BECKER FRIEDMAN INSTITUTE
FOR RESEARCH IN ECONOMICS
THE UNIVERSITY OF CHICAGO

Working Paper Series
No. 2017-17

Identifying the Discount Factor in Dynamic Discrete Choice Models

Jaap H. Abbring, Øystein Daljord

September 2017

Becker Friedman Institute
for Research in Economics

Contact:
773.702.5599
bfi@uchicago.edu
bfi.uchicago.edu

Identifying the Discount Factor in Dynamic Discrete Choice Models*

Jaap H. Abbring[†] Øystein Daljord[‡]

September 26, 2017

Abstract

Empirical applications of dynamic discrete choice models usually either take the discount factor to be known or rely on high level exclusion restrictions that are difficult to interpret and hard to satisfy. We provide identification results under an intuitively appealing exclusion restriction on primitive utilities that is more directly useful in applied research. We show that while such exclusion restrictions are not sufficient for point identification, they identify the discount factor up to a finite set. The identified set is characterized as the solutions to a single, well-behaved moment condition. We also show that our and existing exclusion restrictions limit the choice and state transition probability data in different ways; that is, they give the model nontrivial and distinct empirical content.

*We thank Áureo de Paula, Jean-Pierre Dubé, Christian Hansen, Günter Hitsch, John-Eric Humphries, 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, and seminars at Tilburg, Toronto, Copenhagen, Chicago, UCL, LSE, Penn State, DIW Berlin, Georgetown, EIEF Rome, Cambridge, and Warwick 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. 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

The 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). Common intuition suggests that the discount factor can be identified from observed choice responses to variation that shifts expected discounted future utilities, but not the current utilities (e.g. Lee, 2013; Ching et al., 2014; Bayer et al., 2016). This intuition can be formalized as an exclusion restriction on primitive utilities. We show that, in contrast to the common intuition, these exclusion restrictions do not generally give point identification. They do however narrow down the identified set— the set of observationally equivalent discount factors— to a discrete and, if we exclude values near one in the infinite horizon case, finite set that is easy to compute.

Magnac and Thesmar's Proposition 4 established point identification based on a related, but different exclusion restriction: the existence of a pair of states that affects, in some specific way, expected future values, but not the "current value," which is a difference in expected discounted utilities between two particular choice sequences. Work on identification in stationary models, e.g. Norets and Tang (2014) and Bajari et al. (2015), usually takes the discount factor to be known, perhaps because Magnac and Thesmar's exclusion restriction is difficult to interpret and hard to verify in applications. We discuss these limitations of Magnac and Thesmar's result and show how more intuitively appealing identifying assumptions generally come at the cost of more complicated conditions for identification and inference.

The standard dynamic discrete choice model, without exclusion restrictions, cannot be falsified using data on state transition distributions and choice probabilities (Magnac and Thesmar, 2002, Proposition 2). In that sense, it has no empirical content. We show that our and Magnac and Thesmar's exclusion restrictions have nontrivial and distinct empirical contents. We give an example of data that falsify the model under one restriction, but can be rationalized under the other restriction. This implies that there is some scope for testing both restrictions using data on state transition distributions and choice probabilities.

After introducing the model in Section 2 and discussing Magnac and Thesmar's result in Section 3, we develop our main identification result for stationary models in Section 4. We then analyze the empirical content of Magnac and Thesmar's and our exclusion restrictions in Section 5 and discuss set identification and estimation with multiple exclusion restrictions in Section 6. In Section 7, we show that point identification of the discount factor does not follow as a special case of Fang and Wang's (2015) generic identification results for dynamic discrete choice models with partially naive hyperbolic time preferences. We end in Section 8 by extending our analysis to nonstationary models, which are commonly used in labor applications (e.g. Keane and Wolpin, 1997).

2 Model

Consider a stationary version of Magnac and Thesmar's model. Time is discrete with an infinite horizon. In each period, agents first observe state variables x and ε, where x takes discrete values in $\mathcal{X} = \{x_1, \ldots, x_J\}$ and $\varepsilon = \{\varepsilon_1, \ldots, \varepsilon_K\}$ is continuously distributed on $\mathbb{R}^K$; for $J, K \geq 2$. Then, they choose d from the set of alternatives $\mathcal{D} = \{1, 2, \ldots, K\}$ 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. Following Magnac and Thesmar, 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 $\varepsilon_1, \ldots, \varepsilon_K$ are independently drawn from mean zero type-1 extreme value distributions.[1] 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 $u_k(x, \varepsilon)$ and conditional independence imply that $v_k(x, \varepsilon) = v_k^*(x) + \varepsilon_k$, with v_k^* the unique solution to

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

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

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

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

[1] 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 $\mathbb{R}^K$.

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 5, there exist unique (up to a standard utility normalization) values of the primitives that rationalize the data for any given discount factor $\beta \in [0, 1)$. Thus, we can and will focus our identification analysis on β.

The choice probabilities are fully determined by

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

Thus, with 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 $\tilde{x} \in \mathcal{X}$ 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

$$v_k^*(x) = u_k^*(x) + \beta \int \left(m(x') + v_K^*(x')\right) dQ_k(x'|x), \tag{4}$$

where, for given $\tilde{x} \in \mathcal{X}$, $m(\tilde{x}) = \mathbb{E}\left[\max_{k' \in \mathcal{D}}\{v_{k'}^*(\tilde{x}) - v_K^*(\tilde{x}) + \varepsilon'_{k'}\}\right]$ 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 $m(\tilde{x}) = -\ln\left(p_K(\tilde{x})\right)$.

3 Magnac and Thesmar's identification result

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

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

where $U_k(x) = u_k^*(x) - u_K^*(x) + \beta\left[\mathbf{Q}_k(x) - \mathbf{Q}_K(x)\right]\mathbf{v}_K$ 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

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

Provided that Magnac and Thesmar's rank condition

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

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 $\ln\left(p_k(x)/p_K(x)\right)$.

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 $m(x_1) = \cdots = m(x_J)$ 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 thus 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

$$\mathbf{v}_K = \left[\mathbf{I} - \beta\mathbf{Q}_K\right]^{-1}\left[\mathbf{u}_K + \beta\mathbf{Q}_K\mathbf{m}\right]. \tag{8}$$

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 (Dubé et al., 2014). 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 thus violating the exclusion restriction, except for special cases.

A lack of good economic intuition for the exclusion restrictions is problematic when they are assumptions that are not themselves tested in the analysis. Without a clear understanding of the economic substance of the current value concept, it is both hard to assess how plausible identifying assumptions based on that concept are, and to find variation that satisfies the restrictions. By defining the exclusion restrictions directly on the primitives, which we do in the next section, we can generally improve the intuition for the economic substance of the assumptions.[2] A better understanding of the economic substance of the restrictions is helpful when evaluating the quality of the identifying assumptions and aids the search for sources of identifying variation.

4 A new identification result

4.1 Main result

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

$$\ln \mathbf{p}_k - \ln \mathbf{p}_K = \beta \left[\mathbf{Q}_k - \mathbf{Q}_K\right] \left[\mathbf{m} + \mathbf{v}_K\right] + \mathbf{u}_k - \mathbf{u}_K. \tag{9}$$

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

$$\begin{aligned} \ln \mathbf{p}_k - \ln \mathbf{p}_K = \beta \left[\mathbf{Q}_k - \mathbf{Q}_K\right] \left[\mathbf{I} - \beta \mathbf{Q}_K\right]^{-1} \mathbf{m} + \mathbf{u}_k \\ - \left[\mathbf{I} - \beta \mathbf{Q}_k\right] \left[\mathbf{I} - \beta \mathbf{Q}_K\right]^{-1} \mathbf{u}_K. \end{aligned} \tag{10}$$

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 (2015) and Bajari et al. (2015), we set $\mathbf{u}_K = \mathbf{0}$.[3] This normalization cannot be refuted by data without

[2] Alternatively, the restrictions could be defined on functions of primitives that correspond to economically meaningful concepts.

