

WORKING PAPER · NO. 2017-17

Identifying the Discount Factor in Dynamic Discrete Choice Models

Jaap H. Abbring and Øystein Daljord

MAY 2019

Identifying the Discount Factor in Dynamic Discrete Choice Models*

Jaap H. Abbring[†] Øystein Daljord[‡]

May 2019

Abstract

Empirical research often cites observed choice responses to variation that shifts expected discounted future utilities, but not current utilities, as an intuitive source of information on time preferences. We study the identification of dynamic discrete choice models under such economically motivated exclusion restrictions on primitive utilities. We show that each exclusion restriction leads to an easily interpretable moment condition with the discount factor as the only unknown parameter. The identified set of discount factors that solves this condition is finite, but not necessarily a singleton. Consequently, in contrast to common intuition, an exclusion restriction does not in general give point identification. Finally, we show that exclusion restrictions have nontrivial empirical content: The implied moment conditions impose restrictions on choices that are absent from the unconstrained model.

*We thank Nikhil Agarwal, Áureo de Paula, Jean-Pierre Dubé, Hanming Fang, Christian Hansen, Günter Hitsch, John-Eric Humphries, Robert Miller, Whitney Newey, John Rust, Azeem Shaikh, Eduardo Souza-Rodrigues, Elie Tamer, Michela Tincani, Thomas Wollmann, participants in the 2016 Cowles Commission Summer Conference, the 69th European Meeting of the Econometric Society, the 2nd Conference on Structural Dynamic Models in Copenhagen, the BRIQ Workshop on Structural Analysis of Inequality in Bonn, and seminars at Tilburg, Toronto, Copenhagen, Chicago, UCL, LSE, Penn State, DIW Berlin, Georgetown, EIEF Rome, Cambridge, Warwick, Bonn, Zürich, UC Irvine, CalTech, and USC for helpful discussions and comments. Jaap Abbring's research is financially supported by the Netherlands Organisation for Scientific Research (NWO) through Vici grant 453-11-002.

[†]CentER, Department of Econometrics & OR, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; and CEPR. E-mail: jaap@abbring.org. Web: jaap.abbring.org.

[‡]Booth School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA. E-mail: Oeystein.Daljord@chicagobooth.edu. Web: faculty.chicagobooth.edu/oystein.daljord.

Keywords: discount factor, dynamic discrete choice, empirical content, identification.

JEL codes: C25, C61.

1 Introduction

Identification of the discount factor in dynamic discrete choice models is crucial for their application to the evaluation of agents' responses to dynamic interventions. It is, however, well known that the discount factor is not identified from choice data without further restrictions (Rust, 1994, Lemma 3.3, and Magnac and Thesmar, 2002, Proposition 2). Consequently, empirical researchers usually fix the discount factor at some a priori plausible value, e.g. 0.95, or impose ad hoc functional form assumptions that allow it to be identified and estimated. These approaches solve the identification problem, but often lack economic justification. Inferring the time preferences in the specific context of an application is important as discount factors have been estimated to vary substantially across choice contexts and populations (Frederick et al., 2002).¹

In this paper, we explore identification from observed choice responses to variation that shifts expected discounted future utilities, but not current utilities. Such variation is commonly cited in applications as an intuitive source of information on time preferences. For example, in studies of green technology adoption, Bollinger (2015) and De Groote and Verboven (2018) argued that firms' and households' current choice responses to regulation that shifts their future expenses, but not their current expenses, are informative about discount factors. In a study of demand for game consoles, Lee (2013) assumed that the discount factor is identified from variation in the expected quality of future releases, which shifts future values without affecting current payoffs. In an application to cellphone plan choice, Yao et al. (2012) argues informally that utilities can be identified in a terminal period when the choice problem is static. The discount factor can subsequently be identified from choices in the next to last period. Chung et al. (2014) appeals to Yao et al. (2012)'s idea in a study of salesforce compensation plans. We give further examples from the literature in Section 3.5.

In Section 3, we formalize the intuition in these studies as an exclusion restriction on primitive utilities. We first consider a stationary model with infinite horizon (introduced in Section 2). We prove that, in contrast to common intuition, an exclusion restriction on primitive utilities does not generally point identify the discount factor. It does however narrow the identified set—the set of observationally equivalent discount factors—to a discrete and, if we exclude values near one, finite set. This set contains the solutions to a moment condition that only involves the discount

¹Frederick et al. also showed that geometric discounting is often rejected in data in favor of present biased time preferences. We study the identification and estimation of hyperbolic discount functions in Abbring et al. (2018).

factor and that has a straightforward interpretation in terms of choice responses to variation in expected discounted future utilities. The moment condition can be used directly in estimation, independently of the rest of the model parameters.

We subsequently provide a finite upper bound on the number of discount factors in the identified set for the case in which the states display finite dependence, as defined by [Arcidiacono and Miller \(2011, 2017\)](#). Examples include optimal stopping and renewal problems, which we show to be point identified.

We extend our analysis to nonstationary models with finite horizons, which are commonly used in labor applications ([Eckstein and Wolpin, 1989](#) and [Keane and Wolpin, 1997](#) are early examples). We show that, with exclusion restrictions, the discount factor is generally identified up to a finite set in these models.

In Section 4, we explore the empirical content of exclusion restrictions. [Magnac and Thesmar](#)’s Proposition 2 implies that dynamic discrete choice models without exclusion restrictions cannot be falsified with data on choices and states. In that sense, the models have no empirical content. We show that exclusion restrictions impose nontrivial restrictions on the data, which can be tested.

Finally, in Section 5, we argue that common intuition often supports *multiple* exclusion restrictions, which imply multiple moment conditions. These moment conditions share the true discount factor (if one exists that rationalizes the data) as one solution, but may have individually more solutions. We discuss how standard (set) estimators can be applied to this case.

This paper’s main contribution is to provide a simple and intuitively appealing analysis of identification of discount factor in dynamic discrete choice models under economically motivated exclusion restrictions. Our analysis complements a substantial literature in econometrics (see [Rust, 1994](#) and [Abbring, 2010](#), for reviews). [Magnac and Thesmar](#)’s Proposition 4 established point identification based on a different type of exclusion restriction than ours: the existence of a pair of states that affects, in some specific way, expected discounted future utilities, but not the “current value,” which is a difference in expected discounted utilities between two particular choice sequences. This is a high level exclusion restriction that is difficult to interpret and hard to verify in applications. In particular, unlike our exclusion restriction, it does not formalize the common intuition that is given in applications like those discussed above. Empirical applications often incorrectly cite [Magnac and Thesmar](#)’s result as one for an exclusion restriction on primitive utility. For example, in a study of housing location choice, [Bayer et al. \(2016, p. 921\)](#) wrote

Magnac and Thesmar (2002) . . . showed that dynamic models are identified with an appropriate exclusion restriction— in particular, a variable

that shifts expectations but not current utility. In the context of our framework, lagged amenities provide exactly this sort of exclusion restriction: while current utility depends on the current level of the amenities provided in a neighborhood, lagged amenity levels help predict how amenities will evolve going forward and thus contribute to expectations about the future utility associated with that choice of neighborhood.

We show how Bayer et al.’s exclusion restriction can be used to set identify and estimate the discount factor, even if it is insufficient for point identification.²

Magnac and Thesmar’s identification result relies on a rank condition that ensures sufficient variation in expected discounted future utilities. This rank condition does not suffice for point identification with our exclusion restriction on primitive utilities. We do however use natural extensions of this condition to ensure local identification of myopic preferences, which is needed for our discrete set identification result.

Magnac and Thesmar’s Proposition 2 implies that, without further restrictions, not only the discount factor, but also the utility of one reference choice can be normalized without restricting the observed choice and transition probabilities. Intuitively, discrete choices only identify utility contrasts, not levels. However, *counterfactual* choice probabilities, which are often the objects of interest in dynamic discrete choice analysis, are generally not invariant to the choice of reference utility (Norets and Tang, 2014; Kalouptsi et al., 2016). This suggests that we do not only treat the discount factor, but also the utility of the reference choice as a free parameter that should be determined from data. Indeed, we view the identification of the reference utility as an important, but separate problem from the identification of the discount factor. For expositional convenience, we derive our main results under the normalization that the reference utility equals zero. In the appendix, we show that our results straightforwardly extend to the case in which the reference utility is known up to a constant shift.

We emphasize that the idea of using exclusion restrictions to identify time preferences in choice models is not ours, but has circulated in the literature for a while. One early example is Chevalier and Goolsbee (2009), which studied demand for textbooks. Its choice model implicitly excluded the expected future resale price of a textbook from the current period pay-off to identify a discount factor. Fang and Wang (2015) studied identification of dynamic discrete choice models with partially naive hyperbolic time preferences under explicit exclusion restrictions similar to the ones we use. In Abbring and Daljord (2019), we show that its main generic iden-

²We thank John Rust for this example.

tification result is void: It uses a definition of generic identification that does not preclude the possibility that the model is nowhere identified. We also show that its proof is incorrect. It therefore has no implications for the identification of the model with hyperbolic discounting or its special case with exponential discounting, which we study.

[Komarova et al. \(2017\)](#) showed point identification of the discount factor under parametric assumptions on the utility function in a model like ours, but without exclusion restrictions. [Norets and Tang \(2014\)](#) demonstrated that in a model with parametric utility, point identification is lost to set identification when the distribution of unobservables is allowed to deviate from a known one such as the type-1 extreme value specification that underlies logit choice probabilities. Without any restrictions on the distribution of unobservables beyond conditional independence and absolute continuity, all identification of the discount factor is lost, i.e. the identified set of discount factors is the unit interval. We instead focus on identification for a non-parametric utility function under economically motivated exclusion restrictions. We map each exclusion restriction to an easily interpretable and computable moment condition that directly informs the identification and estimation of the discount factor, and the model’s empirical content.

2 Model

Consider a stationary dynamic discrete choice model (e.g. [Rust, 1994](#)). Time is discrete with an infinite horizon.³ In each period, agents first observe state variables x and ε , where x takes discrete values in $\mathcal{X} = \{x_1, \dots, x_J\}$ and $\varepsilon = \{\varepsilon_1, \dots, \varepsilon_K\}$ is continuously distributed on \mathbb{R}^K ; for $J, K \geq 2$. Then, they choose d from the set of alternatives $\mathcal{D} = \{1, 2, \dots, K\}$ and collect utility $u_d(x, \varepsilon) = u_d^*(x) + \varepsilon_d$. Finally, they move to the next period with new state variables x' and ε' drawn from a Markov transition distribution controlled by d . We assume that a version of [Rust’s \(1987\)](#) conditional independence assumption holds. Specifically, x' is drawn independently of ε from the transition distribution $Q_k(\cdot|x)$ for any choice $k \in \mathcal{D}$; and $\varepsilon_1, \dots, \varepsilon_K$ are independently drawn from mean zero type-1 extreme value distributions.⁴ Agents maximize the rationally expected utility flow discounted with factor $\beta \in [0, 1)$.

Each choice d equals the option k that maximizes the choice-specific expected discounted utility (or, simply, “value”) $v_k(x, \varepsilon)$. The additive separability of $u_k(x, \varepsilon)$

³Section 3.6 considers an extension to a nonstationary model with a finite horizon.

⁴[Magnac and Thesmar](#) showed that the distribution of ε cannot be identified and took it to be known. Our type-1 extreme value assumption leads to the canonical multinomial logit case. Our results extend directly to any other known, continuous distribution on \mathbb{R}^K .

and conditional independence imply that $v_k(x, \varepsilon) = v_k^*(x) + \varepsilon_k$, with v_k^* the unique solution to

$$\begin{aligned} v_k^*(x) &= u_k^*(x) + \beta \mathbb{E} \left[\max_{k' \in \mathcal{D}} \{v_{k'}^*(x') + \varepsilon'_{k'}\} \middle| d = k, x \right] \\ &= u_k^*(x) + \beta \int \mathbb{E} \left[\max_{k' \in \mathcal{D}} \{v_{k'}^*(\tilde{x}) + \varepsilon'_{k'}\} \right] dQ_k(\tilde{x}|x) \end{aligned} \quad (1)$$

for all $k \in \mathcal{D}$. Here, for each given $\tilde{x} \in \mathcal{X}$,

$$\mathbb{E} \left[\max_{k' \in \mathcal{D}} \{v_{k'}^*(\tilde{x}) + \varepsilon'_{k'}\} \right] = \ln \left(\sum_{k' \in \mathcal{D}} \exp(v_{k'}^*(\tilde{x})) \right) \quad (2)$$

is the McFadden surplus for the choice among $k' \in \mathcal{D}$ with utilities $v_{k'}^*(\tilde{x}) + \varepsilon'_{k'}$.

Suppose we have data on choices d and state variables x that allow us to determine $Q_k(\cdot|\tilde{x})$ and the choice probabilities $p_k(\tilde{x}) = \Pr(d = k|x = \tilde{x})$ for all $k \in \mathcal{D}$ and $\tilde{x} \in \mathcal{X}$. The model is point identified if and only if we can uniquely determine its primitives from these data. As we discuss in Section 4, a version of [Magnac and Thesmar's](#) Proposition 2 holds: There exist unique (up to a standard utility normalization) values of the primitives that rationalize the data for any given discount factor $\beta \in [0, 1)$. We therefore focus our identification analysis on β .

