Partial Identification and Inference in Binary Choice and Duration Panel Data Models*

JASON R. BLEVINS

Department of Economics, Ohio State University

March 28, 2011†

Abstract. Many semiparametric fixed effects panel data models, such as binary choice models and duration models, are known to be point identified when at least one regressor has full support on the real line. It is common in practice, however, to have only discrete or continuous, but possibly bounded, regressors. This paper addresses identification and inference for the identified set in such cases, when the parameters of interest may only be partially identified. We first develop a set of general results for criterion-function-based inference in partially identified models which can be applied to both regular and irregular models. We then apply these general results to several specific models. In the fixed effects binary choice panel data model, we obtain a sharp characterization of the identified set and propose a consistent set estimator, establishing its rate of convergence under different conditions. Rates arbitrarily close to $n^{1/3}$ are possible when a continuous regressor is present. When all regressors are discrete the estimates converge arbitrarily fast to the identified set. We also propose a subsampling-based procedure for constructing confidence regions. Finally, we carry out a series of Monte Carlo experiments to illustrate and evaluate the proposed procedures. We also consider extensions to other fixed effects panel data models such as binary choice models with lagged dependent variables and duration models.

Keywords: partial identification, set estimation, rates of convergence, cube-root asymptotics, panel data, fixed effects, binary choice, duration, transformation.

JEL Classification: C13, C14, C25, C41.

*I am grateful to the members of my dissertation committee, Han Hong, Shakeeb Khan, Paul Ellickson, and Andrew Sweeting, for their guidance and support, to Arie Beresteanu, Federico Bugni, Joseph Hotz, and Xiaoxia Shi for many helpful suggestions, and to seminar participants at Duke University, the University of Illinois at Urbana-Champaign, the University of Pittsburgh, the 2009 Triangle Econometrics Conference and the 2010 Annual Meeting of the Midwest Econometrics Group for useful comments.

†First version: October 18, 2009. This is a preliminary and incomplete draft. Comments are welcome.

1. Introduction

This paper develops set estimators for potentially irregular models in which the parameters of interest are partially identified. It provides general conditions for establishing consistency and rates of convergence and considers methods for constructing confidence regions for the identified set. These results are then applied to panel data binary choice and duration models with fixed effects under weak semiparametric assumptions, a class of models which motivates our analysis.

The literature on estimation of partially identified models has primarily focused on regular models, such as moment condition models, for which estimators of the identified set exist that are essentially $\sqrt{n}$-consistent[1] set estimators. Our results extend the criterion-function-based framework for set inference developed by Chernozhukov, Hong, and Tamer (2007), allowing estimation of parameter sets in a more general class of models in which non-standard rates of convergence may arise. This includes models such as those studied by Kim and Pollard (1990) in which a sharp-edge effect leads to cube-root convergence. We show that a similar mechanism drives what amounts to cube-root convergence in a large class of partially identified models as well. We also describe a class of models in which a particular type of discontinuity in the limiting objective function results in an arbitrarily fast rate of convergence.

These general results are motivated by the case of panel data binary choice and duration models with fixed effects. Without making any additional parametric assumptions, we analyze these models in the absence of a full-support condition which has been widely used in the literature to guarantee point identification (Manski, 1975, 1985; Han, 1987; Horowitz, 1992; Abrevaya, 2000; Chen, 2002). We illustrate our general results by proposing set estimators for these models, which are based on rank conditions similar to those used in maximum score estimation, and showing that the estimators are consistent for the identified set at non-standard rates. Existing conditions in the literature are not well-suited for analyzing these models because of their irregular features. We also carry out a series of Monte Carlo experiments using these models in order to provide evidence for our theoretical findings.

This paper contributes to several areas of the econometrics literature including the growing literature on partial identification, the long literature on semiparametric estimation binary choice and transformation models, and an emerging literature on estimating semiparametric models with limited-support regressors. Work on criterion-function-based estimation and inference in partially identified models started with Manski and Tamer (2002), who analyzed a semiparametric binary choice model with interval-valued data

[1] That is, they can achieve rates of convergence arbitrarily close to $\sqrt{n}$.

under a conditional quantile restriction. They derived the sharp identified set for the model, proposed a set estimator, defined as an appropriately-chosen contour set of a modified maximum score objective function, and showed that it was consistent. Chernozhukov et al. (2007) generalized this approach and developed a broad framework for criterion-function-based estimation of partially identified models. They established conditions for consistency and rates of convergence of estimators in this class along with methods for constructing confidence regions for the identified set using subsampling. Romano and Shaikh (2010) further explored subsampling-based inference in partially identified models while Bugni (2008) introduced a bootstrap procedure.

Confidence regions can also be constructed via set expansion, for example, when the identified set is an interval on the real line (Horowitz and Manski, 2000; Imbens and Manski, 2004). Beresteanu and Molinari (2008) extended this method to more general settings and develop inference procedures based on the theory of random sets for partially identified models where the identified set can be expressed as the Aumann expectation of a set valued random variable.

Although this paper focuses on inference for the identified set, Θ_0, one may also be interested in inference about individual points within the identified set, including the true parameter θ_0. This distinction was raised by Imbens and Manski (2004), who proposed confidence regions for θ_0 in the case where Θ_0 is an interval whose endpoints can be estimated. Stoye (2009) later observed that the conditions of Imbens and Manski (2004) imply the existence of a superefficient estimator of the width of the identified interval and proposed alternative conditions to avoid this implicit assumption. Although some of the estimators proposed in this paper are indeed superefficient, this arises due to the inherent properties of the model, rather than as an implicit assumption. Romano and Shaikh (2008) also showed that subsampling can be applied, uniformly, in the criterion-function-based framework to construct confidence sets for individual elements of Θ_0.

There are numerous other areas where partially identified econometric models have arisen including games with multiple equilibria (Tamer, 2003; Andrews, Berry, and Jia, 2004; Pakes, Porter, Ho, and Ishii, 2006; Aradillas-Lopez and Tamer, 2008; Ciliberto and Tamer, 2009; Beresteanu, Molchanov, and Molinari, 2009) and models characterized by conditional moment inequalities (Khan and Tamer, 2009; Kim, 2009; Andrews and Shi, 2009). See Manski (2003) and Tamer (2009) for surveys of partial identification in econometric models.

In the context of the specific models we consider, our work is also related to the literature on nonlinear semiparametric panel data models, particularly the work of Manski (1987) on the static fixed effects model and Honoré and Kyriazidou (2000) for dynamic models. Our characterizations of the identified sets in the models we consider are based in part on known necessary conditions for point identification established in these papers, however, establishing sharpness in partially identified models requires additional work.

This paper is also related to a growing literature concerned with semiparametric estimation of models with limited support regressors, typically involving either discrete or interval-valued regressors. In terms of cross-sectional models, Bierens and Hartog (1988) showed that there are infinitely many single-index representations of the mean regression of a dependent variable when all covariates are discrete. Horowitz (1998) discussed the generic non-identification of single-index and binary response models with only discrete regressors, a result which serves to motivate our analysis. Manski and Tamer (2002) and Magnac and Maurin (2008) considered partial identification and estimation of binary choice models with an interval-valued regressor. Honoré and Lleras-Muney (2006) estimated a partially identified competing risk models with interval outcome data and discrete explanatory variables. Komarova (2008) proposed consistent estimators, based on a linear programming procedure, of the identified set in a cross-sectional binary response model with discrete regressors. In this paper, we analyze a panel data version of this model with both discrete and continuous regressors and determine the asymptotic properties of the corresponding set estimators.

Finally, other authors have also considered partial identification in panel data models, but with different quantities of interest. In particular, Honoré and Tamer (2006) analyzed dynamic random effects panel data models and discuss how to calculate the identified set using minimum distance, maximum likelihood, and linear programming methods. More recently, Chernozhukov, Fernández-Val, Hahn, and Newey (2009) derive bounds on marginal effects in nonlinear panel models with discrete regressors and Rosen (2009) considers partial identification in fixed effects panel data models under conditional quantile restrictions.

The remainder of this paper is organized as follows. In Section 2, we present an overview of our main results and provide examples based on the specific models that motivate our analysis. Section 3 contains our main results on consistency, rates of convergence, and confidence regions in general models, with more practical sufficient conditions given in Section 4. We apply these results to several specific fixed effects panel data models in Section 5. The results from a series of Monte Carlo experiments based on these models are presented in Section 6. Finally, Section 7 concludes.

2. Overview of the Main Results

In a broad sense, this paper concerns econometric models characterized by a finite vector of parameters θ which lie in some parameter space $\Theta \subset \mathbb{R}^k$. Although our particular focus is on semiparametric models with infinite-dimensional components $\psi \in \Psi$, they are not of interest themselves. For example, ψ might include unknown functionals such as the distribution of disturbances in a particular model. Let P denote the data generating process, the true distribution of observables, and suppose that the model is well-specified in the sense that there exist primitives (θ_0, ψ_0) such that $P_{\theta_0,\psi_0} = P$. Both θ_0 and ψ_0 are unknown to the researcher, but we shall focus on the case where θ_0 is of primary interest.

A model is *point identified* if θ_0 is the only element of Θ such that the model would be consistent with the population distribution P for some $\psi \in \Psi$. On the other hand, the model is *partially identified* if there are multiple elements $\theta \in \Theta$ that could have generated P in the sense that there is some ψ such that $P_{\theta,\psi} = P$. The set of all such θ is the *identified set* and is denoted by Θ_0. Formally, Θ_0 depends on P and is defined as

$$\Theta_0(P) = \left\{\theta \in \Theta : \exists \psi \in \Psi \text{ such that } P_{\theta,\psi} = P\right\}. \tag{1}$$

We will simply write Θ_0 when the context is clear, with the dependence on P assumed. Note that this definition encompasses point identification, since Θ_0 may be a singleton, with consistent point or set estimators converging in probability to the same singleton.

Our leading example, which motivates the general results obtained throughout the paper, is the basic panel data binary choice model with fixed effects, estimated under only weak semiparametric assumptions.

Example 1 (Fixed Effects Binary Choice Model)**.** Suppose that for each period $t = 0, \dots, T-1$, we observe a binary response generated according to the model

$$y_t = 1\{x_t'\beta + c + u_t \geq 0\},$$

where x_t is a vector of explanatory variables, c is an unobserved individual-specific effect, and u_t are time-varying unobserved disturbances. Here, $1\{\cdot\}$ denotes the indicator function, equal to one when the event $\{\cdot\}$ is true and zero otherwise. The distribution of c is unrestricted. We make the standard assumption that u_t is stationary conditional on x and c, but allow u_t to be serially dependent. The only finite-dimensional parameters of interest here are the index coefficients β. Manski (1987) showed that when $T = 2$, if at least one component of $x_1 - x_0$ has full support on $\mathbb{R}$, then the model is point identified. Without this assumption, which fails if all regressors have finite or bounded support, the model is only partially identified (Horowitz, 1998).

This paper focuses on set inference in models where the identified set is characterized by some criterion function $Q(\theta)$. Let $\Theta_1 \equiv \arg\max_\Theta Q$ denote the set of maximizers of Q. Using the analogy principle, we will use the finite sample objective function $Q_n(\theta)$ to obtain a set estimator $\hat{\Theta}_n$ for Θ_1. Although we state our results in terms of Θ_1 for generality, in all cases we consider either Θ_1 is a sharp characterization of the identified set ($\Theta_0 = \Theta_1$) or it is known to contain it ($\Theta_0 \subseteq \Theta_1$). Our main general results concern the consistency of $\hat{\Theta}_n$ and its rate of convergence as well as methods for constructing confidence regions for Θ_1 with some prespecified coverage probability.

Example 2 (Objective Function)**.** Returning to the model of Example 1, for $T = 2$, Manski (1987) showed that the conditional maximum score objective function,

$$Q(\theta) = \mathrm{E}[(y_1 - y_0) \cdot \mathrm{sgn}((x_1 - x_0)'\beta)],$$

can be maximized to consistently estimate θ. In Section 5, we show that this function also provides a sharp characterization of the identified set in the model without the full support assumption. The finite-sample analog objective function is

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y_{i1} - y_{i0}) \cdot \mathrm{sgn}((x_{i0} - x_{i1})'\beta).$$

The analogy principle suggests estimating Θ_1 using the set of maximizers of the sample objective function Q_n. However, in general, taking only the set of maximizers may result in an inconsistent estimator. Instead, we analyze a class of estimators defined in terms of contour sets of Q_n. Let $C_n(\tau_n)$ denote such a contour set, defined as

$$(2) \qquad C_n(\tau_n) \equiv \left\{\theta \in \Theta : Q_n(\theta) \geq \sup_{\Theta} Q_n - \tau_n\right\},$$

where τ_n is a non-negative slackness sequence which converges zero in probability. Estimators of this form were introduced by Manski and Tamer (2002) and have been used by Chernozhukov et al. (2007), Romano and Shaikh (2010), Romano and Shaikh (2008), Bugni (2008), and Kim (2009).

To discuss notions of convergence and consistency, we must be precise about which metric space we are working in. We define convergence in terms of the *Hausdorff distance*, a generalization of Euclidean distance to spaces of sets. Let (Θ, d) be a metric space where d is the standard Euclidean distance. For a pair of subsets $A, B \subset \Theta$, the Hausdorff distance between A and B is

$$(3) \qquad d_H(A, B) = \max\left\{\sup_{\theta \in B} \rho(\theta, A), \sup_{\theta \in A} \rho(\theta, B)\right\},$$

where $\rho(\theta, A) \equiv \inf_{\tilde{\theta} \in A} d(\theta, \tilde{\theta})$ is the shortest distance from the point θ to the set A. Intuitively, the Hausdorff distance between A and B is the farthest distance between an arbitrary point in one of the sets to the nearest neighbor in the other set.

In Theorem 1 of Section 3, we provide conditions on Q, Q_n, and the sequence τ_n to ensure that $\hat{\Theta}_n \equiv C_n(\tau_n)$ is consistent for Θ_1. In particular, we require Q_n to converge uniformly to Q in probability at a known rate b_n and that the sequence τ_n approaches zero in probability at a rate slower than that at which b_n approaches infinity (i.e., $\tau_n \xrightarrow{p} 0$ and $b_n \tau_n \xrightarrow{p} \infty$).

Example 3 (Consistency)**.** For the model of Example 1, we show that that Q_n converges uniformly in probability to Q at the rate $b_n = n^{1/2}$. Therefore, for slackness sequences such that $n^{1/2}\tau_n \xrightarrow{p} \infty$, $\hat{\Theta}_n$ will be consistent for Θ_0. One valid choice of the slackness sequence is $\tau_n \propto \sqrt{\ln n / n}$. In general, however, we can choose τ_n to be a sequence that converges to zero at a rate arbitrarily close to, but slower than, $n^{1/2}$.

The rate of convergence of $\hat{\Theta}_n$ to Θ_1 is shown to depend on the shape of the objective function near the identified set. We show that if Q_n can be bounded in probability from above by a polynomial in $\rho(\theta, \Theta_1)$ outside of a shrinking neighborhood of Θ_1, then the rate of convergence is polynomial. The particular rate depends on the degree of the bounding polynomial, γ_1, and the rate at which the neighborhood shrinks, $b_n^{\gamma_2}$, where $\gamma_1\gamma_2 \geq 1$. When this condition holds, Theorem 3 establishes that $d_H(\hat{\Theta}_n, \Theta_1) = O_p(\tau_n^{\gamma_2})$. So the rate depends on γ_1, γ_2, and the slackness sequence, τ_n, which, as we saw above, is in turn restricted by the rate of uniform convergence, b_n.

In an interesting extreme case, including many models with discrete regressors, we show that if Q has a discontinuity at the boundary of Θ_1 (for example, when Q is a step function) then $\hat{\Theta}_n$ converges arbitrarily fast to Θ_1. The following example illustrates applications of both of these results.

Example 4 (Rates of Convergence)**.** In the model of Example 1, when at least one component of $x_1 - x_0$ is continuous, but potentially bounded, we can show that Q_n is bounded in probability by a second-order polynomial in $\rho(\theta, \Theta_1)$ outside of an $O(n^{-1/3})$ neighborhood of Θ_1. That is, $\gamma_1 = 2$ and $\gamma_2 = 2/3$.

Theorem 3 establishes that in this case the rate of convergence of $\hat{\Theta}_n$ to Θ_1 can be made arbitrarily close to $n^{1/3}$, which is the rate of convergence of the maximum score estimator in the point identified case. The exact rate depends on the particular choice of the slackness sequence τ_n.

On the other hand, when all regressors are discrete, the limiting objective function is a step function, with a nonzero jump at the boundary of the identified set. This function satisfies the conditions of Theorem 2 which states that the rate of convergence is arbitrarily fast, or equivalently, $\hat{\Theta}_n$ equals Θ_1 with probability approaching one.[2]

Finally, we also discuss methods for constructing confidence regions for Θ_1. In the case of smooth models, such as the model of Example 1 with continuous regressors, existing

[2]See Lemma 9 in the appendix for a formal statement of this equivalence.

procedures from the literature can be used (Chernozhukov et al., 2007; Romano and Shaikh, 2010). In models with discontinuous limiting objective functions, such as the step function encountered in Example 4, we provide conditions under which a subsampling procedure can be used to obtain conservative confidence regions with asymptotic coverage probability greater than $1 - \alpha$. We also provide an alternative characterization of confidence regions in this case which shows that any sequence of sets with this confidence property must eventually contain $\hat{\Theta}_n$ with probability at least $1 - \alpha$. This result is confirmed by our Monte Carlo experiments, which provide evidence that for large n, the confidence regions obtained via subsampling are equal to set estimates $\hat{\Theta}_n$.

3. Estimation and Inference in General Models

This section develops conditions under which the sequence of set estimates $\hat{\Theta}_n$ converges in probability to Θ_1 in the Hausdorff metric. We derive the rate of convergence of this sequence under two different conditions on the curvature of the objective function and then discuss methods for constructing confidence regions which cover Θ_1 with some prespecified probability. All proofs are reserved for Appendix B.

3.1. Consistency

Let B^c denote the complement of a set B in Θ. In a slight abuse of notation, we write $B^\varepsilon \equiv \{\theta \in \Theta : \rho(\theta, B) < \varepsilon\}$ to denote an ε-expansion of a set $B \subseteq \Theta$. The following conditions are required for consistency, as established by the theorem which follows.

Assumption A1. Θ is a nonempty, compact subset of $\mathbb{R}^k$.

Assumption A2. There exists a function $Q : \Theta \to \mathbb{R}$ such that for all $\varepsilon > 0$, there exists a $\delta_\epsilon > 0$ such that $\sup_{\Theta \setminus \Theta_1^\varepsilon} Q \leq \sup_\Theta Q - \delta_\varepsilon$, where $\Theta_1 \equiv \arg\max_\Theta Q$.

Assumption A3. There exists a real-valued function $Q_n(\theta)$ and a sequence $b_n \to \infty$ such that $\sup_\Theta |Q_n - Q| = O_p(1/b_n)$.

Assumptions A1 and A3 are the standard compactness and uniform convergence conditions for consistency of M-estimators for singletons. Assumption A2 is a regularity condition which requires the population objective function to have a well-separated maximum. This serves to rule out pathological cases that can arise in the absence of continuity. It plays the role of the identification condition and is satisfied, for example, when Q is either continuous or a step function.

Theorem 1 (Consistency)**.** *Suppose that Assumption A1–A3 hold and let τ_n be a nonnegative sequence of random variables such that $\tau_n \xrightarrow{p} 0$. Then, $\sup_{\theta \in \hat{\Theta}_n} \rho(\theta, \Theta_1) \xrightarrow{p} 0$. Furthermore, if $b_n \tau_n \xrightarrow{p} \infty$, then $\lim_{n\to\infty} P(\Theta_1 \subseteq \hat{\Theta}_n) = 1$ and $d_H(\hat{\Theta}_n, \Theta_1) \xrightarrow{p} 0$.*

Note that the first conclusion of Theorem 1 actually holds without the slackness sequence: $\hat{\Theta}_n$ becomes arbitrarily close to being a subset of Θ_1 in probability for any sequence $\tau_n = o_p(1)$, including $\tau_n = 0$. The slackness sequence is introduced to ensure that the converse holds, that $\hat{\Theta}_n$ covers Θ_1 in probability, by slightly expanding the contour sets by an amount which becomes negligible as $n \to \infty$. By expanding it at the right rate—with τ_n converging to zero in probability, but not faster than $1/b_n$—we ensure that $\hat{\Theta}_n$ is just large enough to cover Θ_1 with probability approaching one. Combining these two results yields consistency in the Hausdorff metric.

3.2. Rates of Convergence

The rate of convergence of the Hausdorff distance $d_H(\hat{\Theta}_n, \Theta_1)$ is the slowest rate at which the component distances, $\sup_{\theta \in \Theta_1} \rho(\theta, \hat{\Theta}_n)$ and $\sup_{\theta \in \hat{\Theta}_n} \rho(\theta, \Theta_1)$, converge to zero. The second part of Theorem 1 establishes that with only Assumptions A1–A3, the first distance converges arbitrarily fast in probability, since it eventually equals zero. In what follows, we consider two different shape restrictions on Q and Q_n which can be used to determine the rate of convergence of the second distance, which determines the overall rate.

