

Identification of Counterfactuals and Payoffs in Dynamic Discrete Choice with an Application to Land Use*

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

August 2015

Abstract

Dynamic discrete choice models are non-parametrically not identified without restrictions on payoff functions, yet counterfactuals may be identified even when payoffs are not. We provide necessary and sufficient conditions for the identification of a wide range of counterfactuals for models with nonparametric payoffs, as well as for commonly used parametric functions, and we obtain both positive and negative results. We show that access to extra data of asset resale prices (when applicable) can solve non-identifiability of both payoffs and counterfactuals. The theoretical findings are illustrated empirically in the context of agricultural land use. First, we provide identification results for models with unobserved market-level state variables. Then, using a unique spatial dataset of land use choices and land resale prices, we estimate the model and investigate two policy counterfactuals: long run land use elasticity (identified) and a fertilizer tax (not identified, affected dramatically by restrictions).

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

*We are grateful to Victor Aguirregabiria, Peter Arcidiacono, Lanier Benkard, Bo Honoré, Claudio Lucinda, Thierry Magnac, Junichi Suzuki, Ali Yurukoglu, and seminar participants at various universities for many helpful comments. Adrian Torchiana provided excellent research assistance. The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme Grant Agreement no. 230589 and Agence Nationale de la Recherche Projet ANR-12-CHEX-0012-01. All remaining errors are our own.

†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

Dynamic considerations are key in many decisions made by individuals and firms. For example, when choosing how to use their land, farmers face both severe uncertainty (volatile commodity prices, ever-changing government policies), as well as substantial switching costs (moving from forest to crop production is costly, if not irreversible). In a wide range of applied settings, researchers use dynamic discrete choice models with the goal of conducting positive and normative analyses (most often in the form of counterfactuals).¹ However, dynamic discrete choice models are non-parametrically not identified without “normalizing” the payoffs of one chosen action at every state (Rust (1994), Magnac and Thesmar (2002)). Whether the “normalizations” play a fundamental role in the identification of behavior and welfare in the counterfactuals of interest remains an open question.

In this paper, we fully characterize when counterfactuals are nonparametrically identified. Specifically, we provide necessary and sufficient conditions for the identification of a wide range of counterfactuals for models with nonparametric payoffs, as well as for commonly used parametric functions, and we obtain both positive and negative results. We show that access to extra data of asset resale prices (when applicable) resolves non-identifiability of both payoffs and counterfactuals, while substantially facilitating estimation. We then demonstrate our theoretical findings in the context of agricultural land use (Scott (2013)). We provide a dynamic model that allows for unobserved market-level state variables and establish identification for this set of models. We then construct a unique spatial dataset of land use choices and land resale prices, estimate the model, and conduct two policy relevant counterfactuals; the long-run elasticity of land use which is identified and a fertilizer tax, which is not identified and is found to be substantially affected by the normalization.

The desire for identification of models with as few assumptions as possible, as well as for clarity regarding the impact of each assumption in shaping identified parameters, has led to a growing literature on nonparametric identification (e.g. see Matzkin (2007) and references therein). It is well-known that payoffs in dynamic discrete choice models are not identified nonparametrically without *a priori* restrictions such as prespecifying the payoff of one action in *all* states (Rust (1994), Magnac and Thesmar (2002)). We call this restriction a “strong normalization.” A common strong normalization is to make the payoff of one action equal to zero; we show that alternative payoff restrictions often employed (e.g. exclusion restrictions) are in fact equivalent to strong normalizations. Unlike static models, strong normalizations in dynamic models are not without loss, as returns in some states can affect incentives in

¹Applications include occupational choice (e.g. Miller (1984)); fertility (e.g. Wolpin (1984)); patent renewal (e.g. Pakes (1986)); machine replacement (e.g. Rust (1987)); job search (e.g. Wolpin (1987)); firm entry and exit (e.g. Aguirregabiria, Mira and Roman (2007), Collard-Wexler (2013), Dunne, Klimek, Roberts and Xu (2013)); agricultural policies (e.g. Rosenzweig and Wolpin (1993), Scott (2013)); environmental policy regulation (e.g. Ryan (2012)); demand for durable goods (e.g. Hendel and Nevo (2006), Gowrisankaran and Rysman (2012), Conlon (2010)); learning (e.g. Crawford and Shum (2005)); insurance (e.g. Jeziorski, Krasnokutskaya and Ceccarini (2014)); export dynamics (e.g. Das, Roberts and Tybout (2007)); housing (e.g. Suzuki (2013), Bayer, Murphy, McMillan and Timmins (2011)); health (e.g. Gilleskie (1998), Fang and Wang (2015), Chan, Hamilton and Papageorge (2015)); retirement decisions (e.g. Rust and Phelan (1997)); schooling (e.g. Eckstein and Wolpin (1999)); education (e.g. Todd and Wolpin (2006), Duflo, Hanna and Ryan (2012)); labor market participation (e.g. Eckstein and Lifshitz (2011)); incentives to politicians (e.g. Diermeier, Keane and Merlo (2005)), among others. For recent surveys of this literature see Aguirregabiria and Mira (2010), Arcidiacono and Ellickson (2011), and Keane, Todd and Wolpin (2011).

other states. Therefore, if the value of the normalized payoff changes over time or over states, an important source of misspecification (even in nonparametric settings) may have been introduced with nontrivial implications on objects of interest such as counterfactuals.

Adopting a novel approach, we characterize the identification of a general class of counterfactuals commonly used in applied work. A counterfactual is defined as a transformation of payoffs and transition probabilities. We show that counterfactual choice probabilities are identified if and only if the counterfactual transformation satisfies certain restrictions: its Jacobian needs to lie in a subspace determined by state transitions. In other words, to identify counterfactual choice probabilities, the counterfactual transformation can only interact with model primitives in specific ways. Even though the payoff function is not identified without restrictive assumptions, some common counterfactuals are.² For instance, counterfactuals involving only lump-sum transfers are identified. However, numerous counterfactuals of interest are not identified, such as interventions that change payoffs proportionally with different proportions for different actions. Furthermore, we consider counterfactual welfare and show that welfare changes may not be identified even when counterfactual choice probabilities are. We illustrate our findings in a numerical exercise based on a simplified version of the bus engine replacement problem (Rust (1987)).³ In short, to the best of our knowledge, we offer the first general set of necessary and sufficient conditions to nonparametrically identify counterfactual behavior and welfare in dynamic discrete choice models.

When a counterfactual of interest is not identified, the researcher can consider additional restrictions. For example, it is well known that parametric restrictions can aid the identification of dynamic models (Magnac and Thesmar (2002), Arcidiacono and Miller (2015)). At the same time, economic theory, institutional details and computational constraints often justify them. Here, we show that parametric assumptions also enlarge the set of counterfactuals that are identified. Indeed, we consider payoff parameterizations that involve two components, one that is identifiable (e.g. variable profit) and one that is not identified, except under some strong normalization (e.g. fixed, entry or exit costs). Counterfactual transformations that affect only the identifiable component of payoffs result in identified counterfactual choice probabilities. Imposing a strong normalization does not bias results in this case. In contrast, counterfactuals that affect only the non-identifiable component result in non-identified counterfactuals. We see both types of counterfactuals in applied work; for example, in an entry model, counterfactuals that affect variable profits are identified, while counterfactuals that change entry costs are not.

Access to asset resale prices (when applicable) can also be helpful. In applications where the dynamic optimization problem involves a durable asset (e.g. a farmer and his field), asset resale prices can provide substantial information about value functions. In fact, in setups with constant returns to scale and homogeneous buyers and sellers, resale prices are equal to value functions (Kalouptsi, 2014a). We show that in this case, resale prices allow for identification of payoffs without the need for strong normalizations.⁴

Before turning to our empirical exercise, we consider the case of unobservable market-level

²This is reminiscent of Marschak (1953), who pointed out that the impact of hypothetical policies of interest may be identified even when the full model is not.

³Aguirregabiria and Suzuki (2014) also present substantial bias in a numerical exercise involving a simple dynamic entry and exit model.

⁴Aguirregabiria and Suzuki (2014) consider a more general bargaining model and also show that resale prices allow for payoff identification.

state variables. In our land use application, as well as several other settings, the available observed market states are likely insufficient to capture the perceived market heterogeneity and its evolution (e.g. changing government policy, unobserved costs). We thus extend the (non)identification results to settings with partially-observed market states. In summary, we provide identification theorems for dynamic discrete choice models in four different data environments: (i) agent actions and states are observed (a strong normalization is necessary); (ii) agent actions, states and resale prices are observed (a strong normalization is *not* necessary); (iii) agent actions and part of the state are observed (a strong normalization *and* extra restrictions, such as the presence of a renewal or a terminal action, are necessary); (iv) agent actions, part of the state, and resale prices are observed (a strong normalization and extra restrictions are *not* necessary). Although identification in case (i) has long been established in the literature, to the best of our knowledge, the results for the other three cases are new. Our identification arguments are constructive and lead naturally to estimators.

Finally, we demonstrate our theoretical findings in the context of agricultural land use. Field owners decide whether to plant crops or not and face uncertainty regarding commodity prices, weather shocks, and government interventions. We construct a unique dataset by spatially merging data from several sources. We employ high-resolution annual land use, obtained from the Cropland Data Layer database (CDL). We merged this dataset with NASA’s Shuttle Radar Topography Mission database, which provides extremely detailed topographical information. Land resale transaction data from DataQuick are merged with the above, as well as fine soil information from GAEZ. Our dataset is the first to allow for such rich field heterogeneity, combined with land resale prices.

We consider three estimators for our agricultural land use model. The first, which we call the CCP estimator (after conditional choice probability), is adapted from Scott (2013); following the tradition of Hotz and Miller (1993), it employs observed choices and states. The second estimator, which we call the “joint estimator,” considers the moments of the CCP estimator, plus the moment restrictions obtained from resale prices; all moments are used jointly to estimate payoffs. Resale prices are bound to affect the model estimates beyond the strong normalization. To isolate the impact of strong normalizations on counterfactuals, we consider a third estimator, which we call the “hybrid.” The hybrid estimator employs the CCP moments plus a number of resale price moments sufficient to obtain payoffs that otherwise would be strongly normalized.

After estimating the model, we implement two policy counterfactuals: the long-run elasticity (LRE) of land use and an increase in the costs to prepare land to replant crops. The LRE measures the long-run sensitivity of land use to an (exogenous) change in crop returns; it is an important input to the analysis of several policy interventions, including agricultural subsidies and biofuel mandates (Roberts and Schlenker (2013), Scott (2013)). This counterfactual is identified and the relevant estimators produce the same elasticity. For the second counterfactual, we increase replanting costs based on the estimated benefits of leaving land out of crops for a year. Because one such benefit is to allow soil nutrient levels to recover, it lessens the need for fertilizer inputs after a year of fallow land. This counterfactual can therefore be interpreted as a fertilizer tax. This is relevant: as the production of nitrogenous fertilizer involves high levels of greenhouse gas inputs, the cost of fertilizer would probably increase in response to comprehensive greenhouse gas pricing. Yet, because it is difficult to know the fertilizer saved by leaving land fallow, we use the estimated switching cost param-

eters to implement the counterfactual. This counterfactual is not identified and the strong normalization has a substantial impact on the results, reversing the direction of the change in cropland.

Related Literature. As already noted, Rust (1994) showed that payoffs in dynamic discrete choice models are nonparametrically not identified, while Magnac and Thesmar (2002) characterized the degree of underidentification. There is very little work on the identification of counterfactuals: Aguirregabiria (2010), Aguirregabiria and Suzuki (2014), Norets and Tang (2014), and Arcidiacono and Miller (2015) are the exceptions and focus on special cases. They investigate counterfactuals that add a vector to payoffs or arbitrarily change the stochastic process for the state variables, and provide sufficient conditions to identify counterfactual choice probabilities. Aguirregabiria (2010) focuses on a finite-horizon binary choice model; Aguirregabiria and Suzuki (2014) and Norets and Tang (2014), on infinite-horizon binary choice models; and Arcidiacono and Miller (2015), on both stationary and nonstationary multinomial choice models with short and long panel data. We consider stationary infinite-horizon models and multinomial choices, as well as any (nonlinear) differentiable changes in payoffs. We provide both necessary and sufficient conditions to identify both counterfactual choice probabilities and welfare evaluations.

The use of asset resale price data to identify the dynamic model builds on the insights of Kalouptside (2014a, 2014b). In a different context, Keane and Wolpin (1997) and Heckman and Navarro (2007) also make use of extra data to secure identification in finite-horizon models (in their case, labor outcomes such as future earnings). Finally, despite a significant literature on individual level heterogeneity (e.g. Kasahara and Shimotsu (2009), Norets (2009), Arcidiacono and Miller (2011), Hu and Shum (2012), and Connault (2014)), little work exists relating to the presence of serially correlated unobserved market states in single-agent (and industry) dynamics.

The paper is organized as follows: In Section 2, we set out the standard dynamic discrete choice framework; in Section 3, we present identification results for the standard model, as well as for common parametrizations. Section 4 investigates when counterfactuals are identified. In Section 5, we discuss the use of resale prices. Section 6 provides identification results in the presence of unobserved market-level state variables (with and without resale price data). Section 7 adapts the standard framework to our agricultural land use application, while in Section 8, we describe the data and provide summary statistics. Section 9 presents the estimation and counterfactual results. Section 10 concludes.

2 Dynamic Discrete Choice Framework

Time is discrete and indexed by t ; the time horizon is infinite. In period t , agent i chooses an action a_{it} from a finite set of possible actions $A = \{1, \dots, |A|\}$. The agent's state is denoted by s_{it} and follows a controlled Markov process with transition distribution function $F(s_{it+1}|a_{it}, s_{it})$. Every period t , the agent observes the state s_{it} and chooses an action a_{it} to maximize the discounted expected payoff

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

where $\pi(\cdot)$ is the current payoff function, and $\beta \in (0, 1)$ is the discount factor.

Let $V(s_{it})$ be the value function of the dynamic programming problem. By Bellman's principle of optimality,⁵

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

We follow the literature in splitting the state into two components, $s_{it} = (x_{it}, \varepsilon_{it})$, that satisfy the following standard conditions:⁶

Condition 1 (*Additive separability*) The current payoff is given by

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

where $\varepsilon_{it} = (\varepsilon_{it}(1), \dots, \varepsilon_{it}(|A|))$, and $\sigma > 0$ is a scale parameter.

Condition 2 (*i.i.d. unobservables*) The vector ε_{it} is i.i.d. across agents and time with distribution function G that is absolutely continuous with respect to Lebesgue measure in $\mathbb{R}^{|A|}$.

Condition 3 (*Conditional Independence*) The transition distribution function for $s_{it} = (x_{it}, \varepsilon_{it})$ factors as

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

Condition 4 (*Discrete Support*) The support of x_{it} is finite: $x_{it} \in X = \{x_{(1)}, x_{(2)}, \dots, x_{(|X|)}\}$, with $|X| < \infty$.

Given these assumptions, agent i 's Bellman equation becomes:

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

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*:

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

so that

$$V(x_{it}) = \int \max_{a \in A} \{ v_a(x_{it}) + \sigma \varepsilon_{it}(a) \} dG(\varepsilon_{it}).$$

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

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

⁵We can allow the set of possible actions A to depend on the state s ; i.e. $a \in A(s)$.