The choice probabilities are fully determined by

$$\ln(p_k(\tilde{x})) - \ln(p_K(\tilde{x})) = v_k^*(\tilde{x}) - v_K^*(\tilde{x}), \quad k \in \mathcal{D}/\{K\}, \tilde{x} \in \mathcal{X}. \quad (3)$$

The transition probabilities $Q_k(\cdot|\tilde{x})$, the value contrasts $v_k^*(\tilde{x}) - v_K^*(\tilde{x})$ for $k \in \mathcal{D}/\{K\}$ and $\tilde{x} \in \mathcal{X}$ therefore capture all the model's implications for the data. [Hotz and Miller \(1993\)](#) pointed out that (3) can be inverted to identify the value contrasts from the choice probabilities. To use this, we first rewrite (1) as

$$v_k^*(x) = u_k^*(x) + \beta \int (m(x') + v_K^*(x')) dQ_k(x'|x), \quad (4)$$

where, for given $\tilde{x} \in \mathcal{X}$, $m(\tilde{x}) = \mathbb{E}[\max_{k' \in \mathcal{D}} \{v_{k'}^*(\tilde{x}) - v_K^*(\tilde{x}) + \varepsilon'_{k'}\}]$ is the “excess surplus” (over $v_K^*(\tilde{x})$), the McFadden surplus for the choice among $k' \in \mathcal{D}$ with utilities $v_{k'}^*(\tilde{x}) - v_K^*(\tilde{x}) + \varepsilon'_{k'}$. By (2) and (3), it follows that $m(\tilde{x}) = -\ln(p_K(\tilde{x}))$.

3 Identification

Let \mathbf{v}_k , \mathbf{p}_k , \mathbf{u}_k , and \mathbf{m} be $J \times 1$ vectors with j -th elements $v_k^*(x_j)$, $p_k(x_j)$, $u_k^*(x_j)$, and $m(x_j)$, respectively. Let \mathbf{Q}_k be the $J \times J$ matrix with (j, j') -th entry $Q_k(x_{j'}|x_j)$ and \mathbf{I} be a $J \times J$ identity matrix. Note that the $J \times 1$ vector $\mathbf{m} + \mathbf{v}_K$ stacks the McFadden surpluses in (2). In this notation, the data are $\{\mathbf{p}_k, \mathbf{Q}_k; k \in \mathcal{D}\}$ and directly identify $\mathbf{m} = -\ln \mathbf{p}_K$ (Arcidiacono and Miller, 2011, Lemma 1 and Section 3.3).

3.1 Magnac and Thesmar's result

We can rewrite (4) as $v_k^*(x) = u_k^*(x) + \beta \mathbf{Q}_k(x) [\mathbf{m} + \mathbf{v}_K]$, where $\mathbf{Q}_k(x)$ is the j -th row of \mathbf{Q}_k . Subtracting the same expression for $v_K^*(x)$, rearranging, and substituting (3), we get

$$\ln(p_k(x)) - \ln(p_K(x)) = \beta [\mathbf{Q}_k(x) - \mathbf{Q}_K(x)] \mathbf{m} + U_k(x), \quad (5)$$

where $U_k(x) = u_k^*(x) - u_K^*(x) + \beta [\mathbf{Q}_k(x) - \mathbf{Q}_K(x)] \mathbf{v}_K$ is Magnac and Thesmar's "current value" of choice k in state x . Its Proposition 4 assumes the existence of a known option $k \in \mathcal{D}/\{K\}$ and a known pair of states $\tilde{x}_1, \tilde{x}_2 \in \mathcal{X}$ such that $\tilde{x}_1 \neq \tilde{x}_2$ and $U_k(\tilde{x}_1) = U_k(\tilde{x}_2)$. Under this exclusion restriction, differencing (5) evaluated at \tilde{x}_1 and \tilde{x}_2 yields

$$\begin{aligned} \ln(p_k(\tilde{x}_1)/p_K(\tilde{x}_1)) - \ln(p_k(\tilde{x}_2)/p_K(\tilde{x}_2)) \\ = \beta [\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_k(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)] \mathbf{m}. \end{aligned} \quad (6)$$

Provided that Magnac and Thesmar's rank condition

$$[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_k(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)] \mathbf{m} \neq 0 \quad (7)$$

holds, this linear (in β) equation uniquely determines β in terms of the data.

This identification argument can be interpreted in terms of an experiment that shifts the expected excess surplus contrast $[\mathbf{Q}_k(x) - \mathbf{Q}_K(x)] \mathbf{m}$ by changing the state x from \tilde{x}_2 to \tilde{x}_1 , while keeping the current value $U_k(\tilde{x}_1) = U_k(\tilde{x}_2)$ constant. The discount factor is the per unit effect of that observed shift on the observed log choice probability ratio $\ln(p_k(x)/p_K(x))$.

A shift in the expectation contrast $\mathbf{Q}_k(x) - \mathbf{Q}_K(x)$ does not suffice for identification. For example, suppose that the exclusion restriction holds for some $\tilde{x}_1, \tilde{x}_2 \in \mathcal{X}$, but that the excess surplus $m(x_1) = \dots = m(x_J)$ is constant, so that the expected excess surplus contrast $[\mathbf{Q}_k(x) - \mathbf{Q}_K(x)] \mathbf{m} = 0$. Then, a shift in the expectation

contrast does not shift the expected excess surplus contrast and hence does not change the decision problem. Consequently, this shift is not informative on β and [Magnac and Thesmar](#)'s rank condition (7) fails.

Rank condition (7) has a meaningful interpretation and is verifiable in data. The exclusion restriction $U_k(\tilde{x}_1) = U_k(\tilde{x}_2)$, however, is more problematic, because it imposes opaque conditions on the primitives that are hard to verify in applications. The current values depend on both current utilities and discounted expected future values. Specifically, they involve elements of \mathbf{v}_K , which by (4) equals

$$\mathbf{v}_K = [\mathbf{I} - \beta \mathbf{Q}_K]^{-1} [\mathbf{u}_K + \beta \mathbf{Q}_K \mathbf{m}]. \quad (8)$$

The current value is in fact a value contrast between two sequences of choices: choose k now, K in the next period, and choose optimally ever after, relative to choose K now, K in the next period, and choose optimally ever after. Because this particular value contrast does not correspond to common economic choice sequences, the applied value of [Magnac and Thesmar](#)'s restriction is limited. It is hard to think of naturally occurring experiments that shift the expected contrasts in excess surplus, i.e. satisfy the rank condition, without also shifting the current value and consequently violating the exclusion restriction, except for special cases. Indeed, the intuitive identification arguments in the introduction's empirical examples do not involve current values, but exclusion restrictions on primitive utility.

3.2 An exclusion restriction on primitive utility

Like [Magnac and Thesmar](#), we start with (5) or, equivalently,

$$\ln \mathbf{p}_k - \ln \mathbf{p}_K = \beta [\mathbf{Q}_k - \mathbf{Q}_K] [\mathbf{m} + \mathbf{v}_K] + \mathbf{u}_k - \mathbf{u}_K. \quad (9)$$

Instead of controlling the contribution of \mathbf{v}_K to the right hand side with an exclusion restriction on the current value, we exploit that it can be expressed in terms of the model primitives. Substituting (8) in (9) and rearranging gives

$$\ln \mathbf{p}_k - \ln \mathbf{p}_K = \beta [\mathbf{Q}_k - \mathbf{Q}_K] [\mathbf{I} - \beta \mathbf{Q}_K]^{-1} [\mathbf{m} + \mathbf{u}_K] + \mathbf{u}_k - \mathbf{u}_K. \quad (10)$$

Intuition from static discrete choice analysis and [Magnac and Thesmar](#)'s results for dynamic models suggest that, for identification, we need to fix utility in one reference alternative, say \mathbf{u}_K . Intuitively, choices only depend on, and thus inform about, utility contrasts. Thus, following e.g. [Fang and Wang](#) and [Bajari et al.](#)

(2015), we set $\mathbf{u}_K = \mathbf{0}$.⁵ This normalization cannot be refuted by data without further restrictions (see Section 4). Despite this lack of empirical content, it is not completely innocuous, as it may affect the model’s counterfactual predictions (see e.g. [Norets and Tang](#), Lemma 2, and [Kalouptsi et al.](#)). In the appendix, we demonstrate that our analysis applies without change to the case in which $u_K^*(x)$ is constant, but not necessarily zero, and can straightforwardly be extended to the case in which $u_K^*(x)$ is known up to a constant shift, but not necessarily constant. Thus, our analysis of the identification of the discount factor complements identification results for the reference utility u_K^* .⁶

Now suppose that we know the value of $u_k^*(\tilde{x}_1) - u_l^*(\tilde{x}_2)$ for some known choices $k \in \mathcal{D}/\{K\}$ and $l \in \mathcal{D}$ and known states $\tilde{x}_1 \in \mathcal{X}$ and $\tilde{x}_2 \in \mathcal{X}$; with either $k \neq l$, $\tilde{x}_1 \neq \tilde{x}_2$, or both. For expositional convenience only (see the appendix for the general case), we take this known value to be zero, and simply focus on the exclusion restriction

$$u_k^*(\tilde{x}_1) = u_l^*(\tilde{x}_2). \quad (11)$$

An advantage of this exclusion restriction over [Magnac and Thesmar](#)’s current value restriction is that it is a direct constraint on primitive utility with a clear economic interpretation. It also extends [Magnac and Thesmar](#) by allowing for restrictions on primitive utilities across combinations of choices and states.

3.3 The identified set

Under exclusion restriction (11), (10) implies

$$\begin{aligned} & \ln(p_k(\tilde{x}_1)/p_K(\tilde{x}_1)) - \ln(p_l(\tilde{x}_2)/p_K(\tilde{x}_2)) \\ &= \beta [\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_l(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)] [\mathbf{I} - \beta \mathbf{Q}_K]^{-1} \mathbf{m}. \end{aligned} \quad (12)$$

Like (6), this moment condition on β carries all the information in the data about the discount factor and can be used directly for its identification and estimation.⁷

⁵Note that this normalization does not collapse [Magnac and Thesmar](#)’s exclusion restriction on current values to an easily interpretable restriction on primitives.

⁶[Chou \(2015\)](#) recently provided identification results for dynamic discrete choice models without a normalization of u_K^* . [Chou](#)’s results for the stationary model that we study here take the discount factor to be known. [Chou](#)’s Propositions 3, 7, and 8 for a nonstationary model like the one we study in Section 3.6 provide high-level sufficient conditions for point identification, whereas we focus on set identification under intuitive conditions. A general difference is that we emphasize the economic interpretation of the identifying conditions and that we provide results on their empirical content.

⁷Obviously, any discount factor that is consistent with the data needs to solve (12); conversely, for any discount factor that does, primitive utilities can be found that satisfy the exclusion restric-

Unlike the right hand side of (6), the right hand side of (12) is not linear in β . Nevertheless, given data on transition and choice probabilities, it is a well-behaved, function of β . It is therefore easy to characterize the “identified set” \mathcal{B} of values of $\beta \in [0, 1)$ that equate it to the known left hand side of (12).

Theorem 1. Suppose that the exclusion restriction in (11) holds for some $k \in \mathcal{D}/\{K\}$, $l \in \mathcal{D}$, $\tilde{x}_1 \in \mathcal{X}$, and $\tilde{x}_2 \in \mathcal{X}$; with either $k \neq l$, $\tilde{x}_1 \neq \tilde{x}_2$, or both. Moreover, suppose that either the left hand side of (12) is nonzero (that is, $p_k(\tilde{x}_1)/p_K(\tilde{x}_1) \neq p_l(\tilde{x}_2)/p_K(\tilde{x}_2)$) or a generalization of [Magnac and Thesmar’s](#) rank condition (7) holds:

$$[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_l(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)] \mathbf{m} \neq 0. \quad (13)$$

Then, the identified set \mathcal{B} is a closed discrete subset of $[0, 1)$.

Proof. We need to show that, under the stated conditions, $\mathcal{B} \subseteq [0, 1)$ has no limit points in $[0, 1)$. First note that $[\mathbf{I} - \beta \mathbf{Q}_K]^{-1}$ exists for $\beta \in (-1, 1)$ and equals

$$\mathbf{I} + \beta \mathbf{Q}_K + \beta^2 \mathbf{Q}_K^2 + \cdots. \quad (14)$$

This is trivial for $\beta = 0$. If $|\beta| \in (0, 1)$, it follows from the facts that $|\beta^{-1}| > 1$ and that \mathbf{Q}_K is a Markov transition matrix, with eigenvalues no larger than 1 in absolute value. Consequently, the determinant of $\mathbf{Q}_K - \beta^{-1} \mathbf{I}$ is nonzero, so that $\mathbf{I} - \beta \mathbf{Q}_K$ is invertible and the power series in (14) converges.

It follows that, for given choice and transition probabilities, the right hand side of (12) minus its left hand side is a real-valued power series in β that converges on $(-1, 1)$. Denote the function of β this defines with $f : (-1, 1) \rightarrow \mathbb{R}$. Corollary 1.2.4 in [Krantz and Parks \(2002\)](#) ensures that f is real analytic.

Denote $\mathcal{B}^* = \{\beta \in (-1, 1) \mid f(\beta) = 0\}$. Note that $\mathcal{B} = \mathcal{B}^* \cap [0, 1)$. First, suppose that f has no zeros ($\mathcal{B}^* = \emptyset$). Then, $\mathcal{B} = \emptyset$ has no limit point in $[0, 1)$.

Finally, suppose that f has at least one zero ($\mathcal{B}^* \neq \emptyset$). Then, f cannot be constant (and thus equal zero) under the stated conditions: If the left hand side of (12) is nonzero then, because its right hand side equals zero at $\beta = 0$, $f(0)$ is nonzero; if rank condition (13) holds, then the derivative of the right hand side of (12) at $\beta = 0$, and therefore of f at 0, is nonzero. Because f is a nonconstant real-analytic function, its zero set \mathcal{B}^* has no limit point in $(-1, 1)$ ([Krantz and Parks, Corollary 1.2.7](#)). Because $\mathcal{B} = \mathcal{B}^* \cap [0, 1)$, this implies that \mathcal{B} has no limit point in $[0, 1)$. \square

tion used and that rationalize the data (see Section 4).