In particular, we show that when Q has a discontinuity at the boundary of Θ_1, then $\hat{\Theta}_n$ converges arbitrarily fast in probability to Θ_1. On the other hand, when $Q_n(\theta)$ is stochastically bounded from above by a polynomial in $\rho(\theta, \Theta_1)$ outside of a shrinking neighborhood of Θ_1, we show that the rate of convergence depends on the curvature of the bounding polynomial, the rate at which the neighborhood shrinks, and the rate at which τ_n converges to zero.

Assumption A4. There exists a $\delta > 0$ such that $Q(\theta) \leq \sup_\Theta Q - \delta$ for all $\theta \in \Theta_1^c$.

Theorem 2 (Rate of Convergence with a Constant Majorant)**.** *Suppose that Assumptions A1–A4 hold. If $\tau_n \xrightarrow{p} 0$ and $b_n \tau_n \xrightarrow{p} \infty$, then $\hat{\Theta}_n = \Theta_1$ with probability approaching one.*

Thus, when Q exhibits a discrete jump at the boundary of the identified set, $\hat{\Theta}_n$ converges arbitrarily fast to Θ_1. That is, for *any* sequence r_n, including powers of n and exponential forms, $r_n d_H(\hat{\Theta}_n, \Theta_1) \xrightarrow{p} 0$.[3] We refer to Assumption A4 as a constant majorant condition. As we will see in Section 5, this condition is satisfied when the regressors in the binary choice model of Example 1 have finite support.

[3] See Lemma 9 in Appendix A for a proof that these two statements are equivalent in this setting.

Intuitively, under the conditions of Theorem 2, with probability approaching one we are able to perfectly distinguish which values of θ belong to Θ_1. This happens because Q_n is converging uniformly to Q at a rate that's faster than the rate at which τ_n approaches zero, while at the same time τ_n will eventually become smaller than δ, the size of the discrete jump. The result is that the contour sets $C_n(\tau_n)$ become identically equal to Θ_1.

Next we consider models for which Q_n satisfies a polynomial curvature condition.

Assumption A5. There exist positive constants $(\delta, c_0, c_1, \gamma_1, \gamma_2)$ with $\gamma_1\gamma_2 \geq 1$ such that for any $\varepsilon \in (0,1)$ there are $(c_\varepsilon, n_\varepsilon)$ such that for all $n \geq n_\varepsilon$,

$$Q_n(\theta) \leq c_0 - c_1 \cdot (\rho(\theta, \Theta_1) \wedge \delta)^{\gamma_1}$$

uniformly on $\{\theta \in \Theta : \rho(\theta, \Theta_1) \geq (c_\varepsilon / b_n)^{\gamma_2}\}$ with probability at least $1 - \varepsilon$.

Theorem 3 (Rate of Convergence with a Polynomial Majorant)**.** *Suppose Assumptions A1–A3 and A5 hold. If* $\tau_n \xrightarrow{p} 0$ *and* $b_n \tau_n \xrightarrow{p} \infty$*, then* $d_H(\hat{\Theta}_n, \Theta_1) = O_p(\tau_n^{\gamma_2})$.

Assumption A5 is analogous to conditions used to obtain rates of convergence in point identified models and is a generalization of a similar condition in Chernozhukov et al. (2007). By allowing the degree of the bounding polynomial to differ from the exponent determining the size of the sequence neighborhoods of Θ_1, this theorem is able to characterize the rate of convergence of $\hat{\Theta}_n$ in models where irregular rates obtain. This is important in analyzing the model of Example 1, for instance, where $\hat{\Theta}_n$ is shown to be essentially $\sqrt[3]{n}$-consistent.

3.3. Confidence Regions

In this section we consider the problem of constructing a sequence of sets B_n which have the asymptotic confidence property

$$\lim_{n\to\infty} P(\Theta_1 \subseteq B_n) = 1 - \alpha \tag{4}$$

for a given value of α. This is a complex problem in general, if the sets B_n are not restricted to be members of a more tractable family of sets. So far, the literature has focused on the case where $B_n = C_n(\kappa_n)$ with B_n being a sequence of contour sets of Q_n. Under this restriction, the problem of choosing a sequence of arbitrary sets is reduced to that of choosing a sequence κ_n. In smooth models, which satisfy Assumption A5 and where $\hat{\Theta}_n$ converges at a polynomial rate, Chernozhukov et al. (2007) develop a subsampling-based algorithm for constructing such a sequence.

Since their conditions are not satisfied when the objective function is discontinuous, as in Assumption A4, we approach the problem in two alternative ways. First, we provide a general characterization of sequences B_n which satisfy (4). Then, we provide conditions under which a similar subsampling procedure is valid for constructing such a sequence from the class of contour sets of Q_n.

3.4. General Confidence Regions in Discrete Models

In models for which there exists an arbitrarily fast estimator of Θ_1, for a general sequence B_n satisfying (4) (not necessarily a sequence of contour sets), we have the following characterization.

Lemma 1. *Let $\hat{\Theta}_n$ and B_n be sequences of subsets of Θ and suppose that $\lim_{n\to\infty} P(\hat{\Theta}_n = \Theta_1) = 1$. Then B_n has the confidence property $\lim_{n\to\infty} P\left(\Theta_1 \subseteq B_n\right) \geq 1-\alpha$ if and only if $\lim_{n\to\infty} P\left(\hat{\Theta}_n \subseteq B_n\right) \geq 1-\alpha$.*

The conclusion of this lemma is stark in the sense that in models for which we have an estimator with an arbitrarily fast rate of convergence, any sequence of sets B_n which asymptotically covers Θ_1 with probability at least $1-\alpha$ must also contain $\hat{\Theta}_n$ with probability at least $1-\alpha$. Since $\hat{\Theta}_n$ is itself essentially a probability-one confidence region, it is not reasonable for B_n to be any larger than $\hat{\Theta}_n$. So, the practical conclusion is that any such sequence B_n should be equal to $\hat{\Theta}_n$ with probability $1-\alpha$. Thus, one possible sequence is $B_n = \hat{\Theta}_n$ with probability $1-\alpha$ with B_n unrestricted with probability α. It seems reasonable to simply set $B_n = \hat{\Theta}_n$ with probability one.

3.5. Constructing Confidence Regions via Subsampling

When B_n is restricted to be a sequence contour sets, the coverage of a particular B_n can be inferred using the statistic

$$R_n \equiv \sup_{\Theta} b_n Q_n - \inf_{\Theta_1} b_n Q_n,$$

since for a sequence κ_n, $P\left(\Theta_1 \subseteq C_n(\kappa_n/b_n)\right) = P(R_n \leq \kappa_n)$. In defining R_n the objective function is scaled by b_n, the rate of uniform convergence, so that we might use subsampling to approximate quantiles of R_n.

An appropriate sequence $\hat{\kappa}_n$, and corresponding conservative confidence regions $C_n(\hat{\kappa}_n/b_n)$ with asymptotic coverage probability of at least $1-\alpha$, can be constructed using the algorithm below. We assume that an iid sample is available and that R_n converges in distribution.

Assumption A6. The sample consists of independent draws from P.

Assumption A7. Suppose that $P\{R_n \leq c\} \to P\{R \leq c\}$ for each $c \in \mathbb{R}$ for some random variable R.

Algorithm 1. For a given asymptotic coverage probability $1-\alpha$:

1. Choose a subsample size $m < n$ such that $m \to \infty$ and $m/n \to 0$ as $n \to \infty$. Let M_n denote the number of subsets of size m and let τ_n be any sequence such that such that $C_n(\tau_n)$ is consistent for Θ_1.

2. Compute $\hat{\kappa}_n$ as the $1-\alpha$ quantile of the values $\{\hat{R}_{n,m,j}\}_{j=1}^{M_n}$ where

$$\hat{R}_{n,m,j} \equiv \sup_{\theta \in \Theta} b_m Q_{n,m,j}(\theta) - \inf_{\theta \in C_n(\tau_n)} b_m Q_{n,m,j}(\theta)$$

   and $Q_{n,m,j}$ denotes the sample objective function constructed using the j-th subsample of size m.

3. Report $C_n(\tau_n)$ as a consistent estimate of Θ_1 and $C_n(\hat{\kappa}_n/b_n)$ as a conservative confidence region.

The following theorem addresses the validity of this algorithm for obtaining the desired sequence $\hat{\kappa}_n$.

Theorem 4. *Suppose that Assumptions A1–A4, A7, and A6 hold and that $m \to \infty$, and $m/n \to 0$ as $n \to \infty$. Let $1-\alpha$ denote the desired coverage level, where the distribution of R is continuous at $c(1-\alpha)$. Then,*

$$\hat{\kappa}_n \xrightarrow{p} c(1-\alpha) \equiv \inf\{c : P\{R \leq c\} \geq 1-\alpha\}$$

and

$$P\{\Theta_1 \subseteq C_n(\hat{\kappa}_n)\} \geq (1-\alpha) + o_p(1).$$

4. Sufficient Conditions

This section derives sufficient conditions for the more primitive conditions of the general theorems of Section 3. These sufficient conditions are verified in the context of the applications in Section 5, providing several examples of their use. Many of these conditions are stated in terms of empirical process concepts—restrictions on the indexing class of functions which generate the finite sample and limiting objective functions. We summarize the standard notation and definitions below, but refer the reader to Section 2 of Pakes and Pollard (1989) for further details.

Let P denote the joint distribution of observables and P_n denote the empirical measure associated with n independent and identically distributed from P. Whenever there is no ambiguity we write Pg instead of $\int g(z)P(dz)$ or $\int g dP$ to denote the integral of some function g with respect to a measure P.

We focus on models in which the objective functions can be expressed in terms of a class of functions $\mathcal{F}$ so that $Q(\theta) = Pf(\cdot, \theta)$ and $Q_n(\theta) = P_n f(\cdot, \theta)$ for all θ for $f(\cdot, \theta) \in \mathcal{F}$. As such, we work with empirical processes indexed by classes of functions $\mathcal{F} = \{f(\cdot, \theta) : \theta \in \Theta\}$. Alternatively, we use parameter space Θ as the indexing set when convenient.

Establishing the asymptotic properties of $\hat{\Theta}_n$ in these models amounts to verifying certain properties of $\mathcal{F}$. An envelope for $\mathcal{F}$ is a function F such that $\sup_{\mathcal{F}} |f| \leq F$. For consistency, a sufficient condition for the uniform convergence required by Assumption A3 is that $\mathcal{F}$ is manageable in the sense of Pollard (1989) for a square integrable envelope F.

Assumption C1. Θ is a nonempty, compact subset of $\mathbb{R}^k$ and there exists a class of functions $\mathcal{F} = \{f(\cdot, \theta) : \theta \in \Theta\}$ such that $Q(\theta) = Pf(\cdot, \theta)$ and $Q_n(\theta) = P_n f(\cdot, \theta)$ for all $\theta \in \Theta$.

Assumption C2. $Q(\theta)$ is piecewise continuous on Θ.

Assumption C3. $\mathcal{F}$ is manageable for some envelope F such that $PF^2 < \infty$.

Lemma 2. *If Assumptions C1–C3 hold, then for any sequence τ_n such that $n^{1/2}\tau_n \xrightarrow{p} 0$, $d_H(\hat{\Theta}_n, \Theta_1) \xrightarrow{p} 0$.*

Proof. The compactness assumption implies A1 and piecewise continuity of Q implies A2. Since $\mathcal{F}$ is manageable with $PF^2 = 1 < \infty$, it follows from Corollary 3.2 of Kim and Pollard (1990) that A3 holds with $b_n = n^{1/2}$. The result follows by Theorem 1. ■

Continuity of Q can be established by the uniform law of large numbers when the functions $f(\cdot, \theta)$ are continuous in θ with probability one under P and dominated by some bounded function F (Newey and McFadden, 1994, Lemma 2.4). Furthermore, in many models it is easy to verify that $\mathcal{F}$ is a Vapnik-Chervonenkis (VC) subgraph class, in the sense of Dudley (1987), with constant envelope $F < \infty$. That is, when $\mathcal{F}$ is a class of functions such that $\{\text{subgraph}(f) : f \in \mathcal{F}\}$ is a VC class of sets and $\sup_{\mathcal{F}} |f| \leq F < \infty$, $\mathcal{F}$ is necessarily manageable and $PF^2 < \infty$.

Lemma 3. *Suppose that Assumption C1 holds. If $\mathcal{F}$ is a VC subgraph class such that $|f(\cdot, \theta)| \leq M$ for all $\theta \in \Theta$ for the constant function $M < \infty$, then Assumption C3 holds. In addition, if $f(z, \theta)$ is continuous in θ with probability one, then Assumption C2 holds.*

Proof. Since $\mathcal{F}$ is a VC subgraph class, Lemma 2.12 of Pakes and Pollard (1989) implies that $\mathcal{F}$ is Euclidean for any valid envelope, including the constant function $F = M$. Since $\mathcal{F}$ is Euclidean, it is also manageable for $F = M$ (cf. Pakes and Pollard, 1989, p. 1033). Since $PF^2 = M^2 < \infty$, this verifies Assumption C3.

Furthermore, if $f(z, \theta)$ is continuous in θ with probability one, since it is dominated by $F = M$ for all θ, continuity of Q follows from Lemma 2.4 of Newey and McFadden (1994), verifying Assumption C2. ∎

For smooth models, Kim and Pollard's heuristic for cube-root consistency translates well to the set identified case. Let $\Gamma(\theta) \equiv Q(\theta) - Q(\theta_0)$ and $\Gamma_n(\theta) \equiv Q_n(\theta) - Q_n(\theta_0)$. We can decompose $\Gamma_n(\theta)$ into two components, a trend and a stochastic component: $\Gamma_n(\theta) = \Gamma(\theta) + [\Gamma_n(\theta) - \Gamma(\theta)]$. Suppose that near Θ_1, the limiting objective function is approximately quadratic in the distance $\rho(\theta, \Theta_1)$: $\Gamma(\theta) = O(\rho^2(\theta, \Theta_0))$. For models with a feature Kim and Pollard call the sharp edge effect, the variance of the empirical process component is $O_p(\rho(\theta, \Theta_0)/n)$. Only when the trend overtakes the noise is Γ_n likely to be below the maximum. Thus, the maximum is likely to occur when the standard deviation of the random component is of the same magnitude or larger than the trend. That is, when $\sqrt{\rho(\theta, \Theta_0)/n} > \rho^2(\theta, \Theta_0)$ or $\rho(\theta, \Theta_0) < n^{-1/3}$. Therefore, $\hat{\Theta}_n$, the set of near maximizers of Γ_n, should be within a neighborhood of Θ_0 on the order of $n^{-1/3}$. In the set identified case, however, $\rho(\theta, \Theta_0)$ is only one component of the Hausdorff distance. Conveniently, the other component was shown to converge arbitrarily fast under fairly weak assumptions (Theorem 1) and therefore it does not hinder the rate of convergence.

In terms of Theorem 3, the above argument corresponds to the case where $\gamma_1 = 2$ and $\gamma_2 = 2/3$. Since τ_n can be chosen arbitrarily close to $n^{-1/2}$, the rate of convergence, r_n, can be made arbitrarily close to $O_p(n^{-1/2})^{\gamma_2} = O_p(n^{-1/3})$. The following conditions are sufficient for $\hat{\Theta}_n$ to obtain near cube root rates of convergence.

Assumption C4. There exists a neighborhood Θ_1^ν of Θ_1 with $\nu > 0$ and positive constants c_0 and c_1 such that $Q(\theta) \leq c_0 - c_1 \cdot \rho^2(\theta, \Theta_1)$ for all $\theta \in \Theta_1^\nu$.

Assumption C5. There exists an $\eta_0 > 0$ such that for all $\eta \leq \eta_0$, the classes $\mathcal{F}_\eta \equiv \{f(\cdot, \theta) : \rho(\theta, \Theta_1) \leq \eta\}$ are uniformly manageable with $PF_\eta^2 = O(\eta)$ where $F_\eta \equiv \sup_{\mathcal{F}_\eta} |f(\cdot, \theta)|$ is the natural envelope of $\mathcal{F}_\eta$.

Lemma 4. *Suppose that Assumptions C1-C5 hold. For any sequence $r_n = o(n^{1/3})$, if $\tau_n \propto r_n^{-3/2}$, then $d_H(\hat{\Theta}_n, \Theta_1) = O_p(r_n)$.*

Finally, note that the constant majorant condition of Assumption A4 is satisfied when Q is a step function. Therefore, the following theorem gives sufficient conditions for an arbitrarily fast rate of convergence.

Lemma 5. *Suppose that Assumptions C1–C3 hold and that Q is a step function. If $\tau_n \xrightarrow{p} 0$ and $n^{1/2}\tau_n \xrightarrow{p} \infty$, then for any positive sequence r_n with $r_n \to \infty$, $r_n d_H(\hat{\Theta}_n, \Theta_1) \xrightarrow{p} 0$.*

Proof. As established by Lemma 2, Assumptions C1–C3 are sufficient for Assumptions A1–A3 with $b_n = n^{1/2}$. Since Q is a step function, Assumption A4 holds for all $\delta < \sup_\Theta Q - \sup_{\Theta \setminus \Theta_1} Q$. The result follows from Assumption Theorem 2. ∎

5. Analysis of Panel Data Binary Choice and Duration Models

In this section we apply the general results of Sections 3 and 4 to several fixed effects panel data models: two binary choice variants and a class of transformation and duration models. We focus on models where observations are available at times $t = 0, \ldots, T-1$ for each individual. A panel member in these models is described completely by a random vector $(y_0, x_0, u_0, \ldots, y_{T-1}, x_{T-1}, u_{T-1}, c)$, where y_t is a response variable in period t, x_t is a vector of k observed explanatory variables, u_t is an unobserved disturbance in period t, and c is a time invariant individual-specific unobserved effect. Let $y \equiv (y_0, \ldots, y_{T-1})$ and define x and u similarly.

5.1. Panel Data Binary Choice Models

Discrete response models have become a standard tool in applied econometrics and their properties have been studied thoroughly in the econometrics literature (McFadden, 1974; Maddala, 1983; Amemiya, 1985). Semiparametric methods, such as maximum score, emerged to estimate such models without making tenuous parametric assumptions, however, these methods typically assume the existence of an exogenous explanatory variable with full support (Manski, 1975, 1985; Horowitz, 1992). Similar rank conditions have been successful in estimating more general regression models, but the known conditions for point identification still include a full support condition (Han, 1987; Abrevaya, 2000). In practice, however, it is not uncommon to encounter datasets with genuinely discrete or bounded variables. Without a regressor with full support on the real line, under semiparametric assumptions, the models we consider are only partially identified in general (Horowitz, 1998).

Here we formalize the basic linear-index fixed effects binary choice model introduced in Example 1. Proofs for all results in this section are given in Appendix D.

Model 1 (Fixed Effects Binary Choice)**.** *For $t = 0, 1,$*

(5) $\quad y_t = 1\{x_t'\beta + c + u_t \geq 0\}$

where x_t is a random variable with support $\mathcal{X} \subseteq \mathbb{R}^k$, c is a real-valued random variable, and β is the parameter of interest, a member of some parameter space $\Theta \subseteq \mathbb{R}^k$. In addition, for all x and c, $F_{u_t|xc}$ satisfies the following:

a. $F_{u_t|xc} = F_{u_0|xc}$ *for all t.*

b. The support of u_t is $\mathbb{R}$.

Condition a is a substantive restriction, necessary for the estimation methods we introduce below. It requires u_t to be is stationary conditional on the identity of the panel member—that is, conditional on (x, c). Note, however, that it does not restrict the form of serial dependence of u_t in any way. Condition b is a regularity condition which serves to ensure that for each c and x, the event $y_1 \neq y_0$ occurs with positive probability. Note that although we focus on the case where $T = 2$, Charlier, Melenberg, and van Soest (1995) have shown that maximum score estimation can be applied to panels with $T > 2$.

5.1.1. Identification

We begin by reviewing existing conditions for point identification before deriving the identified that results after relaxing some of these assumptions. In the cross-sectional model with a conditional median restriction, analogous to the fixed effects model above, Manski (1985) showed that a full rank, full support condition on x was sufficient to point identify β up to scale. That is, he assumes that x is not contained in a proper linear subspace of $\mathbb{R}^k$ and that the first component of x has positive density everywhere on $\mathbb{R}$ for almost every value of the remaining components. The same conditions were invoked by Han (1987) for the maximum rank correlation estimator and Horowitz (1992) for the smoothed maximum score estimator. The panel version of this assumption (for $T = 2$) was used by Manski (1987) to establish point identification of β up to scale in a semiparametric fixed effects panel data model of the kind considered in the present paper.

Thus, modulo assumptions on the disturbances, point identification of β hinges on what one knows, or is willing to assume, about the distribution of x. The validity of a full support assumption is application-specific. Many common variables such as age, number of children, years of education, and gender are inherently discrete and so it is clearly inappropriate in many cases. Similarly, variables such as income have only partial support on the real line. One advantage of the estimators proposed in this paper is that they do not distinguish between the point identified and partially identified cases. That is, although they do not require a regressor with full support, they will still exploit the additional information provided by one when available.

We consider two alternatives to the full support condition. The first applies when x is a discrete random variable with finite support, while the second applies when at least one component of x is continuous but may fail to have full support on $\mathbb{R}$.

Assumption B1. x_t is a discrete random vector with finite support $\mathcal{X} \subset \mathbb{R}$ for all t.

