1 is an $X \times 1$ vector of ones.

In other 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.

3.5 Restrictions in Non-Identified Counterfactuals

So far we have explored whether counterfactual behavior is identified when the researcher does not impose any extra restriction on the model to identify payoffs. Next, we consider the following natural question: what is the minimum number and form of payoff restrictions that one has to impose to secure identification of counterfactual CCPs?

Consider a set of $d \leq X$ linearly independent payoff restrictions:

$$R\pi = r \tag{22}$$

with $R \in \mathbb{R}^{d \times AX}$, or in block-form, $R = \begin{bmatrix} R_{-J} & R_J \end{bmatrix}$, e.g. exclusion restrictions. Note that if $d < X$, the restrictions are not sufficient to identify π. We can rewrite (22) as:

$$R_{-J}\pi_{-J} + R_J\pi_J = r \tag{23}$$

and combine it with out main relationship, $\pi_{-J} = A_{-J}\pi_J + b_{-J}(p)$:

$$\left(R_{-J}A_{-J} + R_J\right)\pi_J = r - R_{-J}b_{-J}\left(p\right).$$

This is of the form:

$$Q\pi_J = q \tag{24}$$

with $Q = R_{-J}A_J + R_J \in \mathbb{R}^{d \times X}$ and $q = r - R_{-J}b_{-J}\left(p\right) \in \mathbb{R}^d$.

Now consider, for simplicity, an affine counterfactual, $\widetilde{\pi} = \mathcal{H}\pi + g$. As shown in Section 3.3, if equation (20) of Corollary 2 holds, the counterfactual CCPs are identified with no extra restrictions (R, r). Stack the left hand side of (20) for all a and denote the resulting matrix by C. For instance, in the case of action-diagonal counterfactuals (assuming $J = A$),

$$C \equiv \begin{bmatrix} H_1 & \ldots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \ldots & H_{J-1} \end{bmatrix} A_{-J} - \widetilde{A}_{-J}H_J,$$

or, more compactly, $C = H_{-J}A_{-J} - \widetilde{A}_{-J}H_J$, where C has dimension $(A-1)X \times X$. So, if (20) holds, $C = 0$ and $\widetilde{p}$ is identified regardless of extra restrictions (R, r). At the other extreme, if restrictions (22) are correctly specified with $d = X$, then the model is identified, and so is $\widetilde{p}$ for any C.

The following proposition and corollary shed light on the form of the linear restrictions that need to be imposed to obtain identification.

Proposition 4 *Given the matrix C formed by a counterfactual transformation $\mathcal{H}$ and the linear restrictions (22), the counterfactual CCP is identified if and only if there exists an $(A-1)X \times d$ matrix M such that $C = MQ$.*

Corollary 9 *If the restrictions are linearly independent (i.e. $rank\left(Q\right) = d$) then the counterfactual CCP is identified if and only if*

$$C\left(I - Q'\left(QQ'\right)^{-1}Q\right) = 0 \tag{25}$$

and necessarily $rank\left(C\right) \leq d$.

The above results establish a direct relationship between the counterfactual transformation $\mathcal{H}$ and the restrictions Q. Take a counterfactual $\mathcal{H}$, form the matrix C and compute its rank. Suppose $rank\left(C\right) = d$. Proposition 4 implies that d provides the minimum number of linearly independent restrictions needed to secure identification of $\widetilde{p}$, even when $C \neq 0$ and with restrictions that would not identify the model. Second, the results inform us on the nature of the required restrictions: the proposition tells us that each of these restrictions must be a linear combination of the rows of C.[25] Finally, Corollary 9 indicates a simple way to check in practice whether a combination of restrictions and counterfactual transformation identifies $\widetilde{p}$.[26]

3.6 Identification of Counterfactual CCP for Parametric Models

Parametric restrictions can aid identification of some parameters; can they also enlarge the set of counterfactuals that are identified? We show that they can. Consider again the model presented in Section 2.2. Recall that $\theta_0(a,k)$ is not identified, while $\theta_1(a,k)$ is identified when there is "sufficient variation" in w.

As previously mentioned, a counterfactual may change the term $\theta_0(a,k)$, or the term $R'\theta_1$, or both. We show here that transformations in $R'\theta_1$ result in identified counterfactual CCPs, while transformations in $\theta_0(a,k)$ may not. This is fairly intuitive. When $\theta_1(a,k)$ is identified, a counterfactual that changes $R'\theta_1$ resembles an additive transfer, and as shown in Subsection 3.3, "additive transfers" counterfactuals are always identified. 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 CCP is identified or not.

To illustrate counterfactuals that change $\theta_0(a,k)$, we consider the case of affine action-diagonal counterfactuals as an example. In particular, consider the following transformation:

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

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 function of θ_0 is straightforward; the only requirement is that the function is differentiable.

Proposition 5 *(Parametric Model, Change in Payoffs Counterfactual) Assume* $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = X$, *and that the conditions of Proposition 2 hold.*

(i) The counterfactual CCP corresponding to a counterfactual that only changes the term $R'(a,w)\theta_1(a,k)$ *of* $\pi(a,k,w)$ *is identified.*

(ii) The counterfactual CCP corresponding to the transformation (26) is identified if and only if for all $a \neq J$

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

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

Similar results apply to changes in transitions $F_a = F^w \otimes F_a^k$, as the next corollary states.

Corollary 10 *(Parametric Model, Change in Transition Counterfactual) Assume* $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = X$, *and that the conditions of Proposition 2 hold. The counterfactual CCP corresponding*

---

[25] Indeed, to see this, pick d linearly independent rows of C and place them in the matrix $\widetilde{C}$. Then, the set of allowable restrictions consists of all $d \times X$ matrices of the form $Q = \widetilde{M}\widetilde{C}$ where $\widetilde{M}$ is a $d \times d$ invertible matrix.

[26] One can view the results in this section from a different perspective. Suppose that the restrictions (22) are natural and a researcher considers them to be true. For instance, consider the constraint that $\pi(\text{exit}, \text{not active}) = 0$ in our monopolist example. Proposition 4 and Corollary 9 then informs us on what can be attained by relying only on such natural constraints and allows us to explore the range of possibilities for identifiable counterfactuals.

to a counterfactual that only changes the transition F^w is identified, but the counterfactual CCP corresponding to a change in F_a^k is identified if and only if $A_a^k = \widetilde{A}_a^k$, where A_a^k is defined similarly to A_a with F_a^k in the place of F_a, for all $a \in \widetilde{\mathcal{A}}$.

In a nonparametric setting, changes in the transition process result in non-identified counterfactual behavior. However, Corollary 10 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 monopolist 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 (2007), Collard-Wexler (2013)) and so these are covered by us.

3.7 Identification of Counterfactual Welfare

Finally, we discuss identification of counterfactual welfare and provide the minimal set of sufficient conditions to identify the magnitude of welfare changes. For simplicity, we only consider affine action-diagonal counterfactuals; i.e. $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = X$, and $h_a(\pi_a) = H_a\pi_a + g_a$, 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 and V is the true value functions, respectively.

Proposition 6 *(Welfare) Assume $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = X$, and $h_a(\pi_a) = H_a\pi_a + g_a$, all a. The welfare difference ΔV is identified if, for all $a \neq J$,*

$$H_a A_a - \widetilde{A}_a H_J = 0,$$

and

$$H_J = \left(I - \beta \widetilde{F}_J\right)(I - \beta F_J)^{-1}. \tag{27}$$

Proposition 6 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: $H_J = I$ and $\widetilde{F}_J = F_J$. For "proportional changes" counterfactuals the two conditions are satisfied only when all matrices H_a equal the identity matrix; i.e. $\widetilde{\pi} = \pi$, which is equivalent to saying that ΔV is not identified in this case. On a positive note, an immediate implication of Proposition 6 is that the welfare effect of an "additive transfers" counterfactual is identified. "Additive transfers" counterfactuals are robust to nonidentification of the model primitives: both $\widetilde{p}$ and ΔV are identified.

Proposition 6 is an immediate consequence of Lemma 9 in the Appendix. The Lemma provides the full set of necessary and sufficient conditions to identify ΔV, and shows that identification of $\widetilde{p}$ together with (27) are "almost" necessary to identify welfare.[27]

[27] "Almost" here is in the following sense: when at least one of the two conditions does not hold, the nonlinear system of equations necessary to identify ΔV is generically inconsistent. So, in practice, identification of ΔV is unlikely when either $\widetilde{p}$ is not identified or (27) fails.

Finally, the next corollary considers identification of ΔV for the parametric model of Section 2.2. As expected, identification is guaranteed when counterfactuals change $R'\theta_1$ and/or F^w.

Corollary 11 *(Welfare, Parametric Model) Assume the conditions of Proposition 2 hold. Suppose $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = X$, 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

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

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 $R'(a,w)\theta_1(a,k)$ and F^w.

4 Numerical Example: Monopolist Entry and Exit Problem

This section illustrates some of our theoretical results using the monopolist entry/exit problem of Example 2. We assume that w is observable and captures aggregate demand fluctuating over time. Variable profits $\overline{\pi}(w_t)$ are determined by static profit maximization over quantity: the monopolist faces the demand curve $P_t = w_t - \eta Q_t$ and has constant marginal cost c, so that $\overline{\pi}(w_t;\eta,c) = (w_t - c)^2/4\eta$. We assume w_t follows a first-order Markov process, $\Pr(w_{t+1}|w_t)$; this is the only source of uncertainty in the model. To simplify, we assume that demand can be high, medium or low: $w \in \left\{w^H, w^M, w^L\right\}$.

We assume the econometrician knows (or estimates): (i) the true CCP, $\Pr(active|k,w)$; (ii) the transition probabilities for demand shocks, $\Pr(w_{t+1}|w_t)$; and (iii) the variable profits $\overline{\pi}(w)$, estimated outside of the dynamic problem using price and quantity data.[28]

First, we solve the true model and obtain the baseline CCPs and value functions. Then, we recover π using Proposition 1 and imposing two different identifying restrictions: the first fixes the scrap value to zero, $\phi^s = 0$, so that $\pi(inactive,k,w) = 0$ for all (k,w); identification of π follows directly from (4). The second restriction assumes the fixed cost is zero, $fc = 0$. As $\pi(active,1,w) = \overline{\pi}(w)$ is assumed known, this is sufficient to recover the remaining elements of π.

Table 1 presents the true and the two estimated payoff functions. Note that under the first restriction ($\phi^s = 0$) entry costs change sign. This is because if we wrongly fix $\pi(inactive,k,w) = 0$, then $\pi(active,k,w)$ must capture all desired data patterns, which can turn a true positive number (entry costs) into an "identified" negative number. If there is no scrap value, entering the market becomes less attractive and entry costs must become low (in fact, negative) to capture the observed entry patterns.

[28] We ignore sampling variation for simplicity and set: $c = 11, \eta = 1.5, w = (20,17,12), \beta = 0.95, fc = 5.5, \phi^s = 10, \phi^e = 9$, while the transition matrix for w is

$$F\left(w'|w\right) = \begin{bmatrix} 0.4 & 0.35 & 0.25 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}.$$

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
$\pi(a, k=1, w_H) = \phi^s$	10	0	120
$\pi(a, k=1, w_M) = \phi^s$	10	0	120
$\pi(a, k=1, w_L) = \phi^s$	10	0	120
$a =$ active			
$\pi(a, k=0, w_H) = -\phi^e$	-9	0.5	-113.5
$\pi(a, k=0, w_M) = -\phi^e$	-9	0.5	-113.5
$\pi(a, k=0, w_L) = -\phi^e$	-9	0.5	-113.5
$\pi(a, k=1, w_H) = \overline{\pi}(w_H; \eta, c) - fc$	8	7.5	13.5
$\pi(a, k=1, w_M) = \overline{\pi}(w_M; \eta, c) - fc$	0.5	0	6
$\pi(a, k=1, w_L) = \overline{\pi}(w_L; \eta, c) - fc$	-5.33	-5.83	0.167

Under the second restriction ($fc = 0$), there is no change in the sign of payoffs, but both entry costs and scrap values are considerably larger in magnitude than their true values. To see why, consider first the case in which the firm is already in the market ($k = 1$). For a given probability of being active observed in the data, fixing $fc = 0$ implies higher profits when active, which gives incentives to stay more often in the market. To match the observed CCP, scrap values must increase to provide incentives to exit. Similarly, when the firm is out ($k = 0$), increasing profits when active increases incentives to enter. Entry costs must then increase to compensate for this incentive.

Given the recovered payoffs, we implement four counterfactuals and compare the true 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: $\widetilde{\pi}(active, 0, w) = \pi(active, 0, w) + g$. Counterfactual 2 is a proportional entry subsidy: $\widetilde{\pi}(active, 0, w) = H\pi(active, 0, w)$. In both cases we leave $\pi(inactive, k, w)$ and F unchanged and we choose the additive and proportional subsides so that the true counterfactual CCP and welfare are the same.[29] As shown in Section 3, while the counterfactual CCPs and welfare are identified in the first scenario, they are not identified in the second scenario.

Table 2 presents the results for both counterfactuals 1 and 2 for the true model and the two estimated models. In both scenarios, the true counterfactual probability of entering increases compared to the baseline because of the subsidy; and the probability of staying in the market decreases because it is cheaper to re-enter in the future. So, the monopolist enters and exits more often than the baseline case. In counterfactual 1 (additive subsidy), as expected, the counterfactual CCPs and welfare are identical in the true model and under both restricted models.

