Identifying the Discount Factor in Dynamic Discrete Choice Models

Jaap H. Abbring† Øystein Daljord‡

August 1, 2018

Abstract

Empirical applications of dynamic discrete choice models usually either take the discount factor to be known or rely on ad hoc functional form assumptions to identify and estimate it. We give identification results under economically motivated exclusion restrictions on primitive utilities. We show that each such 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 provide point identification. In applications, multiple exclusion restrictions are often available. The discount factor can be set estimated from the corresponding moment conditions using existing methods without solving the choice model. 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.

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.

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 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 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 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 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 intuitive strategy to identify the discount factor in dynamic discrete choice models with economically motivated exclusion restrictions. This allows empirical researchers to study time preferences in the specific context of their application, which is important because discount factors have been estimated to vary substantially across choice contexts and populations (Frederick et al., 2002).

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

\footnote{Frederick 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).}
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.

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; Kalouptsidi 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 Appendix A, we show that our results straightforwardly extend to the case in which the reference utility is known up to a constant shift.

Fang and Wang (2015) studied identification of dynamic discrete choice models with partially naive hyperbolic time preferences and exclusion restrictions. In Appendix B, we show that its main generic identification result is void: It uses a definition of generic identification that does not preclude the possibility that the model is nowhere identified, and the proof is incorrect. It therefore has no implications for identification of the model with hyperbolic discounting or its special case with exponential discounting that we study. Komarova et al. (2017) investigated identification under parametric assumptions on the utility function in a model like ours. We instead focus on nonparametric identification under economically motivated exclusion restrictions. We deviate from both papers by mapping each exclusion restriction to an easily interpretable and computable moment condition that directly informs the model’s identification, estimation, and empirical content.
2 Model

Consider a stationary dynamic discrete choice model (e.g. Rust, 1994). Time is discrete with an infinite horizon. In each period, agents first observe state variables \( x \) and \( \varepsilon \), where \( x = \{x_1, \ldots, x_J\} \) 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 \( D = \{1, 2, \ldots, K\} \) and collect utility \( u_d(x, \varepsilon) = u_\ast_d(x) + \varepsilon_d \). Finally, they move to the next period with new state variables \( x' \) and \( \varepsilon' \) 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 \( \varepsilon \) from the transition distribution \( Q_k(\cdot|x) \) for any choice \( k \in D \); and \( \varepsilon_1, \ldots, \varepsilon_K \) are independently drawn from mean zero type-1 extreme value distributions.

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 \( v_k(x, \varepsilon) = v_\ast_k(x) + \varepsilon_k \), with \( v_\ast_k \) the unique solution to

\[
v_\ast_k(x) = u_\ast_k(x) + \beta \mathbb{E} \left[ \max_{k' \in D} \{v_\ast_{k'}(x') + \varepsilon_{k'}\} \right] d = k, x
\]

\[
= u_\ast_k(x) + \beta \int \mathbb{E} \left[ \max_{k' \in D} \{v_\ast_{k'}(\tilde{x}) + \varepsilon_{k'}\} \right] dQ_k(\tilde{x}|x)
\]

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

\[
\mathbb{E} \left[ \max_{k' \in D} \{v_\ast_{k'}(\tilde{x}) + \varepsilon_{k'}\} \right] = \ln \left( \sum_{k' \in D} \exp \left( v_\ast_{k'}(\tilde{x}) \right) \right)
\]

is the McFadden surplus for the choice among \( k' \in D \) with utilities \( v_\ast_{k'}(\tilde{x}) + \varepsilon_{k'} \).

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 \( k \in D \) and \( \tilde{x} \in X \). 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 \( \beta \).

\[\text{Section 3.6 considers an extension to a nonstationary model with a finite horizon.}\]

\[\text{Magnac and Thesmar showed that the distribution of } \varepsilon \text{ cannot be identified and took it to be known. Our type-1 extreme value assumption leads to the canonical multinomial logit case. Our results extend directly to any other known, continuous distribution on } \mathbb{R}^K.\]
The choice probabilities are fully determined by
\[ \ln \left( \frac{p_k(\tilde{x})}{p_K(\tilde{x})} \right) = v_k^*(\tilde{x}) - v_K^*(\tilde{x}), \quad k \in \mathcal{D}/\{K\}, \quad \tilde{x} \in \mathcal{X}. \tag{3} \]

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
\[ v_k^*(x) = u_k^*(x) + \beta \int (m(x') + v_k^*(x')) dQ_k(x'|x), \tag{4} \]
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 \( v_k, p_k, u_k, \) and \( m \) 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 \( Q_k \) be the \( J \times J \) matrix with \((j,j')\)-th entry \( Q_k(x_{j'}|x_j) \) and \( I \) be a \( J \times J \) identity matrix. Note that the \( J \times 1 \) vector \( m + v_K \) stacks the McFadden surpluses in (2). In this notation, the data are \( \{p_k, Q_k; k \in \mathcal{D}\} \) and directly identify \( m = -\ln p_K \) (Arcidiacono and Miller, 2011, Lemma 1 and Section 3.3).

3.1 Magnac and Thesmar’s result

We can rewrite (4) as \( v_k^*(x) = u_k^*(x) + \beta Q_k(x) [m + v_K] \), where \( Q_k(x_{j}) \) is the \( j \)-th row of \( Q_k \). Subtracting the same expression for \( v_K^*(x) \), rearranging, and substituting (3), we get
\[ \ln(p_k(x)) - \ln(p_K(x)) = \beta [Q_k(x) - Q_K(x)] m + U_k(x), \tag{5} \]
where \( U_k(x) = u_k^*(x) - u_K^*(x) + \beta [Q_k(x) - Q_K(x)] v_K \) 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

\[
\ln \left( \frac{p_k(\tilde{x}_1)}{p_K(\tilde{x}_1)} \right) - \ln \left( \frac{p_k(\tilde{x}_2)}{p_K(\tilde{x}_2)} \right) = \beta \left[ Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1) - Q_k(\tilde{x}_2) + Q_K(\tilde{x}_2) \right] m. 
\]  

(6)

Provided that Magnac and Thesmar’s rank condition

\[
[Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1) - Q_k(\tilde{x}_2) + Q_K(\tilde{x}_2)] m \neq 0
\]

(7)

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

This identification argument can be interpreted in terms of an experiment that shifts the expected excess surplus contrast \([Q_k(x) - Q_K(x)] m\) 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 \(\ln \left( \frac{p_k(x)}{p_K(x)} \right)\).

A shift in the expectation contrast \(Q_k(x) - Q_K(x)\) 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 \([Q_k(x) - Q_K(x)] m = 0\). 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 \(\beta\) 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 \(v_K\), which by (4) equals

\[
v_K = [I - \beta Q_K]^{-1} [u_K + \beta Q_K m].
\]

(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,

\[ \ln p_k - \ln p_K = \beta [Q_k - Q_K] [m + v_K] + u_k - u_K. \]  

(9)

Instead of controlling the contribution of \( v_K \) 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

\[ \ln p_k - \ln p_K = \beta [Q_k - Q_K] [I - \beta Q_K]^{-1} [m + u_K] + u_k - u_K. \]  

(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 \( u_K \). 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 \( u_K = 0 \). 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 Kalouptsidi et al.). In Appendix A, 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 \( U^*_K \).

