Pairwise-difference estimation of incomplete information games

Andres Aradillas-Lopez [1]

Department of Economics, University of Wisconsin, Madison WI 53706, United States

A R T I C L E   I N F O

A B S T R A C T

This paper contributes to the literature on econometric estimation of incomplete information games with Nash equilibrium behavior by introducing a two-step estimation procedure that makes no parametric assumptions about the distribution of unobservable payoffs shocks. Instead, its asymptotic properties rely on assuming only that these distributions satisfy an invertibility condition, and that the underlying equilibrium selection mechanism is degenerate. Our methodology relies on a pairwise-differencing procedure which, unlike Aradillas-Lopez (2010), does not require computing the equilibria of the game. Furthermore, if normal-form payoffs are linear in the parameters of interest, our procedure results in an estimator with a closed-form expression. We contribute to the pairwise-differencing econometric literature by introducing the first model, where both the control variables being matched *and* the regressors in the index function parameterized by θ contain nonparametric functions. In particular, the asymptotic theory developed in Aradillas-Lopez et al. (2007) does not cover this setting. We describe conditions under which nonparametrically estimated plug-ins yield a $\sqrt{N}$-consistent and asymptotically normal estimator for the parameter of interest. A consistent specification test based on semiparametric residuals is also developed. It appears to be the first test of this type for a model involving nonparametric or "generated" regressors. Several extensions of our method are also discussed. A series of Monte Carlo experiments are used to investigate the properties of our estimator and our specification test.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

This paper studies the estimation of static games with incomplete information under the assumption of Bayesian–Nash equilibrium (BNE) behavior with additively separable payoff shocks and beliefs conditioned on observable covariates. These assumptions characterize existing related work, including Seim (2006), Pesendorfer and Schmidt-Dengler (2008), Sweeting (2009) and Bajari et al. (2010). Unlike those papers and most of the existing work on the subject, our method will leave the distribution of unobservable payoff shocks unknown except for the assumption that it is strictly increasing everywhere. Such an inferential setting was examined in Section 4 of Aradillas-Lopez (2010). However, the method prescribed requires the computation of equilibria at each step of the procedure. Here we describe a two-step procedure which does not require the computation of equilibria. Its validity relies on the assumption that the underlying equilibrium selection mechanism is degenerate and hence the data (choices) in each observation is generated from a single equilibrium. This assumption is common in existing two-step estimation procedures

in game-theoretic models (e.g., see Pesendorfer and Schmidt-Dengler (2008), Bajari et al. (2010)), where conditional choice probabilities are estimated in a first-step and plugged into a second stage to recover the payoff parameters of interest. The assumption that the data in each observation (e.g., market) is generated by a single equilibrium can also be found, either explicitly or implicitly, in Moro (2003), Seim (2006) and Jia (2008). Examples of dynamic models that also rely on this assumption include Aguirregabiria and Mira (2002) and Aguirregabiria (2007). By the nature of the repeated interaction through time of the same set of agents, the assumption that the same equilibrium is being selected has a more solid justification in dynamic vis-à-vis static games. In other instances (see, e.g., Pakes et al. (2007)), modeling assumptions can ensure uniqueness of equilibrium. This paper contributes to the literature by introducing a two-step procedure that does not rely on assuming a known functional form for the distribution of unobservable payoff shocks.

The main idea behind our method is the following. If unobservable shocks are additively separable and their distribution is strictly increasing, a degenerate equilibrium selection mechanism implies a one-to-one mapping between conditional choice probabilities and the observable portion of players' expected utility. After suitable location-and-scale normalizations, this will allow us to identify the payoff (expected utility) parameters of interest provided that prototypical full-rank and exclusion restriction (i.e., that each player has an exclusive payoff shifter) hold for the

---

E-mail address: aaradill@ssc.wisc.edu.

[1] Part of this research was conducted while the author was visiting El Colegio de México (Centro de Estudios Económicos) and the Centro de Investigación y Docencia Económicas (División de Economía).

payoff covariates included. If we could match observations (markets) based on their conditional choice probabilities, their corresponding (observable) expected utilities should also coincide. Our method achieves this matching asymptotically through the use of kernel weights and a bandwidth sequence converging to zero. This enables us to deal with the case where conditional choice probabilities are continuous. Informally speaking, the resulting method boils down to minimizing a kernel-weighted "sum of squared residuals".

More precisely, our estimation technique relies on a pairwise-differencing procedure which, given our parameterization of normal-form payoffs, yields an estimator with a closed-form expression that is trivial to compute. Pairwise differencing estimation has been prominently studied, e.g. in Aradillas-Lopez et al. (2007) and Honoré and Powell (2005, 1994) (henceforth AHP). As in AHP, the control variables being matched in our model are nonparametric functions (conditional choice probabilities in our case). However, the model we study here is *not* a special case of AHP because the index parameterized by the structural parameter of interest θ (in our case, this index corresponds to the observable portion of players' expected utility) *also* contains nonparametric functions (players' beliefs in our case). As a result, we have a model with a nonparametric control function *and* nonparametric, or "generated regressors". This setting is not contemplated in the asymptotic theory results in AHP, and the results presented there cannot be used for inference here. Moreover, the conditions and assumptions in AHP are not designed to provide guidance regarding the estimation of the control functions and the generated regressors to ensure $\sqrt{N}$-consistency and asymptotic normality of the resulting estimator $\widehat{\theta}$. Our paper contributes to the pairwise-differencing literature by examining a model where both the control functions and the parameterized index function include unknown functions that must be nonparametrically estimated. In a more general econometric context, the estimation procedure we study is related to semiparametric models where an unknown function is treated as an (infinite-dimensional) nuisance parameter in the quest for estimating a finite-dimensional parameter of interest θ. Identification strategies in such settings have included exclusion restrictions, invertibility, as well as moment or quantile restrictions. A few examples related to our paper in various degrees include Manski (1985, 1975), Han (1987), Powell et al. (1989), Horowitz (1992), Klein and Spady (1993), Ahn and Manski (1993), Ahn and Powell (1993), Ichimura and Lee (1991), Ichimura (1993), Ahn et al. (1997) and Blundell and Powell (2004). Our estimator (as well as most of the ones just cited) relies on kernel weights. Recently, Sieve-based procedures have received increasing attention and they could be suitable to analyze a model like ours. A recent survey of that literature can be found in Chen (2007).

The assumption that the data in each observation is generated from a single equilibrium is a driving identification condition in this paper as well as in the ones cited above. In other instances, point-identification has been achieved by modeling the selection mechanism explicitly (Sweeting, 2009; Bajari et al., 2005). Alternatively, point-identification has been achieved without modeling the selection mechanism explicitly, but ruling out that certain types of equilibria can be selected. For instance, Bresnahan and Reiss (1990), Bresnahan and Reiss (1991a), Tamer (2003) and Davis (2006) establish identification results in complete information games when only pure-strategy Nash equilibria is played. A few examples of papers that deal with the econometrics of partially identified game-theoretic models under the assumption of equilibrium behavior include Ciliberto and Tamer (2009), Pakes et al. (2006), Andrews et al. (2004), Galichon and Henry (2011) and Beresteanu et al. (2008). A recent survey of the econometrics of discrete, static models with equilibrium

behavior can be found in Berry et al. (2006). A thorough analysis of moment-inequality econometric methods arising from multiple equilibria in empirical work is presented in Pakes (2008). Recently, identification results with non-equilibrium behavior in static games has been studied in Aradillas-Lopez and Tamer (2008).

The class of static models we study here can be applied to a variety of empirical problems. One of the most common applications involves entry decisions. Particular examples – some of which assume complete information – include Bresnahan and Reiss (1991a,b), Berry (1992), Cohen and Manuszak (2006), Davis (2006), Seim (2006) and Ciliberto and Tamer (2009). A recent survey of entry/exit applications of game-theoretic models can be found in Berry and Reiss (2007). Other applications include labor participation decisions (Bjorn and Vuong, 1984), social interactions models (Brock and Durlauf, 2001a,b), empirical model of competition in the supermarket industry (Davis, 2006) and recommendation decisions by stock analysts (Bajari et al., 2010) and the coordination in the timing of radio commercials (Sweeting, 2009). Other recent applications of game-theoretic models related to ours in differing degrees include Augereau et al. (2006), Ishii (2008) and Ho (2009). More recently, de Paula and Tang (forthcoming) have shown conditions under which the sign of strategic interaction can be nonparametrically identified in binary games. In experimental settings, the models we study here are particularly well suited to analyze quantal-response equilibrium (QRE) behavior as described in McKelvey and Palfrey (1995). Empirical applications have also received increasing attention in other fields. Some examples in the fields of political science and international relations include Signorino (1999, 2002) and Signorino and Tarar (2006). The methods we develop here for static models of strategic interaction can be adapted to many of the applications examined in the papers cited above. We hope that its computational simplicity and robustness features will make it an attractive tool for practitioners.

The paper proceeds as follows. Section 2 describes the structure of the underlying game along with the behavioral assumptions and the identification conditions that result from them. Section 4 describes a pairwise-differencing estimation procedure based on these identification conditions and it describes the asymptotic features of the resulting estimator. Section 4 describes a residual-based consistent specification test for our model. Section 5 discusses several extensions of our methodology and results. Section 6 examines the properties of our estimator and our specification test for a series of Monte Carlo experiments. Section 7 concludes. Unless noted otherwise, all proofs can be found in the Mathematical Appendix.

2. The model

We will illustrate our methods for estimation and specification testing by focusing on a 2×2 game. We do this for two reasons. First, because our approach extends naturally to games with multiple actions or players. We describe such extension in Section 5.3. Various other extensions will be described throughout Section 5. The second reason why we use the 2×2 game to introduce our methodology is because this has been one of the prototypical examples in the econometric literature on static discrete games to model entry in monopoly markets (see, e.g., Bresnahan and Reiss, 1990 or Tamer, 2003). We will use $p = 1, 2$ to denote a particular player, and $-p$ will denote his opponent. Each p chooses a binary action $Y_p \in \{0, 1\}$. The payoffs are given by

$$U_p(Y_p) = Y_p \times \left\{ X_p' \beta_p + \Delta_p Y_{-p} - \zeta_p \right\}.$$

$X_p \in \mathbb{R}^{L_p}$ and $\zeta_p \in \mathbb{R}$ denote observable and unobservable payoff covariates to the researcher. β_p and Δ_p are the payoff parameters of interest. This payoff parameterization was used

in Bresnahan and Reiss (1991a) and Tamer (2003) to model entry/exit decisions. The payoffs described above have three distinctive features: (i) Additive separability in observables and unobservables (to the researcher). (ii) A strategic-interaction effect that is not covariate-dependent (captured entirely in our case by the constant parameter Δ_p for player p). (iii) Observable covariates enter payoffs through a linear index. Other instances of applied-oriented work that assumed payoffs with these three features include Berry (1992) and Ciliberto and Tamer (2009) who analyzed entry decisions in airline markets, Seim (2006) who studied location decisions by video stores, Sweeting (2009) who analyzed commercial break timing decisions by radio stations, and both Bjorn and Vuong (1984) and Kooreman (1994) who studied labor participation decisions. The stock recommendation decision model in Section 5 of Bajari et al. (2010) assumes payoffs that satisfy all three of the above features.[2] This is also true for the empirical model of competition in the supermarket industry in Davis (2006) (although more general identification results are also presented there). In the experimental economics literature, the quantal response equilibrium (QRE) behavioral model (see McKelvey and Palfrey, 1995) is used to explain why players fail to choose best-responses with certainty. QRE replaces the deterministic predictions of Nash equilibrium behavior are replaced with a statistical version where players' utility for each action are augmented by a privately observed random error. QRE corresponds to the Bayesian–Nash equilibrium of the augmented or "perturbed" incomplete information game. As in our payoff characterization, this unobserved error is assumed to enter additively. In applications outside of economics, the extensive-form game estimated in Signorino (1999) to study international conflict assumes payoffs (for each terminal node of the extensive-form) that are additively separable in observables and unobservables. Notice that the payoffs of choosing $Y_p = 0$ are normalized to zero. This is done because expected utility maximization (a maintained assumption here) will imply that the only differences in payoffs are relevant. Payoff normalizations of this type are present for identification purposes in every econometric model of discrete games.[3] All the aforementioned articles assume a known parametric distribution (possibly up to a finite-dimensional parameter) for unobservable payoff shocks. In contrast, we will study the case in which these distributions are completely unknown except for an invertibility restriction. Of the three general payoff features mentioned above, only additive separability is indispensable for the potential applicability of a pairwise-differencing approach to estimation. The linear index restriction on observables is perhaps the easiest assumption to drop. Identification concerns aside, the only immediate cost would be that the estimator that results from our approach would no longer have a simple, closed-form expression. Subject to the identification implications for the specific parameterization employed, our approach has the potential of being extended to models where strategic interaction is a function of observable payoff covariates. We will describe these issues in detail in Section 5.1, where we discuss the extension of our methods to games with more general payoff functional forms. Sections 5.2–5.3 will describe other relevant extensions aimed at making our methodology applicable to more complicated and flexible games including many of the empirical applications cited above.

2.1. Basic assumptions

This paper concentrates on an incomplete information environment where actions are taken simultaneously. The source of incomplete information will be assumed to be the realization of the payoff shock ζ_p, which will be assumed to be only privately observed by player p. All other components of the game are assumed to be known to both players, including the realization of $X \equiv X_1 \cup X_2$, as well as the true values of the payoff parameters and the true distributions of all covariates involved.[4] Both players choose simultaneously (i.e., before observing the choice of their opponent) the action that maximizes their expected utility. Players' beliefs are unobserved by the researcher, but they are nonparametrically estimated in a first stage based on the following behavioral assumption.[5]

Assumption A0. (i) Players' behavior corresponds to a Bayesian–Nash equilibrium (BNE). In this context, we refer to an *equilibrium* as a pair of beliefs for both players (probability distributions over opponents' actions) that satisfy the self-consistency requirements of a BNE. The game is equipped with a selection mechanism $\mathcal{S}$ that chooses among the existing equilibria if more than one exists. We will maintain the assumption that $\mathcal{S}$ is *degenerate*, meaning that it selects only one of the existing equilibria with probability one. This rules out a selection mechanism $\mathcal{S}$ that can select two or more equilibria with positive probability.

(ii) The privately observed payoff shocks ζ_1, ζ_2 are independent of each other, independent of $X \equiv X_1 \cup X_2$ and of the selection mechanism $\mathcal{S}$. We denote the distribution of ζ_p by F_p, which is unknown to the researcher but assumed to be strictly increasing everywhere in $\mathbb{R}$.

Rational players should condition their beliefs on everything that is informative for their opponent's payoffs and behavior. From Assumption A0, the realization of ζ_p conveys no information for ζ_{-p}. It follows that both players should condition their beliefs on the realization of the publicly observed payoff covariates $X \equiv X_1 \cup X_2$ and their knowledge of the true payoff parameter values and the distributions involved (see Footnote 3). Let x be a given realization of X and let $\{\widetilde{\mu}^r(x)\}_{r=1}^{R(x)} \equiv \left\{\left(\widetilde{\mu}_1^r(x), \widetilde{\mu}_2^r(x)\right)\right\}_{r=1}^{R(x)}$ denote all the pairs (μ_1, μ_2) that solve the system

$$\mu_1 = F_1(x_1'\beta_1 + \Delta_1\mu_2), \qquad \mu_2 = F_2(x_2'\beta_2 + \Delta_2\mu_1),$$

where the payoff parameters are evaluated at their true values. From Assumption A0, every element in $\{\widetilde{\mu}^r(x)\}_{r=1}^{R(x)}$ is an equilibrium. By the continuity properties of F_p and the fact that it is bounded in $[0, 1]$ it is easy to show (e.g., using Brouwer's Fixed Point Theorem) that at least one equilibrium must exist for any realization x. By the stochastic properties of ζ_1, ζ_2 (namely, their independence of X and $\mathcal{S}$), iterated expectations and BNE behavior yield

$$\Pr[Y_1 = 1|X = x] = \sum_{r=1}^{R(x)} \Pr\left[\mathcal{S} = \widetilde{\mu}^r(x)\big| X = x\right]$$
$$\times F_1\left(x_1'\beta_1 + \Delta_1\widetilde{\mu}_2^r(x)\right),$$
$$\Pr[Y_2 = 1|X = x] = \sum_{r=1}^{R(x)} \Pr\left[\mathcal{S} = \widetilde{\mu}^r(x)\big| X = x\right]$$
$$\times F_2\left(x_2'\beta_2 + \Delta_2\widetilde{\mu}_1^r(x)\right).$$

---

[2] Bajari et al. (2010) also present identification results for payoff functions that are additively separable in unobservable, but where observable covariates do not necessarily enter through a linear index.

[3] Econometric models of experimental games stand apart here because the researcher has full control over the observable components of payoffs.

[4] The extension of our results to asymmetric information environments will be discussed in Section 5.3.

[5] See Manski (2004) for a discussion on measuring subjective expectations.

Degeneracy of $\mathscr{S}$ implies that only one equilibrium is selected with positive probability. Let us denote that equilibrium simply as $\mu(x)$. It follows that $\Pr[\mathscr{S} = \mu(x)| X = x] = 1$, and $\Pr[\mathscr{S} = \widetilde{\mu}^r(x)| X = x] = 0$ for all $\widetilde{\mu}^r(x) \neq \mu(x)$. Altogether, this implies that Assumption A0 yields

$$
\begin{aligned}
\Pr[Y_1 = 1| X = x] &\equiv \mu_1(x) = F_1\left(x_1'\beta_1 + \Delta_1\mu_2(x)\right); \\
\Pr[Y_2 = 1| X = x] &\equiv \mu_2(x) = F_2\left(x_2'\beta_2 + \Delta_2\mu_1(x)\right).
\end{aligned}
\tag{1}
$$

The inferential setting in this paper is one where the researcher does not observe players' beliefs. The behavioral conditions in Assumption A0 (BNE behavior and degenerate selection mechanism) allows us to identify beliefs. The invertibility features of F_p coupled with Eq. (1) will lead to a constructive identification result for the payoff coefficients following a reparameterization. The identification strategy will match conditional choice probabilities with the observable portion of players' expected utility. By (1), players' behavior is summarized as

$$
\begin{aligned}
Y_1 &= \mathbb{1}\left\{X_1'\beta_1 + \Delta_1\Pr[Y_2 = 1|X] - \zeta_1 \geq 0\right\}; \\
Y_2 &= \mathbb{1}\left\{X_2'\beta_2 + \Delta_2\Pr[Y_1 = 1|X] - \zeta_2 \geq 0\right\}.
\end{aligned}
\tag{2}
$$

Optimal decision rules can be represented as threshold-crossing conditions because the distribution of ζ_p rules out optimal indifference between $Y_p = 1$ and $Y_p = 0$ with probability one.[6]

The degeneracy property of the selection mechanism implies that the data (choices) in each observation (e.g., market) is generated from a single equilibrium. This is a common assumption in two-step estimation of game theoretic models. Examples in both dynamic and static settings include Moro (2003), Pesendorfer and Schmidt-Dengler (2008), Bajari et al. (2010) and Aguirregabiria (2007). In another instance, the modeling assumptions in Pakes et al. (2007) ensure that there is a unique equilibrium associated with a given data generating process. Assuming that the data in each observation is generated from a single equilibrium allows the different procedures described in those papers to use nonparametric conditional choice probability estimates as plug-ins for unobserved beliefs in order to recover the structural parameters of interest. A degenerate selection rule was also assumed in Seim (2006) (who also studied conditions for uniqueness of equilibria), and Jia (2008), who imposed the stronger assumption that the equilibrium selected was an extremal one. In contrast, Assumption A0 does not require knowledge of which equilibrium is being selected, although the smoothness conditions that we will impose below on μ_1 and μ_2 as functions of X will in turn imply "smoothness" properties of the selection mechanism $\mathscr{S}$. These boil down, for instance, to assuming that only stable equilibria are selected w.p.1.[7] Our econometric framework differs from existing work in that we do not impose any parametric assumptions at all on the distribution of unobservable payoff shocks. Dropping the assumption that the data observed is generated by a single equilibrium typically

presents the researcher with two general alternatives. He could either impose specific assumptions on the selection mechanism such as ruling out certain kinds of equilibria (Bresnahan and Reiss, 1991a,b; Tamer, 2003), or model the mechanism explicitly (Bjorn and Vuong, 1984; Sweeting, 2009; Bajari et al., 2005). Alternatively, one could make no assumptions on the selection rule and approach the model explicitly as being only partially identified and do inference accordingly (Ciliberto and Tamer, 2009; Pakes et al., 2006; Andrews et al., 2004; Galichon and Henry, 2011; Beresteanu et al., 2008; Pakes, 2008). The approaches vary from moment-inequality based, to set-estimation and the properties of nonadditive likelihoods and capacity functionals. The assumption of equilibrium behavior is common to all the above cited papers except for Section 6.1 in Beresteanu et al. (2008), where rationalizable behavior as in Aradillas-Lopez and Tamer (2008) is considered.

