Accounting for Expectational and Structural Error in Binary Choice Problems: A Moment Inequality Approach*

Michael J. Dickstein

Stanford University and NBER

Eduardo Morales

Princeton University and NBER

September 20, 2013

Abstract

Many economic decisions involve a binary choice - for example, when consumers decide to purchase a good or when firms decide to enter a new market. In such settings, agents' choices often depend on imperfect expectations of the future payoffs from their decision (expectational error) as well as factors that the econometrician does not observe (structural error). In this paper, we show that expectational error, under an assumption of rational expectations, is a source of classical measurement error, and we propose a novel moment inequality estimator that accounts for both expectational error and structural error in a binary choice model. With simulated data and Chilean firm-level customs data, we illustrate the identifying power of our inequalities and show the biases that arise when one ignores either source of error. We use the customs data to estimate the fixed costs exporters face when entering a new market.

Keywords: discrete choice models, moment inequalities, entry models

*We would like to thank Steve Berry, Jesús Carro, Phil Haile, Bo Honoré, Guido Imbens, Ariel Pakes, Christoph Rothe, Bernard Salanie and seminar participants at Princeton University and the SITE Summer Workshop for helpful comments. All errors are our own. Email: mjd@stanford.edu, ecmorales@princeton.edu.

1 Introduction

When firms or consumers face a choice –for example, when choosing to introduce a new product, consume a good, switch jobs, or enter new geographic markets– they seldom possess complete information about the benefits that will result from their decision. These decision-makers must therefore form expectations of the likely benefits, which may differ from the returns realized ex-post. For the econometrician, this distinction poses a problem: either find direct measures of an agent's ex-ante expected returns, a rarity, or devise an approximation to the unobserved expectation. In the latter case, researchers typically specify the covariates in the agent's information set or choose a proxy for the expectations, such as observed ex-post returns. When expectations are imperfect, these approximations introduce error.

We propose a new method to estimate the parameters of a binary discrete choice problem in the presence of expectational error. Our strategy employs moment inequalities that permit two types of errors: (1) an individual structural error, which captures choice variables known to the agent when she makes a decision but not observed by the econometrician, and (2) expectational error, which reflects a mismatch between the agents expectations of payoff-relevant variables and the ex-post realizations of these variables. For example, consider a firm's decision to enter a market, a prototypical setting we use later as an application of our approach. At the time of its entry decision, the firm knows the costs it must pay to enter the market. However, it has only an expectation of the sales revenue it will earn. In this setting, expectational error is the difference between observed ex-post sales revenue and the agent's unobserved expectations; the structural error includes variables like advertising expenditure, which are known to the agent when she makes a decision but are not in the econometrician's data.

The expectational error has a parallel in regression analysis: if we think of the agents' expectations as a true unobserved covariate and use future observed realizations as a proxy for that unobserved covariate, then the rational expectations assumption implies that what we label expectational error can also be thought of as a source of classical measurement error. That is, the measurement or expectational error is mean independent of the true covariate and correlated with the observed covariate. Chesher (2010), Chesher (2011) and Blundell and Powell (2004) introduce identification frameworks that allow the researcher to identify the parameters of binary choice models with endogenous regressors. None of these frameworks, however, accounts for the two types of errors that our model of an agent's choice generates.

The goal of this paper is to estimate the parameters of the deterministic portion of our binary choice model, which takes the form of an index function that is linear in observable covariates. We allow two distinct errors by relying on a restriction that arises naturally from the rational expectations assumption. Under rational expectations, the distribution of the expectational error is mean independent of any variable the agent uses to form expectations. Thus, any variable observed by the econometrician and used by the agent to form her expectations is orthogonal to the expectational error. Our benchmark statistical model combines this mean independence assumption on the expectational error with a standard parametric assumption on the structural error.[1] We show that the parameters of the index function are only partially identified in this setting and demonstrate the identifying power of moment inequalities generated from the model's restrictions. Our proposed identification and estimation approach can be thought of as an extension of binary probit and logit models, adjusted to permit distribution-free measurement error in the covariates.