[3] Our identification analysis of β applies without change to the case in which $u^*_K(x)$ is constant, but not necessarily zero. It can be straightforwardly extended to the case in which $u^*_K(x)$ is known up to a constant shift, but not necessarily constant. Note that none of these normalizations collapses Magnac and Thesmar's exclusion restriction on current values to an easily interpretable restriction on primitives.

further restrictions.[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, 2014, Lemma 2; Kalouptsidi et al., 2016). It is however standard and allows us to focus on the identification of the discount factor.[5]

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, we take this known value to be zero, and simply focus on the exclusion restriction[6]

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

An advantage of this exclusion restriction over Magnac and Thesmar's current value restriction is that it is a direct restriction on primitive utility. The exclusion restriction extends Magnac and Thesmar by allowing for restrictions on primitive utilities across combinations of choices and states.

Equation (11) resembles the exclusion restrictions in Fang and Wang's (2015) Assumption 5. However, (11) is a weaker restriction that allows for a broader range of identifying variation. It is weaker, since it only needs to hold for some $k = l \in \mathcal{D}$, unlike Fang and Wang's Assumption 5, which must hold for all possible choices $k \in \mathcal{D}$. It is more general than their Assumption 5 by allowing for restrictions on primitive utilities across choices, holding states fixed, e.g. $u_k^*(\tilde{x}) = u_l^*(\tilde{x})$ for some $\tilde{x} \in X$ and $k \neq l$. In Section 7, we further discuss Fang and Wang's (2015) results and show that these do not cover the identification problems we study here.

Under (11), (10) implies

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

Like (6), this moment condition on β carries all the information in the data about

[4] In Section 5, we note that the normalized model can rationalize any choice and state transition probability data.

[5] Chou (2015) recently provided identification results for dynamic discrete choice models without this normalization. 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 8 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.

[6] If $u_k^*(\tilde{x}_1) - u_l^*(\tilde{x}_2)$ is known but not zero, the moment condition on β that we derive from (11), which is (12), applies with a known additive shift of its right hand side. This can be implemented as a known adjustment to its left hand side, so that the identification analysis that follows applies with a straightforward reinterpretation of the scalar in the moment condition's left hand side.

the discount factor and can be used directly for its identification and estimation.[7] 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, known 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 exclusion restriction (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:

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

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

$$\mathbf{I} + \beta\mathbf{Q}_K + \beta^2\mathbf{Q}_K^2 + \cdots. \tag{14}$$

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 $f : (-1,1) \to \mathbb{R}$. 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,

[7] Obviously, 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 restriction used and that rationalize the data (see also Section 5).

2002, Corollary 1.2.7). Because $\mathcal{B} = \mathcal{B}^* \cap [0, 1)$, this implies that $\mathcal{B}$ has no limit point in $[0, 1)$. $\square$

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.

We have moved the terms that frustrated the interpretation of Magnac and Thesmar's exclusion restriction from the realm of untestable assumptions to the moment condition (12), which is verifiable in data.[8] Specifically, as is clear from the proof of Theorem 1, the right hand side of (12) equals β times the sum of two terms,

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

and

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

The first term is the known shift in the expected excess surplus contrast from changing the state and choice from $\tilde{x}_2$ and l to $\tilde{x}_1$ and k. For $k = l$, it equals the change in incentives for choosing k over K that is used in Magnac and Thesmar's identification argument. The second term is the corresponding shift in choice K's expected value contrast, which is not directly known because it depends on β. The two terms add up to a shift in the expected surplus contrast that depends on β. In the special case that $k = l$, this shift can again be interpreted as a change in the incentives for

[8] Section 5 shows that the two exclusion restrictions have nontrivial and distinct empirical contents. That is, the data carry some information on them. However, we also show that data are often consistent with both. Hence, no uniformly consistent test on either restriction exists.

choosing k over K in an experiment that moves x from $\tilde{x}_2$ to $\tilde{x}_1$. However, because this change itself now depends on β, its effect on the log choice probability ratio does not directly identify β.

The derivative of the right hand side of (12) at $\beta = 0$ equals the first term (15) (the second term (16) vanishes because choice K has zero value if the agent is myopic). Consequently, our generalized rank condition (13) ensures that this derivative is nonzero. In economic terms, it guarantees that myopic agents are incentivized to change their behavior if they start caring a little bit about their future. This suffices to locally identify myopic preferences in the case no behavioral response is observed (the left hand side of (12) equals zero).[9] Of course, local identification of myopic preferences does not rule out that the data are also consistent with positive discount factors in the case no behavioral response is observed. These discount factors, if any, can easily be found by searching for the solutions to the moment condition in (12). Note that Theorem 1 does not rely on further rank conditions to establish local identification of these positive discount factors. Instead, it exploits that the moment condition sets a infinite power series in β (a real-analytic function of β) to zero to establish local identification without further conditions. The same is true for local identification in the case a behavioral response is observed (the left hand side of (12) is nonzero).

The proof of Theorem 1 only uses that the coefficients in the moment condition's power series are such that it converges on a domain that contains $[0, 1)$. It does not rely on the fact that these countably many coefficients are fully determined by the finite number of choice and transition probabilities that appear in (12). In applications with small numbers of states (J) and choices (K), this may further restrict the number of possible discount factors that rationalize the data under our exclusion restriction. We have no general results on this. However, in the examples in the next subsection, many of which have $J = 3$ and $K = 2$, we typically find zero, one, or two solutions. In Section 5, we show that there is no solution if $J = 2$, $K = 2$, and choice probabilities are state dependent.

4.2 Examples

Theorem 1 shows that the identified set of discount factors is discrete and, away from one, finite, but does not establish point identification. In some special cases,

[9]Here, β is locally identified at some β_0 if $\beta = \beta_0$ uniquely solves (12) in a neighborhood of β_0. Of course, Magnac and Thesmar's rank condition 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).

the discount factor is point identified. In particular, Magnac and Thesmar's identification result applies whenever the current value collapses to the current period's primitive utility.

Example 1. 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 observe, and an additive and independent shock. Renewal incurs a cost that is independent of mileage and resets mileage to $x_1 = 0$:

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

Consequently, Zurcher's expected discounted payoffs from renewal do not depend on mileage. In particular, with our normalization $\mathbf{u}_K = \mathbf{0}$, $v_K^*(\tilde{x}) = \beta\left(m(x_1) + v_K^*(x_1)\right)$ 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 $u_1^*(\tilde{x}_1) = u_1^*(\tilde{x}_2)$, Magnac and Thesmar's exclusion restriction holds and its identification result applies. Their rank condition (7) simplifies to

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

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).

Because mileage is naturally ordered, the exclusion restriction $u_1^*(\tilde{x}_1) = u_1^*(\tilde{x}_2)$ may be derived from a local shape restriction on Zurcher's utility function. Suppose that the payoffs (relative to those from renewal) from operating a bus with x miles equal $u_1^*(x) = b - c(x)$, for some constant b (which may reflect both the benefits from operating the bus and the costs from possible renewal) and weakly increasing operating cost function c. Let $x_1 < x_2 < \cdots < x_J$ and suppose that Q_1 is increasing, so that buses with more miles now, if operated, tend to have more miles in the future.[10] Figure 1 plots a hypothetical cost function c. It has $c(x_1) = c(x_2)$, which reflects that operating costs are flat for buses with new engines and implies the exclusion restriction $u_1^*(x_1) = u_1^*(x_2)$. If Harold Zurcher is myopic ($\beta = 0$), then his choice problem, and hence the renewal probability, is the same for a bus with x_1 miles and one with x_2 miles. However, mileage is more likely to transition to states

[10] To be precise, Q_1 is such that the distribution of $x'|x = x_i$ first order stochastically dominates that of $x'|x = x_j$ for all $i > j$.

with higher maintenance costs from x_2 than from x_1. Therefore, if Zurcher cares about the future ($\beta > 0$), he is more likely to replace the engine at x_2 miles than at x_1 miles. Thus, under this shape restriction, differences in the log choice probability ratios between states x_1 and x_2 are informative about the discount factor.