[29] As $\pi(active, 0, w) = -\phi^e$, and we choose the true $\phi^e = 9$, we opt for the additive transfer $g = 0.9$, and the proportional change $H = 0.9$, 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 Prop Entry Subsidies

States: (k, w)	**Baseline**	**True CF**	**Estimated CF** *scrap value = 0*	**Estimated CF** *fixed cost = 0*
			CF1: $\widetilde{\pi}_0 = \pi_0, \widetilde{\pi}_1 = \pi_1 + g$	
CCP: $\Pr(active\|x)$				
$(k = 0, w_H)$	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: $\widetilde{\pi}_0 = \pi_0, \widetilde{\pi}_1 = H\pi_1$	
CCP: $\Pr(active\|x)$				
$(k = 0, w_H)$	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

Counterfactual 2 (proportional entry subsidy) results in very different outcomes under the two restrictions. In the first restriction, $\phi^s = 0$, the changes in the CCPs are all in the wrong direction: while the true probability of entering increases relative to the baseline, the predicted counterfactual probability of entering decreases. Similarly, the counterfactual probability of exiting 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 identified entry cost under this restriction has the wrong sign: in the true model, multiplying $\pi(active, 0, w)$ by H 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.[30]

Under the second identifying restrictions ($fc = 0$), both entry costs and scrap values have the correct sign, but are magnified. As a result, it is profitable to enter and exit the market repeatedly when the entry cost is reduced by 10% in the counterfactual scenario. Predicted turnover is therefore excessive and predicted welfare is also exaggerated.

Counterfactual 3 changes the transition process $\Pr(w_{t+1}|w_t)$. Because $\overline{\pi}(w)$ is known, the counterfactual behavior and welfare are identified (Proposition 5 and Corollaries 10 and 11).[31] Top panel of Table 3 confirms the results.[32]

Finally, counterfactual 4 implements a "change in types" counterfactual. 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 profit. 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.[33]

The bottom panel of Table 4 presents the results. Similar to counterfactuals 1 and 2, turnover increases in the true counterfactual compared to the baseline; and again, the two identifying restrictions generate very different outcomes. This is expected since the matrices H_a are not similar; so by Corollary 4(i), counterfactual behavior is not identified.

Under the first restriction ($\phi^s = 0$), counterfactual CCPs and welfare are all in the right direction, even though entry costs have the wrong sign in both markets. This happens because replacing the market 1 entry cost by the market 2 entry cost amounts to an increase in entry costs in the restricted model. 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 restrictions ($fc = 0$), turnover and welfare are again exag-

[30] Even though a subsidy implemented on the estimated model generates a counterfactual CCP that can also be obtained with *some* subsidy implemented on the true model, the researcher cannot possibly know the value of such a subsidy unless she knows the true payoff function.

[31] 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, this counterfactual is not identified.

[32] We set $\widetilde{\Pr}(w'|w) = 1/3$, for all (w', w). The qualitative results hold for other transformations on the transition process $\Pr(w'|w)$. We also implemented a counterfactual that reduces the marginal cost of production, c, by 10% (not shown in the paper). Again, because $\overline{\pi}(w)$ is known, this counterfactual is identified (Proposition 5 and 11).

[33] Market 1 is the same as before. For market 2, we set: $c_2 = 9, \eta_2 = 1.7, w_2 = (18, 15, 11), fc_2 = 3, \phi_2^s = 8, \phi_2^e = 6$. The discount factor and transition matrix in market 2 is the same as in market 1. The estimated profit under the first restriction ($\phi_2^s = 0$) is: $\phi_2^e = 1.6, \pi^2(active, 1, w) = (8.52, 1.89, -2.82)$; and under the second restriction ($fc_2 = 0$) is: $\phi_2^e = -63, \phi_2^s = 68, \pi^2(active, 1, w) = (11.91, 5.29, 0.59)$.

Table 3: Counterfactuals 3 and 4 – Change in F(w'/w) 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: $\Pr(active \mid x)$				
$(k = 0, w_H)$	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: $\Pr(active \mid x)$				
$(k = 0, w_H)$	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

gerated, to the point that counterfactual choice probabilities are either zero or one.

5 Extensions

With our agricultural land use application in mind, we present two extensions to the framework. The first extension incorporates extra data on resale markets that, under some conditions, can help identify payoff functions and counterfactuals. The second extension allows for unobserved market-level state variables and we discuss identification of payoffs in this context.

5.1 Identification of Payoffs using Resale Market Data

The sensitivity of certain counterfactuals to identifying restrictions on payoffs calls out for some means to assess the accuracy of these restrictions. In this section, we consider how adding data on resale prices can allow us to bypass identifying restrictions. Suppose that the agent's dynamic problem involves a durable asset whose value in state x_{it} is given by $V(x_{it})$. Sometimes, one can obtain data on asset prices, such as firm acquisition prices (e.g. price of ships in the bulk shipping industry), or real estate values from transactions or appraisals. Such asset prices plausibly contain information about the expected discounted sum of future returns associated with that asset (i.e. $V(x_{it})$). Because value functions are the key unobserved object that makes dynamic models difficult to estimate, value function measurement is very powerful.[34]

There are numerous ways to model resale markets, and different models may imply different mappings between transaction prices and agents' value function. Here, we 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. As agents have the same valuation for the asset, $V(x_{it})$, a seller is willing to sell it at price p_{it}^{RS} only if $p_{it}^{RS} \geq V(x_{it})$; similarly, a buyer is willing to buy if $p_{it}^{RS} \leq V(x_{it})$. In this setup, the equilibrium resale price of asset i in state x_{it} must equal its value and agents are always indifferent between selling the asset or holding on to it:

$$p_{it}^{RS} = V(x_{it}). \tag{28}$$

Combining (28) with the system (1)-(3), one can see that value functions immediately inform us on payoffs and their shape; in fact, they deliver payoffs nonparametrically (see Kalouptsidi (2014*a*) and (2014*b*) for an implementation in the context of shipping). Using (28) we can estimate $V(x_{it})$ by (nonparametrically) regressing p_{it}^{RS} on x_{it}. The only difficulty is that, because V is measured with respect to a specific scale (e.g. dollars), a restriction on the scale parameter σ is no longer innocent. Fortunately, as shown in Proposition 7 below, with little information on payoffs for one action in any state (or on average), we can identify σ.

Proposition 7 *Given the joint distribution of observables* $\Pr(y)$, *where* $y_{it} = \left(a_{it}, x_{it}, p_{it}^{RS}\right)$, *the flow payoffs* π_a *are identified provided the primitives* (β, G) *are known and (i)* σ *is known,*

[34] In a different context, Heckman and Navarro (2007) make use of extra data on labor outcomes, such as future earnings, to obtain identification.

or (ii) the (cardinal) payoff $\pi_a(x)$ is known for some (a, x); or (iii) the expected payoff for one action is known.

We close this subsection by noting that it is crucial that a rich set of asset characteristics are observed and that we are in a world of thick resale markets where the owner and asset heterogeneity do not interact much – we provide evidence for these in our land use application (see Appendix B and C). Aguirregabiria and Suzuki (2014) consider a more general bargaining model with resale costs and show that our result below generalizes to that case. Further generalizations incorporating other frictions on resale markets are also possible, but are beyond the scope of this paper.

5.2 Identification of Payoffs with Unobservable Market-level States

In this section, we consider the identification of models when there are unobservable market-level state variables. This setting is relevant for many applications, such as when the available observed market states are insufficient to capture the agents' full information set, or where it is difficult to capture the evolution of market states adequately, or when the dimensionality of the state space is large. This is the case for our land use application in Section 6.[35]

We assume that there are $m = 1, ..., M$ markets. In addition, we maintain the state decomposition $x_{imt} = (k_{imt}, \omega_{mt})$ from Section 2.2. Importantly, the aggregate state ω_{mt} is not fully observed; but it does have an observed component w_{mt}. Note the transition of k_{imt} can be recovered from the data even though ω_{mt} is not fully observed. Indeed, one can estimate $F^k_{mt}(k_{imt+1}|a, k_{imt}) = F^k(k_{imt+1}|a, k_{imt}, \omega_{mt})$ for each time period in each market with a rich cross section of agents.[36]

We add two new assumptions here. First, we restrict the unobserved aggregate state to enter payoffs in an additively separable fashion:

$$\pi(a, k_{imt}, \omega_{mt}) = \overline{\pi}(a, k_{imt}, w_{mt}) + \xi(a, k_{imt}, \omega_{mt}), \tag{29}$$

where $\xi(a, k, \omega)$ has zero mean. The unobservable $\xi(a, k, \omega)$ may capture mismeasured profits or unobservable costs. It is important to stress that ξ is a function of state variables, not a state variable itself. On its own, ξ need not evolve according to a first-order Markov process although ω does. Note that ξ may be serially correlated, unlike the idiosyncratic shocks ε. In addition, $\overline{\pi}$ and ξ are likely correlated because they may depend on the same state variables. For this reason, we need to make use of instrumental variables to obtain identification.

[35] Hu and Shum (2012) also consider identification under unobserved states. Our results are neither a special case of nor more general than theirs. Hu and Shum suggest the following indirect approach: (i) identify and estimate state transitions, then (ii) use a nested fixed-point maximum likelihood (as in Rust (1987)), or a two-step estimator (as in Hotz-Miller (1993)) to estimate payoffs. Their focus thus is on (i). Their results rely on the following assumptions: (i) the unobserved state is scalar; (ii) the unobserved state is realized before the observed state; (iii) in the case of continuous unobserved states, actions are continuous as well; (iv) certain invertibility conditions are required, which are not a priori testable and checking them involves nontrivial computations. In contrast, our proposed method does not need these restrictions and takes a direct approach: it deals with the identification and estimation of payoffs directly. However, we need to impose additively separable payoffs (see below), and we are not able to recover individual level unobserved heterogeneity in our theorem (although that is possible in practice). We view our results as complements rather than substitutes.

[36] We assume k_{imt} is finite, as in the case of fully observed states. We allow ω_{mt} to be continuous. Neither assumption is important and our results apply to both discrete and continuous states.

Access to valid instrumental variables is our second new assumption. Formally, we assume there exist instruments at the time-t information set, z_{mt}, such that (i)

$$E\left[\xi\left(a,k,\omega_{mt}\right)|z_{mt}\right]=0,$$

for all a, k; and that (ii) for all functions $q\left(w_{mt}\right)$, $E\left[q\left(w_{mt}\right)|z_{mt}\right]=0$ implies $q\left(w_{mt}\right)=0$ (the completeness condition). If it is reasonable to assume that $\overline{\pi}$ and ξ are not correlated, one can take observed state variables w_{mt} as instruments. In other cases, it may be reasonable to use (sufficiently) lagged w_{mt}.

The available dataset now is $y=\{(a_{imt},k_{imt},w_{mt},z_{mt}):i=1,...,N;m=1,...,M;t=1,...,T\}$. Identification with partially observed states cannot make direct use of the main equations (1)-(3). Our identification results are instead based on the following expression which replaces (1)-(3):

$$\begin{aligned}&\pi\left(a,k_{imt},\omega_{mt}\right)+\beta\varepsilon^{V}\left(a,k_{imt},\omega_{mt},\omega_{mt+1}\right)\\=\;&V\left(k_{imt},\omega_{mt}\right)-\beta\sum_{k'}V\left(k',\omega_{mt+1}\right)F^{k}\left(k'|a,k_{imt},\omega_{mt}\right)-\sigma\psi_{a}\left(k_{imt},\omega_{mt}\right)\end{aligned}\tag{30}$$

where $\varepsilon^{V}\left(\cdot\right)$ is an "expectational error" according to the following definition:

Definition 1 *(Expectational error) For any function $\zeta\left(k,\omega\right)$ and particular realization $\omega^{*}\in\Omega$,*

$$\begin{aligned}\varepsilon^{\zeta}\left(k',\omega,\omega^{*}\right)&\equiv E_{\omega'|\omega}\left[\zeta\left(k',\omega'\right)|\omega\right]-\zeta\left(k',\omega^{*}\right),\\\varepsilon^{\zeta}\left(a,k,\omega,\omega^{*}\right)&\equiv\sum_{k'}\varepsilon^{\zeta}\left(k',\omega,\omega^{*}\right)F^{k}\left(k'|a,k,\omega\right).\end{aligned}$$

To derive (30) note that $E\left[V\left(k_{imt+1},\omega_{mt+1}\right)|a,k_{imt},\omega_{mt}\right]$ is given by:

$$\begin{aligned}&\sum_{k'}\int_{\omega'}V\left(k',\omega'\right)dF^{\omega}\left(\omega'|\omega_{mt}\right)F^{k}\left(k'|a,k_{imt},\omega_{mt}\right)\\=\;&\sum_{k'}\left(E_{\omega'|\omega_{mt}}\left[V\left(k',\omega'\right)|\omega_{mt}\right]\right)F^{k}\left(k'|a,k_{imt},\omega_{mt}\right)\\=\;&\sum_{k'}V\left(k',\omega_{mt+1}\right)F^{k}\left(k'|a,k_{imt},\omega_{mt}\right)+\varepsilon^{V}\left(a,k_{imt},\omega_{mt},\omega_{mt+1}\right)\end{aligned}\tag{31}$$

The first term of the right hand side of (31) is the expected ex ante value function at time $t+1$ for agent i in state k_{imt} who selected action a at time t for the *actual realization* of ω_{mt+1} (the conditional expectation is taken over k'). The use of expectational errors allow us to rewrite (1) as follows: for $a=1,...,J$

$$\begin{aligned}\pi\left(a,k_{imt},\omega_{mt}\right)&=v_{a}\left(k_{imt},\omega_{mt}\right)-\beta\sum_{k'}V\left(k',\omega_{mt+1}\right)F^{k}\left(k'|a,k_{imt},\omega_{mt}\right)\\&\quad-\beta\varepsilon^{V}\left(a,k_{imt},\omega_{mt},\omega_{mt+1}\right).\end{aligned}\tag{32}$$

which in turn leads to (30). Importantly, the expectational error is mean independent of all past state variables (k,ω) (see Lemma 10 in Appendix A).