Under the conditions of Theorem 1, each $\beta \in [0, 1)$ that is consistent with (12) is an isolated point in $[0, 1)$ and thus locally identified. Note that $\beta = 1$ is excluded from the model to ensure convergence of the discounted utility flows. Theorem 1 does not exclude that 1 is a limit point of the identified set. So, the identified set may contain countably many discount factors near 1. However, because a closed discrete set is finite on compact subsets, only finitely many discount factors in the identified set lie outside a neighborhood of 1.

Corollary 1. Under the conditions of Theorem 1, $\mathcal{B} \cap [0, 1 - \epsilon]$ is finite for $0 < \epsilon < 1$. In many applications, one may be able to argue against discount factors that are arbitrarily close to 1. Corollary 1 shows that, in such applications, it suffices to search for the finite number of discount factors in a compact set $[0, 1 - \epsilon]$ that solve (12), which is computationally easy.

The right hand side of (12) is the log choice probability difference implied by the model with an exclusion restriction across choices k and l and states \tilde{x}_1 and \tilde{x}_2 . From the proof of Theorem 1, we know it equals the discount factor β , which represents how much the agent cares about the next period, multiplied by the sum of two terms that capture how much relevant variation in next period's expected discounted utility there is for the agent to care about:

$$[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_l(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)] \mathbf{m} \quad (15)$$

and

$$\begin{aligned} & [\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_l(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)] \mathbf{v}_K = \\ & [\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_l(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)] [\beta \mathbf{Q}_K + \beta^2 \mathbf{Q}_K^2 + \dots] \mathbf{m}. \end{aligned} \quad (16)$$

The first term (15) does not depend on β . It is nonzero if the generalized rank condition (13) holds. It corresponds to the leading, linear term in the right hand side of (12), which extends the right hand side of Magnac and Thesmar's (6) to the possibility of comparing across distinct choices k and l .

The next section gives conditions under which the second term (16) vanishes. Section 3.5 gives economic examples in which these conditions hold. If they hold, the right hand side of (12) is linear in β , so that (12) uniquely determines β under the generalized rank condition (13).

In general, the second term (16) does not vanish and depends on β . Then, the right hand side of (12) is not linear in β , but its derivative at $\beta = 0$ still equals the first term (15).⁸ Therefore, if the generalized rank condition (13) holds, this

⁸The derivative corresponding to the second term (16) vanishes because choice K has zero value

derivative is nonzero and myopic preferences ($\beta = 0$) are locally identified.⁹ In economic terms, the rank condition ensures that there is variation in next period's expected discounted utility for a myopic agent to care about, so that only myopic preferences can explain a lack of choice response. In Theorem 1, the rank condition excludes the trivial case that a zero choice response is observed and the right hand side of (12) equals zero for all β .

In the case that a zero choice response is observed, local identification of myopic preferences does not rule out that the data are also consistent with some positive discount factors, as there may be $\beta \in (0, 1)$ such that the sum of (15) and (16) is zero (that is, there is no variation in next period's expected discounted utility for the agent to care about). These discount factors, if any, can easily be found by searching for the solutions of (12). In particular, if the sum of (15) and (16) is nonzero for all $\beta \in (0, 1)$, only myopic preferences can explain the lack of choice response.

More generally, rank condition (13) does not suffice for point identification of β . As the next example demonstrates, the same observed choice response may arise from a combination of a low β (little care about the next period) and a large absolute sum of (15) and (16) (lots of variation in the next period to care about) and from a combination of a high β and a small absolute sum of (15) and (16).

Example 1. Figure 1 plots the left hand side of (6) and (12) (solid black line) and the right hand sides of (6) (dashed red line) and (12) (solid blue curve) for a specific example with $K = 2$ choices, $k = l = 1$, and $J = 3$ states. The example's data satisfy the rank condition in (13). Under the current value restriction, there is a unique discount factor that rationalizes the data (the intersection of the black and the dashed red curve). Under the primitive utility restriction, there are two discount factors that rationalize the same data (the intersections of the black and the solid blue curve).

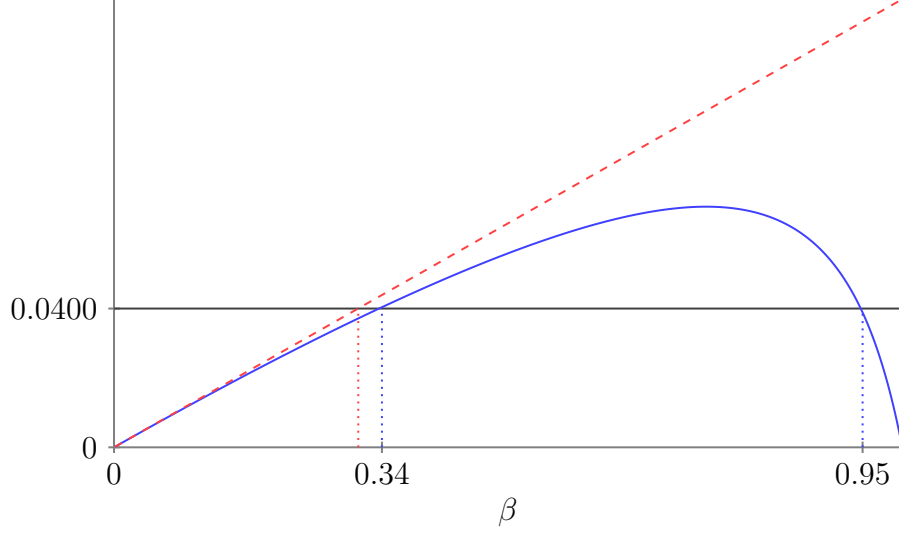
Finally, the rank condition in (13) is not necessary for point identification either.

Example 2. Figure 2 presents an example in which (15) equals zero, so that the right hand side of (6) and the first (excess surplus) term in the right hand side of (12) are zero, but the second (value of choice K) term in the right hand side of (12) is positive and increasing with β . There exists exactly one $\beta \in [0, 1)$ that solves (12), despite the violation of the rank condition.

if the agent is myopic.

⁹Here, β is locally identified at some β_0 if $\beta = \beta_0$ uniquely solves (12) in a neighborhood of β_0 . Rank condition (13) is not *necessary* for local identification of β at zero; for that, higher order variation of the right hand side of (12) in β at zero would suffice (Sargan, 1983).

Figure 1: Example in Which the Rank Condition Holds, but Identification Fails



Note: For $J = 3$ states, $K = 2$ choices, $k = l = 1$, $\tilde{x}_1 = x_1$, and $\tilde{x}_2 = x_2$, this graph plots the left hand side of (6) and (12) (solid black horizontal line) and the right hand sides of (6) (dashed red line), and (12) (solid blue curve) as functions of β . The data are $\mathbf{Q}_1(x_1) = [0.25 \ 0.25 \ 0.50]$, $\mathbf{Q}_1(x_2) = [0.00 \ 0.25 \ 0.75]$,

$$\mathbf{Q}_K = \begin{bmatrix} 0.90 & 0.00 & 0.10 \\ 0.00 & 0.90 & 0.10 \\ 0.00 & 1.00 & 0.00 \end{bmatrix}, \mathbf{p}_1 = \begin{bmatrix} 0.50 \\ 0.49 \\ 0.10 \end{bmatrix}, \text{ and } \mathbf{p}_K = \begin{bmatrix} 0.50 \\ 0.51 \\ 0.90 \end{bmatrix}.$$

Consequently, the left hand side of (6) and (12) equals $\ln(p_1(x_1)/p_K(x_1)) - \ln(p_1(x_2)/p_K(x_2)) = 0.0400$. Moreover, $\mathbf{m}' = [0.69 \ 0.67 \ 0.11]$ and $\mathbf{Q}_1(x_1) - \mathbf{Q}_K(x_1) - \mathbf{Q}_1(x_2) + \mathbf{Q}_K(x_2) = [-0.65 \ 0.90 \ -0.25]$, so that the slope of the dashed red line equals $[\mathbf{Q}_1(x_1) - \mathbf{Q}_K(x_1) - \mathbf{Q}_1(x_2) + \mathbf{Q}_K(x_2)] \mathbf{m} = 0.1291$. A unique value of β , 0.31, solves (6), but two values of β solve (12): 0.34 and 0.95.

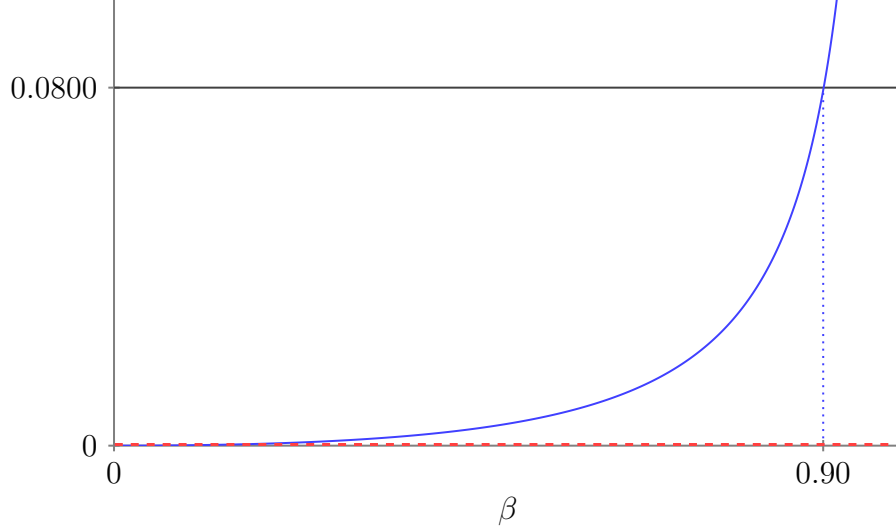
Also note that there is no value of β that satisfies (6). Even though the data can be rationalized by some specification of the model, they are not consistent with the current value restriction. In other words, this restriction has empirical content. We return to this point in Section 4.

3.4 Finite dependence

Some of the examples in the next subsection display a variant of Arcidiacono and Miller's (2011) "finite dependence". Finite dependence is a property of dynamic discrete choice models that can considerably simplify estimation and is widely used in applications (see Arcidiacono and Miller, 2015, for references).

In our context, finite dependence implies that the moment condition is of finite and known polynomial order. This order provides an upper bound on the number of solutions for the discount factor in \mathbb{R} , and therefore in $[0, 1)$. For example, in the

Figure 2: Example in Which the Rank Condition Fails, but the Discount Factor is Identified



Note: For $J = 3$ states, $K = 2$ choices, $k = l = 1$, $\tilde{x}_1 = x_1$, and $\tilde{x}_2 = x_2$, this graph plots the left hand side of (6) and (12) (solid black horizontal line) and the right hand sides of (6) (dashed red line) and (12) (solid blue curve) as functions of β . The data are $\mathbf{Q}_1(x_1) = [0.00 \ 0.25 \ 0.75]$, $\mathbf{Q}_1(x_2) = [0.25 \ 0.25 \ 0.50]$,

$$\mathbf{Q}_K = \begin{bmatrix} 0.00 & 1.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}, \mathbf{p}_1 = \begin{bmatrix} 0.50 \\ 0.48 \\ 0.50 \end{bmatrix}, \text{ and } \mathbf{p}_K = \begin{bmatrix} 0.50 \\ 0.52 \\ 0.50 \end{bmatrix}.$$

Consequently, the left hand side of (6) and (12) equals $\ln(p_1(x_1)/p_K(x_1)) - \ln(p_1(x_2)/p_K(x_2)) = 0.0800$. Moreover, $\mathbf{m}' = [0.69 \ 0.65 \ 0.69]$ and $\mathbf{Q}_1(x_1) - \mathbf{Q}_K(x_1) - \mathbf{Q}_1(x_2) + \mathbf{Q}_K(x_2) = [-0.25 \ 0.00 \ 0.25]$, so that the slope of the dashed red line equals $[\mathbf{Q}_1(x_1) - \mathbf{Q}_K(x_1) - \mathbf{Q}_1(x_2) + \mathbf{Q}_K(x_2)] \mathbf{m} = 0.0000$. A unique value of β , 0.90, solves (12), but (6) has no solution.

case with $k \neq l = K$, (16) reduces to

$$[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1)] \mathbf{v}_K = [\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1)] [\beta \mathbf{Q}_K + \beta^2 \mathbf{Q}_K^2 + \dots] \mathbf{m}. \quad (17)$$

Suppose that $\mathbf{Q}_k(\tilde{x}_1) \mathbf{Q}_K^\rho = \mathbf{Q}_K(\tilde{x}_1) \mathbf{Q}_K^\rho$ for some $\rho \in \{1, 2, \dots\}$. That is, the distribution of the state $\rho + 1$ periods from now does not depend on whether the agent chooses k or K now, provided that she follows up in both cases by choosing K in the next ρ periods (independently of whether this is optimal or not). Under this “single action (K) ρ -period dependence” (Arcidiacono and Miller, 2017) on choices k and K in state \tilde{x}_1 , $\mathbf{Q}_k(\tilde{x}_1) \mathbf{Q}_K^r = \mathbf{Q}_K(\tilde{x}_1) \mathbf{Q}_K^r$ for all $r \in \{\rho, \rho + 1, \dots\}$.¹⁰ Now assume that Theorem 1’s conditions hold. If $\rho = 1$, the right hand side of (17) equals zero,

¹⁰Throughout, we focus on this special case of Arcidiacono and Miller’s (2011) finite dependence, which turns out to be particularly powerful in our specific context.

the current value

$$U_k(\tilde{x}_1) = u_k^*(\tilde{x}_1) + \beta [\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1)] \mathbf{v}_K = u_k^*(\tilde{x}_1),$$

the right hand side of (12) is linear in β , and β is point identified. If instead $\rho \geq 2$, then the right hand side of (17) equals

$$[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1)] [\beta \mathbf{Q}_K + \cdots + \beta^{\rho-1} \mathbf{Q}_K^{\rho-1}] \mathbf{m},$$

the right hand side of (12) is a ρ -th order polynomial in β , and the identified set \mathcal{B} holds no more than ρ discount factors. This example straightforwardly extends to the general exclusion restriction in (11), which we state without further proof.