⁶Together with the assumptions that agents are expected utility maximizers and have rational expectations, Conditions 1-4 form a set of *maintained assumptions*. Maintained assumptions are taken as given (not questioned nor altered) when investigating model identification (Rust (2014)).

where $1\{\cdot\}$ is the indicator function. Define the $|A| \times 1$ vector of conditional choice probabilities $p(x) = \{p_a(x) : a \in A\}$, and the corresponding $|A| |X| \times 1$ vector $p = \{p(x) : x \in X\}$. It is most often assumed that ε_{it} follows the extreme value distribution; in that case, the conditional choice probabilities are given in closed form as

$$p_a(x) = \frac{\exp(v_a(x)/\sigma)}{\sum_{j \in A} \exp(v_j(x)/\sigma)}.$$

2.1 Value Function and CCP Relationships

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

Lemma 5 (*Hotz-Miller inversion*): *Assume Conditions 1-4 hold. For all $(a, x) \in A \times X$, and for some given reference action j ,*

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

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

Naturally, for $a = j$ we have $\phi_{jj}(p(x)) = 0$. The Hotz-Miller inversion allows us to recover the conditional value functions $v_a(x)$, $a \neq j$, from the choice probabilities $p(x)$, given $\{G, \sigma, v_j(x)\}$.

Arcidiacono and Miller (2011) obtain an implication of the Hotz-Miller inversion:

Lemma 6 (*Arcidiacono-Miller*): *Assume Conditions 1-4 hold. For any $(a, x) \in A \times X$, 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 .

In the case of logit errors, the Hotz-Miller inversion becomes:

$$\phi_{aj}(p(x)) = \log p_a(x) - \log p_j(x),$$

while the Arcidiacono-Miller lemma becomes:

$$\psi_a(p(x)) = \log \left(1 + \sum_{j \neq a} \frac{p_j(x)}{p_a(x)} \right) + \gamma,$$

where γ is the Euler constant.

2.2 Empirical Model

Before closing this section, we augment our discussion of general dynamic discrete choice models by a typical empirical treatment; the agricultural land use model presented in Section 7 also fits this empirical framework.

In numerous applications, the state variable x can be decomposed into $x = (k, \omega)$, where $k \in K$ are states whose evolution can be affected by individuals' choices, and $\omega \in \Omega$ are states not affected by agent's choices (e.g., market-level states), with K, Ω finite. In other words, it is common to assume:⁷

$$F(x'|a, x) = F^k(k'|a, k) F^\omega(\omega'|\omega). \quad (2)$$

In addition, we consider the following common separable payoff function

$$\pi(a, k, \omega) = \theta_0(a, k) + R(a, \omega)' \theta_1(a, k), \quad (3)$$

where $R(a, \omega)$ is a known vector-valued function and $'$ denotes transpose (often, $R(a, \omega) = \omega$). As the following two examples from the literature indicate, $\theta_0(a, k)$ is often interpreted as fixed costs and $R(a, \omega)$, as a measure of returns. In Section 7, we interpret $\theta_0(a, k)$ as land use switching costs, while $R(a, \omega)$ measures crop/non-crop returns.

Example: 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. The agent has two actions: to replace or keep the engine, $A = \{\text{replace, keep}\}$. The state variable, k , is the bus mileage which evolves stochastically and is renewed upon replacement. The payoff is decomposed as in (3): $\theta_0(\text{replace}, k)$, captures the cost of obtaining a new engine, while $\theta_0(\text{keep}, k) = 0$; the second term is a flexible function of mileage and captures the engine operating costs (see Section 4.4 for a similar model).

Example: Monopolist Entry/Exit Problem. Consider a monopolist deciding whether to be active in or exit from a market, so that $A = \{\text{active, exit}\}$. Let $k_{it} = a_{it-1}$, and let ω be a vector of market characteristics relevant for the firm's variable profits $\bar{\pi}_1$, fixed costs FC , scrap values ϕ^s , and entry costs ϕ^e (more firm-specific states can be added). The firm's flow payoff is

$$\pi(a, k, \omega) = \begin{cases} k\phi^s(\omega) & \text{if } a = 0 \text{ (exit)} \\ k(\bar{\pi}_1(\omega) - FC(\omega)) - (1 - k)\phi^e(\omega), & \text{if } a = 1 \text{ (active)} \end{cases}$$

Often, these functions are parametric and linear in the parameters (e.g. $\bar{\pi}_1(\omega) = \theta_0^\pi + \theta_1^\pi \omega$) and the above payoff maps into (3).

3 Identification of the Standard Dynamic Discrete Choice Model

A dynamic discrete choice model consists of the primitives $b = (\pi, \sigma, \beta, G, F)$. We refer to the set of primitives b as a *structure*. A structure gives rise to the endogenous objects $\{p_a, v_a, V : a \in A\}$. The question of interest is whether we can identify the structure b , or some feature of b , from the data. A *feature* ζ is a function of b , such as an element of it (e.g. $\zeta(b) = \pi$). We assume throughout this paper that (β, G) are known.

Denote the dataset available to the econometrician by $\{y_{it} : i = 1, \dots, N; t = 1, \dots, T_i\}$, where T_i is the number of periods over which i is observed. In the standard dynamic discrete choice dataset, it consists of agents' actions and states, $y_{it} = (a_{it}, x_{it})$. We assume the joint

⁷Many of our results hold if we allow F^k to depend on ω as well.

distribution of y_{it} , $\Pr(y)$, is known, which implies the conditional choice probabilities $p_a(x)$ and the transition distribution function F are also known.

It is well known that in static discrete choice models utility functions are identified only up to positive affine transformations. Without loss, the researcher must fix the location and scale of the payoff function. Typically, a normalization takes one value of (a, x) for which $\pi(a, x)$ is set to zero (location), while the variance of ε is fixed (scale). We call a positive affine transformation like this a *weak normalization*.

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

$$\pi_a = v_a - \beta F_a V, \quad \text{for } a = 1, \dots, |A| \quad (4)$$

$$v_a - v_j = \sigma \phi_{aj}, \quad \text{for } a = 1, \dots, |A|, a \neq j \quad (5)$$

$$V = v_a + \sigma \psi_a, \quad \text{for } a = 1, \dots, |A| \quad (6)$$

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 (4) restates the definition of the conditional value function found in (1); (5) restates the Hotz-Miller lemma; and (6), 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 .

The objective is to identify the feature $\zeta(b) = (\pi, \sigma)$. Equations (4)-(6) form a set of $(3|A| - 1)|X|$ linear restrictions in $(2|A| + 1)|X| + 1$ unknowns: the vector $[\pi_a, V, v_a]'$ and the scale parameter σ . To solve this system, assume first that the value of σ is known (fixed by normalizing the variance of ε_{it} , as is done in the literature).

Consider a binary choice for illustration. The system of equations can be written in matrix form:⁸

$$\begin{bmatrix} I & 0 & \beta F_1 & -I & 0 \\ 0 & I & \beta F_2 & 0 & -I \\ 0 & 0 & 0 & I & -I \\ 0 & 0 & I & -I & 0 \\ 0 & 0 & I & 0 & -I \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ V \\ v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \sigma \phi_{12} \\ \sigma \psi_1 \\ \sigma \psi_2 \end{bmatrix} \quad (7)$$

Note that the last three block rows of the matrix above are linearly dependent. Therefore, the above system has no unique solution and π is not identified. The dimension of the set of solutions is given by the cardinality of the state space: there are $|X|$ free parameters in the system. This is summarized in the following proposition:

Proposition 7 *Assume Conditions 1-4 hold. Suppose the joint distribution of observables $\Pr(y)$, where $y_{it} = (a_{it}, x_{it})$, is known. Provided the primitives (σ, β, G) are known, the payoff function π_a , for each $a \neq J$, can be represented without loss of generality as an affine transformation of π_J :*

$$\pi_a = A_a \pi_J + b_a \quad (8)$$

where

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

$$b_a = \sigma(A_a \psi_J - \psi_a).$$

⁸ I is the $|X| \times |X|$ identity matrix and (with some abuse of notation) 0 on the left-hand-side is the $|X| \times |X|$ zero matrix, while 0 in the right-hand-side is the $|X| \times 1$ zero vector.

Proof. 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 have⁹

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

■

Before discussing Proposition 7, we fix some notation. Assume without loss that $J = |A|$. Define $\pi = [\pi'_1, \pi'_2, \dots, \pi'_J]' \in \mathbb{R}^{|X||A|}$; and $\pi_{-J} = [\pi'_1, \pi'_2, \dots, \pi'_{J-1}]' \in \mathbb{R}^{|X|(|A|-1)}$. Define also the matrix $A_{-J} = [A'_1, A'_2, \dots, A'_{J-1}]' \in \mathbb{R}^{|X|(|A|-1) \times |X|}$ and $b_{-J} = [b'_1, b'_2, \dots, b'_{J-1}]' \in \mathbb{R}^{|X|(|A|-1)}$. Then, (8) becomes:

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

The non-identifiability of current payoffs in dynamic discrete choice has long been established in the literature (Rust (1994), Magnac and Thesmar (2002)).¹⁰ To obtain a unique solution to the system, and point identify π , we need to add $|X|$ extra restrictions. Suppose that the researcher imposes a set of $|X|$ linear restrictions:

$$R\pi = r \tag{9}$$

where $R \in \mathbb{R}^{|X| \times (|A||X|)}$, $r \in \mathbb{R}^{|X|}$. We can write $R = [R_{-J} \ R_J]$ and stack (8) and (9) to obtain:

$$\begin{bmatrix} I & -A_{-J} \\ R_{-J} & R_J \end{bmatrix} \begin{bmatrix} \pi_{-J} \\ \pi_J \end{bmatrix} = \begin{bmatrix} b_{-J} \\ r \end{bmatrix}. \tag{10}$$

Define $\mathbf{C} = \begin{bmatrix} I & -A_{-J} \\ R_{-J} & R_J \end{bmatrix}$ and $\mathbf{c} = \begin{bmatrix} b_{-J} \\ r \end{bmatrix}$. The following result is immediate:

Lemma 8 *Suppose the conditions of Proposition 7 hold. For any restriction (R, r) such that $\text{rank}(\mathbf{C}) = |A||X|$, the payoff function π is identified.*

Any given restriction (R, r) satisfying this Lemma selects one point in the set of possible solutions for π . A common solution is to fix, or “normalize”, the payoff of some action for all states; in which case $R = [0 \ I]$ and r is fixed at some level. Typically, the payoff of the action we have least information about is set to zero (usually referred to as the “outside option”).¹¹

We prefer to call this solution a *strong normalization*. While a weak normalization requires fixing, for example, $\pi(a, x) = 0$ for a single (a, x) pair, a strong normalization requires

⁹ $(I - \beta F_J)$ is invertible since F_J is a stochastic matrix.

¹⁰Arcidiacono and Miller (2015) provide recent results for nonstationary models; Pesendorfer and Schmidt-Dengler (2008), for dynamic games.

¹¹In principle, one can fix any $|X|$ elements of the vector $[\pi_1, \pi_2, V, v_1, v_2]'$ to some prespecified values. However, it is undesirable to fix instead v_1 , say, because it is an endogenous object and interpreting π in terms of this normalization is both difficult and not transferable to counterfactuals.

$\pi(a, x) = 0$ for all x . This is in fact an additional restriction on the model. Different strong normalizations, although observationally equivalent given data on choices and states, can result in different expected continuation values, and so in different behavior and welfare in counterfactual scenarios.

Strong normalization is one type of restriction that leads to identification. In several applications, researchers use different restrictions; for example, “exclusion restrictions” assume that some payoffs do not depend on all state variables (for instance, firm entry/exit costs are often assumed state invariant). It is clear from Proposition 7 and the Lemma, that any set of extra linear restrictions that suffices to identify π is equivalent to some form of strong normalization. If the system (10) has a unique solution, π^* , then (8) along with the strong normalization $\pi_J = \pi_J^*$ leads to the same solution (the converse also is true).¹²

Before ending our discussion on the identification of standard discrete choice models, we discuss two other prevalent empirical practices: parametric payoffs and the use of renewal actions. Economic theory and institutional details often justify these practices, while computational constraints (curse of dimensionality) favors parsimonious specifications. In the agricultural land use model discussed in Section 7, we assume both a parametric profit function and the presence of a renewal action. Below, we investigate whether these extra restrictions help identification of payoffs; when they do, we show in Section 4 that they also help identifying some types of counterfactual exercises.

3.1 Parametric Payoffs

As the dimension of payoffs $|A| \times |X|$ is often large, empirical work usually resorts to parameterizations (see also Section 2.1), which imposes $\pi(a, x) = \pi(a, x, \theta)$ with $\dim(\theta) \ll |A| \times |X|$. We discuss identification of parametric payoffs through the lens of equation (8). In this case, (8) becomes:

$$\pi_{-J}(\theta) = A_{-J}\pi_J(\theta) + b_{-J}.$$

We differentiate with respect to θ :

$$\frac{\partial \pi_{-J}(\theta)}{\partial \theta} = A_{-J} \frac{\partial \pi_J(\theta)}{\partial \theta}$$

or, more compactly,

$$\begin{bmatrix} I & -A_{-J} \end{bmatrix} \frac{\partial \pi(\theta)}{\partial \theta} = 0.$$

If the matrix above has full rank, θ is locally identified.

For example, consider a linear-in-parameters payoff function:

$$\pi_a(\theta) = \bar{\pi}_a \theta.$$

Then, θ is determined by the linear system of equations:

$$[\bar{\pi}_{-J} - A_{-J}\bar{\pi}_J] \theta = b_{-J} \tag{11}$$

which has a unique solution if $[\bar{\pi}_{-J} - A_{-J}\bar{\pi}_J]$ has full rank.

¹²For instance, an exclusion restriction assuming that $\pi_J(x)$ does not depend on x , and weakly normalizes $\pi_2(x') = c$ for some x' , is equivalent to a strong normalization that fixes $\pi_J(x) = c$ for all x .

Next, consider the empirical model (2)-(3) of Section 2.1. In this case, the matrix in (11) does not have full rank and thus a form of normalization is needed. Indeed, the next proposition shows that, although one can identify $\theta_1(a, k)$ when there is “sufficient variation” in ω , $\theta_0(a, k)$ is not identified. A strong normalization is required (e.g. $\theta_0(J, k)$ needs to be known or prespecified at all k). For notational simplicity, we focus on binary choice with $A = \{a, J\}$ and assume $R(a, \omega)$ is scalar. The transition matrix in this case is $F_a = F^\omega \otimes F_a^k$, where \otimes is the Kronecker product.¹³

Proposition 9 *Consider the empirical model of Section 2.2. Assume (σ, β, G) are known. Suppose $|\Omega| \geq 3$ and there exist $\omega, \tilde{\omega}, \bar{\omega}$ such that the matrix*

$$\begin{bmatrix} (e'_\omega - e'_{\tilde{\omega}}) D_a R_a & (e'_{\tilde{\omega}} - e'_\omega) D_J R_J \\ (e'_\omega - e'_{\bar{\omega}}) D_a R_a & (e'_{\bar{\omega}} - e'_\omega) D_J R_J \end{bmatrix} \quad (12)$$

is invertible; where $D_a = [I - \beta (F^\omega \otimes F_a^k)]^{-1}$, $R_a = [R_a(\omega_1)I_k, \dots, R_a(\omega_\Omega)I_k]'$ (the same for R_J), I_k the identity matrix of size K and $e'_\omega = [0, 0, \dots, I_k, 0, \dots, 0]$ with I_k in the ω position. Then given the joint distribution of observables $\Pr(y)$, where $y_{it} = (a_{it}, k_{it}, \omega_{it})$, the parameters $\theta_1(a, k)$ are identified. The parameters $\theta_0(a, k)$ are not identified, unless $\theta_0(J, k)$ for some action J and all k is known or strongly normalized.