Before turning to identification, we note that even though ω_{mt} is not fully observed, we can still recover the conditional choice probabilities $p_{a}\left(k,\omega_{mt}\right)$. Like F^{k} they can be estimated separately for each market m in each t (or with flexible market and time dummies). Of course we need a large number of agents i in each m and t to obtain accurate estimates. For our results regarding the identification of π in settings with unobserved states, we treat $p_{at}(.)$, $F_t^k(.)$ and $\psi_{at}(.)$ as known objects.[37]

Next, we simplify the notation and use (m,t) subscripts to denote functions that depend on ω_{mt}. We rewrite payoffs as $\pi_{mt}(a, k_{imt}) = \pi(a, k_{imt}, \omega_{mt})$, while $V_{mt}(k_{imt})$, $p_{amt}(k_{imt})$ and $\psi_{amt}(k_{imt})$ are similarly defined. Therefore, (30) in matrix form becomes:

$$\pi_{amt} + \beta\varepsilon^V_{am,t,t+1} = V_{mt} - \beta F^k_{amt} V_{mt+1} - \sigma\psi_{amt} \tag{33}$$

for all a, where $\varepsilon^V_{am,t,t+1}$ stacks $\varepsilon^V_{mt,t+1}(a,k)$ for all k.

We now turn to our identification results. In the case of the fully observed states, V_{mt} and V_{mt+1} represented the same vector V, and we could solve for V after inverting the $\left(I - \beta F^k_{amt}\right)$ matrix. Now, V_{mt} and V_{mt+1} differ because of the unobservable state ω, and we cannot solve for V in the same way. Consequently, obtaining identification is more difficult here and we must impose more restrictions than in a setting with fully observable states.

Terminal or renewal actions are natural candidates. Terminal actions terminate the decision making and impose a finite horizon often facilitating estimation considerably (e.g. a worker retires, a mortgage owner defaults). Renewal actions also facilitate estimation. Formally, if action J is a renewal action, then for all t, τ and all a, j:

$$F^k_{amt} F^k_{Jm\tau} = F^k_{jmt} F^k_{Jm\tau}. \tag{34}$$

Intuitively, choosing J at any time after today's actions leads to the same distribution of states (e.g. replacing the bus engine (Rust (1987)) or planting crops (Scott (2013)).

We now return to (33) and exploit the fact that J is a renewal action. For any a and j, we then have

$$\begin{aligned} & \pi_{amt} - \pi_{jmt} + \beta\left(\varepsilon^V_{am,t,t+1} - \varepsilon^V_{jm,t,t+1}\right) + \sigma\left(\psi_{amt} - \psi_{jmt}\right) \\ = \; & \beta\left(F^k_{amt} - F^k_{jmt}\right)\left(\pi_{Jmt+1} + \varepsilon^V_{Jm,t+1,t+2} + \sigma\psi_{Jmt+1}\right) \end{aligned} \tag{35}$$

because the V_{t+2} portions of the value function cancel conditional on the renewal action being used in period $t+1$. Thus the effect of the terminal value V_{mT} has been eliminated and (35) forms our base equation for identification.[38]

[37] Formally, ω_{mt} is a market-level shock affecting all agents i in m at t. If the data $\{a_{imt}, k_{imt} : i = 1, ..., N\}$ is i.i.d. conditional on the market level shock ω_{mt}, then, by the law of large numbers for exchangeable random variables (see, e.g., Hall and Heyde, 1980),

$$\widehat{p}_{amt}(k) = \frac{\sum_{i=1}^N 1\{a_{imt} = a, k_{imt} = k\}}{\sum_{i=1}^N 1\{k_{imt} = k\}} \xrightarrow{p} E\left[a_{imt} = a | k_{imt} = k, \omega_{mt}\right] = p_a(k, \omega_{mt})$$

as $N \to \infty$. The same argument applies to the estimator of $F^k(.)$.

[38] Formally, use (33) for J in $t+1$ to solve for V_{mt+1} and replace the latter in (33) for any a in t:

$$\begin{aligned} \pi_{amt} + \beta\varepsilon^V_{am,t,t+1} &= V_{mt} - \sigma\psi_{amt} \\ &\quad -\beta F^k_{amt}\left[\pi_{Jmt+1} + \beta\varepsilon^V_{Jmt+1,t+2} + \beta F^k_{Jmt+1} V_{mt+2} + \sigma\psi_{Jmt+1}\right]. \end{aligned}$$

Next, evaluate the above at a and j and subtract to obtain (35) using the renewability property (34).

Proposition 8 *Suppose (σ, β, G) are known. Given the joint distribution of observables $\Pr(y)$, where $y_{imt} = (a_{imt}, k_{imt}, w_{mt}, z_{mt})$, the flow payoff $\overline{\pi}(a, k_{imt}, w_{mt})$ is identified provided the following conditions hold:*

(a) For finite T: either (i) the terminal value V_{mT} is known and the payoff $\overline{\pi}(j, k, w)$ is known for some j and all (k, w); or (ii) there is a renewal action J with known flow payoff for all (k, w).

(b) For large T: there is a renewal action J, and the flow payoff of some action j (not necessarily action J) is known for all (k, w).

Proposition 8 shows that, when market-level states are partially observed, identification requires restrictions on payoffs as discussed in Section 2.1, as well as an extra restriction: the presence of a renewal action.

Finally, simple inspection of (33) shows that payoffs can be identified with resale prices: the right hand side can be recovered from the data. Because the expectational errors and the unobservable ξ have zero mean given the instrumental variables, we can treat the model as a (nonparametric) regression model. There is no need to impose extra identifying restrictions nor renewability.

Proposition 9 *Suppose the primitives (β, G) are known and either (i) σ is known, or (ii) the (cardinal) payoff $\overline{\pi}(a, k_{imt}, w_{mt})$ is known for one combination of (a, k, w). Given the joint distribution of observables $\Pr(y)$, where $y_{imt} = \left(a_{imt}, k_{imt}, w_{mt}, z_{mt}, p^{RS}_{imt}\right)$, the payoff function $\overline{\pi}(a, k_{imt}, w_{mt})$ is identified.*

6 Application: Agricultural Land Use Model

The empirical application is based on our dynamic land use problem of Example 3. Here, we add market-level unobservables as in Scott (2013); our model is therefore a special case of the model presented in Section 5.2. We also make use of land resale prices to empirically investigate the impact of identifying restrictions on counterfactual behavior.

Recall that in each period field owners decide whether to plant crops or not; i.e. $\mathcal{A} = \{c, nc\}$, where c stands for "crop" and nc stands for "no crop." 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, $k_{imt} \in \mathcal{K} = \{0, 1, \ldots, K\}$; and[39]
3. aggregate state, ω_{mt} (e.g. input and output prices, government policies) with an observed component w_{mt}.

The payoff combines (7), with $\theta_1(a, k) = 1$, and (29), so that:

$$\pi(a, k_{imt}, \omega_{mt}, s_{im}, \varepsilon_{imt}) = \theta_0(a, k_{imt}, s_{im}) + R(a, w_{mt}) + \xi(a, k_{imt}, \omega_{mt}, s_{im}) + \sigma\varepsilon_{imt} \quad (36)$$

[39] Ideally, k_{imt} would include detailed information on past land use. We consider the years since the field was in crop (bounded by K) for computational tractability and due to data limitations.

where $R(a,w)$ and $\xi(a,k,\omega,s)$ are observable and unobservable measures of returns. We construct returns $R^a_{mt} \equiv R(a, w_{mt})$ using county-year information (expected prices and realized yields for major US crops, as well as USDA cost estimates) as in Scott (2013).[40]

The transition of state variables follows the decomposition (6), which implies that farmers are small (price takers) and that there are no externalities across fields. The transition rule of k is: $k'(a,k) = 0$ if $a = crop$, and $k'(a,k) = \min\{k+1, K\}$ if $a = no\ crop$. I.e. if "no crop" is chosen, the number of years since last crop increases by one, until K. If "crop" is chosen, the number of years since last crop is reset to zero. Planting crops is therefore a renewal action.

Data. We performed a spatial merge of a number of datasets to create a uniquely rich database. First, we collected high-resolution annual land use data in the United States obtained from the Cropland Data Layer (CDL) database. The CDL was merged with an extensive dataset of land transactions obtained from DataQuick (which includes information on price, acreage, field address and other characteristics). Then, we incorporated 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.

Our dataset is the first to allow for such rich field heterogeneity; s_{im} includes slope, altitude, soil type, as well as latitude and longitude. A field's slope affects the difficulty of preparing it for crops. Altitude and soil type are crucial for its planting productivity. The field's distance to close urban centers, as well as its nearby commercial property values impact both land use and land values.

Further details about the construction of the dataset, as well as some summary statistics, are presented in Appendix 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 noncrop ranges from 27% to 43% on average depending on the year, which suggests some farmers enjoy benefits from leaving land fallow for a year.[41]

6.1 Estimators for the Land Use Model

The parameters of interest are σ and $\theta_0(a,k,s)$, all a,k,s. We present and compare two estimators. First, we employ Scott (2013)'s method which relies on data for actions and states. It is similar in spirit, but differs from the nested fixed-point (Rust (1987)) and the two-step estimator of Hotz and Miller (1993) as it allows for unobservable market states and can be estimated using a linear regression. We call this estimator, the "CCP estimator." It requires

[40] We refer the interested reader there for details. Due to data limitations, we restrict R to depend only on (a, w_{mt}). One important difference from Scott (2013) is that we have field level observable characteristics s_{im} and they affect land use switching costs.

[41] One may worry that transacted fields are selected. In Table 8 of Appendix B we compare the transacted fields (in DataQuick) to all US fields (in the CDL). Overall, the two sets of fields look similar. Finally, we explore whether land use changes upon resale and find no such evidence (see Table 10 in Appendix C).

restrictions on $\theta_0(a,k,s)$. The second estimator, which we call the "V-CCP estimator", makes use of resale prices to avoid the restrictions on $\theta_0(a,k,s)$. Both estimators require first stage estimates of the conditional choice probabilities, $p_{mt}(a,k,s)$, while the V-CCP estimator also requires estimating the value function $V_{mt}(k,s)$ from resale prices.

The CCP estimator. Scott (2013) derives a regression estimator using the feature that planting crops is a renewal action. Indeed, adapting (35) to the land use model, we obtain:

$$Y^b_{mt}(k,s) = \theta^*_0(k,s) + \frac{1}{\sigma}\left(R^c_{mt} - R^{nc}_{mt}\right) + \xi^*_{mt}(k,s) + \varepsilon^{*V}_{mt}(k,s) \tag{37}$$

for all k, s, where[42]

$$Y^b_{mt}(k,s) \equiv \ln\left(\frac{p_{mt}(c,k,s)}{p_{mt}(nc,k,s)}\right) + \beta \ln\left(\frac{p_{mt+1}(c,k'(c,k),s)}{p_{mt+1}(c,k'(nc,k),s)}\right),$$

$$\theta^*_0(k,s) = \frac{1}{\sigma}\left[\theta_0(c,k,s) - \theta_0(nc,k,s) + \beta\left[\theta_0(c,k'(c,k),s) - \theta_0(c,k'(nc,k),s)\right]\right]. \tag{38}$$

We estimate the parameters in two steps. First, we estimate σ alone via instrumental variables (IV) regression on first differences of (37). As discussed in Section 5.2, one should expect $(R^{nc}_{mt} - R^c_{mt})$ and $\xi^*_{mt}(k,s)$ to be correlated. Following Scott (2013), we employ lagged returns and caloric yields to instrument for the first difference of $(R^{nc}_{mt} - R^c_{mt})$. Once σ is estimated, we move to $\theta_0(a,k,s)$. We obtain the residuals of (37) and take their average over time in order to remove $\xi^*_{mt}(k,s)$ and $\varepsilon^{*V}_{mt}(k,s)$. Then we use (38) to obtain the switching costs parameters. As shown in Section 2.2, we cannot identify all $\theta_0(a,k,s)$ when only actions and states are observed. We follow Scott (2013) again and restrict $\theta_0(nc,k,s) = 0$ for all k,s to recover $\theta_0(c,k,s)$.

The V-CCP Estimator. To isolate the impact of identifying restrictions on counterfactuals, we consider a second estimator. The V-CCP estimator uses (37) alone to estimate σ (following the CCP estimator) and only uses resale price data to recover $\theta_0(a,k,s)$. This estimator is designed so that the only role of the resale price data is to avoid the identifying restrictions $\theta_0(nc,k,s) = 0$ for all k,s.

To obtain $\theta_0(a,k,s)$, we first adapt (30) to the present model:

$$Y^v_{mt}(a,k,s) = \theta_0(a,k,s) + R^a_{mt} + \xi_{mt}(a,k,s) + \beta\varepsilon^V_{mt+1}(a,k), \tag{39}$$

where

$$Y^v_{mt}(a,k,s) \equiv V_{mt}(k,s) - \beta V_{mt+1}(k'(a,k),s) - \sigma\psi_{mt}(a,k,s).$$

Similar to the CCP estimator, given σ we obtain the residuals of (39) and take their average over time in order to remove ξ_{mt} and ε^V_{mt+1}. We can obtain both $\theta_0(c,k,s)$ and $\theta_0(nc,k,s)$ in

[42] We refer the interested reader to Scott (2013) for the detailed derivation. Remember that (i) F^k_{amt} evolves deterministically, (ii) $p_{mt}(a,k) = p(a,k,\omega_{mt})$, and that (iii) for the binary choice model with logit shocks, $\psi_a(p(x)) = -\log(p_a(x)) + \gamma$. The term $\xi^*_{mt}(k,s)$ is given by

$$\xi^*_{mt}(k,s) = \frac{1}{\sigma}\left[\xi_{mt}(c,k,s) - \xi_{mt}(nc,k,s) - \beta E\left[\xi_{mt+1}\left(c,k'(c,k),s\right)\right] + \beta E\left[\xi_{mt+1}\left(c,k'(nc,k),s\right)\right]\right],$$

with the expectations taken over ω_{mt+1}; and $\varepsilon^{*V}_{mt}(k,s)$ is defined similarly to $\xi^*_{mt}(k,s)$.

this fashion. However, in practice, we found that not employing the "no crop" value function moments led to more stable results, likely because R^{nc} is somewhat poorly measured (see Appendix B). So, we recover $\theta_0(c,k,s)$ from (39), and given $\theta_0(c,k,s)$ we recover $\theta_0(nc,k,s)$ from (38).