Theorem 2. Suppose that the conditions of Theorem 1 hold and that $\{\mathbf{Q}_k; k \in \mathcal{D}\}$ satisfies single action (K) ρ -period dependence on choices k and K in state \tilde{x}_1 ,

$$\mathbf{Q}_k(\tilde{x}_1) \mathbf{Q}_K^\rho = \mathbf{Q}_K(\tilde{x}_1) \mathbf{Q}_K^\rho,$$

and single action (K) ρ -period dependence on choices l and K in state \tilde{x}_2 ,

$$\mathbf{Q}_l(\tilde{x}_2) \mathbf{Q}_K^\rho = \mathbf{Q}_K(\tilde{x}_2) \mathbf{Q}_K^\rho,$$

for some $\rho \in \{1, 2, \dots\}$. Then there are no more than ρ points in the identified set \mathcal{B} .

Theorem 2 applies finite dependence to cancel differences in expected discounted utilities across pairs of choices twice, once for each of the two states that appear in the exclusion restriction. In the special case that the exclusion restriction concerns a comparison across states \tilde{x}_1 and \tilde{x}_2 for a given choice $k = l$, the right hand side of (16) reduces to

$$[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_k(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)] [\beta \mathbf{Q}_K + \beta^2 \mathbf{Q}_K^2 + \cdots] \mathbf{m}. \quad (18)$$

By Theorem 2, single action (K) ρ -period dependence on choices k and K in states \tilde{x}_1 and \tilde{x}_2 implies that the identified set contains at most ρ discount factors. If $\rho = 1$, then both $U_k(\tilde{x}_1) = u^*(\tilde{x}_1)$ and $U_k(\tilde{x}_2) = u^*(\tilde{x}_2)$, (18) equals 0, and the discount factor is point identified.

In this case with $k = l$, the consequent of Theorem 2 would also hold if, alternatively,

$$\mathbf{Q}_k(\tilde{x}_1) \mathbf{Q}_K^\rho = \mathbf{Q}_k(\tilde{x}_2) \mathbf{Q}_K^\rho \quad \text{and} \quad \mathbf{Q}_K(\tilde{x}_1) \mathbf{Q}_K^\rho = \mathbf{Q}_K(\tilde{x}_2) \mathbf{Q}_K^\rho,$$

for some $\rho \in \{1, 2, \dots\}$. This is a form of single action (K) ρ -period dependence on the initial state (instead of the initial choice) under, respectively, choices k and K . Under one-period dependence on initial states \tilde{x}_1 and \tilde{x}_2 , current values do not necessarily reduce to primitive utilities, but it is still true that $U_k(\tilde{x}_1) - U_k(\tilde{x}_2) = u_k^*(\tilde{x}_1) - u_k^*(\tilde{x}_2)$, (18) equals 0, and the discount factor is point identified.

3.5 Examples

Theorem 1 shows that the identified set of discount factors is discrete and, away from one, finite, but does not establish point identification. Indeed, Example 1 demonstrated that point identification may fail, even if rank condition (13) holds.

Our first two examples below (Examples 3 and 4) illustrate applications from the literature in which the exclusion restriction is plausibly met, but the discount factor is not necessarily point identified. We then give two examples (Examples 5 and 6) of optimal stopping problems with single action one-period dependence, which are point identified by Theorem 2. Finally, Example 7 demonstrates that single action one-period dependence is not necessary for point identification. It is a labor supply model that does not exhibit such one-period dependence, but in which monotonicity of the moment condition in the discount factor gives point identification.

A number of empirical studies of demand for health care insured under Medicare Part D base their identification on nonlinearities in the price schedules. Our first example describes an empirical strategy from this literature in which the primitive utility exclusion restriction seems plausibly met.

Example 3. As part of an informal identification argument, [Finkelstein et al. \(2015\)](#) observed that changes in purchase behaviour around kinks in the price of insurance are informative on time preferences. In the data, insurees pay 25% of additional expenditures out of pocket as long as their total yearly expenditures range between \$275 and \$2510, but contribute 100% to all expenditures between \$2510 and \$5726. A myopic insuree would change her spending only after her total spending hits \$2510 and her out-of-pocket contributions increase. In contrast, a forward looking insuree who is close to the kink late in the year would limit her spending before hitting the increase in contributions. Changes in the propensity to spend towards the end of the year for those close to the kink are therefore taken to be informative on the discount factor.

This argument can be represented as an exclusion restriction. Let x be the yearly expenditure, a state controlled by the choice of filling prescriptions. The utility $u_k(x)$ of a particular drug purchase k is assumed constant for two expenditure levels

$\tilde{x}_1 < \tilde{x}_2$ in $[275, 2510)$. Along with variation in expected future expenses because of the kinked price schedule, the exclusion restriction gives set identification by Theorem 1.¹¹

The next example is from Rossi (2018) which studied the effect of reward programs on gasoline sales using a dynamic discrete choice model.

Example 4. In each period, a consumer can choose to buy gasoline ($k = 1$), or not ($k = K$). Consumers accumulate reward points x by registering their gasoline purchases. The accumulated points can be traded against non-pecuniary rewards at various point thresholds \underline{x} . Rossi observed that the purchase frequency is accelerating in the accumulated points. By assuming that the current period payoff of a gasoline purchase at any price \tilde{y} is independent of the accumulated points, i.e. that $u_k(\tilde{y}, \tilde{x}_1) = u_k(\tilde{y}, \tilde{x}_2)$ for all \tilde{y} and $\tilde{x}_1, \tilde{x}_2 \in [0, \underline{x})$, the purchase acceleration is informative on the discount factor. The closer the consumer is to qualify for a given reward, the less the future reward is discounted. This makes a current purchase more attractive and predicts a purchase frequency that is increasing in the reward points.

We next turn to optimal stopping problems. The first example is the bus engine replacement problem of Rust (1987). Though the plausibility of the exclusion restriction is questionable in this particular application, it illustrates how one-period dependence gives point identification in a well-known application of an optimal stopping model.

Example 5. Rust (1987) studied Harold Zurcher’s management of a fleet of (independent) buses. In each period, Zurcher can either operate a bus as usual ($d = 1$) or renew its engine ($d = K = 2$). The payoff from operating the bus as usual depend on its mileage x since last renewal, which both Zurcher and Rust (1987) observe, and an additive and independent shock. Renewal incurs a cost that is independent of mileage and resets mileage to $x_1 = 0$:

$$\mathbf{Q}_K = \begin{bmatrix} 1 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{bmatrix}.$$

In terms of Section 3.4’s finite dependence, mileage is single action (K) one-period dependent on both initial mileage and the initial renewal choice. Consequently,

¹¹Dalton et al. (2015) used a similar argument to identify salience and myopia in a dynamic discrete choice model with parametric utility applied to Medicare Part D data.

Zurcher's expected discounted payoffs from renewal do not depend on mileage. In particular, with our normalization $\mathbf{u}_K = \mathbf{0}$, $v_K^*(\tilde{x}) = \beta(m(x_1) + v_K^*(x_1))$ for all $\tilde{x} \in \mathcal{X}$. Since $v_K^*(\tilde{x})$ does not vary with \tilde{x} , $[\mathbf{Q}_1 - \mathbf{Q}_K]\mathbf{v}_K = \mathbf{0}$, and $U_1(\tilde{x}) = u_1^*(\tilde{x})$. Therefore, if we assume $u_1^*(\tilde{x}_1) = u_1^*(\tilde{x}_2)$, which may be questionable in this application, [Magnac and Thesmar](#)'s exclusion restriction holds and its identification result applies.¹² Its rank condition (7) simplifies to

$$[\mathbf{Q}_1(\tilde{x}_1) - \mathbf{Q}_1(\tilde{x}_2)]\mathbf{m} \neq \mathbf{0}.$$

That is, it simply requires that the expected next period's excess surplus differs between states \tilde{x}_1 and \tilde{x}_2 under continued operation of the bus (choice 1).

Example 5's analysis of optimal renewal extends to optimal stopping problems in which stopping ends the decision problem. For example, in [Hopenhayn's \(1992\)](#) model of firm dynamics with free entry, active firms solve optimal stopping problems in which they value exit K at $\mathbf{v}_K = \mathbf{0}$. As in Example 5, the fact that $v_K^*(\tilde{x})$ is constant in \tilde{x} ensures that the expectation contrast $[\mathbf{Q}_1 - \mathbf{Q}_K]\mathbf{v}_K = \mathbf{0}$, so that $U_1(\tilde{x}) = u_1^*(\tilde{x})$.

Of course, $[\mathbf{Q}_1 - \mathbf{Q}_K]\mathbf{v}_K$ may equal zero even if $v_K^*(\tilde{x})$ varies with \tilde{x} , in particular if the state is single action (K) one-period dependent on choices 1 and K .

Example 6. Consider a discrete time, econometric implementation of [Dixit's \(1989\)](#) model of firm entry and exit.¹³ In each period, a firm chooses to either serve the market ($d = 1$) or not ($d = K = 2$). Its payoffs from serving the market depend on $x = (y, d_{-1})$, where y is a profit shifter that follows an exogenous Markov process (that is, y may affect choices but is not controlled by them) and d_{-1} is the firm's choice in the previous period. The entry costs in profit state \tilde{y} equal the difference between an incumbent's profit from serving the market and a new entrant's profit from doing so, $u_1^*(\tilde{y}, 1) - u_1^*(\tilde{y}, K)$, which we assume to be nonnegative. As before, we set $\mathbf{u}_K = \mathbf{0}$, so that the exit costs $u_K^*(\tilde{y}, K) - u_K^*(\tilde{y}, 1)$ are zero.

The firm's value $v_K^*(y', k)$ from choosing inactivity (K) next period after choosing $d = k$ now may vary with next period's profit state y' , because the firm will have the option to reenter the market and this option's value may depend on y' . However, because exit costs are zero, this value does not depend on the current choice k : $v_K^*(y', 1) = v_K^*(y', K)$. Moreover, by the assumption that y follows an exogenous

¹²Since mileage is the only observed state variable in this application, $u_1^*(\tilde{x}_1) = u_1^*(\tilde{x}_2)$ requires that the current payoffs from operating a bus are the same at \tilde{x}_1 and \tilde{x}_2 miles, for example because \tilde{x}_1 and \tilde{x}_2 lie on a known flat segment of Harold Zurcher's cost curve.

¹³[Abbring and Klein \(2015\)](#) presented this example's model with state independent entry costs, code for its estimation, and exercises that can be used in teaching dynamic discrete choice models.

Markov process, the distribution of y' given $(y, d_{-1}, d = k)$ is independent of the current choice k and the past choice d_{-1} , so that

$$\mathbf{Q}_1(\tilde{x})\mathbf{v}_K = \mathbb{E}[v_K^*(y', 1) | y = \tilde{y}] = \mathbb{E}[v_K^*(y', K) | y = \tilde{y}] = \mathbf{Q}_K(\tilde{x})\mathbf{v}_K \quad (19)$$

for all $\tilde{x} = (\tilde{y}, \tilde{d}_{-1}) \in \mathcal{X}$. Consequently, as in Example 5, $[\mathbf{Q}_1(\tilde{x}) - \mathbf{Q}_K(\tilde{x})]\mathbf{v}_K = 0$ and $U_1(\tilde{x}) = u_1^*(\tilde{x})$. Note that, in this case, the state $x = (y, d_{-1})$ is single action (K) one-period dependent on choices 1 and K (but generally not on initial states).

An exclusion restriction $u_1^*(\tilde{x}_1) = u_1^*(\tilde{x}_2)$ implies (6) and, under rank condition (7), point identification of β . Because y evolves independently of current and past choices,

$$\mathbf{Q}_k(\tilde{x})\mathbf{m} = \mathbb{E}[m(y', k) | y = \tilde{y}]. \quad (20)$$

Thus, the rank condition is equivalent to

$$\mathbb{E}[m(y', 1) - m(y', K) | y = \tilde{y}_1] \neq \mathbb{E}[m(y', 1) - m(y', K) | y = \tilde{y}_2]. \quad (21)$$

It immediately follows that identification requires that $\tilde{y}_1 \neq \tilde{y}_2$ in this case. A difference in lagged choices alone would not suffice, because these do not help predict next period's profit state y' given the current profit state y and choice $d = k$ nor directly affect next period's excess surplus.