Our partial identification approach relies on two novel types of moment inequalities, *score function* inequalities and *revealed preference* inequalities. Using these inequalities, we are able to obtain a set with the following properties: (a) it contains the true values of the index coefficients we aim to estimate; and (b) its size increases in the variance of the expectational error. Our moment inequalities are also very simple to compute.[2] They may be combined with standard inference methods for moment inequalities to provide set estimates of the index-function coefficients.[3]

We implement our estimator in two settings. First, we design a simulation exercise to illustrate the properties of our moment inequalities. Here, we test the performance of our estimator by comparing our output to the true parameters that generated the dataset for the exercise. Second, we use actual data to estimate a version of the single-agent static entry model in Morales et al. (2011).

Our simulation results illustrate the failure of standard estimation techniques in the presence of both structural and expectational errors. First, we demonstrate that max-

[1]We also show how to estimate the parameters of our model when we do not impose a parametric assumption on the structural error. Dropping this parametric assumption greatly reduces the identification power of our model. Therefore, we focus our analysis on the model that combines parametric assumptions on the structural error with non-parametric restrictions on the expectational error.

[2]Accompanying Matlab code to implement our methodology is posted on `https://www.stanford.edu/~mjd` and `https://sites.google.com/site/edumoralescasado/`.

[3]For references on how to compute confidence intervals in partially identified settings, see Chernozhukov et al. (2007), Andrews and Soares (2010), Pakes et al. (2011), Andrews and Shi (2011a), and Andrews and Shi (2011b).

imum likelihood estimation, which assumes no expectational error, results in estimates that are biased. We confirm that the size of the bias is increasing in the variance of the omitted error. Second, we apply moment inequality estimators defined in Pakes (2010) and Pakes et al. (2011), which rule out a choice-specific and individual-specific structural error; our results reveal that these inequalities are biased towards zero, generating identified sets that are too small and may not contain the true value of the parameter. In contrast, our score function and revealed preference inequalities guarantee that the true value of the parameter vector is contained in the identified set defined by our moment inequalities, even in the presence of both types of errors. Moreover, our simulation results show that our score function inequalities define bounds for the unknown parameter vector that are far more informative than bounds generated using only revealed preference inequality restrictions.

In our analysis of the single-agent static entry model, we use firm-level export data to estimate the parameters that determine the costs firms incur upon exporting to a new foreign market. Here, firms determine export destinations in each period based on (a) expected potential net revenue obtained by exporting to each location, and (b) the fixed costs of exporting. The fixed costs may depend on factors known to the firm when it decides whether to export but that are not in the data, generating structural error. As in our benchmark statistical model, we assume that the distribution of this structural error is known up to a scale parameter.

We estimate the model using firm-level export data for the Chilean chemical sector during 1996-2004. The dataset contains information on firms' export flows disaggregated by destination country and year, production inputs, and other firm-specific characteristics.[4] Although the data does not provide information on firms' expectations of potential export revenues, it does report realized export revenues that firms obtain in country-years with positive exports. Using this data, we estimate a reduced-form equation that allows us to predict firm-specific export revenues for each potential country-year pair.

Our moment inequality estimator indicates that firms' profit margins in export markets are between 7.8% and 14.2% and that their per-year fixed costs of exporting from Chile to the US, a distance of $5,000 - 6,000$ miles, are about three times as large as the costs of exporting to neighboring Argentina. A comparison of the maximum likelihood estimates with our moment inequality estimates indicates that ignoring the

[4]The data were provided by the Chilean Customs Agency and the Chilean Annual Industrial Survey.

expectational error firms' make when deciding whether to export will bias the parameter estimates in two directions: the maximum likelihood estimator (a) underpredicts the number of exporters in countries that are close to the country of origin of the firm and overpredicts entry in countries that are more distant; and (b) underpredicts the probability of entry by large firms and overpredicts this probability for small firms.

We structure the remainder of the paper as follows. In Section 2, we show that expectational error is a source of classical measurement error, and introduce a static binary choice model with both structural and expectational errors. In Section 3, we introduce both score function and revealed preference inequalities and examine the properties of these inequalities. Section 4 studies how the presence of both structural and measurement error affects the performance of estimators that neglect one of the two types of error. Section 5 generalizes the results on the benchmark model introduced in Section 2: we eliminate the parametric restrictions on the distribution of the structural error and introduce moment inequalities that identify the true value of the parameter vector without these assumptions. Section 6 describes the export decision problem and applies our inequality framework to estimate the structural parameters of a firm's export decision. Section 7 concludes.[5]

2 Static Binary Choice Model

In this section, we present our statistical model. We first demonstrate that expectational errors affecting agents with rational expectations are a special case of classical measurement error (Section 2.1). We then introduce a binary choice model that includes both structural and classical measurement error (Section 2.2). Finally, we compare our approach with alternative models in the literature (Section 2.3), and present a setting in which we carry out our simulation exercise (Section 2.4).