Assumption B2. The last component of $x_1 - x_0$ has positive density everywhere on a set $\mathcal{W}_k \subseteq \mathbb{R}$ for almost every value of the remaining components.

Note that this assumption does not rule out the possibility that $\mathcal{W}_k = \mathbb{R}$, but it also includes cases where the support of w is bounded in some sense.

The primitives of Model 1 are β, $F_{u_0|xc}$, and $F_{c|x}$, but β is the only finite-dimensional parameter of interest. We now provide a tractable characterization of Θ_0 in terms of observables and show that it is equivalent to the identified set defined above. Since c is unobserved, in order to estimate β we must find implications of the model that are independent of c.

Theorem 5. *In Model 1, the identified set is*

$$\Theta_0 = \left\{\beta \in \Theta : \operatorname{sgn}\left(P(y_1 = 1 \mid x) - P(y_0 = 1 \mid x)\right) = \operatorname{sgn}((x_1 - x_0)'\beta)\ F_x - a.s.\right\}. \tag{6}$$

5.1.2. Consistent Estimation of the Identified Set

Now, we use the sufficient conditions of Section 4 to establish the consistency of the set estimator for Θ_0. We first propose population and finite sample criterion functions and show that the population criterion function characterizes the identified set exactly. Then, we verify the sufficient conditions of Lemma 2 to show that the estimator is consistent. In the sections that follow, we obtain the rate of convergence in two cases: under Assumption B1 $\hat{\Theta}_n$ converges arbitrarily fast to Θ_0 and, under Assumption B2 it is possible to achieve rates arbitrarily close to $n^{1/3}$.

The population objective function we propose for use in estimating Model 1 is the maximum score objective function of Manski (1987), a panel data analog of the cross-sectional maximum score objective function of Manski (1975, 1985):

$$Q(\beta) = \mathrm{E}\left[(y_1 - y_0)\operatorname{sgn}\left((x_1 - x_0)\beta\right)\right].$$

The corresponding finite sample analog objective function is

$$Q_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}(y_{i1} - y_{i0})\operatorname{sgn}\left((x_{i1} - x_{i0})\beta\right).$$

Note that although the same objective function can be used for maximum score estimation in the point identified case, the set estimators proposed here are fundamentally different since they are defined as contour sets of this function.

Lemma 6 below establishes the equivalence between the identified set, Θ_0, and the set of maximizers of the population objective function, Θ_1.

Lemma 6. *Under the maintained assumptions of Model 1, $\Theta_1 = \Theta_0$.*

To show that $\hat{\Theta}_n$ is consistent in this model, we verify the sufficient conditions given in Lemma 2. First, note that Assumption C1 holds with $f(x, y, \beta) = (y_1 - y_0) \cdot (2 \cdot 1\{(x_1 - x_0)'\beta \geq 0\} - 1)$ and $\mathcal{F} = \{f(\cdot, \beta) : \beta \in \Theta\}$ since then $Q(\beta) = Pf(\cdot, \beta)$ and $Q_n(\beta) = P_n f(\cdot, \beta)$ for all $\beta \in \Theta$ with Θ being a nonempty and compact subset of $\mathbb{R}^k$ by assumption.

Now, under Assumption B1, Q is a step function which is piecewise continuous. Under Assumption B2, Q is continuous by the law of large numbers. Since the last component of $x_1 - x_0$ has positive density everywhere on $\mathcal{W}_k$, it follows that $f(x, y, \beta)$ is continuous in β with probability one. Furthermore, f is dominated by the constant function $F = 1$ for all $\beta \in \Theta$. Therefore, Assumption C2 holds in both cases.

It turns out that $\mathcal{F}$ is also a VC subgraph class of functions which, along with the domination condition above, is sufficient for Assumption C3. We verify this in the proof of the following lemma, which summarizes the consistency result.

Theorem 6. *In Model 1, under either Assumption B1 or B2 for any sequence $\tau_n \xrightarrow{p} 0$ with $n^{1/2}\tau_n \xrightarrow{p} \infty$, $d_H(\hat{\Theta}_n, \Theta_0) \xrightarrow{p} 0$.*

5.1.3. Rates of Convergence

The rate of convergence of $\hat{\Theta}_n$ to Θ_0 in Model 1 depends on the support of w. We obtain the rate under both Assumption B1 and B2. We show that when the support of x is finite, $\hat{\Theta}_n$ converges arbitrarily fast in probability to Θ_0. On the other hand, when at least one component of $x_1 - x_0$ is continuous, the estimator can achieve rates arbitrarily close to $n^{1/3}$. The rate depends on τ_n and, although the exact rate $n^{1/3}$ is not achievable, in practice, one can achieve rates close to $n^{1/3}$ by choosing, for example, $\tau_n \propto \sqrt{\ln n / n}$.

Discrete Regressors Here, we verify the constant majorant in the context of Model 1. We can then apply Theorem 2 to show that in this case, $\hat{\Theta}_n$ converges arbitrarily fast to Θ_0.

When the support of (x_0, x_1) is a finite set, henceforth $\mathcal{X}$, the objective function $Q(\theta)$ can be rewritten as follows:

$$
\begin{aligned}
Q(\theta) &= \mathrm{E}_x \, \mathrm{E}_{y|x} \left[(y_1 - y_0) \operatorname{sgn} \left((x_1 - x_0)'\beta \right) \right] \\
&= \sum_{x \in \mathcal{X}} P(x) \left[P(y_1 = 1 \mid x) - P(y_0 = 1 \mid x) \right] \operatorname{sgn} \left((x_1 - x_0)'\beta \right).
\end{aligned}
$$

Therefore, $Q(\theta)$ is a step function and there exists a real number $\delta > 0$ such that for all $\theta \in \Theta \setminus \Theta_0$, $Q(\theta) \leq \sup_\Theta Q - \delta$. In particular, δ is bounded below by the smallest nonzero value of $P(x)\,[P(y_1 = 0 \mid x) - P(y_0 = 1 \mid x)]$ for any $x \in \mathcal{X}$. Thus, applying Theorem 2, we have the following result.

Theorem 7. *Suppose that Assumption B1 holds in Model 1. For any sequence τ_n such that $\tau_n \xrightarrow{p} 0$ and $n^{1/2}\tau_n \xrightarrow{p} \infty$, then $\hat{\Theta}_n$ equals Θ_0 with probability approaching one.*

Continuous Regressors The properties of the maximum score objective function in the continuous covariate case have been studied by Kim and Pollard (1990), Abrevaya and Huang (2005), and others. For simplicity, define $w \equiv x_1 - x_0$. We follow Abrevaya and Huang (2005) in normalizing the coefficient on the last component of w, w_k, to be either 1 or -1 and consider β to be a vector in $\mathbb{R}^{k-1}$. Since this component of β converges arbitrarily fast, without loss of generality we only consider the case where $\beta_k = 1$. Let $\tilde{w}$ denote the remaining components of w.

The following theorem formalizes the cube-root consistency result for this model. We will need several additional assumptions on the distribution of w, which, for comparison, are intentionally close to those made by Abrevaya and Huang (2005) in analyzing the cross-sectional model in the point identified case. Let G and g denote generic cdfs and density functions, with subscripts denoting the associated random variables in cases where there is ambiguity. Finally, let w_k denote the k-th component of w and let $\tilde{w}$ denote the remaining $k-1$ components.

Theorem 8. *Suppose that Assumptions A6 and B2 hold in Model 1. In addition, suppose the following:*

a. The components of $\tilde{w}$ and $\tilde{w}\tilde{w}'$ have finite first absolute moments.

b. The function $g'(w_k \mid \tilde{w})$ exists and, for some $M > 0$, $|g'(w_k \mid \tilde{w})| < M$ and $|g(w_k \mid \tilde{w})| < M$ for all w_k and almost every $\tilde{w}$.

c. For all v in a neighborhood of 0, all w_k in a neighborhood around $-\tilde{w}'\beta_0$, almost every $\tilde{w}$, and some $M > 0$, the function $g(v \mid \tilde{w}, w_k)$ exists and $g(v \mid \tilde{w}, w_k) < M$.

d. For all v in a neighborhood of 0, all w_k in a neighborhood of $-\tilde{w}'\beta_0$, almost every $\tilde{w}$, and some $M > 0$, the function $\partial G(v \mid \tilde{w}, w_k)/\partial w_k$ exists and $|\partial G(v \mid \tilde{w}, w_k)/\partial w_k| < M$.

e. Θ_0 is contained in the interior of Θ.

f. The matrix $V(\theta) \equiv \mathrm{E}\left[2 g_v(0 \mid \tilde{w}, -\tilde{w}'\beta) g_{w_k}(-\tilde{w}'\beta \mid \tilde{w})\tilde{w}\tilde{w}'\right]$ is positive semidefinite for all $\theta \in \mathrm{bd}(\Theta_0)$.

Then for any sequence τ_n such that $\tau_n \xrightarrow{p} 0$ and $n^{1/2}\tau_n \xrightarrow{p} \infty$, $d_H(\hat{\Theta}_n, \Theta_0) = O_p(\tau_n^{2/3})$.

5.1.4. Confidence Regions

In this section we verify the conditions required for constructing confidence regions in the context of Model 1 under Assumption B1. The following lemma verifies the convergence of R_n required by Assumption A7 and establishes the validity of Algorithm 1 for constructing conservative confidence regions in this model.

Lemma 7. *In Model 1, if Assumption A6 holds, then Assumption A7 is satisfied.*

5.2. Panel Data Binary Choice Models with a Lagged Dependent Variable

In this section we focus on an extension of the basic fixed effects binary choice model which introduces a lagged dependent variable in order to control for state dependence. Since the results in this section parallel those from previous sections and are derived in a similar manner, all proofs are reserved for Appendix E. The formal model specification follows. Note that since we do not observe y_t in periods prior to the sample, the model is left unspecified in the first period.

Model 2 (Lagged Dependent Variable Model)**.** *The choice probabilities in the first period are $P(y_0 = 0 \mid x, c) = p_0(x, c)$, where p_0 is unknown and $0 < p_0(x, c) < 1$ for all x and c. In subsequent periods $t = 1, \ldots, T$,*

$$y_t = 1\{x_t'\beta + \gamma y_{t-1} + c + u_t \geq 0\} \tag{7}$$

where x_t is a random vector with support $\mathcal{X}$, c is a real-valued random variable, and $\theta = (\beta, \gamma)$ are the parameters of interest which lie in some parameter space $\Theta \subseteq \mathbb{R}^{k+1}$. In addition, the unobservables u_t are serially independent, identically distributed with cdf $F_{u_t|xc} = F_{u_0|xc}$ for all t, and have full support on $\mathbb{R}$.

Note that in this model, as opposed to the basic fixed effects model, serial correlation in the disturbances is prohibited. The full support assumption on u_t is a regularity condition which guarantees that certain events used for estimation occur with positive probability.

First we must characterize the identified set, and in doing so our analysis follows along the lines of Chamberlain (1985) and Honoré and Kyriazidou (2000) and we focus on the case where $T = 4$. Restrictions on the identified set in this model that are independent of the fixed effect can be found by comparing events with the same outcome in periods 0 and 3 but different outcomes in periods 1 and 2. Let A and B denote these events, for given values of $d_0, d_3 \in \{0, 1\}$:

$$A = \{y_0 = d_0, y_1 = 0, y_2 = 1, y_3 = d_3\},$$
$$B = \{y_0 = d_0, y_1 = 1, y_2 = 0, y_3 = d_3\}.$$

It follows that for all values of d_0 and d_3,

$$\operatorname{sgn}\left(P(A \mid x, x_2 = x_3) - P(B \mid x, x_2 = x_3)\right) = \operatorname{sgn}\left((x_2 - x_1)'\beta + \gamma(d_3 - d_0)\right).$$

A characterization of Θ_0 using this condition is formalized in the following theorem.

Theorem 9. *In Model 2,*

$$\begin{aligned}\text{(8)} \qquad \Theta_0 \subseteq \tilde{\Theta}_0 = \big\{\theta \in \Theta : &\operatorname{sgn}\left(P(A \mid x, x_2 = x_3) - P(B \mid x, x_2 = x_3)\right) \\ &= \operatorname{sgn}\left((x_1 - x_2)'\beta + \gamma(d_3 - d_0)\right) \;\; F_x - a.s.\ \forall d_0, d_3 \in \{0,1\}\big\}.\end{aligned}$$

Next, we propose a set estimator for $\tilde{\Theta}_0$ and apply the general results from Section 3 to show that it is consistent and derive the rate of convergence. For simplicity we only consider the lagged dependent variable model under Assumption B1.

We use the population objective function

$$Q(\theta) = \mathrm{E}\left[1\{x_2 = x_3\} \cdot (y_2 - y_1) \cdot \operatorname{sgn}((x_2 - x_1)'\beta + \gamma(y_3 - y_0))\right].$$

This function was used by Honoré and Kyriazidou (2000) for estimation in point identified models. The finite sample objective function is

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} 1\{x_{i2} = x_{i3}\} \cdot (y_{i2} - y_{i1}) \cdot \operatorname{sgn}((x_{i2} - x_{i1})'\beta + \gamma(y_{i3} - y_{i0})).$$

The following lemma establishes that Q is maximized exactly on $\tilde{\Theta}_0$.

Lemma 8 (Objective Function Characterization of Θ_1)**.** *Under the maintained assumptions of Model 2, $\Theta_1 = \tilde{\Theta}_0$, where $\tilde{\Theta}_0$ is defined in (8).*

Next, we verify Assumptions C1–C3 to establish the consistency of $\hat{\Theta}_n$ for $\tilde{\Theta}_0$. As before, Q is a step function under Assumption B1 since

$$\begin{aligned}Q(\theta) = \sum_{y_0 \in \{0,1\}} \sum_{y_3 \in \{0,1\}} \sum_{x \in \mathcal{X}} &P(x) P(y_0 \mid x) P(y_3 \mid x, y_0)\, \mathrm{E}[y_2 - y_1 \mid x, y_0, y_3] \\ &\times \operatorname{sgn}\left((x_2 - x_1)'\beta + \gamma(y_3 - y_0)\right).\end{aligned}$$

Furthermore, both Q and Q_n are generated by a class of functions $\mathcal{F}$ indexed by θ and that this class is manageable for a square integrable envelope.

Theorem 10. *Suppose that the conditions of Model 2 and Assumption B1 hold. Let $\mathcal{F}$ be the class of functions*

$$f(x, y, \theta) = 1\{x_2 = x_3\}(y_2 - y_1)\left[2 \cdot 1\{(x_2 - x_1)'\beta + \gamma(y_3 - y_0) \geq 0\} - 1\right]$$

indexed by $\theta \in \Theta$. Then $Q(\theta) = Pf(\cdot, \cdot, \theta)$ and $Q_n(\theta) = P_n f(\cdot, \cdot, \theta)$ for all $\theta \in \Theta$, Q is piecewise continuous, and $\mathcal{F}$ is manageable for the constant envelope $F = 1$. It follows that for any sequence τ_n such that $\tau_n \xrightarrow{\mathrm{P}} 0$ and $n^{1/2}\tau_n \xrightarrow{\mathrm{P}} \infty$, $d_H(\hat{\Theta}_n, \tilde{\Theta}_0) \xrightarrow{\mathrm{P}} 0$.

It is immediate from Theorem 10 and the fact that Q is a step function under Assumption B1 that $\hat{\Theta}_n$ converges arbitrarily fast to $\tilde{\Theta}_0$ in probability.

Corollary. *Under the conditions of Theorem 10, for any positive sequence r_n with $r_n \to \infty$, $r_n d_H(\hat{\Theta}_n, \tilde{\Theta}_0) \xrightarrow{\text{p}} 0$.*

Proof. The result follows directly from Theorem 10, noting that Q is a step function, by applying Lemma 5. ∎

Turning to confidence regions in models with discrete regressors, the arguments to establish the validity of the subsampling procedure of Algorithm 1 are essentially identical to those for Model 1. This follows since both objective functions are of the same form in the respective classes of functions and both functions satisfy the constant majorant assumption. Since $\mathcal{F}$ is Euclidean, it follows that Lemma 7 also applies to Model 2 under Assumption B1. Therefore, conservative confidence regions can be constructed using Algorithm 1.

5.3. Panel Data Transformation and Duration Models

This section focuses on fixed effects panel data duration models, with multiple spells, which are members of a more general class of transformation models.

Model 3 (Panel Data Transformation Model)**.** *For all t,*

$$\Lambda(y_t) = x_t'\beta + c + u_t \tag{9}$$

where Λ is a strictly increasing function, x_t is a random vector with support $\mathcal{X} \subseteq \mathbb{R}^k$, c is a real-valued random variable, and β, the vector of parameters of interest, lies in a compact subset of R^k. The disturbances, u_t, are iid conditional on x and c with positive density almost everywhere on $\mathbb{R}$ and with cdf $F_{u_0|xc}$.

Here, t denotes a single spell, the finite-dimensional parameter of interest is β, and the parameter space is $\Theta \subseteq \mathbb{R}^k$. The covariates x_t remain constant within a spell, but vary may across spells. As before, c is a time-invariant individual-specific unobserved variable.

This model is quite general and contains many common duration models in their panel data forms with individual-specific time-invariant unobserved heterogeneity. For example, the generalized accelerated failure time (GAFT) model of Ridder (1990) arises when $\Lambda(y) = \ln z(y)$. Similarly, the accelerated failure time (AFT) model arises when $\Lambda(y) = \ln y$. The mixed proportional hazard (MPH) model arises when u_t has the minus extreme value distribution with $F_{u_0|xc}(u) = 1 - \exp(-\mathrm{e}^u)$ and $\Lambda(y) = \ln H_0(y)$ is the log integrated baseline hazard function. Writing the model as in (9), without specifying Λ, includes all of these models without imposing unnecessary functional form restrictions on the baseline hazard function or parametric distributional assumptions on u_t.

Identification of this model and similar ones has been considered by a number of authors under a wide variety of conditions. For example, Ridder (1990) shows that the GAFT model with continuous covariates is nonparametrically identified both with continuous duration data and with discrete duration data and an additional parametric assumption on the regression function. Model 3 is slightly more general than the GAFT model, and we focus instead on the case when the observed durations are continuous but the covariates are discrete. Han (1987), Chen (2002), Abrevaya (2000) and others have considered point identification and estimation of various components of generalized regression models, which contain models of this type, but invariably the maintained assumptions include a full-support condition, which we relax. An exception is Honoré and Lleras-Muney (2006), who estimate bounds on parameters in a competing risks model with discrete regressors and interval-valued durations.

5.3.1. Estimating the Index Coefficients

In many ways, this model is very similar to Model 1 and so many of the results will be familiar. When the disturbances are independent, we can carry out a similar ranking procedure relating the ordering of y_1 and y_0 to that of $x_1'\beta$ and $x_0'\beta$:

$$
\begin{aligned}
&P(y_1 \geq y_0 \mid x, c) \geq P(y_0 \geq y_1 \mid x, c) \\
&\quad \iff P(x_1'\beta + u_1 \geq x_0'\beta + u_0 \mid x, c) \geq P(x_0'\beta + u_0 \geq x_1'\beta + u_1 \mid x, c) \\
&\quad \iff P(u_0 - u_1 \leq (x_1 - x_0)'\beta \mid x, c) \geq P(u_1 - u_0 \leq (x_0 - x_1)'\beta \mid x, c) \\
&\quad \iff P(u_0 - u_1 \leq (x_1 - x_0)'\beta \mid x, c) \geq P(u_0 - u_1 \leq (x_0 - x_1)'\beta \mid x, c) \\
&\quad \iff (x_1 - x_0)'\beta \geq 0
\end{aligned}
$$

Note that we are able to exchange u_1 and u_0 due to the independence assumption.

Here we consider estimating the set suggested by the rank condition above:

$$
\tilde{\Theta}_0 = \left\{ \beta \in \Theta : \operatorname{sgn}\left(P(y_1 \geq y_0 \mid x) - P(y_0 \geq y_1 \mid x)\right) = \operatorname{sgn}\left((x_1 - x_0)'\beta\right) \right\}.
$$

This set is guaranteed to contain Θ_0. The intuition underlying this set is that, due to the structure of the model, whenever $x_1'\beta \geq x_0'\beta$ it is likely also the case that $y_1 \geq y_0$.

Consider the following population objective function and sample analog:

$$
\begin{aligned}
Q(\beta) &= \mathrm{E}\left[\operatorname{sgn}(y_1 - y_0) \cdot \operatorname{sgn}\left((x_1 - x_0)'\beta\right)\right] \\
Q_n(\beta) &= \frac{1}{n} \sum_{i=1}^{n} \operatorname{sgn}(y_{i1} - y_{i0}) \cdot \operatorname{sgn}\left((x_{i1} - x_{i0})'\beta\right)
\end{aligned}
$$

Due to the similarity of the objective functions, it follows from the proof of Lemma 6 for Model 1 that Q is maximized exactly on $\tilde{\Theta}_0$.