Moreover, identification fails if entry costs are zero; that is, if $u_1^*(\tilde{y}, 1) = u_1^*(\tilde{y}, K)$. In this case, payoffs do not depend on past choices and, more specifically, $m(y', 1) = m(y', K)$. Intuitively, without entry and exit costs, firms can ignore past and future when deciding on entry and exit and simply maximize the current profits in each period. Consequently, their entry and exit choices carry no information on their discount factor. As an aside, note that the entry costs are directly identified from

$$\ln(p_1(\tilde{x}_1)/p_K(\tilde{x}_1)) - \ln(p_1(\tilde{x}_2)/p_K(\tilde{x}_2)) = u_1^*(\tilde{y}, 1) - u_1^*(\tilde{y}, K)$$

for $\tilde{x}_1 = (\tilde{y}, 1)$ and $\tilde{x}_2 = (\tilde{y}, K)$. Intuitively, for given profit state \tilde{y} , lagged choices only affect current payoffs through the entry costs and have no effect on expected future payoffs, as is clear from (19) and (20).

Finally, if both $\tilde{y}_1 \neq \tilde{y}_2$ and entry costs are strictly positive, (21) will generally be satisfied. In specific applications, we can verify (21) using that both the distribution of y' conditional on y and $m(y', k) = -\ln(p_K(y', k))$ can directly be estimated from choice and profit state transition data.

As in Zurcher's problem, profit states are typically ordered, so that an exclusion

restriction like $u_1^*(\tilde{x}_1) = u_1^*(\tilde{x}_2)$ may be justified as a local shape restriction on the firm's utility function. Alternatively, because the firm's utility is a cardinal payoff, we may be able to exploit that $u_1^*(\tilde{x})$ is known in some state \tilde{x} . For example, if $u_1^*(\tilde{x}) = 0$, then (12) holds with $k = 1$, $l = K$, and $\tilde{x}_1 = \tilde{x}_2 = \tilde{x} = (\tilde{y}, \tilde{d}_{-1})$ and reduces to

$$\ln(p_1(\tilde{x})) - \ln(p_K(\tilde{x})) = \beta \mathbb{E}[m(y', 1) - m(y', K) \mid y = \tilde{y}],$$

so that β is identified if $\mathbb{E}[m(y', 1) - m(y', K) \mid y = \tilde{y}] \neq 0$. This rank condition is generally satisfied if entry costs are positive.

In Examples 5 and 6, the rank condition ensures that the shift in expected surplus contrasts that multiplies β in the right hand side of (12) is nonzero. Because these examples satisfy one-period dependence, this shift does not depend on β itself, and this suffices for point identification. More generally, even if the state is not one-period dependent, strict monotonicity of the right hand side of (12), as in Example 2, suffices for point identification (that is, ensures that a solution is unique if it exists). It is easy to derive conditions that imply such strict monotonicity, and thus point identification, and that do not involve β . Without loss of generality— we can freely interchange states \tilde{x}_1 and \tilde{x}_2 and switch choices k and l — we focus on conditions under which it is strictly increasing or, equivalently, its derivative with respect to β is positive:¹⁴

$$[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_l(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)] [\mathbf{I} - \beta \mathbf{Q}_K]^{-2} \mathbf{m} > 0.$$

For this, it suffices that

$$[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_l(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)] \mathbf{Q}_K^r \mathbf{m} \geq 0 \text{ for all } r \in \{0, 1, 2, \dots\}, \quad (22)$$

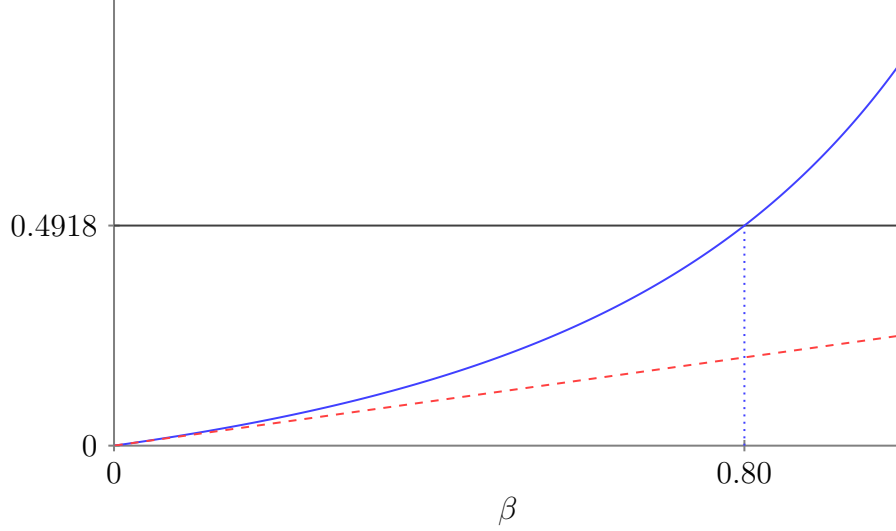
with the inequality strict for at least one r . Like Magnac and Thesmar's rank condition (7), these conditions do not depend on β . It is easy to verify that they hold in Example 2 (which is specified in the Note to Figure 2).

The final example relies on a type of payoff monotonicity that is common in

¹⁴Denoting $\Delta^2 \mathbf{Q} \equiv \mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_l(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)$, we have that

$$\begin{aligned} \frac{\partial}{\partial \beta} \beta \Delta^2 \mathbf{Q} [\mathbf{I} - \beta \mathbf{Q}_K]^{-1} \mathbf{m} &= \Delta^2 \mathbf{Q} [\mathbf{I} - \beta \mathbf{Q}_K]^{-1} \mathbf{m} + \beta \Delta^2 \mathbf{Q} \frac{\partial [\mathbf{I} - \beta \mathbf{Q}_K]^{-1}}{\partial \beta} \mathbf{m} \\ &= \Delta^2 \mathbf{Q} [\mathbf{I} - \beta \mathbf{Q}_K]^{-1} \mathbf{m} + \beta \Delta^2 \mathbf{Q} [\mathbf{I} - \beta \mathbf{Q}_K]^{-1} \mathbf{Q}_K [\mathbf{I} - \beta \mathbf{Q}_K]^{-1} \mathbf{m} \\ &= \Delta^2 \mathbf{Q} [\mathbf{I} + (\mathbf{I} - \beta \mathbf{Q}_K)^{-1} \beta \mathbf{Q}_K] [\mathbf{I} - \beta \mathbf{Q}_K]^{-1} \mathbf{m} = \Delta^2 \mathbf{Q} [\mathbf{I} - \beta \mathbf{Q}_K]^{-2} \mathbf{m}. \end{aligned}$$

Figure 3: Example of a Dynamic Labor Supply Model that Gives a Monotone Moment Condition



Note: For $J = 3$ states, $K = 2$ choices, $k = l = 1$, $\tilde{x}_1 = x_2$, and $\tilde{x}_2 = x_1$, this graph plots the left hand side of (6) and (12) (solid black horizontal line) and the right hand sides of (6) (dashed red line) and (12) (solid blue curve) as functions of β (we switched the roles of x_1 and x_2 to ensure a positive choice response and visually line up this example with the others). The data are generated from Example 7's stylized dynamic labor supply model, which gives $\mathbf{Q}_1(x_2) = [0.00 \ 0.25 \ 0.75]$, $\mathbf{Q}_1(x_1) = [0.25 \ 0.75 \ 0.00]$,

$$\mathbf{Q}_K = \begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.50 & 0.50 & 0.00 \\ 0.00 & 0.50 & 0.50 \end{bmatrix}, \mathbf{p}_1 = \begin{bmatrix} 0.44 \\ 0.56 \\ 0.71 \end{bmatrix}, \text{ and } \mathbf{p}_K = \begin{bmatrix} 0.56 \\ 0.44 \\ 0.29 \end{bmatrix}.$$

Consequently, the left hand side of (6) and (12) equals $\ln(p_1(x_2)/p_K(x_2)) - \ln(p_1(x_1)/p_K(x_1)) = 0.4918$. Moreover, $\mathbf{m}' = [0.57 \ 0.82 \ 1.23]$ and $\mathbf{Q}_1(x_2) - \mathbf{Q}_K(x_2) - \mathbf{Q}_1(x_1) + \mathbf{Q}_K(x_1) = [0.25 \ -1.00 \ 0.75]$, so that the slope of the dashed red line equals $[\mathbf{Q}_1(x_2) - \mathbf{Q}_K(x_2) - \mathbf{Q}_1(x_1) + \mathbf{Q}_K(x_1)] \mathbf{m} = 0.2465$. A unique value of β , 0.80, solves (12), but (6) has no solution.

models with ordered states.

Example 7. In [Eckstein and Wolpin's \(1989\)](#) dynamic model of female labor force participation, women work both to directly earn wages and to invest in work experience that pays off later. Consider a highly stylized and stationary variant of this model. Each period, a woman either works ($d = 1$) or shirks ($d = 2 = K$). Work experience takes three levels, “novice” (x_1), “learning” (x_2), and “seasoned” (x_3). If a woman works and is not yet seasoned, her experience increases one level with probability 0.75 and stays the same with the complementary probability. If instead she shirks, and is not a novice, she falls back one level of experience with probability 0.50 and keeps her experience otherwise. Work gives utility $u_1(x_1) = u_1(x_2) = -0.50$ if novice or learning and $u_1(x_3) = 0.50$ if seasoned. Women maximize their flow of

expected utility, discounted with a factor 0.80.

Figure 3 gives the data implied by this example and plots the moment condition corresponding to the constraint that $u_1(x_1) = u_1(x_2)$. This constraint implies that novices and learning workers earn the same current utility. Nevertheless, work is more attractive to a learning woman, because she has a good shot at earning the higher wage for seasoned workers next period if she works now; moreover, unlike a novice, she may lose experience if she shirks. Seasoned workers, despite the fact that they cannot further increase their experience, are sufficiently motivated by the higher earnings and the risk that their human capital depreciates to work even more. Consequently, $p_K(x_1) > p_K(x_2) > p_K(x_3)$, so that $m(x_1) < m(x_2) < m(x_3)$. More generally, because \mathbf{Q}_K is increasing, the expected excess surplus after r rounds of shirking and human capital depreciation, $\mathbf{Q}_K^r \mathbf{m}$, is increasing in initial experience.

In this example, the dependence of (the distribution of) a worker's experience on initial choices and experience levels does not disappear in a finite number of periods of, e.g., shirking.¹⁵ In particular, experience is not single action (K) one-period dependent on initial choices in states x_1 and x_2 and Magnac and Thesmar's current value restriction and linear moment condition do not hold. Nevertheless, this example's monotonicity ensures that the discount factor is point identified. Because working, compared to shirking, affects the experience of a learning worker more than that of a novice with nothing to lose ($[\mathbf{Q}_1(x_2) - \mathbf{Q}_K(x_2) - \mathbf{Q}_1(x_1) + \mathbf{Q}_K(x_1)] = [0.25 \ -1.00 \ 0.75]$) and $\mathbf{Q}_K^r \mathbf{m}$ is increasing for all r , (22) holds. Consequently, the moment condition is monotone in β and has only one solution, 0.80.

3.6 Extension to nonstationary models

Our analysis extends to nonstationary models, such as that in Keane and Wolpin (1997), with minor modifications. In fact, nonstationary models offer useful identification strategies that are not available for stationary models. Unlike in stationary models, an assumption of stationary utilities has identifying power in nonstationary models. A common version of this argument is that the utilities can be identified in the last period, say T , so that the discount factor is subsequently identified in the next to last period (e.g. Yao et al., 2012). This argument assumes stationary utilities, which can be cast as an exclusion restriction on time as a state variable,

¹⁵In a similar context, Altuğ and Miller (1998) impose such finite dependence by assuming that wages and the utility cost from work only depend on a finite employment history. Our example would display single action (K) one-period dependence on initial choices in state x_1 and two-period dependence in state x_2 if shirking women would for sure see their experience drop by one level. Note that this would still not suffice to reduce the moment condition to Magnac and Thesmar's linear moment condition.

i.e. $u_{i,T-1}(\tilde{x}) = u_{i,T}(\tilde{x})$, where time shifts the continuation values without shifting the primitive utilities.

[Bajari et al. \(2016\)](#) used the assumption of stationary utilities to formally establish identification in a finite-horizon optimal stopping model. Theorem 3 below extends [Bajari et al.](#)'s result beyond optimal stopping problems and also allows for identification of models with nonstationary utilities.¹⁶

Denote time by $t \in \{1, 2, \dots, T\}$, with terminal period $T < \infty$, and index $u_{k,t}^*$, $\mathbf{u}_{k,t}$, \mathbf{m}_t , and $\mathbf{v}_{k,t}$ by time. For ease of exposition, we maintain the assumption of stationary Markov transition matrices \mathbf{Q}_k , but the results extend to nonstationary distributions. The choice- k specific values now satisfy

$$\mathbf{v}_{k,t} = \mathbf{u}_{k,t} + \beta \mathbf{Q}_k [\mathbf{m}_{t+1} + \mathbf{v}_{K,t+1}] \quad (23)$$

for $t = 1, \dots, T-1$; with terminal condition $\mathbf{v}_{k,T} = \mathbf{u}_{k,T}$. With the normalization $\mathbf{u}_{K,t} = \mathbf{0}$ for all t , this gives

$$\ln(p_{k,t}(\tilde{x})) - \ln(p_{K,t}(\tilde{x})) = u_{k,t}^*(\tilde{x}) + \beta[\mathbf{Q}_k(\tilde{x}) - \mathbf{Q}_K(\tilde{x})][\mathbf{m}_{t+1} + \mathbf{v}_{K,t+1}] \quad (24)$$

for all $k \in \mathcal{D} \setminus \{K\}$ and $\tilde{x} \in \mathcal{X}$. Finally, using (23) and the normalization $\mathbf{u}_{K,t} = \mathbf{0}$ for all t , we can write the value of the reference choice K as

$$\mathbf{v}_{K,t} = \sum_{\tau=t+1}^T (\beta \mathbf{Q}_K)^{\tau-t} \mathbf{m}_\tau, \quad (25)$$

where we use the convention that $\sum_{\tau=T+1}^T \cdot = 0$ (so that indeed $\mathbf{v}_{K,T} = \mathbf{u}_{K,T} = \mathbf{0}$).