Proof. See the Appendix A. ■

The term $e'_\omega D_a R_a$ is the expected discounted present value of R_a given today’s value of ω when the agent always chooses action a . Existence of the inverse of (12) requires ω to significantly change the conditional expected values of R_a and R_J .

To see why Proposition 9 holds, note first that (8) can be rewritten as

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

One can therefore identify the difference between the expected discounted present values of two sequences of choices not necessarily optimal: always choose a versus always choose J (Magnac and Thesmar (2002)). In the empirical model of Section 2.2, additive separability implies

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

Since k and ω evolve independently, ω does not help predicting future $\theta_0(a, k)$. We can therefore exploit variation in ω (ensured by the existence of the inverse of (12)) to eliminate $\theta_0(a, k)$ in (8) 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 $\theta_0(a, k)$, unless we impose a strong normalization.

Let us briefly revisit the empirical examples of Section 2.2. Rust (1987) adopts an exclusion restriction and a weak normalization: the cost of scrapping and replacing an engine is invariable over states and the operating cost at zero mileage is zero. In the monopolist’s entry/exit model, variable profits are often estimated outside of the dynamic problem using price and quantity data. As shown in Aguirregabira and Suzuki (2014), identifying other parameters, however, requires strongly normalizing either the scrap value, or both the fixed

¹³This proposition also holds in the more general case of $F^\omega(\omega'|\omega, a)$ and multivariate $R_a(\omega)$.

and entry costs. Indeed, most often in applied work, fixed costs are normalized to zero, while scrap/entry costs are invariant across states. Whether these restrictions are reasonable will depend on the application; it is difficult to know a priori, as data on entry/scrap costs are extremely rare. Kalouptside (2014a) 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 time. Note that in dynamic games, the presence of potential entrants provides an extra degree of freedom. Typically, their payoffs are normalized to zero when they do not enter the market.

3.2 Renewal Action

In several empirical applications, one encounters renewal or terminal actions. Terminal actions terminate the decision making and impose a finite horizon often facilitating estimation considerably (e.g. a worker retires, a student drops out of school, a firm does not renew a patent). Renewal actions also facilitate estimation. Intuitively, action J is a renewal action when it resets the state variable, so that 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))). Formally:

Definition 10 J is a renewal action if for all $a, j \in A$:

$$F_a F_J = F_j F_J \tag{13}$$

The following lemma shows that renewability simplifies the expressions for identified payoffs.

Lemma 11 If J is a renewal action, then for all $a \in A$

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

Proof. See the Appendix A. ■

Even though renewability is a further restriction, it does not aid identification. Indeed, even if we have a renewal action, we still need to impose a strong normalization.

Conclusion 12 *Renewability does not aid identification of payoffs in standard models. But the retrieved payoffs π_a are easier to compute.*¹⁴

4 Identification of Counterfactuals

Counterfactuals consist of transformations of model primitives, notably payoffs and transitions. A counterfactual that changes payoffs π to $\tilde{\pi}$, is described by a known function, $h : R^{|A||X|} \rightarrow R^{|A||X|}$ (so that $\tilde{\pi} = h(\pi)$, or $\tilde{\pi}_a = h_a(\pi)$, $a \in A$). A counterfactual can also change transitions F to \tilde{F} via a function $h^F : R^{|A| \times |X| \times |X|} \rightarrow R^{|A| \times |X| \times |X|}$. This general setup allows for counterfactuals in which $\tilde{\pi}_a(x)$ is affected by payoffs at all other actions and states. A typical simplification imposes some form of diagonal structure on h . Indeed, empirical work often employs what we call “action diagonal counterfactuals”, where $\tilde{\pi}_a$ depends

¹⁴The same conclusion holds for terminal actions.

solely on π_a (i.e. $\tilde{\pi}_a = h_a(\pi_a)$), as well as “action-state diagonal counterfactuals”, where $\tilde{\pi}_a(x) = h_a(\pi_a(x))$. In addition, several counterfactuals of interest result in affine functions.

The feature of interest here is the counterfactual CCP vector \tilde{p} , i.e. $\zeta(b; h, h^F) = \tilde{p}$. We provide a theorem with necessary and sufficient conditions to identify counterfactual CCPs for general functions h and h^F . We then proceed to the case of affine transformations of payoffs and the special cases of “action diagonal” and “action-state diagonal” counterfactuals. In Subsection 4.3 we discuss counterfactual welfare and in Subsection 4.4 we illustrate the results with a simple numerical example.

Our starting point is relationship (8). It expresses all payoffs π_a , $a \neq J$ as a function of π_J , the known primitives (F, β, σ, G) and the conditional choice probability vector $p \in \mathbb{R}^{|A||X|}$. This system of equations is useful for counterfactual analysis because it does not involve non-primitive objects such as continuation values. Rewriting (8) at the counterfactual scenario, we get

$$\tilde{\pi}_a = h_a(\pi) = \tilde{A}_a h_J(\pi) + \tilde{b}_a(\tilde{p})$$

where $\tilde{A}_a = (I - \beta \tilde{F}_a)(I - \beta \tilde{F}_J)^{-1}$ and $\tilde{b}_a(\tilde{p}) = \sigma(\tilde{A}_a \psi_J(\tilde{p}) - \psi_a(\tilde{p}))$.¹⁵ We stack all payoff vectors, π_a , $a \neq J$, to obtain:

$$\tilde{\pi}_{-J} = \tilde{A}_{-J} \tilde{\pi}_J + \tilde{b}_{-J}(\tilde{p}). \quad (14)$$

The counterfactual CCP \tilde{p} is identified if and only if it does not depend on the free parameter π_J . To determine whether or not this is the case, we apply the implicit function theorem to (14). We make 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; and ∇ represents the differential operator.¹⁶

Theorem 13 *Assume the conditions of Proposition 7 hold. Suppose h is differentiable. Provided the matrix $\begin{bmatrix} I & -\tilde{A}_{-J} \end{bmatrix} \nabla \psi(p)$ is invertible, the counterfactual conditional choice probability, \tilde{p} , is identified if and only if for all π ,*

$$G(A, \tilde{A}) vecbr(\nabla h(\pi)) = 0 \quad (15)$$

where

$$G(A, \tilde{A}) = \begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes I, \begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes \tilde{A}_{-J}.$$

¹⁵Note that a unique \tilde{p} , is guaranteed from (14): since the Bellman is a contraction mapping, \tilde{V} is unique; from (4) so are \tilde{v}_a and thus so is \tilde{p} .

¹⁶Suppose the matrix C is partitioned in block form as $\begin{bmatrix} C_{11} & \dots & C_{1b} \\ \vdots & \ddots & \vdots \\ C_{c1} & \dots & C_{cb} \end{bmatrix}$, where C_{ij} are matrices of the same size. Then $vecbr(C) = [C_{11}, \dots, C_{1b}, C_{21}, \dots, C_{cb}]'$. The block Kronecker product, \boxtimes , of two partitioned matrices B and C is defined by (Koning, Neudecker and Wansbeek (1991)):

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

Proof. The implicit function theorem allows us to locally solve (14) with respect to \tilde{p} provided the matrix

$$\frac{\partial}{\partial \tilde{p}} \left[h_{-J}(\pi) - \tilde{A}_{-J} \tilde{\pi}_J - \tilde{b}_{-J}(\tilde{p}) \right] = -\frac{\partial}{\partial \tilde{p}} \tilde{b}_{-J}(\tilde{p}) = \sigma \begin{bmatrix} I & -\tilde{A}_{-J} \end{bmatrix} \nabla \psi(\tilde{p})$$

is invertible. Under this condition, \tilde{p} does not depend on the free parameter π_J if and only if

$$\frac{\partial}{\partial \pi_J} \left[h_a(\pi_1, \pi_2, \dots, \pi_{J-1}, \pi_J) - \tilde{A}_a h_J(\pi_1, \pi_2, \dots, \pi_{J-1}, \pi_J) - \tilde{b}_a(\tilde{p}) \right] = 0$$

for all $a \neq J$ and all π . But, the above yields

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

or, using (8),

$$\sum_{l \neq J} \frac{\partial h_a}{\partial \pi_l} A_l + \frac{\partial h_a}{\partial \pi_J} = \tilde{A}_a \left(\sum_{l \neq J} \frac{\partial h_J}{\partial \pi_l} A_l + \frac{\partial h_J}{\partial \pi_J} \right)$$

or,

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

Stacking the above expressions for all $a \neq J$ we obtain

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

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

$$\begin{aligned} \left(\begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes I \right) vecbr(\nabla h_{-J}(\pi)) - \left(\begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes \tilde{A}_{-J} \right) vecbr(\nabla h_J(\pi)) &= 0 \\ \left[\begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes I, \begin{bmatrix} A'_{-J} & I \end{bmatrix} \boxtimes \tilde{A}_{-J} \right] \begin{bmatrix} vecbr(\nabla h_{-J}(\pi)) \\ vecbr(\nabla h_J(\pi)) \end{bmatrix} &= 0 \end{aligned}$$

which is (15). ■

Theorem 13 shows that counterfactual choice probabilities \tilde{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 \tilde{A}_{-J} , which in turn are determined by the transition matrices F and \tilde{F} . An implication of Theorem 13 is that, if the Jacobian of $G(A, \tilde{A}) vecbr(\nabla h(\pi))$ has an inverse, the set of payoffs that satisfy (15) has Lebesgue measure zero and therefore \tilde{p} is not identified for almost all π . Thus, under the implicit function theorem requirements, non-identification of counterfactual CCPs is a generic property for nonlinear h . An example of a nonlinear transformation is a change in the agents' level of risk aversion.¹⁷

Another immediate implication of the theorem is that adding a known vector to $\tilde{\pi}$ does not affect the Jacobian matrix of h . Therefore whether the counterfactual CCP is identified does not depend on this vector. For example, lump-sum transfers (e.g. in the form of taxes or subsidies) result in identified counterfactual CCPs.

If the counterfactual of interest is “action diagonal”, (15) becomes simpler:

¹⁷If $\left[\begin{bmatrix} I & -\tilde{A}_{-J} \end{bmatrix} \nabla \psi(p) \right]$ is not invertible, \tilde{p} is not identified even when the full model is.

Corollary 14 (“Action Diagonal” Counterfactuals) In “action diagonal” counterfactuals, provided $\left[\begin{bmatrix} I & -\tilde{A}_{-J} \end{bmatrix} \nabla\psi(p)\right]$ is invertible, \tilde{p} is identified if and only if for all π and $a \neq J$

$$\frac{\partial h_a}{\partial \pi_a} A_a = \tilde{A}_a \frac{\partial h_J}{\partial \pi_J}. \quad (16)$$

Proof. Since h is action diagonal, $\frac{\partial h_a}{\partial \pi_l} = 0$, $l \neq a$, and (16) stems directly from (15). ■

Note that if transitions F_a are transformed by a counterfactual h^F , payoffs have to change appropriately in order to satisfy (15) or (16). Thus, \tilde{p} is nonparametrically not identified for counterfactuals that only change transitions (e.g. a change in the volatility or long-run mean of some state); this result is also documented by Aguirregabiria and Suzuki (2014), Norets and Tang (2014) and Arcidiacono and Miller (2015).

4.1 Affine Counterfactuals

We now consider affine payoff transformations:

$$\tilde{\pi} = \mathcal{H}\pi + g \quad (17)$$

where $\mathcal{H} \in \mathbb{R}^{|A||X| \times |A||X|}$ and $g \in \mathbb{R}^{|A||X|}$. Because $\nabla h(\pi) = \mathcal{H}$, it is clear from Theorem 13 that \tilde{p} is not identified unless \mathcal{H} is restricted to lie in the nullspace of $G(A, \tilde{A})$.

A special case of affine counterfactuals that are prevalent in applied work consist of “action diagonal” and “action-state diagonal” payoff transformations and unaffected transitions (i.e. $F_a = \tilde{F}_a$, for all a). “Action diagonal” counterfactuals restrict \mathcal{H} to be block-diagonal so that:

$$\tilde{\pi}_a = H_a \pi_a + g_a$$

all a , with $H_a \in \mathbb{R}^{|X| \times |X|}$ and $g_a \in \mathbb{R}^{|X|}$.

Lemma 15 (“Action Diagonal” Affine Counterfactuals) In “action diagonal” affine counterfactuals, to identify \tilde{p} it is necessary that all H_a are similar matrices.

Proof. From (16), \tilde{p} is identified if and only if

$$H_a = A_a H_J A_a^{-1}, \quad \text{for all } a \neq J. \quad (18)$$

■

As an example, consider a counterfactual that sets $g_a = 0$ for all a and scales proportionally the payoff of one action j , so that $H_j = \lambda I$. To identify \tilde{p} , we must have $H_J = \lambda I$ and thus $H_a = \lambda I$, for all a , because of (18). In particular if $\lambda = 1$ for one action, all matrices H_a must equal the identity matrix and thus the counterfactual in fact must not change payoffs.

Finally, we turn to “action-state diagonal” affine counterfactuals and fully characterize them. Such counterfactuals involve diagonal matrices H_a , all a , and essentially implement percentage changes in the original payoffs, where the percentage terms may differ for different states x .

Lemma 16 (“Action-State Diagonal” Affine Counterfactuals I) In “action-state diagonal” affine counterfactuals, to identify \tilde{p} it is necessary that $H_a = H$, for all a and that it satisfies $H = A_a H A_a^{-1}$, all a .

Proof. Immediate implication of (18). ■

Lemma 16 places a strong restriction on H_a : the payoffs of all actions need to be affected by the counterfactual in the same fashion. For example, if we change the payoff of action a in state x by $\lambda(x)$, $\tilde{\pi}(a, x) = \lambda(x) \pi(a, x)$, then we also need to change the payoff of any other action a in state x by the same proportion $\lambda(x)$ and check if the transition matrices F_a are such that $H = A_a H A_a^{-1}$ holds. Otherwise, the counterfactual CCP is not identified.

The next proposition presents restrictions that the transition matrices F_a must satisfy to obtain identification in “action-state diagonal” affine counterfactuals.

Proposition 17 (“Action-State Diagonal” Affine Counterfactuals II) Suppose $H_a = H$ is a diagonal matrix with pairwise distinct diagonal entries $\lambda_1, \lambda_2, \dots, \lambda_k$ and corresponding multiplicities n_1, n_2, \dots, n_k . Let $k > 1$. We partition F_a , $a \in A$, in block form $(F_a)_{ij}$ where each (ij) block has size $n_i \times n_j$. Suppose (18) holds so that \tilde{p} is identified. Then, the transition matrices are restricted as

$$F_{ij}^a = (I - \beta F_{ii}^a) (I - \beta F_{ii}^j)^{-1} F_{ij}^j \quad (19)$$

for $i \neq j$ and

$$(I - \beta F_{ii}^a) (I - \beta F_{ii}^j)^{-1} \mathbf{1} = \mathbf{1} \quad (20)$$

where $\mathbf{1} \in \mathbb{R}^{n_i}$ consists of ones. Furthermore, the right hand side of (19) must be between zero and one.

Corollary 18 If $H_a = H = \lambda I$, \tilde{p} is identified.

Corollary 19 If H is diagonal and has a simple eigenvalue, a necessary condition to identify \tilde{p} is that the rows of F_a corresponding to the simple eigenvalue have to be equal for all a .

Corollary 20 If H is diagonal with pairwise distinct eigenvalues, \tilde{p} is not identified, unless the transition probabilities are action invariant.