As before, we can write $Q(\beta) = Pf(\cdot, \beta)$ and $Q_n(\beta) = P_n f(\cdot, \beta)$ for all β where

$$f(x, y, \beta) = 1\{y_1 - y_0 > 0\} \cdot 1\{(x_1 - x_0)'\beta \geq 0\} - 1\{y_1 - y_0 < 0\} \cdot 1\{(x_1 - x_0)'\beta < 0\}.$$

Since x_t has finite support for each t, Q is again a step function and, as with the previous models, $\mathcal{F}$ is a VC subgraph class with envelope $F = 1$. It follows from Lemmas 3 and 5 that if $\tau_n \xrightarrow{\text{p}} 0$ and $n^{1/2}\tau_n \xrightarrow{\text{p}} \infty$, then for any positive sequence r_n with $r_n \to \infty$, $r_n d_H(\hat{\Theta}_n, \Theta_1) \xrightarrow{\text{p}} 0$.

5.3.2. Estimating the Transformation Function

In addition to β, one might also be interested in estimating the transformation function Λ at particular values of y. This section introduces a set estimator for $\Lambda(y)$ which is based on rank conditions similar to those exploited above for estimating β.

The location of $\Lambda(y)$ is not identified,[4] so we focus on estimating differences $\Lambda(y) - \Lambda(\bar{y})$. In practice, it is useful to locate $\bar{y}$ in an informative area of the domain of Λ, such as the median of the sample (Chen, 2002). Now, for a generic value of y, the model implies the following two restrictions on $\Lambda(y) - \Lambda(\bar{y})$:

$$\begin{aligned} P(y_1 \geq y \mid x) \geq P(y_0 \geq \bar{y} \mid x) &\iff (x_1 - x_0)'\beta \geq \Lambda(y) - \Lambda(\bar{y}), \\ P(y_1 \geq \bar{y} \mid x) \geq P(y_0 \geq y \mid x) &\iff (x_1 - x_0)'\beta \geq \Lambda(\bar{y}) - \Lambda(y). \end{aligned}$$

As before, these rank conditions suggest an estimator based on the maximum score principle. These two restrictions cannot be used directly because β is unknown, but we can make use of the consistent estimator $\hat{\Theta}_n$ defined above. For notational simplicity, we now impose the location normalization $\Lambda(\bar{y}) = 0$. We propose constructing a set estimate $\widehat{\Lambda_n(y)}$ for $\Lambda(y)$ which consists of all values of λ which maximize the following objective function, within a tolerance of τ_n, for some $\hat{\beta} \in \hat{\Theta}_n$:

$$\begin{aligned} \Gamma_n(\lambda \mid \hat{\beta}, y, \bar{y}) = \frac{1}{n} \sum_{i=1}^{n} \Big[&(1\{y_{i1} > y\} - 1\{y_{i0} > \bar{y}\})\, 1\{(x_{i1} - x_{i0})'\hat{\beta} \geq \lambda\} \\ &+ (1\{y_{i1} > \bar{y}\} - 1\{y_{i0} > y\})\, 1\{(x_{i1} - x_{i0})'\hat{\beta} \geq -\lambda\} \Big]. \end{aligned} \tag{10}$$

That is, the set estimate is defined as

$$\widehat{\Lambda_n(y)} = \bigcup_{\hat{\beta} \in \hat{\Theta}_n} \left\{ \lambda : \Gamma_n(\lambda \mid \hat{\beta}, y, \bar{y}) \geq \sup_{\lambda'} \Gamma_n(\lambda' \mid \hat{\beta}, y, \bar{y}) - \tau_n \right\},$$

[4] Note that replacing $\Lambda(y)$ by $\tilde{\Lambda}(y) = \Lambda(y) + \alpha$ and u_t by $\tilde{u}_t = u_t + \alpha$ results in an observationally equivalent model.

where $\tau_n \xrightarrow{p} 0$ is a slackness sequence.

This seems to be a reasonable estimator, since $\hat{\Theta}_n$ is consistent, and we find that it has good finite sample performance in Section 6. Studying the asymptotic properties of two-stage estimators such as this one is a clear direction for future research.

6. Monte Carlo Experiments

In this section we summarize the results of a series of Monte Carlo experiments intended to shed light on the finite sample properties of the proposed estimators and inference procedures defined in Section 3.[5] First, we consider the estimator for Model 1 by replicating the following model.

Specification FE1. Observations are generated according to the Model 1 with

$$y_{it} = 1\{x_{i1t} + \beta x_{i2t} + c_i + u_{it} \geq 0\},$$

where x_{i1t} and x_{i2t} are uniformly distributed for each t with $x_{i1t} \in \{-1, 0, 1\}$ and $x_{i2t} \in \{0, 1, 2, 3, 4\}$. The individual effect is generated as $c_i = (x_{i11} + x_{i12} + x_{i21} + x_{i22})/4$ and the disturbances are iid standard Normal draws. The population parameter used in the experiments is $\beta_0 = -0.15$ which yields the identified set $\Theta_0 = [-0.248, -0.003]$.

Figure 1 displays one realization of $Q_n(\beta)$ for this model, with $n = 500$, along with the population objective function $Q(\beta)$. We compare the estimates for several sample sizes in Table 1, which lists the mean estimated set over 1000 replications for each sample size with $\tau_n = C\sqrt{\ln n/n}$ for $C \in \{0.20, 0.10, 0.05, 0.00\}$. For each sample size, the standard deviation of the endpoints of the estimated sets and the coverage frequency are also reported. By definition of consistency, the coverage probability should asymptotically approach one. The reported estimates illustrate the trade-off faced in choosing C in finite samples. Higher mean coverage comes at the cost of a mean lower Hausdorff distance.

As seen in Table 1, smaller constants C used to construct τ_n produce smaller estimated sets, but only at the expense of lower empirical coverage for small sample values of n. One interesting point to note about the estimates in the first panel of Table 1, with $C = 0.20$, is that the upper bound of the estimated reaches a plateau at -0.003 for the small sample sizes shown. This corresponds to the large jump in the objective function at $\beta = -0.003$ that can be seen in Figure 1. Since the sequence $\tau_n = 0.20\sqrt{\ln n/n}$ is large relative to the other panels, the cutoff value does not rise above this jump as quickly.

Table 2 lists, for $m = n^{2/5}$, $m = n^{3/5}$, and $m = n^{4/5}$ respectively, the empirical coverage frequencies of 1000 confidence regions for $1 - \alpha \in \{0.75, 0.90, 0.95\}$. For each of the

[5] Fortran 95 source code to reproduce all figures and tables in this section is available from the author's website at `http://jblevins.org/research/panel`.

1000 datasets used for estimation and for each value of $1-\alpha$, a confidence region was constructed using Algorithm 1. These regions are based on the estimated sets from the same 1000 datasets as before.

Although the confidence regions in this discrete-regressor specification have roughly the desired coverage in small samples, the coverage appears to converge to 1.0 in all cases. This reflects the alternate characterization discussed in Section 3.3 and indicates that R_n has a degenerate limiting distribution (with equal quantiles). The asymptotic confidence regions constructed via subsampling are still valid, but they are conservative in the extreme sense that they are asymptotic probability one confidence sets.

Specification FE2. This specification is identical to Specification FE1, except with $\beta_0 = 1.000$. The resulting identified set is $\Theta_0 = [0.756, 1.327]$.

In Specification FE1, $\hat{\Theta}_n$ appears to be consistent even with $\tau_n = 0$. However, Specification FE2 shows that this is not the case in general since the estimator appears to be inconsistent with $\tau_n = 0$. As shown in Table 3, for very large sample sizes (up to one million observations), the coverage of $\hat{\Theta}_n$ appears to converge to zero (instead of one) and the Hausdorff distance appears to converge to 0.29 instead of zero. The confidence regions, shown in Table 4, also have poor empirical coverage when $\tau_n = 0$ but behave as expected for appropriately-constructed sequences τ_n.

We also consider two binary choice specifications with a continuous but bounded regressor.

Specification FE3. This specification is identical to Specification FE1 with the exception that $x_{i1t} \sim \mathrm{U}(0,1)$ is continuously distributed and bounded on the interval $(0,1)$, $x_{i2t} \sim \mathrm{U}(\{1,2,3\})$ is discrete and uniformly distributed, and $\beta_0 = -0.600$. This model is point identified with $\Theta_0 = [-0.600, -0.600]$.

Specification FE4. This specification is identical to Specification FE3 with the exception that $\beta_0 = -1.000$. The resulting identified set is $\Theta_0 = [-3.000, -0.986]$, where the compact parameter space is $\Theta = [-3, 3]$.

The population objective function for Specification FE3 has a maximum at $\beta_0 = -0.6$, as shown in Figure 2, but this maximum is not very well pronounced. When we change $\beta_0 = -1.0$, the model is no longer point identified, as can be seen in the limiting objective function plotted in Figure 3. The estimates and confidence regions are given in Tables 5 and 6 for Specification FE3 and in Tables 7 and 8 for Specification FE4.

One interesting point is that in the bounded regressor case of Specification FE3, even when Θ_0 is a singleton, the maximum score objective function may still be maximized on a set. Since there is no guidance about how to pick a particular point from this set, one must pick an arbitrary point. The proposed set estimators avoid this problem and remain valid even in the point identified case.

Specification MPH1. This specification replicates a mixed proportional hazards version of Model 3 with

$$\Lambda(y_{it}) = x_{i1t} + \beta x_{i2t} + c_i + u_{it}$$

where

$$\begin{aligned} x_{i1t} &\sim \mathrm{U}(\{-1, 0, 1\}), \\ x_{i2t} &\sim \mathrm{U}(\{-1, -0.5, -0.2, 0.5, 1.0\}), \\ c_i &= (x_{i11} + x_{i12} + x_{i21} + x_{i22})/4, \end{aligned}$$

and u_{it} follows the standard minus type I extreme value distribution. The population parameter is $\beta_0 = 0.100$ and the resulting identified set satisfies $\Theta_0 \subseteq \tilde{\Theta}_0 = [0.003, 0.498]$.

The population objective function and one realization of $Q_n(\beta)$ with $n = 250$ observations for specification MPH1 are plotted in Figure 4. Table 9 displays the corresponding estimates. We also estimate bounds for the transformation function $\Lambda(y)$. The true transformation and estimated bounds are displayed in Figure 5 for $n = 250$ observations.

7. Conclusion

We have developed new conditions for establishing consistency and both regular and irregular rates of convergence for set estimators in partially identified econometric models and proposed methods for performing inference in models whose estimators exhibit arbitrarily fast convergence. We have applied these general results to a standard binary choice panel data models with fixed effects. First we characterize the sharp identified set and we propose a consistent estimator which converges arbitrarily fast with fully discrete regressors and can achieve rates arbitrarily close to $n^{1/3}$ when a continuous regressor is present. We also consider extensions to a lagged dependent variable and panel data duration models. Finally, a series of Monte Carlo experiments illustrates the estimation and inference procedures, which perform as expected.

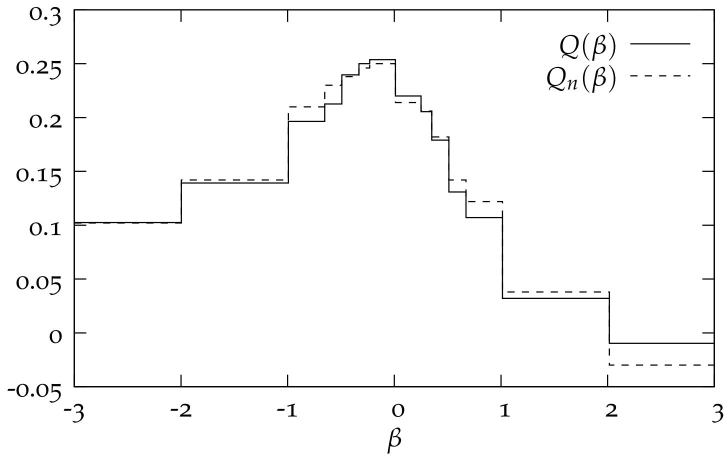

FIGURE 1. $Q(\beta)$ and one realization of $Q_n(\beta)$ for specification FE1 with $n = 500$.

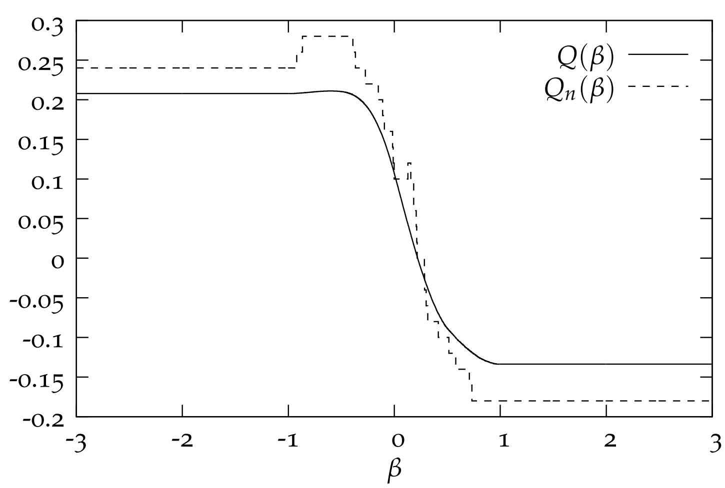

FIGURE 2. $Q(\beta)$ and one realization of $Q_n(\beta)$ for specification FE3 with $n = 100$.

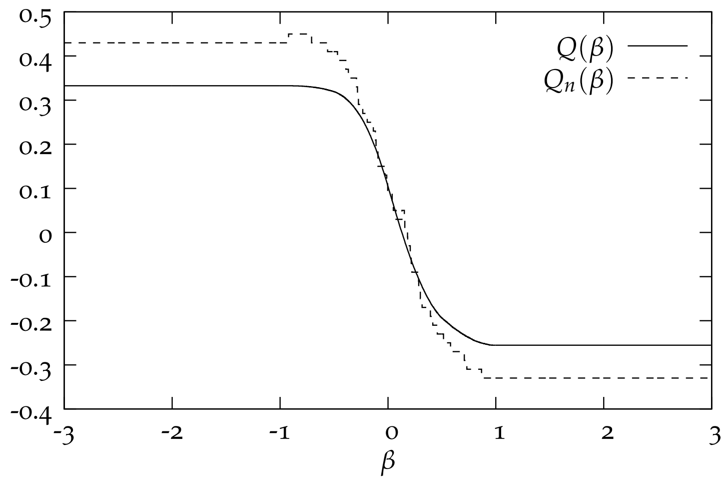

FIGURE 3. $Q(\beta)$ and one realization of $Q_n(\beta)$ for specification FE4 with $n = 100$.

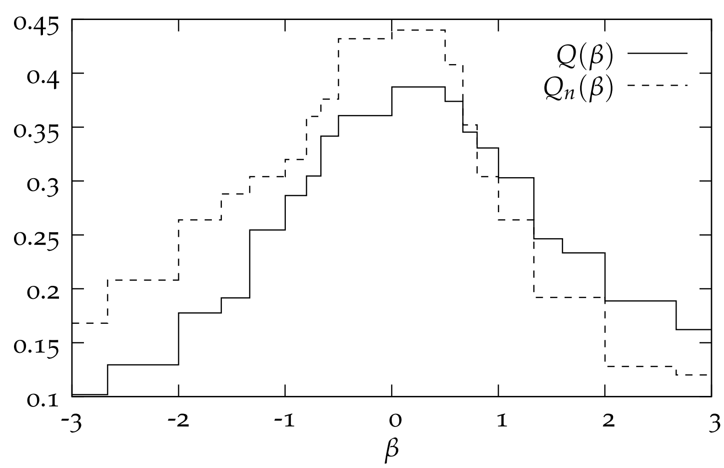

FIGURE 4. $Q(\beta)$ and one realization of $Q_n(\beta)$ specification MPH1 with $n = 250$.

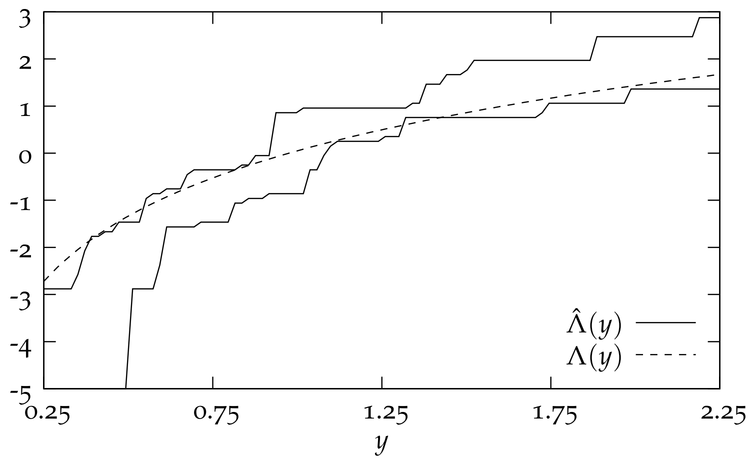

Figure 5. $\Lambda(y)$ and estimated bounds for specification MPH1 with $n = 250$.

C	n	Mean $\hat{\Theta}_n$	St. Dev.	Coverage	d_H
0.20	125	[-0.596, 0.194]	[0.304, 0.232]	0.94	0.45
	250	[-0.520, 0.128]	[0.207, 0.172]	0.97	0.33
	500	[-0.472, 0.085]	[0.127, 0.137]	0.99	0.26
	1000	[-0.436, 0.036]	[0.086, 0.094]	1.00	0.20
	2000	[-0.395, 0.004]	[0.081, 0.039]	1.00	0.15
	4000	[-0.350, -0.003]	[0.053, 0.000]	1.00	0.10
	8000	[-0.329, -0.003]	[0.019, 0.000]	1.00	0.08
	16000	[-0.318, -0.003]	[0.033, 0.000]	1.00	0.07
0.10	125	[-0.473, 0.112]	[0.282, 0.211]	0.85	0.33
	250	[-0.383, 0.044]	[0.180, 0.156]	0.86	0.20
	500	[-0.368, 0.025]	[0.118, 0.113]	0.93	0.15
	1000	[-0.356, 0.010]	[0.085, 0.059]	0.99	0.12
	2000	[-0.329, -0.001]	[0.060, 0.025]	1.00	0.08
	4000	[-0.306, -0.003]	[0.045, 0.000]	1.00	0.06
	8000	[-0.283, -0.003]	[0.042, 0.000]	1.00	0.04
	16000	[-0.260, -0.003]	[0.030, 0.000]	1.00	0.01
0.05	125	[-0.339, 0.012]	[0.240, 0.230]	0.63	0.22
	250	[-0.310, -0.029]	[0.167, 0.181]	0.65	0.16
	500	[-0.327, 0.000]	[0.113, 0.121]	0.87	0.12
	1000	[-0.314, -0.001]	[0.076, 0.061]	0.97	0.08
	2000	[-0.292, -0.006]	[0.054, 0.035]	0.98	0.05
	4000	[-0.271, -0.004]	[0.038, 0.018]	0.99	0.02
	8000	[-0.256, -0.003]	[0.025, 0.000]	1.00	0.01
	16000	[-0.250, -0.003]	[0.012, 0.000]	1.00	0.00
0.00	125	[-0.339, 0.012]	[0.240, 0.230]	0.63	0.22
	250	[-0.310, -0.029]	[0.167, 0.181]	0.65	0.16
	500	[-0.289, -0.037]	[0.111, 0.141]	0.75	0.10
	1000	[-0.275, -0.037]	[0.063, 0.100]	0.83	0.06
	2000	[-0.262, -0.027]	[0.036, 0.077]	0.90	0.03
	4000	[-0.253, -0.012]	[0.020, 0.047]	0.96	0.01
	8000	[-0.249, -0.005]	[0.008, 0.024]	0.99	0.00
	16000	[-0.248, -0.003]	[0.004, 0.008]	1.00	0.00

TABLE 1. Estimates for specification FE1 with $\beta_0 = -0.150$ and $\Theta_0 = [-0.248, -0.003]$.

C	n	$m = n^{2/5}$			$m = n^{3/5}$			$m = n^{4/5}$		
		0.750	0.900	0.950	0.750	0.900	0.950	0.750	0.900	0.950
0.20	125	0.804	0.922	0.963	0.909	0.941	0.948	0.940	0.960	0.967
	250	0.845	0.930	0.966	0.942	0.968	0.976	0.966	0.981	0.982
	500	0.860	0.953	0.976	0.972	0.987	0.991	0.973	0.992	0.994
	1000	0.871	0.973	0.983	0.981	0.996	0.998	0.981	0.994	0.998
	2000	0.936	0.984	0.998	0.993	0.997	0.997	0.996	1.000	1.000
	4000	0.977	0.994	1.000	1.000	1.000	1.000	0.999	1.000	1.000
	8000	0.990	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	16000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.10	125	0.710	0.804	0.879	0.812	0.897	0.910	0.870	0.883	0.891
	250	0.726	0.804	0.841	0.802	0.878	0.907	0.873	0.888	0.897
	500	0.785	0.866	0.905	0.886	0.930	0.945	0.913	0.941	0.958
	1000	0.829	0.909	0.951	0.948	0.982	0.989	0.953	0.979	0.990
	2000	0.912	0.959	0.995	0.983	0.992	0.994	0.991	1.000	1.000
	4000	0.966	0.991	1.000	0.999	1.000	1.000	0.998	1.000	1.000
	8000	0.990	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	16000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.05	125	0.624	0.641	0.654	0.648	0.660	0.660	0.650	0.650	0.651
	250	0.664	0.683	0.695	0.667	0.684	0.684	0.666	0.666	0.667
	500	0.759	0.804	0.834	0.834	0.878	0.896	0.871	0.887	0.894
	1000	0.814	0.870	0.916	0.922	0.963	0.966	0.941	0.960	0.970
	2000	0.908	0.953	0.987	0.978	0.985	0.986	0.984	0.988	0.988
	4000	0.964	0.987	0.994	0.995	0.995	0.995	0.995	0.996	0.996
	8000	0.990	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	16000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.00	125	0.624	0.641	0.654	0.648	0.660	0.660	0.650	0.650	0.651
	250	0.664	0.683	0.695	0.667	0.684	0.684	0.666	0.666	0.667
	500	0.737	0.743	0.748	0.759	0.765	0.765	0.731	0.734	0.737
	1000	0.802	0.804	0.805	0.837	0.837	0.837	0.845	0.846	0.847
	2000	0.905	0.905	0.905	0.899	0.899	0.899	0.918	0.918	0.918
	4000	0.964	0.964	0.964	0.964	0.964	0.964	0.957	0.957	0.957
	8000	0.990	0.990	0.990	0.991	0.991	0.991	0.986	0.986	0.986
	16000	0.999	0.999	0.999	0.999	0.999	0.999	1.000	1.000	1.000

TABLE 2. Empirical coverage of confidence regions for specification FE1.