2.1 Expectational and Measurement Error

In this section, we demonstrate the similarities between expectational error in a rational expectations model and measurement error in a classical errors-in-variables model. Although this holds for decision problems in which the decision variable is continuous or

[5]We include all proofs in an appendix to the main text. Details of the empirical application appear in a separate online appendix.

discrete, we show these similarities in the context of a binary choice problem to remain consistent with our statistical model of interest.

Suppose individual i faces a choice between two alternatives, $j \in \{0, 1\}$. Firm i's difference in payoff between alternative 1 and alternative 0 is:[6]

$$U = \beta \mathcal{E}[X|\mathcal{J}] + \nu, \tag{1}$$

where $\mathcal{J}$ is the information set the agent uses for her decision, $\mathcal{E}[\cdot]$ denotes the expectation operator based on the subjective believes of the agent, and β is the parameter vector representing the agent's preferences. The variable X may capture, for example: characteristics of alternatives that the agent imperfectly observes at the time of her decision (e.g. demand level in a market), decisions by other agents (e.g. in simultaneous move games), or a continuation value function (e.g. dynamic discrete problems). The variable ν captures characteristics that the econometrician does not observe but that are included in $\mathcal{J}$. While the parameter vector β is assumed to be constant across individuals, all the variables may vary with i. Denote the expectational error as $\epsilon = \beta X - \beta \mathcal{E}[X|\mathcal{J}]$ and express the difference in payoffs as:

$$U = \beta X + \nu + \varepsilon, \tag{2}$$

where $\varepsilon = -\beta\epsilon$. Define $\mathbb{E}[\cdot]$ as the expectation operator of the data generating process (DGP) for X. The DGP captures the distribution of X across individuals in the population. Under the rational expectations assumption, agents' subjective beliefs coincide with the true DGP (i.e. $\mathcal{E}[\cdot] = \mathbb{E}[\cdot]$), hence: $\mathbb{E}[\varepsilon|\mathcal{J}] = 0$ and $\mathbb{E}[\varepsilon|X] \neq 0$. Therefore, for any variable $Z \in \mathcal{J}$,

$$\mathbb{E}[\varepsilon|Z] = 0. \tag{3}$$

An alternative way to derive the statistical model defined by equations (2) and (3) is to assume that the decision maker observes the true value of some characteristic, X^*, and selects her preferred alternative on the basis of the utility function

$$U = \beta X^* + \nu, \tag{4}$$

Suppose that, instead of X^*, the econometrician observes some random variable X such

[6]For notational simplicity: (1) we eliminate the subindex i except in cases for which it is strictly necessary; (2) we express all variables as the difference between their magnitude for alternative 1 minus their value for alternative 0.

that $\epsilon = X - X^*$, so that we can rewrite the payoff function as:

$$U = \beta X + \nu + \varepsilon, \tag{5}$$

where $\varepsilon = -\beta\epsilon$. If we impose the classical errors-in-variables assumptions, it holds that $\mathbb{E}[\varepsilon|X^*] = 0$ and $\mathbb{E}[\varepsilon|X] \neq 0$, and we define a variable Z as an instrumental variable (IV) if and only if it verifies

$$\mathbb{E}[\varepsilon|Z] = 0. \tag{6}$$

Equations (2) and (3) are identical to equations (5) and (6). The interpretation of the expectational error of an agent with rational expectations as a special case of the measurement error in a classical errors-in-variables model is obtained when we think of the expectation as the unobserved true covariate (i.e. $\mathcal{E}[X|\mathcal{J}] = X^*$) and of the realized future value as the observed mismeasured covariate (i.e. X).[7]