Alternatively, we could use that Zurcher's utility is a cardinal payoff, on which some direct data may be available. In particular, the discount factor is identified if the renewal costs and the operating benefits and costs, and thus the (relative) payoffs $u_1^*(\tilde{x})$, are known for some mileage $\tilde{x}$.[11]

Note that neither approach excludes mileage from the operating cost function. We only need that the difference in maintenance costs between either a pair of states or a pair of choices is excluded from the identifying moment condition. In this example, the distinction is essential. Identification could not be obtained by excluding mileage from the operating cost function as it is the only state variable of the problem.

Example 1'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 1, 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 $\tilde{x}$.

Example 2. Consider a discrete time econometric implementation of Dixit's (1989) model of firm entry and exit.[12] 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,

[11] Recall from Footnote 6 that our identification analysis straightforwardly extends to the case in which $u_l^*(\tilde{x}_1) - u_k^*(\tilde{x}_2)$ is known, but not necessarily zero. Specifically, for $l = 1$, $k = K$, and $\tilde{x}_1 = \tilde{x}_2 = \tilde{x}$, it extends to the case in which $u_1^*(\tilde{x})$ is known.

[12] 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.

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 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

$$\mathbf{Q}_1(\tilde{x})\mathbf{v}_K = \mathbb{E}\left[v_K^*(y',1) \,|\, y = \tilde{y}\right] = \mathbb{E}\left[v_K^*(y',K) \,|\, y = \tilde{y}\right] = \mathbf{Q}_K(\tilde{x})\mathbf{v}_K \tag{17}$$

for all $\tilde{x} = (\tilde{y}, \tilde{d}_{-1}) \in \mathcal{X}$. Consequently, as in Example 1, $[\mathbf{Q}_1(\tilde{x}) - \mathbf{Q}_K(\tilde{x})]\mathbf{v}_K = 0$ and $U_1(\tilde{x}) = u_1^*(\tilde{x})$.

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,

$$\mathbf{Q}_k(\tilde{x})\mathbf{m} = \mathbb{E}\left[m(y',k) \,|\, y = \tilde{y}\right]. \tag{18}$$

Thus, the rank condition is equivalent to

$$\mathbb{E}\left[m(y',1) - m(y',K) \,|\, y = \tilde{y}_1\right] \neq \mathbb{E}\left[m(y',1) - m(y',K) \,|\, y = \tilde{y}_2\right]. \tag{19}$$

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

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

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 (17) and (18).

Finally, if both $\tilde{y}_1 \neq \tilde{y}_2$ and entry costs are strictly positive, (19) will generally be satisfied. In specific applications, we can verify (19) using that both the distribution of y' conditional on y and $m(y',k) = -\ln\left(p_K(y',k)\right)$ 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 again 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

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

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

In Examples 1 and 2, the rank condition ensures that the shift in expected surplus contrasts that multiplies β in the right hand side of (12) is nonzero. Because this shift does not depend on β itself, this suffices for identification. Even in cases in which it depends on β, it may suffice for identification that it is nonzero.

Example 3. Consider (12) with $k = l = 1$ and suppose that its left hand side equals zero, i.e. that the choice probability ratio does not change between states $\tilde{x}_2$ and $\tilde{x}_1$. Then, (12) requires that either $\beta \in (0, 1)$ is such that this shift in surplus contrasts is zero, or $\beta = 0$. Consequently, in this special case, it is sufficient and necessary for identification that the shift in expected surplus contrasts is nonzero for all $\beta \in (0, 1)$. Intuitively, this ensures that the incentives (from future payoffs) for choosing 1 over K differ between states $\tilde{x}_2$ and $\tilde{x}_1$, so that a lack of response in choices can only be explained by myopia, $\beta = 0$.

In general, however, neither Magnac and Thesmar's rank condition nor a nonzero shift in expected surplus contrasts suffices for identification.

Example 4. Figure 2 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 Magnac and Thesmar's rank condition and imply that the right hand side of (12) is positive on $(0, 1)$.

The choice probabilities imply a relatively low excess surplus $m(x_3)$ in state x_3. Because the experiment underlying Magnac and Thesmar's rank condition moves probability mass away from state x_3, the right hand side of (6), and the first (excess surplus) term in the right hand side of (12), slope upward and equal the left hand side for only one value of β. Under the current value restriction, this is the only discount factor consistent with the data.

Under the primitive utility restriction, we also need to take account of the second (value of choice K) term in the right hand side of (12). In contrast to the excess surplus $m(x_3)$, the value $v_K(x_3)$ is relatively high, because $\mathbf{Q}_K(x_3)$ puts a relatively low (zero) probability on ending up in the low excess surplus state x_3. Consequently, the move of probability mass away from state x_3 renders the second term in the right hand side of (12) negative, and increasingly so with increasing β. It follows that the right hand side of (12) first equals its left hand side at a slightly higher discount factor than the one identified under Magnac and Thesmar's condition. Moreover, the negative contribution of the second term eventually grows so large that the right hand side of (12) again equals the left hand side at a discount factor closer to one. Thus, two distinct discount factors are consistent with the data under the primitive utility restriction.

Magnac and Thesmar's rank condition is not necessary for identification either.

Example 5. Figure 3 presents an example in which the shift in expected excess surplus is 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 Magnac and Thesmar's rank condition.

Also note that there is no value of β that satisfies (6). Thus, 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 5.

More generally, strict monotonicity of the right hand side of (12), as in Example 5, suffices for point identification. 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:

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

For this, it suffices that

$$[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_l(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)] \mathbf{Q}_K^r \mathbf{m} \geq 0 \text{ for all } r \in \{0, 1, 2, \ldots\}, \quad (20)$$

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 5 (which is specified in the Note to Figure 3).

The final example relies on a type of payoff monotonicity that is common in models with ordered states.

Example 6. Suppose that the states in $\mathcal{X}$ are ordered so that the excess surplus is increasing in the state: $m_1 < m_2 < \cdots < m_J$. Moreover, let $\mathbf{Q}_K$ be increasing. Consider our exclusion restriction on primitive utility for $\tilde{x}_1 \neq \tilde{x}_2$ and $k = l$. There again is no reason for Magnac and Thesmar's current value restriction to hold under this restriction. The structure however allows us to formulate easy-to-handle sufficient conditions.

First, suppose that the transitions under choice K are not affected by a change of state from $\tilde{x}_2$ to $\tilde{x}_1$: $\mathbf{Q}_K(\tilde{x}_1) = \mathbf{Q}_K(\tilde{x}_2)$. Note that the expected excess surplus after r state transitions under choice K, $\mathbf{Q}_K^r \mathrm{m}$, is increasing in the initial state. Thus, sufficient condition (20) is satisfied if $\mathbf{Q}_k(\tilde{x}_1)$ first-order stochastically dominates $\mathbf{Q}_k(\tilde{x}_2)$; that is, if a move of state from $\tilde{x}_2$ to $\tilde{x}_1$ under choice k first-order stochastically increases the next period's state.