C	n	Mean $\hat{\Theta}_n$	St. Dev.	Coverage	d_H
0.20	125	[0.404, 2.334]	[0.322, 0.755]	0.83	1.15
	250	[0.513, 2.040]	[0.264, 0.728]	0.83	0.86
	500	[0.590, 1.871]	[0.192, 0.595]	0.88	0.64
	1000	[0.648, 1.615]	[0.128, 0.372]	0.92	0.36
	2000	[0.700, 1.460]	[0.090, 0.210]	0.92	0.19
	4000	[0.735, 1.382]	[0.062, 0.109]	0.94	0.09
	8000	[0.754, 1.327]	[0.041, 0.070]	0.94	0.03
	16000	[0.758, 1.324]	[0.041, 0.052]	0.95	0.01
	32000	[0.759, 1.325]	[0.042, 0.050]	0.95	0.01
	64000	[0.757, 1.327]	[0.037, 0.041]	0.96	0.01
	128000	[0.755, 1.327]	[0.028, 0.042]	0.97	0.01
	256000	[0.754, 1.329]	[0.026, 0.033]	0.98	0.01
	512000	[0.753, 1.329]	[0.022, 0.033]	0.98	0.01
	1024000	[0.754, 1.329]	[0.025, 0.035]	0.98	0.01
0.00	125	[0.789, 1.453]	[0.285, 0.618]	0.23	0.49
	250	[0.841, 1.276]	[0.210, 0.368]	0.16	0.34
	500	[0.874, 1.215]	[0.161, 0.227]	0.10	0.29
	1000	[0.867, 1.181]	[0.134, 0.177]	0.07	0.28
	2000	[0.869, 1.174]	[0.128, 0.169]	0.06	0.28
	4000	[0.872, 1.168]	[0.125, 0.168]	0.03	0.29
	8000	[0.870, 1.162]	[0.125, 0.168]	0.02	0.29
	16000	[0.877, 1.170]	[0.125, 0.168]	0.02	0.29
	32000	[0.868, 1.160]	[0.125, 0.168]	0.02	0.29
	64000	[0.874, 1.164]	[0.125, 0.168]	0.01	0.29
	128000	[0.877, 1.167]	[0.125, 0.168]	0.01	0.29
	256000	[0.872, 1.163]	[0.125, 0.168]	0.01	0.29
	512000	[0.870, 1.158]	[0.125, 0.168]	0.01	0.29
	1024000	[0.878, 1.167]	[0.125, 0.168]	0.00	0.29

TABLE 3. Estimates for specification FE2 with $\beta_0 = 1.000$ and $\Theta_0 = [0.756, 1.327]$.

		Empirical Coverage		
C	n	0.750	0.900	0.950
0.20	125	0.819	0.834	0.841
	250	0.834	0.838	0.863
	500	0.768	0.884	0.903
	1000	0.835	0.900	0.917
	2000	0.807	0.905	0.918
	4000	0.760	0.898	0.926
	8000	0.761	0.900	0.927
	16000	0.721	0.881	0.926
	32000	0.728	0.871	0.927
	64000	0.740	0.899	0.943
	128000	0.753	0.895	0.949
	256000	0.767	0.904	0.952
	512000	0.723	0.896	0.948
	1024000	0.746	0.886	0.941
0.00	125	0.248	0.275	0.275
	250	0.184	0.230	0.230
	500	0.110	0.145	0.148
	1000	0.081	0.089	0.090
	2000	0.060	0.061	0.061
	4000	0.028	0.028	0.028
	8000	0.021	0.021	0.021
	16000	0.016	0.016	0.016
	32000	0.020	0.020	0.020
	64000	0.007	0.007	0.007
	128000	0.007	0.007	0.007
	256000	0.014	0.014	0.014
	512000	0.005	0.005	0.005
	1024000	0.002	0.002	0.002

TABLE 4. Empirical coverage of confidence regions for specification FE2.

C	n	Mean $\hat{\Theta}_n$	St. Dev.	Coverage	d_H
0.20	125	[-2.370, -0.218]	[1.054, 0.148]	0.89	1.88
	250	[-2.422, -0.273]	[1.013, 0.127]	0.90	1.90
	500	[-2.475, -0.289]	[0.958, 0.106]	0.95	1.93
	1000	[-2.582, -0.312]	[0.871, 0.083]	0.98	2.01
	2000	[-2.612, -0.344]	[0.838, 0.067]	0.99	2.03
	4000	[-2.558, -0.364]	[0.878, 0.053]	1.00	1.97
	8000	[-2.427, -0.390]	[0.952, 0.046]	1.00	1.83
	16000	[-2.177, -0.409]	[1.042, 0.038]	1.00	1.58
0.10	125	[-1.921, -0.296]	[1.209, 0.171]	0.72	1.50
	250	[-1.672, -0.390]	[1.198, 0.162]	0.63	1.23
	500	[-1.607, -0.393]	[1.155, 0.139]	0.69	1.13
	1000	[-1.813, -0.388]	[1.143, 0.106]	0.86	1.29
	2000	[-1.708, -0.419]	[1.116, 0.089]	0.89	1.16
	4000	[-1.559, -0.434]	[1.071, 0.074]	0.92	1.00
	8000	[-1.367, -0.452]	[0.985, 0.064]	0.94	0.80
	16000	[-1.048, -0.467]	[0.730, 0.053]	0.97	0.47
0.05	125	[-1.254, -0.427]	[1.154, 0.216]	0.33	0.91
	250	[-1.058, -0.489]	[0.991, 0.198]	0.26	0.66
	500	[-1.213, -0.444]	[1.037, 0.156]	0.50	0.77
	1000	[-1.218, -0.452]	[0.998, 0.133]	0.59	0.74
	2000	[-1.143, -0.482]	[0.917, 0.109]	0.65	0.63
	4000	[-0.976, -0.501]	[0.754, 0.098]	0.65	0.44
	8000	[-0.866, -0.509]	[0.596, 0.081]	0.67	0.32
	16000	[-0.748, -0.516]	[0.327, 0.068]	0.76	0.19
0.00	125	[-1.254, -0.427]	[1.154, 0.216]	0.33	0.91
	250	[-1.058, -0.489]	[0.991, 0.198]	0.26	0.66
	500	[-0.843, -0.516]	[0.777, 0.182]	0.19	0.43
	1000	[-0.748, -0.545]	[0.615, 0.161]	0.15	0.31
	2000	[-0.688, -0.576]	[0.444, 0.140]	0.12	0.21
	4000	[-0.652, -0.589]	[0.307, 0.121]	0.10	0.14
	8000	[-0.615, -0.592]	[0.153, 0.106]	0.08	0.10
	16000	[-0.608, -0.597]	[0.088, 0.087]	0.05	0.08

TABLE 5. Estimates for specification FE3 with $\beta_0 = -0.600$ and $\Theta_0 = [-0.600, -0.600]$.

C	n	$m = n^{2/5}$			$m = n^{3/5}$			$m = n^{4/5}$		
		0.750	0.900	0.950	0.750	0.900	0.950	0.750	0.900	0.950
0.20	125	0.762	0.955	0.985	0.894	0.939	0.966	0.906	0.954	0.966
	250	0.836	0.973	0.988	0.935	0.969	0.982	0.929	0.960	0.973
	500	0.907	0.975	0.981	0.941	0.970	0.989	0.967	0.984	0.990
	1000	0.968	0.986	0.989	0.973	0.994	0.996	0.985	0.995	0.998
	2000	0.980	0.991	0.997	0.987	0.994	0.996	0.987	0.996	0.998
	4000	0.979	0.991	0.996	0.989	0.998	0.999	0.998	0.999	1.000
	8000	0.963	0.989	0.995	0.993	0.997	0.999	0.997	0.998	0.999
	16000	0.982	0.998	1.000	0.998	1.000	1.000	1.000	1.000	1.000
0.10	125	0.555	0.814	0.934	0.778	0.898	0.919	0.781	0.837	0.840
	250	0.505	0.755	0.900	0.767	0.899	0.929	0.711	0.769	0.789
	500	0.610	0.870	0.953	0.843	0.895	0.912	0.840	0.898	0.921
	1000	0.823	0.974	0.983	0.847	0.926	0.965	0.900	0.946	0.963
	2000	0.907	0.981	0.985	0.922	0.969	0.982	0.937	0.958	0.972
	4000	0.956	0.980	0.983	0.947	0.969	0.979	0.953	0.975	0.987
	8000	0.955	0.961	0.973	0.959	0.983	0.988	0.972	0.985	0.994
	16000	0.957	0.969	0.989	0.969	0.984	0.992	0.986	0.995	0.998
0.05	125	0.365	0.443	0.491	0.441	0.471	0.473	0.443	0.456	0.456
	250	0.300	0.368	0.442	0.385	0.468	0.476	0.363	0.379	0.382
	500	0.403	0.672	0.828	0.690	0.833	0.868	0.651	0.731	0.770
	1000	0.515	0.845	0.951	0.740	0.809	0.846	0.754	0.828	0.862
	2000	0.582	0.905	0.963	0.799	0.871	0.912	0.830	0.885	0.909
	4000	0.661	0.935	0.971	0.829	0.896	0.933	0.831	0.893	0.916
	8000	0.778	0.942	0.956	0.860	0.925	0.955	0.903	0.938	0.954
	16000	0.849	0.948	0.961	0.888	0.933	0.958	0.933	0.961	0.976
0.00	125	0.365	0.443	0.491	0.441	0.471	0.473	0.443	0.456	0.456
	250	0.300	0.368	0.442	0.385	0.468	0.476	0.363	0.379	0.382
	500	0.226	0.295	0.379	0.319	0.412	0.417	0.359	0.383	0.394
	1000	0.150	0.222	0.316	0.264	0.358	0.389	0.303	0.329	0.343
	2000	0.135	0.201	0.287	0.263	0.354	0.433	0.339	0.360	0.381
	4000	0.125	0.184	0.261	0.218	0.294	0.362	0.300	0.319	0.337
	8000	0.107	0.174	0.245	0.163	0.240	0.302	0.256	0.293	0.309
	16000	0.064	0.123	0.162	0.154	0.227	0.287	0.198	0.225	0.249

Table 6. Empirical coverage of confidence regions for specification FE3.

C	n	Mean $\hat{\Theta}_n$	St. Dev.	Coverage	d_H
0.20	125	[-2.822, -0.371]	[0.613, 0.134]	0.92	0.74
	250	[-2.924, -0.428]	[0.406, 0.113]	0.97	0.61
	500	[-2.947, -0.456]	[0.336, 0.099]	0.98	0.57
	1000	[-2.987, -0.490]	[0.170, 0.083]	0.99	0.51
	2000	[-2.994, -0.537]	[0.114, 0.072]	1.00	0.45
	4000	[-3.000, -0.575]	[0.000, 0.064]	1.00	0.41
	8000	[-3.000, -0.620]	[0.000, 0.052]	1.00	0.37
	16000	[-3.000, -0.653]	[0.000, 0.042]	1.00	0.33
0.10	125	[-2.530, -0.452]	[0.934, 0.152]	0.80	0.88
	250	[-2.387, -0.550]	[1.003, 0.142]	0.73	0.91
	500	[-2.561, -0.564]	[0.888, 0.125]	0.80	0.76
	1000	[-2.830, -0.581]	[0.582, 0.103]	0.92	0.54
	2000	[-2.889, -0.632]	[0.473, 0.086]	0.95	0.44
	4000	[-2.952, -0.662]	[0.315, 0.073]	0.98	0.36
	8000	[-2.992, -0.700]	[0.129, 0.057]	1.00	0.29
	16000	[-2.998, -0.730]	[0.064, 0.049]	1.00	0.26
0.05	125	[-1.849, -0.575]	[1.185, 0.183]	0.51	1.32
	250	[-1.705, -0.647]	[1.145, 0.161]	0.43	1.41
	500	[-2.215, -0.623]	[1.073, 0.137]	0.65	0.99
	1000	[-2.472, -0.654]	[0.937, 0.116]	0.76	0.76
	2000	[-2.532, -0.698]	[0.890, 0.095]	0.78	0.68
	4000	[-2.586, -0.733]	[0.843, 0.082]	0.81	0.61
	8000	[-2.716, -0.767]	[0.716, 0.066]	0.86	0.47
	16000	[-2.840, -0.791]	[0.554, 0.055]	0.92	0.34
0.00	125	[-1.849, -0.575]	[1.185, 0.183]	0.51	1.32
	250	[-1.705, -0.647]	[1.145, 0.161]	0.43	1.41
	500	[-1.579, -0.706]	[1.103, 0.150]	0.37	1.49
	1000	[-1.468, -0.757]	[1.025, 0.126]	0.30	1.58
	2000	[-1.437, -0.804]	[0.989, 0.111]	0.27	1.59
	4000	[-1.336, -0.832]	[0.908, 0.093]	0.22	1.68
	8000	[-1.261, -0.866]	[0.821, 0.077]	0.17	1.75
	16000	[-1.299, -0.893]	[0.829, 0.063]	0.16	1.71

TABLE 7. Estimates for specification FE4 with $\beta_0 = -1.000$ and $\Theta_0 = [-3.000, -0.986]$.

C	n	$m = n^{2/5}$			$m = n^{3/5}$			$m = n^{4/5}$		
		0.750	0.900	0.950	0.750	0.900	0.950	0.750	0.900	0.950
0.20	125	0.714	0.966	0.986	0.926	0.952	0.970	0.943	0.982	0.982
	250	0.775	0.983	0.992	0.945	0.983	0.994	0.968	0.991	0.996
	500	0.829	0.983	0.993	0.956	0.990	0.994	0.974	0.989	0.995
	1000	0.862	0.992	0.997	0.969	0.995	0.998	0.996	0.999	1.000
	2000	0.887	0.991	0.997	0.986	0.998	1.000	0.998	0.998	0.998
	4000	0.950	0.998	1.000	0.992	0.997	0.998	0.999	1.000	1.000
	8000	0.964	0.997	1.000	0.996	1.000	1.000	0.999	1.000	1.000
	16000	0.983	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.10	125	0.599	0.845	0.966	0.847	0.932	0.938	0.829	0.876	0.879
	250	0.547	0.788	0.924	0.776	0.934	0.944	0.817	0.879	0.886
	500	0.556	0.882	0.966	0.858	0.928	0.953	0.872	0.940	0.953
	1000	0.616	0.945	0.986	0.900	0.968	0.987	0.952	0.977	0.988
	2000	0.627	0.947	0.984	0.937	0.978	0.994	0.981	0.990	0.995
	4000	0.673	0.980	0.998	0.960	0.990	0.993	0.988	0.997	0.998
	8000	0.710	0.984	0.998	0.984	0.993	0.998	0.989	0.997	0.999
	16000	0.735	0.989	0.997	0.988	0.999	1.000	0.994	0.996	0.999
0.05	125	0.507	0.572	0.624	0.580	0.610	0.610	0.580	0.588	0.591
	250	0.441	0.509	0.581	0.502	0.586	0.588	0.530	0.544	0.545
	500	0.457	0.716	0.864	0.724	0.894	0.917	0.774	0.825	0.849
	1000	0.434	0.790	0.924	0.793	0.896	0.922	0.834	0.895	0.926
	2000	0.414	0.798	0.927	0.830	0.924	0.952	0.931	0.962	0.973
	4000	0.394	0.828	0.950	0.880	0.953	0.975	0.916	0.957	0.975
	8000	0.420	0.855	0.964	0.912	0.970	0.989	0.945	0.975	0.985
	16000	0.370	0.874	0.973	0.923	0.981	0.992	0.962	0.983	0.986
0.00	125	0.507	0.572	0.624	0.580	0.610	0.610	0.580	0.588	0.591
	250	0.441	0.509	0.581	0.502	0.586	0.588	0.530	0.544	0.545
	500	0.384	0.444	0.513	0.478	0.544	0.551	0.513	0.528	0.534
	1000	0.309	0.367	0.436	0.412	0.517	0.539	0.459	0.481	0.488
	2000	0.278	0.327	0.393	0.331	0.425	0.488	0.491	0.508	0.519
	4000	0.229	0.270	0.337	0.322	0.424	0.495	0.393	0.412	0.421
	8000	0.199	0.232	0.272	0.279	0.360	0.434	0.359	0.380	0.391
	16000	0.147	0.183	0.220	0.213	0.297	0.368	0.298	0.317	0.334

TABLE 8. Empirical coverage of confidence regions for specification FE4.

C	n	Mean $\hat{\Theta}_n$	St. Dev.	Coverage	d_H
0.20	125	[-0.399, 0.594]	[0.411, 0.424]	0.74	0.56
	250	[-0.329, 0.568]	[0.339, 0.332]	0.80	0.45
	500	[-0.292, 0.554]	[0.298, 0.258]	0.85	0.38
	1000	[-0.206, 0.540]	[0.270, 0.208]	0.89	0.28
	2000	[-0.129, 0.555]	[0.226, 0.126]	0.97	0.19
	4000	[-0.063, 0.516]	[0.169, 0.088]	0.98	0.09
	8000	[-0.019, 0.499]	[0.101, 0.029]	1.00	0.02
	16000	[0.001, 0.498]	[0.032, 0.000]	1.00	0.00
0.10	125	[-0.298, 0.477]	[0.401, 0.413]	0.64	0.46
	250	[-0.210, 0.438]	[0.333, 0.317]	0.67	0.33
	500	[-0.199, 0.445]	[0.274, 0.253]	0.77	0.26
	1000	[-0.143, 0.464]	[0.235, 0.208]	0.85	0.18
	2000	[-0.084, 0.485]	[0.192, 0.138]	0.94	0.10
	4000	[-0.040, 0.488]	[0.140, 0.085]	0.97	0.05
	8000	[-0.007, 0.497]	[0.068, 0.022]	1.00	0.01
	16000	[0.002, 0.498]	[0.016, 0.000]	1.00	0.00
0.05	125	[-0.168, 0.375]	[0.427, 0.409]	0.52	0.38
	250	[-0.139, 0.372]	[0.353, 0.323]	0.57	0.29
	500	[-0.161, 0.408]	[0.279, 0.252]	0.71	0.23
	1000	[-0.120, 0.428]	[0.234, 0.213]	0.80	0.16
	2000	[-0.071, 0.460]	[0.179, 0.146]	0.91	0.08
	4000	[-0.033, 0.480]	[0.129, 0.097]	0.96	0.04
	8000	[-0.003, 0.496]	[0.052, 0.035]	0.99	0.01
	16000	[0.003, 0.498]	[0.000, 0.000]	1.00	0.00
0.00	125	[-0.168, 0.375]	[0.427, 0.409]	0.52	0.38
	250	[-0.139, 0.372]	[0.353, 0.323]	0.57	0.29
	500	[-0.127, 0.372]	[0.294, 0.256]	0.64	0.21
	1000	[-0.094, 0.400]	[0.233, 0.213]	0.76	0.13
	2000	[-0.055, 0.444]	[0.168, 0.160]	0.88	0.06
	4000	[-0.022, 0.474]	[0.110, 0.108]	0.95	0.03
	8000	[-0.001, 0.495]	[0.039, 0.039]	0.99	0.00
	16000	[0.003, 0.498]	[0.000, 0.000]	1.00	0.00

TABLE 9. Estimates for specification MPH1 with $\beta_0 = 0.100$ and $\tilde{\Theta}_0 = [0.003, 0.498]$.

A. Notation and Preliminary Results

First we introduce some notation. We shall make use of a modified signum function $\operatorname{sgn}(x)$ where

$$\operatorname{sgn}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0. \end{cases}$$

This definition, which is standard in the maximum score literature, differs from the common definition only at zero, where we define $\operatorname{sgn}(0) = 1$ instead of $\operatorname{sgn}(0) = 0$. We write $a \vee b$ to denote $\max\{a, b\}$ and $a \wedge b$ to denote $\min\{a, b\}$.

Let $l^\infty(B)$ denote the space of uniformly bounded, real-valued functions on B endowed with the uniform metric $d_\infty(g, h) \equiv \sup_{\theta \in \Theta} |g(\theta) - h(\theta)|$. Let $\mathbb{G}_n = \sqrt{n}(P_n - P)$ denote the standardized empirical process.