Theorem 3. Suppose that

$$u_{k,t}^*(\tilde{x}_1) = u_{l,t'}^*(\tilde{x}_2) \quad (26)$$

for $k \in \mathcal{D} \setminus \{K\}$, $l \in \mathcal{D}$, $\tilde{x}_1 \in \mathcal{X}$, $\tilde{x}_2 \in \mathcal{X}$, $1 \leq t' < T$, and $t' \leq t \leq T$; with either $k \neq l$, or $\tilde{x}_1 \neq \tilde{x}_2$, or $t' < t$, or a combination of the three. If either $p_{k,t}(\tilde{x}_1)/p_{K,t}(\tilde{x}_1) \neq p_{l,t'}(\tilde{x}_2)/p_{K,t'}(\tilde{x}_2)$ or

$$[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1)]\mathbf{m}_{t+1} - [\mathbf{Q}_l(\tilde{x}_2) - \mathbf{Q}_K(\tilde{x}_2)]\mathbf{m}_{t'+1} \neq \mathbf{0}, \quad (27)$$

then there are no more than $T - t'$ points in the identified set.

¹⁶[Yao et al.](#) showed identification of the discount factor in a dynamic model with continuous controls under the assumption of stationary utilities and conjectured a similar result for discrete controls. Theorem 3 proves its conjecture.

Proof. Differencing (24) corresponding to (26) and substituting in (25) gives

$$\begin{aligned} \ln(p_{k,t}(\tilde{x}_1)/p_{K,t}(\tilde{x}_1)) - \ln(p_{l,t'}(\tilde{x}_2)/p_{K,t'}(\tilde{x}_2)) = \\ \beta \left([\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1)] \left[\sum_{\tau=t+1}^T (\beta \mathbf{Q}_K)^{\tau-t-1} \mathbf{m}_\tau \right] - \right. \\ \left. [\mathbf{Q}_l(\tilde{x}_2) - \mathbf{Q}_K(\tilde{x}_2)] \left[\sum_{\tau=t'+1}^T (\beta \mathbf{Q}_K)^{\tau-t'-1} \mathbf{m}_\tau \right] \right). \end{aligned} \quad (28)$$

For given choice and transition probabilities, the right hand side of (28) minus its left hand side is a polynomial of order $T - t'$ in β . If this polynomial is nonconstant, then by the fundamental theorem of algebra, it has up to $T - t'$ real roots, which is an upper bound on the number of points in the identified set. To show that (28) is nonconstant under the stated assumptions, note first that the right hand side of (28) is zero at $\beta = 0$. If the left hand side is nonzero, the polynomial is nonconstant. If the left hand side is zero, then the rank condition (27) ensures that the derivative of the right hand side is nonzero at $\beta = 0$, so that the right hand side, and thus the polynomial, is nonconstant. \square

Rank condition (27) adapts (13) to the nonstationary case. Unlike the stationary dynamic choice problem, the nonstationary problem does not require that the discount factor lies in $[0, 1)$. We leave the definition of the domain of the discount factor to the reader.

In a study of identification in nonstationary models, Arcidiacono and Miller (2017) distinguished between identification in long panels, which include the terminal period, and short panels, which do not. In general, Theorem 3 requires long panels. However, for models with ρ -period dependence, it also applies to short panels that extend to at least period $t + \rho$. For instance, in Zurcher's renewal problem with a finite horizon, mileage is still single action (K) one-period dependent, so that the discount factor can be point identified in short panels until period $t + 1$.

4 Empirical content

The previous section focused on identification and gave conditions under which the primitives can be recovered from the data. In applications, we need to entertain the possibility that the model is misspecified and did not generate the data to begin with. It is well known that the unrestricted model has no empirical content: It can rationalize any choice data $\{\mathbf{p}_k, \mathbf{Q}_k; k \in \mathcal{D}\}$. This section shows that the model under exclusion restrictions can be rejected by data.

The standard result for the unrestricted stationary model follows from a version of [Magnac and Thesmar](#)'s Proposition 2: For any given data $\{\mathbf{p}_k, \mathbf{Q}_k; k \in \mathcal{D}\}$, $\mathbf{u}_K = \mathbf{0}$, and $\beta \in [0, 1)$, there exists a unique set of primitive utilities $\{\mathbf{u}_k, k \in \mathcal{D}/\{K\}\}$ that rationalizes the data. Specifically, $\mathbf{m} = -\ln \mathbf{p}_K$. Then, \mathbf{v}_K follows from $\mathbf{u}_K = \mathbf{0}$ and (8). Next, by (3), $\mathbf{v}_k = \mathbf{v}_K + \ln \mathbf{p}_k - \ln \mathbf{p}_K$ for $k \in \mathcal{D}/\{K\}$ ensures that the value functions are compatible with the choice probability data. In turn, by (4), these value functions are uniquely generated by the primitive utilities $\mathbf{u}_k = \mathbf{v}_k - \beta \mathbf{Q}_k [\mathbf{m} + \mathbf{v}_K]$ for $k \in \mathcal{D}/\{K\}$ (note that \mathbf{v}_K was already set to be consistent with $\mathbf{u}_K = \mathbf{0}$).

This result justifies our focus on the identification of the discount factor β in the previous section: Once the discount factor is identified, we can find unique primitive utilities that rationalize the data. The empirical consequences of a violation of the assumed exclusion restriction can manifest themselves in two distinct ways.

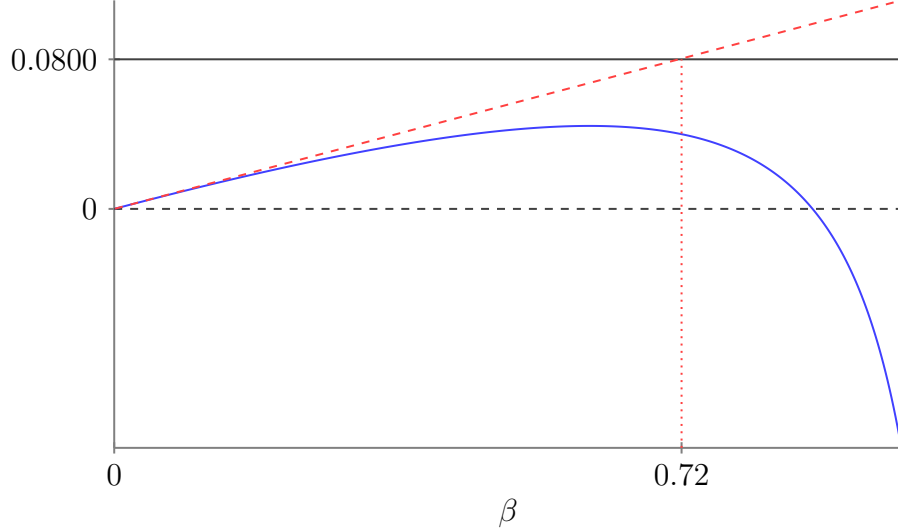
First, in some cases, it may be possible to find primitives that satisfy the false exclusion restriction. If so, these primitives will in general not equal the true primitives. Because we can find primitive utilities that rationalize the data for any discount factor, the data can be of no help in determining the right restriction in this case. Instead, we need to argue for the identifying assumption on other grounds.

Second, there may not exist discount factors in their domain that are compatible with the data under the assumed exclusion restriction. The subset of the possible data that can be rationalized under an exclusion restriction can be very small. For instance, in a binary choice model with $J = 2$ and $u_1^*(x_1) = u_1^*(x_2)$, the model cannot generate any state-dependent value contrasts. It follows that this model cannot rationalize any state-dependent choice data. In empirical practice, this may force parameter estimates to lie outside their theoretical domains. In turn, this may lead researchers to statistically reject the model and conclude that at least one of its assumptions is violated. While some solution methods, such as typical nested fixed point algorithms, impose the restriction that $\beta \in [0, 1)$, it is easy to use the moment conditions in (12) for model testing as their computation do not restrict the values β can take.

The empirical content of the identified model also gives some scope to test nonnested identifying assumptions against each other. For example, the data in Example 2 cannot be rationalized under [Magnac and Thesmar](#)'s current value restriction, but are consistent with an exclusion restriction on primitive utility. Conversely, it is easy to construct data that are inconsistent with the primitive utility restriction, yet can be rationalized by primitives that satisfy the current value restriction.

Example 8. Figure 4 displays the left and right hand sides of [Magnac and Thesmar](#)'s

Figure 4: Example of Data that are Consistent with an Exclusion Restriction on Current Values but Not with One on Primitive Utility



Note: For $J = 3$ states, $K = 2$ choices, $k = l = 1$, $\tilde{x}_1 = x_1$, and $\tilde{x}_2 = x_2$, this graph plots the left hand side of (6) and (12) (solid black horizontal line) and the right hand sides of (6) (dashed red line) and (12) (solid blue curve) as functions of β . The data are $\mathbf{Q}_1(\tilde{x}_1) = [0.25 \ 0.25 \ 0.50]$, $\mathbf{Q}_1(\tilde{x}_2) = [0.00 \ 0.25 \ 0.75]$,

$$\mathbf{Q}_K = \begin{bmatrix} 0.90 & 0.00 & 0.10 \\ 0.00 & 0.90 & 0.10 \\ 0.00 & 1.00 & 0.00 \end{bmatrix}, \mathbf{p}_1 = \begin{bmatrix} 0.50 \\ 0.48 \\ 0.10 \end{bmatrix}, \text{ and } \mathbf{p}_K = \begin{bmatrix} 0.50 \\ 0.52 \\ 0.90 \end{bmatrix}.$$

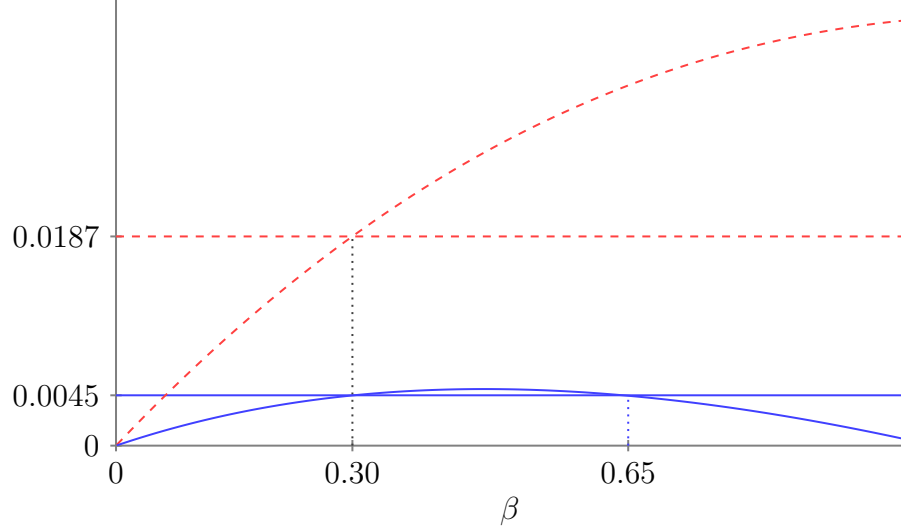
Consequently, the left hand side of (6) and (12) equals $\ln(p_1(x_1)/p_K(x_1)) - \ln(p_1(x_2)/p_K(x_2)) = 0.0800$. Moreover, $\mathbf{m}' = [0.69 \ 0.65 \ 0.11]$ and $\mathbf{Q}_1(x_1) - \mathbf{Q}_K(x_1) - \mathbf{Q}_1(x_2) + \mathbf{Q}_K(x_2) = [-0.65 \ 0.90 \ -0.25]$, so that the slope of the dashed red line equals $[\mathbf{Q}_1(x_1) - \mathbf{Q}_K(x_1) - \mathbf{Q}_1(x_2) + \mathbf{Q}_K(x_2)] \mathbf{m} = 0.1116$. A unique value of β , 0.72, solves (6), but (12) has no solution.

moment condition in (6) and ours in (12). There is a $\beta \in [0, 1)$ that solves (6), but the moment condition in (12) cannot be met. Intuitively, the increasingly negative contribution of the second (value of choice K) term in the right hand side of (12) limits the possible log choice probability ratio response to the change in states to a level below the observed response.

In practice, we can easily establish whether given data are consistent with one exclusion restriction or the other by verifying whether the corresponding moment condition, (6) or (12), or its empirical analog has a solution $\beta \in [0, 1)$.

Finally, note that the empirical content of the nonstationary model depends on the chosen domain of the discount factor. Therefore, we limit our discussion of this model's empirical content to noting that Theorem 3 does not guarantee a real root (and less so one in a specified domain for β) for general choice and state probabilities.