Proof. See Appendix A. ■

Proposition 17 and its corollaries show that in general, counterfactual CCPs under “action-state diagonal” affine transformations are identified *only if* proper restrictions on the transition matrices F_a hold. The degree of freedom of these restrictions is controlled by the number and level of multiplicities of the diagonal matrix H . For example, if all eigenvalues of H are pairwise distinct, \tilde{p} is not identified unless the state transitions are action invariant. Action-invariant transition matrices is a clearly uninteresting case for dynamic models.

Let us summarize the key takeaways. Counterfactual CCPs are not identified unless some stringent conditions are met. As shown also in Aguirregabiria and Suzuki (2014) and Norets and Tang (2014), one such condition is that $\tilde{\pi}_a = \lambda \pi_a + g_a$ for all a . There are some counterfactuals of interest within this category: lump-sum transfers and proportional changes in payoffs that do not depend on actions and states. In contrast, counterfactual CCPs under

transformations that change payoffs proportionally with different proportions for different actions are not identified.¹⁸ In addition, unless very stringent conditions on the transition matrices are satisfied, transformations in which payoffs are distorted differentially for some states result in nonidentified \tilde{p} even when $H_a = H$ for all a .

4.2 Counterfactuals for Parametric Payoffs

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 empirical model of Section 2.2. As shown in Section 3.1, $\theta_0(a, k)$ is not identified, while $\theta_1(a, k)$ is generally identified. Take the vector $\theta_0(J, k)$, all k , as the free parameter and represent the payoffs as

$$\pi(a, k, \omega; \theta_J) = \theta_0(a, k; \theta_0(J)) + R_a(\omega)' \theta_1(a, k).$$

where $\theta_0(J)$ stacks $\theta_0(J, k)$ for all k .

Proposition 21 *Assume the conditions of Proposition 9 hold. The counterfactual CCP corresponding to a counterfactual that only changes the term $R_a(\omega)' \theta_1(a, k)$ of $\pi(a, k, \omega)$ is identified.*

Proposition 21 is a direct consequence of Proposition 9; indeed following the proof of the latter, one immediately sees that $\theta_1(a, k)$ are determined independently of $\theta_0(a, k)$ and hence are unaffected by the strong normalization on $\theta_0(J)$. For instance, counterfactual CCPs from transformations that affect variable profits in our monopolist's entry problem of Example 2.2 are identified regardless of the strong normalization imposed on fixed and entry costs. In contrast, counterfactuals that proportionally reduce the cost of entry (e.g. entry subsidies) are not identified and thus the strong normalization may lead to severe bias.

More generally, suppose payoffs can be decomposed into two parts: one that is not identified (except under some strong normalization) and one that is. Counterfactuals that transform only the second part of the payoff function are identified regardless of the strong normalization imposed on the first part.

4.3 Counterfactual Welfare

Finally, we discuss counterfactual welfare. We show that even when counterfactual CCPs are identified, counterfactual welfare evaluations may not be. Consider the commonly used "action diagonal counterfactual" $\tilde{\pi}_a = h_a(\pi_a)$, and define the value function difference $\Delta V = \tilde{V} - V$, where \tilde{V} is the counterfactual value function corresponding to $(\tilde{\pi}, \tilde{F})$. The feature of interest here is $\zeta(b; h, h^F) = \Delta V$.

Proposition 22 *Assume the conditions of Theorem 13 hold. The welfare difference ΔV is not identified unless for all a :*

$$\frac{\partial h_a(\pi_a)}{\partial \pi_a} = (I - \beta \tilde{F}_a) (I - \beta F_a)^{-1}. \quad (21)$$

¹⁸Suppose, for example, that one strongly normalizes $\pi_J = 0$, when the true π_J is not zero; then a counterfactual of the type $\tilde{\pi}_a = \lambda_a \pi_a$ and $\tilde{\pi}_J = \lambda_J \pi_J$, with say $\lambda_a > 0 > \lambda_J$, should increase the probability of choosing a in any state in the true counterfactual (with the true π_J) relatively more than in the strongly normalized counterfactual (with $\pi_J = 0$).

Proof. Recall from Proposition 7 that for any a , $V = (I - \beta F_J)^{-1} (\pi_J + \sigma \psi_a(p))$ and similarly for \tilde{V} . Since \tilde{p} is identified,

$$\frac{\partial \Delta V}{\partial \pi_J} = \left(I - \beta \tilde{F}_J \right)^{-1} \frac{\partial h_J(\pi_J)}{\partial \pi_J} - (I - \beta F_J)^{-1}.$$

Thus π_J does not affect ΔV if and only if

$$\frac{\partial h_J(\pi_J)}{\partial \pi_J} = \left(I - \beta \tilde{F}_J \right) (I - \beta F_J)^{-1}.$$

Since (16) must be satisfied, we obtain (21).¹⁹ ■

For example, consider the case of an affine counterfactual that sets $g_a = 0$ for all a and does not affect transitions. The above proposition implies that ΔV is not identified unless the counterfactual does not change payoffs either. Indeed, note that in this case (21) becomes $\frac{\partial h_a(\pi_a)}{\partial \pi_a} = I$; in the linear case, this occurs when $H_a = I$ for all a , since by Proposition 17, H_a are similar. In contrast, if the counterfactual involves only lump-sum transfers, both \tilde{p} and ΔV are identified.

In conclusion, imposing strong normalizations may lead to over/underestimation of welfare changes. Suppose for instance that the true $\pi_J > 0$, that (21) does not hold and rather $\frac{\partial \Delta V}{\partial \pi_J} > 0$. Then the true change in welfare ΔV (with the true π_J) is larger for every state x than the strongly normalized one (with $\pi_J = 0$).

4.4 Numerical Example: A Simplified Bus Engine Replacement Problem

We illustrate the main results of this section with a simple numerical exercise. We perform four counterfactuals using both the true and strongly normalized model. We employ a simplified version of example 2.2 (Rust (1987)), so that $A = \{\text{replace, keep}\}$ and x is the bus mileage. We assume the following payoff function

$$\pi(a, x, \theta) = \begin{cases} -\theta_0 + \phi(x, \theta_1), & \text{if } a = 1 \text{ (replace)} \\ -c(x, \theta_2), & \text{if } a = 2 \text{ (keep)} \end{cases}$$

where θ_0 is the fixed cost of installing a new engine; $\phi(x, \theta_1)$ is the scrap value of the old engine which depends on x and $c(x, \theta_2)$ is the operating cost at mileage x . We consider a deterministic mileage accumulation rule: if the engine is replaced ($a = 1$), $x_{t+1} = 0$, while if the engine is kept ($a = 2$), $x_{t+1} = \min\{x_t + 1, \bar{x}\}$, with $\bar{x} = 2$.

First, we solve the true model and obtain the baseline CCPs and value functions. Then we assume the econometrician knows (or estimates) the true CCP, imposes the strong normalization $\pi_1 = 0$ and identifies the payoff function π_2 from (8).²⁰

The first counterfactual is a lump-sum tax the agent pays if he opts for keeping the old engine (i.e. we subtract $g_2 = (1, 1, 1)'$ from π_2). This counterfactual does not affect the

¹⁹When \tilde{p} is identified, the equality $\frac{\partial h_a}{\partial \pi_a} A_a = \tilde{A}_a \frac{\partial h_J}{\partial \pi_J}$ is satisfied, or, $\frac{\partial h_a}{\partial \pi_a} = \tilde{A}_a \frac{\partial h_J}{\partial \pi_J} A_a^{-1}$. So, $\frac{\partial h_J(\pi_J)}{\partial \pi_J} = \left(I - \beta \tilde{F}_J \right) (I - \beta F_J)^{-1}$ implies $\frac{\partial h_a}{\partial \pi_a} = \left(I - \beta \tilde{F}_a \right) (I - \beta F_a)^{-1}$; the converse also is true.

²⁰We assume that $\phi(x, \theta_1) = \theta_{10} + \theta_{11}x$ and $c(x, \theta_2) = \theta_{20} + \theta_{21}x + \theta_{22}x^2$. We take $\theta_0 = 8$, $\phi(x, \theta_1) = 6 + x$, and $c(x, \theta_2) = 1 + x + 0.1x^2$, so that $\pi_1 = (-2, -1, 0)'$ and $\pi_2 = (-1, -2.1, -3.4)'$. When we strongly normalize $\pi_1 = 0$, we obtain $\pi_2 = (1.95, 0.8, -1.5)'$. To simplify the exercise, we ignore sampling variation and assume the econometrician estimates the CCPs perfectly.

counterfactual CCPs, nor welfare changes. The second counterfactual is a proportional tax of 20% regardless of actions and states (i.e. we multiply both π_1 and π_2 by $H = 1.2 \times I$, noting that the true π is negative). This does not affect CCPs, but biases the welfare change. The third counterfactual is another proportional tax of 20% for both actions, but is charged only when $x = 1$ (i.e. we multiply both π_1 and π_2 by a diagonal matrix with diagonal $(1, 1.2, 1)$). Finally, the fourth counterfactual is yet another proportional tax of 20% charged only when the agent decides to keep the old engine at $x = 1$ (i.e. we multiply *only* π_2 by the same diagonal matrix). The strong normalization must bias both the third and fourth counterfactual CCP and welfare change because the payoff functions are changed proportionally with proportions that differ across states and actions.

Table 1 presents the results. As expected the true and the strongly normalized counterfactuals are identical in the first case. In the second case, the counterfactual CCP is identical in the true model and in the normalized model, but the welfare change is different and has the opposite sign. Cases three and four exhibit the bias created by the strong normalization: the CCP bias can be substantial and different across counterfactuals, while the bias in welfare changes can produce the wrong sign for all values of state variables.

Table 1: Numerical Example – True vs. Strongly Normalized Counterfactuals

	Baseline	True Counterfactual	Str. Normalized Counterfactual	True Counterfactual	Str. Normalized Counterfactual
		<i>CF1</i> : $\tilde{\pi}_1 = \pi_1, \tilde{\pi}_2 = \pi_2 + g_2$		<i>CF2</i> : $\tilde{\pi}_1 = \lambda\pi_1, \tilde{\pi}_2 = \lambda\pi_2$	
CCP: Pr ($a = 1 x$)					
$x = 0$	25.0%	40.5%	40.5%	21.1%	21.1%
$x = 1$	61.3%	73.6%	73.6%	61.8%	61.8%
$x = 2$	94.0%	96.5%	96.5%	96.2%	96.2%
Welfare: $\tilde{V} - V$					
$x = 0$	-	-9.662	-9.662	-4.607	3.393
$x = 1$	-	-9.362	-9.362	-4.585	3.215
$x = 2$	-	-9.205	-9.205	-4.399	3.200
		<i>CF3</i> : $\tilde{\pi}_1 = H\pi_1, \tilde{\pi}_2 = H\pi_2$		<i>CF4</i> : $\tilde{\pi}_1 = \pi_1, \tilde{\pi}_2 = H_2\pi_2$	
CCP: Pr ($a = 1 x$)					
$x = 0$	25.0%	27.6%	24.4%	26.3%	24.4%
$x = 1$	61.3%	64.0%	58.0%	69.5%	58.0%
$x = 2$	94.0%	93.4%	94.2%	93.7%	94.2%
Welfare: $\tilde{V} - V$					
$x = 0$	-	-2.011	0.461	-1.049	0.461
$x = 1$	-	-2.155	0.493	-1.123	0.493
$x = 2$	-	-1.904	0.436	-0.993	0.436

5 Identification using Resale Market Data

When a counterfactual of interest is not identified, the researcher is often faced with a significant challenge. In this section we show that data on resale markets can substantially aid identification and estimation of dynamic models (Kalouptsi, 2014a). Suppose we can relate agent i with an asset so that the value of the asset i 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. The asset prices contain information about value functions $V(x_{it})$. In particular, 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:

$$p_{it}^{RS} = V(x_{it}). \quad (22)$$

In other words, in a world with a large number of homogeneous buyers, resale prices can be treated as observations of the value function. As such, they provide direct, raw information on the main dynamic object of interest. Indeed, inspecting (22) in combination with the agent's Bellman equation (1), one can easily see that value functions immediately inform us on payoffs and their shape; in fact, they deliver payoffs nonparametrically (see Kalouptsi (2014a) and (2014b) for an implementation). In addition, as shown below, we can avoid strong normalizations. It is of course 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.²¹ This is the simplest model allowing the use of resale prices; 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 are also possible.

The dataset now includes resale prices, p_{it}^{RS} , so that $y_{it} = (a_{it}, x_{it}, p_{it}^{RS})$. Using (22) we can estimate $V(x_{it})$ by (nonparametrically) regressing p_{it}^{RS} on x_{it} :

$$p_{it}^{RS} = V(x_{it}) + \eta_{it} \quad (23)$$

where η_{it} is measurement error. We therefore assume for identification purposes that V is known (in addition to p_a and F). When V is known, the identification of flow payoffs follows almost immediately; the only difficulty is that the variance of the idiosyncratic shocks σ needs to be determined, and then equations (4), (5) and (6) have a unique solution.

Proposition 23 *Assume Conditions 1-4 hold. Given the joint distribution of observables $\Pr(y)$, where $y_{it} = (a_{it}, x_{it}, p_{it}^{RS})$, the flow payoffs π_a are identified provided the primitives (β, G) are known and (i) σ is known, or (ii) the (cardinal) payoff $\pi_a(x)$ is known for some (a, x) ; or (iii) the expected payoff for one action is known.*

²¹One way to relax homogeneity is to assume a competitive resale market with a large number of potential buyers drawn from a finite type space. In this case, the only buyers active in the market for the asset are the ones with the highest valuation, and the resale price will be equal to their valuation.

Proof. Case (i): Substitute V in (6) into (4):

$$\pi_a = (I - \beta F_a) V - \sigma \psi_a. \quad (24)$$

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

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

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. ■

The use of resale prices reduces the indeterminacy of the system and does not require a strong normalization.²² Because V is measured with respect to a specific scale (e.g. dollars), a normalization on σ is no longer innocent. In other words, we are working with cardinal measurements of the value function; thus we need to measure the payoff function in the same units. Fortunately, with little information on payoffs for any one action in any state, we can identify σ . Alternatively, measures of average payoffs can be used, as the latter are readily available in a variety of settings (e.g. below we use public data on costs and returns for agriculture).

6 Identification with Unobservable Market-level States

With our agricultural land use application in mind, we extend the identification results to the case where the state variables contain unobservable components. Here, we explicitly assume that there are $m = 1, \dots, M$ markets. In addition, we borrow the state decomposition $x_{imt} = (k_{imt}, \omega_{mt})$, as well as a state transition similar but slightly more general to (2) from Section 2.2: the transition function of k is $F^k(k_{imt+1}|a, k_{imt}, \omega_{mt})$. Importantly, the aggregate state ω_{mt} is not fully observed; but it does have an observed component w_{mt} .²³ Note the transition of k_{imt} can be recovered from the data even though ω_{mt} is not fully observed. Indeed, one can estimate $F_{mt}^k(k_{imt+1}|a, k_{imt}) = F^k(k_{imt+1}|a, k_{imt}, \omega_{mt})$ for each time period in each market with a rich cross section of agents. In contrast, $F^\omega(\omega_{mt+1}|\omega_{mt})$ cannot be estimated.

We add the following two assumptions:

Condition 24 (*Additive Separability II*) *Per period payoffs are expressed as follows:*

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

²²In addition, if the payoff function is also known for another combination (j, \tilde{x}) , it is possible to identify the discount factor β .

²³We 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.

Condition 25 (*Instrumental Variables*) *There exist instruments at the time- t information set, z_{mt} , such that*

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

for all a, k ; and that for all functions $q(w_{mt})$, $E [q(w_{mt}) | z_{mt}] = 0$ implies $q(w_{mt}) = 0$.