Lemma 9. *Let $\{X_n\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ and let X be a random variable defined on the same space. Then $X_n = X$ with probability approaching one*[6] *if and only if for any positive sequence r_n with $r_n \to \infty$, $r_n(X_n - X) \xrightarrow{P} 0$.*

Proof of Lemma 9. Suppose that $X_n = X$ with probability approaching one and let $\varepsilon > 0$. Given any positive sequence r_n, for all n we have

$$\begin{aligned} P(X_n \neq X) &= P(|X_n - X| > \varepsilon/r_n) + P(|X_n - X| \leq \varepsilon/r_n) \\ &\geq P(|X_n - X| > \varepsilon/r_n) \\ &= P(|r_n(X_n - X)| > \varepsilon) \end{aligned}$$

Since $\lim_{n\to\infty} P(X_n \neq X) \to 0$ and $P(X_n \neq X) \geq P(|r_n(X_n - X)| > \varepsilon) \geq 0$, it follows that $\lim_{n\to\infty} P(|r_n(X_n - X)| > \varepsilon) = 0$.

Now, suppose that for all positive sequences r_n with $r_n \to \infty$, $r_n(X_n - X) \xrightarrow{P} 0$. Let $\varepsilon > 0$ and observe that

$$\begin{aligned} P(X_n = X) &= P(r_n X_n = r_n X) \\ &= P(r_n(X_n - X) = 0) \\ &= 1 - P(r_n(X_n - X) \neq 0) \\ &= 1 - P(|r_n(X_n - X)| \geq 0) \\ &= 1 - P(|r_n(X_n - X)| \geq \varepsilon) - P(0 < |r_n(X_n - X)| < \varepsilon) \end{aligned}$$

The first probability above is $o(1)$ while the second can be made arbitrarily small by choice of ε. Therefore, $\lim_{n\to\infty} P(X_n = X) = 1$. ■

Lemma 10. *Let f and g be bounded real functions on $A \subset \mathbb{R}^n$. Then*

$$\left| \sup_{x \in A} f(x) - \sup_{x \in A} g(x) \right| \leq \sup_{x \in A} |f(x) - g(x)| .$$

Proof of Lemma 10. First, note that for all $x \in A$,

$$f(x) - \sup_{y \in A} g(y) \leq f(x) - g(x) \leq |f(x) - g(x)| \tag{11}$$

and

$$\sup_{y \in A} f(y) - g(x) \geq f(x) - g(x) \geq - |f(x) - g(x)| . \tag{12}$$

[6] That is, $\lim_{n\to\infty} P(\{\omega \in \Omega : X_n(\omega) = X(\omega)\}) = 1$.

We prove the result by showing that

$$-\sup_{x\in A}|f(x)-g(x)| \leq \sup_{x\in A} f(x) - \sup_{x\in A} g(x) \leq \sup_{x\in A}|f(x)-g(x)|.$$

For the right hand side:

$$\sup_{x\in A} f(x) - \sup_{x\in A} g(x) = \sup_{x\in A}\left[f(x) - \sup_{y\in A} g(y)\right] \leq \sup_{x\in A}|f(x)-g(x)|.$$

The equality holds since $\sup g$ is constant with respect to x and the inequality follows from (11), since it holds for all x. Similarly, the left hand side follows from (12):

$$\begin{aligned}
\sup_{x\in A} f(x) - \sup_{x\in A} g(x) &= \sup_{x\in A} f(x) + \inf_{x\in A}(-g(x)) \\
&= \inf_{x\in A}\left[\sup_{y\in A} f(y) - g(x)\right] \\
&\geq \inf_{x\in A} -|f(x)-g(x)| \\
&= -\sup_{x\in A}|f(x)-g(x)|
\end{aligned}$$

Together, these two inequalities imply the result. ■

B. Estimation and Inference in General Models

B.1. *Proof of Theorem 1*

The proof proceeds in two steps. In the first step, we show that $\sup_{\theta\in\hat{\Theta}_n}\rho(\theta,\Theta_1)\xrightarrow{p}0$. The second step shows that $\lim_{n\to\infty}P(\Theta_1\subset\hat{\Theta}_n)=1$. Combining these steps and using the definition of the Hausdorff distance yields the final conclusion of the theorem.

Step 1 For any $\varepsilon>0$,

$$\sup_{\Theta\backslash\Theta_1^\varepsilon} Q_n \leq \sup_{\Theta\backslash\Theta_1^\varepsilon} Q + o_p(1) \leq \sup_{\Theta} Q - \delta_\varepsilon + o_p(1),$$

where $\delta_\varepsilon>0$. The first inequality above follows from A3, giving uniform convergence in probability of Q_n to Q. The second inequality follows from A2, since Θ_1 maximizes Q. Similarly,

$$\inf_{\hat{\Theta}_n} Q_n \geq \sup_{\Theta} Q_n - \tau_n \geq \sup_{\Theta} Q - \tau_n + o_p(1)$$

The first inequality follows from the definition of $\hat{\Theta}_n$ and the second follows again from uniform convergence. By assumption, $\tau_n=o_p(1)$, and since $\delta_\varepsilon>0$, with probability approaching one, $\tau_n<\delta_\varepsilon$, or equivalently, $\sup_\Theta Q-\tau_n+o_p(1)\geq\sup_\Theta Q-\delta_\varepsilon+o_p(1)$. Given the inequalities above, this implies $\inf_{\hat{\Theta}_n}Q_n\geq\sup_{\Theta\backslash\Theta_1^\varepsilon}Q_n$, which in turn implies that $\Theta_n\subseteq\Theta_1^\varepsilon$, and so $\sup_{\theta\in\hat{\Theta}_n}\rho(\theta,\Theta_1)\leq\varepsilon$.

Step 2 By definition of $\hat{\Theta}_n$ and τ_n, we know that if $b_n\tau_n \geq \sup_\Theta b_n Q_n - \inf_{\Theta_1} b_n Q_n$, then $\Theta_1 \subseteq \hat{\Theta}_n$. We have

$$\begin{aligned}
\sup_\Theta Q_n - \inf_{\Theta_1} Q_n &= \left[\sup_\Theta Q_n - \sup_\Theta Q\right] + \left[\sup_\Theta Q - \inf_{\Theta_1} Q_n\right] \\
&\leq \left|\sup_\Theta Q_n - \sup_\Theta Q\right| + \left|\sup_\Theta Q - \inf_{\Theta_1} Q_n\right| \\
&= \left|\sup_\Theta Q_n - \sup_\Theta Q\right| + \left|\sup_{\Theta_1} Q - \inf_{\Theta_1} Q_n\right| \\
&\leq \sup_\Theta |Q_n - Q| + \sup_{\Theta_1} |Q_n - Q| \\
&\leq \sup_\Theta |Q_n - Q| + \sup_{\Theta} |Q_n - Q|
\end{aligned}$$

These steps follow by, respectively, adding and subtracting $\sup_\Theta Q$, taking the absolute value, noting that Θ_1 maximizes Q, using the fact that $\inf f = -\sup -f$, and applying Lemma 10 twice (see Appendix A), noting that $\Theta_1 \subseteq \Theta$. By A3, $\sup_\Theta |Q_n - Q| = O_p(1/b_n)$ and so the requirement that $b_n\tau_n \xrightarrow{\text{p}} \infty$ (i.e., that τ_n approaches zero in probability slower than $1/b_n$) implies that $\tau_n \geq 2\sup_\Theta |Q_n - Q| \geq \sup_\Theta Q_n - \inf_{\Theta_1} Q_n$ with probability approaching one.

B.2. Proof of Theorem 2

From Theorem 1, $\lim_{n\to\infty} P(\Theta_1 \subseteq \hat{\Theta}_n) = 1$. We will prove the result by showing that $\lim_{n\to\infty} P(\hat{\Theta}_n \subseteq \Theta_1) = 1$ and therefore the Hausdorff distance $d_H(\hat{\Theta}_n, \Theta_1)$ eventually equals zero with probability approaching one.

Uniform convergence at the b_n rate, from A3, implies $Q_n(\theta) \leq Q(\theta) + O_p(1/b_n)$ and $Q(\theta) \leq Q_n(\theta) + O_p(1/b_n)$. It follows that

$$\sup_{\Theta\setminus\Theta_1} Q_n \leq \sup_{\Theta\setminus\Theta_1} Q + O_p(1/b_n) \leq \sup_\Theta Q - \delta + O_p(1/b_n) \leq \sup_\Theta Q_n - \delta + O_p(1/b_n),$$

where the second inequality follows from the constant majorant condition.

Since τ_n converges to zero in probability and $\delta > 0$ is constant, with probability approaching one, $\tau_n < \delta$. Thus, with probability approaching one, $-\delta < -\tau_n$, $\sup_{\Theta\setminus\Theta_1} Q_n \leq \sup_\Theta Q_n - \tau_n + O_p(1/b_n) \leq \inf_{\hat{\Theta}_n} Q_n + O_p(1/b_n)$, and therefore, $\hat{\Theta}_n \subseteq \Theta_1$.

B.3. Proof of Theorem 3

For any $\varepsilon > 0$, let δ, κ_0, κ_1, γ_1, γ_2, κ_ε, and n_ε satisfy A5 and define

$$\nu_n \equiv \left(\frac{\kappa_1 \cdot \kappa_\varepsilon \vee 2\tau_n \cdot b_n}{b_n \cdot \kappa_1}\right)^{1/\gamma_1}$$

where b_n is given by A3 and $b_n\tau_n \xrightarrow{\text{p}} \infty$ by construction. Furthermore, by Assumption A3, $\kappa_0 = \sup_\Theta Q$. By definition of ν_n, we have $\nu_n = o_p(1)$ and

$$\nu_n^{1/\gamma_2} \geq \left(\frac{\kappa_\varepsilon}{b_n}\right)^{\frac{1}{\gamma_1\gamma_2}} \geq \frac{\kappa_\varepsilon}{b_n},$$

since $\gamma_1\gamma_2 \geq 1$. Therefore, there exists an $n'_\varepsilon \geq n_\varepsilon$ such that for all $n \geq n'_\varepsilon$, with probability at least $1-\varepsilon$ we have both $\nu_n \leq \delta$ and $\nu_n \geq (\kappa_\varepsilon/b_n)^{\gamma_2}$. Therefore, by A3 and A5,

$$\sup_{\Theta\setminus\Theta_1^{\nu_n}} Q_n \leq \sup_\Theta Q - \kappa_1 \cdot (\nu_n \wedge \delta)^{\gamma_1} \leq \sup_\Theta Q - \kappa_1 \cdot \nu_n^{\gamma_1} = \sup_\Theta Q - 2\cdot\tau_n \leq \sup_\Theta Q_n - \tau_n \leq \inf_{\hat{\Theta}_n} Q_n.$$

This implies that $\hat{\Theta}_n \cap (\Theta \setminus \Theta_1^{\nu_n})$ is empty, or equivalently, that $\hat{\Theta}_n \subseteq \Theta_1^{\nu_n}$. Therefore, in light of Step 1 of the proof of Theorem 1, which shows that $\lim_{n\to\infty} P(\Theta_1 \subseteq \hat{\Theta}_n) = 1$, we have $d_H(\hat{\Theta}_n, \Theta_1) \le \nu_n$ and therefore $d_H(\hat{\Theta}_n, \Theta_1) = O_p(\tau_n^{\gamma_2})$.

B.4. Proof of Theorem 4

Before proceeding with the proof, we first establish the following two lemmas.

Lemma 11. *If A7 holds, then for any sequence $\hat{c}_n$ such that $\hat{c}_n \xrightarrow{p} c(1-\alpha) \equiv \inf\{c : P\{R \le c\} \ge 1-\alpha\}$ for some $\alpha \in (0,1)$,*

$$P\{\Theta_1 \subseteq C_n(\hat{c}_n)\} \ge (1-\alpha) + o_p(1).$$

Proof of Lemma 11. Observe that

$$P\{\Theta_1 \subseteq C_n(\hat{c}_n)\} = P\{R_n \le \hat{c}_n\} = P\{R \le c(1-\alpha)\} + o_p(1) \ge (1-\alpha) + o_p(1).$$

The first equality holds by definition of C_n and R_n, the second by A7 and $\hat{c}_n \xrightarrow{p} c(1-\alpha)$, and the third by definition of $c(1-\alpha)$. ■

Lemma 12. *Suppose A7 holds, let Θ_n be a sequence of subsets of Θ such that $d_H(\Theta_n, \Theta_1)$ converges to zero arbitrarily fast in probability, and let $R_n' = \sup_\Theta b_n Q_n - \inf_{\Theta_n} b_n Q_n$. Then $P(R_n' \le c) \to P(R \le c)$ for each $c \in \mathbb{R}$.*

Proof of Lemma 12. For all $\varepsilon > 0$, there exists an n_ε such that for all $n \ge n_\varepsilon$, $P(\Theta_n = \Theta_1) \ge 1-\varepsilon$. Then, $P\left(\inf_{\Theta_n} n^{1/2} Q_n = \inf_{\Theta_1} n^{1/2} Q_n\right) \ge 1-\varepsilon$. From A7, we have that that $\inf_{\theta\in\Theta_1} n^{1/2} Q_n \xrightarrow{d} R$. Therefore, $R_n' \xrightarrow{d} R$. ■

The proof of Theorem 4 proceeds in three steps. First, we derive upper and lower bounds for $\hat{R}_{n,m,j}$ such that $\underline{R}_{n,m,j} \le \hat{R}_{n,m,j} \le \overline{R}_{n,m,j}$ with probability approaching one. Next, we prove that the empirical distribution function of $\hat{R}_{n,m,j}$ converges in probability to the distribution function of R, the limiting distribution of R_n. Finally, we show that $\hat{c}_n$ converges in probability to $c(1-\alpha)$, the desired quantile of the distribution of R.

Step 1 By Theorem 2, we have $d_H(C_n(\kappa_n), \Theta_1) = 0$ with probability approaching one and thus, $d_H(C_n(\kappa_n), \Theta_1) \le \varepsilon_n$ for some sequence ε_n which converges to zero arbitrarily fast in probability. For a fixed subsample j, let $\underline{R}_{n,m,j} \equiv \sup_{\theta\in\Theta} b_m Q_{n,m,j}(\theta) - \inf_{\theta\in\Theta_1^{\varepsilon_n}} b_m Q_{n,m,j}(\theta)$. Let $\mathcal{K}_n$ be the collection of all subsets $K \subseteq \Theta$ such that $d_H(K, \Theta_1) \le \varepsilon_n$ and define $\overline{R}_{n,m,j} \equiv \sup_{K\in\mathcal{K}_n} \left[\sup_{\theta\in\Theta} b_m Q_{n,m,j}(\theta) - \inf_{\theta\in K} b_m Q_{n,m,j}(\theta)\right]$. There exists a set $\Theta_{n,m,j} \in \mathcal{K}_n$ such that $R_{n,m,j}$ is equal to $\inf_{\theta\in\Theta_{n,m,j}} b_m Q_{n,m,j}(\theta)$. With probability approaching one, since $C_n(\kappa_n) \subseteq \Theta_1^{\varepsilon_n}$ and $C_n(\kappa_n) \in \mathcal{K}_n$, we have $\underline{R}_{n,m,j} \le \hat{R}_{n,m,j} \le \overline{R}_{n,m,j}$ for all $j = 1, \ldots, M_n$.

Step 2 From Step 1, with probability approaching one,

$$\underline{G}_{n,m}(x) \equiv M_n^{-1} \sum_{j=1}^{M_n} 1\{\overline{R}_{n,m,j} \le x\} \le \hat{G}_{n,m}(x) \equiv M_n^{-1} \sum_{j=1}^{M_n} 1\{\hat{R}_{n,m,j} \le x\}$$
$$\le \overline{G}_{n,m}(x) \equiv M_n^{-1} \sum_{j=1}^{M_n} 1\{\overline{R}_{n,m,j} \le x\}.$$

We will show that $\underline{G}_{n,m}(x) \xrightarrow{p} P\{R \le x\}$ and $\overline{G}_{n,m}(x) \xrightarrow{p} P\{R \le x\}$ as $n \to \infty$ (and thus, $m \to \infty$). Therefore, $\hat{G}_{n,m}(x) \xrightarrow{p} P\{R \le x\}$ for each $x \in \mathbb{R}$.

Let $\overline{J}_m(x)$ denote the cdf of $\overline{R}_{n,m,j}$. Note that $\underline{G}_{n,m}(x)$ is a U-statistic of degree m with $0 \leq \underline{G}_{n,m}(x) \leq 1$ (i.e., it is bounded). Furthermore, $\mathrm{E}[\underline{G}_{n,m}(x)] = \mathrm{E}[1\{\overline{R}_{n,m,j} \leq x\}] = \overline{J}_m(x)$, where the last equality holds by nonreplacement sampling, since each subsample of size m is itself an iid sample. Under A6, by the Hoeffding inequality for bounded U-statistics for iid data (Serfling, 1980, Theorem A, p. 201), for any $t > 0$,

$$P\left\{\underline{G}_{n,m}(x) - \overline{J}_m(x) \geq t\right\} \leq \exp\left[-2t^2 \frac{n}{m}\right].$$

A similar inequality follows for $t < 0$ by considering the U-process $-\underline{G}_{n,m}(x)$. Therefore, $\underline{G}_{n,m}(x) = \overline{J}_m(x) + o_p(1)$ for fixed m. Finally, since $\overline{R}_{n,m,j}$ is obtained from sets satisfying the assumptions of Lemma 12, we have $\overline{J}_m(x) = P\{\overline{R}_{n,m,j} \leq x\} = P\{\overline{R} \leq x\} + o_p(1)$.

A similar argument shows that $\overline{G}_{n,m}(x) \xrightarrow{\mathrm{P}} P\{R \leq x\}$ as well, and therefore, $\hat{G}_{n,m}(x) \xrightarrow{\mathrm{P}} P\{R \leq x\}$.

Step 3 Convergence of the distribution function at continuity points implies convergence of the quantile function at continuity points (cf. Shorack, 2000, Proposition 3.1). Therefore, $\hat{c}_n = \inf\{x : \hat{G}(x) \geq 1 - \alpha\} \xrightarrow{\mathrm{P}} c(1 - \alpha)$.

C. Sufficient Conditions

C.1. Proof of Lemma 4

We show the result by verifying the conditions of Theorem 3. Since C1 implies A1, C2 implies A2, and, as shown in Lemma 2, C3 is sufficient for A3 with $b_n = n_{1/2}$, it suffices to establish A5.

By definition of $\mathbb{G}_n(\theta)$, we can always write

$$Q_n(\theta) = (P_n - P)f(\cdot, \theta) + Pf(\cdot, \theta) = n^{-1/2}\mathbb{G}_n(\theta) + Q(\theta). \tag{13}$$

Choose ζ smaller than the minimum of ν and η_0. Recall that C2 implies A2, so there exists a $\delta_\zeta > 0$ such that

$$\sup_{\Theta \setminus \Theta_1^\zeta} Q \leq \sup_\Theta Q - 2\delta_\zeta. \tag{14}$$

Combining (13) and (14) and using A3 gives, for all $\theta \in \Theta \setminus \Theta_1^\zeta$,

$$Q_n(\theta) \leq n^{-1/2}\mathbb{G}_n(\theta) + \sup_\Theta Q - 2\delta_\zeta.$$

$\mathcal{F}$ is P-Donsker by C3 and since $\sup_\Theta |\cdot|$ is continuous on $l^\infty(\Theta)$, $\sup_\Theta |\mathbb{G}_n(\theta)| = O_p(1)$ by the continuous mapping theorem. It follows that for any $\varepsilon_1 \in (0, 1)$ there exists an n_1 so that for all $n \geq n_1$,

$$Q_n(\theta) \leq \sup_\Theta Q - \delta_\zeta \tag{15}$$

uniformly on $\Theta \setminus \Theta_1^\zeta$ with probability at least $1 - \varepsilon_1$.

Now, by C4, there is a neighborhood Θ_1^ν of Θ_1 such that Q is approximately quadratic in the distance $\rho(\theta, \Theta_1)$. That is,

$$Q(\theta) \leq \sup_\Theta Q - K_1 \cdot \rho^2(\theta, \Theta_1)$$

for all $\theta \in \Theta_1^\nu$. Similarly, by C5 and Lemma 4.1 of Kim and Pollard (1990), for all $K_2 > 0$ there exists a sequence of random variables $M_n = O_p(1)$ such that

$$(P_n - P)f(\cdot, \theta) \leq K_2 \cdot \rho^2(\theta, \Theta_1) + n^{-2/3}M_n^2$$

for $\theta \in \Theta_1^{\eta_0}$. Combining these results for $K_2 = K_1/2$ and using (13) and A3 yields

$$Q_n(\theta) \leq \sup_{\Theta} Q - \frac{K_1}{2} \cdot \rho^2(\theta, \Theta_1) + n^{-2/3} M_n^2$$

for all $\theta \in \Theta_1^{\nu \wedge \eta_0}$, where M_n and R_n are both $O_p(1)$.