Note that if there is a unique BNE for a given x, then the selection mechanism $\mathscr{S}$ plays no role whatsoever and the data (for that x) is trivially generated by a single equilibrium. The cardinality of equilibria will depend on the stability of the Jacobian matrix (with respect to μ_1, μ_2) of the BNE system in the set $(\mu_1, \mu_2) \in [0, 1] \times [0, 1]$. These features in turn depend on the magnitude and direction of the mutual strategic interaction effect, and the relative importance of private information, the latter being measured by how "flat" the distributions F_p are. For instance, for a given x it is easy to verify that if $\Delta_1\Delta_2 F_1'(x_1'\beta_1 + \Delta_1 c_2)F_2'(x_2'\beta_2 + \Delta_2 c_1) < 1$ for all constants $(c_1, c_2) \in [0, 1] \times [0, 1]$, then there will be a unique BNE for x. If the derivatives of $F_1(\cdot)$ and $F_2(\cdot)$ are assumed to be uniformly bounded by two constants $\overline{F'}_1$ and $\overline{F'}_2$ respectively, then $\Delta_1\Delta_2 < 1/(\overline{F'}_1\overline{F'}_2)$ would ensure uniqueness everywhere in $\mathbb{S}(X)$ (the support of X). This would be immediately satisfied for instance, if $\Delta_1\Delta_2 \leq 0$ which corresponds to a game where strategic interaction has opposite signs between both players. We will add the following assumption.

Assumption A1. The distributions of X_1 and X_2 are absolutely continuous with respect to Lebesgue measure. The supports of X_1 and X_2 are not contained in any proper linear subspace of $\mathbb{R}^{K_1}$ and $\mathbb{R}^{K_2}$ respectively. There exist regressors $X_{1\ell} \in X_1$ and $X_{2\ell} \in X_2$ such that $\beta_{1\ell} \neq 0$, $\beta_{2\ell} \neq 0$ and conditional on $X_1 \setminus X_{1\ell}$ and $X_2 \setminus X_{2\ell}$, we have $\Pr[X_{1\ell} \neq X_{2\ell}] > 0$. We will also assume that we observe a random sample of N games where $(Y_i, X_i, \zeta_i)_{i=1}^N$ comes from the population described here.

The requirement that all variables in X_1 and X_2 be continuously distributed is not essential and our methodology can be readily adapted accordingly. The exclusion restriction in Assumption A1 which assumes the existence of individual-specific observable payoff shifters will allow identification of the strategic-interaction parameter α_p. This type of restriction can be also found in Bajari et al. (2010). In Section 4.1 we will discuss inference in experimental data sets when this exclusion restriction is not satisfied. We will argue that in those types of settings, having precise knowledge of observable components of payoffs can allow the researcher to consistently test our behavioral model under weaker versions of our assumptions.

2.2. Reparameterization

We are interested in a procedure to estimate the payoff parameters that relies only on Eq. (1) and on the invertibility properties of $F_p(\cdot)$ in an inferential setting where the researcher observes an iid sample from the population of games described above. Exploiting these features, our estimation procedure will match (asymptotically) conditional choice probabilities with the observable components of players' expected utility. We will exploit

---

[6] This is an important distinction with complete information models, where knowing the exact realization of everybody's payoffs enables agents to mix their actions in a way that makes their opponents optimally indifferent between more than one choice. Moreover, even ruling out mixed-strategy equilibria does not yield threshold-crossing decision rules (e.g., see Bresnahan and Reiss, 1991a or Tamer, 2003). As a result, ex-ante knowledge of the direction of strategic interaction (i.e., the signs of Δ_1 and Δ_2 in our case) is necessary in order to do inference. This feature is true both for identified models (Bresnahan and Reiss, 1991a) or partially identified models (Ciliberto and Tamer, 2009).

[7] In our model, an equilibrium $\widetilde{\mu}^r$ is stable if the Jacobian with respect to μ_1, μ_2 of the BNE equilibrium system has full rank evaluated at $\widetilde{\mu}^r$. Assuming that the determinant of such Jacobian is nonzero w.p.1. would suffice. A discussion on smooth equilibrium selection and stable equilibria can also be found in Bajari et al. (2010).

the fact that, with probability one, for any pair of observations i, j in the sample,

$$\mu_p(X_i) = \mu_p(X_j)$$
$$\Longleftrightarrow X'_{pi}\beta_p + \Delta_p\mu_{-p}(X_i) = X'_{pj}\beta_p + \Delta_p\mu_{-p}(X_j),$$

where the above parameters are evaluated at their true value. Consequently, for identification purposes our parameter space Θ and the stochastic properties of X need to be such that there does not exist $(\widetilde{\beta}_p, \widetilde{\Delta}_p) \neq (\beta_p, \Delta_p)$ such that w.p.1., $\mu_p(X_i) = \mu_p(X_j) \Longleftrightarrow X'_{pi}\widetilde{\beta}_p + \widetilde{\Delta}_p\mu_{-p}(X_i) = X'_{pj}\widetilde{\beta}_p + \widetilde{\Delta}_p\mu_{-p}(X_j)$. Given our previous results, this situation will be ruled out if and only if there *does not exist an invertible mapping* $\Gamma : \mathbb{R} \to \mathbb{R}$ such that $\Gamma(X'_p\beta_p + \Delta_p\mu_{-p}(X)) = X'_p\widetilde{\beta}_p + \widetilde{\Delta}_p\mu_{-p}(X)$ w.p.1. for some pair of parameter values $(\beta_p, \Delta_p) \neq (\widetilde{\beta}_p, \widetilde{\Delta}_p)$ in Θ. For the "only if" part, suppose such a mapping Γ does exist for the true parameter value (β_p, Δ_p). Then w.p.1. we would have $F_p(X'_p\beta_p + \Delta_p\mu_{-p}(X)) = F_p(\Gamma(X'_p\widetilde{\beta}_p + \widetilde{\Delta}_p\mu_{-p}(X))) \equiv \widetilde{F}_p(X'_p\widetilde{\beta}_p + \widetilde{\Delta}_p\mu_{-p}(X))$. Note that $\widetilde{F}_p$ would be invertible by the properties of F_p, and w.p.1. we would have $\mu_p(X_i) = \mu_p(X_j) \Longleftrightarrow X'_{pi}\widetilde{\beta}_p + \widetilde{\Delta}_p\mu_{-p}(X_i) = X'_{pj}\widetilde{\beta}_p + \widetilde{\Delta}_p\mu_{-p}(X_j)$. For the "if" part, suppose such a mapping Γ does not exist. Then for any $(\widetilde{\beta}_p, \widetilde{\Delta}_p) \neq (\beta_p, \Delta_p)$, with positive probability we can have $\mu_p(X_i) = \mu_p(X_j)$ but $X'_{pi}\widetilde{\beta}_p + \widetilde{\Delta}_p\mu_{-p}(X_i) \neq X'_{pj}\widetilde{\beta}_p + \widetilde{\Delta}_p\mu_{-p}(X_j)$.

By the linear index nature of expected payoffs in our model, it is easy to see that we only need to look at linear mappings of the form $\Gamma(\psi) = a + b\psi$. It follows immediately that our parameter space needs to be such that for both $p = 1, 2$,

$$\nexists a, b \in \mathbb{R} : a + b \cdot (\beta'_p X_p + \Delta_p\mu_{-p}(X)) = \widetilde{\beta}'_p X_p + \widetilde{\Delta}_p\mu_{-p}(X)$$
    w.p.1. for some $(\beta_p, \Delta_p) \neq (\widetilde{\beta}_p, \widetilde{\Delta}_p)$.

For the above condition to be satisfied, X_p cannot include a constant term (i.e., an intercept cannot be identified), and our parameter space must be such that the norm of at least a subset of parameters in (β_p, Δ_p) is fixed and known. In addition, the covariates in X_p must satisfy a full-rank condition and $\mu_p(X)$ needs to have a rich enough (i.e., continuous) support. The scale normalization will be addressed by assuming the existence, for each p, of a continuously distributed regressor W_p (whose identity is known to the researcher) with a strictly positive slope coefficient. Our parameter space will normalize this coefficient to 1. Thus, our model will be reparameterized as follows: We will split $X_p = (W_p, V_p)$ and we will have $\beta_p = (1, \gamma_p)$. The unknown payoff parameter vector for p will hence be denoted by $\theta_p = (\gamma_p, \alpha_p)$, with α_p denoting the strategic-interaction parameter in our normalized parameter space. We will define $Z_p = (V_p, \mu_{-p}(X))$. Therefore, (1) becomes

$$\Pr[Y_p = 1|X] \equiv \mu_p(X) = F_p(W_p + Z'_p\theta_p) \quad \text{for } p = 1, 2.$$

The regressor W_p will be assumed to be continuously distributed conditional on all other covariates and V_p will be assumed to have full-column rank with positive probability. Normalizations of scale and location are common in semiparametric index models. A partial list of well-known examples includes Manski (1975, 1985), Han (1987), Horowitz (1992), Klein and Spady (1993), Ichimura (1993), Ichimura and Lee (1991) and Sherman (1993). Instances in which the absolute value of an individual coefficient was normalized to 1 include Sherman (1993) and Horowitz (1992). In our context, normalizing an individual coefficient to 1 will have the added computational advantage of resulting in an estimator with a closed-form expression.

Proposition 1. *As before let* $\theta = (\theta_1, \theta_2)$ *denote the true value of the payoff parameters and let* $\widetilde{\theta} = (\widetilde{\theta}_1, \widetilde{\theta}_2)$ *denote an arbitrary value for it. Consider an i.i.d sample from the population described above and denote* $\mu_p(X_i) \equiv \mu_{pi}$ *for the ith observation in the sample. Take any two observations* $i \neq j$. *If Assumptions* A0 *and* A1 *are satisfied, then under our reparameterization*

$$E\left[\left((W_{1i} - W_{1j}) + (Z_{1i} - Z_{1j})'\widetilde{\theta}_1\right)^2 \middle| \mu_{1i} = \mu_{1j}\right] \quad \text{and}$$

$$E\left[\left((W_{2i} - W_{2j}) + (Z_{2i} - Z_{2j})'\widetilde{\theta}_2\right)^2 \middle| \mu_{2i} = \mu_{2j}\right]$$

are each uniquely minimized at $\widetilde{\theta}_1 = \theta_1$ *and* $\widetilde{\theta}_2 = \theta_2$ *respectively.*

The result of Proposition 1 is a direct consequence of A0 and A1 and the properties of our parameter space. Invertibility of $F_p(\cdot)$ implies

$$E\left[\left((W_{pi} - W_{pj}) + (Z_{pi} - Z_{pj})'\widetilde{\theta}_p\right)^2\right]$$
$$= E\left[\left((Z_{pi} - Z_{pj})'(\widetilde{\theta}_p - \theta_p) + F_p^{-1}(\mu_{pi}) - F_p^{-1}(\mu_{pj})\right)^2\right].$$

Therefore whenever $\mu_{pi} = \mu_{pj}$ we have $E[((W_{pi} - W_{pj}) + (Z_{pi} - Z_{pj})'\widetilde{\theta}_p)^2] = E\left[\left((Z_{pi} - Z_{pj})'(\widetilde{\theta}_p - \theta_p)\right)^2\right]$. Next, note that by the full-rank assumption of V_{pi} and the exclusion restriction in Assumption A1, element-wise we have $\Pr[Z_{pi} \neq Z_{pj}|\mu_{pi} = \mu_{pj}] > 0$. In particular, the exclusion restriction in Assumption A1 (i.e., the presence of individual-specific observable payoff shifters) implies that μ_{1i} and μ_{2i} are not deterministic conditional on each other. This allows for identification of the strategic interaction coefficient α_p. It follows that the above conditional expectation is uniquely minimized at $\widetilde{\theta}_p = \theta_p$.[8] The estimation procedure we develop here exploits these conditional moment restrictions.

3. Estimation

Given Proposition 1, if μ_{1i} and μ_{2i} were exactly known and if we had an i.i.d sample of the population described above, a pairwise-difference estimator for θ_p in the spirit of Honoré and Powell (2005) would minimize the objective function $\binom{N}{2}^{-1}\frac{1}{h_a}\sum_{i<j}K_a(\frac{\mu_{pi}-\mu_{pj}}{h_a})[(W_{pi} - W_{pj}) + (Z_{pi} - Z_{pj})'b_p]^2$, where $K_a(\cdot)$ and h_a would be appropriately chosen kernel and bandwidth respectively.[9] The intuition is that conditional on X_i, if the kernel $K_a(\cdot)$ and bandwidth h_a are chosen appropriately then for any given vector of constants b_p,

$$\frac{1}{h_a}\binom{N}{2}^{-1}\sum_{i<j}K_a\left(\frac{\mu_{pj} - \mu_{pi}}{h_a}\right)\left[(W_{pj} - W_{pi}) + (Z_{pj} - Z_{pi})'b_p\right]^2$$

is a consistent estimator for $E[[(W_{pj} - W_{pi}) + (Z_{pj} - Z_{pi})'b_p]^2 \mid \mu_{pj} = \mu_{pi}, X_i]$. Since both μ_{pi} and μ_{-pi} (and therefore, Z_{pi}) are unknown, this strategy unfeasible. We will replace these unknown functions with nonparametric estimates.

---

[8] In Section 4.1 we discuss inference in experimental data sets without individual-specific observable payoff shifters. There we argue that having exact knowledge of the true properties of the observable payoff components allows the researcher to do consistent behavioral specification testing.

[9] The notation $\sum_{i<j}(v_i, v_j)$ describes a summation over all combinations of pairs of indices (i, j) out of $(1, 2, \ldots, n)$.

3.1. Nonparametric estimators for $\widehat{\mu}_1(\cdot)$ and $\widehat{\mu}_2(\cdot)$

Lack of knowledge about the functional forms for $F_p(\cdot)$, $p = 1, 2$ implies that $\mu_p(\cdot)$ is also unknown. As before, let $X \equiv X_1 \cup X_2$ and denote the dimension of X by L. We will employ the usual kernel-smoothed estimators, for $p = 1, 2$ and $x \in \mathbb{S}(X)$,

$$\widehat{\mu}_p(X_i) \equiv \widehat{\mu}_{pi} = \frac{1}{Nh_b^L} \sum_{j=1}^{N} Y_{pj} K_b \left(\frac{X_j - X_i}{h_b} \right) \bigg/ \frac{1}{Nh_b^L}$$

$$\times \sum_{j=1}^{N} K_b \left(\frac{X_j - X_i}{h_b} \right). \tag{3}$$

The properties of the bandwidth sequence h_b and the kernel $K_b(\cdot)$ will be carefully detailed below. From now on, we will use $\widehat{\mu}_p(X_i)$ and μ_{pi} interchangeably. Let

$$\widehat{Z}_{1i} \equiv \left(V'_{1i}, \widehat{\mu}_{2i} \right)'; \qquad \widehat{Z}_{2i} \equiv \left(V'_{2i}, \widehat{\mu}_{1i} \right)'.$$

$\widehat{Z}_{pi}$ constitute *generated regressors* used instead of their unknown nonparametric population counterparts. The estimator analyzed here will plug-in these nonparametric estimates into a kernel-weighed objective function. The model we study here constitutes the first known instance of a pairwise differencing procedure with nonparametric matching control variables (i.e., those plugged into the kernel function K_a) *and* nonparametric or "generated" regressors. The following section describes the estimator and its properties.

3.2. Pairwise-difference estimator

Let $\phi : \mathbb{R}^L \to \mathbb{R}_+$ be a function such that $\phi(X) > 0$ if $X \in \mathcal{X}$ and $\phi(X) = 0$ if $X \in \mathbb{S}(X) \setminus \mathcal{X}$ and $\mathcal{X}$ is a compact set in the interior of $\mathbb{S}(X)$. The estimator we study here is given by:

$$\widehat{\theta}_p = \underset{b}{\text{argmin}} \; \frac{1}{h_a} \binom{N}{2}^{-1} \sum_{i<j} K_a \left(\frac{\widehat{\mu}_{pi} - \widehat{\mu}_{pj}}{h_a} \right)$$

$$\times \left[\left(W_{pi} - W_{pj} \right) + \left(\widehat{Z}_{pi} - \widehat{Z}_{pj} \right)' b \right]^2 \phi(X_i)\phi(X_j). \tag{4}$$

This is a pairwise-difference procedure with nonparametric control variables (the unknown functions inside $K_a(\cdot)$). Aradillas-Lopez et al. (2007) (AHP) study the asymptotic properties of such procedures in a context that differs from (4). To see how, note that all the asymptotic results in AHP are obtained for estimators that minimize with respect to b an objective function of the form

$$\binom{N}{2}^{-1} \sum_{i<j} K \left(\frac{\widehat{\mu}(W_i, \widehat{\gamma}) - \widehat{\mu}(W_j, \widehat{\gamma})}{h_N} \right) s(V_i, V_j; b)\phi(W_i)\phi(W_j)$$

(see Eq. (18) and Section 3.3 in AHP), where $\widehat{\mu}$ is nonparametrically estimated and $\widehat{\gamma}$ is a finite-dimensional parameter that is either known (or not present at all), or a $\sqrt{N}$-consistent asymptotically normal estimator. $\phi(\cdot)$ is a trimming function. The crucial feature of the objective function in AHP is that the parameterized index function $s(v_i, v_j; b)$ is not assumed to include any unknown functions or "generated regressors". In contrast, the estimator in (4) involves an objective function of the form

$$\binom{N}{2}^{-1} \sum_{i<j} K \left(\frac{\widehat{\mu}(W_i, \widehat{\gamma}) - \widehat{\mu}(W_j, \widehat{\gamma})}{h_N} \right) s(\widehat{V}_i, \widehat{V}_j; b)\phi(W_i)\phi(W_j),$$

where the generated regressors in $\widehat{V}_i$ and $\widehat{V}_j$ are nonparametrically estimated (and $\widehat{\gamma}$ is not present). For asymptotic purposes, this distinction is as relevant as the one we would make between a linear regression model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$ where X_{1i}

and X_{2i} are observed, and $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 E[Z_i|X_{2i}] + \varepsilon_i$, where $E[Z_i|X_{2i}]$ has unknown functional form and it is replaced with a nonparametric estimator $\widehat{E}[Z_i|X_{2i}]$. Another instance that highlights the asymptotic relevance of dealing with nonparametric regressors is Ahn and Manski (1993), who study the asymptotic distribution of β in a binary choice model of the form $Y_i = \mathbb{1}\left\{X'_{1i}\beta_1 + \beta_2 E[Z_i|X_{2i}] + \varepsilon_i \geq 0\right\}$ where the distribution of ε_i is assumed to be known up to a finite dimensional parameter and $E[Z_i|X_{2i}]$ has unknown functional form and is replaced with a kernel-weighed nonparametric estimate. Just like the standard asymptotic theory of linear regression or that of binary choice models would be inappropriate to do inference in counterpart cases that include nonparametric or "generated" regressors,[10] the asymptotic results in AHP are not useful to us. Furthermore, the conditions and assumptions in AHP do not tell us whether a $\sqrt{N}$-consistent asymptotically normal estimator can be constructed in a situation where both the control variables μ and the index function s contain nonparametric functions. We will describe exactly why this is the case below, in the discussion following Theorem 1. To our knowledge, the model we study here constitutes the first instance of a pairwise-differencing procedure with nonparametric matching control variables *and* nonparametric or "generated" regressors.[11, 12]

The optimization problem in (4) yields a simple, closed-form expression for $\widehat{\theta}_p$,

$$\widehat{\theta}_p = \left[-\binom{N}{2}^{-1} \frac{1}{h_a} \sum_{i<j} K_a \left(\frac{\widehat{\mu}_{pi} - \widehat{\mu}_{pj}}{h_a} \right) \right.$$

$$\times \left(\widehat{Z}_{pi} - \widehat{Z}_{pj} \right) \left(\widehat{Z}_{pi} - \widehat{Z}_{pj} \right)' \phi(X_i)\phi(X_j) \right]^{-1}$$

$$\times \left[\binom{N}{2}^{-1} \frac{1}{h_a} \sum_{i<j} K_a \left(\frac{\widehat{\mu}_{pi} - \widehat{\mu}_{pj}}{h_a} \right) \right.$$

$$\times \left(\widehat{Z}_{pi} - \widehat{Z}_{pj} \right) \left(W_{pi} - W_{pj} \right) \phi(X_i)\phi(X_j) \right] \tag{5}$$

which resemble the expression for weighted-least squares estimators. We use a trimming function $\phi(\cdot)$ in order to stay away from points near the boundary of the support of X. This will enable us to keep under control the order of magnitude of the bias of $\widehat{\mu}_p(\cdot)$ uniformly in our sample. We will study the properties of the proposed estimator under the following assumptions.