Now suppose that we know the value of \( U^*_k(\tilde{x}_1) - U^*_l(\tilde{x}_2) \) for some known choices \( k \in D/\{K\} \) and \( l \in D \) and known states \( \tilde{x}_1 \in X \) and \( \tilde{x}_2 \in X \); with either \( k \neq l \), \( \tilde{x}_1 \neq \tilde{x}_2 \), or both. For expositional convenience only (see Appendix A for the general case),

\footnote{Note that this normalization does not collapse Magnac and Thesmar’s exclusion restriction on current values to an easily interpretable restriction on primitives.}

\footnote{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.}
we take this known value to be zero, and simply focus on the exclusion restriction

\[ u_k^*(\tilde{x}_1) = u_l^*(\tilde{x}_2). \]  

(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

\[
\ln \left( \frac{p_k(\tilde{x}_1)}{p_K(\tilde{x}_1)} \right) - \ln \left( \frac{p_l(\tilde{x}_2)}{p_K(\tilde{x}_2)} \right) = \beta \left[ Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1) - Q_l(\tilde{x}_2) + Q_K(\tilde{x}_2) \right] \left[ I - \beta Q_K \right]^{-1} \mathbf{m}. 
\]

(12)

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

Unlike the right hand side of (6), the right hand side of (12) is not linear in \( \beta \). Nevertheless, given data on transition and choice probabilities, it is a well-behaved, function of \( \beta \). 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:

\[
[Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1) - Q_l(\tilde{x}_2) + Q_K(\tilde{x}_2)] \mathbf{m} \neq 0. 
\]

(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 \( I - \beta Q_K \) exists for \( \beta \in (-1, 1) \) and equals

\[
I + \beta Q_K + \beta^2 Q_K^2 + \cdots. 
\]

\(^6\)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 Section 4).
This is trivial for $\beta = 0$. If $|\beta| \in (0, 1)$, it follows from the facts that $|\beta^{-1}| > 1$ and that $Q_K$ is a Markov transition matrix, with eigenvalues no larger than 1 in absolute value. Consequently, the determinant of $Q_K - \beta^{-1}I$ is nonzero, so that $I - \beta Q_K$ 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 $\beta$ that converges on $(-1, 1)$. Denote the function of $\beta$ 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 $B^* = \{ \beta \in (-1, 1) \mid f(\beta) = 0 \}$. Note that $B = B^* \cap [0, 1)$. First, suppose that $f$ has no zeros ($B^* = \emptyset$). Then, $B = \emptyset$ has no limit point in $[0, 1)$.

Finally, suppose that $f$ has at least one zero ($B^* \neq \emptyset$). 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 $B^*$ has no limit point in $(-1, 1)$ (Krantz and Parks, Corollary 1.2.7). Because $B = B^* \cap [0, 1)$, this implies that $B$ has no limit point in $[0, 1)$.

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, $B \cap [0, 1 - \epsilon]$ 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.

From the proof of Theorem 1, the right hand side of (12) equals $\beta$ times the sum of two terms,

$$[Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1) - Q_l(\tilde{x}_2) + Q_K(\tilde{x}_2)]m$$

(15)
and

\[
\begin{align*}
[Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1) - Q_l(\tilde{x}_2) + Q_K(\tilde{x}_2)] v_K = \\
[Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1) - Q_l(\tilde{x}_2) + Q_K(\tilde{x}_2)] \left[ \beta Q_K + \beta^2 Q_K^2 + \cdots \right] m.
\end{align*}
\]  

(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 \(\beta\). The two terms add up to a shift in the expected surplus contrast that depends on \(\beta\). In the special case that \(k = l\), this shift can again be interpreted as a change in the incentives for 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 \(\beta\), its effect on the log choice probability ratio does not directly identify \(\beta\).

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).\footnote{Here, \(\beta\) is locally identified at some \(\beta_0\) if \(\beta = \beta_0\) uniquely solves (12) in a neighborhood of \(\beta_0\). Magnac and Thesmar’s rank condition is not necessary for local identification of \(\beta\) at zero; for that, higher order variation of the right hand side of (12) in \(\beta\) at zero would suffice (Sargan, 1983).} Now, 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 an infinite power series in \(\beta\) (a real-analytic function of \(\beta\)) 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).
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. Indeed, there is no solution if \(J = 2, K = 2\), and choice probabilities are state dependent (see Section 4); and zero, one, or two solutions in the examples in the next subsection, many of which have \(J = 3\) and \(K = 2\).

### 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 would imply 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 case with \(k \neq l = K\), (16) reduces to

\[
[Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1)] v_K = [Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1)] \left[ \beta Q_K + \beta^2 Q_K^2 + \cdots \right] m. \tag{17}
\]

Suppose that \(Q_k(\tilde{x}_1) Q_K^\rho = Q_K(\tilde{x}_1) Q_K^\rho\) for some \(\rho \in \{1, 2, \ldots\}\). 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 \(\rho\) periods (independently of whether this is optimal or not). Under this “single action \((K)\) \(\rho\)-period dependence” (Arcidiacono and Miller, 2017) on choices \(k\) and \(K\) in state \(\tilde{x}_1\), \(Q_k(\tilde{x}_1) Q_K^r = Q_K(\tilde{x}_1) Q_K^r\) for all \(r \in \{\rho, \rho + 1, \ldots\}\).\(^8\) Now assume that Theorem 1’s conditions hold. If \(\rho = 1\), the right hand side of (17) equals zero, the current value

\[
U_k(\tilde{x}_1) = u_k^*(\tilde{x}_1) + \beta [Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1)] v_K = u_k^*(\tilde{x}_1),
\]

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

\[
[Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1)] \left[ \beta Q_K + \cdots + \beta^{\rho-1} Q_K^{\rho-1} \right] m,
\]

the right hand side of (12) is a \(\rho\)-th order polynomial in \(\beta\), and the identified set \(\mathcal{B}\) holds no more than \(\rho\) discount factors. This example straightforwardly extends to

---

\(^8\)Throughout, 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 general exclusion restriction in (11), which we state without further proof.

**Theorem 2.** Suppose that the conditions of Theorem 1 hold and that \( \{ Q_k; k \in D \} \) satisfies single action \((K)\) \(\rho\)-period dependence on choices \(k\) and \(K\) in state \(\tilde{x}_1\),

\[
Q_k(\tilde{x}_1)Q_K^\rho = Q_K(\tilde{x}_1)Q_K^\rho,
\]

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

\[
Q_l(\tilde{x}_2)Q_K^\rho = Q_K(\tilde{x}_2)Q_K^\rho,
\]

for some \(\rho \in \{1, 2, \ldots\}\). Then there are no more than \(\rho\) points in the identified set \(B\).