Note that we could also use (39) to estimate σ. We opted not to do so because the finite sample differences in the estimates of σ from (39) would prevent us from isolating the impact of the identifying restrictions on $\theta_0(a,k,s)$ on counterfactual land use patterns.[43] The only difference between the CCP and V-CCP estimators is in how they recover estimates of $\theta_0(a,k,s)$. The CCP estimator relies on the restriction $\theta_0(nc,k,s)=0$ to derive $\theta_0(c,k,s)$ from estimates of $\theta_0^*(k,s)$. The V-CCP estimator relies on direct estimates of $\theta_0(c,k,s)$ to derive $\theta_0(nc,k,s)$ from estimates of $\theta_0^*(k,s)$.

6.2 Results

We now turn to our results. We focus on the parameters of interest σ and $\theta_0(a,k,s)$ and give the details of the first stage estimates in Appendix C. Table 4 presents the estimated parameters using the CCP and V-CCP estimators. For brevity we only present the average $\sigma\theta_0(a,k,s)$ across field types s (we multiply by σ so that the parameters can be interpreted in dollars per acre). We set $K=2$ due to data limitations and because after 2 years out of crops there are very few conversions back to crops in the data.[44]

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 $k=0$, crop was 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 to produce crop 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 such interpretation hinges on the assumption that $\theta_0(nocrop,k)=0$ for all k. As is typical in switching cost models, estimated switching costs are somewhat large in order to explain the observed persistence in choices; unobserved heterogeneity – which is beyond the scope of this paper – can alleviate this (see Scott (2013)).

The estimated parameters of the V-CCP estimator do not impose $\theta_0(nocrop,k)=0$. When $k=0$, switching out of crops is now expensive on average (not zero anymore). In fact, we reject the null hypothesis of $\overline{\theta}(nocrop,k)=0$, for all k. This is reasonable because the "no crop" 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 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 $\theta_0(nocrop,k)$. Given that the probability of planting

[43] In particular, we can use moments involving resale price data for all values of $\theta_0(a,k,s)$ and also equation (39) jointly with the CCP estimator to estimate σ. This estimator adds more moments from the resale data than are needed to avoid identifying restrictions, and it produces a different estimate of σ than the CCP estimator. However, our main findings on the impact of the identifying restrictions on counterfactual behavior are qualitatively robust to the use of this third estimator.

[44] We weight observations as in Scott (2013) and cluster standard errors by year. We construct the confidence intervals for $\sigma\overline{\theta}_0(a,k)$ by sampling from the estimated asymptotic distribution of $(\widehat{\sigma},\widehat{\theta}_0)$.

Table 4: Empirical Results

Estimator:	CCP	V-CCP
$\sigma\bar{\theta}_0\,(crop,0)$	-721.93	-1228.9
	(-1150,-562)	(-2250,-852)
$\sigma\bar{\theta}_0\,(crop,1)$	-2584.4	-1119.4
	(-4610,-1840)	(-3130,-379)
$\sigma\bar{\theta}_0\,(crop,2)$	-5070.8	-4530.4
	(-9220,-3541)	(-8340,-3120)
$\sigma\bar{\theta}_0\,(nocrop,0)$	0	-2380.3
		(-3540,-1950)
$\sigma\bar{\theta}_0\,(nocrop,1)$	0	470.05
		(-393,788)
$\sigma\bar{\theta}_0\,(nocrop,2)$	0	-454.58
		(-996,-255)
σ	734.08	734.08
	(418,1050)	(418,1050)

θ_0 values are means across all fields in the sample.

90% confidence intervals in parentheses.

crops after one year of fallow is lower than the probability of planting crops after crops in the data (in most markets), in order to rationalize the choice probabilities, the restricted model (imposing $\theta_0\,(nocrop, k) = 0$) 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.[45]

6.3 Policy Counterfactuals

We consider the two counterfactuals discussed in Example 3 of Section 3.1: 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, R^c. To calculate it, we compare the steady-state acreage distribution in the data obtained when R^c is held fixed at their average recent levels and when R^c is held fixed at 10% higher levels. The LRE is defined as the arc elasticity between the total acreage in the two steady states.[46]

As shown in Table 5, the CCP and V-CCP estimators give *exactly* the same LRE. This is no coincidence. Rescale payoffs (36) by dividing by $1/\sigma$ (i.e. $\theta_0/\sigma + R^a_{mt}/\sigma$ plus errors); then, by Proposition 2 σ is identified and by Proposition 5(i), a counterfactual that changes only the identified part of payoffs is also identified. 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.

The second counterfactual increases the crop replanting costs as

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

The difference $\theta\,(crop, 1) - \theta\,(crop, 0)$ 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.[47] Formally, $\theta_0\,(a)$ is a 3×1 vector, and we take $\widetilde{\theta}_0\,(a) = H_0\,(a)\,\theta_0\,(a)$, with $H_0\,(nocrop) = I$,

[45] 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.

[46] See Scott (2013) for a formal definition and further discussion. The LREs estimated here are somewhat 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 unoberseved heterogeneity, LREs are very similar to those presented here.

[47] As with the LRE, we fix R^c and R^{nc} at their mean level for each county.

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

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

Because $H_0\,(nocrop)$ is diagonal but $H_0\,(crop)$ is not, these are not similar matrices. By Corollary 4(i) and Proposition 5(ii), the counterfactual choice probability is not identified.

Indeed, 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 relax 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 relax the identifying restrictions.

7 Conclusions

This paper studies the identification of counterfactuals in dynamic discrete choice models. We ask (i) whether counterfactual behavior and welfare can be identified when the model parameters are not; and if so, (ii) precisely which counterfactuals are identified and which are not. We provide a minimal set of sufficient conditions that determine whether a counterfactual is identified and that are straightforward to verify in practice.

For the applied examples of a monopolist's entry/exit decisions and a farmer's land use decisions, we explore relevant counterfactuals. 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, counterfactuals can often be designed in a way that makes them robust to such restrictions.

References

[1] 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.

[2] Aguirregabiria, V. and P. Mira (2010) "Dynamic Discrete Choice Structural Models: A Survey," *Journal of Econometrics*. 156(1), 38-67.

[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.

[4] Arcidiacono, P. and P. B. Ellickson (2011). "Practical Methods for Estimation of Dynamic Discrete Choice Models," *Annual Review of Economics*, vol. 3, 363-394.

[5] Arcidiacono, P. and R. Miller (2011). "CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, 79, 1823-1867.

[6] Arcidiacono, P. and R. Miller (2015). "Identifying Dynamic Discrete Choice Models off Short Panels," Mimeo.

[7] Chan, T. Y., B. H. Hamilton and N. Papageorge (2015). "Health, Risky Behavior and the Value of Medical Innovation for Infectious Diseases," Working Paper, Johns Hopkins University.

[8] Collard-Wexler, A. (2013). "Demand Fluctuations in the Ready-mix Concrete Industry," *Econometrica*, 81(3), 1003–1007.

[9] Crawford, G. S. and M. Shum (2005). "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 73, 4, 1137-1173.

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

[11] Diermeier, D., M. Keane, and A. Merlo (2005). "A Political Economy Model of Congressional Careers," *American Economic Review*, 95, 347–373.

[12] Duflo, E., R. Hanna, and S. P. Ryan (2012). "Incentives Work: Getting Teachers to Come to School," *The American Economic Review*, 102, 1241–1278.

[13] Dunne, T., S. Klimek, M. Roberts, and Y. Xu (2013). "Entry and Exit and the Determinants of Market Structure," *The RAND Journal of Economics*, 44 (3), 462–487.

[14] Eckstein, Z., and K. I. Wolpin (1989). "Dynamic Labour Force Participation of Married Women and Endogenous Work Experience," *Review of Economic Studies*, Vol. 56, No. 3, 375-390.

[15] Eckstein, Z., and K. I. Wolpin (1999). "Why youths drop out of high school: the impact of preferences, opportunities, and abilities," *Econometrica*, 67 (6), 1295–1339.

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

[17] Gilleskie, D. (1998). "A Dynamic Stochastic Model of Medical Care Use and Work Absence," *Econometrica*, 66, 1-45.

[18] Gould, E. D. (2008). "Marriage and Career: The Dynamic Decisions of Young Men," *Journal of Human Capital*, Vol. 2, No. 4 (Winter 2008), 337-378.

[19] Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Application.* Academic Press, New York-London.

[20] Heckman, J., and S. Navarro (2007). "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, 136, 341–396.

[21] Hendel, I., and A. Nevo (2006). "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, 74(6), 1637-1674.

[22] 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.

[23] Hu, Y., and M. Shum (2012). "Nonparametric Identification of Dynamic Models with Unobserved State Variables," *Journal of Econometrics*, 171, 32–44.

[24] Igami, M. (2015) "Estimating the Innovator's Dilemma: Structural Analysis of Creative Destruction in the Hard Disk Drive Industry, 1981–1998," *Journal of Political Economy*, forthcoming.

[25] Jeziorski, P., E. Krasnokutskaya, and O. Ceccariniz (2015). "Adverse Selection and Moral Hazard in the Dynamic Model of Auto Insurance," Mimeo.

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

[27] Li, S. and C. Wei (2014). "The Cost of Green Stimulus: A Dynamic Discrete Analysis of Vehicle Scrappage Programs," Working Paper, The George Washington University.

[28] Kalouptsidi, M. (2014*a*). "Time to Build and Fluctuations in Bulk Shipping," *American Economic Review*, 104(2): 564-608.

[29] Kalouptsidi, M. (2014*b*). "Detection and Impact of Industrial Subsidies: The Case of World Shipbuilding," NBER Working Paper 20119.

[30] Kalouptsidi, M., P. Scott, and E. Souza-Rodrigues (2016). "On the Non-Identification of Counterfactuals in Dynamic Discrete Games." *International Journal of Industrial Organization (Special issue EARIE)*, forthcoming.

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

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

[33] Keane, M. P., P. E. Todd, and K. I. Wolpin (2011). "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications," in *Handbook of Labor Economics*, volume 4, chapter 4, Elsevier, 331-461.

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

[35] Keane, M. P., and Wolpin, K. I. (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*, Vol. 51, No. 3, 851-892.

[36] Koning, R. H., H. Neudecker and T. Wansbeek. (1991). "Block Kronecker Products and the vecb Operator". *Linear Algebra and Its Applications*. 165-184.

[37] Magnac, T. and D. Thesmar (2002). "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70(2), 801–816.

[38] Norets, A. (2009). "Inference in Dynamic Discrete Choice Models With Serially Correlated Unobserved State Variables," *Econometrica*, 77, 1665–1682.

[39] Norets, A. and X. Tang (2014). "Semiparametric Inference in Dynamic Binary Choice Models," *Review of Economic Studies*, 81(3), 1229-1262.

[40] Pesendorfer, M. and Schmidt-Dengler (2008). "Asymptotic Least Square Estimators for Dynamic Games," *Review of Economic Studies*, 75, 901–908.

[41] Phillips, P. C. B. and H. Moon (1999). "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67 (5), 1057-1111.

[42] 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), 2265-2295.

[43] 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.

[44] Rust, J. (1987). "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999-1033.

[45] Rust, J. (1994). "Structural Eestimation of Markov Decision Processes," In: Engle, R., McFadden, D. (Eds.), *Handbook of Econometrics*. North-Holland, New York, 3081-3143.

[46] Rust, J., and C. Phelan (1997). "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica*, 65, 781-832.

[47] Ryan, S. P. (2012). "The Costs of Environmental Regulation in a Concentrated Industry," *Econometrica*, 80. 1019-1061.

[48] Schiraldi, P. (2011). "Automobile Replacement: a Dynamic Structural Approach," *The RAND Journal of Economics*, Volume 42, Issue 2, pages 266–291.

[49] Scott, P. T. (2013). "Dynamic Discrete Choice Estimation of Agricultural Land Use," Working Paper, Toulouse School of Economics.

[50] Sweeting, A. (2013). "Dynamic Product Positioning in Differentiated Product Markets: The Effect of Fees for Musical Performance Rights on the Commercial Radio Industry," *Econometrica*, Volume 81, Issue 5, pages 1763–1803.

[51] Todd, P. and K. I. Wolpin (2006). "Assessing the Impact of a School Subsidy Program in Mexico." *American Economic Review*, 96 (5), 1384-1417.

[52] Varela, M. (2013). "The costs of Growth: Estimating Entry Costs During Rollouts," Working Paper, University of Arizona.

A Appendix: Proofs of Propositions

A.1 Identification of the Dynamic Discrete Choice Model

A.1.1 Proof of Proposition 1

Fix the vector $\pi_J \in \mathbb{R}^X$. Then,

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

where for $a = J$

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

After substituting for V, we have[48]

$$\pi_a = (I - \beta F_a)(I - \beta F_J)^{-1} (\pi_J + \sigma\psi_J) - \sigma\psi_a.$$

A.1.2 Proof of Remark 1

Action J is a renewal action if for all a, $j \in \mathcal{A}$

$$F_a F_J = F_j F_J. \tag{40}$$

Lemma 3 *If J is a renewal action, then for all $a \in A$*

$$A_a = I + \beta (F_J - F_a).$$

Proof. Using the definition of A_a from Proposition 1,

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

[48] $(I - \beta F_J)$ 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 $1 - \beta\lambda$, where λ are the eigenvalues of F_J. Because $\beta < 1$ and $\lambda \leq 1$, we have $1 - \beta\lambda > 0$.

Given the renewal action property, $F_aF_J = F_J^2$, we have

$$A_a = \sum_{t=0}^{\infty} \beta^t F_J - \beta F_a - \sum_{t=1}^{\infty} \beta^{t+1} F_J^{t+1} = I + \beta \left(F_J - F_a\right).$$

■

We conclude that, even though renewability is a further restriction, it does not aid identification. It only simplifies the expression for A_a, $a \neq J$.

A.1.3 Proof of Proposition 2

We make use of two lemmas.

Lemma 4 *If $\pi\left(a, x; \theta\right)$ satisfies*

$$\pi(a, x; \theta) = \overline{\pi}_a\left(x\right)\theta, \tag{41}$$

θ is identified provided rank $\left[\overline{\pi}_{-J} - A_{-J}\overline{\pi}_J\right] = \dim\left(\theta\right)$, where $\overline{\pi}_{-J} = \left[\overline{\pi}_1, ..., \overline{\pi}_{J-1}, \overline{\pi}_{J+1}, ..., \overline{\pi}_A\right]$.