Next, if $\mathbf{Q}_K(\tilde{x}_1) = \mathbf{Q}_K(\tilde{x}_2)$ does not hold, condition (20) is satisfied if the effect of a change of state from $\tilde{x}_2$ to $\tilde{x}_1$ on next period's state under choice k first-order dominates that same effect under choice K.

5 Empirical content

The previous two sections 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.

First note that a version of Magnac and Thesmar's (2002) Proposition 2 holds: For any given data $\{\mathbf{p}_k, \mathbf{Q}_k; k \in \mathcal{D}\}$, $\mathbf{u}_K = 0$, 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 $\mathbf{u}_K = 0$ 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 $\mathbf{u}_K = 0$).

This result justifies our focus on the identification of the discount factor β in the previous two sections: Once the discount factor has been identified, we can find unique primitive utilities that rationalize the data. It also implies that the data cannot tell us whether the model without exclusion restrictions is false or not; that is, the unrestricted model has no empirical content. Therefore, we now turn to the empirical consequences of a violation of the assumed exclusion restriction. Such a violation can manifest itself in two distinct ways.

First, in some cases, it may be possible to find primitives that both satisfy the false exclusion restriction and are compatible with the data. If so, these primitives will in general not equal the true primitives. In Example 4, falsely assuming Magnac and Thesmar's current value restriction when the primitive utility restriction is true identifies a discount factor strictly below the true one. Because we can find primitive utilities that rationalize the data for any discount factor, the data can be of no help to determine the right restriction in this case. Instead, we need to argue in favor of one exclusion restriction or the other on other grounds. In Section 4, we presented a novel identification analysis for exactly this reason: the primitive utility restriction is comparatively easy to motivate and justify in applications.

Second, the data may be incompatible with the assumed exclusion restriction. For example, the data in Example 5 cannot be rationalized under the current value restriction, even though they are compatible with some specification of the model. In that example, the data are however consistent with an exclusion restriction on primitive utility. Conversely, there exist data that are inconsistent with the primitive utility restriction, but that can be rationalized by primitives that satisfy the current value restriction.

Example 7. Figure 4 displays the left and right hand sides of (6) and (12) for a variant of Example 4's data in which the shift in the log choice probability ratio when moving the state from $\tilde{x}_2$ to $\tilde{x}_1$, and therefore the left hand side of (6) and (12), is twice as large. At the same time, the right hand sides are similar to those in Example 4 (as is easily verified by comparing Figure 4 to Figure 2). There is still a $\beta \in [0,1)$ that solves (6), but (12) can no longer 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.

Examples 4 and 7 establish that the two exclusion restrictions have nontrivial and distinct empirical contents, so that, to some extent, data can distinguish them. 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 $\beta \in [0,1)$.

As we have already noted, without restrictions, the primitives are free to generate value contrasts $\mathbf{v}_k - \mathbf{v}_K$ that are compatible with any given choice data (that is, satisfy (3)). By extension, any observed shift in the log choice probability ratio from $\tilde{x}_2$ to $\tilde{x}_1$ can be rationalized by setting $u_k^*(\tilde{x}_1)$, $u_k^*(\tilde{x}_2)$, and β such that

$$\begin{aligned} &v_k^*(\tilde{x}_1) - v_K^*(\tilde{x}_1) - v_l^*(\tilde{x}_2) + v_K^*(\tilde{x}_2) = \\ &u_k^*(\tilde{x}_1) - u_l^*(\tilde{x}_2) + \beta[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_l(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)]\,[\mathbf{m} + \mathbf{v}_K] \end{aligned} \tag{21}$$

matches this observed shift.

Under the primitive utility restriction across states, the range of the right hand side of (21) is limited by forcing $u_k^*(\tilde{x}_1) - u_l^*(\tilde{x}_2) = 0$, while under the current value restriction across states, the range is limited by requiring that $u_k^*(\tilde{x}_1) - u_l^*(\tilde{x}_2)$ is exactly offset by $[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_l(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)]\mathbf{v}_K$. The set of choice data that fall outside the range of the model under either restriction can be substantial. We give some examples in the Appendix.

6 Multiple exclusion restrictions and inference

Sometimes more than one exclusion restriction is available. For instance, a state variable that is observed to take at least three values and that is excluded from the utility function gives multiple exclusion restrictions. With multiple exclusion restrictions, point identification can be obtained even if each individual moment condition only gives set identification. We give two examples of identification with two exclusion restrictions.

Example 8. 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 point identifies the discount factor individually, while the solid moment condition individually set identifies the discount factor. In this case, the solid moment condition is redundant for point identification.

Example 9. 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 may to some extent vary freely with e.g. 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.

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. Rather than to insist on point identification, an alternative approach is to accept set identification and use estimators that consistently estimate the identified set, which may contain one or more points. We give one example.

Example 10. Suppose the population moment condition are as given in Figure 6. Though each individual moment condition is equally consistent with one low discount factor, at 0.07 and 0.17, respectively, and one high 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 can not be determined a priori without further assumptions, the discount factor is set identified and we may look for 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 $S : [0, 1) \to [0, \infty)$. 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:

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

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.

For point identified problems, such as the optimal stopping problems in Examples 1 and 2 or the monotonic problems in Example 6, standard inference for extremum estimators applies (e.g. Newey and McFadden, 1994).

7 Relation to Fang and Wang (2015)

Fang and Wang (2015) considered generic identification of a dynamic discrete choice model with hyperbolic discounting, which encompasses geometric discounting is a special case. At first glance, its Proposition 2 seems to accomplish this paper's improvements on Magnac and Thesmar and more. However, in Section 4, we have shown that point identification can be had under weaker conditions; hence, Fang and Wang's conditions are not necessary. Furthermore, Fang and Wang's condition $\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_k(\tilde{x}_2) \neq 0$ in its Assumption 5 is not sufficient for point identification of the geometric discount factor, under either Magnac and Thesmar's or our exclusion restrictions. In particular, this condition does not preclude that $\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_k(\tilde{x}_1) + \mathbf{Q}_K(\tilde{x}_1) = \mathbf{0}$, in which case the right hand sides of both (6) and (12) are zero, and therefore uninformative about β.

Finally, Fang and Wang's Proposition 2 does not apply to geometric discounting.[13] Proposition 2 only establishes generic identification, where "generic" is defined over the space of possible values of the data $\{\mathbf{p}_k, \mathbf{Q}_k; k \in \mathcal{D}\}$. First, Fang and Wang's model generically does not have geometric discounting (that is, "present bias" and

[13] In Section E of its online appendix, Fang and Wang treated the geometric discounting case explicitly. Paralleling Magnac and Thesmar's Proposition 2, Section E shows that conditional on β, there is a unique $\mathbf{u}$ that rationalizes the data. That falls short of identification of the discount factor, which is the concern of this paper.

"partial naivety" parameters that both equal one). Consequently, Fang and Wang's Proposition 2 has no bearing on its identification. Second, its "rank condition," which only requires that $\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_k(\tilde{x}_2) \neq 0$, is generically satisfied in the data space, and therefore cannot be necessary for its generic identification result.

Our results do not imply that Fang and Wang's generic identification result for hyperbolic discounting is false. Rather, they show that Fang and Wang's generic identification result for hyperbolic discounting sheds no light on the identification of the geometric discount factor.

8 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. Time shifts the continuation values over the life cycle, while stationary utilities can be cast as exclusion restrictions.

Similar ideas have been used in previous literature. Bajari et al. (2016) used the assumption of stationary utilities to formally establish identification in a finite horizon optimal stopping model. Theorem 2 below extends Bajari et al.'s result beyond optimal stopping problems and allows for identification with nonstationary utilities.

Yao et al. (2012) showed identification of the discount factor of a dynamic model with continuous controls under a time homogeneity assumption and conjectured a similar result for discrete controls. Theorem 2 proves its conjecture.