Condition 24 restricts the unobserved aggregate state to enter payoffs in an additively separable fashion. The unobservable $\xi(a, k, \omega)$ may reflect mismeasured profits or unobservable costs. It is important to stress that ξ is not a state variable, but a function of the state variables. 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, $\bar{\pi}$ 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. Condition 25 assumes access to valid instruments and imposes the completeness condition. If it is reasonable to assume that $\bar{\pi}$ 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 $y = \{(a_{imt}, k_{imt}, w_{mt}, z_{mt}) : i = 1, \dots, N; m = 1, \dots, M; t = 1, \dots, T_i\}$. Identification with partially observed states cannot make direct use of the main equations (4)-(6), since the transition matrix is unknown. Our identification results are instead based on the following expression which replaces (4)-(6):

$$\begin{aligned} & \pi(a, k_{imt}, \omega_{mt}) + \beta \varepsilon^V(a, k_{imt}, \omega_{mt}, \omega_{mt+1}) \\ &= V(k_{imt}, \omega_{mt}) - \beta \sum_{k'} V(k', \omega_{mt+1}) F^k(k'|a, k_{imt}, \omega_{mt}) - \sigma \psi_a(k_{imt}, \omega_{mt}) \end{aligned} \quad (25)$$

where $\varepsilon^V(\cdot)$ is an “expectational error” according to the following definition:

Definition 26 (*Expectational error*) *For any function $\zeta(k, \omega)$ and particular realization $\omega^* \in \Omega$,*

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

To derive (25) note that $E[V(k_{imt+1}, \omega_{mt+1}) | a, k_{imt}, \omega_{mt}]$ is given by:

$$\begin{aligned} & \sum_{k'} \int_{\omega'} V(k', \omega') dF^\omega(\omega' | \omega_{mt}) F^k(k'|a, k_{imt}, \omega_{mt}) \\ &= \sum_{k'} \left(E_{\omega'|\omega_{mt}} [V(k', \omega') | \omega_{mt}] \right) F^k(k'|a, k_{imt}, \omega_{mt}) \\ &= \sum_{k'} V(k', \omega_{mt+1}) F^k(k'|a, k_{imt}, \omega_{mt}) + \varepsilon^V(a, k_{imt}, \omega_{mt}, \omega_{mt+1}) \end{aligned} \quad (26)$$

The first term of the right hand side of (26) 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 (4) as follows: for $a = 1, \dots, J$

$$\begin{aligned} \pi(a, k_{imt}, \omega_{mt}) &= v_a(k_{imt}, \omega_{mt}) - \beta \sum_{k'} V(k', \omega_{mt+1}) F^k(k'|a, k_{imt}, \omega_{mt}) \\ &\quad - \beta \varepsilon^V(a, k_{imt}, \omega_{mt}, \omega_{mt+1}). \end{aligned} \quad (27)$$

which in turn leads to (25). Importantly, the expectational error is mean independent of all past state variables (k, ω) (see Lemma 30 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 $p_a(k, \omega_{mt})$. 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 $p_a(\cdot)$, $F^k(\cdot)$ and $\psi_a(\cdot)$ as known objects.²⁴

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, (25) in matrix form becomes:

$$\pi_{amt} + \beta \varepsilon_{am,t,t+1}^V = V_{mt} - \beta F_{amt}^k V_{mt+1} - \sigma \psi_{amt} \quad (28)$$

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

We now turn to our identification results. In contrast to the case of fully observed states, the value function enters (28) in a recursive fashion at ω_{mt} and ω_{mt+1} . Repeated substitution of $V_{m\tau}$, all τ , in (28) leads to an expression that includes V_{mT} as well as all expectational errors. Although we can average out the latter, we are still left with the unobserved V_{mT} – this is a barrier to identification. As we show formally in Proposition 27, when states are partially observed, fixing payoffs of an action for all states is not sufficient for identification (as in the case of fully observed states). To obtain identification we must impose extra restrictions. Renewability is a natural candidate; we take action J to be a renewal action. Adapting Definition 10 in Section 2 for partially observed states, the condition states that for all t, τ and all a, j :

$$F_{amt}^k F_{Jm\tau}^k = F_{jmt}^k F_{Jm\tau}^k. \quad (29)$$

For any a and j , we then have

$$\begin{aligned} & \pi_{amt} - \pi_{jmt} + \beta \left(\varepsilon_{am,t,t+1}^V - \varepsilon_{jm,t,t+1}^V \right) + \sigma (\psi_{amt} - \psi_{jmt}) \\ &= \beta \left(F_{amt}^k - F_{jmt}^k \right) \left(\pi_{Jmt+1} + \varepsilon_{Jm,t+1,t+2}^V + \sigma \psi_{Jmt+1} \right) \end{aligned} \quad (30)$$

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 (30) forms our base equation for identification.²⁵

²⁴Formally, ω_{mt} is a market-level shock affecting all agents i in m at t . If the data $\{a_{imt}, k_{imt} : i = 1, \dots, N\}$ 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),

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

as $N \rightarrow \infty$. The result extends to non-i.i.d. data, provided the cross-section dependence dies out with the distance across the agents. It is clear that depending on the realization of ω_{mt} , $\hat{p}_{amt}(k)$ converges in probability to the realization of $p_a(k, \omega_{mt})$. The same argument applies to the estimator of $F^k(\cdot)$.

²⁵Formally, use (28) for J in $t+1$ to solve for V_{mt+1} and replace the latter in (28) for any a in t :

$$\begin{aligned} \pi_{amt} + \beta \varepsilon_{am,t,t+1}^V &= V_{mt} - \sigma \psi_{amt} \\ &\quad - \beta F_{amt}^k \left[\pi_{Jmt+1} + \beta \varepsilon_{Jm,t+1,t+2}^V + \beta F_{Jmt+1}^k V_{mt+2} + \sigma \psi_{Jmt+1} \right] \end{aligned}$$

Proposition 27 *Assume the Conditions 1-4 and 24-25 hold. 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 payoffs $\bar{\pi}(a, k_{imt}, w_{mt})$ are identified provided the market level stochastic process (z_{mt}, ω_{mt}) is stationary and:*

(a) *Finite T : (i) the terminal value V_{mT} is known and $\bar{\pi}(j, k_{imt}, w_{mt})$ is known or prespecified for some j and all (k, w) ; or (ii) there is a renewal action J with known or prespecified flow payoff for all (k, w) ; and the spatial correlation between (z_{mt}, ω_{mt}) and $(z_{m't}, \omega_{m't})$ dies out as $M \rightarrow \infty$.*

(b) *Large T : there is a renewal action J , the flow payoff of some action j (not necessarily action J) is known or prespecified for all (k, w) and the serial correlation between (z_{mt}, ω_{mt}) and $(z_{m't'}, \omega_{m't'})$ dies out as $T \rightarrow \infty$.*

Proof. See Appendix A. ■

Proposition 27 shows that, when market-level states are partially observed, identification requires a strong normalization on payoffs, like Proposition 7, as well as an extra restriction: the presence of a renewal action. In related work, Arcidiacono and Miller (2015) consider non-stationary settings with fully observed states, where strong normalizations fixing the payoff of the renewal action in the last period observed is needed.

6.1 Identification with Resale Prices and Unobservable Market-level States

Simple inspection of (28) shows that payoffs can be identified with resale prices: the right hand side is essentially observed. Because the expectational errors and the unobservable ξ have zero mean given the instrumental variables, we can treat the model as a (non-parametric) regression model. There is no need to impose a strong normalization nor renewability.

Proposition 28 *Assume the Conditions 1-4 and 24-25 hold. Suppose the primitives (β, G) are known and either (i) σ is known, or (ii) the (cardinal) payoff $\bar{\pi}(a, k_{imt}, w_{mt})$ is known for one combination of (a, k, w) . Given the joint distribution of observables $\Pr(y)$, where $y_{imt} = (a_{imt}, k_{imt}, w_{mt}, z_{mt}, p_{imt}^{RS})$, $\bar{\pi}(a, k_{imt}, w_{mt})$ is identified provided (z_{mt}, ω_{mt}) is stationary and the spatial and/or serial correlation between (z_{mt}, ω_{mt}) and $(z_{m't'}, \omega_{m't'})$ dies out as $M \rightarrow \infty$ and/or $T \rightarrow \infty$.*

Proof. See Appendix A. ■

7 Agricultural Land Use Model

Our model of agricultural land use closely follows Scott (2013) and is a special case of the empirical model of Section 6. Each year, field owners decide whether to plant crops or not; i.e. $A = \{c, nc\}$, where c stands for “crops” and nc stands for “no crops” (e.g. pasture, hay, grassland, forests, and other forms of non-managed land). Fields are indexed by i and counties are indexed by m . We partition the state x_{imt} into:

1. time-invariant field and county characteristics, ζ_{im} , e.g. slope, soil composition;

Next, evaluate the above at a and j and subtract to obtain (30) using the renewability property (29).

2. number of years since field was last in crops, $k_{imt} \in K = \{0, 1, \dots, \bar{k}\}$; and²⁶
3. aggregate state, ω_{mt} (e.g. input and output prices, government policies) with an observed component w_{mt} .

Per period payoffs are specified as follows:

$$\pi(a, k_{imt}, \omega_{mt}, \zeta_{im}, \varepsilon_{imt}) = \theta_0(a, k_{imt}, \zeta_{im}) + R(a, w_{mt}, \zeta_{im}) + \xi(a, k_{imt}, \omega_{mt}, \zeta_{im}) + \sigma \varepsilon_{imt} \quad (31)$$

where $R(a, w, \zeta)$ and $\xi(a, k, \omega, \zeta)$ are observable and unobservable measures of returns, while $\theta_0(a, k, \zeta)$ captures switching costs between land uses. We construct returns R using county-year information (expected prices and realized yields for major US crops, as well as USDA cost estimates) as in Scott (2013).²⁷ Due to data limitations we only allow R to depend on (a, w_{mt}) , so that we have $R_{mt}^a \equiv R(a, w_{mt})$.

The dependence of θ_0 on k is what creates dynamic incentives for landowners. The action of “no crops” leaves the land idle, slowly reverting it to natural vegetation, rough terrain, etc. The farmer needs to clear the land in order to convert to crop and start planting. The costs of switching to crop may be rising as the terrain gets rougher. At the same time, however, there may be benefits to switching, e.g. planting crops may be more profitable after the land is left fallow for a year. In summary, we expect $\theta_0(a, k, \zeta)$ to differ across k .

To complete the specification of this model, we determine state transitions. We follow the decomposition (2), which implies that farmers are small and that there are no externalities across fields. The transition rule of k is:

$$k'(a, k) = \begin{cases} 0, & \text{if } a = c \\ \min\{k + 1, \bar{k}\}, & \text{if } a = nc \end{cases} \quad (32)$$

for all (k, ω, ζ) . If “no crops” is chosen, the number of years since last crop increases by one, until \bar{k} . If “crops” is chosen, the number of years since last crop is reset to zero. Planting crops is therefore a renewal action.

7.1 Estimators for the Land Use Model

The parameters of interest are σ and $\theta_0(a, k, \zeta)$, all a, k, ζ . We present and compare three estimators. First, we employ Scott (2013)’s method which relies on data for actions and states. It differs from the nested fixed-point (Rust (1987)) and the 2-step estimator of Hotz and Miller (1993) as it allows for partially-observed market states and is in regression form. We call this estimator, the “CCP estimator.” It requires a strong normalization on $\theta_0(a, k, \zeta)$. The second estimator, which we call the “joint estimator,” considers the moments of the CCP estimator, plus the moment restrictions obtained from resale prices; all moments are used jointly to estimate payoffs. Finally, because resale prices are bound to affect the model estimates beyond the strong normalization, we consider a third estimator. The third estimator, named the “hybrid estimator,” is based on the CCP estimator and only uses the resale prices to drop

²⁶Ideally, k_{imt} would include detailed information on past land use. We consider the years since the field was in crop (bounded by \bar{k}) for computational tractability and due to data limitations.

²⁷We refer the interested reader there for details. One important difference from Scott (2013) is that we have field level characteristics ζ and they affect land use switching costs.

the strong normalization on $\theta_0(a, k, \zeta)$. All three estimators require first stage estimates of the conditional choice probabilities, $p_{mt}(a, k, \zeta)$, while the joint and the hybrid also require estimating the value function $V_{mt}(k, \zeta)$ (from resale prices). The details of the first stage are presented in Appendix C.

The “CCP estimator” Scott (2013) derives a regression estimator using the renewability of action “crops”. Indeed, adapting (30) to the land use model, we obtain:

$$\sigma Y_{mt}^b(k, \zeta) = \tilde{\theta}_0(k, \zeta) + (R_{mt}^c - R_{mt}^{nc}) + \tilde{\xi}_{mt}(k, \zeta) + \tilde{\varepsilon}_{mt}^V(k, \zeta) \quad (33)$$

for all k, ζ , where

$$Y_{mt}^b(k, \zeta) \equiv \ln \left(\frac{p_{mt}(c, k, \zeta)}{p_{mt}(nc, k, \zeta)} \right) + \beta \ln \left(\frac{p_{mt+1}(c, k'(c, k), \zeta)}{p_{mt+1}(c, k'(nc, k), \zeta)} \right),$$

$$\tilde{\theta}_0(k, \zeta) = \theta_0(nc, k, \zeta) - \theta_0(c, k, \zeta) + \beta [\theta_0(c, k'(c, k), \zeta) - \theta_0(c, k'(nc, k), \zeta)], \quad (34)$$

$\tilde{\xi}_{mt}(k, \zeta) = \xi_{mt}(nc, k, \zeta) - \xi_{mt}(c, k, \zeta) - \beta E[\xi_{mt+1}(c, k'(c, k), \zeta)] + \beta E[\xi_{mt+1}(c, k'(nc, k), \zeta)]$, with the expectations taken over ω_{mt+1} ; $\tilde{\varepsilon}_{mt}^V(k, \zeta)$ is defined similarly to $\tilde{\xi}_{mt}(k, \zeta)$.²⁸

As shown in Section 3.1, we cannot identify all $\theta_0(a, k, \zeta)$ when only actions and states are observed. Following Scott (2013), we strongly normalize $\theta_0(nc, k, \zeta) = 0$ for all k, ζ . It is important to note that this specification has already (weakly) normalized scale, as R_{mt}^a is measured in dollars; therefore, we can estimate σ .²⁹ Note also that the profit function has the same functional form as (3) in Subsection 2.2.

We estimate (33) in two steps: first, we estimate σ alone via instrumental variables (IV) regression on first differences of (33); then we obtain $\theta_0(a, k, \zeta)$ by averaging the residuals. As discussed in Section 6, one should expect $(R_{mt}^{nc} - R_{mt}^c)$ and $\tilde{\xi}_{mt}(k, \zeta)$ to be correlated. Thus, for the regression estimating σ , we need instruments. Following Scott (2013) again, we employ lagged returns and caloric yields to instrument for the first difference of $(R_{mt}^{nc} - R_{mt}^c)$. Once σ is estimated, we move to $\theta_0(a, k, \zeta)$. We obtain the residuals and take their average over time in order to remove $\xi(k, \zeta)$. Finally, having normalized the intercepts for “no crops,” we invert (34) to obtain $\theta_0(c, k, \zeta)$.