Notice that when $n^{-2/3} M_n^2$ is smaller than $K_1/2 \cdot \rho^2(\theta, \Theta_1)$, we have

$$Q_n(\theta) \leq \sup_{\Theta} Q - \frac{K_1}{4} \rho^2(\theta, \Theta_1), \tag{16}$$

which is a polynomial majorant of the form required by A5. This is true whenever $\rho(\theta, \Theta_1) \geq 2K_1^{-1/2} n^{-1/3} M_n$. Since $M_n = O_p(1)$, for any $\varepsilon_2 \in (0,1)$, there exists a K_3 and n_2 such that for all $n \geq n_2$, $\rho(\theta, \Theta_1) \geq K_3 n^{-1/3} \geq 2K_1^{-1/2} n^{-1/3} M_n$ and the bound in (16) holds uniformly on $\Theta_1^{\zeta} \setminus \Theta_1^{K_3 n^{-1/3}}$. with probability at least $1 - \varepsilon_2$. (Note that we can always choose n_2 large enough so that $K_3 n^{-1/3}$ is smaller than $\zeta < \nu \wedge \eta_0$, ensuring that the relevant region of the domain is nonempty.)

To show that A5 holds, let $\varepsilon \in (0,1)$ be given. For $\varepsilon_1 = \varepsilon/2$, choose n_1 and δ_ζ as above so that (15) holds uniformly on $\Theta \setminus \Theta_1^{\zeta}$ with probability at least $1 - \varepsilon_1$, where $\zeta < \nu \wedge \eta_0$. Then, for $\varepsilon_2 = \varepsilon/2$, choose n_2 and K_3 such that (16) holds uniformly on $\Theta_1^{\zeta} \setminus \Theta_1^{K_3 n^{-1/3}}$ with probability at least $1 - \varepsilon_2$.

To summarize, we have shown that

$$Q_n(\theta) \leq \sup_{\Theta} Q - \max\left\{ \frac{K_1}{4} \rho^2(\theta, \Theta_1), \delta_\zeta \right\}$$

uniformly on $\Theta \setminus \Theta_1^{K_3 n^{-1/3}}$ with probability at least $1 - \varepsilon$. It follows that A5 holds with $b_n = n^{1/2}$, $\delta = \delta_\zeta$, $c_0 = \sup_\Theta Q$, $c_1 = \frac{K_1}{4}$, $\gamma_1 = 2$, $\gamma_2 = 2/3$, $c_\varepsilon = K_3$, and $n_\varepsilon = \max\{n_1, n_2\}$. Therefore, for any sequence r_n such that $r_n = o(n^{1/3})$, let $\tau_n \propto r_n^{-3/2}$. Since $n^{1/3} r_n^{-1} \to \infty$, $(n^{1/3} r_n^{-1})^{3/2} = n^{1/2} r_n^{-3/2} \propto n^{1/2} \tau_n \xrightarrow{p} \infty$ and therefore Theorem 3 implies $d_H(\hat{\Theta}_n, \Theta_1) = O_p(\tau_n^{\gamma_2}) = O_p(\tau_n^{2/3}) = O_p(r_n)$.

D. Fixed Effects Model

D.1. Proof of Theorem 5

For the proof, let Θ_0 denote the true identified set defined in (1) and let $\tilde{\Theta}_0$ denote the set on the right side of (6). We first show $\Theta_0 \subseteq \tilde{\Theta}_0$, and then $\tilde{\Theta}_0 \subseteq \Theta_0$.

Step 1 Let $\theta \in \Theta_0$. By definition of Θ_0, there exist distributions $F_{u_0|xc}$ and $F_{c|x}$ such that $\pi(y_t = 1 \mid x; \beta, F_{u_0|xc}, F_{c|x}) = P(y_t = 1 \mid x)$ F_x-almost surely for $t = 0, 1$. Conditioning on c, we have $P(y_0 = 1 \mid x, c) = 1 - F_{u_0|xc}(-x_0'\beta - c)$ and $P(y_1 = 1 \mid x, c) = 1 - F_{u_0|xc}(-x_1'\beta - c)$. By the monotonicity of $F_{u_0|xc}$,

$$\begin{aligned} P(y_1 = 1 \mid x, c) \geq P(y_0 = 1 \mid x, c) &\iff 1 - F_{u_0|xc}(-x_1'\beta - c) \geq 1 - F_{u_0|xc}(-x_0'\beta - c) \\ &\iff F_{u_0|xc}(-x_1'\beta - c) \leq F_{u_0|xc}(-x_0'\beta - c) \\ &\iff -x_1'\beta - c \leq -x_0'\beta - c \\ &\iff (x_1 - x_0)'\beta \geq 0 \end{aligned}$$

The third line follows from the assumption that the support of u_t is $\mathbb{R}$.[7] Since this event is independent of c, we have

$$P(y_1 = 1 \mid x) - P(y_0 = 1 \mid x) \geq 0 \iff (x_1 - x_0)'\beta \geq 0,$$

[7] For any random variable Z, $F_Z(z_1) \leq F_Z(z_2)$ implies $z_1 \leq z_2$ only on the support of Z.

or, equivalently,

$$\operatorname{sgn}\left(P(y_1 = 1 \mid x) - P(y_0 = 1 \mid x)\right) = \operatorname{sgn}\left((x_1 - x_0)'\beta\right).$$

Therefore, $\theta \in \Theta_0 \Rightarrow \theta \in \tilde{\Theta}_0$.

Step 2 Now, suppose $\theta \in \tilde{\Theta}_0$. We will show that for each such θ, given population distributions $P(y_t \mid x)$ for $t = 0, 1$, there are values of the remaining free model primitives—the cdfs $F_{u_0|xc}$ and $F_{c|x}$—such that the implications of the model coincide with the true population values $P(y_0 = 0 \mid x)$ and $P(y_1 = 0 \mid x)$.

First, note that we do not need to consider the events $y_0 = 1$ or $y_1 = 1$ since in each time period, the (binary) choice probabilities must sum to one. Thus, we need to show that there exist distributions $F_{u_0|xc}$ and $F_{c|x}$ such that for F_x-almost every x the model implications align with the population choice probabilities:

$$\begin{aligned} P(y_0 = 0 \mid x) &= \pi(y_0 = 0 \mid x; \theta, F_{u_0|xc}, F_{c|x}) \\ P(y_1 = 0 \mid x) &= \pi(y_1 = 0 \mid x; \theta, F_{u_0|xc}, F_{c|x}) \end{aligned}$$

For a given x and for primitives $(\theta, F_{u_0|xc}, F_{c|x})$, the model implications are:

$$\begin{aligned} \pi(y_0 = 0 \mid x; \theta, F_{u_0|xc}, F_{c|x}) &= \int F_{u_0|xc}(-x_0'\beta - c)\, dF_{c|x} \\ \pi(y_1 = 0 \mid x; \theta, F_{u_0|xc}, F_{c|x}) &= \int F_{u_0|xc}(-x_1'\beta - c)\, dF_{c|x} \end{aligned}$$

Fix x. It will suffice to construct a distribution $F_{c|x}$ with only a single mass point $c^*(x)$ (conditional on each fixed value of x):

$$F_{c|x}(c) = \begin{cases} 0 & \text{if } c < c^*(x), \\ 1 & \text{if } c \geq c^*(x). \end{cases}$$

Suppose that $P(y_1 = 1 \mid x) < P(y_0 = 1 \mid x)$ (the opposite case follows similarly). Then our choice of $\theta \in \tilde{\Theta}_0$ guarantees that β is such that $x_1'\beta < x_0'\beta$. We can rewrite these two inequalities equivalently as $P(y_0 = 0 \mid x) < P(y_1 = 0 \mid x)$ and $-x_0'\beta < -x_1'\beta$. Thus, the following choice for $F_{u_0|xc}$ is a valid cdf:

$$F_{u_0|xc}(u) = \begin{cases} 0 & \text{if } u < -x_1'\beta - c^*(x), \\ P(y_0 = 0 \mid x) & \text{if } -x_0'\beta - c^*(x) \leq u \leq -x_1'\beta - c^*(x), \\ P(y_1 = 0 \mid x) & \text{if } -x_1'\beta - c^*(x) \leq u < \bar{u}, \\ 1 & \text{if } u \geq \bar{u}, \end{cases}$$

for any $\bar{u} > -x_1'\beta - c^*(x)$. Essentially, we only need to choose a cdf that passes through the two points $\left(-x_0'\beta - c^*(x), P(y_0 = 0 \mid x)\right)$ and $\left(-x_1'\beta - c^*(x), P(y_1 = 0 \mid x)\right)$ and there are an infinite number of such cdfs, as illustrated by Figure 6.

Given the above cdfs, we have:

$$\begin{aligned} \pi(y_0 = 0 \mid x; \theta, F_{u_0|xc}, F_{c|x}) &= F_{u_0|xc}(-x_0'\beta - c^*(x)) = P(y_0 = 0 \mid x), \\ \pi(y_1 = 0 \mid x; \theta, F_{u_0|xc}, F_{c|x}) &= F_{u_0|xc}(-x_1'\beta - c^*(x)) = P(y_1 = 0 \mid x). \end{aligned}$$

Therefore $\theta \in \Theta_0$, and since $\theta \in \tilde{\Theta}_0$ was chosen arbitrarily, $\tilde{\Theta}_0 \subseteq \Theta_0$.

D.2. Proof of Lemma 6

Define $w = x_1 - x_0$, $z = y_1 - y_0$, and $\Theta_1 = \arg\max_{\theta \in \Theta} Q(\theta)$.

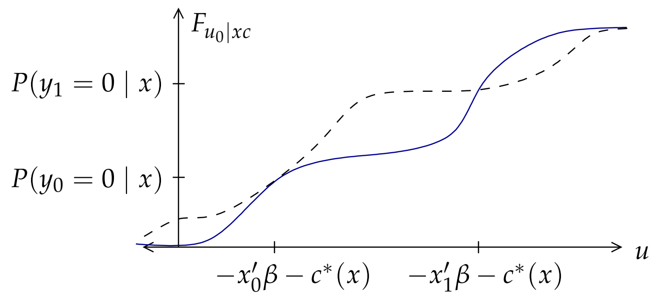

Figure 6. Two distributions $F_{u_0|xc}$ with equivalent observable implications under $F_{c|x}$.

Step 1 Let $\theta_1 \in \Theta_0$ and $\theta_2 \in \Theta$. We will show that $\Theta_0 \subseteq \Theta_1$ by proving that, for arbitrary choices of θ_1 and θ_2, $Q(\theta_1) \geq Q(\theta_2)$.

Consider the difference

$$\begin{aligned} Q(\theta_1) - Q(\theta_2) &= \mathrm{E}\left[z\,\mathrm{sgn}(w'\beta_1)\right] - \mathrm{E}\left[z\,\mathrm{sgn}(w'\beta_2)\right] \\ &= \mathrm{E}\left[z\left(\mathrm{sgn}(w'\beta_1) - \mathrm{sgn}(w'\beta_2)\right)\right] \\ &= 2\int_{D(\theta_1,\theta_2)} \mathrm{sgn}(w'\beta_1)\,\mathrm{E}\left[z \mid x, c\right]\, dF_{xc} \end{aligned}$$

where $D(\theta_1, \theta_2) = \{(x, c) : \mathrm{sgn}(w'\beta_1) \neq \mathrm{sgn}(w'\beta_2)\}$ is the set of values of x and c where $\mathrm{sgn}(w'\beta_1)$ and $\mathrm{sgn}(w'\beta_2)$ differ. The last equality above follows from the fact that the integrand vanishes on complement of $D(\theta_1, \theta_2)$, and that on $D(\theta_1, \theta_2)$ we have $\mathrm{sgn}(w'\beta_1) = -\,\mathrm{sgn}(w'\beta_2)$, implying that $\mathrm{sgn}(w'\beta_1) - \mathrm{sgn}(w'\beta_2) = 2\,\mathrm{sgn}(w'\beta_1)$. Since $\theta_1 \in \Theta_0$, Theorem 5 guarantees that

$$\mathrm{sgn}(w'\beta_1) = \mathrm{sgn}\left(P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c)\right) = \mathrm{sgn}\,\mathrm{E}(z \mid x, c)$$

F_{xc}-almost surely. Rewriting the above difference,

$$Q(\theta_1) - Q(\theta_2) = 2\int_{D(\theta_1,\theta_2)} |\mathrm{E}[z \mid x, c]|\, dF_{xc} \geq 0$$

for all θ_2. Therefore, $\Theta_0 \subseteq \Theta_1$.

Step 2 Now, let $\theta_1 \in \Theta_0$ and suppose there exists a $\theta_2 \in \Theta_0^c \cap \Theta_1$. We will use the definition of Θ_0 to show that $Q(\theta_2) < Q(\theta_1)$, contradicting the assumption that $\theta_2 \in \Theta_1$, and guaranteeing that $\Theta_0^c \cap \Theta_1 = \emptyset$, or equivalently, $\Theta_1 \subseteq \Theta_0$.

First, note that we can rewrite $Q(\theta)$ as follows:

$$\begin{aligned} Q(\theta) &= \mathrm{E}[z\,\mathrm{sgn}(w'\beta)] \\ &= \mathrm{E}_{xc}\,\mathrm{E}_{z|wc}[z\,\mathrm{sgn}(w'\beta)] \\ &= \mathrm{E}_{xc}\left[\left(P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c)\right)\left(1\{w'\beta \geq 0\} - 1\{w'\beta < 0\}\right)\right] \\ &= \int_{\{w'\beta \geq 0\}} \left(P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c)\right) dF_{xc} \\ &\quad + \int_{\{w'\beta < 0\}} \left(P(y_0 = 1 \mid x, c) - P(y_1 = 1 \mid x, c)\right) dF_{xc} \end{aligned}$$

The first equality is definitional, the second is an application of the law of iterated expectations, and the third follows from the definition of z and the signum function. In the fourth line, the expectations of the indicator functions are expressed as integrals over the corresponding regions of the support of x.

Now, consider the difference $Q(\theta_2) - Q(\theta_1)$:

$$\begin{aligned}
Q(\theta_2) - Q(\theta_1) = &\int_{\{w'\beta_2 \geq 0\}} (P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c))\, dF_{xc} \\
&+ \int_{\{w'\beta_2 < 0\}} (P(y_0 = 1 \mid x, c) - P(y_1 = 1 \mid x, c))\, dF_{xc} \\
&- \int_{\{w'\beta_1 \geq 0\}} (P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c))\, dF_{xc} \\
&- \int_{\{w'\beta_1 < 0\}} (P(y_0 = 1 \mid x, c) - P(y_1 = 1 \mid x, c))\, dF_{xc}
\end{aligned}$$

Over regions where $w'\beta_2$ and $w'\beta_1$ have the same sign, the difference is zero, therefore

$$\begin{aligned}
Q(\theta_2) - Q(\theta_1) = &\int_{\{w'\beta_2 \geq 0, w'\beta_1 < 0\}} (P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c))\, dF_{xc} \\
&- \int_{\{w'\beta_2 < 0, w'\beta_1 \geq 0\}} (P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c))\, dF_{xc}
\end{aligned}$$

From the proof of Theorem 5, we know that for $\theta_1 \in \Theta_0$,

$$P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c) \geq 0 \iff w'\beta_1 \geq 0$$

and for $\theta_2 \in \Theta_0^c$,

$$P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c) < 0 \iff w'\beta_2 \geq 0.$$

This implies that the first term in the difference above is strictly negative and the second term, which is being subtracted, is weakly non-negative. Thus, $Q(\theta_2) < Q(\theta_1)$. This contradicts the choice of θ_2, meaning that $\Theta_0^c \cap \Theta_1 = \emptyset$ and therefore it must be the case that $\Theta_1 \subseteq \Theta_0$.

D.3. *Proof of Theorem 6*

Assumptions C1 and C2 were verified in the text. It remains to show that $\mathcal{F}$ is manageable for an envelope F for which $PF^2 < \infty$.

Let $\mathcal{D} \subset \mathbb{R}^d$ denote the support of w and let $\mathcal{X} = \{-1, 0, 1\} \times \mathcal{D}$ denote the support of (z, w). For each $(z, w) \in \mathcal{X}$ and for each real number t, α, and γ, and real vector $\delta \in \mathbb{R}^d$, define

$$g(z, w, t, \alpha, \gamma, \delta) = \alpha t + \gamma z + \delta' w$$

and define

$$\mathcal{G} = \left\{ g(\cdot, \cdot, \cdot, \alpha, \gamma, \delta) : \alpha, \gamma \in \mathbb{R} \text{ and } \delta \in \mathbb{R}^d \right\}.$$

Since $\mathcal{G}$ is a vector space of real-valued functions on $\mathcal{X} \times \mathbb{R}$, by Lemma 2.4 of Pakes and Pollard (1989), classes of sets of the form $\{g \geq r\}$ or $\{g > r\}$ with $g \in \mathcal{G}$ and $r \in \mathbb{R}$ are VC classes. We will show that $\mathcal{F}$ is Euclidean by showing that it is a VC subgraph class, that is, that the collection of subgraphs of functions in $\mathcal{F}$ is a VC class. To accomplish this, we will use Lemma 2.5 of Pakes and Pollard (1989) which states that, in particular, if $\mathcal{C}_1$ and $\mathcal{C}_2$ are VC classes, then so are $\{C_1 \cap C_2 : C_1 \in \mathcal{C}_1, C_2 \in \mathcal{C}_2\}$, $\{C_1 \cup C_2 : C_1 \in \mathcal{C}_1, C_2 \in \mathcal{C}_2\}$, and $\{C_1^c : C_1 \in \mathcal{C}_1\}$.

First, note that we can rewrite f as

$$\begin{aligned}
f(z, w, \theta) &= (1\{z > 0\} - 1\{z < 0\}) \cdot \left(1\{w'\beta \geq 0\} - 1\{w'\beta < 0\}\right) \\
&= 1\{z > 0,\, w'\beta \geq 0\} - 1\{z > 0,\, w'\beta < 0\} \\
&\quad - 1\{z < 0,\, w'\beta \geq 0\} + 1\{z < 0,\, w'\beta < 0\}.
\end{aligned}$$

Now, for any $\theta \in \Theta$,

$$\begin{aligned}
\operatorname{subgraph}(f(\cdot,\cdot,\theta)) &= \{(z,w,t) \in \mathcal{X} \times \mathbb{R} : 0 < t < f(z,w,\theta) \text{ or } 0 > t > f(z,w,\theta)\} \\
&= \big(\{z > 0\} \cap \{w'\beta \geq 0\} \cap \{t \geq 1\}^c \cap \{t > 0\}\big) \\
&\quad \cup \big(\{z > 0\} \cap \{w'\beta \geq 0\}^c \cap \{t \geq -1\} \cap \{t \geq 0\}^c\big) \\
&\quad \cup \big(\{z \geq 0\}^c \cap \{w'\beta \geq 0\} \cap \{t \geq -1\} \cap \{t \geq 0\}^c\big) \\
&\quad \cup \big(\{z \geq 0\}^c \cap \{w'\beta \geq 0\}^c \cap \{t \geq 1\}^c \cap \{t > 0\}\big) \\
&= (\{g_1 > 0\} \cap \{g_2 \geq 0\} \cap \{g_3 \geq 1\}^c \cap \{g_3 > 0\}) \\
&\quad \cup (\{g_1 > 0\} \cap \{g_2 \geq 0\}^c \cap \{g_3 \geq -1\} \cap \{g_3 \geq 0\}^c) \\
&\quad \cup (\{g_1 \geq 0\}^c \cap \{g_2 \geq 0\} \cap \{g_3 \geq -1\} \cap \{g_3 \geq 0\}^c) \\
&\quad \cup (\{g_1 \geq 0\}^c \cap \{g_2 \geq 0\}^c \cap \{g_3 \geq 1\}^c \cap \{g_3 > 0\})
\end{aligned}$$

where $g_k(z,w,t) = \alpha_k t + \gamma_k z + \delta_k' w \in \mathcal{G}$ for each k with, $\alpha_1 = 0$, $\gamma_1 = 1$, $\delta_1 = 0$, $\alpha_2 = 0$, $\gamma_2 = 0$, $\delta_2 = \beta$, $\alpha_3 = 1$, $\gamma_3 = 0$, and $\delta_3 = 0$. The collection of sets of the form $\{g \geq 0\}$ or $\{g > 0\}$ is a VC class by Lemma 2.4 of Pakes and Pollard (1989). Furthermore, this property is preserved over complements, unions, and intersections of VC classes by their Lemma 2.5. Therefore, $\{\operatorname{subgraph}(f) : f \in \mathcal{F}\}$ is a VC class, and by Lemma 2.12 of Pakes and Pollard (1989), $\mathcal{F}$ is Euclidean for every envelope. In particular, $\mathcal{F}$ is Euclidean for the constant envelope $F = 1$. Since it is Euclidean, $\mathcal{F}$ is also manageable in the sense of Pollard (1989) (cf. Pakes and Pollard, 1989, p. 1033) as required by C3.

D.4. *Proof of Lemma 7*

First, note that we can rewrite $n^{1/2}Q_n$ as

$$n^{1/2}Q_n(\theta) = n^{1/2}(P_n f_\theta - P f_\theta) + n^{1/2} P f_\theta = \mathbb{G}_n(f_\theta) + n^{1/2} P f_\theta,$$

and therefore,

$$R_n \equiv \inf_{\theta \in \Theta_0} n^{1/2} Q_n(\theta) = \inf_{\theta \in \Theta_0} \left(\mathbb{G}_n(f_\theta) + n^{1/2} P f_\theta \right).$$

Supposing, without loss of generality, that Q is normalized so that it is identically zero on Θ_0, since the map $\inf_{\Theta_0}$, which takes real functions on Θ into $\mathbb{R}$, is continuous in $\ell^\infty(\mathcal{F})$, the continuous mapping theorem gives $R_n \xrightarrow{d} \inf_{\theta \in \Theta_0} \mathbb{G}(f_\theta) \equiv R$.