Figure 5: Example with Two Moment Conditions of Which One Identifies the Discount Factor



Note: For $J = 4$ states, $K = 2$ choices, and $k = l = 1$, this graph plots the left (horizontal lines) and right hand sides (curves) of (12) as functions of β , for $\tilde{x}_1 = x_1$ and $\tilde{x}_2 = x_2$ (corresponding to $u_1(x_1) = u_1(x_2)$; dashed red) and $\tilde{x}_1 = x_3$ and $\tilde{x}_2 = x_4$ (corresponding to $u_1(x_3) = u_1(x_4)$; solid blue). The data are

$$\mathbf{Q}_1 = \begin{bmatrix} 0.40 & 0.26 & 0.18 & 0.18 \\ 0.33 & 0.29 & 0.36 & 0.27 \\ 0.19 & 0.26 & 0.18 & 0.45 \\ 0.08 & 0.18 & 0.29 & 0.09 \end{bmatrix}, \mathbf{Q}_K = \begin{bmatrix} 0.17 & 0.26 & 0.13 & 0.43 \\ 0.13 & 0.07 & 0.20 & 0.60 \\ 0.20 & 0.30 & 0.10 & 0.40 \\ 0.25 & 0.15 & 0.50 & 0.10 \end{bmatrix},$$

$$\mathbf{p}'_1 = [0.60 \quad 0.59 \quad 0.88 \quad 0.88], \text{ and } \mathbf{p}'_K = [0.40 \quad 0.41 \quad 0.12 \quad 0.12].$$

Consequently, the left hand sides of (12) equal $\ln(p_1(x_1)/p_K(x_1)) - \ln(p_1(x_2)/p_K(x_2)) = 0.0187$ and $\ln(p_1(x_3)/p_K(x_3)) - \ln(p_1(x_4)/p_K(x_4)) = 0.0045$. A unique value of β , 0.30, solves (12) for $\tilde{x}_1 = x_1$ and $\tilde{x}_2 = x_2$ (dashed red). Two values of β solve (12) for $\tilde{x}_1 = x_3$ and $\tilde{x}_2 = x_4$ (solid blue), of which one coincides with the solution to the first moment condition.

5 Multiple exclusion restrictions and inference

Often, more than one exclusion restriction is available. In particular, economic intuition for an exclusion restriction across states typically suggests the exclusion of a state *variable* from the utility function. For example, the state variable x can be partitioned as (y, z) , where z does not affect utilities: $u_k(\tilde{y}, \tilde{z}_1) = u_k(\tilde{y}, \tilde{z}_2)$ for all $k \in \mathcal{D}/\{K\}$, \tilde{y} , \tilde{z}_1 , and $\tilde{z}_2 > \tilde{z}_1$.¹⁷ This typically gives multiple exclusion restrictions like (11). For example, if choices, y , and z are all binary, we have two exclusion restrictions, one for each possible value of y .

With multiple exclusion restrictions, point identification can be obtained even if

¹⁷We provide a more formal statement of the exclusion of state variables in our discussion of Fang and Wang in Abbring and Daljord (2019).

each individual moment condition set identifies β . We give two examples of identification with two exclusion restrictions.

Example 9. In Figure 5, the moment condition represented by the solid blue line and curve and the one in red dashes have two and one solutions, respectively. Both moment conditions are consistent with a discount factor of 0.30, while the solid moment condition is also consistent with a discount factor of 0.65. The dashed moment condition by itself point identifies the discount factor, while the solid moment condition only set identifies it. In this case, the solid moment condition is redundant for point identification.

Example 10. In Figure 6, the dashed red moment condition holds for discount factors 0.17 and 0.90, while the solid blue moment condition is solved by discount factors 0.07 and 0.90. Each individual moment condition is consistent with two discount factors, but only one discount factor solves both moment conditions.

With choice and transition probabilities generated from a model that satisfies two (or more) exclusion restrictions, the implied two (or more) moment conditions will always share one solution, the discount factor that was used to generate the data. We conjecture that, generically, the moments will not share any further solutions, because different choice and transition probabilities, which vary freely with the primitive utilities, enter the various moment conditions.

Generic point identification is of limited practical value in our context. First, we are not able to a priori characterize the subset of the model space on which point identification fails in terms of economic concepts. Though this subset is small, it may, for all we know, contain economically important models.¹⁸

Second, we may not learn whether the discount factor is point or set identified in finite samples. While finding the shared solutions to multiple moment conditions is easy if we know the population choice and transition probabilities, locating the shared solutions in finite samples can be difficult due to sampling variation. This suggests that we do not insist on point identification, but accept set identification and use a consistent estimator of the identified set, which may contain one or more points. Set estimators are easy to implement for single parameter problems. We give one example.

Example 11. Suppose the population moment conditions are as given in Figure 6. Though each individual moment condition is equally consistent with one small

¹⁸For example, Ekeland et al.'s (2004) generic identification result for the hedonic model is particularly instructive because it shows that identification fails exactly for the linear-quadratic special case that is at the center of most applied work.

discount factor, at 0.07 and 0.17, respectively, and one large discount factor at the true value of 0.90, only the latter is a common solution to both moment conditions. The discount factor is therefore point identified in this population.

In the top panel of Figure 7, the same two moment conditions are plotted with sampling variation in the choice data. One sample moment condition is solved by discount factors 0.16 and 0.91 and the other by discount factors 0.25 and 0.68. The data do not clearly reveal that the point-identified true discount factor is 0.90. If anything, the data suggest point identification in the lower region. Even if point identification cannot be determined a priori without further assumptions, the discount factor is still set identified and we can use consistent set estimators.

Following [Chernozhukov et al. \(2007\)](#) and [Romano and Shaikh \(2010\)](#), suppose that the identified set $\mathcal{B} = \{\beta \in [0, 1) : S(\beta) = 0\}$ for some population criterion function $S : [0, 1) \rightarrow [0, \infty)$. Note that we can alternatively write $\mathcal{B} = \arg \min_{\beta \in [0, 1)} S(\beta)$. This suggests that we estimate \mathcal{B} by a random contour set $\mathcal{C}_n(s) = \{\beta \in [0, 1) : a_n S_n(\beta) \leq s\}$ for some level $s > 0$ and normalizing sequence $\{a_n\}$, where $S_n(\beta)$ is the sample equivalent of $S(\beta)$ and n is the sample size. For a given confidence level $\alpha \in (0, 1)$, s is set to equal a consistent estimator s_n of the α -quantile of $\sup_{\beta \in \mathcal{B}} a_n S_n(\beta)$, so that the estimator $\mathcal{C}_n(s_n)$ asymptotically contains the identified set with probability α :

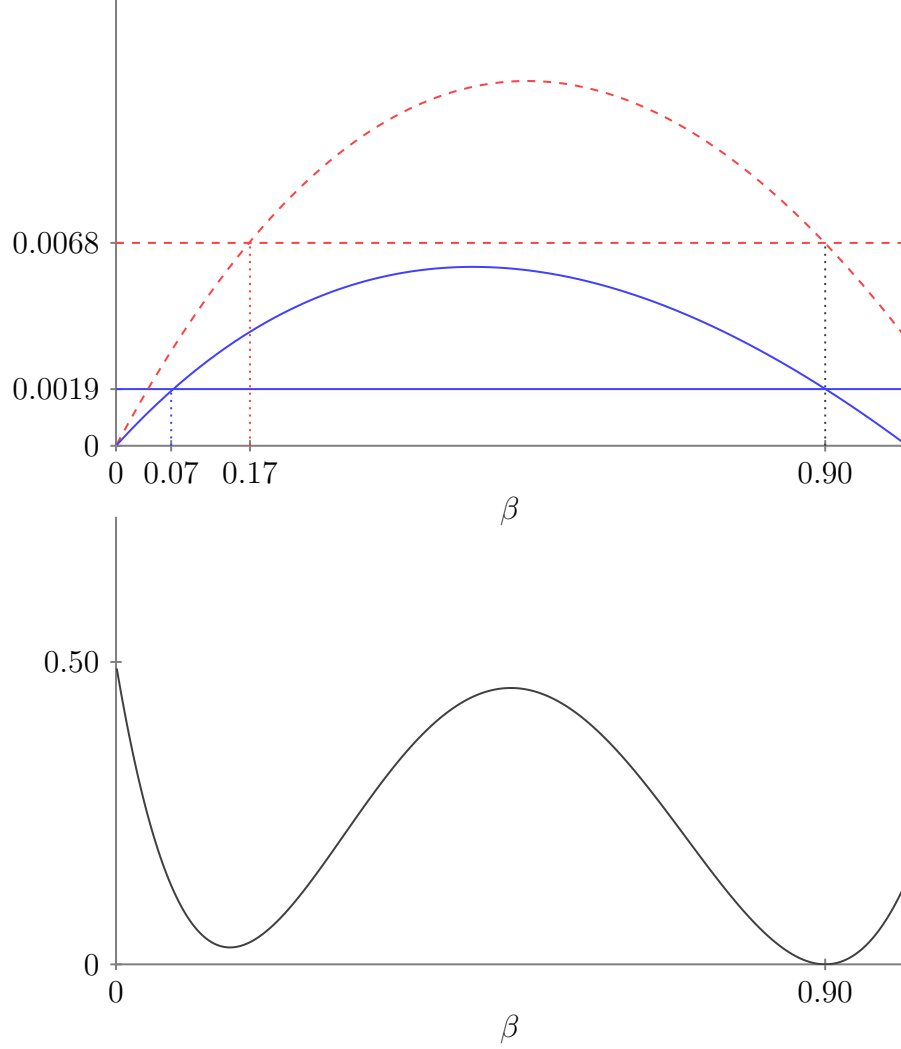
$$\lim_{n \rightarrow \infty} \Pr\{\mathcal{B} \subseteq \mathcal{C}_n(s_n)\} = \alpha.$$

The bottom panel of Figure 7 illustrates one such estimator. The criterion $S_n(\beta)$ is here a quadratic form in the difference between the left and right hand sides of (12) evaluated at consistent estimators of the choice and transition probabilities using equal weights. The critical value s_n is given as the horizontal line. The estimated set is $\mathcal{C}_n(s_n) = [0.10, 0.28] \cup [0.79, 0.91]$. The data are equally consistent with a range of small discount factors and a range of large discount factors, but an intermediate range (0.28, 0.79) is rejected at the α -level, along with discount factors smaller than 0.10 and larger than 0.91.

Under some regularity conditions, the set estimator converges to the identified set as the sample size grows. Since the identified set is a point in this example, in the limit, the subset of $\mathcal{C}_n(s_n)$ with small discount factors vanishes and its subset with large discount factors degenerates to the population discount factor 0.90. While these set estimators are computationally demanding for parameter spaces with even just a handful of dimensions, they are easy to implement in a one-dimensional case such as ours.

For point identified problems, standard inference for extremum estimators applies (e.g. [Newey and McFadden, 1994](#)).

Figure 6: Example with Two Moment Conditions that Jointly Identify the Discount Factor but Individually Do Not



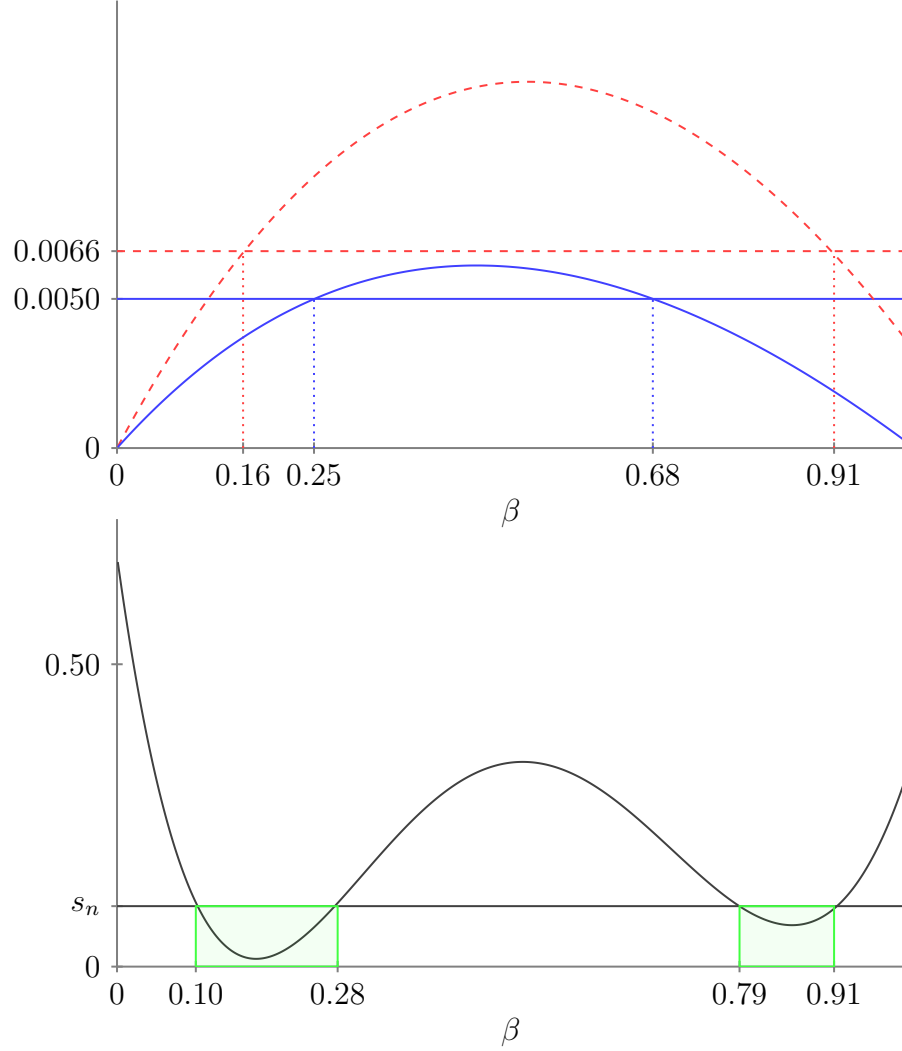
Note: For $J = 4$ states, $K = 2$ choices, and $k = l = 1$, the graph in the top panel plots the left (horizontal lines) and right hand sides (curves) of (12) as functions of β , for $\tilde{x}_1 = x_1$ and $\tilde{x}_2 = x_2$ (corresponding to $u_1(x_1) = u_1(x_2)$; dashed red) and $\tilde{x}_1 = x_3$ and $\tilde{x}_2 = x_4$ (corresponding to $u_1(x_3) = u_1(x_4)$; solid blue). The graph in the bottom panel plots the corresponding squared Euclidian distance between the left and right hand sides of (12) as a function of β (in multiples of 10^{-4}). The data are

$$\mathbf{Q}_1 = \begin{bmatrix} 0.43 & 0.26 & 0.18 & 0.18 \\ 0.33 & 0.29 & 0.36 & 0.27 \\ 0.19 & 0.26 & 0.18 & 0.45 \\ 0.05 & 0.18 & 0.29 & 0.09 \end{bmatrix}, \quad \mathbf{Q}_K = \begin{bmatrix} 0.17 & 0.26 & 0.13 & 0.43 \\ 0.13 & 0.07 & 0.20 & 0.60 \\ 0.20 & 0.30 & 0.10 & 0.40 \\ 0.25 & 0.15 & 0.50 & 0.10 \end{bmatrix},$$

$$\mathbf{p}'_1 = [0.92 \quad 0.92 \quad 0.63 \quad 0.63], \text{ and } \mathbf{p}'_K = [0.08 \quad 0.08 \quad 0.37 \quad 0.37].$$

Consequently, the left hand sides of (12) equal $\ln(p_1(x_1)/p_K(x_1)) - \ln(p_1(x_2)/p_K(x_2)) = 0.0068$ and $\ln(p_1(x_3)/p_K(x_3)) - \ln(p_1(x_4)/p_K(x_4)) = 0.0019$. A unique value of β , 0.90, solves (12) for both $\tilde{x}_1 = x_1$ and $\tilde{x}_2 = x_2$ (dashed red) and $\tilde{x}_1 = x_3$ and $\tilde{x}_2 = x_4$ (solid blue). In addition, each of these two moment conditions has one other solution.