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

\[
[Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1) - Q_k(\tilde{x}_2) + Q_K(\tilde{x}_2)] [\beta Q_K + \beta^2 Q_K^2 + \cdots] m.
\]

(18)

By Theorem 2, single action \((K)\) \(\rho\)-period dependence on choices \(k\) and \(K\) in states \(\tilde{x}_1\) and \(\tilde{x}_2\) implies that the identified set contains at most \(\rho\) 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,

\[
Q_k(\tilde{x}_1)Q_K^\rho = Q_k(\tilde{x}_2)Q_K^\rho \quad \text{and} \quad Q_K(\tilde{x}_1)Q_K^\rho = Q_K(\tilde{x}_2)Q_K^\rho,
\]

for some \(\rho \in \{1, 2, \ldots\}\). This is a form of single action \((K)\) \(\rho\)-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. In some special cases, the dis-
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 $x$ 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 $c(x)$.

count 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. This will be the case if the state exhibits single action ($K$) one-period dependence on initial choices, as in the following renewal and optimal stopping examples.

**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 (1987) observe, and an additive and independent shock. Renewal incurs a cost that is independent of mileage and resets mileage to $x_1 = 0$:

$$Q_K = \begin{bmatrix} 1 & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 1 & \ldots & 0 & 0 \end{bmatrix}.$$  

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, 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
\( x \in X \). Since \( v_{K}^{\ast}(\tilde{x}) \) does not vary with \( \tilde{x} \), \([Q_{1} - Q_{K}]v_{K} = 0\), and \( U_{1}(\tilde{x}) = u_{1}^{\ast}(\tilde{x})\). Therefore, if \( u_{1}^{\ast}(\tilde{x}_{1}) = u_{1}^{\ast}(\tilde{x}_{2})\), Magnac and Thesmar’s exclusion restriction holds and its identification result applies. Its rank condition (7) simplifies to

\[
\begin{align*}
[Q_{1}(\tilde{x}_{1}) - Q_{1}(\tilde{x}_{2})]m & \neq 0.
\end{align*}
\]

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 \( u_{1}^{\ast}(\tilde{x}_{1}) = u_{1}^{\ast}(\tilde{x}_{2})\) 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 \( u_{1}^{\ast}(x) = a - c(x) \), for some constant \( a \) (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.\(^{9}\) 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 \( u_{1}^{\ast}(x_{1}) = u_{1}^{\ast}(x_{2}) \). 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 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. So 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. Suppose that the renewal costs and the operating benefits and costs are known for some mileage \( \tilde{x} \). This implies that \( u_{1}^{\ast}(\tilde{x}_{1}) - u_{k}^{\ast}(\tilde{x}_{2}) \) is known, but not necessarily zero, for \( l = 1, k = K \), and \( \tilde{x}_{1} = \tilde{x}_{2} = \tilde{x} \). In Appendix A, we demonstrate that our identification analysis extends to this case. Hence, the discount factor is identified.

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. Since mileage is the only state variable of the problem, identification could not be obtained by excluding mileage from the

---

\(^{9}\)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 \( i > j \).
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 \( v_K = 0 \). As in Example 1, the fact that \( v_k^*(\hat{x}) \) is constant in \( \hat{x} \) ensures that the expectation contrast \( [Q_1 - Q_K]v_K = 0 \), so that \( U_1(\hat{x}) = u_1^*(\hat{x}) \).

Of course, \( [Q_1 - Q_K]v_K \) may equal zero even if \( v_k^*(\hat{x}) \) varies with \( \hat{x} \), in particular if the state is single action (\( K \)) one-period dependent on choices 1 and \( K \).

**Example 2.** Consider a discrete time, econometric implementation of Dixit’s (1989) model of firm entry and exit. 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 \( u_K = 0 \), 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 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

\[
Q_1(\hat{x})v_K = \mathbb{E}[v_K^*(y', 1) | y = \hat{y}] = \mathbb{E}[v_K^*(y', K) | y = \hat{y}] = Q_K(\hat{x})v_K \tag{19}
\]

for all \( \hat{x} = (\tilde{y}, \tilde{d}_{-1}) \in \mathcal{X} \). Consequently, as in Example 1, \( [Q_1(\hat{x}) - Q_K(\hat{x})]v_K = 0 \) and \( U_1(\hat{x}) = u_1^*(\hat{x}) \). 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 \( u_1^*(\hat{x}_1) = u_1^*(\hat{x}_2) \) implies (6) and, under rank condition (7), point identification of \( \beta \). Because \( y \) evolves independently of current and past

---

10Section 5 discusses exclusion of state variables, which typically leads to multiple exclusion restrictions like (11).

11Abbring 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.
choices,

$$Q_k(\tilde{x})m = \mathbb{E}[m(y', k) | y = \tilde{y}] .$$

(20)

Thus, the rank condition is equivalent to

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

(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

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

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 (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 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

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

so that $\beta$ is identified if $\mathbb{E}[m(y', 1) - m(y', K) | y = \tilde{y}] \neq 0$. 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 $\beta$ in the right hand side of (12) is nonzero. Because these examples satisfy one-period dependence, this shift does not depend on $\beta$ itself, and this suffices for point identification. Even if the state is not one-period dependent and the shift does depend on $\beta$, 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 $\beta$. 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 $Q_K(x_3)$ 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 $\beta$. 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.
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 $\beta$. The data are $Q_1(x_1) = [0.25\ 0.25\ 0.50]$, $Q_1(x_2) = [0.00\ 0.25\ 0.75]$, $Q_K = \begin{bmatrix} 0.90 & 0.00 & 0.10 \\ 0.00 & 0.90 & 0.10 \\ 0.00 & 1.00 & 0.00 \end{bmatrix}$, $p_1 = [0.50\ 0.49\ 0.10]$, and $p_K = [0.50\ 0.51\ 0.90]$. 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, $m' = [0.69\ 0.67\ 0.11]$ and $Q_1(x_1) - Q_K(x_1) - Q_1(x_2) + Q_K(x_2) = [-0.65\ 0.90\ -0.25]$, so that the slope of the dashed red line equals $|Q_1(x_1) - Q_K(x_1) - Q_1(x_2) + Q_K(x_2)| m = 0.1291$. A unique value of $\beta$, 0.31, solves (6), but two values of $\beta$ solve (12): 0.34 and 0.95.

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 $\beta$. 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 $\beta$ 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
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, $k = l = \bar{x}_1 = x_1$, and $\bar{x}_2 = x_2$, 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 $\beta$. The data are $Q_1(x_1) = \begin{bmatrix} 0.00 & 0.25 & 0.75 \end{bmatrix}$, $Q_1(x_2) = \begin{bmatrix} 0.25 & 0.25 & 0.50 \end{bmatrix}$, $Q_K = \begin{bmatrix} 0.00 & 1.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$, $p_1 = \begin{bmatrix} 0.50 \\ 0.50 \end{bmatrix}$, and $p_K = \begin{bmatrix} 0.50 \\ 0.50 \end{bmatrix}$.