3.3. Additional assumptions

In addition to A0 and A1, we will consider the following:

Assumption A2. There exists a constant $\overline{M} \geq L + 1$ (recall that L is the dimension of $X \equiv X_1 \cup X_2$) such that $F_1(\cdot)$ and $F_2(\cdot)$ are $\overline{M}$-times differentiable with bounded derivatives. Let $f_X(x)$ be the

---

[10] Rilstone (1993) characterizes semiparametric efficiency bounds in simple models with nonparametric regressors.

[11] The model we study here is briefly mentioned in Section 2.3.1 of AHP as an extension of their setup, but no attempt was made there to obtain asymptotic results for such a model.

[12] Hong and Shum (2010) describe a pairwise-difference estimation procedure for a dynamic optimization model. Strong assumptions about the accumulation equation in their model yields a control variable that has a known functional form. Section 2.3.2 of AHP briefly suggests how to potentially extend their setup to one with nonparametric control variables, but once again no asymptotic results are presented.

joint density of X. Then $f_X(\cdot)$ is $\overline{M}$-times differentiable with respect to all of its arguments, with bounded derivatives. There exists a random variable ν such that $\|X\| \leq \nu$ w.p.1 and $E[\nu^4] < \infty$.

Assumption A3. The trimming function $\phi(\cdot)$ is bounded, $\overline{M}$-times differentiable with bounded derivatives everywhere in $\mathbb{R}$, it is positive for all $x \in \mathcal{X} \subset \text{int}(\mathbb{S}(X))$ and zero for all $x \in \mathbb{S}(X) \setminus \mathcal{X}$. For $p = 1, 2$, let $F_p'(\cdot)$ denote the first derivative of $F_p(\cdot)$, then $\exists \underline{\ell} > 0, \underline{f} > 0$ such that $\inf \{F_p'(x_p'\beta_p + \alpha_p\mu) : x \in \mathcal{X}, \mu \in [0, 1]\} > \underline{\ell}$ for $p = 1, 2$ and $\inf \{f_X(x) : x \in \mathcal{X}\} > \underline{f}$. In addition, the trimming function $\phi(\cdot)$ is such that $E[\phi(X)|\mu_p(X) = \mu]$ and $E[Z_p\phi(X)|\mu_p(X) = \mu]$ are both $\overline{M}$-times differentiable functions of μ with bounded derivatives for $p = 1, 2$.

To be more precise, given the continuously-distributed nature of X, the trimming function $\phi(\cdot)$ only needs to be differentiable almost everywhere in $\mathbb{S}(X)$.[13] Given the previous assumptions, the last part of A3 is satisfied if neither $\phi(X_1)$ nor $\phi(X_2)$ are deterministic conditional on $\mu_p(X)$ for $p = 1, 2$. Note also that Assumptions A0–A3 imply that the density $f_{\mu_p}(\mu)$ is $\overline{M}$-times differentiable with bounded derivatives for $p = 1, 2$.

Assumption A4. Let $\overline{M}$ be as defined above. The kernels $K_a(\cdot), K_b(\cdot)$ and bandwidths h_a, h_b satisfy

(i) The kernel $K_a(\cdot) : \mathbb{R} \to \mathbb{R}$ is symmetric around zero, bounded and bias-reducing of order $\overline{M}$. In addition, $K_a(\cdot)$ has bounded first and second derivatives everywhere in $\mathbb{R}$ denoted by $K_a^{(1)}(\cdot)$ and $K_a^{(2)}(\cdot)$ which satisfy

$$\int K_a^{(1)} d\psi = 0; \qquad \int \psi K_a^{(1)}(\psi) d\psi = -1;$$

$$\int \psi^j K_a^{(1)}(\psi) d\psi = 0 \quad \text{for } j = 2, \ldots, \overline{d} - 1$$

and $\int \psi^{\overline{d}} K_a^{(1)}(\psi) d\psi < \infty$, with $\overline{d} > 2$.

(ii) The kernel $K_b(\cdot) : \mathbb{R}^L \to \mathbb{R}$ is symmetric around zero, bounded and bias-reducing of order $\overline{M}$.

(iii) There exists $\overline{\delta} > 0$ such that $N^{1/2-\overline{\delta}} h_b^L h_a^2 \to \infty$. Let $\overline{d}$ and $\overline{M}$ be as described above, then

$$N^{1/2} h_a^{\overline{d}} \to 0, \qquad N^{1/2} h_a^{\overline{M}} \to 0, \qquad N^{1/2} h_b^{\overline{M}} / h_a^2 \to 0.$$

Using bandwidths of the form $h_a \propto N^{-a}$ and $h_b \propto N^{-b}$ for some $a > 0, b > 0$ it is easy to show that $\overline{M}$ is bounded above by a function that increases with L. The following full-rank assumption is the last piece to ensure asymptotic normality of our estimator.

Assumption A5. The following matrix has full rank for $p = 1, 2$

$$D_p \equiv E\left[\left(E[Z_p Z_p'\phi(X)|\mu_p] \cdot E[\phi(X)|\mu_p] \right. \right.$$
$$\left. \left. - E[Z_p\phi(X)|\mu_p]E[Z_p\phi(X)|\mu_p]'\right) f_{\mu_p}(\mu_p)\right].$$

Note that given the definition of Z_p, a necessary condition for asmA5 to hold is that for $p \neq q$ with $p, q \in \{1, 2\}$, $\Pr\left(E[\mu_p|\mu_q] \neq \mu_p | X \in \mathcal{X}\right) > 0$. The exclusion restrictions in Assumption A1 are sufficient to ensure this.

Theorem 1. Let D_p be as described in Assumption A5. Let

$$\tau_{pi} = \left(Z_{pi}E\left[\phi(X)|\mu_{pi}\right] - E\left[Z_p\,\phi(X)|\,\mu_{pi}\right]\right)\left(Y_{-pi} - \mu_{-pi}\right)\phi(X_i)$$

$$\upsilon_{pi} = \frac{\left(Z_{pi}E\left[\phi(X)|\mu_{pi}\right] - E\left[Z_p\phi(X)|\mu_{pi}\right]\right) f_{\mu_p}(\mu_{pi})}{F_p'\left(W_{pi} + Z_{pi}'\theta_p\right)}$$
$$\times \left(Y_{pi} - \mu_{pi}\right)\phi(X_i)$$

$$\psi_i^{\theta_p} = D_p^{-1} \times \left(\tau_{pi}\alpha_p + \upsilon_{pi}\right). \tag{6}$$

Then, if Assumptions A0–A5 are satisfied

$$\widehat{\theta}_p - \theta_p = \frac{1}{N} \sum_{i=1}^{N} \psi_i^{\theta_p} + o_p(N^{-1/2}), \quad \text{and consequently}$$

$$\sqrt{N}(\widehat{\theta}_p - \theta_p) \xrightarrow{d} \mathcal{N}\left(0, E\left[\psi^{\theta_p}\psi_p^{\theta_p\prime}\right]\right).$$

The term τ_{pi} reflects the researcher's lack knowledge of the true functional form of Z_{pi}. This term is not present in the influence function described in Theorem 4 of AHP. Therefore, using the asymptotic results in that paper would lead us to incorrect standard errors and inconsistent inference here. The reason why τ_{pi} is absent in AHP is because that paper specializes on the case where the index function parameterized by θ does not include unknown functions. The second term in our influence function is υ_{pi}, and it reflects the nonparametric nature of the control variables used in K_a. A quick look at the influence function ψ^{θ_p} reveals once again the importance of the condition $\Pr\left(E\left[\mu_p|\mu_{-p}\right] \neq \mu_p | X \in \mathcal{X}\right) > 0$ for $p = 1, 2$. If this condition does not hold, the interaction parameter α_p would not be estimable. The exclusion restrictions in Assumption A1 are sufficient for this condition to be satisfied.[14] The next section describes the properties of a consistent specification test that uses the estimator $\widehat{\theta}_p$ just described.

4. A consistent specification test

This section describes a consistent specification test which asymptotically rejects our model if Eq. (1) is violated with nonzero probability. In the spirit of Fan and Li (1996) and Zheng (1998), our test-statistic is based on semiparametrically estimated residuals. To the best of our knowledge, our test is the first of this kind that involves nonparametric or "generated" regressors. As we will see the presence of these regressors is nontrivial since we will need to carefully determine the relative rates of convergence of the various bandwidths involved in order to ensure an asymptotically pivotal distribution for our test-statistic under the null hypothesis that the model is correctly specified. This issue is not present in either Fan and Li (1996) or Zheng (1998) and, to the best of our knowledge, has not been documented before in consistent specification tests of this type. As before, denote $X_i \equiv (W_{1i}, V_{1i}, W_{2i}, V_{2i})$. Now let

$$E[Y_p|X_i] \equiv \mu_p(X_i) \equiv \mu_{pi};$$
$$t_{pi} \equiv W_{pi} + V_{pi}'\gamma_p + \alpha_p\mu_{-pi} \equiv W_{pi} + Z_{pi}'\theta_p; \tag{7}$$
$$E[Y_p|t_{pi}] \equiv F_p(t_{pi}).$$

If our model is correct, Eq. (1) implies that $\mu_p(X_i) = F_p(t_{pi})$ w.p.1. Let $\widehat{\theta}_p$ be the estimator we developed above. Let $\varphi, \overline{\phi} : \mathbb{R}^L \to \mathbb{R}_+$ be a pair of nonnegative trimming functions function which are nonzero for all X in a compact set $\mathcal{X}$. Take two kernel functions, $\mathcal{H}_b : \mathbb{R}^L \to \mathbb{R}$ and $\mathcal{H}_c : \mathbb{R} \to \mathbb{R}$ and bandwidth sequences $\widetilde{h}_b, \widetilde{h}_c$. Use the notation $\Delta\psi_{ij} \equiv \psi_i - \psi_j$ and let

---

[13] Thus, everything that follows would work for example if $\phi(X) = \mathbb{1}\{X \in \mathcal{X}\}$.

[14] See Section 4.1, where we outline how the design of payoffs by the researcher in experimental settings could potentially allow certain types of inference even if the exclusion restriction in Assumption A1 fails.

$$\widetilde{\mu}_p(X_i) \equiv \widetilde{\mu}_{pi}$$

$$= \frac{1}{N\widetilde{h}_b^L} \sum_{j=1}^N Y_{pj} \mathcal{H}_b\left(\frac{\Delta X_{ji}}{\widetilde{h}_b}\right) \Big/ \frac{1}{N\widetilde{h}_b^L} \sum_{j=1}^N \mathcal{H}_b\left(\frac{\Delta X_{ji}}{\widetilde{h}_b}\right);$$

$$\widehat{t}_{pi} \equiv \underbrace{W_{pi} + V'_{pi}\widehat{\gamma}_p + \widehat{\alpha}_p\widetilde{\mu}_{-pi}}_{\equiv W_{pi} + \widetilde{Z}'_{pi}\widehat{\theta}_p};$$

$$T_{p_N}(\widehat{t}_{pi}) = \frac{1}{N\widetilde{h}_c} \sum_{j=1}^N \mathcal{H}_c\left(\frac{\Delta \widehat{t}_{pji}}{\widetilde{h}_c}\right) \varphi(X_j); \qquad (8)$$

$$S_{p_N}(\widehat{t}_{pi}) = \frac{1}{N\widetilde{h}_c} \sum_{j=1}^N Y_{pj} \mathcal{H}_c\left(\frac{\Delta \widehat{t}_{pji}}{\widetilde{h}_c}\right) \varphi(X_j);$$

$$\widehat{F}_p(\widehat{t}_{pi}) = \frac{S_{p_N}(\widehat{t}_{pi})}{T_{p_N}(\widehat{t}_{pi})}.$$

Finally let $\overline{H} : \mathbb{R}^L \to \mathbb{R}$ and $\widetilde{b}$ be another kernel function and bandwidth sequence. Define, for $p = 1, 2$,

$$\widehat{\varepsilon}_{pi} \equiv Y_{pi} - \widehat{F}_p(\widehat{t}_{pi});$$

$$U_{p_N} \equiv \binom{N}{2}^{-1} \sum_{i<j} \frac{\widehat{\varepsilon}_{pj}\widehat{\varepsilon}_{pi}\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L} \overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right). \qquad (9)$$

The bandwidth sequences involved in the construction of our test-statistic must satisfy very particular relative rates of convergence (to be detailed in Eqs. (10)–(11), below) in order to achieve an asymptotic pivotal distribution if the model is correctly specified. In particular, we will require $\widetilde{b} \neq \widetilde{h}_b \neq \widetilde{h}_c$. Since only two of these three bandwidths (the first and the third one) would be present in the general setting of Fan and Li (1996), the conditions in that paper are not useful to us and there is no known result in the literature on these type of consistent specification tests that we can use. We will derive the relevant conditions here. We will use U_{p_N} to construct a test statistic that will become asymptotically unbounded if $\Pr\left[\mu_p(X) \neq F_p(t_p)\,\middle|\, X \in \mathcal{X}\right] > 0$ for $p = 1$ or $p = 2$ (i.e., if Eq. (1) is violated with positive probability in $\mathcal{X}$). Conversely, if the model is correctly specified our test-statistic will have a χ^2_2 asymptotic distribution. The following additional conditions will yield the desired result.

Assumption B1. The trimming functions $\varphi(\cdot)$ and $\overline{\phi}(\cdot)$ are bounded, nonnegative for all $X \in \mathbb{R}^L$, strictly positive everywhere inside a compact set $\mathcal{X}$, and exactly equal to zero for all $X \notin \mathcal{X}$. The trimming set $\mathcal{X}$ is assumed to be in the interior of $\mathbb{S}(X)$. In particular, $\mathcal{X}$ is such that $f_{t_p}(t_p)$, $E\left[\varphi(X)\,\middle|\, t_p\right]$, and $f_x(X)$ are uniformly bounded away from zero for all $X \in \mathcal{X}$.[15] In addition, there exists a constant $\overline{M} > L$ such that the following objects are $\overline{M}$-times differentiable with respect to $X \in \mathbb{R}^L$, with bounded derivatives whenever $X \in \mathcal{X}$: $\varphi(X), \overline{\phi}(X), \mu_p(X)$ (for $p = 1, 2$), and $f_x(X)$. Let t_p and $F_p(t_p)$ be as defined in (7) and let $f_{t_p}(\cdot)$ denote the density of t_p. The following objects are $\overline{M}$-times differentiable with respect to t_p, with bounded derivatives whenever $X \in \mathcal{X}$[16]:

$$Q_p(t_p) = \nabla_{t_p}\left(E\left[\varphi(X)Z_p\,\middle|\, t_p\right]f_{t_p}(t_p)\right);$$

$$R_p(t_p) = \nabla_{t_p}\left(E\left[\varphi(X)\,\middle|\, t_p\right]f_{t_p}(t_p)\right);$$

$$\widetilde{Q}_p(t_p) = \nabla_{t_p}\left(E\left[\varphi(X)Z_p\,\middle|\, t_p\right]F_p(t_p)f_{t_p}(t_p)\right);$$

$$\widetilde{R}_p(t_p) = \nabla_{t_p}\left(E\left[\varphi(X)\,\middle|\, t_p\right]F_p(t_p)f_{t_p}(t_p)\right).$$

---

[15] Note that t_p is a deterministic, real-valued function of X.

[16] See footnote 14.

The scalar $\overline{M}$ can be thought of as a "measure of smoothness" of the unknown functions involved. Further restrictions involving $\overline{M}$ and the rates of convergence of the bandwidths $\widetilde{b}, \widetilde{h}_b$ and $\widetilde{h}_c$ will be described below in Eq. (11).

Assumption B2. $\mathcal{H}_b(\cdot)$ and $\overline{H}(\cdot)$ are bounded, symmetric around zero, bias-reducing kernels of order $\overline{M}$. They also satisfy $\int \overline{H}^2(\Psi)\,\mathrm{d}\Psi < \infty$ and $\int \mathcal{H}_b(\Psi)^2\mathrm{d}\Psi < \infty$. In addition, $\mathcal{H}_b(\cdot)$ is $\overline{M}$-times differentiable, with bounded derivatives. The kernel $\mathcal{H}_c$ is also symmetric around zero, bounded, bias-reducing of order $\overline{M}$. Like $\mathcal{H}_b$, the kernel $\mathcal{H}_c$ is also $\overline{M}$-times differentiable with bounded derivatives and satisfies $\int \mathcal{H}_c(\psi)^2\mathrm{d}\psi < \infty$. Let $\mathcal{H}_c^{(1)}(\cdot)$ denote the first derivative of the kernel $\mathcal{H}_c(\cdot)$. Then $\mathcal{H}_c^{(1)}(\cdot)$ satisfies

$$\int_{-\infty}^\infty \mathcal{H}_c^{(1)}(\psi)\mathrm{d}\psi = 0; \qquad \int_{-\infty}^\infty \psi\mathcal{H}_c^{(1)}(\psi)\mathrm{d}\psi = -1;$$

$$\int_{-\infty}^\infty \psi^j\mathcal{H}_c^{(1)}(\psi)\mathrm{d}\psi = 0 \quad \text{for } m = 2, \ldots, \overline{M} - 1,$$

and $\int_{-\infty}^\infty \psi^{\overline{M}}\mathcal{H}_c^{(1)}(\psi)\mathrm{d}\psi < \infty$. The bandwidths $\widetilde{b} \to 0, \widetilde{h}_b \to 0$ and $\widetilde{h}_c \to 0$ satisfy

$$N\widetilde{b}^L \longrightarrow \infty; \qquad N\widetilde{h}_b^L \longrightarrow \infty; \qquad N\widetilde{h}_c \longrightarrow \infty;$$

$$N^{1/2}\widetilde{b}^{L/2}\widetilde{h}_b^L \longrightarrow \infty; \qquad N^{1/2}\widetilde{b}^{L/2}\widetilde{h}_c^2 \longrightarrow \infty. \qquad (10)$$

In addition, there exists a $\delta > 0$ such that

$$N^\delta\left(\frac{\widetilde{b}^{L/2}}{\widetilde{h}_c^4}\right) \longrightarrow 0; \qquad N^\delta\left(\frac{\widetilde{b}^{1/2}}{\widetilde{h}_b^2}\right)^L \longrightarrow 0;$$

$$\frac{1}{N^{1-2\delta}}\left[\frac{\widetilde{b}^{L/2}}{\widetilde{h}_b^{2L}\widetilde{h}_c^6}\right] \longrightarrow 0,$$

and the "smoothness measure" $\overline{M}$ is such that

$$\left(\frac{N}{\widetilde{b}^{L/2}}\right)\widetilde{h}_b^{\overline{M}} \longrightarrow 0; \qquad \left(\frac{N}{\widetilde{b}^{L/2}}\right)\widetilde{h}_c^{\overline{M}} \longrightarrow 0;$$

$$\left(\frac{N\widetilde{b}^{L/2}}{\widetilde{h}_c^2}\right)\left(\frac{\widetilde{b}}{\widetilde{h}_c}\right)^{\overline{M}} \longrightarrow 0; \qquad \left(\frac{N\widetilde{b}^{L/2}}{\widetilde{h}_b^L}\right)\left(\frac{\widetilde{b}}{\widetilde{h}_b}\right)^{\overline{M}} \longrightarrow 0; \qquad (11)$$

$$\left(\frac{N\widetilde{b}^{L/2}}{\widetilde{h}_c^4}\right)\left(\frac{\widetilde{h}_b}{\widetilde{h}_c}\right)^{2\overline{M}} \longrightarrow 0.$$

Eqs. (10)–(11) describe the relative convergence rates among the bandwidths involved in the construction of our test-statistic. They provide precise guidelines to extend the asymptotically pivotal property of the type of test-statistics used in Fan and Li (1996) to a model where the covariates involved include nonparametric functions or "generated regressors". The main result is presented next.