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

$$\mathbf{v}_{k,t} = \mathbf{u}_{k,t} + \beta \mathbf{Q}_k \left[\mathbf{m}_{t+1} + \mathbf{v}_{t+1}\right] \tag{22}$$

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

$$\ln\left(p_{k,t}(\tilde{x})\right) - \ln\left(p_{K,t}(\tilde{x})\right) = u^*_{k,t}(\tilde{x}) + \beta[\mathbf{Q}_k(\tilde{x}) - \mathbf{Q}_K(\tilde{x})][\mathbf{m}_{t+1} + \mathbf{v}_{K,t+1}] \tag{23}$$

for all $k \in \mathcal{D}\backslash\{K\}$ and $\tilde{x} \in \mathcal{X}$. Finally, using (22) and the normalization $\mathbf{u}_{K,t} = \mathbf{0}$

for all t, we can write the value of the reference choice K as

$$\mathbf{v}_{K,t} = \sum_{\tau=t+1}^{T} (\beta \mathbf{Q}_K)^{\tau-t} \mathbf{m}_\tau, \tag{24}$$

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

Theorem 2. Suppose that

$$u^*_{k,t}(\tilde{x}_1) = u^*_{l,t'}(\tilde{x}_2) \tag{25}$$

for $k \in \mathcal{D}/\{K\}$, $l \in \mathcal{D}$, $\tilde{x}_1 \in \mathcal{X}$, $\tilde{x}_2 \in \mathcal{X}$, $1 \leq t' < T$, 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

$$[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1)]\mathbf{m}_{t+1} - [\mathbf{Q}_l(\tilde{x}_2) - \mathbf{Q}_K(\tilde{x}_2)]\mathbf{m}_{t'+1} \neq 0, \tag{26}$$

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

Proof. Differencing (23) corresponding to (25) and substituting (24) gives

$$\begin{aligned} \ln\left(p_{k,t}(\tilde{x}_1)/p_{K,t}(\tilde{x}_1)\right) - \ln\left(p_{l,t'}(\tilde{x}_2)/p_{K,t'}(\tilde{x}_2)\right) = \\ \beta\Bigg([\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1)] \Big[\sum_{\tau=t+1}^{T} (\beta \mathbf{Q}_K)^{\tau-t-1} \mathbf{m}_\tau \Big] - \\ [\mathbf{Q}_l(\tilde{x}_2) - \mathbf{Q}_K(\tilde{x}_2)] \Big[\sum_{\tau=t'+1}^{T} (\beta \mathbf{Q}_K)^{\tau-t'-1} \mathbf{m}_\tau \Big] \Bigg). \end{aligned} \tag{27}$$

For given choice and transition probabilities, the right hand side of (27) 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 has up to $T - t'$ real roots, which is an upper bound on the number of points in the identified set. To show that (27) is nonconstant under the stated assumptions, note first that the right hand side of (27) 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 (26) 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. □

The rank condition (26) is our extension (13) of Magnac and Thesmar's rank condition, with time denoted by indices rather than included as states. 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. The empirical content of the identified model depends on the chosen domain, so we do not extend the analysis of Section 5, but note that Theorem 2 does not guarantee a real root (and less so one in a specified domain for β) for general choice and state probabilities.

In a study of identification in nonstationary models, Arcidiacono and Miller (2015) distinguished between identification in long panels, panels that include the terminal period, and short panels, which do not. In general, Theorem 2 requires long panels, but some classes of problems that restrict the continuation value are identified also in short panels. Optimal stopping problems like Zurcher's in Example 1 and in Bajari et al. (2016) form one such class of problems. Since $v_{K,t}(\tilde{x})$ is state independent in these problems, it follows directly from (23) and (25) that (27) simplifies to

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

Since (28) only involves data for periods t, $t+1$, t', and $t'+1$, the discount factor is identified in short panels.

References

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. 12

Arcidiacono, P. and R. Miller (2015). Identifying dynamic discrete choice models of short panels. Mimeo, Duke University. 23

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. 2, 6

Bajari, P., C. S. Chu, D. Nekipelov, and M. Park (2016). Identification and semiparametric estimation of a finite horizon dynamic discrete choice model with a terminating action. *Quantitative Marketing and Economics 14*(4), 271–323. 21, 23

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. 2

Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica 75*(5), 1243–1284. 19

Ching, A., T. Erdem, and M. P. Keane (2014). A simple method to estimate the roles of learning, inventories and category consideration in consumer choices. *Journal of Choice Modelling 13*, 60–72. 2

Chou, C. (2015, October). Identification and linear estimation of general dynamic programming discrete choice models. Mimeo, University of Southern California, Los Angeles, CA. 7

Dixit, A. K. (1989). Entry and exit decisions under uncertainty. *Journal of Political Economy 97*(3), 620–638. 12

Dubé, J.-P., G. J. Hitsch, and P. Jindal (2014). The joint identification of utility and discount functions from stated choice data: An application to durable goods adoption. *Quantitative Marketing and Economics 12*(4), 331–377. 5

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. 3, 6, 7, 20, 21

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica 60*(5), 1127–1150. 12

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. 4

Kalouptsidi, M., P. T. Scott, and E. Souza-Rodrigues (2016, June). Identification of counterfactuals in dynamic discrete choice models. Mimeo, Princeton University. 7

Keane, M. P. and K. I. Wolpin (1997, June). The career decisions of young men. *Journal of Political Economy 115*(3), 473–522. 3, 21

Krantz, S. G. and H. R. Parks (2002). *A Primer of Real Analytic Functions* (Second ed.). Boston: Birkhäuser Verlag. 8

Lee, R. (2013). Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review 103*(7), 2960–3000. 2

Magnac, T. and D. Thesmar (2002). Identifying dynamic discrete choice processes. *Econometrica 70*, 801–816. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 15, 16, 17, 20, 22, 28, 31

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. 20

Norets, A. and X. Tang (2014). Semiparametric inference in dynamic binary choice models. *Review of Economic Studies 81*(3), 1229–1262. 2, 7

Romano, J. and A. Shaikh (2010). Inference for the identified set in partially identified econometric models. *Econometrica 78*(1), 169–211. 19

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica 55*, 999–1033. 3, 11, 29

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. 2

Sargan, J. D. (1983). Identification and lack of identification. *Econometrica 51*(6), 1605–1633. 10

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. 21

Appendix: Examples of empirical content

For $K = J = 2$, the restriction $u_1^*(\tilde{x}_1) = u_1^*(\tilde{x}_2)$ and the normalization $u_K^*(\tilde{x}_1) = u_K^*(\tilde{x}_2)$ together imply that $v_1^*(\tilde{x}_1) - v_K^*(\tilde{x}_1) = v_1^*(\tilde{x}_2) - v_K^*(\tilde{x}_2)$. Thus, by (3), the model cannot rationalize data with state dependent choice probabilities. It can be compatible with state independent choice probabilities, but then β is not identified: Since $v_1^*(\tilde{x}_1) - v_K^*(\tilde{x}_1) = v_1^*(\tilde{x}_2) - v_K^*(\tilde{x}_2)$, it follows that $m(\tilde{x}_1) = m(\tilde{x}_2)$ and the experiment is uninformative, as seen in Example 3. A third choice with state dependent utility for the added choice or, alternatively, a third state with $u_1^*(x_3) \neq u_1^*(\tilde{x}_1) = u_1^*(\tilde{x}_2)$ is necessary to generate state dependence of the value contrasts.