The difference between the rational expectations assumption and the classical errors-in-variables assumption is that the former implies a stronger orthogonality condition. While the classical errors-in-variables assumption exclusively imposes $\mathbb{E}[\varepsilon|X^*] = 0$, the rational expectations assumption implies $\mathbb{E}[\varepsilon|\mathcal{J}] = 0$, with the set $\mathcal{J}$ including but not necessarily restricted to the variable $X^* = \mathcal{E}[X|\mathcal{J}]$. For example, the interpretation of ε as expectational error implies that $\mathbb{E}[\varepsilon|\nu] = 0$. On the contrary, this orthogonality restriction is not imposed by the interpretation of ε as measurement error. Analogously, as long as $Z \in \mathcal{J}$, equation (3) is an implication of the rational expectations assumption. However, equation (6) is an additional restriction that is not automatically implied by the classical errors-in-variables assumption. Given that the notion of measurement error implies fewer assumptions than that of expectational error, we present our model in Section 2.2 in terms of the former and we make explicit which additional assumptions we need for our estimator to identify the true value of the parameter vector β.

2.2 Model

In this section, we introduce a binary choice model in which agents' choices depend on three components: an *index function* that is linear in observable variables, structural error, and measurement error.

[7] As the comparison above shows, ε potentially captures both reporting error (pure measurement error) and differences between agents' expectations of each alternative's characteristics and the actual values of these characteristics (expectational error). In this case $\varepsilon = -\beta[(X^{obs} - X) + (X - \mathcal{E}[X|\mathcal{J}])]$, where X^{obs} is the observed measure of the realization of the variable X.

Agents' decisions. For each individual i, we normalize the utility of one alternative to zero and express the utility of the other alternative as

$$U = \beta X^* + \nu, \tag{7}$$

where $\beta \in \Gamma_\beta \in R^K$, $X^* \in \mathcal{X}^* \in R^K$ and $\nu \in R$. We define a dummy variable d as $d = \mathbb{1}\{U \geq 0\}$, where $\mathbb{1}\{\cdot\}$ is the indicator function that takes a value of 1 when $\{\cdot\}$ is true and a value of 0 otherwise. Therefore, we can write the individual revealed preference inequality as:

$$(d - (1 - d)) \cdot (\beta X^* + \nu) \geq 0. \tag{8}$$

The vector β groups the index coefficients and is defined up to a scalar normalization (possibly involving the distribution of ν). The term βX^* is the single index determining the choice dummy, d. Both X^* and ν belong to the information set of the agent at the time she makes a decision: $(X^*, \nu) \in \mathcal{J}$.

Measurement model. The econometrician does not observe ν and might observe X^* with error. We denote Z_1 as the $P \times 1$ subvector of X^* that is measured without error,

$$Z_1 = X_1^*, \tag{9}$$

and X_2^* as the $(K - P) \times 1$ subvector of X^* that may be measured with error

$$X = X_2^* + \epsilon, \tag{10}$$

where $\epsilon \in R^{K-P}$, and $0 \leq P \leq K$.[8] Once we incorporate equations (9) and (10) into equation (8), the individual-level revealed preference inequality becomes

$$(d - (1 - d)) \cdot (\beta_1 Z_1 + \beta_2 X + \nu + \varepsilon) \geq 0, \tag{11}$$

with $\varepsilon = -\beta_2 \epsilon$. In equation (11), only the vector (d, Z_1, X) is observed by the econometrician.[9] The econometrician also observes a vector of instrumental variables, $Z_2 \in \mathcal{Z}_2 \in R^T$, $T \geq K - P$. We stack both the true covariates that are observed without

[8]This model is consistent with either all or none of the K covariates being observed with error.

[9]This measurement model assumes that X and Z_1 are observed independently of the alternative chosen by the agent. Section 6 considers an alternative measurement model in which, for each agent, we only observe X if she chooses one particular alternative (i.e. exports).

error and the instruments into a single vector $Z = (Z_1, Z_2)$, $Z \in \mathcal{Z} \in R^{P+T}$.

For our moment inequality estimator, a key restriction of equation (7) is that the index function, βX^*, is linear in the unobserved true covariate, X_2^*. While equation (7) assumes that this index function is also linear in the observed true covariates, X_1^*, our moment inequality estimator also applies in cases in which this index function is nonlinear in X_1^*.