Theorem 2. 1. *Let U_{p_N} be as defined in (9). Let*

$$\widehat{\Sigma}_p = \binom{N}{2}^{-1} \sum_{i<j} \frac{\widehat{\varepsilon}_{pi}^2\widehat{\varepsilon}_{pj}^2\overline{\phi}(X_i)^2\overline{\phi}(X_j)^2}{\widetilde{b}^L} \overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right)^2,$$

$$\widehat{\Sigma}_{1,2} = \binom{N}{2}^{-1} \sum_{i<j} \frac{\widehat{\varepsilon}_{1i}\widehat{\varepsilon}_{1j}\widehat{\varepsilon}_{2i}\widehat{\varepsilon}_{2j}\overline{\phi}(X_i)^2\overline{\phi}(X_j)^2}{\widetilde{b}^L} \overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right)^2,$$

$$\mathcal{T}_{p_N} = N\widetilde{b}^{L/2}U_{p_N}, \qquad \widehat{\Sigma} = \begin{pmatrix} \widehat{\Sigma}_1 & \widehat{\Sigma}_{1,2} \\ \widehat{\Sigma}_{1,2} & \widehat{\Sigma}_2 \end{pmatrix}, \quad and$$

$$\mathcal{T}_N = N^2\widetilde{b}^L\left(U_{1_N}, U_{2_N}\right)\widehat{\Sigma}^{-1}\left(U_{1_N}, U_{2_N}\right)'.$$

Suppose our model is correctly specified, so that Eq. (1) is satisfied w.p.1, and Proposition 1 and Theorem 1 hold. Then if Assumptions B1 and B2 are also satisfied,

$$\mathcal{T}_{p_N} \xrightarrow{d} \mathcal{N}(0, 1) \quad \text{for } p = 1, 2, \text{ and } \mathcal{T}_N \xrightarrow{d} \chi_2^2.$$

2. Suppose $\Pr\left(\mu_p(X) \neq F_p(t_p)\big| X \in \mathcal{X}\right) > 0$, so that Eq. (1) is violated with positive probability in the set $\mathcal{X}$. Let θ_p^* denote the probability limit of $\widehat{\theta}_p$. Let $t_p^* = W_p + Z_p'\theta_p^*$ and

$$F_p^*(t_p^*) = \frac{E\left[\mu_p(X)\varphi(X)|t_p^*\right]}{E[\varphi(X)|t_p^*]}.$$

Rule out the case in which $\Pr\left(\mu_p(X) \neq F_p(t_p)\big| X \in \mathcal{X}\right) > 0$, but $\Pr\left(\mu_p(X) = F_p^*(t_p^*)\big| X \in \mathcal{X}\right) = 1$. Also maintain the exclusion restriction conditions in Assumption A1 and rule out a perfect correlation between $Y_1 - F_1^*(t_1^*)$ and $Y_2 - F_2^*(t_2^*)$ conditional on $X \in \mathcal{X}$. Then, the statistic $\mathcal{T}_N$ diverges w.p.1.

For a prespecified size α, we would reject our model if $\mathcal{T}_N \geq c_\alpha$, where $\Pr[\chi_2^2 \geq c_\alpha] = \alpha$. If the model is correct, the size α will be achieved asymptotically. Otherwise if Eq. (1) is violated with positive probability in the set $\mathcal{X}$, our test will always reject the model asymptotically for any size α. If the model is rejected, the test itself is incapable of determining which assumption failed. It could be because: (i) The semiparametric assumptions concerning payoffs and the distributions involved are incorrect, (ii) our informational assumptions are incorrect, (iii) behavior does not correspond to BNE, or (iv) the selection mechanism is not degenerate. It is not difficult to see how our specification test would detect violations to our assumptions in cases (i)–(iii). The case where the selection mechanism does not satisfy our assumptions but all other behavioral assumptions are correct is a bit subtler. To understand what happens in this case, recall from our discussion following Assumption A0 that if multiple BNE exist for X_i, a nondegenerate selection rule would yield

$$E[Y_{1i}|X_i] = \sum_{r=1}^{R(X_i)} \Pr\left[\mathscr{S} = \widetilde{\mu}^r(X_i)\big| X_i\right] \cdot F_1\left(X_{1i}'\beta_1 + \alpha_1\widetilde{\mu}_2^r(X_i)\right),$$

$$E[Y_{2i}|X_i] = \sum_{r=1}^{R(X_i)} \Pr\left[\mathscr{S} = \widetilde{\mu}^r(X_i)\big| X_i\right] \cdot F_2\left(X_{2i}'\beta_2 + \alpha_2\widetilde{\mu}_1^r(X_i)\right),$$

where $\{\widetilde{\mu}^r(X_i)\}_{r=1}^{R(X_i)} \equiv \left\{(\widetilde{\mu}_1^r(X_i), \widetilde{\mu}_2^r(X_i))\right\}_{r=1}^{R(X_i)}$ is the collection of BNE equilibria that corresponds to X_i. In particular, the above equations show that if there exists multiple BNE with positive probability in the set $\mathcal{X}$ and if the equilibrium selection mechanism violates our degeneracy assumptions, using the reparameterization in Section 2.2 we will have $E[Y_{pi}|X_i] \neq F_p(W_{pi} + V_{pi}'\gamma_p + \alpha_p E[Y_{-pi}|X_i])$ for at least one of the players. Therefore if the selection mechanism does not satisfy our degeneracy assumptions, then with positive probability we will have

$$E[Y_{1i}|X_i] \neq E\left[Y_{1i}\big| W_{1i} + V_{1i}'\gamma_1 + \alpha_1 E[Y_{2i}|X_i]\right], \quad \text{and/or}$$

$$E[Y_{2i}|X_i] \neq E\left[Y_{2i}\big| W_{2i} + V_{2i}'\gamma_2 + \alpha_2 E[Y_{1i}|X_i]\right].$$

Our specification test relies on the semiparametrically estimated residuals

$$Y_{1i} - \widehat{E}\left[Y_{1i}\big| W_{1i} + V_{1i}'\gamma_1 + \alpha_1\widehat{E}[Y_{2i}|X_i]\right] \quad \text{and}$$

$$Y_{2i} - \widehat{E}\left[Y_{2i}\big| W_{2i} + V_{2i}'\gamma_2 + \alpha_2\widehat{E}[Y_{1i}|X_i]\right].$$

This stands in contrasts with the general setting in Fan and Li (1996), who use semiparametric residuals that do not include nonparametric regressors. The various conditions we described above concerning the construction of the generated regressors

$\widehat{E}[Y_{pi}|X_i]$ are sufficient to guarantee that our test-statistic is able to capture asymptotically if Eq. (1) is violated with positive probability or, more generally, if with positive probability we have $E[Y_{pi}|X_i] \neq E\left[Y_{pi}\big| W_{pi} + V_{pi}'\gamma_p + \alpha_p E[Y_{-pi}|X_i]\right]$. From our discussion above it follows therefore that our test is capable of detecting not only incorrect parameterizations, incorrect informational assumptions or departures from Nash equilibrium behavior, but also violations to our assumptions about equilibrium selection.

4.1. Some remarks on experimental data set environments

Data generated by experiments allows the researchers full control over important observable components of the model. This knowledge can be used to do consistent inference under weaker versions of our assumptions. To illustrate this consider the hypothetical experimental design of a 2×2 game

		PLAYER 2	
		$Y_2 = 1$	$Y_2 = 0$
PLAYER 1	$Y_1 = 1$	$X_a + X_b + \Delta_1, X_a + X_b + \Delta_2$	$X_a + X_b, 0$
	$Y_1 = 0$	$0, X_a + X_b$	$0, 0$

Both X_a and X_b are randomly drawn by the researcher, and their distributions are such that neither is deterministic conditional on the other (e.g., they could be mutually independent) and Δ_p is a constant chosen by the researcher. Notice that these normal-form payoffs do not have individual-specific shifters, contradicting the exclusion restriction in Assumption A1. In some instances, normal-form payoffs with the above features are appropriate to study behavior in coordination games. Suppose each pair of subjects in the experiment observe the realization of this payoff matrix and simultaneously choose their action. Suppose that the choices observed in the laboratory correspond to a quantal response equilibrium (QRE). As in McKelvey and Palfrey (1995), suppose the payoff shock in the perturbed game is assumed to enter additively. QRE corresponds to the BNE of the augmented or perturbed game. Suppose the design of the experiment is such that the labels $p = 1$ and $p = 2$ are meaningful. Suppose the goal is to test whether this BNE satisfies the conditions in Assumption A0 (including those concerning equilibrium selection). If this is the case Eq. (1) would hold and

$$\mu_p(X) = F_p(X_a + X_b + \Delta_p\mu_{-p}(X)) \equiv F_p(W_p + X_b),$$
$$\text{where } X \equiv (X_a, X_b), \ W_p \equiv X_a + \Delta_p\mu_{-p}(X),$$

$\mu_1(X) = Pr[Y_1 = 1|X]$ and $\mu_2(X) = Pr[Y_2 = 1|X]$. A crucial difference between experimental and non-experimental data sets is that in the former, the researcher *knows the true value of* Δ_p. Suppose we have a sample of N realizations of this experiment, and that the distribution F_p is assumed to be the same for each p across all realizations observed.[17] Under these conditions, our methodology allows consistent testing without making any parametric assumptions about F_p. Suppose we proceed by estimating γ_p in a model parameterized as $\mu_p(X) = F_p(W_p + \gamma_p X_b)$ where, as before, we have normalized the coefficient of one of the payoff covariates $-W_p-$ to 1. Note that since payoffs depend on both X_a and X_b, neither W_p nor X_b are deterministic conditional on $\mu_p(X)$. As before let $\mu_p(X_i) \equiv \mu_{pi}$. By the same arguments as in Proposition 1 and the researcher's precise knowledge of payoffs we know that $E\left[\left((W_{pi} - W_{pj}) + (X_{bi} - X_{bj}) \cdot \widetilde{\gamma}_p\right)^2\big| \mu_{pi} = \mu_{pj}\right]$ is

---

[17] Haile et al. (2008) characterize observable implications of QRE when payoff disturbances are no longer assumed to be identically distributed across observations.

uniquely minimized at $\widetilde{\gamma}_p = \gamma_p = 1$. We will use this to estimate γ_p. Note that W_{pi} is not exactly known to the researcher because the functional forms of $\mu_1(\cdot)$, $\mu_2(\cdot)$ are unknown, but as before they can be consistently estimated under our assumptions. Thus, having exact knowledge of Δ_p enables the researcher to consistently estimate W_{pi} by using $\widehat{W}_{pi} = X_{ai} + \Delta_p\widehat{\mu}_{-p}(X_i)$. This is a very advantageous feature of experimental data sets. Using our approach we estimate

$$\widehat{\gamma}_p = \underset{b}{\mathrm{argmin}} \binom{N}{2}^{-1} \sum_{i<j} K_a\left(\frac{\widehat{\mu}_{pi} - \widehat{\mu}_{pj}}{h_a}\right)$$
$$\times \left((\widehat{W}_{pi} - \widehat{W}_{pj}) + (X_{bi} - X_{bj})\cdot b\right)^2 \phi(X_i)\phi(X_j),$$

which yields

$$\widehat{\gamma}_p = \left[-\binom{N}{2}^{-1}\frac{1}{h_a}\sum_{i<j}K_a\left(\frac{\widehat{\mu}_{pi} - \widehat{\mu}_{pj}}{h_a}\right)\right.$$
$$\times \left. (X_{bi} - X_{bj})^2 \phi(X_i)\phi(X_j)\right]^{-1}$$
$$\times \left[\binom{N}{2}^{-1}\frac{1}{h_a}\sum_{i<j}K_a\left(\frac{\widehat{\mu}_{pi} - \widehat{\mu}_{pj}}{h_a}\right)\right.$$
$$\times \left. (X_{bi} - X_{bj})\left(\widehat{W}_{pi} - \widehat{W}_{pj}\right)\phi(X_i)\phi(X_j)\right].$$

The expression for this estimator is a bit different from (5). Still, its asymptotic distribution[18] can be characterized using the same arguments as those leading to the results in Theorem 1.

Note that the setting considered above is one where all payoff parameters are fixed and known ex-ante in the experimental design, yet we proceed to estimate γ_p (whose true value in the previous formulation is $\gamma_p = 1$) *as if* it were unknown. If we reject the null hypothesis $H_0 : \gamma_p = 1$, we would also reject the behavioral implications of QRE in this simple game. The specification test from Theorem 2 would provide an even more powerful test. Identifying which QRE assumption (BNE beliefs, i.i.d nature of payoff shocks across observations, etc.) is being violated in the data may not be possible in general, but our methodology allows a test for the QRE model which does not require parametric specification of the unobserved payoff perturbations. It can also allow us to test QRE under alternative information structures (i.e., we could test whether agents in the experiment condition their beliefs only on a specific subset of observable covariates).

5. Extensions

At the beginning of Section 2 we claimed that focusing on a 2×2 game was done mainly for illustrative purposes. The approach and methods described in Sections 3–4 can be generalized and adapted straightforwardly to various other settings. We describe some of those extensions here. Section 5.1 studies how to deal with more general payoff parameterizations under the maintained assumption that unobservable shocks enter additively. Section 5.2 describes a way to allow for endogenous payoff covariates. In Section 5.3 we describe how to estimate models with more actions or players as well as asymmetric information. Finally, our approach can be applied to some non-strategic interaction econometric models, we discuss this briefly in Section 5.4.

---

[18] Furthermore, the researcher can choose the distribution of X so that all the assumptions pertaining to it in Theorem 1 are satisfied. Knowing this distribution can also be used to improve the estimation of standard errors.

5.1. More general payoff functional forms

Additive separability in observables and unobservables in players' payoff functions is a crucial assumption for our approach. However, as long as this restriction is satisfied our methods have the potential to be extended to more general payoff specifications. Suppose payoffs are given by

$$U_p(Y_p) = Y_p \times \left\{\pi_p(X_p, Y_{-p}; \theta_p) - \zeta_p\right\}.$$

For a given $\mu \in [0, 1]$ define

$$\mathcal{U}_p(X_p, \mu; \theta_p) = \mu \cdot \pi_p(X_p, 1; \theta_p) + (1 - \mu) \cdot \pi_p(X_p, 0; \theta_p).$$

If we maintain our general assumptions about ζ_p, then for a given x a BNE is any pair $\widetilde{\mu}(x) \equiv (\widetilde{\mu}_1(x), \widetilde{\mu}_2(x))$ in $[0, 1] \times [0, 1]$ that satisfies $\widetilde{\mu}_p(x) = F_p\left(\mathcal{U}_p(x_p, \widetilde{\mu}_{-p}(x); \theta_p)\right)$ for $p = 1, 2$. If our degenerate equilibrium selection assumptions are maintained, we obtain a more general version of Eq. (1). Namely,

$$\mu_p(x) = F_p\left(\mathcal{U}_p(x_p, \mu_{-p}(x); \theta_p)\right)$$
$$\text{for } p = 1, 2, \text{ with } \mu_p(x) = \Pr[Y_p = 1|X = x].$$

Maintain the assumption of an iid sample produced by this model and keep denoting $\mu_p(X_i) \equiv \mu_{pi}$. Our identification and estimation strategy would once again exploit invertibility of F_p, which implies

$$\mu_{pi} = \mu_{pj} \iff \mathcal{U}_p(X_{pi}, \mu_{-pi}; \theta_p) = \mathcal{U}_p(X_{pj}, \mu_{-pj}; \theta_p).$$

Analogously to the reparameterization in Section 2.2, for identification purposes we need to impose conditions on the distribution of X and the parameter space Θ that preclude the existence of an invertible mapping Γ such that $\Gamma\left(\mathcal{U}_p(X_p, \mu_{-p}(X); \theta_p)\right) = \mathcal{U}_p(X_p, \mu_{-p}(X); \widetilde{\theta}_p)$ w.p.1. for some pair $\theta \neq \widetilde{\theta}$ in Θ. In Section 2.2 we saw that when $\mathcal{U}_p(X_p, \mu_{-p}(X); \widetilde{\theta}_p)$ is a linear index, we only need to focus on linear mappings $\Gamma(\psi) = a + b\psi$ and from here, location and scale normalizations suffice (along with full-rank, exclusion and support conditions). The type of normalizations that would be required in alternative settings would depend on the specific payoff parameterization used for $\pi_p(\cdot)$. Once the parameter space satisfies this, we would have to impose any additional assumption needed to ensure that

$$\Pr\left[\mathcal{U}_p(X_{pi}, \mu_{-pi}; \widetilde{\theta}_p) \neq \mathcal{U}_p(X_{pj}, \mu_{-pj}; \widetilde{\theta}_p)\,\big|\,\mu_{pi} = \mu_{pj}\right] > 0$$
$$\forall \widetilde{\theta} \in \Theta : \widetilde{\theta} \neq \theta \text{ (true parameter value)}.$$

In the case of linear-index payoffs, once X_p excludes a constant and a scale normalization in performed in Θ, the exclusion restrictions in Assumption A1 and the full-rank condition on X_p were sufficient for the above condition to hold. Analogous conditions for non-linear payoffs would depend on the specific functional form and parameterization used. Once all these conditions are satisfied we would be able to show that $E\left[\left(\mathcal{U}_p(X_{pi}, \mu_{-pi}; b) - \mathcal{U}_p(X_{pj}, \mu_{-pj}; b)\right)^2\,\big|\,\mu_{pi} = \mu_{pj}\right]$ is uniquely minimized at $b = \theta_p$. Accordingly, our proposed estimator would be of the form

$$\widehat{\theta}_p = \underset{b}{\mathrm{argmin}} \sum_{i<j} K_a\left(\frac{\widehat{\mu}_{pi} - \widehat{\mu}_{pj}}{h_a}\right)$$
$$\times \left(\mathcal{U}_p(X_{pi}, \widehat{\mu}_{-pi}; b) - \mathcal{U}_p(X_{pj}, \widehat{\mu}_{-pj}; b)\right)^2 \phi(X_i)\phi(X_j).$$

In general $\widehat{\theta}_p$ may no longer have a closed-form expression. As in the case we studied here, conditions concerning the way in which the unknown functions involved are estimated and their resulting asymptotic features, as well as smoothness assumptions involving unknown functionals would determine the conditions under which the estimator $\widehat{\theta}_p$ is $\sqrt{N}$-consistent and asymptotically normal.

5.2. Conditional invertibility

Here we describe how, under certain conditions, our results can be extended to settings where there exists endogenous payoff covariates. Suppose we replace Eq. (1) with the following more general version

$$\Pr\left[Y_p = 1 | X = x\right] \equiv \mu_p(x) = F_p\left(x'_p\beta_p + \alpha_p\mu_{-p}(x); \delta_p(x)\right),$$

such that for any fixed $\delta \in \mathbb{S}(\delta_p(X))$ the transformation $F_p(\cdot; \delta)$ is strictly increasing. That is, for any given δ we have $F_p(a; \delta) = F_p(b; \delta) \iff a = b$. This could correspond to a case of endogeneity which is entirely captured (or controlled) by δ_p. For the ith observation in our sample let $\delta_p(X_i) \equiv \delta_{pi}$. Suppose δ_{pi} can be expressed as $\delta_{pi} = E[\xi_{pi}|X_i]$ for some observable $\xi_{pi} \in \mathbb{R}^\ell$ and we have an estimator

$$\widehat{\delta}_{pi} = \frac{1}{Nh_b^L}\sum_{j=1}^{N}\xi_{pj}K_b\left(\frac{X_j - X_i}{h_b}\right), \quad \text{or}$$

$$\widehat{\delta}_{pi} = \xi_{pi} \quad \text{if } E[\xi_{pi}|X_i] = \xi_{pi},$$

with a linear representation of the type

$$\widehat{\delta}_{pi} - \delta_{pi} = \frac{1}{Nh_b^L}\sum_{j=1}^{N}\frac{[\xi_{pj} - \delta_{pi}]}{f_X(X_i)}K_b\left(\frac{X_j - X_i}{h_b}\right) + \upsilon_N(X_i),$$

$$\sup_{x\in\mathcal{X}}\|\upsilon_N(x)\| = O_p(N^{\epsilon-1}h_b^{-L})$$

for any $\epsilon > 0$. The set $\mathcal{X}$ is the trimming set used previously. Define

$$Y_{pi}^* \equiv (Y_{pi}, \xi_{pi}), \qquad \mu_{pi}^* \equiv (\mu_{pi}, \delta_{pi}), \qquad \widehat{\mu}_{pi}^* \equiv (\widehat{\mu}_{pi}, \widehat{\delta}_{pi}).$$

Now let Z_{pi}^* be the subset of elements in Z_{pi} which are not deterministic conditional on μ_{pi}^*. Let θ_p^* be the subset of elements in θ_p that correspond to Z_p^*. This is the subset of parameters that can be identified under our assumptions. We propose an estimator for θ_p^* of the form

$$\widehat{\theta}_p^* = \underset{b}{\mathrm{argmin}} \binom{N}{2}^{-1}\sum_{i<j}K_a\left(\frac{\widehat{\mu}_{1i}^* - \widehat{\mu}_{1j}^*}{h_a}\right)\Big[(W_{1i} - W_{1j})$$
$$+ \left(\widehat{Z}_{1i}^* - \widehat{Z}_{1j}^*\right)' b\Big]^2 \phi(X_i)\phi(X_j).$$

Maintain Assumption A1 and consider the following modifications to Assumptions A0 and A2–A5.

Assumption A0*. Maintain Assumption A0 replacing the invertibility of $F_p(\cdot)$ with the conditional invertibility described above.

Assumption A2*. Abbreviate $t_p \equiv x'_p\beta_p + \alpha_p\mu_{-p}$. Modify Assumption A2 to hold for $F_p(t_p; \delta_p)$ for its $\ell+1$ arguments. Replace $\overline{M}$ with $\overline{M}^*$, which will be characterized in Assumption A4*, below.