D.5. *Proof of Theorem 8*

Assumptions C1–C3 have been established by Theorem 6. We show that Assumptions C4 and C5 hold, and the conclusion follows from Lemma 4.

Under the maintained assumptions, it follows by generalizing the arguments of Abrevaya and Huang (2005) to the set identified case that $\nabla_{\theta\theta'} Q(\theta) = -V(\theta)$ for all $\theta \in \operatorname{bd}(\Theta_0)$. Therefore, in a neighborhood $\mathcal{N}$ of Θ_0, Q is approximately quadratic and for some $C > 0$, $Q(\theta) \leq \sup Q - C \cdot \rho^2(\theta, \Theta_0)$.

To show C5, let $\eta > 0$ and define $\mathcal{F}_\eta \equiv \{f(\cdot,\theta) \in \mathcal{F} : \rho(\theta, \Theta_0) \leq \eta\}$. We will show that $\mathcal{F}_\eta$ is a VC subgraph class with envelope F_η such that $PF_\eta^2 = O(\eta)$. For any $f \in \mathcal{F}_\eta$ and any $\bar{\beta} \in \Theta_0$,

$$
\begin{aligned}
F_\eta^2(w, y, \theta) &\leq 1\{\tilde{w}'\beta \geq -w_k > \tilde{w}'\bar{\beta}\} + 1\{\tilde{w}'\bar{\beta} \geq -w_k > \tilde{w}'\beta\} \\
&\leq 1\{\tilde{w}'(\beta - \bar{\beta}) \geq -w_k - \tilde{w}'\bar{\beta} > 0\} + 1\{0 \geq -w_k - \tilde{w}'\bar{\beta} > \tilde{w}'(\beta - \bar{\beta})\} \\
&\leq 1\{-\tilde{w}'(\beta - \bar{\beta}) \leq \tilde{w}'\bar{\beta} + w_k < 0\} + 1\{0 \leq \tilde{w}'\bar{\beta} + w_k < \tilde{w}'(\beta - \bar{\beta})\} \\
&\leq 1\{-\tilde{w}'(\beta - \bar{\beta}) \leq \tilde{w}'\bar{\beta} + w_k < \tilde{w}'(\beta - \bar{\beta})\} \\
&\leq 1\{-\|\tilde{w}\| \cdot \|\beta - \bar{\beta}\| \leq \tilde{w}'\bar{\beta} + w_k < \|\tilde{w}\| \cdot \|\beta - \bar{\beta}\|\}.
\end{aligned}
$$

In particular, the above inequality holds when $\bar{\beta}$ is the nearest neighbor to β in Θ_0, with $\|\beta - \bar{\beta}\| = \rho(\beta, \Theta_0) \leq \eta$. Therefore $F_\eta^2(x, y, \theta) \leq 1\{-|\eta| \|\tilde{w}\| \leq \tilde{w}'\bar{\beta} + w_k < \eta \|\tilde{w}\|\}$ and

$$
\begin{aligned}
PF_\eta^2 &\leq \int_{\tilde{\mathcal{W}}} \int_{\mathcal{W}_1 \cup \{w_k : -\eta\|\tilde{w}\| \leq \tilde{w}'\bar{\beta} + w_k < \eta\|\tilde{w}\|\}} dG(w_k \mid \tilde{w})\, dG(\tilde{w}) \\
&\leq \int_{\tilde{\mathcal{W}}} \sup g(w_k \mid \tilde{w}) \cdot 2\eta \|\tilde{w}\| \, dG(\tilde{w}).
\end{aligned}
$$

Since w has finite first absolute moments and bounded density, this integral is finite and the envelope satisfies $PF_\eta^2 = O(\eta)$.

E. Lagged Dependent Variable Model

E.1. Proof of Theorem 9

Letting G denote $F_{u_0|xc}$ for simplicity, the probabilities of the events A and B can be written as follows:

$$
\begin{aligned}
P(A \mid x, c, x_2 = x_3) &= p_0(x, c)^{1-d_0}(1 - p_0(x, c))^{d_0} G(-x_1'\beta - \gamma d_0 - c) \\
&\quad \times \left[1 - G(-x_2'\beta - c)\right] G(-x_2'\beta - \gamma - c)^{1-d_3} \\
&\quad \times \left[1 - G(-x_2'\beta - \gamma - c)\right]^{d_3}, \\
P(B \mid x, c, x_2 = x_3) &= p_0(x, c)^{1-d_0}(1 - p_0(x, c))^{d_0} \left[1 - G(-x_1'\beta - \gamma d_0 - c)\right] \\
&\quad \times G(-x_2'\beta - \gamma - c) G(-x_2'\beta - c)^{1-d_3} \\
&\quad \times \left[1 - G(-x_2'\beta - c)\right]^{d_3}.
\end{aligned}
$$

Note that the latter probability is nonzero since u_t has full support on $\mathbb{R}$ for all t and since $p_0(x, c) > 0$. Dividing, we have

$$
\begin{aligned}
\frac{P(A \mid x, c, x_2 = x_3)}{P(B \mid x, c, x_2 = x_3)} &= \frac{G(-x_1'\beta - \gamma d_0 - c)}{G(-x_2'\beta - \gamma - c)} \times \frac{1 - G(-x_2'\beta - c)}{1 - G(-x_1'\beta - \gamma d_0 - c)} \\
&\quad \times \left[\frac{G(-x_2'\beta - \gamma - c)}{G(-x_2'\beta - c)}\right]^{1-d_3} \times \left[\frac{1 - G(-x_2'\beta - \gamma - c)}{1 - G(-x_2'\beta - c)}\right]^{d_3}.
\end{aligned}
$$

When $d_3 = 0$,

$$
\frac{P(A \mid x, c, x_2 = x_3)}{P(B \mid x, c, x_2 = x_3)} = \frac{G(-x_1'\beta - \gamma d_0 - c)}{G(-x_2'\beta - \gamma d_3 - c)} \times \frac{1 - G(-x_2'\beta - \gamma d_3 - c)}{1 - G(-x_1'\beta - \gamma d_0 - c)},
$$

and when $d_3 = 1$,

$$
\frac{P(A \mid x, c, x_2 = x_3)}{P(B \mid x, c, x_2 = x_3)} = \frac{G(-x_1'\beta - \gamma d_0 - c)}{G(-x_2'\beta - \gamma d_3 - c)} \times \frac{1 - G(-x_2'\beta - \gamma d_3 - c)}{1 - G(-x_1'\beta - \gamma d_0 - c)}.
$$

We have used the fact that when $d_3 = 0$, $\gamma d_3 = 0$, and when $d_3 = 1$, $\gamma d_3 = \gamma$. In both cases, by the monotonicity of G,

$$P(A \mid x, c, x_2 = x_3) \geq P(B \mid x, c, x_2 = x_3) \iff -x_1'\beta - \gamma d_0 - c \geq -x_2'\beta - \gamma d_3 - c.$$

The result follows since this event is independent of c.

E.2. *Proof of Lemma 8*

This proof parallels the proof of Lemma 6, the corresponding result for Model 1. Let $\Theta_1 \equiv \arg\max_\Theta Q$ and define $w_t \equiv x_t - x_{t-1}$, $z \equiv y_2 - y_1$, and $v \equiv y_3 - y_0$.

Step 1 Let $\theta_1 \in \Theta_0$ and $\theta_2 \in \Theta$. We will show that $Q(\theta_1) \geq Q(\theta_2)$ and therefore, $\theta_1 \in \Theta_1$. We have

$$\begin{aligned}
Q(\theta_1) - Q(\theta_2) &= \mathrm{E}\left[1\{w_3 = 0\} \cdot z \cdot \left(\operatorname{sgn}(w_2'\beta_1 + \gamma_1 v) - \operatorname{sgn}(w_2'\beta_2 + \gamma_2 v)\right)\right] \\
&= \int \mathrm{E}[z \mid x, c, y_0, y_3, w_3 = 0] \left(\operatorname{sgn}(w_2'\beta_1 + \gamma_1 v) - \operatorname{sgn}(w_2'\beta_2 + \gamma_2 v)\right) dF_{x,c,y_0,y_3|w_3=0} \\
&= 2 \int_{D(\theta_1,\theta_2)} \operatorname{sgn}(w_2'\beta_1 + \gamma_1 v)\, \mathrm{E}[z \mid x, c, y_0, y_3, w_3 = 0]\, dF_{x,c,y_0,y_3|w_3=0}
\end{aligned}$$

where $D(\theta_1, \theta_2)$ is defined as the set of all (x, c, v) where $\operatorname{sgn}(w_2'\beta_1 + \gamma_1 v)$ and $\operatorname{sgn}(w_2'\beta_2 + \gamma_2 v)$ differ. The first equality follows by definition of Q, the second is an application of the law of iterated expectations, and the third is due to the fact that on $D(\theta_1, \theta_2)$, $\operatorname{sgn}(w_2'\beta_2 + \gamma_2 v) = -\operatorname{sgn}(w_2'\beta_1 + \gamma_1 v)$. Note that on the integrand above vanishes on the complement of $D(\theta_1, \theta_2)$.

Now, since $\theta_1 \in \Theta_0$, from Theorem 9 we have that for all d_0, d_3,

$$\begin{aligned}
\operatorname{sgn}(w_2'\beta_1 + \gamma_1 v) &= \operatorname{sgn}\left(P(A \mid x, x_2 = x_3) - P(B \mid x, x_2 = x_3)\right) \\
&= \operatorname{sgn}\big(P(y_1 = 0, y_2 = 1 \mid x, x_2 = x_3, y_0 = d_0, y_3 = d_3) \\
&\qquad - P(y_1 = 1, y_2 = 0 \mid x, x_2 = x_3, y_0 = d_0, y_3 = d_3)\big)
\end{aligned}$$

for all $d_0, d_3 \in \{0, 1\}$. The second line follows because the common factor which was removed, $P(y_0 = d_0, y_3 = d_3 \mid x, x_2 = x_3)$, is always positive. Furthermore, we can write

$$\begin{aligned}
\mathrm{E}[z \mid x, c, y_0, y_3, w_3 = 0] = P(y_1 = 0, y_2 = 1 \mid x, c, y_0, y_3, w_3 = 0)& \\
- P(y_1 = 1, y_2 = 0 \mid x, c, y_0, y_3, w_3 = 0)&.
\end{aligned}$$

So, the sign above times the conditional expectation of z simplifies to the absolute value of the conditional expectation. Returning to the objective function,

$$Q(\theta_1) - Q(\theta_2) = 2 \int_{D(\theta_1,\theta_2)} |E[z \mid x, c, y_0, y_3, w_3 = 0]|\, dF_{x,c,y_0,y_3|w_3=0} \geq 0.$$

Step 2 Let $\theta_1 \in \Theta_0$ and suppose there exists a $\theta_2 \in \Theta_0^c \cap \Theta_1$. We will show that this implies $Q(\theta_2) < Q(\theta_1)$, which is a contradiction of the choice of $\theta_2 \in \Theta_1$, and therefore $\Theta_0^c \cap \Theta_1$ must be empty.

Note that we can express Q as

$$\begin{aligned}
Q(\theta) = &\int_{\{w_3'\beta + \gamma v \geq 0\}} \big[P(y_1 = 0, y_2 = 1 \mid x, c, y_0, y_3, w_3 = 0) \\
&\qquad - P(y_1 = 1, y_2 = 0 \mid x, c, y_0, y_3, w_3 = 0)\big]\, dF_{x,c,y_0,y_3|w_3=0} \\
&+ \int_{\{w_3'\beta + \gamma v < 0\}} \big[P(y_1 = 1, y_2 = 0 \mid x, c, y_0, y_3, w_3 = 0) \\
&\qquad - P(y_1 = 0, y_2 = 1 \mid x, c, y_0, y_3, w_3 = 0)\big]\, dF_{x,c,y_0,y_3|w_3=0}.
\end{aligned}$$

Again we consider a difference $Q(\theta_2) - Q(\theta_1)$. Using the linearity of integrals, we can partition the range of each integral into disjoint sets and subtract the corresponding integrands on each set. When $w_3'\beta_1 + \gamma_1 v$ and $w_3'\beta_2 + \gamma_2 v$ have the same sign, the difference is zero, so we only need to consider regions where the sign differs:

$$\begin{aligned} D_1 &\equiv \{(x, c, y_0, y_3) : w_3'\beta_2 + \gamma_2 v \geq 0, w_3'\beta_1 + \gamma_1 v < 0\}, \\ D_2 &\equiv \{(x, c, y_0, y_3) : w_3'\beta_2 + \gamma_2 v < 0, w_3'\beta_1 + \gamma_1 v \geq 0\}. \end{aligned}$$

Hence,

$$\begin{aligned} Q(\theta_2) - Q(\theta_1) = \int_{D_1} \big[& P(y_1 = 0, y_2 = 1 \mid x, c, y_0, y_3, w_3 = 0) \\ & - P(y_1 = 1, y_2 = 0 \mid x, c, y_0, y_3, w_3 = 0) \big] \, dF_{x,c,y_0,y_3|w_3=0} \\ + \int_{D_2} \big[& P(y_1 = 1, y_2 = 0 \mid x, c, y_0, y_3, w_3 = 0) \\ & - P(y_1 = 0, y_2 = 1 \mid x, c, y_0, y_3, w_3 = 0) \big] \, dF_{x,c,y_0,y_3|w_3=0}. \end{aligned}$$

Since $\theta_1 \in \Theta_0$ and $\theta_2 \notin \Theta_0$, first term is strictly negative and the second term is weakly non-positive.

E.3. Proof of Theorem 10

By Lemma 8, $\Theta_1 = \tilde{\Theta}_0$. Therefore, we verify Assumptions C1–C3 in order to establish consistency of $\hat{\Theta}_n$ for $\tilde{\Theta}_0$ using Lemma 2.

Assumption C1 is satisfied by definition of Model 2 and by inspection of $f(\cdot, \cdot, \theta)$. Assumption C2 is satisfied under B1 because Q is a step function which is piecewise continuous. To verify Assumption C3, since $|f(\cdot, \cdot, \theta)| \leq 1$ for all $f \in \mathcal{F}$, $F = 1$ is a valid envelope and in light of Lemma 3, it suffices to show that $\mathcal{F}$ is a VC subgraph class.

We follow the same strategy as in the proof of Theorem 6. Define $w_t \equiv x_t - x_{t-1}$, $z \equiv y_2 - y_1$, and $v \equiv y_3 - y_0$, and let $f(w_2, w_3, z, v, \theta) = 1\{w_3 = 0\} \cdot z_2 \cdot \left[2 \cdot 1\{w_2'\beta + \gamma v \geq 0\} - 1\right]$. First, note that f can be rewritten as

$$\begin{aligned} f(w_2, w_3, z, v, \theta) = 1\{w_3 \geq 0\} \cdot 1\{w_3 \leq 0\} \cdot (1\{z_2 > 0\} - 1\{z < 0\}) \\ \cdot \left(1\{w_2'\beta + \gamma v \geq 0\} - 1\{w_2'\beta + \gamma v < 0\}\right). \end{aligned}$$

Upon expanding this expression, it is clear that, as before, for any θ we can express subgraph $f(\cdot, \cdot, \cdot, \cdot, \theta)$ as series of intersections and unions (and complements thereof) of the form $\{g \geq 0\}$ and $\{g > 0\}$ for specific coefficient values α of some polynomial

$$g(w_2, w_3, z, v, t, \alpha) = \alpha_1 t + \alpha_2 w_2 + \alpha_3 w_3 + \alpha_4 z + \alpha_5 v.$$

Therefore, $\{\text{subgraph}(f) : f \in \mathcal{F}\}$ is a VC class.

References

Abrevaya, J. (2000). Rank estimation of a generalized fixed-effects regression model. *Journal of Econometrics 95*, 1–23. [2, 15, 23]

Abrevaya, J. and J. Huang (2005). On the bootstrap of the maximum score estimator. *Econometrica 73*, 1175–1204. [19, 49]

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press. [15]

Andrews, D. W. K., S. Berry, and P. Jia (2004). Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store location. Unpublished manuscript, Yale University. [3]

Andrews, D. W. K. and X. Shi (2009). Inference based on conditional moment inequalities. Unpublished manuscript, Yale University. [3]

Aradillas-Lopez, A. and E. Tamer (2008). The identification power of equilibrium in simple games. *Journal of Business and Economic Statistics 26*, 261–283. [3]

Beresteanu, A., I. Molchanov, and F. Molinari (2009). Sharp identification regions in models with convex predictions: games, individual choice, and incomplete data. Working Paper CWP27/09, CEMMAP. [3]

Beresteanu, A. and F. Molinari (2008). Asymptotic properties for a class of partially identified models. *Econometrica 76*, 763–814. [3]

Bierens, H. J. and J. Hartog (1988). Non-linear regression with discrete explanatory variables, with an application to the earnings function. *Journal of Econometrics 38*, 269–299. [4]

Bugni, F. A. (2008). Bootstrap inference in partially identified models. Unpublished manuscript, Northwestern University. [3, 6]

Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In J. J. Heckman and B. S. Singer (Eds.), *Longitudinal Analysis of Labor Market Data*. New York: Cambridge University Press. [20]

Charlier, E., B. Melenberg, and A. H. O. van Soest (1995). A smoothed maximum score estimator for the binary choice panel data model with an application to labour force participation. *Statistica Neerlandica 49*(3), 324–342. [16]

Chen, S. (2002). Rank estimation of transformation models. *Econometrica 70*, 1683–1697. [2, 23, 24]

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2009). Identification and estimation of marginal effects in nonlinear panel models. Unpublished manuscript, Massachusetts Institute of Technology. [4]

Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica 75*, 1243–1284. [2, 3, 6, 8, 10]

Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in the airline industry. *Econometrica 77*, 1791–1828. [3]

Dudley, R. M. (1987). Universal donsker classes and metric entropy. *Annals of Probability 15*, 1306–1326. [13]

Han, A. K. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics 35*, 303–316. [2, 15, 16, 23]

Honoré, B. E. and E. Kyriazidou (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica 68*, 839–874. [3, 20, 21]

Honoré, B. E. and A. Lleras-Muney (2006). Bounds in competing risks models and the war on cancer. *Econometrica 74*, 1675–1698. [4, 23]

Honoré, B. E. and E. Tamer (2006). Bounds on parameters in panel dynamic discrete choice models. *Econometrica 74*, 611–629. [4]

Horowitz, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica 60*, 505–531. [2, 15, 16]

Horowitz, J. L. (1998). *Semiparametric Methods in Econometrics*. New York: Springer. [4, 5, 15]

Horowitz, J. L. and C. F. Manski (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association 95*, 77–88. [3]

Imbens, G. W. and C. F. Manski (2004). Confidence intervals for partially identified parameters. *Econometrica 72*, 1845–1857. [3]

Khan, S. and E. Tamer (2009). Inference on endogenously censored regression models using conditional moment inequalities. *Journal of Econometrics 152*, 104–119. [3]

Kim, J. and D. Pollard (1990). Cube root asymptotics. *The Annals of Statistics 18*, 191–219. [2, 13, 14, 19, 44]

Kim, K. (2009). Set estimation and inference with models characterized by conditional moment inequalities. Unpublished manuscript, University of Minnesota. [3, 6]

Komarova, T. (2008). Binary choice models with discrete regressors: Identification and misspecification. Unpublished manuscript, London School of Economics. [4]

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press. [15]

Magnac, T. and E. Maurin (2008). Partial identification in monotone binary models: Discrete regressors and interval data. *Review of Economic Studies 75*, 835–864. [4]

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*, 205–228. [2, 15, 17]

Manski, C. F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics 27*, 313–333. [2, 15, 16, 17]

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica 55*, 357–362. [3, 5, 6, 16, 17]

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer-Verlag. [3]

Manski, C. F. and E. Tamer (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica 70*, 519–546. [2, 4, 6]

McFadden, D. L. (1974). Conditional logit analysis of qualitative choice analysis. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. Academic Press. [15]

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Amsterdam. New Holland. [13, 14]

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*, 1027–1057. [12, 14, 48, 49]

Pakes, A., J. Porter, K. Ho, and J. Ishii (2006). Moment inequalities and their application. Unpublished manuscript, Harvard University. [3]

Pollard, D. (1989). Asymptotics via empirical processes. *Statistical Science 4*, 341–366. [13, 49]

Ridder, G. (1990). The non-parametric identification of generalized accelerated failure-time models. *Review of Economic Studies 57*, 167–182. [22, 23]

Romano, J. P. and A. M. Shaikh (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference 138*, 2786–2807. [3, 6]

Romano, J. P. and A. M. Shaikh (2010). Inference for the identified set in partially identified econometric models. *Econometrica 78*, 169–211. [3, 6, 8]

Rosen, A. (2009). Set identification via quantile restrictions in short panels. Working Paper CWP26/09, CEMMAP. [4]

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley. [44]

Shorack, G. R. (2000). *Probability for Statisticians*. Springer. [44]

Stoye, J. (2009). More on confidence intervals for partially identified parameters. *Econometrica 77*, 1299–1315. [3]

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies 70*, 147–165. [3]

Tamer, E. (2009). Partial identification in econometrics. Unpublished manuscript, Northwestern University. [3]