Figure 7: Example with Two Moment Conditions that Jointly Identify the Discount Factor but Individually Do Not, Using Noisy Estimates of the Choice Probabilities



Note: This figure redraws Figure 6 for the same values of \mathbf{Q}_1 and \mathbf{Q}_K , but randomly perturbed values of its choice probabilities \mathbf{p}_1 and \mathbf{p}_K . Rounded to two digits, the perturbed choice probabilities equal those reported below Figure 6. Consequently, the perturbation to $\mathbf{m} = -\ln \mathbf{p}_K$ is very small too, so that the right hand sides of (12) are very close to those plotted in Figure 6. The left hand sides of (12), however, now equal $\ln(p_1(x_1)/p_K(x_1)) - \ln(p_1(x_2)/p_K(x_2)) = 0.0066$ (instead of 0.0068) and $\ln(p_1(x_3)/p_K(x_3)) - \ln(p_1(x_4)/p_K(x_4)) = 0.0050$ (instead of 0.0019). The resulting moment conditions again have two solutions. However, they no longer share a common solution and the squared Euclidian distance in the bottom panel never attains zero. The green shaded areas highlight the intervals $[0.10, 0.28]$ and $[0.79, 0.91]$ of values of β at which the distance is below some critical level s_n (which is taken to be 0.10×10^{-4} in this example).

Appendix

Identification with general reference utility

Consider the stationary model of Section 2. Suppose that we know \mathbf{u}_K up to a constant additive shift; that is, $\mathbf{u}_K = \gamma \mathbf{1} + \bar{\mathbf{u}}_K$, with $\gamma \in \mathbb{R}$ unknown, $\mathbf{1}$ the $J \times 1$ vector of ones, and $\bar{\mathbf{u}}_K$ a known $J \times 1$ vector with j -th element $\bar{u}_K(x_j)$. Then, we can rewrite (10) as

$$\ln \mathbf{p}_k - \ln \mathbf{p}_K = \beta [\mathbf{Q}_k - \mathbf{Q}_K] [\mathbf{I} - \beta \mathbf{Q}_K]^{-1} (\mathbf{m} + \bar{\mathbf{u}}_K) + \mathbf{u}_k - \gamma \mathbf{1} - \bar{\mathbf{u}}_K. \quad (29)$$

Note that the constant additive shift $\gamma \mathbf{1}$ drops from the first term, which is a difference in expectations under choices k and K .

Now suppose that $u_k^*(\tilde{x}_1) - u_l^*(\tilde{x}_2)$ is known, but not necessarily zero, for some known choices $k \in \mathcal{D}/\{K\}$ and $l \in \mathcal{D}$, and known states $\tilde{x}_1 \in \mathcal{X}$ and $\tilde{x}_2 \in \mathcal{X}$; with either $k \neq l$, $\tilde{x}_1 \neq \tilde{x}_2$, or both. This is an exclusion restriction that encompasses (11) in the main text as a special case. Under this generalized exclusion restriction, (29) implies

$$\begin{aligned} \ln(p_k(\tilde{x}_1)/p_K(\tilde{x}_1)) - \ln(p_l(\tilde{x}_2)/p_K(\tilde{x}_2)) - \Delta^2 u \\ = \beta [\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_l(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)] [\mathbf{I} - \beta \mathbf{Q}_K]^{-1} \bar{\mathbf{m}}, \end{aligned} \quad (30)$$

with $\Delta^2 u \equiv u_k^*(\tilde{x}_1) - u_l^*(\tilde{x}_2) - \bar{u}_K(\tilde{x}_1) + \bar{u}_K(\tilde{x}_2)$ and $\bar{\mathbf{m}} \equiv \mathbf{m} + \bar{\mathbf{u}}_K$ known. The factor multiplying β in the right hand side of (12) can again be interpreted in terms of incentives related to differences in expected future utilities, which now include the known utilities derived from the reference choice K . Multiplying these “incentives” by the discount factor β gives the log choice probability response, corrected for the known effects of the current utility contrast $\Delta^2 u$, in the left hand side of (12).

The analysis of the main text applies to this generalization with straightforward adaptations. In particular, (30) is a moment condition in only one unknown, the discount factor β , and can be taken directly to data. The following generalization of Theorem 1 can be proved like that theorem.

Theorem 4. Suppose that $u_k^*(\tilde{x}_1) - u_l^*(\tilde{x}_2)$ is known for some $k \in \mathcal{D}/\{K\}$, $l \in \mathcal{D}$, $\tilde{x}_1 \in \mathcal{X}$, and $\tilde{x}_2 \in \mathcal{X}$; with either $k \neq l$, $\tilde{x}_1 \neq \tilde{x}_2$, or both. Moreover, suppose that either the left hand side of (30) is nonzero (that is, $p_k(\tilde{x}_1)/p_K(\tilde{x}_1) - p_l(\tilde{x}_2)/p_K(\tilde{x}_2) \neq \Delta^2 u$) or a generalization of Magnac and Thesmar’s rank condition (7) holds:

$$[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_l(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)] \bar{\mathbf{m}} \neq 0.$$

Then, the identified set \mathcal{B} is a closed discrete subset of $[0, 1)$.

A version of Corollary 1 follows directly and so do the simplifications that arise from finite dependence, in particular those that arise in renewal and optimal stopping problems. Finally, it is easy to adapt the analysis in this appendix to the nonstationary case. We will not pursue that here.

This appendix (in particular, a comparison of moment conditions (12) and (30)) demonstrates that the analysis in the main text extends

- without change to the case in which $u_K^*(x)$ equals a (not necessarily zero or even known) constant;
- with a simple, known adjustment to the choice probability response in the left hand side of (12) to the case that $u_k^*(\tilde{x}_1) - u_l^*(\tilde{x}_2)$ is known, but not necessarily zero; and
- with another such adjustment to the left hand side of (12) *and* a known adjustment to the polynomial in the right hand side of (12) if u_K^* is only known up to a constant additive shift, but not necessarily constant.

This shows that our analysis can directly be applied to problems in which a state independent reference utility exists (as is typically assumed in applied work) and directly complements results on the identification of more general reference utility specifications.

References

- Abbring, J. H. (2010). Identification of dynamic discrete choice models. *Annual Review of Economics* 2, 367–394.
- Abbring, J. H. and Ø. Daljord (2019). Comment on “Estimating dynamic discrete choice models with hyperbolic discounting” by Hanming Fang and Yang Wang. Working paper, Tilburg University and University of Chicago. [arXiv:1905.07048](https://arxiv.org/abs/1905.07048) [econ.EM].
- Abbring, J. H., Ø. Daljord, and F. Iskhakov (2018). Identifying present-biased discount functions in dynamic discrete choice models. Mimeo, CentER, Tilburg.
- Abbring, J. H. and T. J. Klein (2015, November). Dynamic discrete choice models: Methods, Matlab code, and exercises. Lecture notes, CentER, Tilburg. Available at ddc.abbring.org.
- Altuğ, S. and R. A. Miller (1998). The effect of work experience on female wages and labour supply. *Review of Economic Studies* 65(1), 45–85.
- Arcidiacono, P. and R. Miller (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica* 79, 1823–1867.
- Arcidiacono, P. and R. Miller (2015). Nonstationary dynamic models with finite dependence. Mimeo, Duke University.
- Arcidiacono, P. and R. Miller (2017). Identifying dynamic discrete choice models of short panels. Mimeo, Duke University.
- Bajari, P., V. Chernozhukov, H. Hong, and D. Nekipelov (2015). Identification and efficient semiparametric estimation of a dynamic discrete game. Working Paper 21125, National Bureau of Economic Research, Cambridge, MA.
- Bajari, P., C. S. Chu, D. Nekipelov, and M. Park (2016). Identification and semi-parametric estimation of a finite horizon dynamic discrete choice model with a terminating action. *Quantitative Marketing and Economics* 14(4), 271–323.
- Bayer, P., R. McMillan, A. Murphy, and C. Timmins (2016, May). A dynamic model of demand for houses and neighborhoods. *Econometrica* 84(3), 893–942.

- Bollinger, B. (2015). Green technology adoption: An empirical study of the Southern California garment cleaning industry. *Quantitative Marketing and Economics* 13, 319–358.
- Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75(5), 1243–1284.
- Chevalier, J. and A. Goolsbee (2009, November). Are Durable Goods Consumers Forward-Looking? Evidence from College Textbooks*. *The Quarterly Journal of Economics* 124(4), 1853–1884.
- Chou, C. (2015, October). Identification and linear estimation of general dynamic programming discrete choice models. Mimeo, University of Southern California, Los Angeles, CA.
- Chung, D. J., T. Steenburgh, and K. Sudhir (2014). Do bonuses enhance sales productivity? a dynamic structural analysis of bonus-based compensation plans. *Marketing Science* 33(2), 165–187.
- Dalton, C. M., G. Gowrisankaran, and R. Town (2015, April). Salience, myopia, and complex dynamic incentives: Evidence from medicare part d. Working Paper 21104, National Bureau of Economic Research.
- De Groote, O. and F. Verboven (2018, May). Subsidies and myopia in technology adoption: Evidence from solar photovoltaic systems. Technical report, Forthcoming AER.
- Dixit, A. K. (1989). Entry and exit decisions under uncertainty. *Journal of Political Economy* 97(3), 620–638.
- Eckstein, Z. and K. I. Wolpin (1989). Dynamic labour force participation of married women and endogenous work experience. *Review of Economic Studies* 56(3), 375–390.
- Ekeland, I., J. J. Heckman, and L. Nesheim (2004). Identification and estimation of hedonic models. *Journal of Political Economy* 112(1), S60–S109.
- 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.

- Finkelstein, A., L. Einav, and P. Schrimpf (2015, 02). The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D *. *The Quarterly Journal of Economics* 130(2), 841–899.
- Frederick, S., G. Loewenstein, and T. O’Donoghue (2002, June). Time discounting and time preference: A critical review. *Journal of Economic Literature* 40, 351–401.
- Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica* 60(5), 1127–1150.
- 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.
- Kalouptsi, M., P. T. Scott, and E. Souza-Rodrigues (2016, June). Identification of counterfactuals in dynamic discrete choice models. Mimeo, Princeton University.
- Keane, M. P. and K. I. Wolpin (1997, June). The career decisions of young men. *Journal of Political Economy* 115(3), 473–522.
- Komarova, T., F. Sanches, D. Silva, and S. Srisuma (2017). Joint analysis of the discount factor and payoff parameters in dynamic discrete choice models. Mimeo, London School of Economics.
- Krantz, S. G. and H. R. Parks (2002). *A Primer of Real Analytic Functions* (Second ed.). Boston: Birkhäuser Verlag.
- Lee, R. (2013). Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review* 103(7), 2960–3000.
- Magnac, T. and D. Thesmar (2002). Identifying dynamic discrete choice processes. *Econometrica* 70, 801–816.
- Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 36, pp. 2111–2245. Amsterdam: North-Holland.
- Norets, A. and X. Tang (2014). Semiparametric inference in dynamic binary choice models. *Review of Economic Studies* 81(3), 1229–1262.
- Romano, J. and A. Shaikh (2010). Inference for the identified set in partially identified econometric models. *Econometrica* 78(1), 169–211.

- Rossi, F. (2018). Lower price or higher reward? measuring the effect of consumers preferences on reward programs. *Management Science* 64(9), 4451–4470.
- Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica* 55, 999–1033.
- Rust, J. (1994). Structural estimation of Markov decision processes. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, pp. 3081–3143. Amsterdam: North-Holland.
- Sargan, J. D. (1983). Identification and lack of identification. *Econometrica* 51(6), 1605–1633.
- Yao, S., C. F. Mela, J. Chiang, and Y. Chen (2012). Determining consumers discount rates with field studies. *Journal of Marketing Research* 49(6), 822–841.