Assumption A3*. For a given δ, let $F_p^-(\cdot; \delta)$ denote the inverse function of $F_p(\cdot; \delta)$. That is, $F_p(a; \delta) = b \Leftrightarrow F^-(b; \delta) = a$. Denote $F_p^-(\mu_{pi}; \delta_{pi}) \equiv F_p^-(\mu_{pi}^*)$. Let $\nabla_{\mu^*}F_p^-(\mu_p^*) \in \mathbb{R}^{\ell+1}$ denote its Jacobian. We will assume that $\|\nabla_{\mu^*}F_p^-(\mu_p^*)\|$ is uniformly bounded in $\mathcal{X}$. Replace all the statements in Assumption A3 concerning expectations conditional on μ_p with the same statements conditional on μ_p^*.

Assumption A4*. The kernels $K_a : \mathbb{R}^{\ell+1} \to \mathbb{R}$ and $K_b : \mathbb{R}^L \to \mathbb{R}$ are bias-reducing of order $\overline{M}^*$. Each one of the $\ell+1$ components of the Jacobian vector $\nabla K_a(\cdot)$ satisfies the conditions described in Assumption A4 for some $\overline{d}^*$. Modify (A4.iii) and assume now that there exists $\overline{\delta}^* > 0$ such that $N^{1/2-\overline{\delta}^*}h_b^L h_a^{2+\ell} \to \infty$. The constants $\overline{d}^*$ and $\overline{M}^*$ satisfy $N^{1/2}h_a^{\overline{d}^*} \to 0$, $N^{1/2}h_b^{\overline{M}^*} \to 0$, $N^{1/2}h_b^{\overline{M}^*}/h_a^{2+\ell} \to 0$.

Assumption A5*. Let $f_{\mu_p^*}(\mu_p^*)$ denote the density of μ_p^*. This density satisfies the smoothness conditions described in Assumptions A4–A5 for $f_{\mu_p}(\mu_p)$ with $\overline{M}$ replaced with $\overline{M}^*$. Define

$$D_p^* \equiv E\left[\left(E[Z_p^*Z_p^{*\prime}\phi(X)|\mu_p^*]\cdot E[\phi(X)|\mu_p^*]\right.\right.$$
$$\left.\left. - E[Z_p^*\phi(X)|\mu_p^*]E[Z_p^*\phi(X)|\mu_p^*]'\right)f_{\mu_p^*}(\mu_p^*)\right].$$

Then D_p^* has full rank for $p = 1, 2$.

Theorem 3. Suppose A0*, A1*, A1 and A2*– A5* hold. Let

$$\tau_{pi}^* = \left(Z_{pi}^*E\left[\phi(X)|\mu_{pi}^*\right] - E\left[Z_p^*\phi(X)|\mu_{pi}^*\right]\right)$$
$$\times \left(Y_{-pi} - \mu_{-pi}\right)\phi(X_i)$$
$$\upsilon_{pi}^* = \left(Z_{pi}^*E\left[\phi(X)|\mu_{pi}^*\right] - E\left[Z_p^*\phi(X)|\mu_{pi}^*\right]\right) \qquad (12)$$
$$\times f_{\mu_p^*}(\mu_{pi}^*)\nabla_{\mu_p^*}F_p^-(\mu_p^*)'\left(Y_{pi}^* - \mu_{pi}^*\right)\phi(X_i)$$
$$\psi_i^{\theta_p^*} = (D_p^*)^{-1}\times\left(\tau_{pi}^*\alpha_p + \upsilon_{pi}^*\right); \qquad \lambda_i^{\theta_p^*} = (D_p^*)^{-1}\times\upsilon_{pi}^*.$$

Then,

$$(A) - \text{If } \mu_{-p}^* \in Z_p^* : \widehat{\theta}_p^* - \theta_p^* = \frac{1}{N}\sum_{i=1}^{N}\psi_i^{\theta_p^*} + o_p(N^{-1/2}),$$
$$(13)$$
$$(B) - \text{If } \mu_{-p}^* \notin Z_p^* : \widehat{\theta}_p^* - \theta_p^* = \frac{1}{N}\sum_{i=1}^{N}\lambda_i^{\theta_p^*} + o_p(N^{-1/2}).$$

Theorem 3 is a more general version of Theorem 1. Due to their similarity, we omit its proof in the Appendix to save space. Examples of control functions that could arise in the context of endogeneity can be found for example in Blundell and Powell (2004) and Imbens and Newey (2002).

5.3. Games with more actions or players and asymmetric information

Suppose now we have $p = 1, \ldots, P$ players, each of which has $\ell = 1, \ldots L_p$ possible choices for his action (e.g., consider a multiple entry model). Let $Y_{p\ell}$ denote the indicator function that the ℓth choice was selected for player p's action. Let $\mathcal{Y}_{-p\ell}$ be a multinomial real-valued random variable indexing the set of all combinations of possible actions that can be chosen by player p's opponents if he chooses the ℓth action. Let $S_{p\ell}$ denote the vector of signals upon which player p conditions his beliefs about the distribution of $\mathcal{Y}_{-p\ell}$. If $S_{p\ell}$ is observable, we can extend our methods to this more general setting. Define

$$t_{p\ell} = W_{p\ell} + V'_{p\ell}\beta_{p\ell} + \alpha'_{-p\ell}\pi_{-p\ell}(S_{p\ell})$$
$$\equiv W_{p\ell} + Z'_{p\ell}\theta_{p\ell}, \quad \ell = 1, \ldots, L_p - 1,$$

where $\pi_{-p\ell}(S_{p\ell})$ is the pdf of $\mathcal{Y}_{-p\ell}|S_{p\ell}$. Implicit in this formulation is the assumption that there exists a payoff shifter with nonzero coefficient, $W_{p\ell}$ for each $\ell = 1, \ldots, L$ whose slope is normalized to one. This is a multiple-index model scale normalization needed for identification. We normalize the expected utility of choosing the $L_{p\ell}$th action to zero, which amounts to fixing $t_{p_{L_p}} = 0$. Define $X_{p\ell} \equiv (W_{p\ell}, V_{p\ell}, S_{p\ell})$, $X_p = (X_{p_1}, \ldots, X_{p_{L_p-1}})$, $\underline{t}_p \equiv (t_{p_1}, \ldots, t_{p_{L_p-1}})$, $\mu_{p\ell} \equiv \Pr(Y_{p\ell} = 1|X_p)$, and $\underline{\mu}_p \equiv (\mu_{p_1}, \ldots, \mu_{p_{L_p-1}})$. We maintain the degeneracy property of equilibrium selection, leading to

$$\underline{\mu}_p = F_p(\underline{t}_p).$$

Suppose we generalize our assumptions about F_p in the 2×2 case so that for any pair of observations i, j of player p,

$$\underset{\sim}{\mu}_{p_i} = \underset{\sim}{\mu}_{p_j} \iff \underset{\sim}{t}_{p_i} = \underset{\sim}{t}_{p_j}.$$

This type of invertible relation between choice probabilities and expected utilities is achieved, for instance, in Bajari et al. (2010) by assuming stochastic independence across actions and agents of payoff shocks. In their case, the distribution of payoff shocks is assumed known to the researcher, and therefore the functional form of this invertible mapping is also known (see Assumption A1 and Eq. (10) in Bajari et al. (2010)). Let $Z^*_{p_\ell} \in \mathbb{R}^{k_{p_\ell}}$ be the elements of Z_{p_ℓ} that are not deterministic conditional on $\underset{\sim}{\mu}_p$ and denote their coefficients by $\theta^*_{p_\ell}$. Let $k_p \equiv \sum_{\ell=1}^{L_p-1} k_{p_\ell}$ and

$$t^*_{p_\ell} = W_{p_\ell} + \theta^{*\prime}_{p_\ell} Z^*_{p_\ell},$$

$$\underset{k_p \times (L_p-1)}{Z^*_p} \equiv \begin{pmatrix} Z^*_{p_1} & 0 & \cdots & 0 \\ 0 & Z^*_{p_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & Z^*_{p_{L_p-1}} \end{pmatrix},$$

$$\underset{(L_p-1) \times 1}{\underset{\sim}{W}_p} \equiv \begin{pmatrix} W_{p_1} \\ \vdots \\ W_{p_{L_p-1}} \end{pmatrix}.$$

Extending our approach to this setting would lead to an estimator of the form

$$\widehat{\theta}^*_p = \left[-\binom{N}{2}^{-1} \frac{1}{h_a^{L_p-1}} \sum_{i<j} K_a \left(\frac{\widehat{\mu}_{pi} - \widehat{\mu}_{pj}}{h_a} \right) \right.$$

$$\times \left. \left(\widehat{Z^*_{pi}} - \widehat{Z^*_{pj}} \right) \left(\widehat{Z^*_{pi}} - \widehat{Z^*_{pj}} \right)' \phi(X_i)\phi(X_j) \right]^{-1}$$

$$\times \left[\binom{N}{2}^{-1} \frac{1}{h_a^{L_p-1}} \sum_{i<j} K_a \left(\frac{\widehat{\mu}_{pi} - \widehat{\mu}_{pj}}{h_a} \right) \right.$$

$$\times \left. \left(\widehat{Z^*_{pi}} - \widehat{Z^*_{pj}} \right) \left(\underset{\sim}{W}_{pi} - \underset{\sim}{W}_{pj} \right) \phi(X_i)\phi(X_j) \right],$$

which is a generalized version of the estimator studied here. Its asymptotic properties can be analyzed using the same type of conditions and results as in our previous sections. Applying our approach to a larger game will have computational implications, but they are not nearly as severe as the ones that would arise in a procedure that requires the computation of equilibria. Perhaps the most important computational cost would involve the order of the kernel needed to achieve $\sqrt{N}$-consistency of our estimator. This order would increase with the dimension of the (continuously distributed) signals used by players to construct their beliefs. In a game with multiple players, we would expect a larger signal vector and hence a more complicated set of kernels. We also conjecture that in larger games, the finite-sample properties of our estimator would be more sensitive to the choice of bandwidths and kernels.

5.4. Extensions to non-strategic interaction models

Our methodology can be applied to models amenable to pairwise-differencing estimation with nonparametric regressors. Take for instance the transformation model studied in Han (1987), where a binary variable $Y \in \{0, 1\}$ and a covariate vector X are assumed to satisfy $E[Y|X] = G(X'\beta)$ for some unknown transformation $G(\cdot)$ that is assumed to be strictly increasing everywhere in the real line. Once again, β can be identified

up to a proportionality scale, and no intercept can be identified. Once the parameter space has been properly normalized, the maximum rank correlation estimator (MRC) $\widehat{\beta}$ suggested in Han (1987) exploits this monotonicity condition by maximizing the objective function $\sum_{i<j} \mathbb{1}\{Y_i > Y_j\}\mathbb{1}\{X'_i\beta > X'_j\beta\}$. Now suppose one of the regressors is a nonparametric function (e.g., as in the discrete choice model studied in Ahn and Manski (1993)). Specifically suppose $E[Y|X] = G(W'\beta + E[V|Z] \cdot \gamma)$, where $E[V|Z]$ is of unknown functional form, but nonparametrically identified. Let $\theta \equiv (\beta, \gamma)$ and suppose the parameter space Θ reflects the proper location and scale normalizations. Estimating θ by plugging this generated regressor into the MRC objective function may be unattractive for applied researchers interested in doing inference in this model since the asymptotic distribution of

$$\binom{N}{2}^{-1} \sum_{i<j} \mathbb{1}\{Y_i > Y_j\}$$

$$\times \left[\mathbb{1}\left\{ W'_i\beta + \widehat{E}[V_i|Z_i] \cdot \gamma > W'_j\beta + \widehat{E}[V_j|Z_j] \cdot \gamma \right\} \right.$$

$$\left. - \mathbb{1}\left\{ W'_i\beta + E[V_i|Z_i] \cdot \gamma > W'_j\beta + E[V_j|Z_j] \cdot \gamma \right\} \right]$$

can be difficult to characterize due to the discontinuous nature of this objective function (see Chen et al. (2003)). Our estimation procedure can be an attractive alternative. We would have

$$\widehat{\theta} = \underset{b_1, b_2}{\operatorname{argmax}} \sum_{i<j} \left((W_i - W_j)'b_1 \right.$$

$$\left. + \left(\widehat{E}[V_i|Z_i] - \widehat{E}[V_j|Z_j] \right) \cdot b_2 \right)^2 \phi(Z_i)\phi(Z_j).$$

As we can see, this is a straightforward extension of our estimator. The results and conditions in this paper can be readily applied to the above model and provide researchers with a feasible way to do inference. In addition to this example, other non-strategic models that can be studied by extending our results include all examples examined in AHP where the regressors include nonparametric functions. As we mentioned before, the asymptotic theory in AHP does not contemplate such cases. Our paper constitutes the first instance of a pairwise-difference estimator with nonparametric control variables *and* nonparametric regressors.

6. A Monte Carlo study

This section is directed to applied researchers. Our goals are to: (i) Assess the accuracy of our asymptotic results in finite samples. (ii) Compare our procedure with a parametric one where $F_p(\cdot)$ is assumed to be of known functional form. (iii) Evaluate the performance of our methods when the true distribution F_p is invertible but relatively flat. We focus on the 2×2 game with payoffs

$$Y_p \times \left(W_p + V_p\beta_p + \alpha_p Y_{-p} - \zeta_p \right),$$

W_p and V_p mutually independent, standard normal,

(W_p, V_p) independent of ζ_p. $\beta_1 = \beta_2 = -0.5$.

The distributional properties of ζ_p and the values of α_p will vary in the various experimental designs we analyze here. The above features will be maintained.

6.1. Experiment designs

As before, denote $X_p \equiv (W_p, V_p)$ and $X \equiv (X_1, X_2)$. The realization of X and the true values of the parameters are known to both players, but ζ_p is private information. ζ_1, ζ_2 are mutually independent and behavior conforms with BNE. We use three designs.

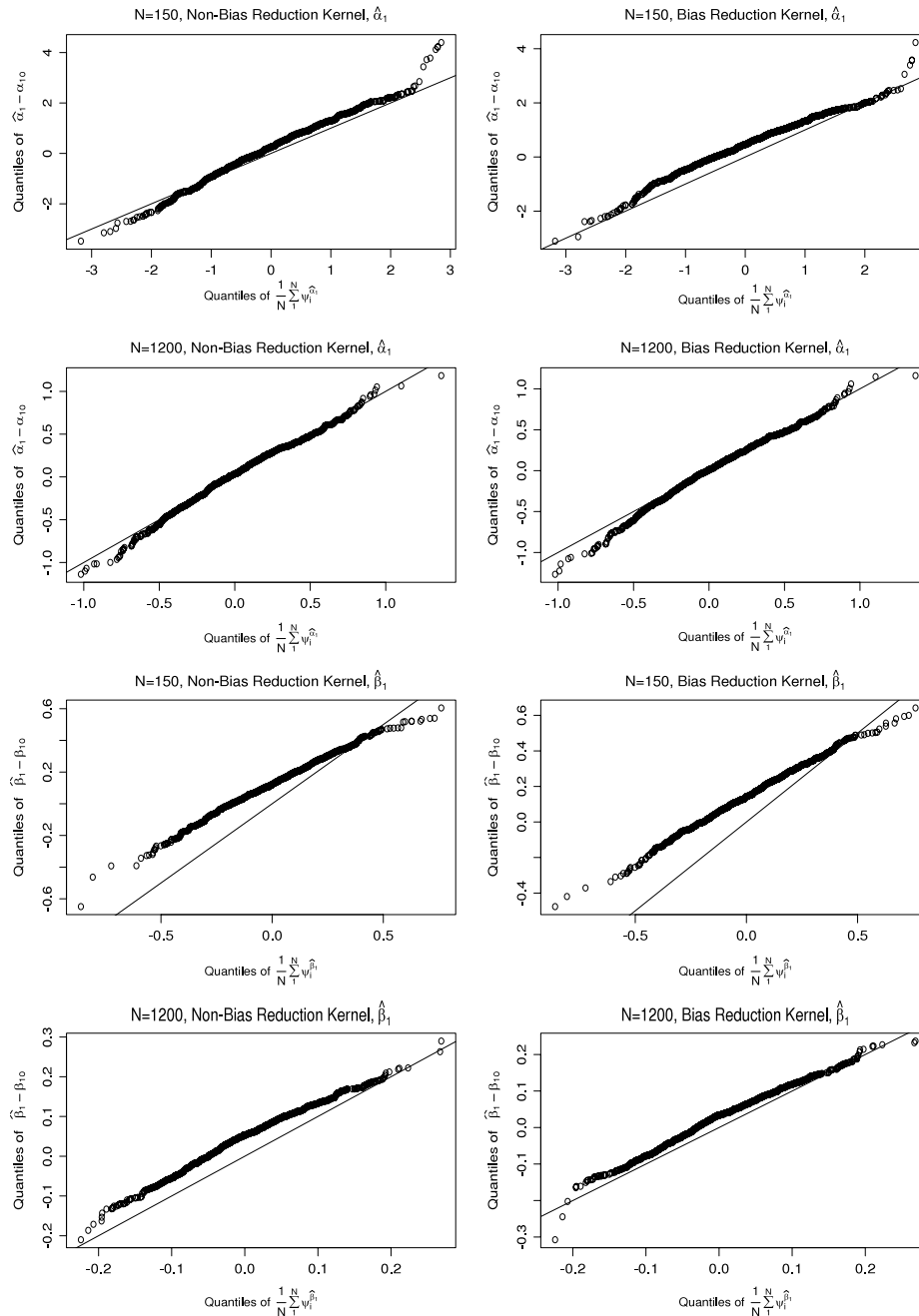*Design 1A.* $\zeta_p \sim$ logistic, $\alpha_1 = \alpha_2 = -1$.

Fig. 1. *Design* 1A. Quantile–quantile comparison between the pairwise-difference estimators and the asymptotic representation $\frac{1}{N}\sum_{i=1}^{N}\psi^{\widehat{\theta}}$ in Theorem 1. Bandwidth constants and rates used were $\mathcal{C}_{h_a} = 0.39$, $\mathcal{C}_{h_b} = 2.37$, $\lambda_{h_a} = \lambda_{h_b} = 1/5$ for non-bias reducing kernels, and $\mathcal{C}_{h_a} = 0.39$, $\mathcal{C}_{h_b} = 2.28$, $\lambda_{h_a} = 127/2000$, $\lambda_{h_b} = 127/1600$ for the bias-reducing case. Bias reducing kernels K_a, K_b were of order six.

Design 1B. $\zeta_p = \varepsilon_p + u_p$, $\varepsilon_p \perp u_p$, $\varepsilon_p \sim \mathcal{N}(0, 1)$, $u_p \sim U[0, 1]$, $\alpha_1 = \alpha_2 = -1$.

Design 1C. Same as 1B, with $\alpha_1 = \alpha_2 = -3$.

Designs 1A and 1B have unique BNE equilibrium with probability one due to the magnitude of $|\alpha_p|$. For the case of 1C, whenever multiple equilibria was observed, a degenerate selection mechanism was used where the solution closest to $(0, 0)$ was chosen. Coupled with these equilibrium and selection mechanism properties, all versions of Design 1 satisfy the behavioral assumptions of this paper. We refer to the shocks in designs 1B-C as *skewed*. Even though the distribution of unobservable shocks is invertible in all our designs, it is easy to verify that the skewed distribution is noticeably flatter than

its logistic counterpart in the tails. As a result, the invertibility property exploited by our methods is more tenuous in design 1A relative to 1B and 1C. One of our goals here is to explore the impact of this in the finite-sample properties of our estimator and test-statistic. The kernels and bandwidths used are discussed in the appendix. (See Fig. 1.)

6.2. Estimator performance

Unless we explicitly state otherwise, in this section we use $\theta_0 = (\theta_{1_0}, \theta_{2_0})$ to denote the true payoff parameter values in the various designs. Unless stated otherwise, all tables and figures referenced here can be found in Appendix B.

Table B.1
Simulation results for design 1A.

N	Kernels used	RMSE	$|\widehat{\theta} - \theta_0|_{0.25}$	$|\widehat{\theta} - \theta_0|_{0.50}$	$|\widehat{\theta} - \theta_0|_{0.75}$	$(\widehat{\theta}_{0.025}, \widehat{\theta}_{0.975})$
$\widehat{\beta}_1$						
150	Non-bias red.	0.2080	0.0705	0.1503	0.2486	$(-0.7251, -0.0698)$
	Bias reduc.	0.2177	0.0756	0.1561	0.2629	$(-0.7034, -0.0413)$
600	Non-bias red.	0.1189	0.0416	0.0865	0.1406	$(-0.6344, -0.2469)$
	Bias reduc.	0.1185	0.0391	0.0846	0.1368	$(-0.6592, -0.2485)$
1200	Non-bias red.	0.0873	0.0318	0.0624	0.1034	$(-0.6078, -0.3174)$
	Bias reduc.	0.0812	0.0298	0.0574	0.0931	$(-0.6302, -0.3293)$
$\widehat{\alpha}_1$						
150	Non-bias red.	1.1842	0.3898	0.8106	1.3617	$(-3.1721, 1.2181)$
	Bias reduc.	1.0295	0.3397	0.7137	1.1814	$(-2.6750, 1.0136)$
600	Non-bias red.	0.5576	0.1879	0.3764	0.6396	$(-2.0621, 0.1173)$
	Bias reduc.	0.5476	0.1672	0.3579	0.6411	$(-2.0510, 0.0878)$
1200	Non-bias red.	0.3721	0.1224	0.2652	0.4255	$(-1.7113, -0.2876)$
	Bias reduc.	0.3744	0.1242	0.2613	0.4287	$(-1.7487, -0.3156)$

Bandwidth constants and rates used were $\mathcal{C}_{h_a} = 0.39$, $\mathcal{C}_{h_b} = 2.37$, $\lambda_{h_a} = \lambda_{h_b} = 1/5$ for non-bias reducing kernels, and $\mathcal{C}_{h_a} = 0.39$, $\mathcal{C}_{h_b} = 2.28$, $\lambda_{h_a} = 127/2000$, $\lambda_{h_b} = 127/1600$ for bias-reducing kernels (of order six). 1000 simulations in every case.