Proof. Assume $\pi_{-J} = \overline{\pi}_{-J}\theta$. Then (5) becomes $\left[\overline{\pi}_{-J} - A_{-J}\overline{\pi}_J\right]\theta = b_{-J}$. So, if the rank of the matrix $\left[\overline{\pi}_{-J} - A_{-J}\overline{\pi}_J\right]$ equals $\dim\left(\theta\right)$, then

$$\theta = \left[\left(\overline{\pi}_{-J} - A_{-J}\overline{\pi}_J\right)'\left(\overline{\pi}_{-J} - A_{-J}\overline{\pi}_J\right)\right]^{-1}\left(\overline{\pi}_{-J} - A_{-J}\overline{\pi}_J\right)' b_{-J}.$$

■

Lemma 5 *Let $D_a = [I - \beta(F^w \otimes F_a^k)]^{-1}$, 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 $A_a^k = \left(I_k - \beta F_a^k\right)\left(I_k - \beta F_J^k\right)^{-1}$. The following properties hold*

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

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

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

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: $\sum_j f_{ij}^w = 1$, where f_{ij}^w is the $(i, j) - th$ element of F^w. By the definition of the Kronecker product,

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

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

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

$$(F^w \otimes F_a^k)^n = (F^w)^n \otimes (F_a^k)^n$$

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

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

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

$$A_a \mathbf{1} = (I - \beta(F^w \otimes F_a^k)) D_J \mathbf{1} = (I - \beta(F^w \otimes F_a^k)) \mathbf{1}(I_k - \beta F_J^k)^{-1} = \mathbf{1}(I_k - \beta F_a^k)(I_k - \beta F_J^k)^{-1} = \mathbf{1} A_a^k.$$

■

We now provide the proof of Proposition 2; 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. $\theta_0^a = [\theta_0(a, k_1), ..., \theta_0(a, k_K)]'$),

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

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

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

and

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

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

$$\overline{\pi}_a = \begin{bmatrix} I_k & 0_k & R_a(w_1) I_k & 0_k \\ \vdots & \vdots & \vdots & \vdots \\ I_k & 0_k & R_a(W) I_k & 0_k \end{bmatrix} \tag{42}$$

and similarly for π_J. In Lemma 4, we showed that identification hinges on the matrix $(\overline{\pi}_a - A_a \overline{\pi}_J)$. This matrix equals:

$$\overline{\pi}_a - A_a \overline{\pi}_J = \left[\begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix}, \quad -A_a \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix}, \quad R_a, \quad -A_a R_J \right] \tag{43}$$

where $R_a = [R_a(w_1) I_k, ..., R_a(w_W) I_k]'$ (the same for R_J).

It follows from Lemma 5 that the first two block columns of (43) 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.[49] 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 (43):

$$\begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} \theta_0^a - \begin{bmatrix} A_a^k \\ \vdots \\ A_a^k \end{bmatrix} \theta_0^J + R_a\theta_1^a - \left[I - \beta\left(F^w \otimes F_a^k\right)\right]\left[I - \beta\left(F^w \otimes F_J^k\right)\right]^{-1} R_J\theta_1^J = b_a.$$

Left-multiplying both sides by $D_a = \left[I - \beta\left(F^w \otimes F_a^k\right)\right]^{-1}$ and using Lemma 5, we obtain:

$$\begin{bmatrix} \left(I_k - \beta F_a^k\right)^{-1} \\ \vdots \\ \left(I_k - \beta F_a^k\right)^{-1} \end{bmatrix} \theta_0^a - \begin{bmatrix} \left(I_k - \beta F_J^k\right)^{-1} \\ \vdots \\ \left(I_k - \beta F_J^k\right)^{-1} \end{bmatrix} \theta_0^J + D_aR_a\theta_1^a - D_JR_J\theta_1^J = D_ab_a.$$

Take the w block row of the above:

$$\left(I_k - \beta F_a^k\right)^{-1}\theta_0^a - \left(I_k - \beta F_J^k\right)^{-1}\theta_0^J + e_w' D_aR_a\theta_1^a - e_w' D_JR_J\theta_1^J = e_w' D_ab_a \tag{44}$$

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:

$$\begin{bmatrix} \left(e_w' - e_{\widetilde{w}}'\right) D_aR_a & \left(e_{\widetilde{w}}' - e_w'\right) D_JR_J \\ \left(e_w' - e_{\overline{w}}'\right) D_aR_a & \left(e_{\overline{w}}' - e_w'\right) D_JR_J \end{bmatrix} \begin{bmatrix} \theta_1^a \\ \theta_1^J \end{bmatrix} = \begin{bmatrix} \left(e_w' - e_{\widetilde{w}}'\right) D_ab_a \\ \left(e_w' - e_{\overline{w}}'\right) D_ab_a \end{bmatrix}$$

which proves the Proposition.

A.2 Identification of Counterfactuals

A.2.1 Proof of Lemma 2

Assume without loss that $J = A$. To prove Lemma 2, we first make use of Lemma 6 below. 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 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

$$P_a = \begin{bmatrix} p_a(x_1) & 0 & \cdots & 0 \\ 0 & p_a(x_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_a(X) \end{bmatrix}$$

for $a = 1, ..., A-1$; and let $P = [P_1, P_2, ..., P_{A-1}]$. Lemma 6 follows.

[49] 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}.

Lemma 6 *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

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

where $p'(x) = [p_1(x), ..., p_{J-1}(x)]$. 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,

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

■

Recall the definition of $b_{-J}(p) : \mathbb{R}^{(A-1)X} \to \mathbb{R}^{(A-1)X}$ in Section 2 (here we ignore σ for simplicity). Because $\psi_a = \psi_J - \phi_{aJ}$, we have

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

where $\psi_J(p)$ 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 6 we have

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

where $\mathbf{A} = \begin{bmatrix} A_1 - I \\ \vdots \\ A_{J-1} - I \end{bmatrix}$ has dimension $(A-1)\,X \times X$.

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

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

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

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

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 $det\,(\mathbf{A}P + I) \neq 0$.

A.2.2 Proof of Proposition 1

Assume without loss that action A belongs to both sets $\mathcal{A}$ and $\widetilde{\mathcal{A}}$, and take $J = A$. The implicit function theorem allows us to locally solve (16) with respect to $\widetilde{p}$ provided the matrix

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

is invertible; this is proved in Lemma 2. The vector $\widetilde{p}$ does not depend on the free parameter π_J if and only if

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

for all $a \in \widetilde{\mathcal{A}}$, with $a \neq J$ and all π. 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 (4),

$$\sum_{l \in \mathcal{A}, l \neq J} \frac{\partial h_a}{\partial \pi_l} A_l + \frac{\partial h_a}{\partial \pi_J} = \widetilde{A}_a \left(\sum_{l \in \mathcal{A}, l \neq J} \frac{\partial h_J}{\partial \pi_l} A_l + \frac{\partial h_J}{\partial \pi_J} \right) \tag{45}$$

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:

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

which is (18). 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 $\left(\widetilde{A} - 1\right) \widetilde{X} X \times \left(\widetilde{A} - 1\right) A \widetilde{X} X$ matrix.[50] Similarly, $\left(\begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes \widetilde{A}_{-J} \right)$ is an $\left(\widetilde{A} - 1\right) \widetilde{X} X \times A \widetilde{X} X$ matrix, and $\widetilde{A}_{-J}$ is an $\left(\widetilde{A} - 1\right) \widetilde{X} \times \widetilde{X}$ matrix.

A.2.3 Proof of Example 1, Section 3.2

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

$$\widetilde{\pi} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \pi$$

where $H_{11} = (1 + \lambda) I$, $H_{21} = 0$, $H_{22} = I$, and $H_{12} = -\lambda [1, 0]$, where 1 is a vector of ones and 0 is a matrix with zeros (see footnote 17). Let $J = 2$. By Corollary 2, $\widetilde{p}$ is identified if and only if

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

[50] 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 $\left(\widetilde{A} - 1\right) \widetilde{X} \times \left(\widetilde{A} - 1\right) \widetilde{X}$.

or

$$(1+\lambda)\, I A_1 - \lambda\, [1, 0] - A_1 I = 0$$

which implies

$$\lambda A_1 = \lambda\, [1, 0] \,.$$

This implies A_1 is non-invertible, which is a contradiction.

A.2.4 Proof of Corollary 1

Since h is action diagonal, $\frac{\partial h_a}{\partial \pi_l} = 0$, $l \neq a$, and (19) stems directly from (18).

A.2.5 Proof of Corollary 2

Because $\frac{\partial h_a}{\partial \pi_l} = H_{al}$, equation (45) 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 $a, J \neq l$.

A.2.6 Proof of Corollary 3

Equation (21) is trivially satisfied since $H_a = I$ and $A_a = \widetilde{A}_a$ for all a.

A.2.7 Proof of Corollary 4

(i) Equation (21) becomes $H_a = A_a H_J A_a^{-1}$, for all $a \neq J$. (ii) Diagonal similar matrices are equal to each other, so $H_a = H$, which implies $H = A_a H A_a^{-1}$, for all a.

A.2.8 Proof of Corollary 5

We make use of the following Lemma:

Lemma 7 *Suppose $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = \mathcal{X}$, $\widetilde{F}_a = F_a$, and $\widetilde{\pi}_a = H \pi_a$ for all $a \in A$. Assume H is diagonal with k pairwise distinct diagonal entries $\lambda_1, ..., \lambda_k$ and corresponding multiplicities $n_1, ..., n_k$. Assume also $H = A_a H A_a^{-1}$ for all a. We partition F_a and F_J in blocks F_{ij}^a and F_{ij}^J of size $n_i \times n_j$ (corresponding to the multiplicities). To identify $\widetilde{p}$, it is necessary that the matrices F_a, $a \neq J$, have diagonal blocks F_{ii}^a given by*

$$(I - \beta F_{ii}^a) \left(I - \beta F_{ii}^J \right)^{-1} 1 = 1, \tag{46}$$

where 1 is a $n_i \times 1$ vector of ones; and off-diagonal blocks F_{ij}^a, $i \neq j$, given by

$$F_{ij}^a = (I - \beta F_{ii}^a)(I - \beta F_{ii}^J)^{-1} F_{ij}^J. \tag{47}$$

Furthermore, the right hand side of (47) must be between zero and one.

Proof. Suppose $A_a H = H A_a$ for all a, and H has the form

$$H = diag\left\{\lambda_1 I_{n_1}, ..., \lambda_k I_{n_k}\right\}.$$

We write A_a in partitioned form, $(A_a)_{ij}$, so that it conforms with the decomposition of H. Then the corresponding off-diagonal blocks of $A_a H - H A_a$ satisfy

$$(\lambda_i - \lambda_j)(A_a)_{ij} = 0$$

for $i \neq j$. Therefore, $(A_a)_{ij} = 0$, for $i \neq j$ and A_a must be block diagonal:

$$A_a = diag\left\{A_1^a, ..., A_k^a\right\}$$

Then the definition of A_a implies:

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

or

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

The left hand side is block diagonal. Therefore, the off-diagonal blocks satisfy

$$F_{ij}^a - A_i^a F_{ij}^J = 0 \tag{48}$$

For the diagonal blocks we have

$$\beta^{-1}(I - A_i^a) = F_{ii}^a - A_i^a F_{ii}^J \tag{49}$$

We can isolate A_i^a from (49) to obtain

$$A_i^a = (I - \beta F_{ii}^a)\left(I - \beta F_{ii}^J\right)^{-1}.$$

Substitute this into (48) and get (47). Note that (47) implies that, given F^J and F_{ii}^a, all remaining blocks of F^a are uniquely determined.

Because F_a are stochastic matrices so that their rows add to 1 and all elements are between 0 and 1, there are further restrictions that the blocks of F_a must satisfy. Indeed, consider without loss the first block row of F_a:

$$[F_{11}^a, F_{12}^a, ..., F_{1k}^a].$$

Then each row belonging to this block row must add to 1. Let 1 be a vector of ones. Then

$$F_{11}^a 1 + F_{12}^a 1 + ... + F_{1k}^a 1 = 1$$

where, abusing notation slightly, the vectors 1 above have varying length. Using (48) and the fact that the rows of F_J add to one, we get

$$F_{11}^a 1 + A_1^a\left(1 - F_{11}^J 1\right) = 1.$$

Using (49) as well,

$$A_1^a 1 - 1 = A_1^a F_{11}^J 1 - \beta^{-1}(I - A_1^a) 1 - A_1^a F_{11}^J 1$$

which only holds if

$$A_1^a 1 = 1$$

So,

$$(I - \beta F_{ii}^a) \left(I - \beta F_{ii}^J\right)^{-1} 1 = 1.$$

■

Next, we return to Corollary 5. (i) The case of $H = \lambda I$ follows immediately from $A_a H = H A_a$ for all a. (ii) Next, suppose that one of the elements of H, say λ_1, is simple, that is $n_1 = 1$ and $k > 1$:

$$H = diag \left\{\lambda_1, \lambda_2 I_{n_2}, ..., \lambda_k I_{n_k}\right\}.$$

Then, the corresponding elements $(A_a)_{1j}$ for $j = 1, 2, ..., k$, are all scalars. Because $(A_a)_{ij} = 0$, for $i \neq j$ and $A_1^a 1 = 1$, it is obvious that $A_1^a = 1$. From equations (48) and (49),

$$\begin{aligned} F_{1j}^a &= A_1^a F_{1j}^J, \quad j = 2, ..., k, \\ \beta^{-1} (1 - A_1^a) &= F_{11}^a - A_1^a F_{11}^J, \end{aligned}$$

we conclude the corresponding rows of F_a and F_J are equal. (iii) Finally, when all elements of H are pairwise distinct, all rows of F_a and F_J must be equal.