Consequently, the left hand side of (6) and (12) equals $\ln \left( \frac{p_1(x_1)}{p_K(x_1)} \right) - \ln \left( \frac{p_1(x_2)}{p_K(x_2)} \right) = 0.0800$. Moreover, $m' = \begin{bmatrix} 0.69 & 0.65 & 0.69 \end{bmatrix}$ and $Q_K(x_1) - Q_1(x_1) - Q_K(x_2) + Q_1(x_2) = \begin{bmatrix} -0.25 & 0.00 & 0.25 \end{bmatrix}$, so that the slope of the dashed red line equals $[Q_1(x_1) - Q_K(x_1) - Q_1(x_2) + Q_K(x_2)]m = 0.0000$. A unique value of $\beta$, 0.90, solves (12), but (6) has no solution.

return to this point in Section 4.

More generally, strict monotonicity of the right hand side of (12), as in Example 5, 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 $\beta$. Without loss of generality—we can freely interchange states $\bar{x}_1$ and $\bar{x}_2$ and switch choices $k$ and $l$—we focus on conditions under which it is strictly increasing or, equivalently, its derivative with
respect to $\beta$ is positive:\(^{12}\)

$$[Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1) - Q_l(\tilde{x}_2) + Q_K(\tilde{x}_2)] [I - \beta Q_K]^{-2} \mathbf{m} > 0.$$  

For this, it suffices that

$$[Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1) - Q_l(\tilde{x}_2) + Q_K(\tilde{x}_2)] Q_K^r \mathbf{m} \geq 0 \text{ for all } r \in \{0, 1, 2, \ldots\}, \quad (22)$$

with the inequality strict for at least one $r$. Like Magnac and Thesmar’s rank condition (7), these conditions do not depend on $\beta$. 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.** 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 4 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.

---

\(^{12}\)Denoting $\Delta^2 Q \equiv Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1) - Q_l(\tilde{x}_2) + Q_K(\tilde{x}_2)$, we have that

$$\frac{\partial}{\partial \beta} \Delta^2 Q [I - \beta Q_K]^{-1} \mathbf{m} = \Delta^2 Q [I - \beta Q_K]^{-1} \mathbf{m} + \beta \Delta^2 Q \frac{\partial [I - \beta Q_K]^{-1}}{\partial \beta} \mathbf{m} = \Delta^2 Q [I - \beta Q_K]^{-1} \mathbf{m} + \beta \Delta^2 Q [I - \beta Q_K]^{-1} Q_K [I - \beta Q_K]^{-1} \mathbf{m} = \Delta^2 Q [I - \beta Q_K]^{-1} \beta Q_K [I - \beta Q_K]^{-1} \mathbf{m} = \Delta^2 Q [I - \beta Q_K]^{-2} \mathbf{m}.$$
Figure 4: 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 $\beta$ (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 6’s stylized dynamic labor supply model, which gives $Q_1(x_2) = \begin{bmatrix} 0.00 & 0.25 & 0.75 \end{bmatrix}$, $Q_1(x_1) = \begin{bmatrix} 0.25 & 0.75 & 0.00 \end{bmatrix}$, $Q_K = \begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.50 & 0.50 & 0.00 \\ 0.00 & 0.50 & 0.50 \end{bmatrix}$, $p_1 = \begin{bmatrix} 0.44 \\ 0.56 \\ 0.71 \end{bmatrix}$, and $p_K = \begin{bmatrix} 0.56 \\ 0.44 \\ 0.29 \end{bmatrix}$.

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, $m' = \begin{bmatrix} 0.57 & 0.82 & 1.23 \end{bmatrix}$ and $Q_K(x_2) - Q_1(x_1) + Q_K(x_1) = \begin{bmatrix} -1.00 & 0.75 \end{bmatrix}$, so that the slope of the dashed red line equals $\begin{bmatrix} 0.25 & -1.00 & 0.75 \end{bmatrix}m = 0.2465$. A unique value of $\beta = 0.80$, solves (12), but (6) has no solution.

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 $Q_K$ is increasing, the expected excess surplus after $r$ rounds of shirking and human capital depreciation, $Q_K^r m$, 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. In particular, experience is not single action ($K$) one-period dependent 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.

\[\text{In a similar context, Altuğ and Miller (1998) impose such finite dependence by assuming that wages and the utility cost from work only depend on a finite employment history. Our example would display single action ($K$) one-period dependence on initial choices in state $x_1$ and two-period dependence in state $x_2$ if shirking women would for sure see their experience drop by one level. Note that this would still not suffice to reduce the moment condition to Magnac and Thesmar’s linear moment condition.}\]
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 (\([Q_1(x_2) - Q_K(x_2) - Q_1(x_1) + Q_K(x_1)] = [0.25 -1.00 0.75]\)) and \(Q^r_K m\) is increasing for all \(r\), (22) holds. Consequently, the moment condition is monotone in \(\beta\) 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, 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.\(^{14}\)

Denote time by \(t \in \{1, 2, \ldots, T\}\), with terminal period \(T < \infty\), and index \(u^*_{k,t}\), \(u_{k,t}\), \(m_t\), and \(v_{k,t}\) by time. For ease of exposition, we maintain the assumption of stationary Markov transition matrices \(Q_k\), but the results extend to nonstationary distributions. The choice-\(k\) specific values now satisfy

\[
v_{k,t} = u_{k,t} + \beta Q_k [m_{t+1} + v_{t+1}] \tag{23}
\]

for \(t = 1, \ldots, T - 1\); with terminal condition \(v_{k,T} = u_{k,T}\). With the normalization \(u_{k,t} = 0\) for all \(t\), this gives

\[
\ln (p_{k,t}(\tilde{x})) - \ln (p_{K,t}(\tilde{x})) = u^*_{k,t}(\tilde{x}) + \beta [Q_k(\tilde{x}) - Q_K(\tilde{x})][m_{t+1} + v_{K,t+1}] \tag{24}
\]

for all \(k \in D \setminus \{K\}\) and \(\tilde{x} \in \mathcal{X}\). Finally, using (23) and the normalization \(u_{K,t} = 0\)

\(^{14}\)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.
for all \( t \), we can write the value of the reference choice \( K \) as

\[ v_{K,t} = \sum_{\tau = t + 1}^{T} (\beta Q_K)^{\tau - t} m_{\tau}, \tag{25} \]

where we use the convention that \( \sum_{\tau = T + 1}^{T} = 0 \) (so that indeed \( v_{K,T} = u_{K,T} = 0 \)).