6.2.1. Finite-sample, asymptotic properties and the use of bias reducing kernels

The use of bias-reducing kernels may have computational implications, in particular for games with multiple players or actions (see Section 5.3). Furthermore, even in 2×2 models, the order of the kernels used in Theorem 1 increase with the dimensionality of observable covariates. A question of interest is how sensitive our results are to the use of higher order kernels. Our results (see Tables B.1, B.2 and B.4) suggest that the use of bias-reducing kernels is more critical for the precision of our results if the distribution of payoff shocks is relatively flat over a large range of values. Conversely, if each F_p is strongly monotonic over the majority of its support, the use of regular kernels has a relatively minor impact in our results.

6.2.2. Pairwise-differencing vs. logistic MLE

We compared a (normalized) MLE estimator $\widehat{\theta}_p^{ML}$ assuming logistic payoff shocks against our estimator. If the logistic parameterization is correct (see Table B.3), $\widehat{\theta}_p^{ML}$ is asymptotically efficient, but our estimator fares comparatively well. On the other hand, if the logistic specification is incorrect (see Fig. 2), our estimator performs considerably better than $\widehat{\theta}_p^{ML}$, especially in the case of the strategic interaction estimator $\widehat{\alpha}_p^{ML}$. The comparative advantage of our procedure becomes more evident in settings where the strategic interaction effect (i.e., the absolute magnitude of $\alpha_1 \cdot \alpha_2$) is stronger.

6.3. Specification test statistic $\mathcal{T}_N$

As above, here we wanted to investigate the extent to which the χ_2^2 asymptotic approximation of $\mathcal{T}_N$ differed from its finite sample properties if: (i) Non-bias reducing kernels were used, and (ii) if F_p is relatively flat. As we discussed above, invertibility of F_p is an increasingly weaker feature as we move from designs 1A to 1B and 1C. The asymptotic χ_2^2 approximation appears to be remarkably good when invertibility is a strong feature of the model such as in Design 1A. As Fig. 3 suggests, the use of bias-reducing kernels was not crucial to preserve this feature. On the other hand, if the true functions F_p are invertible but relatively flat as in Designs 1B and 1C, the asymptotic χ_2^2 approximation is less accurate for finite samples when non-bias reducing kernels are employed. In particular as Table B.5 shows, the empirical size in Designs 1B and 1C was greater than the nominal size asymptotic size. The discrepancy was greater for 1C, the design with the weakest monotonicity features of all.

7. Concluding remarks

We described a two-step estimation procedure for static game-theoretic models with incomplete information that relies only on invertibility properties of the distribution of unobserved payoff shocks and a degenerate equilibrium selection mechanism that is otherwise unspecified. The general semiparametric and behavioral assumptions of our model are compatible with those made in existing work in the econometrics of static games. Unlike existing two-step procedures, the methods analyzed here are the first ones in the literature that make no parametric assumption at all on the distribution of unobservable shocks. Section 4 of Aradillas-Lopez (2010) studies the estimation of static incomplete information games when the distribution of unobserved payoff shocks is unknown. However, the methods suggested there are unattractive to applied researchers because they are computationally demanding even for simple games. This paper contributes by introducing a two-step pairwise differencing procedure that is very simple to implement and is capable of being adapted to a wide variety of extensions of the original model. The paper contributes to the pairwise-differencing estimation literature by examining for the first time a model with nonparametric control functions as well as nonparametric or "generated" regressors. In particular, we showed that the asymptotic theory results in Aradillas-Lopez et al. (2007) (econometrically speaking, the closest to our paper in the literature) are not valid here. A consistent specification test based on semiparametric residuals. To our knowledge, this appears to be the first instance of such a test in a model that includes nonparametric regressors. We described various extensions of our methods to other static models, with and without strategic interaction. We included a Monte Carlo experiment section aimed at comparing the finite sample performance of our estimator and specification tests vis-a-vis their asymptotic approximations, as well as the robustness of our results. That section also described a particular methodology to select the various bandwidths and kernels involved in our procedure. The class of strategic interaction models analyzed here have been used empirically to analyze an increasing variety of decision-making problems, including entry/exit, labor participation, firm coordination, social interaction, recommendations by stock analysts, quantal-response equilibria in experimental data sets as well as applications in the political science literature. We hope that applied researchers will find our methodology to be a computationally convenient approach to do robust inference in these types of models.

The maintained assumption that payoffs were correctly specified was key throughout our approach. As Ponomareva and Tamer
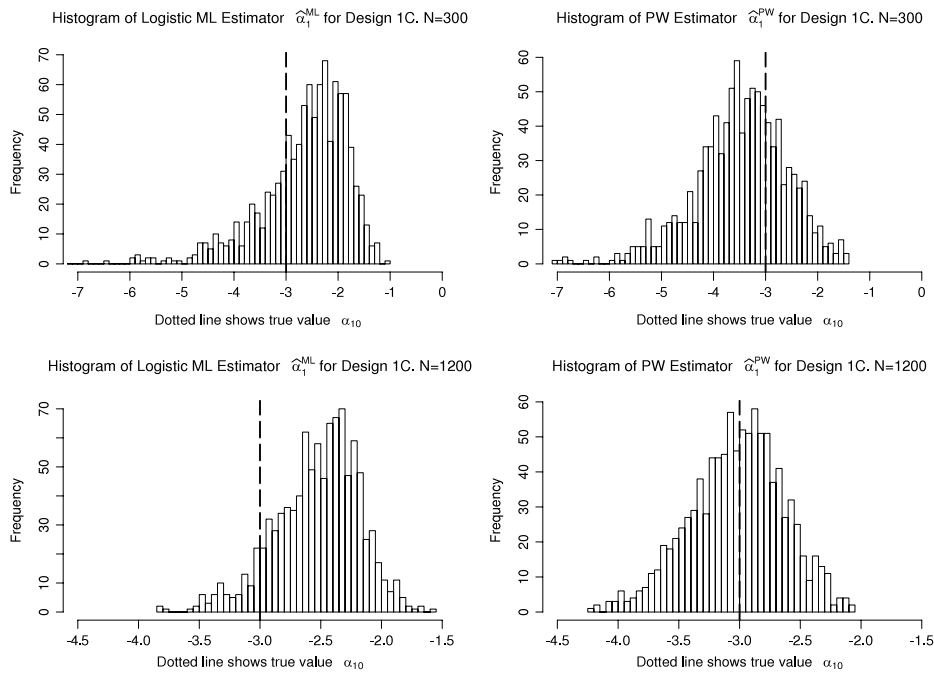
Fig. 2. MLE vs. Pairwise-differencing (non-bias reducing kernels) in Design 1C.

Table B.2
Comparison with asymptotic influence function. Design 1A.

N		Std. dev	Median	25%-quantile	75%-quantile
150	$\frac{1}{N}\sum_{i=1}^{N}\psi_i^{\widehat{\beta}_1}$	0.2318	0.0034	−0.1486	0.1587
	$(\widehat{\beta}_1 - \beta_{10})^{\text{Non-bias red}}$	0.1668	0.1240	0.0215	0.2351
	$(\widehat{\beta}_1 - \beta_{10})^{\text{Biasreduc}}$	0.1638	0.1432	0.0376	0.2541
	$\frac{1}{N}\sum_{i=1}^{N}\psi_i^{\widehat{\alpha}_1}$	1.0143	−0.0409	−0.7250	0.6682
	$(\widehat{\alpha}_1 - \alpha_{10})^{\text{Non-bias red}}$	1.1705	0.2054	−0.5815	0.4661
	$(\widehat{\alpha}_1 - \alpha_{10})^{\text{Biasreduc}}$	0.9500	0.4344	−0.1988	1.0381
600	$\frac{1}{N}\sum_{i=1}^{N}\psi_i^{\widehat{\beta}_1}$	0.1158	0.0015	−0.0759	0.0788
	$(\widehat{\beta}_1 - \beta_{10})^{\text{Non-bias red}}$	0.0961	0.0711	0.0104	0.1352
	$(\widehat{\beta}_1 - \beta_{10})^{\text{Biasreduc}}$	0.1031	0.0629	0.0046	0.1259
	$\frac{1}{N}\sum_{i=1}^{N}\psi_i^{\widehat{\alpha}_1}$	0.4945	−0.0124	−0.3265	0.3248
	$(\widehat{\alpha}_1 - \alpha_{10})^{\text{Non-bias red}}$	0.5564	0.0534	−0.3438	0.4045
	$(\widehat{\alpha}_1 - \alpha_{10})^{\text{Biasreduc}}$	0.5451	0.0424	−0.2841	0.4296
1200	$\frac{1}{N}\sum_{i=1}^{N}\psi_i^{\widehat{\beta}_1}$	0.0796	−0.0008	−0.0501	0.0529
	$(\widehat{\beta}_1 - \beta_{10})^{\text{Non-bias red}}$	0.0733	0.0522	−0.0018	0.0996
	$(\widehat{\beta}_1 - \beta_{10})^{\text{Biasreduc}}$	0.0766	0.0327	−0.0243	0.0779
	$\frac{1}{N}\sum_{i=1}^{N}\psi_i^{\widehat{\alpha}_1}$	0.3580	−0.0004	−0.2483	0.2504
	$(\widehat{\alpha}_1 - \alpha_{10})^{\text{Non-bias red}}$	0.3716	0.0309	−0.2466	0.2857
	$(\widehat{\alpha}_1 - \alpha_{10})^{\text{Biasreduc}}$	0.3842	0.0669	−0.2645	0.2538

Bandwidths as described in Table B.1. 1000 simulations in every case.

(2011) show (see Section 3.3 there), payoff misspecification in simple 2 × 2 games can have important consequences even in partially identified models. There is evidence of a tradeoff between a nonparametric specification for: (i) the distribution of shocks, and (ii) normal-form payoffs. Except in experimental-like settings where payoff parameterizations are set and known ex-ante by the researcher, it is not immediately obvious which robustness direction is better. This question would depend on the specific problem at hand. Finally, while we have shown that estimation and inference can be performed under our semiparametric conditions, we note that performing counterfactual analysis cannot be done

without additional assumptions. In general, while our payoff estimates are robust to misspecification of payoff distributions, counterfactual analysis would involve plugging these estimates into a specific payoff distribution. Other types of pairwise-differencing procedures have been used to examine dynamic choice models in non-strategic interaction settings (Hong and Shum, 2010). An interesting area for future research is the extent to which a variation of our methodology can be used to estimate dynamic models of strategic interaction and investigate the extent to which the main robustness features of our inferential procedures in static games can be preserved in those settings.

Table B.3
Simulation results for Design 1A. Non-bias reducing kernels.

n	$\widehat{\theta}$	RMSE	Bias	$|\widehat{\theta} - \theta_0|_{0.25}$	$|\widehat{\theta} - \theta_0|_{0.50}$	$|\widehat{\theta} - \theta_0|_{0.75}$	$(\widehat{\theta}_{0.025}, \widehat{\theta}_{0.975})$
150	$\widehat{\beta}_1^{ML}$	0.2232	−0.0141	0.0696	0.1426	0.2383	(−1.0171, −0.1449)
	$\widehat{\alpha}_1^{ML}$	1.1054	−0.0799	0.2792	0.5908	1.1045	(−3.6417, 0.6338)
	$\widehat{\beta}_1^{PW}$	0.2080	0.1241	0.0705	0.1503	0.2486	(−0.7251, −0.0698)
	$\widehat{\alpha}_1^{PW}$	1.1842	0.1790	0.3898	0.8106	1.3617	(−3.1721, 1.2181)
600	$\widehat{\beta}_1^{ML}$	0.1024	−0.0021	0.0312	0.0677	0.1198	(−0.7088, −0.3173)
	$\widehat{\alpha}_1^{ML}$	0.4748	−0.0218	0.1478	0.3057	0.5417	(−2.0177, −0.1472)
	$\widehat{\beta}_1^{PW}$	0.1189	0.0701	0.0416	0.0865	0.1406	(−0.6344, −0.2469)
	$\widehat{\alpha}_1^{PW}$	0.5576	0.0365	0.1879	0.3764	0.6396	(−2.0621, 0.1173)
1200	$\widehat{\beta}_1^{ML}$	0.0695	−0.0009	0.0221	0.0454	0.0785	(−0.6511, −0.3732)
	$\widehat{\alpha}_1^{ML}$	0.3184	−0.0121	0.0995	0.2245	0.3708	(−1.6693, −0.4300)
	$\widehat{\beta}_1^{PW}$	0.0873	0.0474	0.0318	0.0624	0.1039	(−0.6078, −0.3174)
	$\widehat{\alpha}_1^{PW}$	0.3721	0.0186	0.1224	0.2652	0.4255	(−1.7113, −0.2876)

Bandwidth rates were $\lambda_{h_a} = \lambda_{h_b} = 1/5$. Proportionality constants were $\mathcal{C}_{h_a} = 0.39$ and $\mathcal{C}_{h_b} = 2.37$. 1000 simulations in every case.
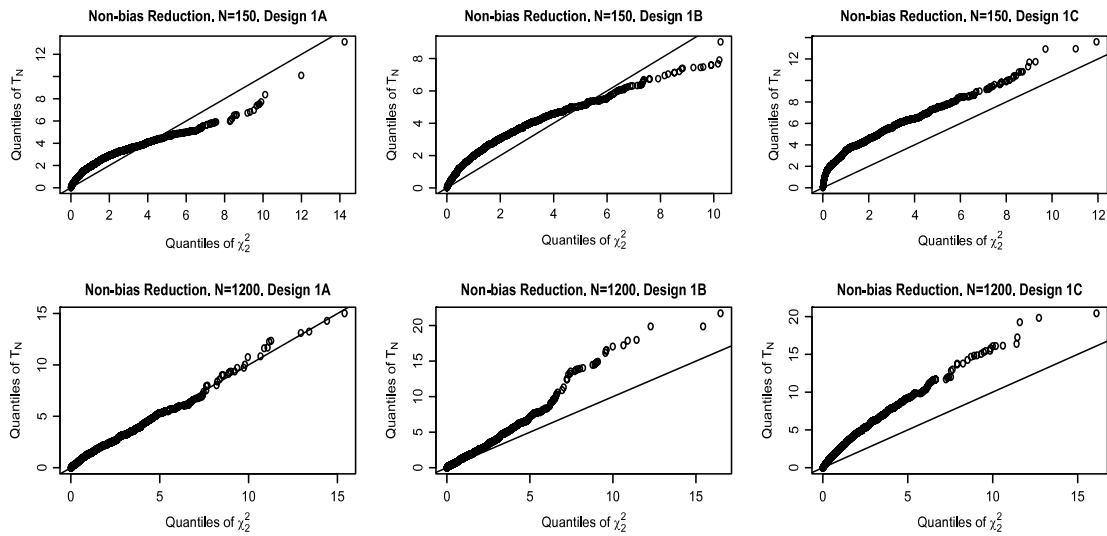

Fig. 3. Quantile–quantile comparison between the test-statistic $\mathcal{T}_N$ without bias reduction, and a χ_2^2 random variable for Designs 1A, 1B and 1C. All bandwidth rates used $\lambda = 1/5$. The bandwidth proportionality constants for *Design* 1A were: $\mathcal{C}_{h_a} = 0.39$, $\mathcal{C}_{h_b} = 2.37$, $\mathcal{C}_{\widetilde{h}_b} = \mathcal{C}_{\widetilde{b}} = 3.8$, $\mathcal{C}_{\widetilde{h}_c} = 0.9$. For *Design* 1B: $\mathcal{C}_{h_a} = 0.50$, $\mathcal{C}_{h_b} = 2.35$, $\mathcal{C}_{\widetilde{h}_b} = \mathcal{C}_{\widetilde{b}} = 3.5$, $\mathcal{C}_{\widetilde{h}_c} = 0.4$. For *Design* 1C: $\mathcal{C}_{h_a} = 0.50$, $\mathcal{C}_{h_b} = 3.00$, $\mathcal{C}_{\widetilde{h}_b} = \mathcal{C}_{\widetilde{b}} = 0.32$, $\mathcal{C}_{\widetilde{h}_c} = 0.28$.

Appendix A. Mathematical appendix

A.1. Proof of Theorem 1

We will focus on $\widehat{\theta}_1$, the proofs for $\widehat{\theta}_2$ are identical in nature. As before, θ_1 will denote the true value of the parameter. Using the invertibility properties of $F_1(\cdot)$, after rearranging we can express $\widehat{\theta}_1 - \theta_1$ as given in (A.1) in Box I, Before proceeding, we state the next result:

Proposition 2. *If X is absolutely continuous with respect to Lebesgue measure and if Assumption* A2 *is satisfied along with the properties about $K_b(\cdot)$ and h_b stated in* A4, *then for any compact set* $\mathfrak{X} \subset$ *int* $(\mathbb{S}(X))$, *any $\delta > 0$ and for $p = 1, 2$,*

$$(N^{\delta-1} h_b^L)^{1/2} \sup_{x \in \mathfrak{X}} \left|\widehat{\mu}_p(x) - \mu_p(x)\right| = O_p(1)$$

$$\widehat{\mu}_p(x) - \mu_p(x) = \frac{1}{N h_b^L} \sum_{i=1}^{N} \frac{\left[Y_{pi} - \mu_p(x)\right]}{f_X(x)} K_b\left(\frac{X_i - x}{h_b}\right) + \xi_N(x)$$

with $(N^{1-\delta} h_b^L) \sup_{x \in \mathfrak{X}} |\xi_N(x)| = O_p(1)$.

Given the assumptions of the proposition, its proof follows from Theorem A-1 in Aradillas-Lopez (2010). Alternatively, Lemma 3 in Collomb and Hardle (1986) could be used. From here we proceed by analyzing the term B_N in (A.1). From Proposition 2, Assumptions A2 and A4 we can express B_N as

$$B_N = \binom{N}{2}^{-1} \frac{1}{h_a} \sum_{i<j} K_a\left(\frac{\mu_{1i} - \mu_{1j}}{h_a}\right) (Z_{1i} - Z_{1j})$$
$$\times \left[(\widehat{Z}_{1i} - Z_{1i}) - (\widehat{Z}_{1j} - Z_{1j})\right]' \theta_1 \phi(X_i)\phi(X_j) + o_p(N^{-1/2}).$$

Note that $\widehat{Z}_{1i} - Z_{1i} = (0, \ldots, 0, \widehat{\mu}_{2i} - \mu_{2i})'$. We will examine the first component of B_N. Let $S \equiv (X, Y_1, Y_2, \mu_1, \mu_2)$ and define

Table B.4
Simulation results for Designs 1B and 1C. Non-bias reducing kernels.

n	$\widehat{\theta}$	RMSE	Bias	$|\widehat{\theta} - \theta_0|_{0.25}$	$|\widehat{\theta} - \theta_0|_{0.50}$	$|\widehat{\theta} - \theta_0|_{0.75}$	$(\widehat{\theta}_{0.025}, \widehat{\theta}_{0.975})$
Design 1B							
150	$\widehat{\beta}_1^{PW}$	0.1639	0.0752	0.0516	0.1125	0.1937	$(-0.7076, -0.1464)$
	$\widehat{\alpha}_1^{PW}$	0.7828	-0.0959	0.2677	0.5154	0.8646	$(-2.6468, 0.4892)$
300	$\widehat{\beta}_1^{PW}$	0.1177	0.0513	0.0395	0.0837	0.1331	$(-0.6720, -0.2539)$
	$\widehat{\alpha}_1^{PW}$	0.5641	-0.0891	0.1809	0.3814	0.6413	$(-2.2730, -0.0651)$
600	$\widehat{\beta}_1^{PW}$	0.0867	0.0414	0.0273	0.0608	0.1012	$(-0.6129, -0.3037)$
	$\widehat{\alpha}_1^{PW}$	0.3507	-0.0655	0.1112	0.2468	0.4177	$(-1.7108, -0.4172)$
1200	$\widehat{\beta}_1^{PW}$	0.0645	0.0328	0.0219	0.0441	0.0761	$(-0.5844, -0.3594)$
	$\widehat{\alpha}_1^{PW}$	0.2498	-0.0499	0.0791	0.1666	0.2934	$(-1.5267, -0.6093)$
Design 1C							
150	$\widehat{\beta}_1^{PW}$	0.1800	0.0557	0.0583	0.1204	0.2104	$(-0.7963, -0.1285)$
	$\widehat{\alpha}_1^{PW}$	1.4321	-0.5181	1.5580	2.4856	3.3217	$(-6.4779, -1.10963)$
300	$\widehat{\beta}_1^{PW}$	0.1307	0.0433	0.0421	0.0897	0.1523	$(-0.7228, -0.2379)$
	$\widehat{\alpha}_1^{PW}$	1.0032	-0.4717	1.8873	2.4181	2.9864	$(-5.4380, -1.9204)$
600	$\widehat{\beta}_1^{PW}$	0.0955	0.0308	0.0316	0.0639	0.1101	$(-0.6563, -0.2955)$
	$\widehat{\alpha}_1^{PW}$	0.5996	-0.2234	1.8464	2.1902	2.5837	$(-4.3789, -2.2340)$
1200	$\widehat{\beta}_1^{PW}$	0.0733	0.0350	0.0250	0.0491	0.0855	$(-0.5965, -0.3373)$
	$\widehat{\alpha}_1^{PW}$	0.3840	-0.0218	1.7598	1.9988	2.2750	$(-3.8187, -2.3206)$

All bandwidth rates used $\lambda = 1/5$. The bandwidth proportionality constants were given by: *Design* 1B.—$\mathcal{C}_{h_a} = 0.50$, $\mathcal{C}_{h_b} = 2.35$. *Design* 1C.—$\mathcal{C}_{h_a} = 0.50$, $\mathcal{C}_{h_b} = 3.00$. 1000 simulations in every case.