For $K = J = 2$ and under the current value restriction, the model can, under some conditions on the transition probabilities, rationalize state dependent $p_1(\tilde{x}_1)$ and $p_1(\tilde{x}_2)$ that are sufficiently large. To see this, rewrite the moment condition in (6) as

$$\ln\left(\frac{p_1(\tilde{x}_1)}{p_1(\tilde{x}_2)}\right) + (1 - \beta\Delta)\ln\left(\frac{1 - p_1(\tilde{x}_2)}{1 - p_1(\tilde{x}_1)}\right) = 0, \tag{29}$$

where $\Delta \in [-2, 2]$ is such that $[\mathbf{Q}_1(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_1(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)] = [\Delta\ , -\Delta]$. Evidently, $\Delta > 1$ is necessary to rationalize any state dependent choice data, but does not suffice. For any $\Delta \in (1, 2]$, (29) imposes further cross-restrictions on the choice probabilities that can be rationalized. If $\Delta = 2$, state dependent $p_1(\tilde{x}_1)$ and $p_1(\tilde{x}_2)$ are compatible with some discount factor $\beta \in [0, 1)$ if and only if $p_1(\tilde{x}_1) + p_1(\tilde{x}_2) > 1$. With lower Δ, only values of $p_1(\tilde{x}_1)$ and $p_1(\tilde{x}_2)$ closer to one are consistent with some $\beta \in [0, 1)$. Interestingly, with state dependent choice data, any such β necessarily takes values in $(\Delta^{-1}, 1) \subseteq (\frac{1}{2}, 1)$.

If either the number of choices $K \geq 3$ or the number of states $J \geq 3$, then both exclusion restrictions are compatible with nontrivial sets of choice and state transition probabilities. Their empirical contents are more subtle in this case and hard to characterize in general. We limit our discussion to one example.

Example 11. Let $\mathbf{Q}_K = \mathbf{I}$ and suppose the data imply a positive shift from $\tilde{x}_2$ to $\tilde{x}_1$ in the expected excess surplus contrast; i.e. $[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_k(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)]\mathbf{m} > 0$. Then, the moment condition in (12) can be rewritten as

$$\frac{\beta}{1-\beta} = \frac{\ln\left(p_k(\tilde{x}_1)/p_K(\tilde{x}_1)\right) - \ln\left(p_k(\tilde{x}_2)/p_K(\tilde{x}_2)\right)}{[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_k(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)]\mathbf{m}}. \tag{30}$$

Because the left hand side of (30) takes all values in $[0, \infty)$ when β takes values in

$[0,1)$, any choice probabilities such that $\ln\left(p_k(\tilde{x}_1)/p_K(\tilde{x}_1)\right) - \ln\left(p_k(\tilde{x}_2)/p_K(\tilde{x}_2)\right) \geq 0$ can be rationalized with some β in this case. The intuition for this is straightforward. Because $\mathbf{Q}_K = \mathbf{I}$, not only the shift in the expected excess surplus contrast, but also the corresponding shift in the expected surplus contrast is positive. Moreover, the exclusion restriction does not allow the current period utility contrast to change between $\tilde{x}_2$ and $\tilde{x}_1$. Consequently, the value contrasts $v_k^*(\tilde{x}_1) - v_K^*(\tilde{x}_1) \geq v_k^*(\tilde{x}_2) - v_K^*(\tilde{x}_2)$, so that the model is only compatible with nonnegative shifts in the log choice probability ratio.

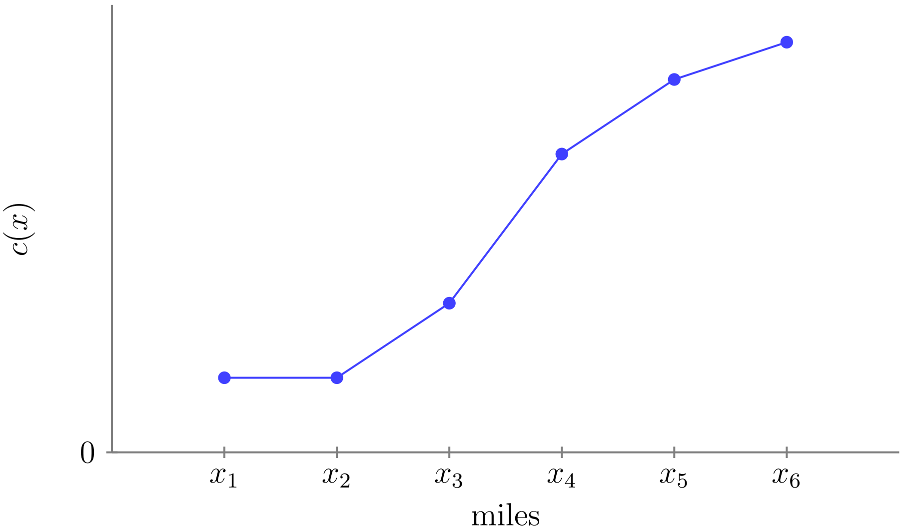
Under the current value restriction, Magnac and Thesmar's moment condition is

$$\beta = \frac{\ln\left(p_k(\tilde{x}_1)/p_K(\tilde{x}_1)\right) - \ln\left(p_k(\tilde{x}_2)/p_K(\tilde{x}_2)\right)}{[\mathbf{Q}_k(\tilde{x}_1) - \mathbf{Q}_K(\tilde{x}_1) - \mathbf{Q}_k(\tilde{x}_2) + \mathbf{Q}_K(\tilde{x}_2)]\mathbf{m}}. \tag{31}$$

The range of the left hand side of (31) is $[0,1)$, which is a strict subset of the range $[0,\infty)$ of the left hand side of (30). The right hand sides of (31) and (30) are the same.[14] Therefore, like (30), (31) cannot rationalize negative shifts in the log choice probability ratio. In addition, the model is not compatible with too large positive shifts in the log choice probability ratio. Thus, the current value restriction has more empirical content than the primitive utility restriction in this case but, as Example 7 shows, this is not generally true.

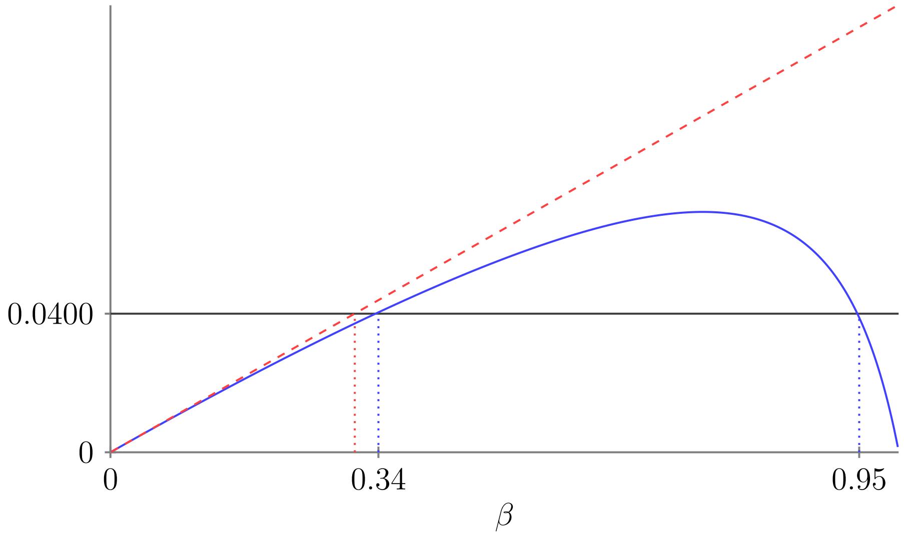
[14] The range of the right hand side can be shown to be $\mathbb{R}$ under the stated assumptions.