A.2.9 Proof of Corollary 6

Because $H_a = I$ for all a, equation (21) is satisfied if and only if $A_a = \widetilde{A}_a$.

A.2.10 Proof of Corollary 7

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 (20) 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.2.11 Proof of Proposition 3

We make use of Lemma 8 below, which holds in the general case. Suppose that $\mathcal{A} = \{1, 2, ..., A\}$. Without loss, take the reference action to be $J = 1$ and suppose action $a = A$ is eliminated, so that $\widetilde{\mathcal{A}} = \{1, 2, ..., A - 1\}$. Also without loss, adjust the set of states accordingly: $\mathcal{X} = \{1, ..., \overline{x}, \overline{x} + 1, ..., X\}$ and $\widetilde{\mathcal{X}} = \{1, ..., \overline{x}\}$.

The counterfactual payoff for action a is the $\overline{x} \times 1$ vector:

$$\widetilde{\pi}_a = \begin{bmatrix} I_{\overline{x}} & 0 \end{bmatrix} \pi_a. \tag{50}$$

If some states are deleted, we need to adjust the state transitions. Indeed, the transition is written as follows:

$$F_a = \begin{bmatrix} \widehat{F}_a & f_a \\ g_a & q_a \end{bmatrix} \tag{51}$$

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

We consider transition counterfactuals that are based on transitions of the maintained states, $\widehat{F}_a$, adjusted additively by the eliminated transitions. More precisely, set for each $a \in \widetilde{\mathcal{A}}$ the new transition matrices

$$\widetilde{F}_a = \widehat{F}_a + f_a r_a' \tag{52}$$

where r_a is a $k \times (X - \overline{x})$ matrix such that $r_a' 1 = 1$. The adjustment $f_a r_a'$ guarantees that $\widetilde{F}_a$ is a stochastic matrix (indeed, $\widetilde{F}_a 1 = \widehat{F}_a 1 + f_a r' 1 = \widehat{F}_a 1 + f_a 1 = F_a 1 = 1$). More specifically, the term $f_a r_a'$ specifies how the probability mass in f_a is redistributed among the remaining states.

Lemma 8 *The counterfactual CCP defined by* (50), (51) *and* (52) *is identified if and only if the following restrictions hold for all $a \in \widetilde{\mathcal{A}}$:*

$$(I - \beta \widehat{F}_a)^{-1} f_a = (I - \beta \widehat{F}_J)^{-1} f_J \tag{53}$$

and

$$f_a r_a' = f_a \left[\left(I - \beta r_a' \left(I - \beta \widehat{F}_J \right)^{-1} f_J \right) \left(I - \beta r_J' \left(I - \beta \widehat{F}_J \right)^{-1} f_J \right)^{-1} \right] r_J'. \tag{54}$$

Proof. The counterfactual identification condition given in Corollary 2 becomes:

$$\begin{bmatrix} I_{\overline{x}} & 0 \end{bmatrix} A_a = \widetilde{A}_a \begin{bmatrix} I_{\overline{x}} & 0 \end{bmatrix} \tag{55}$$

for $a = 1, ..., A-1$ since $H_{aj} = 0$, $a \neq j$. First, focus on the RHS of 55. Note that

$$A_a = (I - \beta F_a)(I - \beta F_1)^{-1} = \begin{bmatrix} I - \beta \widehat{F}_a & -\beta f_a \\ -\beta g_a & 1 - \beta q_a \end{bmatrix} (I - \beta F_1)^{-1}$$

And so

$$\begin{bmatrix} I_{\overline{x}} & 0 \end{bmatrix} A_a = \begin{bmatrix} I - \beta \widehat{F}_a & -\beta f_a \end{bmatrix} (I - \beta F_1)^{-1}$$

Next, note that

$$\begin{aligned} (I - \beta F_1)^{-1} &= \begin{bmatrix} I - \beta \widehat{F}_1 & -\beta f_1 \\ -\beta g_1 & I - \beta q_1 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \left(I - \beta \widehat{F}_1 \right)^{-1} + \beta^2 \left(I - \beta \widehat{F}_1 \right)^{-1} f_1 E^{-1} g_1 \left(I - \beta \widehat{F}_1 \right)^{-1} & \beta \left(I - \beta \widehat{F}_1 \right)^{-1} f_1 E^{-1} \\ \beta E^{-1} g_1 \left(I - \beta \widehat{F}_1 \right)^{-1} & E^{-1} \end{bmatrix} \end{aligned}$$

where $E = (I - \beta q_1) - \beta^2 g_1 \left(I - \beta \widehat{F}_1 \right)^{-1} f_1$.[51] Therefore, the first term of $\begin{bmatrix} I_{\overline{x}} & 0 \end{bmatrix} A_a$ is

$$\left(I - \beta \widehat{F}_a \right) \left(I - \beta \widehat{F}_1 \right)^{-1} + \beta^2 \left(I - \beta \widehat{F}_a \right) \left(I - \beta \widehat{F}_1 \right)^{-1} f_1 E^{-1} g_1 \left(I - \beta \widehat{F}_1 \right)^{-1} - \beta^2 f_a E^{-1} g_1 \left(I - \beta \widehat{F}_1 \right)^{-1},$$

[51] We use the following property:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1} B E^{-1} C A^{-1} & -A^{-1} B E^{-1} \\ -E^{-1} C A^{-1} & E^{-1} \end{bmatrix}$$

with $E = D - CA^{-1}B$.

while the second term is

$$\beta\left(I-\beta\widehat{F}_a\right)\left(I-\beta\widehat{F}_1\right)^{-1}f_1E^{-1}-\beta f_aE^{-1}.$$

Next, we turn to the RHS of (55):

$$\widetilde{A}_a\begin{bmatrix}I_{\overline{x}} & 0\end{bmatrix}=\begin{bmatrix}\widetilde{A}_a & 0\end{bmatrix}=\begin{bmatrix}\left(I-\beta\widetilde{F}_a\right)\left(I-\beta\widetilde{F}_1\right)^{-1} & 0\end{bmatrix}$$

Define $\widehat{A}_a=(I-\beta\widehat{F}_a)(I-\beta\widehat{F}_1)^{-1}$. Then (55) is satisfied if and only if the following two conditions hold:

$$\widehat{A}_a+\beta^2\widehat{A}_af_1E^{-1}g_1\left(I-\beta\widehat{F}_1\right)^{-1}-\beta^2f_aE^{-1}g_1\left(I-\beta\widehat{F}_1\right)^{-1}=\widetilde{A}_a$$

and

$$\beta\left[\widehat{A}_af_1-f_a\right]E^{-1}=0.$$

The latter becomes

$$\widehat{A}_af_1=f_a$$

Next, we substitute the latter into the former.

$$\widehat{A}_a+\beta^2f_aE^{-1}g_1\left(I-\beta\widehat{F}_1\right)^{-1}-\beta^2f_aE^{-1}g_1\left(I-\beta\widehat{F}_1\right)^{-1}=\widetilde{A}_a$$

which implies $\widehat{A}_a=\widetilde{A}_a$. Finally, note that $\widetilde{A}_a$ is given by:

$$\widetilde{A}_a=\left(I-\beta\widehat{F}_a-\beta f_ar_a'\right)\left(I-\beta\widehat{F}_1-\beta f_1r_1'\right)^{-1}$$

Using the property $(A-BC)^{-1}=A^{-1}+A^{-1}B(I-CA^{-1}B)^{-1}CA^{-1}$, we get

$$\begin{aligned}\widetilde{A}_a &= \left(I-\beta\widehat{F}_a-\beta f_ar_a'\right)\\ &\quad\times\left[\left(I-\beta\widehat{F}_1\right)^{-1}+\left(I-\beta\widehat{F}_1\right)^{-1}\beta f_1\left(I-r_1'\left(I-\beta\widehat{F}_1\right)^{-1}\beta f_1\right)^{-1}r_1'\left(I-\beta\widehat{F}_1\right)^{-1}\right]\\ &= \left(I-\beta\widehat{F}_a-\beta f_ar_a'\right)\left(I-\beta\widehat{F}_1\right)^{-1}\\ &\quad+\left(I-\beta\widehat{F}_a-\beta f_ar_a'\right)\left(I-\beta\widehat{F}_1\right)^{-1}\beta f_1\left(I-r_1'\left(I-\beta\widehat{F}_1\right)^{-1}\beta f_1\right)^{-1}r_1'\left(I-\beta\widehat{F}_1\right)^{-1}\end{aligned}$$

If we replace $\hat{A}_af_1=f_a$, we get

$$\begin{aligned}\widetilde{A}_a &= \widehat{A}_a-\beta f_ar_a'\left(I-\beta\widehat{F}_1\right)^{-1}\\ &\quad+\beta f_a\left(I-r_1'\left(I-\beta\widehat{F}_1\right)^{-1}\beta f_1\right)^{-1}r_1'\left(I-\beta\widehat{F}_1\right)^{-1}\\ &\quad-\beta f_ar_a'\left(I-\beta\widehat{F}_1\right)^{-1}\beta f_1\left(I-r_1'\left(I-\beta\widehat{F}_1\right)^{-1}\beta f_1\right)^{-1}r_1'\left(I-\beta\widehat{F}_1\right)^{-1}\end{aligned}$$

So, to satisfy $\widehat{A}_a = \widetilde{A}_a$, we need the sum of the last three terms to be zero. This implies

$$f_a r_a' = f_a \left[\left(I - \beta r_a' \left(I - \beta \widehat{F}_1 \right)^{-1} f_1 \right) \left(I - \beta r_1' \left(I - \beta \widehat{F}_1 \right)^{-1} f_1 \right)^{-1} \right] r_1'.$$

■

Next, we return to Corollary 3. Assume $x = (k, w)$, then

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

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 $a - th$ column. The $a - th$ 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 (51). Because $f_J = 0$ as well, conditions (53) and (54) in Lemma 8 are trivially satisfied and the counterfactual CCP is identified.

A.2.12 Proof of Corollary 8

Suppose that $\mathcal{A} = \{1, 2, ..., A\}$. Without loss, 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 (20) becomes:

$$A_a = \widetilde{A}_a, \ for \ a = 2, ..., A$$

and

$$H_{j1} + \sum_{a=2}^{A} H_{ja} A_a = \widetilde{A}_j, \tag{56}$$

since $H_{al} = 0$ and $H_a = I$ for all $a, l \neq j$. The first set of restrictions are satisfied, since transitions are unaffected. Now, post-multiply (56) by $(I - \beta F_1) = \left(I - \beta \widetilde{F}_1 \right)$ to obtain (recall that the reference action is $J = 1$):

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

or

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

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 $\widetilde{F}_j 1 = 1$, where 1 is a vector of ones, 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.2.13 Proof of Proposition 4

To prove Proposition 4, first note that for affine counterfactuals, $\widetilde{p}$ is a function of π_J through an equation of the form:

$$C\pi_J + c_0 = b\left(\widetilde{p}\right)$$

where C is an $(A-1)X \times X$ matrix and c_0 an $(A-1)X \times 1$ vector that do not depend on π_J or $\widetilde{p}$.[52]

The counterfactual CCP *is identified under the set of linear restrictions* (24), if for all $\pi_J \neq \hat{\pi}_J$ such that $Q\pi_J = Q\hat{\pi}_J = q$, it holds $\widetilde{p}(\pi_J) = \widetilde{p}(\hat{\pi}_J)$. Proposition 4 is a direct consequence of the following proposition:

Proposition 10 *Given the matrix C formed by a counterfactual transformation $\mathcal{H}$ and the linear restrictions (22), the counterfactual CCP is identified if and only if any of the following equivalent statements hold:*

1. *For all $\pi_J \neq \hat{\pi}_J$ such that $Q\pi_J = Q\hat{\pi}_J = q$, it holds $C\left(\pi_J - \hat{\pi}_J\right) = 0$.*
2. *$N\left(Q\right) \subseteq N\left(C\right)$, where* $N(Q)$ denotes the nullspace of Q.
3. *There is an $(A-1)X \times d$ matrix M such that $C = MQ$.*

Proof. For the first statement, take $\pi_J \neq \hat{\pi}_J$ such that $Q\pi_J = Q\hat{\pi}_J = q$. Then $\tilde{p}(\pi_J) = \tilde{p}(\hat{\pi}_J)$ implies $b^{-1}\left(C\pi_J + c_0\right) = b^{-1}\left(C\hat{\pi}_J + c_0\right)$, or $C\pi_J + c_0 = C\hat{\pi}_J + c_0$ or $\Leftrightarrow\Leftrightarrow$ $C\left(\pi_J - \hat{\pi}_J\right) = 0$. It is clear that this argument holds in reverse as well.