Assumptions on error terms. We impose the following assumptions on the distribution of (ν, ε) conditional on the vector (d, X^*, Z_2, X):

Assumption 1 *The random variable ν is independent of the random vector (X^*, Z_2); i.e. $F_\nu(\nu|X^*, Z_2) = F_\nu(\nu)$.*

Assumption 2 *The marginal distribution function of ν is known up to a scale parameter, log concave, has mean zero, and, for any y in the support of ν, verifies the following property:*

$$\frac{\partial^2 \mathbb{E}[\nu|\nu \geq y]}{\partial y^2} \geq 0.$$

Assumption 3 *The distribution of ε conditional on (d, X^*, Z_2) has support $(-\infty, \infty)$ and has mean zero; i.e. $\mathbb{E}[\varepsilon|d, X^*, Z_2] = 0$.*

As in the standard binary probit or logit model, Assumption 1 imposes independence between the structural error, ν, and the true vector of covariates, X^*. Therefore, without expectational error, the error term in equation (11) is exogenous to the vector of observed covariates. Assumption 1 also imposes that the observed vector of instruments, Z_2, is independent of the structural error.

Assumption 2 implies that the distribution of the structural error is known to the econometrician and restricts this distribution to be log concave and have a right-truncated expectation that is convex in the truncation point.[10] Both the normal and

[10]We can relax Assumption 2 and assume instead that the econometrician does not know the true distribution F_ν but just a finite set of distributions $\mathcal{F}_\nu$ to which F_ν belongs. In this case, the econometrician may generate different distribution-specific identified sets for β for each $F_\nu \in \mathcal{F}_\nu$. The union of these distribution-specific identified sets incorporates the econometrician's uncertainty in F_ν.

logistic densities are log concave.[11] As Heckman and Honoré (1990) show, not every log concave distribution has a right-truncated expectation that is convex in the truncation point. Nevertheless, the following remark clarifies that both the normal and the logistic distribution are consistent with Assumption 2.

Remark 1 *If $\nu \sim \mathbb{N}(0, \sigma^2)$, for any $\sigma^2 < \infty$; or $\nu \sim$ Logistic, then $\frac{\partial^2 \mathbb{E}[\nu|\nu \geq y]}{\partial y^2} \geq 0$, for any $y \in (-\infty, \infty)$.*

The proof of Remark 1 is contained in Section A.1. An implication of both Assumptions 1 and 2 is that $\mathbb{E}[\nu|X^*, Z_2] = 0$.

Assumption 3 does not impose a parametric restriction on the distribution of the expectational error, ϵ. In contrast, as discussed in Section 4.3, estimating this binary discrete choice model using maximum-likelihood techniques would require assuming both the distribution of the measurement error conditional on the true covariates as well as the marginal distribution of these unobserved true covariates.

In addition, Assumption 3 does not require full independence between the measurement error and the vector of decisions, true covariates, and instruments. This is particularly important for the case in which the error ε captures expectational error and agents have rational expectations. In that case, given that (X^*, d) $\in \mathcal{J}$ and, by definition, $\mathbb{E}[\varepsilon|\mathcal{J}] = 0$, Assumption 3 can be simplified to $\mathbb{E}[\varepsilon|Z_2] = 0$. This condition will be satisfied as long as the econometrician defines the vector of instruments, Z_2, using variables included in $\mathcal{J}$.

The data are informative about the joint density of (d, Z, X), $\mathcal{P}(d, Z, X)$. Any structure $S^a \equiv \{\beta^a,\ f^a(X|X^*, Z_2)\ f^a(Z_2|X^*)\ f^a(X^*)\}$ is admissible as long as it verifies the restriction in Assumptions 1 to 3 and

$$\mathcal{P}(d, Z, X) = \int f(d|X^*, X; \beta^a) f^a(X|X^*, Z_2) f^a(Z_2|X^*) f^a(X^*) dX^*.$$

In the model described in this section, the parameter vector β is set—not point—identified, because there exists at least two admissible structures S^{a_1} and S^{a_2} such that

[11] A random variable y has a log concave distribution if its density function f_y satisfies that $f_y(\lambda y_1 + (1-\lambda)y_2) \geq [f_y(y_1)]^\lambda\ [f_y(y_2)]^{1-\lambda}$, $0 \leq \lambda \leq 1$, for any given values y_1 and y_2 in the support of y. Some general references on log concave density functions are Pratt (1981), Heckman and Honoré (1990), and Bagnoli and Bergstrom (2005). Heckman and Honoré (1990) clarify that the class of log concave densities also includes uniform, exponential, extreme value and laplace (or double exponential) densities. Under some parameter restrictions, it also includes the power function, Weibull, gamma, chi-squared and beta.