Table B.5
Simulation results for τ_N for designs 1A, B and C with non-bias reducing kernels.

n	Design 1A		Design 1B		Design 1C	
	Empirical size	95% quantile	Empirical size	95% quantile	Empirical size	95% quantile
150	0.015	5.004	0.038	5.555	0.203	8.368
300	0.025	5.198	0.049	5.872	0.197	8.613
600	0.044	5.781	0.063	6.190	0.249	9.314
1200	0.051	5.982	0.111	8.287	0.232	9.987

Empirical size is defined as $\Pr(\mathcal{T}_N \geq 5.99)$. Nominal size with a χ^2_2 approximation is 0.05. All bandwidth rates used $\lambda = 1/5$. The bandwidth proportionality constants were given by: *Design* 1A.—$\mathcal{C}_{h_a} = 0.39$, $\mathcal{C}_{h_b} = 2.37$, $\mathcal{C}_{\widetilde{h}_b} = \mathcal{C}_{\widetilde{b}} = 3.8$, $\mathcal{C}_{\widetilde{h}_c} = 0.9$. *Design* 1B.—$\mathcal{C}_{h_a} = 0.50$, $\mathcal{C}_{h_b} = 2.35$, $\mathcal{C}_{\widetilde{h}_b} = \mathcal{C}_{\widetilde{b}} = 3.50$, $\mathcal{C}_{\widetilde{h}_c} = 0.40$. Design 1C.—$\mathcal{C}_{h_a} = 0.50$, $\mathcal{C}_{h_b} = 3.00$, $\mathcal{C}_{\widetilde{h}_b} = \mathcal{C}_{\widetilde{b}} = 0.32$, $\mathcal{C}_{\widetilde{h}_c} = 0.28$. 1000 simulations in every case.

$$
\widehat{\theta}_1 - \theta_1 = D_N^{-1} \times \underbrace{\binom{N}{2}^{-1} \sum_{i<j} \frac{1}{h_a} K_a\left(\frac{\widehat{\mu}_{1i} - \widehat{\mu}_{1j}}{h_a}\right) \left(\widehat{Z}_{1i} - \widehat{Z}_{1j}\right) \left(F_1^{-1}|_{\mu_{1i}} - F_1^{-1}|_{\mu_{1j}}\right) \phi(X_i)\phi(X_j)}_{\equiv A_N}
$$

$$
+ D_N^{-1} \times \underbrace{\binom{N}{2}^{-1} \sum_{i<j} \frac{1}{h_a} K_a\left(\frac{\widehat{\mu}_{1i} - \widehat{\mu}_{1j}}{h_a}\right) \left(\widehat{Z}_{1i} - \widehat{Z}_{1j}\right) \left[\left(\widehat{Z}_{1i} - Z_{1i}\right) - \left(\widehat{Z}_{1j} - Z_{1j}\right)\right]' \theta_1 \phi(X_i)\phi(X_j)}_{\equiv B_N} \tag{A.1}
$$

where

$$
D_N = \binom{N}{2}^{-1} \sum_{i<j} \frac{1}{h_a} K_a\left(\frac{\widehat{\mu}_{1i} - \widehat{\mu}_{1j}}{h_a}\right) \left(\widehat{Z}_{1i} - \widehat{Z}_{1j}\right) \left(\widehat{Z}_{1i} - \widehat{Z}_{1j}\right)' \phi(X_i)\phi(X_j).
$$

Box I.

$$
\widehat{v}_N(S_i, S_j, S_k) = \frac{1}{h_a h_b^l} K_a\left(\frac{\mu_{1i} - \mu_{1j}}{h_a}\right) (Z_{1i} - Z_{1j})
$$

$$
\times \left\{ \frac{(Y_{2k} - \mu_{2i})}{f_X(X_i)} K_b\left(\frac{X_k - X_i}{h_b}\right) \right.
$$

$$
\left. - \frac{(Y_{2k} - \mu_{2j})}{f_X(X_j)} K_b\left(\frac{X_k - X_j}{h_b}\right) \right\} \phi(X_i)\phi(X_j).
$$

Now let $\widetilde{p}_N(S_i, S_j, S_k) = \sum_{\mathcal{C}} \widehat{v}_N(S_i, S_j, S_k)$, where $\mathcal{C}$ is the set $\{(i, j, k), (i, k, j), (j, k, i)\}$. Note that $\widetilde{p}_N(S_i, S_j, S_k)$ is symmetric with respect to its arguments. Using Proposition 2, Assumptions A2 and A4 we have

$$
\binom{N}{2}^{-1} \frac{1}{h_a} \sum_{i<j} K_a\left(\frac{\mu_{1i} - \mu_{1j}}{h_a}\right) (Z_{1i} - Z_{1j})
$$

$$\times \left[(\widehat{Z}_{1i} - Z_{1i}) - (\widehat{Z}_{1j} - Z_{1j}) \right]' \phi(X_i)\phi(X_j)\theta$$

$$= \alpha_1 \frac{1}{N} \binom{N}{2}^{-1} \sum_{i<j<k} \widetilde{p}_N(S_i, S_j, S_k) + o_p(N^{-1/2})$$

$$= \frac{N-2}{3N} \binom{N}{3}^{-1} \sum_{i<j<k} \widetilde{p}_N(S_i, S_j, S_k)\alpha_1 + o_p(N^{-1/2})$$

where $\sum_{i<j<k}(\cdot)$ denotes the sum over all combinations of three elements (i, j, k) out of $1, \ldots, N$. For $a \in \mathbb{R}^{L_1}$ and $b \in \mathbb{R}$ define

$$\underbrace{\Delta_1(a, b)}_{(L_1 \times 1)} = E\left[(a - Z_1)\phi(X) | \mu_1 = b \right].$$

By Assumptions A0–A3, $\Delta_1(a, b)$ is $\overline{M}$-times differentiable with respect to b, with bounded derivatives.[19] Adding Assumption A4, an $\overline{M}$-th order Taylor approximation yields

$$E\left[\widetilde{p}_N(S_i, S_j, S_k) | S_i \right] = 2\Delta_1(Z_{1i}, \mu_{1i})(Y_{2i} - \mu_{2i})\phi(X_i) + o_p(N^{-1/2})$$

$$\equiv 2\left(Z_{1i}E[\phi(X)|\mu_1 = \mu_{1i}] - E[Z_1\phi(X)|\mu_1 = \mu_{1i}] \right)$$

$$\times (Y_{2i} - \mu_{2i})\phi(X_i) + o_p(N^{-1/2}).$$

Iterated expectations yields $E\left[\widetilde{p}_N(S_i, S_j, S_k) \right] = o_p(N^{-1/2})$. Using Assumptions A2 and A4.(iii), we have $E[\|\widetilde{p}_N(S_i, S_j, S_k)\|^2] = O(N)$. Therefore Lemma A.3 in Ahn and Powell (1993) yields:

$$\binom{N}{3}^{-1} \sum_{i<j<k} \widetilde{p}_N(S_i, S_j, S_k) = \frac{3}{N} \sum_{i=1}^{N} 2\Delta_1(Z_{1i}, \mu_{1i})(Y_{2i} - \mu_{2i})$$

$$\times \phi(X_i) + o_P(N^{-1/2})$$

which finally yields

$$B_N \equiv \binom{N}{2}^{-1} \frac{1}{h_a} \sum_{i<j} K_a\left(\frac{\mu_{1i} - \mu_{1j}}{h_a} \right)(Z_{1i} - Z_{1j})$$

$$\times \left[(\widehat{Z}_{1i} - Z_{1i}) - (\widehat{Z}_{1j} - Z_{1j}) \right]' \theta_1 \phi(X_i)\phi(X_j)$$

$$= \frac{2\alpha_1}{N} \sum_{i=1}^{N} \Delta_1(Z_{1i}, \mu_{1i})(Y_{2i} - \mu_{2i})\phi(X_i) + o_P(N^{-1/2}).$$

Next we examine the term A_N in Eq. (A.1) (see Box II).

We start with A_{1_N}. If Assumptions A0–A4 are satisfied, an $\overline{M}$-th order Taylor expansion and Lemma A3 in Ahn and Powell (1993) yield the equation given in Box III.

Let

$$\overline{v}_N(S_i, S_j, S_k) = \frac{1}{h_a h_b^L} K_a\left(\frac{\mu_{1i} - \mu_{1j}}{h_a} \right)(F_1^{-1}(\mu_{1i}) - F_1(\mu_{1j}))$$

$$\times \left\{ \frac{(Y_{2k} - \mu_{2i})}{f_X(X_i)} K_b\left(\frac{X_k - X_i}{h_b} \right) \right.$$

$$\left. - \frac{(Y_{2k} - \mu_{2j})}{f_X(X_j)} K_b\left(\frac{X_k - X_j}{h_b} \right) \right\} \phi(X_i)\phi(X_j)$$

and define $\overline{p}_N(S_i, S_j, S_k) = \sum_{\mathcal{C}} \overline{v}_N(S_i, S_j, S_k)$, where $\mathcal{C}$ is the set $\{(i, j, k), (i, k, j), (j, k, i)\}$. An argument identical to the one used above to examine B_N can be used to show that we can express

$$\binom{N}{2}^{-1} \frac{1}{h_a} \sum_{i<j} K_a\left(\frac{\mu_{1i} - \mu_{1j}}{h_a} \right) \left[(\widehat{Z}_{1i} - Z_{1i}) - (\widehat{Z}_{1j} - Z_{1j}) \right]$$

$$\times (F_1^{-1}(\mu_{1i}) - F_1^{-1}(\mu_{1j}))\phi(X_i)\phi(X_j)$$

$$= \frac{N-2}{3N} \binom{N}{3}^{-1} \sum_{i<j<k} \overline{p}_N(S_i, S_j, S_k) + o_p(N^{-1/2}).$$

Again, Lemma A.3 in Ahn and Powell (1993) and our assumptions yield, via a Taylor approximation: $\binom{N}{3}^{-1} \sum_{i<j<k} \overline{p}_N(S_i, S_j, S_k) = o_p(N^{-1/2})$. These two results together yield $A_{1_N} = o_p(N^{-1/2})$. Now we analyze A_{2_N}, define

$$v_N(S_i, S_j, S_k) = \frac{1}{h_a^2 h_b^L} K_a^{(1)}\left(\frac{\mu_{1i} - \mu_{1j}}{h_a} \right)(Z_{1i} - Z_{1j})(F_1^{-1}(\mu_{1i})$$

$$- F_1^{-1}(\mu_{1j})) \times \left\{ \frac{(Y_{2k} - \mu_{2i})}{f_X(X_i)} K_b\left(\frac{X_k - X_i}{h_b} \right) \right.$$

$$\left. - \frac{(Y_{2k} - \mu_{2j})}{f_X(X_j)} K_b\left(\frac{X_k - X_j}{h_b} \right) \right\} \phi(X_i)\phi(X_j),$$

and let $p_N(S_i, S_j, S_k) = \sum_{\mathcal{C}} v_N(S_i, S_j, S_k)$, where $\mathcal{C}$ is the set $\{(i, j, k), (i, k, j), (j, k, i)\}$ (note that $p_N(S_i, S_j, S_k)$ is symmetric.[20]) Using Proposition 2, (A2) and (A4) we obtain[21]

$$A_{2_N} = \frac{N-2}{3N} \binom{N}{3}^{-1} \sum_{i<j<k} p_N(S_i, S_j, S_k) + o_p(N^{-1/2}).$$

Assumptions A0–A4 imply that $E\left[\|v_N(S_i, S_j, S_k)\|^2 \right] = O(h_a^{-3} h_b^{-L})$. The existence of $\overline{\delta} > 0$ such that $N^{1/2-\overline{\delta}} h_b^L h_a^2 \to \infty$ by implies $Nh_b^L h_a^3 \to \infty$. Combined, these results yield $E\left[\|p_N(S_i, S_j, S_k)\|^2 \right] = O(N)$. An $\overline{M}$-th order Taylor approximation and our assumptions yield[22]

$$E\left[p_N(S_i, S_j, S_k) | S_i \right] = \frac{2\Delta_1(Z_{1i}, \mu_{1i})f_{\mu_1}(\mu_{1i})\phi(X_i)}{F_1'(W_{1i} + Z_{1i}'\theta_1)}$$

$$\times (Y_{1i} - \mu_{1i}) + o_p(N^{-1/2}).$$

Finally, Lemma A.3 in Ahn and Powell (1993) yields

$$A_{2_N} = \frac{2}{N} \sum_{i=1}^{N} \frac{\Delta_1(Z_{1i}, \mu_{1i})f_{\mu_1}(\mu_{1i})\phi(X_i)}{F_1'(W_{1i} + Z_{1i}'\theta_1)}(Y_{1i} - \mu_{1i}) + o_p(N^{-1/2})$$

$$\equiv \frac{2}{N} \sum_{i=1}^{N} \frac{(Z_{1i}E[\phi(X)|\mu_1 = \mu_{1i}] - E[Z_1\phi(X)|\mu_1 = \mu_{1i}])f_{\mu_1}(\mu_{1i})\phi(X_i)}{F_1'(W_{1i} + Z_{1i}'\theta_1)}$$

$$\times (Y_{1i} - \mu_{1i}) + o_p(N^{-1/2})$$

$$= A_N,$$

where the last line follows from $A_{1_N} = o_p(N^{-1/2})$.

Finally, we analyze the matrix D_N. Given our assumptions (the properties of $K_a(\cdot)$, $K_a^{(1)}(\cdot)$ and $K_a^{(2)}(\cdot)$) and the result from Proposition 2, a dominated convergence argument yields

$$\binom{N}{2}^{-1} \frac{1}{h_a} \sum_{i<j} K_a\left(\frac{\mu_{1i} - \mu_{1j}}{h_a} \right)(Z_{1i} - Z_{1j})$$

$$\times \left[(\widehat{Z}_{1i} - Z_{1i}) - (\widehat{Z}_{1j} - Z_{1j}) \right]' \phi(X_i)\phi(X_j) = o_p(1)$$

and

$$\binom{N}{2}^{-1} \frac{1}{h_a^2} \sum_{i<j} K_a^{(1)}\left(\frac{\mu_{1i} - \mu_{1j}}{h_a} \right)(Z_{1i} - Z_{1j})(Z_{1i} - Z_{1j})'$$

$$\times \left[(\widehat{\mu}_{1i} - \mu_{1i}) - (\widehat{\mu}_{1j} - \mu_{1j}) \right] \phi(X_i)\phi(X_j) = o_p(1).$$

Combining Proposition 2 with the requirement that $N^{1/2-\overline{\delta}} h_b^L h_a^2$ for some $\overline{\delta} > 0$ and the uniform boundedness of $K_a^{(2)}(\cdot)$, we obtain

$$D_N = \binom{N}{2}^{-1} \frac{1}{h_a}$$

---

[19] Note the importance of having $\text{Var}(\mu_{2i}|\mu_{1i}) \neq 0$ with positive probability. With positive probability, we must have that conditional on μ_{2i}, μ_{1i} is not deterministic. See the paragraph following Assumption A1.

[20] Symmetry of $K_a(\cdot)$ implies that $K_a^{(1)}(x) = -K_a^{(1)}(-x)$.

[21] Note the importance of the requirement that $N^{1/2-\overline{\delta}} h_b^L h_a^2 \to \infty$ for some $\overline{\delta} > 0$.

[22] The properties about $K_a^{(1)}(\cdot)$ are crucial here.

$$A_N = \binom{N}{2}^{-1} \frac{1}{h_a} \sum_{i<j} K_a \left(\frac{\widehat{\mu}_{1i} - \widehat{\mu}_{1j}}{h_a} \right) \left(\widehat{Z}_{1i} - \widehat{Z}_{1j} \right) \left(F_1^{-1}(\mu_{1i}) - F_1^{-1}(\mu_{1j}) \right) \phi(X_i)\phi(X_j)$$

$$= \underbrace{\binom{N}{2}^{-1} \frac{1}{h_a} \sum_{i<j} K_a \left(\frac{\mu_{1i} - \mu_{1j}}{h_a} \right) \left[(Z_{1i} - Z_{1j}) + (\widehat{Z}_{1i} - Z_{1i}) - (\widehat{Z}_{1j} - Z_{1j}) \right] \left(F_1^{-1}(\mu_{1i}) - F_1^{-1}(\mu_{1j}) \right) \phi(X_i)\phi(X_j)}_{\equiv A_{1N}}$$

$$+ \underbrace{\binom{N}{2}^{-1} \frac{1}{h_a^2} \sum_{i<j} K_a \left(\frac{\mu_{1i} - \mu_{1j}}{h_a} \right) \left[(\widehat{\mu}_{1i} - \mu_{1i}) - (\widehat{\mu}_{1j} - \mu_{1j}) \right] (Z_{1i} - Z_{1j}) \left(F_1^{-1}(\mu_{1i}) - F_1^{-1}(\mu_{1j}) \right) \phi(X_i)\phi(X_j)}_{\equiv A_{2N}}$$

Box II.

$$\binom{N}{2}^{-1} \frac{1}{h_a} \sum_{i<j} \underbrace{K_a \left(\frac{\mu_{1i} - \mu_{1j}}{h_a} \right) (Z_{1i} - Z_{1j}) \left(F_1^{-1}(\mu_{1i}) - F_1^{-1}(\mu_{1j}) \right) \phi(X_i)\phi(X_j)}_{\text{symmetric}} = o_p(N^{-1/2})$$

Box III.

$$\times \sum_{i<j} \underbrace{K_a \left(\frac{\mu_{1i} - \mu_{1j}}{h_a} \right) (Z_{1i} - Z_{1j})(Z_{1i} - Z_{1j})' \phi(X_i)\phi(X_j)}_{\text{symmetric}}$$
$$+ o_p(1).$$

Using the previous condition about the bandwidths, taking expectations and using the assumption that $E\left[\|Z\|^4 \right] < \infty$, Lemma A.3 in Ahn and Powell (1993) yields

$$D_N = E\left[\left(E[Z_p Z_p' \phi(X)|\mu_p] \cdot E[\phi(X)|\mu_p] - E[Z_p \phi(X)|\mu_p] E[Z_p \phi(X)|\mu_p]' \right) f_{\mu_p}(\mu_p) \right] + o_p(1).$$