For the second statement, take $u \in N(Q)$ and π_J^* a particular solution of $Q\pi_J = q$. Set

[52] To see this, following the binary example in Section 3.2, we obtain:

$$\left(\sum_{\hat{a} \neq J} \left(H_{a\hat{a}} - \widetilde{A}_a H_{J\hat{a}} \right) A_{\hat{a}} + H_{aJ} - \widetilde{A}_a H_{JJ} \right) \pi_J = \widetilde{A}_a g_J - g_{\underset{\sim}{a}} + b_a(\widetilde{p}) - \sum_{\hat{a} \neq J} \left(H_{a\hat{a}} - \widetilde{A}_a H_{J\hat{a}} \right) b_a(p)$$

so that

$$C = \left(\sum_{\hat{a} \neq J} \left(H_{a\hat{a}} - \widetilde{A}_a H_{J\hat{a}} \right) A_{\hat{a}} + H_{aJ} - \widetilde{A}_a H_{JJ} \right)$$

and

$$c_0 = \widetilde{A}_a g_J - g_{\underset{\sim}{a}} - \sum_{\hat{a} \neq J} \left(H_{a\hat{a}} - \widetilde{A}_a H_{J\hat{a}} \right) b_a(p).$$

$\hat{\pi}_J = \pi_J^* + u$ and note that $\hat{\pi}_J$ is also a solution (since $Q\hat{\pi}_J = Qu + Q\pi_J^*$ and $Qu = 0$ since u is in the null-space of Q). Thus,

$$C\left(\hat{\pi}_J - \pi_J^*\right) = C\left(\pi_J^* + u - \pi_J^*\right) = Cu = 0$$

For the converse, suppose that there exist $\pi_J \neq \hat{\pi}_J$ such that $Q\pi_J = Q\hat{\pi}_J = q$, but $C\left(\pi_J - \hat{\pi}_J\right) \neq 0$. Then, $Q\left(\pi_J - \hat{\pi}_J\right) = q - q = 0$. Thus, $\pi_J - \hat{\pi}_J \in N(Q) \subseteq N(C)$ and thus $C\left(\pi_J - \hat{\pi}_J\right) = 0$, contradiction.
We now show that 2 and 3 are equivalent. We rely on the known fact that $N(Q) = [col(Q')]^{\perp}$, where $[X]^{\perp}$ is the orthogonal complement of X. Hence, $N(Q) = [row(Q)]^{\perp}$. It is easy to show that $A \subseteq B$ if and only if $[B]^{\perp} \subseteq [A]^{\perp}$. Therefore,

$$N\left(Q\right) \subseteq N\left(C\right) \Leftrightarrow [row(Q)]^{\perp} \subseteq [row(C)]^{\perp} \Leftrightarrow row(C) \subseteq row(Q) \tag{57}$$

But (57) states that every row of C is a linear combination of the rows of Q. The coefficients in this linear combination form the matrix M. For the converse, take a row of C; from 3, it is a linear combination of the rows of Q and thus belongs to $row(Q)$. Therefore, since all the rows of C belong to $row(Q)$, so does their linear span. ■

A.2.14 Proof of Corollary 9

Take $C = MQ$ or $CQ' = MQQ'$, or $CQ'(QQ'^{-1} = M$. Hence, $C = CQ'(QQ'^{-1}Q$, or $\Rightarrow C\left(I - Q'\left(QQ'\right)^{-1}Q\right) = 0$. For the converse, from (25) we have $C = CQ'\left(QQ'\right)^{-1}Q$. Take $M = CQ'(QQ'^{-1}$ to obtain $C = MQ$. Finally, note that $rankC = X - dim(N(C)) \leq X - dim(N(Q)) = rankQ = d$.

A.2.15 Proof of Proposition 5

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

(ii) Consider the counterfactual

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

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 2, equation (44), we know that for any w, the w block row of (4) is

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

The corresponding w block row for the counterfactual scenario is

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

Lack of identification of θ_0 is represented by the free parameter θ_0^J. So, applying the implicit function theorem with respect to θ_0^J in the equation above, and using (44), we prove the claim.

A.2.16 Proof of Corollary 10

Assume $\widetilde{\pi}_a = \pi_a$, all a, then $H_0(a) = H_0(J) = I$. From Proposition 5, it is clear that $\widetilde{p}$ is identified if and only if for all $a \neq J$, $A_a^k = \widetilde{A}_a^k$. 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.2.17 Proof of Proposition 6

Proposition 6 is a direct consequence of Lemma 9 below.

Lemma 9 *Assume $\widetilde{\mathcal{A}} = \mathcal{A}$, $\widetilde{\mathcal{X}} = X$, and let $h_a(\pi_a) = H_a\pi_a + g_a$, all a. Define the matrices*

$$\begin{aligned} C &= \left[H_{-J}A_{-J} - \widetilde{A}_{-J}H_J\right] \\ D &= \left[\left(I - \beta\widetilde{F}_J\right)(I - \beta F_J)^{-1} - H_J\right]. \end{aligned}$$

Then ΔV is identified if and only if

$$\widetilde{P}\left[\sigma C - \widetilde{\mathbf{A}}D\right] = D. \tag{58}$$

where the matrices $\widetilde{\mathbf{A}}$ and $\widetilde{P}$ are defined as in Lemma 2 (but based on $\widetilde{A}_a$ and $\widetilde{p}$). Furthermore, the following holds:

1. If $C = 0$, $D = 0$, then ΔV is identified.
2. If $C = 0$, $D \neq 0$, then ΔV is not identified.
3. If $C \neq 0$, $D = 0$, then, provided there is no combination of actions a and j such that $\widetilde{p}_a(x) = \widetilde{p}_j(x)$ for some state x, ΔV is not identified.
4. If $C \neq 0$, $D \neq 0$, then ΔV is generically not identified.

Proof. We know that

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

and similarly for $\widetilde{V}$

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

Then,

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

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

$$\sigma\frac{\partial\psi_J(\widetilde{p})}{\partial\widetilde{p}}\frac{\partial\widetilde{p}}{\partial\pi_J} = D \tag{59}$$

From Lemma 6, 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 6. By the Implicit Function Theorem, we know that

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

and, by Lemma 2,

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

Thus (59) becomes:

$$\sigma \widetilde{P}\left(\widetilde{\mathbf{A}}\widetilde{P} + I\right)^{-1} C = D \tag{60}$$

Note that[53]

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

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

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

Then, (60) becomes:

$$\sigma M^{-1}\widetilde{P}C = D$$

or

$$\sigma \widetilde{P}C = MD = \left(I + \widetilde{P}\widetilde{\mathbf{A}}\right)D$$

or

$$\widetilde{P}(\sigma C - \widetilde{\mathbf{A}}D) = D$$

Next, we examine the four mutually exclusive possibilities. Statement 1 is obvious. Statement 2 follows from (60). Statement 3 amounts to $\widetilde{P}C = 0$. Recall $\widetilde{P}$ is $X \times (A-1)X$ and C is $(A-1)X \times X$. If there are no two actions a and j such that $\widetilde{p}_a(x) = \widetilde{p}_j(x)$ for some x, then $\widetilde{P}$ is full rank (recall $\widetilde{p}_a(x) > 0$ for all a, x, by the large support assumption on $G(\varepsilon)$). This implies the matrix $\left(\widetilde{P}'\widetilde{P}\right)^{-1}$ exists, and so $C = 0$. But this is impossible since $C \neq 0$ by assumption.

Finally, for statement 4, suppose $C \neq 0$, $D \neq 0$. Recall that $\widetilde{p}$ as a function of π_J is the unique solution of

$$\widetilde{b}_{-J}(\widetilde{p}) = C\pi_{-J} + H_J b(p)$$

The solution set is X-dimensional. Thus, (58) contains X^2 equations and X unknowns (π_J). By Sard's Theorem, the system has generically no solution. ■

A.2.18 Proof of Corollary 11

Lack of identification of θ_0 is represented by the free parameter θ_0^J. So, applying the same argument as in Lemma 9, but differentiating ΔV with respect to θ_0^J, we prove the claim.

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

A.3 Extensions

A.3.1 Identification of Payoffs with Resale Prices

Case (i): Substitute V in (3) into (1):

$$\pi_a = (I - \beta F_a) V - \sigma \psi_a. \tag{61}$$

Case (ii): Fix an element of the vector π_J at the state $\overline{x}$, then solve for σ:

$$\sigma = \frac{1}{\psi_J(\overline{x})} \left[V(\overline{x}) - \beta \sum_{x' \in X} \Pr(x'|J, \overline{x}) V(x') - \pi_J(\overline{x}) \right]$$

provided the right hand side is positive.

Case (iii): Fix $E[\pi_J(x)]$, where the expectation is taken over x. Then,

$$\sigma = \frac{1}{E[\psi_J(x)]} \left[E[V(x)] - \beta E\left[\sum_{x' \in X} \Pr(x'|J, x) V(x') \right] - E[\pi_J(x)] \right]$$

provided the right hand side is positive.

A.3.2 Identification of Payoffs with Unobservable Market-level States

In order to prove Propositions 8 and 9, we make use of the following Lemmas.

Lemma 10 *For any action a, the expectational error term $\varepsilon^{\zeta}(a, k, \omega, \omega^*)$ is mean independent of k, ω: $E\left[\varepsilon^{\zeta}(a, k, \omega, \omega^*) | k, \omega\right] = 0$.*

Proof. From the definition of $\varepsilon^{\zeta}(a, k, \omega, \omega^*)$,

$$\begin{aligned} E\left[\varepsilon^{\zeta}(a, k, \omega, \omega^*) | k, \omega\right] &= E\left[\sum_{k'} \varepsilon^{\zeta}(k', \omega, \omega^*) F^k(k'|a, k, \omega) | k, \omega\right] \\ = \quad & E\left[\sum_{k'} \left(\int_{\omega'} \zeta(k', \omega') dF^{\omega}(\omega'|\omega) - \zeta(k', \omega^*) \right) F^k(k'|a, k, \omega) | k, \omega\right] \\ = \quad & \sum_{k'} \int_{\omega'} \zeta(k', \omega') dF^{\omega}(\omega'|\omega) F^k(k'|a, k, \omega) - \sum_{k'} \int_{\omega^*} \zeta(k', \omega^*) dF^{\omega}(\omega^*|\omega) F^k(k'|a, k, \omega) \\ = \quad & 0. \end{aligned}$$

■

Note that the expectational error is also mean independent of all past (k, ω) (immediate consequence of the law of iterated expectations).

Lemma 11 *Consider the functions $g(k_{imt}, \omega_{mt})$ and $F(k'|k_{imt}, \omega_{mt})$. Assume w_{mt} is an observable subvector of ω_{mt} and consider the data set $\{(k_{imt}, w_{mt}, z_{mt}) : i = 1, ..., N; m = 1, ..., M; t = 1, ..., T\}$. Assume that for each m and t, one can obtain the estimators $\widehat{g}_{mt}^{N}(k) \xrightarrow{p}$*

$g(k,\omega_{mt})$ and $\widehat{F}^N_{mt}(k',k) \xrightarrow{p} F(k'|k,\omega_{mt})$ as $N\to\infty$.[54] *For any function $q(k,z_{mt})$, define the estimators*

$$\frac{1}{MT}\sum_{m,t=1}^{MT}\left[q(k,z_{mt})\,\widehat{g}^N_{mt}(k)\right], \text{ and}$$

$$\frac{1}{M(T-1)}\sum_{m,t=1}^{M(T-1)}\left[q(k,z_{mt})\,\widehat{g}^N_{mt+1}(k)\,\widehat{F}^N_{mt}(k',k)\right].$$

Assume the following uniform conditions hold: (i)

$$\lim\sup_{M,T,N}\left[\frac{1}{MT}\sum_{m,t=1}^{MT}E\left\|q(k,z_{mt})\,\widehat{g}^N_{mt}(k')\right\|\right]<\infty,$$

$$\lim\sup_{M,T,N}\left[\frac{1}{M(T-1)}\sum_{m,t=1}^{M(T-1)}E\left\|q(k,z_{mt})\,\widehat{g}^N_{mt+1}(k')\,\widehat{F}^N_{mt}(k',k)\right\|\right]<\infty; \tag{62}$$

and (ii)

$$\lim\sup_{M,T,N}\left[\frac{1}{MT}\sum_{m,t=1}^{MT}\left\|E\left[q(k,z_{mt})\left(\widehat{g}^N_{mt}(k')-g(k',\omega_{mt})\right)\right]\right\|\right]=0$$

$$\lim\sup_{M,T,N}\left[\frac{1}{M(T-1)}\sum_{m,t=1}^{M(T-1)}\left\|E\left[q(k,z_{mt})\begin{pmatrix}\widehat{g}^N_{mt+1}(k')\,\widehat{F}^N_{mt}(k',k)\\ -g(k',\omega_{mt+1})\,F(k'|k,\omega_{mt})\end{pmatrix}\right]\right\|\right]=0. \tag{63}$$

If (z_{mt},ω_{mt}) is i.i.d. across (m,t), or it is weakly dependent (in both indices t and m), and if

$$E\left\|q(k,z_{mt})\,g(k',\omega_{mt+1})\right\|<\infty,$$
$$E\left\|q(k,z_{mt})\,g(k',\omega_{mt+1})\,F(k'|k,\omega_{mt})\right\|<\infty,$$

then,

$$\frac{1}{MT}\sum_{m,t=1}^{MT}\left[q(k,z_{mt})\,\widehat{g}^N_{mt}(k)\right]\xrightarrow{p}E\left[q(k,z_{mt})\,g(k,\omega_{mt})\right], \text{ and}$$

$$\frac{1}{M(T-1)}\sum_{m,t=1}^{M(T-1)}\left[q(k,z_{mt})\,\widehat{g}^N_{mt+1}(k')\,\widehat{F}^N_{mt}(k',k)\right]\xrightarrow{p}E\left[q(k,z_{mt})\,g(k',\omega_{mt+1})\,F(k'|k,\omega_{mt})\right]$$

as $(N,M,T)\to\infty$; or as $(N,M)\to\infty$ (if T is fixed); or as $(N,T)\to\infty$ (if M is fixed).

Proof. We only consider the second estimator; the first estimator is handled similarly. The proof makes use of sequential convergence as a way to obtain joint convergence (e.g. Phillips

[54] Note that the asymptotic results $\widehat{g}^N_{mt}(k)\xrightarrow{p}g(k,\omega_{mt})$ and $\widehat{F}^N_{mt}(k',k)\xrightarrow{p}F(k'|k,\omega_{mt})$ as $N\to\infty$ can be obtained using the law of large numbers for exchangeable random variables (see, e.g., Hall and Heyde, 1980), provided the observations $i=1,..N$ for each index (m,t) are i.i.d. conditional on ω_{mt}.

and Moon, 1999, Lemma 6 and Theorem 1). The sequential limit can be obtained directly from two facts: (i) $\widehat{g}^{N}_{mt}(k) \xrightarrow{p} g(k,\omega_{mt})$ and $\widehat{F}^{N}_{mt}(k',k) \xrightarrow{p} F(k'|k,\omega_{mt})$ as $N \to \infty$ implies

$$\begin{aligned}
&\frac{1}{M(T-1)} \sum_{m,t=1}^{M(T-1)} \left[q(k,z_{mt}) \widehat{g}^{N}_{mt+1}(k') \widehat{F}^{N}_{mt}(k',k) \right] \\
&\xrightarrow{p} \frac{1}{M(T-1)} \sum_{m,t=1}^{M(T-1)} \left[q(k,z_{mt}) g(k',\omega_{mt+1}) F(k'|k,\omega_{mt}) \right]
\end{aligned}$$

as $N \to \infty$ for all (M,T). And (ii) provided (z_{mt},ω_{mt}) is i.i.d. across (m,t), or if the correlation of (z_{mt},ω_{mt}) and $(z_{m't'},\omega_{m't'})$ dies out as the distance between (m,t) and (m',t') increases, and provided

$$E\left\| q(k,z_{mt}) g(k',\omega_{mt+1}) F(k'|k,\omega_{mt}) \right\| < \infty,$$

then, by the Weak Law of Large Numbers,

$$\frac{1}{M(T-1)} \sum_{m,t=1}^{M(T-1)} \left[q(k,z_{mt}) g(k',\omega_{mt+1}) F(k'|k,\omega_{mt}) \right] \xrightarrow{p} E\left[q(k,z_{mt}) g(k',\omega_{mt+1}) F(k'|k,\omega_{mt}) \right]$$

as $(M,T) \to \infty$.