Figure 1: Example of a Shape Restriction on Utility that Implies an Exclusion Restriction



Note: In this stylized example of Rust's (1987) bus engine renewal problem, mileage takes $J = 6$ values, the cost $c(x)$ of operating an engine with x miles is constant between x_1 and x_2 miles and increases thereafter, and the utility $u_1^*(x)$ from operating an engine with x miles equals a constant minus the operating cost $c(x)$.

Figure 2: Example in Which an Exclusion Restriction on Current Values Suffices for Identification but One on Primitive Utility Does Not

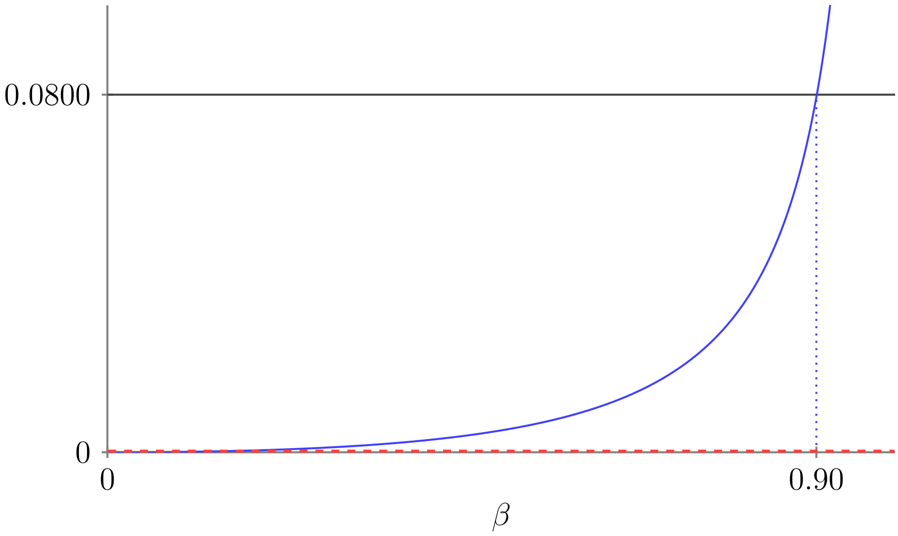


Note: For $J = 3$ states, $K = 2$ choices, $k = l = 1$, $\tilde{x}_1 = x_1$, and $\tilde{x}_2 = x_2$, this graph plots the left hand side of (6) and (12) (solid black horizontal line) and the right hand sides of (6) (dashed red line), and (12) (solid blue curve) as functions of β. The data are $\mathbf{Q}_1(x_1)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

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

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, $\mathbf{m}' = \begin{bmatrix} 0.69 & 0.67 & 0.11 \end{bmatrix}$ and $\mathbf{Q}_1(x_1) - \mathbf{Q}_K(x_1) - \mathbf{Q}_1(x_2) + \mathbf{Q}_K(x_2) = \begin{bmatrix} -0.65 & 0.90 & -0.25 \end{bmatrix}$, 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.

Figure 3: Example in Which Magnac and Thesmar's Rank Condition Fails, but an Exclusion Restriction on Primitive Utility Suffices for Identification

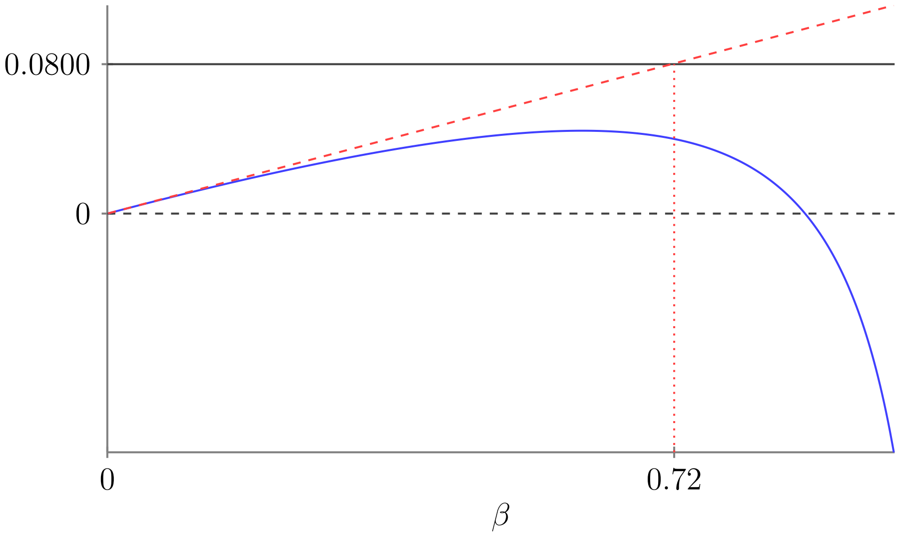


Note: For $J = 3$ states, $K = 2$ choices, $k = l = 1$, $\tilde{x}_1 = x_1$, and $\tilde{x}_2 = x_2$, this graph plots the left hand side of (6) and (12) (solid black horizontal line) and the right hand sides of (6) (dashed red line) and (12) (solid blue curve) as functions of β. The data are $\mathbf{Q}_1(x_1)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

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

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, $\mathbf{m}' = \begin{bmatrix} 0.69 & 0.65 & 0.69 \end{bmatrix}$ and $\mathbf{Q}_1(x_1) - \mathbf{Q}_K(x_1) - \mathbf{Q}_1(x_2) + \mathbf{Q}_K(x_2) = \begin{bmatrix} -0.25 & 0.00 & 0.25 \end{bmatrix}$, 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.

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

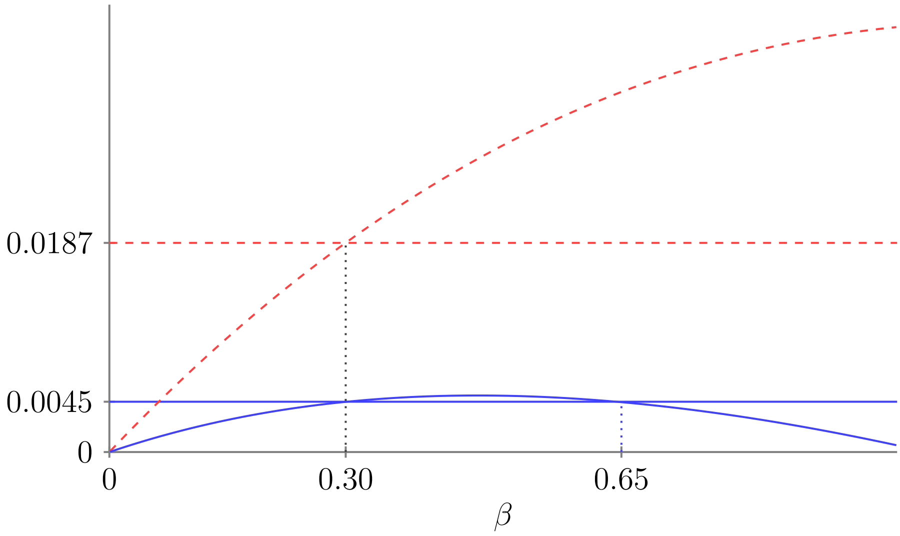


Note: For $J = 3$ states, $K = 2$ choices, $k = l = 1$, $\tilde{x}_1 = x_1$, and $\tilde{x}_2 = x_2$, this graph plots the left hand side of (6) and (12) (solid black horizontal line) and the right hand sides of (6) (dashed red line) and (12) (solid blue curve) as functions of β. The data are $\mathbf{Q}_1(\tilde{x}_1)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

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

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, $\mathbf{m}' = \begin{bmatrix} 0.69 & 0.65 & 0.11 \end{bmatrix}$ and $\mathbf{Q}_1(x_1) - \mathbf{Q}_K(x_1) - \mathbf{Q}_1(x_2) + \mathbf{Q}_K(x_2) = \begin{bmatrix} -0.65 & 0.90 & -0.25 \end{bmatrix}$, 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.