**Theorem 3.** Suppose that

\[ u_{k,t}^*(\tilde{x}_1) = u_{l,t'}^*(\tilde{x}_2) \tag{26} \]

for \( k \in D/\{K\}, \ l \in D, \ \tilde{x}_1 \in X, \ \tilde{x}_2 \in X, \ 1 \leq t' < T, \ and \ t' \leq t \leq T; \) 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

\[ [Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1)] m_{t+1} - [Q_l(\tilde{x}_2) - Q_K(\tilde{x}_2)] m_{t'+1} \neq 0, \tag{27} \]

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

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

\[
\ln \left( \frac{p_{k,t}(\tilde{x}_1)}{p_{K,t}(\tilde{x}_1)} \right) - \ln \left( \frac{p_{l,t'}(\tilde{x}_2)}{p_{K,t'}(\tilde{x}_2)} \right) = \beta \left( [Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1)] \left[ \sum_{\tau = t + 1}^{T} (\beta Q_K)^{\tau - t} m_{\tau} \right] - [Q_l(\tilde{x}_2) - Q_K(\tilde{x}_2)] \left[ \sum_{\tau = t' + 1}^{T} (\beta Q_K)^{\tau - t'} m_{\tau} \right] \right). \tag{28}
\]

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 \( \beta \). 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.

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 \(\rho\)-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, in the sense that it can rationalize any choice data \(\{p_k, Q_k; k \in D\}\), and can therefore not be tested using such data. This section demonstrates that exclusion restrictions give the model empirical content and can, to some extent, be verified in data.

The standard result for the unrestricted stationary model follows from a version of Magnac and Thesmar’s Proposition 2: For any given data \(\{p_k, Q_k; k \in D\}\), \(u_K = 0\), and \(\beta \in [0, 1]\), there exists a unique set of primitive utilities \(\{u_k, k \in D/\{K\}\}\) that rationalizes the data. Specifically, \(m = -\ln p_K\). Then, \(v_K\) follows from \(u_K = 0\) and (8). Next, by (3), \(v_k = v_K + \ln p_k - \ln p_K\) for \(k \in D/\{K\}\) 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 \(u_k = v_k - \beta Q_k [m + v_K]\) for \(k \in D/\{K\}\) (note that \(v_K\) was already set to be consistent with \(u_K = 0\)).

This result justifies our focus on the identification of the discount factor \(\beta\) in the previous section: Once the discount factor has been 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 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 recovers 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 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 primitives, at least not in their theoretical domains (such as [0, 1) for the discount factor), that are compatible with both the data and the assumed exclusion restriction. The subset of the data than 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 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.

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 5 cannot be rationalized for a discount factor in its domain under the current value restriction. In that example, the data are however consistent with an exclusion restriction on primitive utilities.

Conversely, there exist data that are inconsistent with the primitive utility restriction, yet can be rationalized by primitives that satisfy the current value restriction.

Example 7. Figure 5 displays the left and right hand sides of Magnac and Thesmar’s moment condition in (6) and ours in (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 5 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.

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 $\beta$) for general choice and state probabilities.

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 $k \in D\setminus\{K\}$, $\tilde{y}$, $\tilde{z}_1$, and $\tilde{z}_2 > \tilde{z}_1$. 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 $y$.

\footnote{We provide a more formal statement of the exclusion of state variables in our discussion of Fang and Wang in Appendix B.}
Figure 6: Example with Two Moment Conditions of Which One Identifies the Discount Factor

![Graph with two moment conditions and values]

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 $\beta$, 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

$$Q_1 = \begin{bmatrix} 0.40 & 0.26 & 0.18 & 0.18 \\ 0.33 & 0.29 & 0.36 & 0.27 \\ 0.19 & 0.26 & 0.18 & 0.45 \\ 0.08 & 0.18 & 0.29 & 0.09 \end{bmatrix}, \quad Q_K = \begin{bmatrix} 0.17 & 0.26 & 0.13 & 0.43 \\ 0.13 & 0.07 & 0.20 & 0.60 \\ 0.20 & 0.30 & 0.10 & 0.40 \\ 0.25 & 0.15 & 0.50 & 0.10 \end{bmatrix},$$

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

Consequently, the left hand sides of (12) equal $\ln \left( \frac{p_1(x_1)}{p_K(x_1)} \right) - \ln \left( \frac{p_1(x_2)}{p_K(x_2)} \right) = 0.0187$ and $\ln \left( \frac{p_1(x_3)}{p_K(x_3)} \right) - \ln \left( \frac{p_1(x_4)}{p_K(x_4)} \right) = 0.0045$. A unique value of $\beta$, 0.30, solves (12) for $\tilde{x}_1 = x_1$ and $\tilde{x}_2 = x_2$ (dashed red). Two values of $\beta$ 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.

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 6, 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 9.** In Figure 7, 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. In Appendix B, we sketch a proof of this conjecture.

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.\footnote{For 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.}

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 10.** Suppose the population moment conditions are as given in Figure 7. Though each individual moment condition is equally consistent with one small 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 8, 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 $S : [0, 1) \to [0, \infty)$. 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 $\alpha$-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 $\alpha$:

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

The bottom panel of Figure 8 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 $\alpha$-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, standard inference for extremum estimators applies (e.g. Newey and McFadden, 1994).
Figure 7: 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 $\beta$, 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 $\beta$ (in multiples of $10^{-4}$). The data are

$$Q_1 = \begin{bmatrix} 0.43 & 0.26 & 0.18 & 0.18 \\ 0.33 & 0.29 & 0.36 & 0.27 \\ 0.19 & 0.26 & 0.18 & 0.45 \\ 0.05 & 0.18 & 0.29 & 0.09 \end{bmatrix}, \quad Q_K = \begin{bmatrix} 0.17 & 0.26 & 0.13 & 0.43 \\ 0.13 & 0.07 & 0.20 & 0.60 \\ 0.20 & 0.30 & 0.10 & 0.40 \\ 0.25 & 0.15 & 0.50 & 0.10 \end{bmatrix},$$

$$p_1' = [0.92 \quad 0.92 \quad 0.63 \quad 0.63], \quad \text{and} \quad p_K' = [0.08 \quad 0.08 \quad 0.37 \quad 0.37].$$

Consequently, the left hand sides of (12) equal $\ln(\frac{p_1(x_1)}{p_K(x_1)}) - \ln(\frac{p_1(x_2)}{p_K(x_2)}) = 0.0068$ and $\ln(\frac{p_1(x_3)}{p_K(x_3)}) - \ln(\frac{p_1(x_4)}{p_K(x_4)}) = 0.0019$. A unique value of $\beta$, 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 8: 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 7 for the same values of $Q_1$ and $Q_K$, but randomly perturbed values of its choice probabilities $p_1$ and $p_K$. Rounded to two digits, the perturbed choice probabilities equal those reported below Figure 7. Consequently, the perturbation to $m = -\ln p_K$ is very small too, so that the right hand sides of (12) are very close to those plotted in Figure 7. 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 $\beta$ at which the distance is below some critical level $s_n$ (which is taken to be $0.10 \times 10^{-4}$ in this example).
Appendices

A  Identification with general reference utility

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

$$\ln p_k - \ln p_K = \beta [Q_k - Q_K] [I - \beta Q_K]^{-1} (m + \bar{u}_K) + u_k - \gamma 1 - \bar{u}_K. \quad (29)$$

Note that the constant additive shift $\gamma 1$ 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 $k \in D / \{K\}$ and $l \in D$, 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

$$\ln \left(\frac{p_k(\tilde{x}_1)}{p_K(\tilde{x}_1)}\right) - \ln \left(\frac{p_l(\tilde{x}_2)}{p_K(\tilde{x}_2)}\right) - \Delta^2 u = \beta \left[Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1) - Q_l(\tilde{x}_2) + Q_K(\tilde{x}_2)\right] [I - \beta Q_K]^{-1} \bar{m}, \quad (30)$$

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 $\bar{m} \equiv m + \bar{u}_K$ known. The factor multiplying $\beta$ 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 $\beta$ 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 $\beta$, 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 $k \in D / \{K\}$, $l \in D$, $\tilde{x}_1 \in \mathcal{X}$, 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:

$$[Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1) - Q_l(\tilde{x}_2) + Q_K(\tilde{x}_2)] \bar{m} \neq 0.$$
Then, the identified set $B$ 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.