$\beta^{a_1} \neq \beta^{a_2}$. For a restricted version of this model, Appendix A.2 characterizes the set of structures S^a that are consistent with the joint density $\mathcal{P}(d, Z, X)$, and shows that these structures imply different values of the parameter vector β.

Although Assumptions 1 to 3 do not point-identify β, they have nontrivial identification power. In Section 3, we derive moment inequalities based on these assumptions; these inequalities define a set that contains the true value of β and is strictly contained in R^K.

2.3 Related Literature

Section 2.2 introduces a binary choice model in which the observable covariates are endogenous due to the existence of classical errors-in-variables. Our paper is not the first to study identification of index-function parameters in a binary choice model with endogenous observable covariates. But we are, to our knowledge, the first to examine the identifying properties of Assumptions 1 to 3 in an economic model that contains two distinct errors with different economic interpretation.

There are three alternative models that build on conditional independence assumptions to identify the parameters of binary choice models with endogenous regressors: (1) the IV model of Chesher (2010) and Chesher (2011); (2) the triangular system model that motivates the use of control function methods; and, (3) the special regressor approach.

As Blundell and Powell (2003) show, even when the econometrician observes an excluded variable that is independent of the error term in the random utility function, semi-parametric and non-parametric binary response models are generally not point identified. Chesher (2010) shows that this result holds even if we impose parametric restrictions both on the random utility function and on the marginal distribution of the error term. Chesher (2010) provides the inequalities that sharply define the identified set under the assumption that the econometrician observes a excluded variable that is independent of the error term (i.e. fully independent instrument). While Chesher (2010) focuses on the case in which the endogenous variable is continuous, Chesher (2011) performs an analogous exercise for the case in which it is discrete.[12] Following our notation, Chesher (2010) and Chesher (2011) assume that $(\nu + \varepsilon)|Z \sim (\nu + \varepsilon)$.

Our model is stricter than the one proposed in Chesher (2010) in that we formally

[12] Other papers that explore this IV approach are Chesher and Smolinski (2010), Chesher et al. (2011), Chesher and Rosen (2012).

define the error term as the sum of two different unobserved components: structural error, ν, and expectational error, ϵ; we only allow for endogeneity that is due to expectational error. However, our model can be viewed as more flexible than that in Chesher (2010) because our identification strategy does not assume that the aggregate error term, $(\nu + \epsilon)$, is fully independent of the instrument vector, Z_2. We only need to impose mean independence between this instrument and ϵ. This weaker independence assumption of our statistical model matches the assumptions common to economic models of agents with rational expectations.[13]

The triangular system control function model is attractive because, under certain conditions, point identifies the parameters of interest. In particular, Blundell and Powell (2004) obtain point identification by applying a control-function approach. This approach assumes that the endogenous variables are determined by an equation $X = h(Z, e)$ such that there is a one-to-one mapping from the latent variables, e, to the endogenous variables, X, at each value of the instrument vector, Z.[14] Using the notation introduced in Section 2.2, the latent variable e is assumed to verify: $(\nu + \varepsilon)|X, Z \sim (\nu + \varepsilon)|X, e \sim (\nu + \varepsilon)|e$. In contrast, our model is a single-equation model: there is no specification of any structural equation that would imply that the error term $(\nu + \varepsilon)$ is independent of the endogenous regressor X, conditional on some latent variable, e.[15]

The special regressor approach assumes that the aggregate unobservable component, $(\nu + \epsilon)$, is distributed independently of a continuously distributed explanatory variable (i.e. special regressor) and impose a particular index restriction (Lewbel (2000)). This model is point identified only if the special regressor has large support. In an application in which the only source of endogeneity is expectational error, this approach implies that one covariate, the special regressor with large support, is measured without error.[16] Our statistical model allows all the regressors to contain expectational error.

[13] The only type of independence that the rational expectations assumption imposes on the definition of the expectational error is mean independence between this error and any variable contained in the information set of the agent.

[14] This restriction rules out cases in which there are discrete endogenous variables. In our case, we allow for discrete endogenous variables as long as the measurement error can verify the restriction in Assumption 3.

[15] Other papers that explore the use of control function methods for the identification of binary choice models in semi- and non-parametric settings are Blundell and Powell (2003), Chesher (2003), Chesher (2005), Chesher (2007), Vytlacil and Yildiz (2007), Florens et al. (2008), Imbens and Newey (2009), and Shaikh and Vytlacil (2011).