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



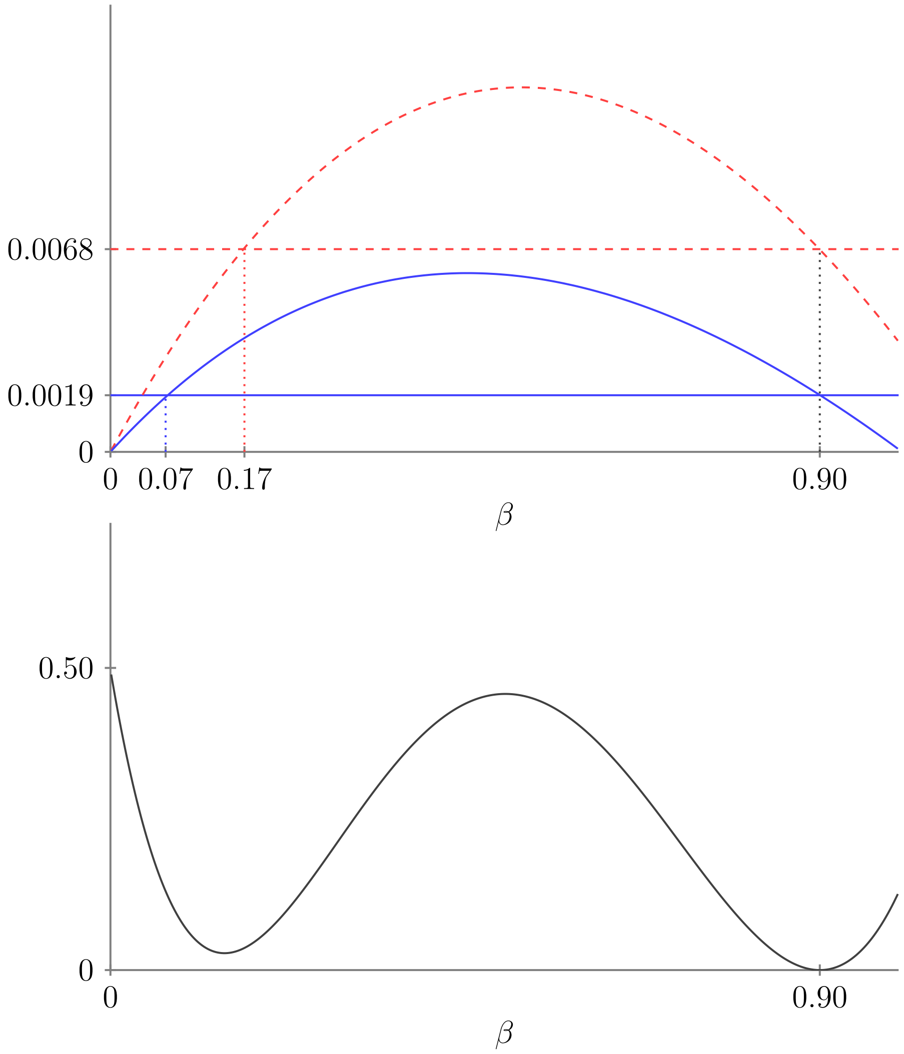
Note: For $J = 4$ states, $K = 2$ choices, and $k = l = 1$, this graph plots the left (horizontal lines) and right hand sides (curves) of (12) as functions of β, for $\tilde{x}_1 = x_1$ and $\tilde{x}_2 = x_2$ (corresponding to $u_1(x_1) = u_1(x_2)$; dashed red) and $\tilde{x}_1 = x_3$ and $\tilde{x}_2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

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

$$\mathbf{p}_1' = \begin{bmatrix} 0.60 & 0.59 & 0.88 & 0.88 \end{bmatrix}, \text{ and } \mathbf{p}_K' = \begin{bmatrix} 0.40 & 0.41 & 0.12 & 0.12 \end{bmatrix}.$$

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.

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



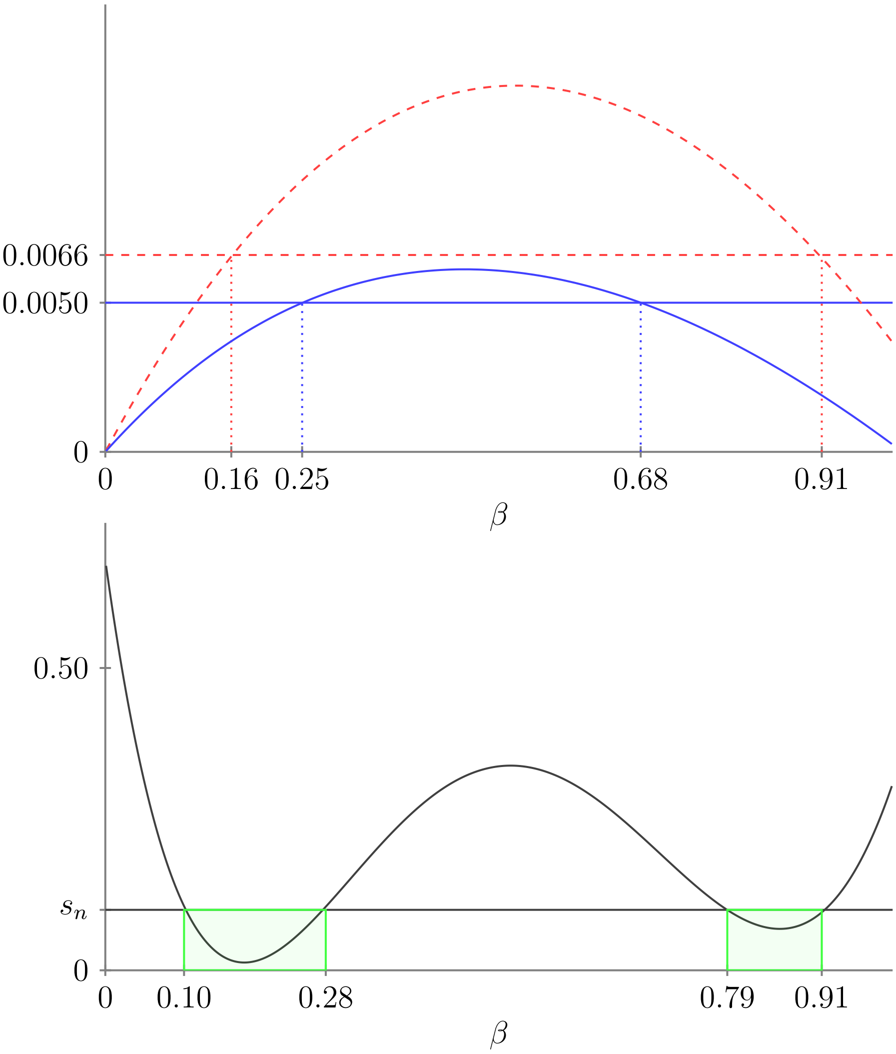
Note: For $J = 4$ states, $K = 2$ choices, and $k = l = 1$, the graph in the top panel plots the left (horizontal lines) and right hand sides (curves) of (12) as functions of β, for $\tilde{x}_1 = x_1$ and $\tilde{x}_2 = x_2$ (corresponding to $u_1(x_1) = u_1(x_2)$; dashed red) and $\tilde{x}_1 = x_3$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

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

$$\mathbf{p}_1' = \begin{bmatrix} 0.92 & 0.92 & 0.63 & 0.63 \end{bmatrix}, \text{ and } \mathbf{p}_K' = \begin{bmatrix} 0.08 & 0.08 & 0.37 & 0.37 \end{bmatrix}.$$

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 condition 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).