The “Joint Estimator” Assuming V is known, (25) can be added as follows:

$$Y_{mt}^v(a, k, \zeta) = \theta_0(a, k, \zeta) + R_{mt}^a + \sigma \psi_{mt}(a, k, \zeta) + \xi_{mt}(a, k, \zeta) + \beta \varepsilon_{mt+1}^V(a, k), \quad (35)$$

where

$$Y_{mt}^v(a, k, \zeta) \equiv V_{mt}(k, \zeta) - \beta V_{mt+1}(k'(a, k), \zeta),$$

which can be calculated using the first stage estimate of $V_{mt}(k, \zeta)$. For our first stage estimate of the value function, we use a polynomial approximation in (k, ζ) for each (m, t) . We employ

²⁸We refer the interested reader to Scott (2013) for the detailed derivation. Remember that (i) F_{amt}^k evolves deterministically, (ii) $p_{mt}(a, k) = p(a, k, \omega_{mt})$, and that (iii) for the binary choice model with logit shocks, $\psi_a(p(x)) = -\log(p_a(x)) + \gamma$.

²⁹To see this, divide both sides of (31) by σ and observe that now $1/\sigma$ transforms the scale of R_{mt}^a (dollars) into “utils”. Scott (2013) adopts an equivalent approach: the profit specification is $\pi(a, k, \omega) = \theta_0(a, k) + \theta_1 R_{mt}^a + \xi(a, k, \omega)$ and he estimates $\theta_1 = 1/\sigma$.

the Generalized Method of Moments (GMM) to obtain σ from the first differences of (33) and (35). Since ψ_{mt} is likely correlated with ξ_{amt} , we employ lagged values of ψ_{mt} , as well as caloric yields as instrument for moments (35). We then follow the same procedure as above to obtain $\theta_0(a, k, \zeta)$, all a, k, ζ from the residuals.

The “Hybrid Estimator” To isolate the impact of strong normalizations on counterfactuals, we consider a third estimator, which we call the “hybrid.” The hybrid estimator uses (33) alone to estimate σ (following the CCP estimator) and only uses (35) to recover the intercepts $\theta_0(a, k, \zeta)$ (following the joint estimator). The idea here is to force resale prices to provide information only on $\theta_0(a, k, \zeta)$; i.e. resale prices allow us to only avoid the strong normalization.

Note that the CCP and hybrid estimators follow identical strategies to estimate σ . Conditional on an estimate of σ , the hybrid and joint estimators follow identical strategies to recover the intercepts θ_0 .

8 Data

We performed a spatial merge of a number of datasets to create a uniquely rich database on land use. First, we employ high-resolution (30-56m) annual land use data, obtained from the Cropland Data Layer database. The CDL covers the entire contiguous United States since 2008. We merged this dataset with NASA’s Shuttle Radar Topography Mission database which 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. To augment our detailed 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. 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. Finally, we use various public databases on agricultural production from the USDA. All spatial merges were done using nearest neighbor interpolation. 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; indeed, ζ_{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. Latitude and longitude capture the field’s location precisely. We also compute each field’s distance to close urban centers, as well as its nearby commercial property values (specifically, the land value of nearby restaurants). These characteristics shape the field’s options for future development, a potentially important determinant of both land use and land values.

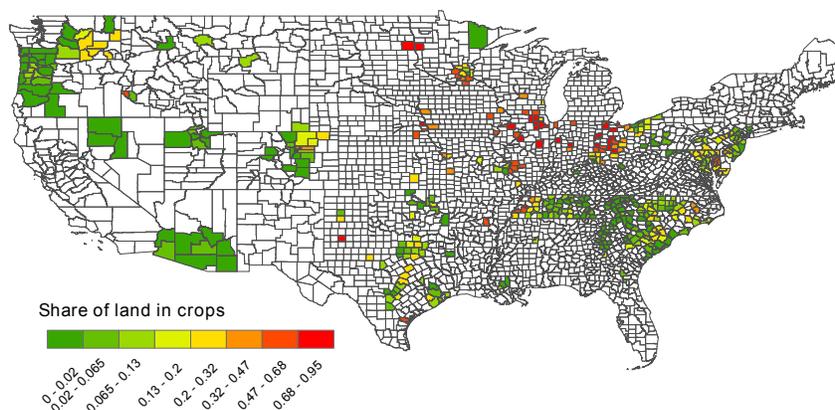
Table 2 presents some summary statistics. The average proportion of cropland in the sample is 15%. Land use exhibits substantial persistence: the probability of keeping the land in crop is about 85%, while the probability of switching to crops (i.e. the probability of planting crops after two years as noncrop) is quite small, at 1.6%. Although not presented in the table, the proportion of fields that switch back to crops after one year as noncrop is significant (ranging from 27% to 43% on average depending on the year), which suggests

Table 2: 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.

Figure 1: Prevalence of Cropland for Counties in Sample, 2013



some farmers enjoy benefits from leaving land fallow for a year. Measurable crop returns are \$214 per acre on average, while measurable non-crop returns are \$13 per acre (not shown). 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.

A total of 91,198 farms were transacted between 2008 to 2013 based on DataQuick. However, after applying the selection criteria (i.e., dropping non-standard transactions and outliers) there remained 24,643 observations. See Appendix C.2 for details.

Figure 1 presents the spatial distribution of land use pattern in the US using the CDL fields for the counties in our sample.

9 Results

We now turn to our empirical results. We provide a brief discussion on the first stage (choice probabilities and value function) and then turn to the main object of interest: fieldowner payoffs. The details of the first stage can be found in Appendix C.

First Stage: Choice Probabilities and Value Function. To obtain an estimate of conditional choice probabilities, we implement a semiparametric strategy, estimating conditional choice probabilities separately for each county, field state, and year. We use slope as a covariate, which proves to be a powerful predictor of land use patterns.

Next, we estimate the value function from resale prices following (23). We regress resale price per acre on field characteristics (k , slope, altitude, distance to urban centers, nearby commercial values), year and county dummies, as well as higher order terms and interactions of these covariates.

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. As discussed in Section 8, the land resale market is arguably thick, with a large number of transactions taking place every year.³⁰ Moreover, we are able to control for a rich set of field characteristics. One may worry that transacted fields are selected. In Table 6 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. In addition, the probability of keeping (switching to) crops is very similar across the two datasets. Finally, we explore whether land use changes upon resale and find no such evidence (see Table 7 in Appendix B).

Payoffs. Table 3 presents the estimated parameters using the three proposed estimators. For brevity we only present the average $\theta_0(a, k, \zeta)$ across field types ζ . Due to data limitations, we set $\bar{k} = 2$.³¹

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 $k = 0$ (i.e., when the land was in crops the previous year). Conversion costs when $k = 2$ are even larger. Of course such interpretation hinges on the assumption that $\theta_0(\text{nocrop}, k) = 0$ 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 hybrid estimator do not normalize $\theta_0(\text{nocrop}, k) = 0$. When $k = 0$, switching out of crops is now expensive (not zero anymore). Furthermore, the absolute value of the estimated $\theta_0(\text{crop}, 0)$ is now larger than the absolute value of $\theta_0(\text{crop}, 1)$,

³⁰The USDA releases information about agricultural land sales in Wisconsin, but seemingly not for other states. There are approximately 100 thousand acres transacted per year (about 1000 transactions), and about 14.5 million acres of farmland in Wisconsin, so there are roughly a little less than 1% of land transacted per year. We obtain slightly larger numbers (between 1.4% and 2% depending on the year) when we compare the number of transacted properties in DataQuick with the number of fields in the CDL data.

³¹We weight observations as in Scott (2013) and cluster standard errors by year.

Table 3: Empirical Results

Estimator:	CCP	Hybrid	Joint
$\bar{\theta}_0(crop, 0)$	-721.93 (800)	-1228.9 (80.9)	-802.93 (741)
$\bar{\theta}_0(crop, 1)$	-2584.4 (3738)	-1119.4 (74.7)	-281.16 (683)
$\bar{\theta}_0(crop, 2)$	-5070.8 (7670)	-4530.4 (136)	-2938.9 (1240)
$\bar{\theta}_0(nocrop, 0)$	0	-2380.3 (1900)	-1896.6 (17421)
$\bar{\theta}_0(nocrop, 1)$	0	470.05 (2510)	830.39 (22939)
$\bar{\theta}_0(nocrop, 2)$	0	-454.58 (1210)	-228.46 (11103)
σ	734.08 (191.98)	734.08 (191.98)	472 (243.1)

θ_0 values are means across all fields in the sample.

Standard errors in parentheses.

which reflects the benefits of leaving land fallow for one year. This potential benefit is not apparent when we strongly normalize $\theta_0(nocrop, k)$. Given that the probability of planting crops after one year of fallow is lower than the probability of planting crops after crops (in most places), in order to rationalize the choice probabilities, the strongly normalized model must assign higher costs to crops after fallow than after crops. We view this as an appealing feature of the hybrid and joint models – 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.³²

Although the numbers differ, the results for the joint estimator are qualitatively similar to the hybrid estimator.³³

9.1 Policy Counterfactuals

We consider two policy relevant counterfactuals: the long-run elasticity (LRE) of land use and a fertilizer tax.

The LRE measures the long-run sensitivity of land use to an (exogenous) change in crop returns, R^c . To calculate it, we first compute the long run steady-state acreage distribution holding returns (R) fixed at their average recent levels. Then, the steady-state acreage

³²One 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, 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., field which have been out of crops for at least two years have typically been out of crops for a long time.

³³In practice, we found that not employing the non-crop value function moments led to more stable results, likely because R^{nc} is somewhat poorly measured.

distribution is obtained with R^c held fixed at 10% higher levels. The LRE is defined as the arc elasticity between the total acreages in the two steady states.³⁴

As shown in Table 4, the LRE varies little across the different estimators. In fact, the CCP and hybrid estimators give *exactly* the same LRE. This is no coincidence. Note that the rescaled (by σ) profit function in (31) is composed of the intercept and R^a multiplied by the coefficient $1/\sigma$ (plus errors). Since σ is identified, any counterfactual that transforms the identified part of the profit is also identified, as discussed in Subsection 4.2. Therefore, the LRE is not affected by strong normalizations, and the *only* difference between the CCP estimator and the hybrid estimator is that the latter relies on land values to identify the profit function while the former makes a strong normalization. The joint estimator results in a higher LRE than the CCP or hybrid because it results in a different estimate for σ .

The second counterfactual makes the following change to the profit function:

$$\tilde{\theta}(\text{crop}, 0) = \theta(\text{crop}, 0) + 0.1(\theta(\text{crop}, 1) - \theta(\text{crop}, 0)).$$

The difference $\theta(\text{crop}, 1) - \theta(\text{crop}, 0)$ 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 after a year of fallow land. Because it is difficult to measure the fertilizer saved by leaving land fallow, we use the estimated switching cost parameters to implement the counterfactual. The increased costs to replant crops therefore resembles a fertilizer tax. Higher cost of fertilizers is a likely result of pricing greenhouse gas emissions.³⁵

As shown in Table 4, the strong normalization does matter when it comes to this counterfactual. Indeed, the fertilizer tax counterfactual is not identified, since it involves a proportional transformation of the unidentified part of profits. The CCP estimator leads to a 32% increase in cropland as a result of this tax. In contrast, the hybrid and the joint estimators predict a decrease in cropland, as expected. Behind the difference in results is the fact that the CCP estimator does not predict benefits from leaving land fallow (on average), as discussed previously. In contrast, the hybrid and joint estimators capture the desired qualitative effect, since the fertilizer tax represents an increase in the costs of planting crops after crops.

To summarize, when we only relax the strong normalization (i.e., moving from the CCP to the hybrid estimator), the LRE does not change, as it involves only a transformation of the identified component of the profit function. However, the fertilizer tax, which involves a transformation of the non-identified part of payoffs, is altered when we relax the strong normalization. If we add moments that change our estimates of the identified part of payoffs (moving to the joint estimator), then either type of counterfactual will be affected.

10 Conclusions

This paper studies the identification of counterfactuals and payoffs in dynamic discrete choice models. In Marschak’s (1953) spirit, we ask (i) whether counterfactuals are identified even

³⁴See 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 smaller set of counties combined with the absence of unobserved heterogeneity: when Scott’s estimation strategy is applied to our sample of counties ignoring unobserved heterogeneity, one finds LREs very similar to those presented here.

³⁵As with the LRE, we fix R^c and R^{nc} at their mean level for each county.

Table 4: Policy Counterfactuals

Estimator:	CCP	Hybrid	Joint
Long-run elasticity	0.57	0.57	0.86
Fertilizer tax	0.32	-0.16	-0.31

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

Long-run elasticity is a 10 percent arc elasticity.

when the model parameters are not; and if so, (ii) which set of counterfactuals are or are not identified; and in the latter case, (iii) whether and how identification can be restored. Our results provide a positive message with qualifications: the behavioral and welfare impacts of some counterfactuals can be identified even when the full set of structural parameters is not. For cases in which the counterfactual impacts are not identified, extra restrictions are necessary. We provide conditions that must be verified to situate a particular counterfactual in one case or another; and we help clarify the role that some of the extra assumptions have in determining what we can learn from data. Finally, since this paper was the result of our interest in policy relevant counterfactuals in the context of agricultural land use, we explore the impact of strong normalizations in long-run land use elasticities and increases in the cost of replanting crops, finding that the former is identified (robust to strong normalizations) while the latter is not.

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., P. Mira, and H. Roman (2007). “An Estimable Dynamic Model of Entry, Exit and Growth in Oligopoly Retail Markets,” *American Economic Review*, 97 (2), 449-454.
- [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] Bayer, P., A. Murphy, R. McMillan, and C. Timmins (2011). “A Dynamic Model of Demand for Houses and Neighborhoods,” NBER Working Paper 17250.

- [8] 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.
- [9] Collard-Wexler, A. (2013). “Demand fluctuations in the ready-mix concrete industry,” *Econometrica*, 81(3), 1003–1007.
- [10] Connault, B. (2014). “Hidden Rust Models,” Working Paper, Princeton University.
- [11] Conlon, C. T. (2010). “A dynamic model of costs and margins in the lcd tv industry,” Unpublished manuscript, Columbia University.
- [12] Crawford, G. S. and M. Shum (2005). “Uncertainty and Learning in Pharmaceutical Demand,” *Econometrica*, 73, 4, 1137–1173.
- [13] Das, S., M.J. Roberts, and J.R Tybout. (2007). “Market entry costs, producer heterogeneity, and export dynamics,” *Econometrica*, 75(3), 837–873.
- [14] Diermeier, D., M. Keane, and A. Merlo (2005). “A Political Economy Model of Congressional Careers,” *American Economic Review*, 95, 347–373.
- [15] Duflo, E., R. Hanna, and S. P. Ryan (2012). “Incentives Work: Getting Teachers to Come to School,” *The American Economic Review*, 102, 1241–1278.
- [16] 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.
- [17] 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.
- [18] Eckstein, Z. and O. Lifshitz (2011). “Dynamic Female Labor Supply,” *Econometrica*, 79(6), 1675–1726.
- [19] Fang, H. and Y. Wang (2015). “Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions,” *International Economic Review*, 56, 2, 565–596.
- [20] Firth D. (1993). “Bias Reduction of Maximum Likelihood Estimates,” *Biometrika*, 80:27–38.
- [21] Gilleskie, D. (1998). “A Dynamic Stochastic Model of Medical Care Use and Work Absence,” *Econometrica*, 66, 1-45.
- [22] Gowrisankaran, G., and M. Rysman (2012). “Dynamics of Consumer Demand for New Durable Goods,” *Journal of Political Economy*, 120, 1173-1219.
- [23] Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic Press, New York-London.
- [24] Heckman, J., and S. Navarro. (2007), “Dynamic Discrete Choice and Dynamic Treatment Effects,” *Journal of Econometrics*, 136, 341–396.