The sequential limit is obtained by first passing the limit $N \to \infty$ and then the limit $(M,T) \to \infty$. Provided conditions (62) and (63) hold, by Phillips and Moon's (1999) Lemma 6 and Theorem 1, the sequential limit equals the simultaneous limit as $(N,M,T) \to \infty$.[55] ■

A.3.3 Proof of Proposition 8

Let b denote the primitives of the model: $b = (\pi,\sigma,\beta,G,F)$.

(a) Finite T. Suppose first that the terminal value V_{mT} is known. Replace V_{mt+1} in (33) for a specific action j to get:

$$\pi_{amt} + \beta\varepsilon^{V}_{am,t,t+1} = V_{mt} - \beta F^{k}_{amt}\left[\pi_{jmt+1} + \beta\varepsilon^{V}_{jm,t+1,t+2} + \sigma\psi_{jmt+1} + \beta F^{k}_{jmt+1} V_{mt+2} \right] - \sigma\psi_{amt}$$

all a. Repeated substitution of $V_{mt+\tau}$ above leads to:

$$\begin{aligned}
\pi_{amt} + \beta\varepsilon^{V}_{am,t,t+1} \quad = \quad & V_{mt} - \beta F^{k}_{amt}\left[\sum_{\tau=1}^{T-1} \beta^{\tau-1} \Lambda_{jm,t,\tau} \left(\pi_{jmt+\tau} + \beta\varepsilon^{V}_{jm,t+\tau,t+\tau+1} + \sigma\psi_{jmt+\tau} \right) \right] \\
& -\beta^{T} F^{k}_{amt} \Lambda_{jm,t,T} V_{mT} - \sigma\psi_{amt} \qquad (64)
\end{aligned}$$

where the matrices $\Lambda_{jm,t,\tau}$ are defined recursively:

$$\begin{aligned}
\Lambda_{jm,t,\tau} &= I, \text{ for } \tau = 1 \\
\Lambda_{jm,t,\tau} &= \Lambda_{jm,t,\tau-1} F^{k}_{jmt+\tau-1}, \text{ for } \tau \geq 2.
\end{aligned}$$

[55] In general, the order of the limits can be misleading in cases in which all indices (N,M,T) pass to infinity simultaneously. We make use of the joint convergence because it holds under a wider range of circumstances than the sequential convergence.

Next, evaluate (64) for $a = j$ and subtract it to obtain:

$$\pi_{amt} - \pi_{jmt} + \beta\left(\varepsilon^V_{am,t,t+1} - \varepsilon^V_{jm,t,t+1}\right) = \beta^T\left(F^k_{jmt} - F^k_{amt}\right)\Lambda_{jm,t,T}V_{mT} - \sigma\left(\psi_{amt} - \psi_{jmt}\right) +$$

$$+\beta\left(F_{jmt} - F_{amt}\right)\left[\sum_{\tau=1}^{T-1}\beta^{\tau-1}\Lambda_{jm,t,\tau}\left(\pi_{jmt+\tau} + \beta\varepsilon^V_{jm,t+\tau,t+\tau+1} + \sigma\psi_{jmt+\tau}\right)\right]. \quad (65)$$

For any known vector function $q(z_{mt})$, with elements $q(k, z_{mt})$, apply the Hadamard multiplication on both sides of (65) and take expectation. We eliminate the error terms, $\varepsilon^V_{am,t+\tau,t+\tau+1}$ and $\xi_{akmt+\tau}$, because z_{mt} is in the time-t information set. Then,

$$\begin{aligned} E\left[q(z_{mt}) \circ \overline{\pi}_{amt}\right] &= E\left[q(z_{mt}) \circ \left(\overline{\pi}_{jmt} - \sigma\left(\psi_{amt} - \psi_{jmt}\right)\right)\right] + \\ &\quad +\beta E\left[q(z_{mt}) \circ \left(F_{jmt} - F_{amt}\right)\left[\sum_{\tau=1}^{T-1}\beta^{\tau-1}\Lambda_{jm,t,\tau}\left(\overline{\pi}_{jmt+\tau} + \sigma\psi_{jmt+\tau}\right)\right]\right] \\ &\quad +\beta^T E\left[q(z_{mt}) \circ \left(F^k_{jmt} - F^k_{amt}\right)\Lambda_{jm,t,T}V_{mT}\right]. \end{aligned} \quad (66)$$

where $\circ$ denotes the Hadamard product, and the expectations are taken over $(z_{mt}, \omega_{mt}, ..., \omega_{mT})$.

If the payoff $\overline{\pi}(j, k_{imt}, w_{mt})$, the scale parameter σ and the terminal value function V_{mT} are known, then the RHS of (66) can be recovered from the data (using the results of Lemma 11). Because the RHS of (66) is known, for any two structures b and b', with corresponding payoffs $\overline{\pi}$ and $\overline{\pi}'$, we have

$$E\left[q(z_{mt}) \circ \left(\overline{\pi}_{amt} - \overline{\pi}'_{amt}\right)\right] = 0$$

for any function q. By the completeness condition, the equality above implies $\overline{\pi}_{amt} - \overline{\pi}'_{amt} = 0$ almost everywhere.

Next, consider the case of a renewal action J. Take (35), multiply both sides by $q(z_{mt})$ and take expectations:

$$\begin{aligned} &E\left[q(z_{mt}) \circ \left(\overline{\pi}_{amt} - \overline{\pi}_{jmt} - \beta\left(F^k_{amt} - F^k_{jmt}\right)\overline{\pi}_{Jmt+1}\right)\right] \\ = \; &\sigma E\left[q(z_{mt}) \circ \left(\beta\left(F^k_{amt} - F^k_{jmt}\right)\psi_{Jmt+1} - \left(\psi_{amt} - \psi_{jmt}\right)\right)\right]. \end{aligned}$$

Similar to the previous case, the RHS can be recovered from data (using Lemma 11). Then, for any two structures b and b' with corresponding payoffs $\overline{\pi}$ and $\overline{\pi}'$,

$$\begin{aligned} &E\left[q(z_{mt}) \circ \left(\overline{\pi}_{amt} - \overline{\pi}_{jmt} - \beta\left(F^k_{amt} - F^k_{jmt}\right)\overline{\pi}_{Jmt+1}\right)\right] \\ = \; &E\left[q(z_{mt}) \circ \left(\overline{\pi}'_{amt} - \overline{\pi}'_{jmt} - \beta\left(F^k_{amt} - F^k_{jmt}\right)\overline{\pi}'_{Jmt+1}\right)\right]. \end{aligned}$$

By the completeness condition,

$$\overline{\pi}_{amt} - \overline{\pi}_{jmt} - \beta\left(F^k_{amt} - F^k_{jmt}\right)\overline{\pi}_{Jmt+1} = \overline{\pi}'_{amt} - \overline{\pi}'_{jmt} - \beta\left(F^k_{amt} - F^k_{jmt}\right)\overline{\pi}'_{Jmt+1} \quad (67)$$

for almost all (w_{mt}, w_{mt+1}). Consider (67) for $j = J$. Because $\overline{\pi}_J(k, w)$ is known for all observed states (k, w), we conclude that $\overline{\pi}_{amt} - \overline{\pi}'_{amt} = 0$ almost everywhere.

(b) Large T. Suppose again J is the renewal action. We do not necessarily assume that the flow payoffs of the renewal action $\overline{\pi}_J$ is known. Instead we now assume $\overline{\pi}_j$ is known for some action j for all k and w. This implies $\overline{\pi}_j = \overline{\pi}'_j$. Take the equation (67) above. Then, for almost all (w_{mt}, w_{mt+1}),

$$\overline{\pi}_{amt} - \overline{\pi}'_{amt} = \beta(F_{amt} - F_{jmt})(\overline{\pi}_{Jmt+1} - \overline{\pi}'_{Jmt+1})$$

It suffices to show that $\overline{\pi}_{Jmt} - \overline{\pi}'_{Jmt} = 0$, since then $\overline{\pi}_{amt} = \overline{\pi}'_{amt}$, all a. Set $a = J$ in the above equation and evaluate recursively for any t,

$$\overline{\pi}_{Jm\tau} - \overline{\pi}'_{Jm\tau} = \beta^T \prod_{t=\tau}^{T} (F_{Jmt} - F_{jmt})(\overline{\pi}_{JmT} - \overline{\pi}'_{JmT}) \tag{68}$$

Because of the renewal property, we have:

$$\prod_{t=\tau}^{T} (F_{Jt} - F_{jt}) = (F_{j\tau} - F_{J\tau}) \prod_{t=\tau}^{T} F_{jt},$$

and note that $\|F_{Jt} - F_{jt}\| \leq 2$ for any t. Because the product of stochastic matrices is stochastic,

$$\left\| \left[\prod_{t=\tau}^{T} F_{jt} \right] \right\| = 1.$$

Putting the claims together,

$$\left\| (F_{j\tau} - F_{J\tau}) \left[\prod_{t=\tau}^{T} F_{jt} \right] \right\| \leq \|F_{j\tau} - F_{J\tau}\| \left\| \left[\prod_{t=\tau}^{T} F_{jt} \right] \right\| \leq 2.$$

Since $\beta < 1$ the sequence in the right hand side of (68) converges to zero provided the flow payoffs $\overline{\pi}_J(k, w)$ and $\overline{\pi}'_J(k, w)$ are bounded for almost all (k, w).

A.3.4 Proof of Proposition 9

The proof is similar to that of Proposition 8. Suppose first that σ is known. Multiply both sides of (30) by a known function $q(k, z_{mt})$ and take expectations:

$$\begin{aligned} E[q(k, z_{mt})\overline{\pi}(a, k, w_{mt})] &= E[q(k, z_{mt})(V(k, \omega_{mt}) - \sigma\psi(a, k, \omega_{mt}))] \\ -\beta E &\left[q(k, z_{mt}) \sum_{k'} V(k', \omega_{mt+1}) F^k(k'|a, k, w_{mt}) \right] \end{aligned} \tag{69}$$

where the expectations are taken over $(z_{mt}, \omega_{mt}, \omega_{mt+1})$. The RHS can be recovered from data (using Lemma 11). Then, for any two structures b and b' with corresponding payoffs $\overline{\pi}$ and $\overline{\pi}'$,

$$E[q(k, z_{mt})(\overline{\pi}(a, k, w_{mt}) - \overline{\pi}'(a, k, w_{mt}))] = 0.$$

By the completeness condition, $\overline{\pi}(a, k, w_{mt}) - \overline{\pi}'(a, k, w_{mt}) = 0$ almost everywhere.

Next, suppose σ is not known, but $\overline{\pi}(a, k, w)$ is known for one combination of (a, k, w). Take (69) for the known $\overline{\pi}(a, k, w)$:

$$\begin{aligned} \sigma E[q(k, z_{mt})\psi(a, k, \omega_{mt})] &= E[h(k, z_{mt}) V(k, \omega_{mt})] \\ -\beta E &\left[q(k, z_{mt}) \sum_{k'} V(k', \omega_{mt+1}) F^k(k'|a, k, w) \right] - E[q(k, z_{mt})\overline{\pi}(a, k, w)]. \end{aligned}$$

Table 6: 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 types	http://www.gaez.iiasa.ac.at/
SRTM	Topographical – altitutude 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.htm

The RHS is known, which implies for any two σ and σ',

$$(\sigma - \sigma')\, E\left[q\left(k, z_{mt}\right) \psi\left(a, k, \omega_{mt}\right)\right] = 0,$$

and so $\sigma = \sigma'$.

B Appendix: Data (for online publication)

Table 6 lists our data sources. All are publicly available for download save DataQuick's land value data. The Cropland Data Layer (CDL) is a high-resolution (30-56m) annual land-use data that covers the entire contiguous United States since 2008.[56]

Next, we merge the above with an extensive database of land transactions in the United States obtained from DataQuick. 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 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 in DataQuick.

[56]The CDL rasters were processed to select an 840m subgrid of the original data, and then points in this grid were matched across years to form a land use panel. The grid scale was chosen for two reasons. First, it provides comprehensive coverage (i.e., most 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 7: 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.

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 much more sparse information on pasture land rental rates (see Scott (2013)).

Table 7 presents some summary statistics. Table 8 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.

C Appendix: Estimation (for online publication)

C.1 Conditional Choice Probabilities

We estimate conditional choice probabilities using a semiparametric logit regression. We are fully flexible over field states and year, but smooth across counties. In particular, we

Table 8: 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

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'. 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-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 (within counties).

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.[57]

C.2 Resale Price Regression

Next, 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

[57]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.

transactions taking place every year.[58] 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:

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

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 7 (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 dummies.

Table 9 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 $\theta_0 (a, k)$.

Finally, we 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 8 of Appendix 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 10). 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%. 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.

[58] 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 transnactions) out of 14.5 million acres of farmland (seemingly information on other states is not available).

Table 9: Hedonic Regression

	(1)
VARIABLES	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.

Table 10: Land use and transactions

	(1)	(2)	(3)	(4)
VARIABLES	incrops2010	incrops2011	incrops2012	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.