**B Relation to Fang and Wang (2015)**

Fang and Wang considered generic identification of a dynamic discrete choice model with hyperbolic discounting. Time preferences in their model are represented by three parameters: a standard discount factor $\beta$, a present bias factor $\delta$, and a partial naivety parameter $\tilde{\delta}$.\(^1\) Our model, in which agents discount exponentially with factor $\beta$, follows as a special case if agents have no present bias ($\delta = 1$) and are aware of that ($\tilde{\delta} = 1$).

Like this paper, Fang and Wang normalized $u_K = 0$, specified that the utility shocks $\varepsilon_1, \ldots, \varepsilon_K$ are independent with type-1 extreme value distributions, and assumed that agents form expectations about the state’s evolution using the actual

\(^1\)In line with the literature on hyperbolic discounting, Fang and Wang instead denoted these parameters with $\delta$, $\beta$, and $\tilde{\beta}$. We have exchanged $\beta$ and $\delta$ to ensure consistency with the notation in our paper, which follows e.g. Magnac and Thesmar. More generally, throughout this appendix, we use notation that is consistent with the main text and that may deviate from Fang and Wang’s.
controlled transition probabilities \(Q_k, k \in D\). Therefore, its model has \((K - 1)J + 3\) free parameters: the \((K - 1)J\) remaining utilities in \(u_k, k \in D/\{K\}\), and the parameters \(\beta, \delta, \tilde{\delta}\) of the discount function.\(^{18}\)

An application of Hotz and Miller’s choice probability inversion to Fang and Wang’s model gives \((K - 1)J\) equations that relate the \((K - 1)J + 3\) parameters to the data (as in this paper, \(\{P_k, Q_k; k \in D\}\)), one for each log choice probability contrast \(\ln p_k(\tilde{x}) - \ln P_k(\tilde{x}), k \in D/\{K\}\), \(\tilde{x} \in X\). Fang and Wang provided the analysis leading to these equations, but not the final equations themselves. They are the equivalent of (10) for Fang and Wang’s extension to hyperbolic discounting.

Like this paper, Fang and Wang studied identification under additional exclusion restrictions. Specifically, it assumed that the state \(x\) contains a variable that affects state transition probabilities, and this way beliefs over future states, but not current utilities. To this end, it supposed that \(x = (y, z)\), where \(y\) has support \(Y\) with \(J_y\) elements, \(z\) has support \(Z\) with \(J_z\) elements, \(X = Y \times Z\) has \(J = J_yJ_z\) elements, and \(z\) does not affect utility.\(^{19}\) This gives \((K - 1)J_y(J_z - 1)\) exclusion restrictions like (11): For all \(\tilde{y} \in Y\), \(\tilde{z}_1 \in Z\), and \(\tilde{z}_2 \in Z\) such that \(\tilde{z}_2 > \tilde{z}_1\),

\[
u_k^*(\tilde{y}, \tilde{z}_1) = u_k^*(\tilde{y}, \tilde{z}_2) \quad \text{for all } k \in D/\{K\}.\tag{31}\]

Altogether, this gives \((K - 1)J + (K - 1)J_y(J_z - 1)\) “identifying” equations in \((K - 1)J + 3\) parameters.

Fang and Wang’s key result, Proposition 2, states that its model is generically identified if the number of equations exceeds the number of parameters. This requires that \((K - 1)J_y(J_z - 1) \geq 4.\(^{20}\) Fang and Wang does not formally define “generic identification,” but paraphrases Proposition 2 as giving identification “for almost all data sets generated by the assumed hyperbolic discounting model” (p. 579). Fang and Wang’s Online Appendix C further clarifies which measure the term “almost all” belongs to. Because probabilities sum to one, the \(KJ + KJ^2\) choice and transition probabilities in \(\{P_k, Q_k; k \in D\}\) can be represented by a vector \(b \in B \subset \mathbb{R}^{(K - 1)J + KJ(J - 1)}\) that stacks \((K - 1)J + KJ(J - 1)\) of these probabilities.\(^{21}\) Given this representation, Fang and Wang implicitly defines “for almost all

---

\(^{18}\)Fang and Wang denoted \(K - 1\) with \(I\) and \(J\) with \(X\) and wrote about \(I(X + 3)\) parameters.

\(^{19}\)Fang and Wang denoted \(y\) with \(x_r\), \(z\) with \(x_e\), \(Y\) with \(X_r\), \(Z\) with \(X_e\), \(J_y\) with \(|X_r|\), and \(J_z\) with \(|X_e|\). We do not adopt this notation because, following Magnac and Thesmar, we have already used \(x_j\) for the \(j\)-th support point of \(x\).

\(^{20}\)Fang and Wang incorrectly counted equations and give a slightly different condition; see (ii) below. This is irrelevant to our main concern about this proposition, which we will discuss first.

\(^{21}\)Fang and Wang does not specify such a representation. However, from the table on page 3 of Online Appendix C it is clear that Fang and Wang thought of \(B\) as spanning the full set of choice and transition probabilities, and not just those in the model’s range, and agreed that
data sets” to mean for all \( b \in B \) outside a subset of \( \mathbb{R}^{(K-1)J+KJ(J-1)} \) with Lebesgue measure zero.\(^{22}\)

At first glance, Fang and Wang’s Proposition 2 may seem to establish a result that is related to our Theorem 1. However, Fang and Wang’s proof of Proposition 2 implies that its key generic identification claim is void. In particular, the logical consequence of this proof up to its very last sentence is that the \((K-1)J+(K-1)J_y(J_z-1) \geq (K-1)J+4 \) identifying equations cannot be solved for the \((K-1)J+3 \) parameters, except for data \( b \in B \) in a subset of \( \mathbb{R}^{(K-1)J+KJ(J-1)} \) that has Lebesgue measure zero.\(^{23}\) That is, the range of the data that the model can generate if at least four exclusion restrictions are imposed has zero measure. Consequently, generic identification, which Fang and Wang defines using this same measure, allows the exception to cover all data that can possibly be generated by the model. Therefore, it does not preclude that the model is nowhere identified.