This completes the proof of Theorem 1. □

A.2. Proof of Theorem 2

For reasons of space, we present here a summary of the proof. A detailed step-by-step derivation can be found in the technical appendix supplement at http://www.ssc.wisc.edu/~aaradill/suppl_pwinfo.pdf or it can be made readily available upon request. The key to the proof is to show that U_{p_N} can be expressed as

$$U_{p_N} = \binom{N}{2}^{-1} \sum_{i<j} \frac{\varepsilon_{p_i}\varepsilon_{p_j}\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L} \overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}} \right)$$
$$+ \alpha_{p_0} \left(U_{p_N}^{B_I} + U_{p_N}^{B_{II}} \right) + U_{p_N}^{B_{III}} + U_{p_N}^{B_{IV}} + \mathcal{R}_{p_N}, \quad (A.2)$$

where $\varepsilon_{p_i} \equiv Y_{p_i} - F_p(t_{p_i})$, $N\widetilde{b}^{L/2}\mathcal{R}_{p_N} \overset{p}{\to} 0$, $N\widetilde{h}_c^{3/2}U_{p_N}^{B_I} \overset{d}{\to} \mathcal{N}(0, V_p^{B_I})$, $N\widetilde{h}_b^{L/2}U_{p_N}^{B_{II}} \overset{d}{\to} \mathcal{N}(0, V_p^{B_{II}})$, $N\widetilde{h}_c^{1/2}U_{p_N}^{B_{IV}} \overset{d}{\to} \mathcal{N}(0, V_p^{B_{IV}})$, and $N U_{p_N}^{B_{III}} \overset{d}{\to} 3\mathcal{Y}$, where $\mathcal{Y}$ can be expressed as a linear combination (corresponding to the general structure described in Lemma 5.1.4.A in Serfling (1980)) of independent χ_1^2 random variables. The proof of (A.2) uses asymptotically linear representation results like those in Thoerem A-1 of Aradillas-Lopez (2010) or Lemma 3 in Collomb and Hardle (1986), as well as Hoeffding decomposition properties of (i) nondegenerate U-statistics such as those studied in Ahn and Powell (1993), (ii) degenerate U-statistics such as those in Lemma 5.1.4.A in Serfling (1980) and (iii) degenerate U-statistics such as those in Theorem 1 of Hall (1984). Once this is established,

the bandwidth conditions in Assumption B2 imply

$$N\widetilde{b}^{L/2}U_{p_N} = N\widetilde{b}^{L/2} \binom{N}{2}^{-1} \sum_{i<j} \frac{\varepsilon_{p_i}\varepsilon_{p_j}\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L} \overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}} \right) + \xi_N,$$
$$\text{where } \xi_N \overset{p}{\to} 0.$$

Next, we use Theorem 1 in Hall (1984) to show that

$$\Sigma_p^{-1} N\widetilde{b}^{L/2} \binom{N}{2}^{-1} \sum_{i<j} \frac{\varepsilon_{p_i}\varepsilon_{p_j}\overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L}$$
$$\times \overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}} \right) \overset{d}{\longrightarrow} \mathcal{N}(0, 1),$$

where $\Sigma_p = E\left[\frac{\varepsilon_{pj}^2 \varepsilon_{pi}^2 \overline{\phi}(X_i)^2 \overline{\phi}(X_j)^2}{\widetilde{b}^L} \overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}} \right)^2 \right].$

This yields $\Sigma_p^{-1} N\widetilde{b}^{L/2}U_{p_N} \equiv \Sigma_p^{-1} \mathcal{T}_{p_N} \overset{d}{\longrightarrow} \mathcal{N}(0, 1)$. We then show that for any pair of constants $\tau_1, \tau_2 \in \mathbb{R}$,

$$\tau_1 \mathcal{T}_{1_N} + \tau_2 \mathcal{T}_{2_N} \overset{d}{\longrightarrow} \mathcal{N}\left(0, \tau_1^2 \Sigma_1 + \tau_2^2 \Sigma_2 + 2\tau_1 \tau_2 \Sigma_{1,2} \right),$$

where $\Sigma_{1,2} = E\left[\frac{\varepsilon_{1j}\varepsilon_{1i}\varepsilon_{2j}\varepsilon_{2i}\overline{\phi}(X_i)^2 \overline{\phi}(X_j)^2}{\widetilde{b}^L} \overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}} \right)^2 \right].$

From here, the Cramer–Wold device and the properties of the normal distribution imply

$$\left(\mathcal{T}_{1_N}, \mathcal{T}_{2_N} \right)' \overset{d}{\longrightarrow} N(0, \Sigma), \quad \text{where } \Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2} & \Sigma_2 \end{pmatrix}.$$

Note that Σ is invertible under the conditions of Theorem 2 which implies (via the continuous mapping theorem) that $N^2 \widetilde{b}^L (U_{1_N}, U_{2_N}) \Sigma^{-1} (U_{1_N}, U_{2_N})' \overset{d}{\longrightarrow} \chi_2^2$. Moreover, under the conditions of Theorem 2 the estimator $\widehat{\Sigma}$ described there is consistent for Σ. This yields part 1 of the theorem. Namely, if the model is correctly specified, $\mathcal{T}_N \equiv N^2 \widetilde{b}^L \left(U_{1_N}, U_{2_N} \right) \widehat{\Sigma}^{-1} \left(U_{1_N}, U_{2_N} \right)' \overset{d}{\longrightarrow} \chi_2^2$.

To prove part 2 of Theorem 2, we let

$$D_p = E\left[\left(E\left[Z_p Z_p' \overline{\phi}(X)|\mu_p \right] E\left[\overline{\phi}(X)|\mu_p \right] - E\left[Z_p \overline{\phi}(X)|\mu_p \right] E\left[Z_p \overline{\phi}(X)|\mu_p \right]' \right) f_{\mu_p}(\mu_p) \right]$$

$$C_p = E\left[\left(E\left[Z_p W_p \overline{\phi}(X)|\mu_p \right] E\left[\overline{\phi}(X)|\mu_p \right] - E\left[Z_p \overline{\phi}(X)|\mu_p \right] E\left[W_p \overline{\phi}(X)|\mu_p \right] \right) f_{\mu_p}(\mu_p) \right].$$

Given our assumptions, a dominated convergence argument easily yields

$$\widehat{\theta}_p \overset{p}{\longrightarrow} -D_p^{-1} C_p \equiv \theta_p^*.$$

Under the conditions of part 2 of Theorem 2, θ_p^* is well-defined even if the model is incorrect and Eq. (1) is violated with positive probability. Conversely, if the model is correctly specified we know that $W_p = F_p^{-1}(\mu_p) - Z_p'\theta_p$ and $\theta_p^* = \theta_p$ (the true structural parameter value). Let $t_{pi}^* = W_{pi} + Z_{pi}'\theta_p^*$. Given our assumptions, conditional on X_j we have[23]

$$\widehat{F}_p(\widehat{t}_{pj}) \overset{p}{\longrightarrow} \frac{E\left[\mu_p(X_i)\varphi(X_i)|t_{pi}^* = t_{pj}^*\right]}{E\left[\varphi(X_i)|t_{pi}^* = t_{pj}^*\right]} \equiv F_p^*(t_{pj}^*).$$

Moreover, this convergence is uniform over $\mathcal{X}$. This yields

$$U_{p_N} = \binom{N}{2}^{-1} \sum_{i<j} \frac{\varepsilon_{pi}^* \varepsilon_{pj}^* \overline{\phi}(X_i)\overline{\phi}(X_j)}{\widetilde{b}^L} \overline{H}\left(\frac{\Delta X_{ij}}{\widetilde{b}}\right) + o_p(1),$$

with $\varepsilon_{pi}^* \equiv Y_{pi} - F_p^*(t_{pi}^*)$.

$$= E\left[\left(\mu_p(X_i) - F_p^*(t_{pi}^*)\right)^2 \overline{\phi}(X_i)^2\right] + o_p(1).$$

Suppose Eq. (1) is violated with positive probability for player p in the set $\mathcal{X}$ and

$$\Pr\left[\mu_p(X_i)E\left[\varphi(X_i)\mid t_{pi}^*\right] \neq E\left[\mu_p(X_i)\varphi(X_i)\mid t_{pi}^*\right] \mid X_i \in \mathcal{X}\right] > 0.$$

Then, under the conditions of part 2 of Theorem 2 we would have

$$\Pr\left[\mu_p(X_i) \neq F_p\left(t_p^*(X_i)\right) | X_i \in \mathcal{X}\right] > 0.$$

Consequently, $E\left[\left(\mu_{pi} - F_p^*(t_{pi}^*)\right)^2 \overline{\phi}(X_i)^2\right] > 0$. It follows that if Eq. (1) is violated with positive probability,

$$\Pr\left[N\widetilde{b}^{L/2} \left|U_{p_N}\right| > m_N\right] \longrightarrow 1$$

for any sequence of scalars such that $m_N/(N\widetilde{b}^{L/2}) \longrightarrow 0$.

If the exclusion restriction in Assumption A1 is satisfied and if $Y_1 - F_1^*(t_1^*)$ and $Y_2 - F_2^*(t_2^*)$ are not perfectly correlated conditional on $X \in \mathcal{X}$ (as it is assumed in part 2 of Theorem 2), it is easy to show that $\widehat{\Sigma}^{-1}$ has a well-defined probability-limit. Combined with the previous result, this yields

$$\Pr\left(|\mathcal{T}_N| > m_N\right) \longrightarrow 1$$

for any sequence of scalars such that $m_N/\left(N^2 \widetilde{b}^L\right) \longrightarrow 0$.

Therefore $\mathcal{T}_N$ diverges w.p.1. This concludes the proof. □

Appendix B. Monte Carlo experiment results

Kernels and bandwidths used

For a random variable ψ let $\widehat{\mathcal{R}}(\psi)$ denote the "rule of thumb" proportionality constant in Silverman (1986, Eq. (3.31)). We used covariate-specific bandwidths of the form $h_b(W_p) = \mathcal{C}_{h_b}\widehat{\mathcal{R}}(W_p) \cdot N^{-\lambda_{h_b}}$, $h_b(V_p) = \mathcal{C}_{h_b}\widehat{\mathcal{R}}(V_p) \cdot N^{-\lambda_{h_b}}$, $\widetilde{h}_b(W_p) = \mathcal{C}_{\widetilde{h}_b}\widehat{\mathcal{R}}(W_p) \cdot N^{-\lambda_{\widetilde{h}_b}}$, $\widetilde{h}_b(V_p) = \mathcal{C}_{\widetilde{h}_b}\widehat{\mathcal{R}}(V_p) \cdot N^{-\lambda_{\widetilde{h}_b}}$, $\widetilde{b}(W_p) = \mathcal{C}_{\widetilde{b}}\widehat{\mathcal{R}}(W_p) \cdot N^{-\lambda_{\widetilde{b}}}$, $\widetilde{b}(V_p) = \mathcal{C}_{\widetilde{b}}\widehat{\mathcal{R}}(V_p) \cdot N^{-\lambda_{\widetilde{b}}}$, $h_a(\widehat{\mu}_p) = \mathcal{C}_{h_a}\widehat{\mathcal{R}}(\widehat{\mu}_p) \cdot N^{-\lambda_{h_a}}$, and $\widetilde{h}_c(\widehat{t}_p) = \mathcal{C}_{h_c}\widehat{\mathcal{R}}(\widehat{t}_p) \cdot N^{-\lambda_{\widetilde{h}_c}}$. The exponents $\lambda_{h_b}, \ldots, \lambda_{\widetilde{b}}$ were chosen to satisfy the convergence rates in our bandwidth assumptions. To determine the proportionality constants $\mathcal{C}_{h_b}$, etc., we employed a procedure based on the asymptotic approximations of our estimator and specification test-statistic. The details of the procedure used can be found in the supplement to this paper, at

http://www.ssc.wisc.edu/~aaradill/suppl_pwinfo.pdf. The results below show the actual values for the proportionality constants used. Bias-reducing kernels used were of the form $K(\psi) = \left(a_0 + a_1\psi^2 + a_2\psi^4 + \cdots + a_p\psi^{2p}\right)\phi(\psi)$, where ϕ is the Standard Normal density function, and the polynomial coefficients were selected to achieve the kernel requirements in the paper. Non-bias reducing kernels (as indicated in the tables below) were simply equal to $\phi(\cdot)$. In what follows, for $p = 1, 2$ we let $\psi^{\widehat{\beta}_p}$ and $\psi^{\widehat{\alpha}_p}$ denote the asymptotic influence functions (see Theorem 1) of $\widehat{\beta}_p$ and $\widehat{\alpha}_p$.

References

Aguirregabiria, V., Mira, P., 2002. Swapping the nested fixed point algorithm: a class of estimators for discrete markov decision models. Econometrica 70, 1519–1543.

Aguirregabiria, V., Mira, P., 2007. Sequential estimation of dynamic discrete games. Econometrica 75 (1), 1–53.

Ahn, H., Ichimura, H., Powell, J., 1997. Simple estimators of monotone index models, Working Paper, UC Berkeley.

Ahn, H., Manski, C., 1993. Distribution theory for the analysis of binary choice under uncertainty with nonparametric estimation of expectations. Journal of Econometrics 56, 291–321.

Ahn, H., Powell, J., 1993. Semiparametric estimation of censored selection models. Journal of Econometrics 58, 3–29.

Andrews, D., Berry, S., Jia, P., 2004. Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store location. Working Paper. Yale University.

Aradillas-Lopez, A., 2010. Semiparametric estimation of a simultaneous game with incomplete information. Journal of Econometrics 157 (2), 409–431.

Aradillas-Lopez, A., Honoré, B., Powell, J., 2007. Pairwise difference estimation with nonparametric control variables. International Economic Review 48 (4), 1119–1158.

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

Augereau, A., Greenstein, S., Rysman, M., 2006. Coordination vs. differentiation in a standards war: 56K modems. Rand Journal of Economics 37, 887–909.

Bajari, P., Hong, H., Kreiner, J., Nekipelov, D., 2010. Estimating static models of strategic interaction. Journal of Business and Economic Statistics 28 (4), 469–482.

Bajari, P., Hong, H., Ryan, S., 2005, Identification and estimation of discrete games of complete information, Working Paper.

Beresteanu, A., Molchanov, I., Molinari, F., 2008, Sharp Identification Regions in Games, CEMMAP Working Paper CWP15/08. Institute for Fiscal Studies. UCL.

Berry, S., 1992. Estimation of a model of entry in the airline industry. Econometrica 60 (4), 889–917.

Berry, S., Reiss, P., 2007. Empirical models of entry and market structure. In: Armstrong, M., Porter, R. (Eds.), Handbook of Industrial Organization, vol. 3. Elsevier Science, pp. 1845–1886.

Berry, S., Tamer, E., 2006. Identification in models of oligopoly entry. In: Blundell, R., Newey, W., Persson, T. (Eds.), Advances in Econometrics. Theory and Applications, Ninth World Congress. In: Econometric Society Monographs, vol. 2. Cambridge University Press, pp. 46–85.

Bjorn, P.A., Vuong, Q., 1984. Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an Application to Labor Force Participation, SSWP No. 537, Caltech.

Blundell, R., Powell, J., 2004. Endogeneity in semiparametric binary response models. Review of Economic Studies 71, 655–679.

Bresnahan, T., Reiss, P., 1990. Entry in monopoly markets. Review of Economic Studies 57, 531–553.

Bresnahan, T., Reiss, P., 1991a. Empirical models of discrete games. Journal of Econometrics 48, 57–81.

Bresnahan, T., Reiss, P., 1991b. Entry and competition in concentrated markets. Journal of Political Economy 99, 977–1009.

Brock, W., Durlauf, S., 2001a. Discrete choice with social interactions. The Review of Economic Studies 68, 235–260.

Brock, W., Durlauf, S., 2001b. Interactions-based models. In: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 5. Elsevier Science, pp. 3297–3380.

Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models. In: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 6B. Elsevier Science, pp. 5549–5632.

Chen, X., Linton, O., VanKeilegom, I., 2003. Estimation of semiparametric models when the criterion function is not smooth. Econometrica 71 (5), 1591–1608.

Ciliberto, F., Tamer, E., 2009. Market structure and multiple equilibria in airline markets. Econometrica 77, 1791–1828.

Cohen, A., Manuszak, M., 2006. Endogenous Market Structure with Discrete Product Differentiation and Multiple Equilibria: An Empirical Analysis of Competition Between Banks and Thrifts. Working Paper.

Collomb, G., Hardle, W., 1986. Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression from dependent observations. Stochastic Processes and their Applications 23, 77–89.

---

[23] We use the notation μ_{pi} and $\mu_p(X_i)$ interchangeably.

Davis, P., 2006. Estimation of quantity games in the presence of indivisibilities and heterogeneous firms. Journal of Econometrics 134, 187–214.

de Paula, A., Tang, X., 2011, Inference of signs of interaction effects in simultaneous games with incomplete information. Econometrica (forthcoming).

Fan, Y., Li, Q., 1996. Consistent specification tests: omitted variables and semiparametric functional forms. Econometrica 64, 865–890.

Galichon, A., Henry, M., 2011. Set identification in models with multiple equilibria. Review of Economic Studies 78 (4), 1264–1298.

Haile, P., Hortacsu, A., Kosenok, G., 2008. On the empirical content of quantal response equilibrium. American Economic Review 98 (1), 180–200.

Hall, P., 1984. Central limit theorem for integrated square error of multivariate nonparametric density estimators. Journal of Multivariate Analysis 14, 1–16.

Han, A., 1987. The maximum rank correlation estimator. Journal of Econometrics 303–316.

Ho, K.E., 2009. Insurer-provider networks in the medical care market. American Economic Review 99 (1), 393–430.

Hong, H., Shum, M., 2010. Pairwise difference estimator of a dynamic optimization model. Review of Economic Studies 77, 273–304.

Honoré, B., Powell, J., 1994. Pairwise difference estimators of censored and truncated regression models. Journal of Econometrics 64, 241–278.

Honoré, B., Powell, J., 2005. Pairwise difference estimation in nonlinear models. In: Andrews, D., Stock, J. (Eds.), Identification and Inference in Econometric Models. Essays in Honor of Thomas Rothenberg. Cambridge University Press, pp. 520–553.

Horowitz, J., 1992. A smoothed maximum score estimator for the binary response model. Econometrica 60 (3).

Ichimura, H., 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. Journal of Econometrics 58 (1–2), 71–120.

Ichimura, H., Lee, L.-F., 1991. Semiparametric Least Squares Estimation of Multiple Index Models: Single Equation Estimation. Cambridge University Press, pp. 3–49.

Imbens, G., Newey, W., 2002. Identification and estimation of triangular simultaneous equations models without additivity. Working Paper. Harvard University and MIT.

Ishii, J., 2008, Interconnection Pricing, Compatibility and Investment in Network Industries: ATM Networks in the Banking Industry. Working Paper. Stanford University.

Jia, P., 2008. What happens when wal-mart comes to town: an empirical analysis of the discount retailing industry. Econometrica 76, 1263–1316.

Klein, R., Spady, W., 1993. An efficient semiparametric estimator of binary response models. Econometrica 61, 387–421.

Kooreman, P., 1994. Estimation of econometric models of some discrete games. Journal of Applied Econometrics 9, 255–268.

Manski, C., 1975. The maximum score estimation of the stochastic utility model of choice. Journal of Econometrics 3, 205–228.

Manski, C., 2004. Measuring expectations. Econometrica 72, 1329–1376.

Manski, C.F., 1985. Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator. Journal of Econometrics 27 (3), 313–333.

McKelvey, R., Palfrey, T., 1995. Quantal response equilibria for normal-form games. Games and Economic Behavior 10, 6–38.

Moro, A., 2003. The effect of statistical discrimination on black–white wage inequality: estimating a model with multiple equilibria. International Economic Review 44 (2), 467–500.

Pakes, A., 2008, Theory and empirical work on imperfectly competitive markets. Working Paper. Harvard University and NBER.

Pakes, A., Ostrovsky, M., Berry, S., 2007. Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). RAND Journal of Economics 38, 373–399.

Pakes, A., Porter, J., Ho, K., Ishii, J., 2006. Moment inequalities and their applications. Working Paper.

Pesendorfer, M., Schmidt-Dengler, P., 2008. Asymptotic least squares estimators for dynamic games. Review of Economic Studies 75, 901–928.

Ponomareva, M., Tamer, E., 2011. Misspecification in moment inequality models: back to moment equalities? The Econometrics Journal 14, 186–203.

Powell, J., Stock, J., Stoker, T., 1989. Semiparametric estimation of index coefficients. Econometrica 1403–1430.

Rilstone, P., 1993. Calculating the local Semiparametric Efficiency Bound for the Generated Regressors Problem 56, 357–370.

Seim, K., 2006. An empirical model of firm entry with endogenous product-type choices. RAND Journal of Economics 37 (3).

Serfling, R., 1980. Approximation Theorems of Mathematical Statistics. Wiley, New York, NY.

Sherman, R., 1993. The limiting distribution of the maximum rank correlation estimator. Econometrica 61, 123–137.

Signorino, C., 1999. Strategic interaction and the statistical analysis of international conflict. American Political Science Review 93 (2), 279–298.

Signorino, C., 2002. Strategy and selection in international relations. International Interactions 28, 93–115.

Signorino, C., Tarar, A., 2006. A unified theory and test of extended immediate deterrence. American Journal of Political Science 50 (3), 586–605.

Silverman, B., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall.

Sweeting, A., 2009. The strategic timing of radio commercials: an empirical analysis using multiple equilibria. RAND Journal of Economics 40 (4), 710–724.

Tamer, E.T., 2003. Incomplete bivariate discrete response model with multiple equilibria. Review of Economic Studies 70, 147–167.

Zheng, J., 1998. Consistent specification testing for conditional symmetry. Econometric Theory 14, 139–149.