- [25] Hendel, I., and A. Nevo (2006). “Measuring the Implications of Sales and Consumer Inventory Behavior,” *Econometrica*, 74(6), 1637-1674.
- [26] 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.
- [27] Hu, Y., and M. Shum (2012). “Nonparametric Identification of Dynamic Models with Unobserved State Variables,” *Journal of Econometrics*, 171, 32–44.
- [28] Jeziorski, P, E. Krasnokutskayay, and O. Ceccariniz (2014). “Adverse Selection and Moral Hazard in the Dynamic Model of Auto Insurance,” Mimeo.
- [29] Kalouptside, M. (2014a). “Time to Build and Fluctuations in Bulk Shipping,” *American Economic Review*, 104(2): 564-608.
- [30] Kalouptside, M. (2014b). “Detection and Impact of Industrial Subsidies: The Case of World Shipbuilding,” NBER Working Paper 20119.
- [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 Wolpin, K. I. (1997). “The career decisions of young men,” *Journal of Political Economy*, 105 (3), 473–522.
- [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] Koning, R. H., H. Neudecker and T. Wansbeek. (1991). “Block Kronecker Products and the vecb Operator”. *Linear Algebra and Its Applications*. 165-184.
- [35] Magnac, T. and D. Thesmar (2002). “Identifying dynamic discrete decision processes,” *Econometrica*, 70(2), 801–816.
- [36] Marschak, J. (1953). “Economic Measurements for Policy and Prediction,” in *Studies in Econometric Method*, eds.W. Hood and T. Koopmans, New York: Wiley.
- [37] Matzkin, R.L. (2007). “Nonparametric Identification,” Chapter 73 in *Handbook of Econometrics*, Vol. 6b, edited by J.J. Heckman and E.E. Leamer, Elsevier B.V., 5307-5368.
- [38] Miller, R.A. (1984). “Job matching and occupational choice,” *Journal of Political Economy*, 92, 1086–1120.
- [39] Norets, A. (2009). “Inference in Dynamic Discrete Choice Models With Serially Correlated Unobserved State Variables,” *Econometrica*, 77, 1665–1682.
- [40] Norets, A. and X. Tang (2014). “Semiparametric Inference in Dynamic Binary Choice Models,” *Review of Economic Studies*, 81(3), 1229-1262.
- [41] Pakes, A. (1986). “Patents as options: some estimates of the value of holding European patent stocks,” *Econometrica*, 54, 755–784.

- [42] Pesendorfer, M. and Schmidt-Dengler (2008). “Asymptotic Least Square Estimators for Dynamic Games,” *Review of Economic Studies*, 75, 901–908.
- [43] Phillips, P. C. B. and H. Moon (1999). “Linear Regression Limit Theory for Nonstationary Panel Data,” *Econometrica*, 67 (5), 1057-1111.
- [44] 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.
- [45] 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.
- [46] Rust, J. (1987). “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher,” *Econometrica*, 55, 999-1033.
- [47] Rust, J. (1994). “Structural estimation of Markov decision processes,” In: Engle, R., McFadden, D. (Eds.), *Handbook of Econometrics*. North-Holland, New York, pp. 3081-3143.
- [48] Rust, J. (2014). “The Limits of Inference with Theory: A Review of Wolpin (2013),” *Journal of Economic Literature*, 52(3). 820-50.
- [49] Rust, J., and C. Phelan (1997). “How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets,” *Econometrica*, 65, 781-832.
- [50] Ryan, S. P. (2012). “The Costs of Environmental Regulation in a Concentrated Industry,” *Econometrica*, 80. 1019-1061.
- [51] Scott, P. T. (2013). “Dynamic Discrete Choice Estimation of Agricultural Land Use,” Working Paper, Toulouse School of Economics.
- [52] Suzuki, J. (2013). “Land Use Regulation as a Barrier to Entry: Evidence from the Texas Lodging Industry,” *International Economic Review*, 54 (2), 495-523.
- [53] Todd, P. and K. I. Wolpin (2006). “Assessing the Impact of a School Subsidy Program in Mexico.” *American Economic Review*, 96 (5), 1384-1417.
- [54] Wolpin, K. I. (1984). “A dynamic stochastic model of fertility and child mortality,” *Journal of Political Economy*, 92, 852–874.
- [55] Wolpin, K. I. (1987). “Estimating a Structural Search Model: The Transition from Schooling to Work,” *Econometrica*, 55, 801-818.

A Appendix: Proofs of Propositions

A.1 Identification of the Dynamic Discrete Choice Model

A.1.1 Proof of Proposition 9

Under the state decomposition, the transition matrix can be written as the Kronecker product $F_a = F^\omega \otimes F_a^k$. We make use of the following Lemma:

Lemma 29 *Let $D_a = [I - \beta (F^\omega \otimes F_a^k)]^{-1}$ and $\mathbf{1}$ the block vector:*

$$\mathbf{1} = [I_k, I_k, \dots, I_k]'$$

Then

$$D_a \mathbf{1} = \mathbf{1} (I - \beta F_a^k)^{-1}$$

That is, the sum of entries on each block-row of D_a is constant for all block-rows.

Proof. Let $q'_{\omega a}$ be the ω -block row of D_a . Then $q'_{\omega a} = e'_\omega D_a$, or $q'_{\omega a} D_a^{-1} = e'_\omega$, or $q'_{\omega a} (I - \beta (F^\omega \otimes F_a^k)) = e'_\omega$, which implies $q'_{\omega a} - \beta q'_{\omega a} (F^\omega \otimes F_a^k) = e'_\omega$. We form the sum of the entries of $q'_{\omega a}$ by multiplying with $\mathbf{1}$:

$$q'_{\omega a} \mathbf{1} - \beta q'_{\omega a} (F^\omega \otimes F_a^k) \mathbf{1} = e'_\omega \mathbf{1}$$

Now,

$$(F^\omega \otimes F_a^k) \mathbf{1} = \begin{bmatrix} F_{11}^\omega F_a^k & F_{12}^\omega F_a^k & \dots & F_{1\Omega}^\omega F_a^k \\ \vdots & \vdots & \vdots & \vdots \\ F_{\Omega 1}^\omega F_a^k & F_{\Omega 2}^\omega F_a^k & \dots & F_{\Omega \Omega}^\omega F_a^k \end{bmatrix} \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} = \begin{bmatrix} F_a^k \\ \vdots \\ F_a^k \end{bmatrix} = \mathbf{1} F_a^k$$

where F_{ij}^ω is the (i, j) element of F^ω , and $\sum_{\omega'} F_{\omega \omega'}^\omega = 1$. So that

$$q'_{\omega a} \mathbf{1} - \beta (q'_{\omega a} \mathbf{1}) F_a^k = I_k$$

or

$$q'_{\omega a} \mathbf{1} (I - \beta F_a^k) = I_k$$

■

We now provide the proof of Proposition 9; we focus on the binary choice $\{a, J\}$ for notational simplicity, the general case is obtained in the same fashion. Let $\theta = [\theta_0^{a'}, \theta_0^{J'}, \theta_1^{a'}, \theta_1^{J'}]'$ be the vector of $4K$ unknown parameters (e.g. $\theta_0^a = [\theta_0(a, k_1), \dots, \theta_0(a, k_K)]'$). The parametric form of interest is linear in the parameters; stacking the payoffs for a given ω and all k we have:

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

and

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

Collecting $\pi_a(\omega)$ for all ω , we get $\pi_a = \bar{\pi}_a \theta$, where

$$\bar{\pi}_a = \begin{bmatrix} I_k & 0_k & R_a(\omega_1)I_k & 0_k \\ \vdots & \vdots & \vdots & \vdots \\ I_k & 0_k & R_a(\omega_\Omega)I_k & 0_k \end{bmatrix} \quad (36)$$

and similarly for π_J . In the main text (see (11)), we showed that identification hinges on the matrix $(\bar{\pi}_a - A_a \bar{\pi}_J)$. This matrix equals:

$$\bar{\pi}_a - A_a \bar{\pi}_J = \begin{bmatrix} \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix}, & -A_a \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix}, & R_a, & -A_a R_J \end{bmatrix} \quad (37)$$

where $R_a = [R_a(\omega_1)I_k, \dots, R_a(\omega_\Omega)I_k]'$ (the same for R_J). Note that

$$\begin{aligned} A_a \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} &= \left(I - \beta \left(F^\omega \otimes F_a^k \right) \right) \left(I - \beta \left(F^\omega \otimes F_J^k \right) \right)^{-1} \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} \\ &= \left(I - \beta \left(F^\omega \otimes F_a^k \right) \right) \begin{bmatrix} \left(I - \beta F_J^k \right)^{-1} \\ \vdots \\ \left(I - \beta F_J^k \right)^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \left(I - \beta F_J^k \right)^{-1} \\ \vdots \\ \left(I - \beta F_J^k \right)^{-1} \end{bmatrix} - \beta \begin{bmatrix} F_a^k \left(I - \beta F_J^k \right)^{-1} \\ \vdots \\ F_a^k \left(I - \beta F_J^k \right)^{-1} \end{bmatrix} = \begin{bmatrix} Q_a^{-1} Q_J \\ \vdots \\ Q_a^{-1} Q_J \end{bmatrix} \end{aligned}$$

where $Q_a = \left(I - \beta F_a^k \right)^{-1}$ and likewise for J ; to go from the first to the second line we use Lemma 29, while the third line employs the Kronecker product definition as well as the fact that F^ω is a stochastic matrix whose rows sum to 1. It follows that the first two block columns of (37) consist of identical blocks each (the first block column has elements I_k , and the second, $Q_a^{-1} Q_J$). As a consequence, the respective block parameters θ_0^a , θ_0^J , are not identified unless a strong normalization is imposed.³⁶ The remaining parameters, θ_1^a , θ_1^J , are identified as follows.

Consider equation (11), which in the binary case becomes $(\bar{\pi}_a - A_a \bar{\pi}_J) \theta = b_a$, or using (37):

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

³⁶In 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} .

Left-multiplying both sides by $D_a = \left[I - \beta \left(F^\omega \otimes F_a^k \right) \right]^{-1}$ and using Lemma 29, we obtain:³⁷

$$\begin{bmatrix} Q_a \\ \vdots \\ Q_a \end{bmatrix} \theta_0^a - \begin{bmatrix} Q_J \\ \vdots \\ Q_J \end{bmatrix} \theta_0^J + D_a R_a \theta_1^a - D_J R_J \theta_1^J = D_a b_a.$$

Take the ω block row of the above:

$$Q_a \theta_0^a - Q_J \theta_0^J + e'_\omega D_a R_a \theta_1^a - e'_\omega D_J R_J \theta_1^J = e'_\omega D_a b_a$$

Since $|\Omega| \geq 3$, take two other distinct block rows corresponding to $\tilde{\omega}, \bar{\omega}$ and difference both from the above to obtain:

$$\begin{bmatrix} (e'_\omega - e'_{\tilde{\omega}}) D_a R_a & (e'_{\tilde{\omega}} - e'_\omega) D_J R_J \\ (e'_\omega - e'_{\bar{\omega}}) D_a R_a & (e'_{\bar{\omega}} - e'_\omega) D_J R_J \end{bmatrix} \begin{bmatrix} \theta_1^a \\ \theta_1^J \end{bmatrix} = \begin{bmatrix} (e'_\omega - e'_{\tilde{\omega}}) D_a b_a \\ (e'_\omega - e'_{\bar{\omega}}) D_a b_a \end{bmatrix}$$

which proves the Proposition.

A.1.2 Proof of Lemma 11

Using the definition of A_a from Proposition 7,

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

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

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

A.2 Identification of Counterfactuals

A.2.1 Proof of Proposition 17

H has the form

$$H = \text{diag} \{ \lambda_1 I_{n_1}, \dots, \lambda_k I_{n_k} \}$$

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

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

³⁷To see this note that:

$$D_a \begin{bmatrix} Q_a^{-1} Q_J \\ \vdots \\ Q_a^{-1} Q_J \end{bmatrix} \theta_0^J = D_a \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} Q_a^{-1} Q_J \theta_0^J = \begin{bmatrix} Q_a \\ \vdots \\ Q_a \end{bmatrix} Q_a^{-1} Q_J \theta_0^J = \begin{bmatrix} I_k \\ \vdots \\ I_k \end{bmatrix} Q_J \theta_0^J.$$

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

$$A_a = \text{diag} \{A_1^a, \dots, A_k^a\}$$

Then the definition of A_a implies:

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

or

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

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

$$F_{ij}^a - A_i^a F_{ij}^J = 0 \quad (38)$$

while for the diagonal blocks we have

$$\frac{1}{\beta} (I - A_i^a) = F_{ii}^a - A_i^a F_{ii}^J \quad (39)$$

Equations (38) and (39) summarize the restrictions placed on transition probabilities. We can isolate A_i^a from (39) and substitute into (38) to obtain

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

and (19).

Note that (19) implies that, given F^J and F_{ii}^a , all remaining blocks of F^a are uniquely determined. We must still guarantee, however, that F^a are stochastic matrices so that their rows add to 1 and all elements are between 0 and 1. Indeed, consider the first block row of F^a :

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

Then each row belonging to this block row must add to 1:

$$F_{11}^a \mathbf{1} + F_{12}^a \mathbf{1} + \dots + F_{1k}^a \mathbf{1} = \mathbf{1}$$

where, abusing notation slightly, the vectors $\mathbf{1}$ above have varying length. Using (38) and the fact that the rows of F^J add to one, we get

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

or, using (39) as well,

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

we obtain (20).

Proof of Corollaries: The case of $H = \lambda I$ follows from (18). Next, suppose that one of the eigenvalues of H , say λ_1 , is simple, that is $n_1 = 1$ and $k > 1$ (this case includes pairwise distinct eigenvalues). Then, equations (38) and (39) give

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

Summing over j and taking into account that row elements of F_a and F_J sum to one, we have that $F_{11}^a = A_1^a F_{11}^J$ and $\frac{1}{\beta} (1 - A_1^a) = 0$. Therefore, $A_1^a = 1$ and the corresponding rows of F_a and F_J are equal.

A.3 Identification with Unobservable Market-level States

In order to prove Propositions 27 and 28, we make use of the following Lemmas.