In the last sentence of the proof of Proposition 2, Fang and Wang indeed concludes that, generically, the identifying equations have no solutions, but qualifies the conclusion with “except the true primitives ... that generated the data.” This qualification does not follow from the preceding mathematical arguments and is incorrect. The identifying equations only have solutions for \( b \in B \) in the subset that Fang and Wang’s application of the transversality theorem excepts, which has Lebesgue measure zero. Moreover, the transversality theorem has no implications for the number of solutions in this subset.

Our argument assumes that Fang and Wang’s proof is otherwise correct. As we discuss in (iii) below, we cannot judge this decisively since the proof is incomplete.\(^{22}\)

\(^{22}\)In particular, Fang and Wang’s Theorem 1 (the transversality theorem) formalizes “generic” and “almost all” as “except for a set of \( b \in B \) of Lebesgue measure zero” and applies this theorem with \( B \subset \mathbb{R}^{(K-1)J+KJ(J-1)} \) the space of choice and transition probabilities. Note that the exact way the data are represented in \( \mathbb{R}^{(K-1)J+KJ(J-1)} \) is irrelevant here, as Lebesgue measure is invariant under affine transformations with determinant 1 or -1. In particular, each representation assigns zero measure to the same sets of choice and transition probabilities.

\(^{23}\)Let \( \hat{\theta} \in \mathbb{R}^{(K-1)J+3} \) collect the unknown parameters: \( \mathbf{u}_k, k \in \mathcal{D}/\{K\} \), and \((\beta, \delta, \hat{\delta})\) (these are the free parameters in Fang and Wang’s “structure” \( \theta \)). Fang and Wang’s Proposition 2 assumes that these are related to the data \( b \in B \) in a system of at least \((K-1)J+4 \) equations \( \hat{\mathcal{G}}(\theta, b) = \mathbf{0} \). The main arguments of its proof are given in the three paragraphs on page 3 of Online Appendix C. The first paragraph verifies that \( \hat{\mathcal{G}} \) satisfies the conditions of the transversality theorem (Theorem 1). In particular, it claims that it can be verified that the Jacobian matrix \( \partial \hat{\mathcal{G}}(\theta, b) \) with partial derivatives of \( \hat{\mathcal{G}} \) with respect to both its arguments has rank equal to the number of equations in \( \hat{\mathcal{G}} \). The second and third paragraphs subsequently appeal to the transversality theorem to conclude that, for all \( b \in B \) outside a subset of \( \mathbb{R}^{(K-1)J+KJ(J-1)} \) that has Lebesgue measure zero, \( \partial \theta \hat{\mathcal{G}}(\theta, b) \) has this same rank, which is at least \((K-1)J+4 \), whenever \( \theta \) solves \( \hat{\mathcal{G}}(\theta, b) = \mathbf{0} \). It then notes that \( \partial \theta \hat{\mathcal{G}}(\theta, b) \) cannot have rank \((K-1)J+4 \) or higher, because it only has as many columns as parameters, \((K-1)J+3 \). This implies that, for all \( b \in B \) outside a subset of \( \mathbb{R}^{(K-1)J+KJ(J-1)} \) that has Lebesgue measure zero, there exists no \( \theta \) that solves \( \hat{\mathcal{G}}(\theta, b) = \mathbf{0} \).
However, the key result that, generically, the identifying equations have no solutions if there are more equations than unknowns, and its possible proof using the transversality theorem, are intuitive. This result is reflected in Section 5’s discussion of inference with two moment conditions (in the context of exponential discounting, two moment conditions suffice to have more equations than unknowns). The empirical analogs of the two identifying equations in Figure 8 do not have a common solution $\beta$, because of independent sampling variation in both equations.

At the root of these problems with Fang and Wang’s Proposition 2 is its focus on identification that is generic in the data space, rather than the parameter space. This is nonstandard and complicates the analysis in two ways. First, the specification of an appropriate measure directly on the data requires knowledge of the model’s empirical content, i.e. the range of data that can be generated by varying the model parameters on their domain. Our discussion of Proposition 2 highlights the perils of ignoring the model’s empirical content in the case in which we have more constraints than unknowns. The discussion in Section 4 of our model with exponential discounting shows that this extends to cases in which we have as many constraints as unknowns. There, we showed that, in the case with binary choice and an exclusion restriction on a binary state variable, the model cannot generate state-dependent choice probabilities. Consequently, that model’s range has measure zero under Fang and Wang’s definition.

Second, genericity in the data space interacts in an opaque way with the property Fang and Wang claims holds generically, identification. An abstract example serves best to illustrates this point. Consider a model that maps a parameter $\theta \in \mathbb{R}$ to a choice probability $p = p(\theta) \in [0, 1]$. Define ”for almost all” $\theta$ (or $p$) to mean for all $\theta$ (or $p$) outside a set of Lebesgue measure 0. If $p(\theta) = 1/(1 + \exp(\theta))$, then $\theta$ is identified for almost all $p$ and almost all $\theta$. If instead $p(\theta)$ equals 0 if $\theta \leq 0$, $\theta$ if $\theta \in (0, 1)$, and 1 if $\theta \geq 1$, then $\theta$ is identified for almost all $p$, but we not for almost all $\theta$.

As we noted in Section 5, we do not think a formal generic identification result would add much practical value to the analysis based on individual moment conditions that we already provide. That being said, we could likely develop such a result, with more substance than Fang and Wang’s Proposition 2, for our model.

We would start by defining “generic” in the parameter space, following e.g. Sargan, McManus (1992), and Ekeland et al. We would then leverage Section 3’s results and concentrate the identification analysis on the scalar discount factor, $\beta$.

\footnote{We would also have to choose between a measure-theoretic definition of “genericity”, like Fang and Wang’s, and a topological one. McManus and Ekeland et al. provided discussion.}
by analyzing two or more moment conditions (12) that relate it to the data. We
would first write the choice probabilities as functions of some “true” unknown pa-
rameters $\beta^\dagger$ and $u^\dagger_k$, $k \in \mathcal{D}/\{K\}$, and the state transition probabilities $Q_k$, $k \in \mathcal{D}$, and use this to rewrite the moment conditions as equations in $\beta^\dagger$; $u^\dagger_k$, $k \in \mathcal{D}/\{K\}$; $Q_k$, $k \in \mathcal{D}$; and $\beta$. Then, one approach would be to fix $Q_k$, $k \in \mathcal{D}$, that meet “rank”
conditions like $Q_k(\tilde{x}_1) - Q_K(\tilde{x}_1) - Q_l(\tilde{x}_2) + Q_K(\tilde{x}_2) \neq 0$ for each assumed exclusion
restriction $u^\dagger_k(\tilde{x}_1) = u^\dagger_l(\tilde{x}_2)$ and apply standard methods to show that generically in
the space of true unknown parameters $\beta^\dagger$ and $u^\dagger_k$, $k \in \mathcal{D}/\{K\}$, the moment condi-
tions have no solution $\beta \neq \beta^\dagger$. Unlike Fang and Wang, we would have imposed that
$\beta = \beta^\dagger$ is actually a solution. This would prove generic identification of the discount
factor for all, and not just almost all, state transition probabilities that meet the
rank conditions.