Lemma 30 *For any action a , the expectational error term $\varepsilon^\zeta(a, k, \omega, \omega^*)$ is mean independent of k, ω : $E[\varepsilon^\zeta(a, k, \omega, \omega^*) | k, \omega] = 0$.*

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

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

■

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

Lemma 31 *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 $\{(k_{imt}, w_{mt}, z_{mt}) : i = 1, \dots, N; m = 1, \dots, M; t = 1, \dots, T\}$. Assume that for each m and t , one can obtain the estimators $\hat{g}_{mt}^N(k) \xrightarrow{p} g(k, \omega_{mt})$ and $\hat{F}_{mt}^N(k', k) \xrightarrow{p} F(k'|k, \omega_{mt})$ as $N \rightarrow \infty$.³⁸ For any function $h(k, z_{mt})$, define the estimators*

$$\begin{aligned} &\frac{1}{MT} \sum_{m,t=1}^{MT} \left[h(k, z_{mt}) \hat{g}_{mt}^N(k) \right], \text{ and} \\ &\frac{1}{M(T-1)} \sum_{m,t=1}^{M(T-1)} \left[h(k, z_{mt}) \hat{g}_{mt+1}^N(k) \hat{F}_{mt}^N(k', k) \right]. \end{aligned}$$

Assume the following uniform conditions hold: (i)

$$\begin{aligned} \limsup_{M,T,N} \left[\frac{1}{MT} \sum_{m,t=1}^{MT} E \left\| h(k, z_{mt}) \hat{g}_{mt}^N(k') \right\| \right] &< \infty, \\ \limsup_{M,T,N} \left[\frac{1}{M(T-1)} \sum_{m,t=1}^{M(T-1)} E \left\| h(k, z_{mt}) \hat{g}_{mt+1}^N(k') \hat{F}_{mt}^N(k', k) \right\| \right] &< \infty; \end{aligned} \quad (40)$$

and (ii)

$$\limsup_{M,T,N} \left[\frac{1}{MT} \sum_{m,t=1}^{MT} \left\| E \left[h(k, z_{mt}) \left(\hat{g}_{mt}^N(k') - g(k', \omega_{mt}) \right) \right] \right\| \right] = 0$$

³⁸Note that the asymptotic results $\hat{g}_{mt}^N(k) \xrightarrow{p} g(k, \omega_{mt})$ and $\hat{F}_{mt}^N(k', k) \xrightarrow{p} F(k'|k, \omega_{mt})$ as $N \rightarrow \infty$ can be obtained using the law of large numbers for exchangeable random variables (see, e.g., Hall and Heyde, 1980), provided the observations $i = 1, \dots, N$ for each index (m, t) are i.i.d. conditional on ω_{mt} .

$$\limsup_{M,T,N} \left[\frac{1}{M(T-1)} \sum_{m,t=1}^{M(T-1)} \left\| E \left[h(k, z_{mt}) \begin{pmatrix} \hat{g}_{mt+1}^N(k') \hat{F}_{mt}^N(k', k) \\ -g(k', \omega_{mt+1}) F(k'|k, \omega_{mt}) \end{pmatrix} \right] \right\| \right] = 0 \quad (41)$$

If (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 if

$$E \|h(k, z_{mt}) g(k', \omega_{mt+1})\| < \infty,$$

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

then,

$$\frac{1}{MT} \sum_{m,t=1}^{MT} [h(k, z_{mt}) \hat{g}_{mt}^N(k)] \xrightarrow{p} E [h(k, z_{mt}) g(k, \omega_{mt})], \text{ and}$$

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

as $(N, M, T) \rightarrow \infty$; or as $(N, M) \rightarrow \infty$ (if T is fixed); or as $(N, T) \rightarrow \infty$ (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 and Moon, 1999, Lemma 6 and Theorem 1). The sequential limit can be obtained directly from two facts: (i) $\hat{g}_{mt}^N(k) \xrightarrow{p} g(k, \omega_{mt})$ and $\hat{F}_{mt}^N(k', k) \xrightarrow{p} F(k'|k, \omega_{mt})$ as $N \rightarrow \infty$ implies

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

as $N \rightarrow \infty$ 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

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

then, by the Weak Law of Large Numbers,

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

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

The sequential limit is obtained by first passing the limit $N \rightarrow \infty$ and then the limit $(M, T) \rightarrow \infty$. Provided conditions (40) and (41) hold, by Phillips and Moon's (1999) Lemma 6 and Theorem 1, the sequential limit equals the simultaneous limit as $(N, M, T) \rightarrow \infty$.³⁹ ■

³⁹In 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.

A.3.1 Proof of Proposition 27

(a) Finite T . Suppose first that the terminal value V_{mT} is known. Replace V_{mt+1} in (28) for a specific action j to get:

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

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

$$\begin{aligned} \pi_{amt} + \beta \varepsilon_{am,t,t+1}^V &= V_{mt} - \beta F_{amt}^k \left[\sum_{\tau=1}^{T-1} \beta^{\tau-1} \Lambda_{jm,t,\tau} \left(\pi_{jmt+\tau} + \beta \varepsilon_{jm,t+\tau,t+\tau+1}^V + \sigma \psi_{jmt+\tau} \right) \right] \\ &\quad - \beta^T F_{amt}^k \Lambda_{jm,t,T} V_{mT} - \sigma \psi_{amt} \end{aligned} \quad (42)$$

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

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

Next, evaluate (42) for $a = j$ and subtract it to obtain:

$$\begin{aligned} \pi_{amt} - \pi_{jmt} + \beta \left(\varepsilon_{am,t,t+1}^V - \varepsilon_{jm,t,t+1}^V \right) &= \beta^T \left(F_{jmt}^k - F_{amt}^k \right) \Lambda_{jm,t,T} V_{mT} - \sigma \left(\psi_{amt} - \psi_{jmt} \right) + \\ &\quad + \beta \left(F_{jmt} - F_{amt} \right) \left[\sum_{\tau=1}^{T-1} \beta^{\tau-1} \Lambda_{jm,t,\tau} \left(\pi_{jmt+\tau} + \beta \varepsilon_{jm,t+\tau,t+\tau+1}^V + \sigma \psi_{jmt+\tau} \right) \right]. \end{aligned} \quad (43)$$

For any known vector function $h(z_{mt})$, with elements $h(k, z_{mt})$, apply the Hadamard multiplication on both sides of (43) and take expectation. We eliminate the error terms, $\varepsilon_{am,t+\tau,t+\tau+1}^V$ and $\xi_{akmt+\tau}$, because z_{mt} is in the time- t information set. Then,

$$\begin{aligned} E[h(z_{mt}) \circ \bar{\pi}_{amt}] &= E[h(z_{mt}) \circ (\bar{\pi}_{jmt} - \sigma(\psi_{amt} - \psi_{jmt}))] + \\ &\quad + \beta E \left[h(z_{mt}) \circ (F_{jmt} - F_{amt}) \left[\sum_{\tau=1}^{T-1} \beta^{\tau-1} \Lambda_{jm,t,\tau} (\bar{\pi}_{jmt+\tau} + \sigma \psi_{jmt+\tau}) \right] \right] \\ &\quad + \beta^T E \left[h(z_{mt}) \circ \left(F_{jmt}^k - F_{amt}^k \right) \Lambda_{jm,t,T} V_{mT} \right]. \end{aligned} \quad (44)$$

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

If the payoff $\bar{\pi}(j, k_{imt}, w_{mt})$, the scale parameter σ and the terminal value function V_{mT} are known, then the RHS of (44) can be recovered from the data (using the results of Lemma 31). Because the RHS of (44) is known, for any two structures b and b' , with corresponding payoffs $\bar{\pi}$ and $\bar{\pi}'$, we have

$$E[h(z_{mt}) \circ (\bar{\pi}_{amt} - \bar{\pi}'_{amt})] = 0$$

for any function h . By the completeness condition (Condition 25), the equality above implies $\bar{\pi}_{amt} - \bar{\pi}'_{amt} = 0$ almost everywhere.

Next, consider the case of a renewal action J . Take (30), multiply both sides by $h(z_{mt})$ and take expectations:

$$\begin{aligned} &E \left[h(z_{mt}) \circ \left(\bar{\pi}_{amt} - \bar{\pi}_{jmt} - \beta \left(F_{amt}^k - F_{jmt}^k \right) \bar{\pi}_{Jmt+1} \right) \right] \\ &= \sigma E \left[h(z_{mt}) \circ \left(\beta \left(F_{amt}^k - F_{jmt}^k \right) \psi_{Jmt+1} - (\psi_{amt} - \psi_{jmt}) \right) \right]. \end{aligned}$$

Similar to the previous case, the RHS can be recovered from data (using Lemma 31). Then, for any two structures b and b' with corresponding payoffs $\bar{\pi}$ and $\bar{\pi}'$,

$$\begin{aligned} & E \left[h(z_{mt}) \circ \left(\bar{\pi}_{amt} - \bar{\pi}_{jmt} - \beta \left(F_{amt}^k - F_{jmt}^k \right) \bar{\pi}_{Jmt+1} \right) \right] \\ &= E \left[h(z_{mt}) \circ \left(\bar{\pi}'_{amt} - \bar{\pi}'_{jmt} - \beta \left(F_{amt}^k - F_{jmt}^k \right) \bar{\pi}'_{Jmt+1} \right) \right]. \end{aligned}$$

By the completeness condition (Condition 25),

$$\bar{\pi}_{amt} - \bar{\pi}_{jmt} - \beta \left(F_{amt}^k - F_{jmt}^k \right) \bar{\pi}_{Jmt+1} = \bar{\pi}'_{amt} - \bar{\pi}'_{jmt} - \beta \left(F_{amt}^k - F_{jmt}^k \right) \bar{\pi}'_{Jmt+1} \quad (45)$$

for almost all (w_{mt}, w_{mt+1}) . Consider (45) for $j = J$. Because $\bar{\pi}_J(k, w)$ is known or prespecified for all observed states (k, w) , we conclude that $\bar{\pi}_{amt} - \bar{\pi}'_{amt} = 0$ 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 $\bar{\pi}_J$ is known. Instead we now assume $\bar{\pi}_j$ is known (or strongly normalized) for some action j for all k and w . This implies $\bar{\pi}_j = \bar{\pi}'_j$. Take the equation (45) above. Then, for almost all (w_{mt}, w_{mt+1}) ,

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

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

$$\bar{\pi}_{Jm\tau} - \bar{\pi}'_{Jm\tau} = \beta^T \prod_{t=\tau}^T (F_{Jmt} - F_{jmt}) (\bar{\pi}_{JmT} - \bar{\pi}'_{JmT}) \quad (46)$$

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 (46) converges to zero provided the flow payoffs $\bar{\pi}_J(k, w)$ and $\bar{\pi}'_J(k, w)$ are bounded for almost all (k, w) .

Table 5: 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 – altitude and slope	http://dds.cr.usgs.gov/srtm/
NASS Quick Stats	Yields, prices, pasture rental rates	http://www.nass.usda.gov/Quick_Stats/
ERS	Operating costs	http://www.ers.usda.gov/data-products/commodity-costs-and-returns.aspx
NOOA Urban Centers	Urban center locations and populations	http://www.nws.noaa.gov/geodata/catalog/national/html/urban.htm

A.3.2 Proof of Proposition 28

The proof is similar to that of Proposition 27. Suppose first that σ is known. Multiply both sides of (25) by a known function $h(k, z_{mt})$ and take expectations:

$$E[h(k, z_{mt}) \bar{\pi}(a, k, w_{mt})] = E[h(k, z_{mt}) (V(k, \omega_{mt}) - \sigma \psi(a, k, \omega_{mt}))] - \beta E \left[h(k, z_{mt}) \sum_{k'} V(k', \omega_{mt+1}) F^k(k'|a, k, w_{mt}) \right] \quad (47)$$

where the expectations are taken over $(z_{mt}, \omega_{mt}, \omega_{mt+1})$. The RHS can be recovered from data (using Lemma 31). Then, for any two structures b and b' with corresponding payoffs $\bar{\pi}$ and $\bar{\pi}'$,

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

By the completeness condition (Condition 25), $\bar{\pi}(a, k, w_{mt}) - \bar{\pi}'(a, k, w_{mt}) = 0$ almost everywhere.

Next, suppose σ is not known, but $\bar{\pi}(a, k, w)$ is known (or prespecified) for one combination of (a, k, w) . Take (47) for the known $\bar{\pi}(a, k, w)$:

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

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

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

and so $\sigma = \sigma'$.

B Appendix: Data (for online publication)

Table 5 lists the data sources. All are publicly available for download save DataQuick's land value data.

The Cropland Data Layer (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.

Table 6: 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

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 “crop” and “non-crops” are defined. When we refer to a “field,” this can mean two things in terms of our data. For analysis based on land use data only (not including land value data), a field refers to one of these points in the CDL grid. For empirical models that include land values, a field refers to a transacted parcel in the Dataquick database.

We know the coordinates of the centroids of transacted parcels in the DataQuick database. To assign transacted parcels a land use, we associate a parcel with the nearest point in the CDL grid. Similarly, slopes and soil categories are assigned to fields/parcels using nearest neighbor interpolation.

From the GAEZ database, we take soil categories and protected status. Protected lands were dropped from all analyses. The SRTM data provides altitudes at a 30m grid scale which are used to calculate slopes.

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).

Table 6 compares the transacted fields (in DataQuick) to all US fields (in the CDL). To investigate whether land use changes upon resale, we used a linear probability model and found no such evidence (see Table 7). 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.

Table 7: Land use and transactions

VARIABLES	(1) incrops2010	(2) incrops2011	(3) incrops2012	(4) incrops2013
soldin2009	0.000647 (0.00604)			
soldin2010	0.000116 (0.00326)	0.00364 (0.00334)		
soldin2011	-0.00117 (0.00316)	0.000629 (0.00324)	-0.00159 (0.00330)	
soldin2012		-0.000620 (0.00306)	-0.00472 (0.00313)	0.00411 (0.00265)
soldin2013			-0.00962*** (0.00306)	-0.000445 (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.

C Appendix: Estimation (for online publication)

C.1 Conditional Choice Probabilities

We estimate conditional choice probabilities using a semiparametric logit regression. We perform a separate regression for each county, field state, and year. In a regression, fields in nearby counties are given weights which are proportional to the inverse square difference between the centroids of the two counties. The sample of fields included in a given county's regression are all other fields within the same state, and potentially some neighboring states.

The logit regression for county m , field state k , and year t maximizes the following log likelihood function:

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

where S_m is the set of counties in the same 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 $p_{mt}(c, k, \zeta_{im}; \theta_{ckt})$ is parameterized as follows:

$$p_{mt}(c, k, \zeta_{im}; \theta_{ckt}) = \frac{\exp(\theta_{0ckt} + \theta_{sckt} slope_{im})}{1 + \exp(\theta_{0ckt} + \theta_{sckt} slope_{im})}$$

where $slope_{im}$ is field i 's slope. Slope proves to be a powerful predictor of agricultural land use decisions. Note that without the slope covariate, this regression would amount to taking

frequency estimates for each county, field state, and year (interacted), with some smoothing across counties. Including slope allows for some within-county heterogeneity.

The set of counties in S_m only includes counties which appear in our sample – that is, counties appearing in the DataQuick database. For some states, the database includes a small number of counties (see Figure 1), 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.⁴⁰

In a few instances, the semiparametric logit regressions converge very slowly or not at all. After twenty maximum likelihood iterations, we switch to a penalized logistic regression (i.e., “firthlogit” in STATA, where large values of coefficients are penalized).⁴¹

C.2 Resale Price Hedonics

Before estimating a hedonic model of land prices, we drop many transactions in the DataQuick database in which the transaction price is not plausibly a reflection of the parcel’s value. 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.

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 hedonic regression equation:

$$\ln V_{it} = X'_{it}\theta_V + \eta_{it}$$

where V_{it} is a transaction price (in dollars per acre), and X_{it} is a vector of characteristics for the corresponding field.⁴² The covariates X_{it} include all variables in Table 2, year dummies, returns interacted with year dummies, field state dummies interacted with year dummies, and county dummies. Table 8 presents the estimated coefficients.

⁴⁰In 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.

⁴¹The “firthlogit” (Firth (1993)), adds to the loglikelihood function the penalization term $\frac{1}{2} \log |\mathbf{I}(\theta)|$, where $\mathbf{I}(\theta)$ is the information matrix evaluated at θ . The penalization term avoids the “separation” problem: in finite samples, the maximum likelihood estimate may not exist because at least one parameter estimate may be infinite (even when the true parameter is not). This situation can occur when the dependent variable can be perfectly predicted by a single regressor or by a non-trivial linear combination of regressors.

⁴²Field acreage is available only in the DataQuick dataset; therefore 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. Therefore, we choose a specification for the value function that regresses price per acre on covariates.

Table 8: Hedonic Regression

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

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

Omitted: soil, county, year, and field state dummies
as well as interactions with returns.