An alternative approach would be to treat the state transition probabilities like
the other parameters and follow a similar path to prove generic identification in
the space of all parameters, including the transition probabilities. This approach
would not require rank conditions on the state probabilities, because these would be
generically satisfied (in (i) below, we discuss the role of similar conditions in Fang
and Wang’s Proposition 2). Adapting either of these approaches to Fang and Wang’s
model would likely be possible, but nontrivial because, due to the formulation of the
time-inconsistent dynamic program as an intrapersonal game in that model, the true
parameters do not necessarily map into unique choice probabilities.

We end by addressing some more minor issues with Fang and Wang’s analysis.

(i). Fang and Wang’s Proposition 2 does not only require exclusion restrictions
but also requires that each set of $K - 1$ exclusion restrictions (31) comes with
a condition reminiscent of Magnac and Thesmar’s rank condition:25

$$Q_l(\tilde{x}_1) - Q_l(\tilde{x}_2) \neq 0 \quad \text{for some } l \in \mathcal{D}. \quad (32)$$

These conditions only exclude a set of $b \in B$ that has Lebesgue measure
zero. Consequently, they are generically satisfied and cannot be necessary for

---

25Specifically, Proposition 2 supposes that “there exist state variables that satisfy Assumption
5” and Assumption 5 includes both exclusion restrictions and a rank condition. It is somewhat
confusing that Assumption 5 is not a condition on state variables but a condition on two states, $\tilde{x}_1$
and $\tilde{x}_2$ in our notation, which Assumption 5 incorrectly refers to as “state variables.” Consequently,
even though it is clear that Proposition 2 requires the exclusion restrictions specified in (31) and
the sentence preceding it, it is ambiguous which rank conditions it requires. In particular, it is not
made explicit whether choice $l$ in (32) can vary with the states $(\tilde{x}_1, \tilde{x}_2)$ or not. Fortunately, that
does not matter for the argument here. Finally, note that Fang and Wang called Assumption 5 an
“exclusion restriction.” In line with this paper’s analysis, and e.g. Magnac and Thesmar, we will
continue to call each equality in (31) an “exclusion restriction” and will never refer to (32) as one.
Proposition 2's consequent, generic identification, nor for our conclusion that, generically, the identifying equations have no solutions. As an aside, note that a rank condition like (32) does not suffice for nontrivial set identification, let alone point identification. In particular, it does not preclude that \( Q_l(\tilde{x}_1) - Q_K(\tilde{x}_2) - Q_l(\tilde{x}_2) + Q_K(\tilde{x}_2) = 0 \), in which case the right hand side of (12) is zero and this moment condition is not informative about \( \beta \).

(ii). Fang and Wang's proof of Proposition 2 uses that the number of equations exceeds the number of parameters, which requires that \((K - 1)J_yJ_z \geq 4\). Proposition 2, however, only demands that \((K - 1)J_yJ_z \geq 26\). For example, if choices, \( y \), and \( z \) are all binary \((K = J_y = J_z = 2)\), \((K - 1)J_yJ_z = 4\), but \((K - 1)J_yJ_z = 2\). In this case, we have only two exclusion restrictions, one for each possible value of \( \tilde{y} \), which does not suffice. Fang and Wang clearly intended to demand that the number of equations exceeds the number of parameters, but incorrectly counted the equations implied by its exclusion restrictions. It correctly noted that the exclusion restrictions give \((K - 1)J_yJ_z\) additional equations: for all \( k \in D/\{K\} \) and \( \tilde{y} \in Y \), \( u^*_k(\tilde{y}, \tilde{z}) = u^*_k(\tilde{y}) \) for all \( \tilde{z} \in Z \) and some \( u^*_k(\tilde{y}) \). However, it failed to appreciate that these \((K - 1)J_yJ_z\) equations come with \((K - 1)J_y\) additional parameters: \( u^*_k(\tilde{y}) \) for all \( k \in D/\{K\} \) and \( \tilde{y} \in Y \). On balance, these equations only introduce \((K - 1)J_yJ_z - (K - 1)J_y = (K - 1)J_y(\tilde{J}_z - 1)\) additional restrictions. Of course, these restrictions are simply the equalities in (31) that can be derived by differencing Fang and Wang's equations \( u^*_k(\tilde{y}, \tilde{z}_1) = u^*_k(\tilde{y}) \) and \( u^*_k(\tilde{y}, \tilde{z}_2) = u^*_k(\tilde{y}) \) between values \( \tilde{z}_1 \) and \( \tilde{z}_2 > \tilde{z}_1 \) of the excluded variable \( z \) to get rid of the extra parameters \( u^*_k(\tilde{y}) \).

(iii). The proof of Proposition 2 is incomplete. Just below Proposition 2, Fang and Wang noted that its formal proof "involves verifying the conditions for the transversality theorem" and that the "details are available in the online Appendix C," but they are not. It is exactly this omitted part of the proof of Proposition 2 that would show (i) the irrelevance of the rank conditions (32) and (ii) the error in counting the equations.

26Recall that Fang and Wang denoted \((K - 1)J_yJ_z\) with \( I \cdot |X_e| \cdot |X_e| \).

27In the first paragraph of its Online Appendix C, Fang and Wang stated that the exclusion restrictions imply "\( I \times |X_e| \times |X_e| \) equations requiring that \( u_i(x, x_r) = u_i(x, x_r) \) for each \( i \in I \in \mathcal{I} \setminus \{0\} \), each \( x_e \in X_e \) and each \( x_r \in X_e \)." Here, Fang and Wang denoted the action set \( D \) with \( I \) and the reference choice \( K \) with 0.

28In particular, in the notation of Footnote 23, the transversality theorem requires that the rank of \( \partial \hat{G}(\theta, b) \) equals the number of equations whenever \( \hat{G}(\theta, b) = 0 \). Instead of actually verifying this condition, Fang and Wang noted that "this can be verified in the same way that we verify [a similar condition in the proof of Proposition 1]." However, Fang and Wang did not verify the latter condition either, but simply noted that it can be verified (p.578).
(iv). Fang and Wang’s Online Appendix E provides an identification result for the special case of Section 2’s stationary model with exponential discounting. This result takes the discount factor to be known and therefore does not address identification of the full model. It is a version of Magnac and Thesmar’s Proposition 2, which we cite in the second paragraph of Section 4.
References


Magnac, T. and D. Thesmar (2002). Identifying dynamic discrete choice processes. *Econometrica* 70, 801–816. 2, 3, 4, 5, 6, 7, 8, 9, 11, 14, 15, 18, 19, 20, 21, 22, 25, 27, 33, 34, 35, 38, 40