[16] Other papers that explore the special regressor approach are Magnac and Maurin (2007) and Magnac and Maurin (2008).

2.4 Simulation Exercise: Setting

We now introduce a simple setting that we will use in Sections 3 and 4 to illustrate the properties of our moment inequality estimator. The simulated data is generated by a structure in which

$$U = \beta_1 X_1^* + \beta_2 X_2^* + \beta_3 X_3^* + \nu, \tag{12}$$

with $\beta = (0.5, 0.5, 0.25)$. The error ν is distributed normally and independently of X^*: $\nu | X^* \backsim \mathbb{N}(0, 2)$.[17] We assume that X_1^* and X_3^* are measured without error: $\{Z_1, Z_3\} = \{X_1^*, X_3^*\}$. On the other hand, the variable X_2^* is measured with error: $X_2 = X_2^* + \epsilon^x$, with ϵ^x distributed normally and independently of (X^*, ν): $\epsilon^x | (X^*, \nu) \backsim \mathbb{N}(0, 0.4)$. We also repeat the estimation for values of $\sigma^2_{\epsilon^x}$ between 0.2 and 1.8, in order to study how the properties of the different estimators change as the variance of the expectational error increases. Finally, we generate our instrumental variable, Z_2, as a second measurement of X_2^*: $Z_2 = X_2^* + \epsilon^z$, with ϵ^z distributed normally and independently from (X^*, ν, ϵ^x): $\epsilon^z | (X^*, \nu, \epsilon^x) \backsim \mathbb{N}(0, 0.04)$.[18]

3 Moment Inequalities

In this section, we introduce two different types of moment inequalities that (partially) identify the vector β of index coefficients under the assumptions of the static model described in Section 2.2. Section 3.1 derives both types of inequalities conditional on the vector of instruments, Z. Section 3.2 converts these conditional moments into unconditional moments. In Section 3.3, we illustrate the identification power of these unconditional moment inequalities in our simulation exercise.

3.1 Conditional Moment Inequalities

For any given value of the instrument vector Z, we derive two different types of inequalities: *score function* moment inequalities and *revealed preference* moment inequalities.

[17] We assume that ν is the difference in the structural error between alternative 1 and alternative 0. With $\nu_1 | X^* \backsim \mathbb{N}(0, 1)$, $\nu_0 | X^* \backsim \mathbb{N}(0, 1)$, and both independent from each other, $\nu | X^* = (\nu_1 - \nu_0) | X^* \backsim \mathbb{N}(0, 2)$.

[18] For our identification approach to define a set that contains the true value of the parameter vector under Assumptions 1 to 3, we need not assume that Z_2 is a second measurement of X_2^* generated independently from X_2. The second measurement assumption here is used simply to generate the simulated data.

3.1.1 Conditional Score Function Moment Inequalities

For any $z \in \mathcal{Z}$, we define the conditional score function moment inequality as

$$\mathcal{M}_s(z;\beta) = \mathbb{E}\left[\begin{array}{c} m_s^-(d, Z_1, X; \beta) \\ m_s^+(d, Z_1, X; \beta) \end{array}\middle| Z = z\right] \geq 0,$$

where the two moment functions are defined as

$$m_s^-(d, Z_1, X; \beta) = d\frac{F_\nu\big(-(\beta_1 Z_1 + \beta_2 X)\big)}{1 - F_\nu\big(-(\beta_1 Z_1 + \beta_2 X)\big)} - (1-d), \tag{13a}$$

$$m_s^+(d, Z_1, X; \beta) = (1-d)\frac{1 - F_\nu\big(-(\beta_1 Z_1 + \beta_2 X)\big)}{F_\nu\big(-(\beta_1 Z_1 + \beta_2 X)\big)} - d. \tag{13b}$$

While m_s^+ is increasing in $\beta_1 Z_1 + \beta_2 X$, the opposite is true for m_s^-. We derive these inequalities by differentiating the log probability of $d = 1$ -hence, the label *score function* inequality.[19] We use $\mathcal{M}_s(\mathcal{Z}^M;\beta)$ to group the conditional moment inequalities corresponding to the set of values of z contained in the vector $\mathcal{Z}^M$: $\mathcal{M}_s(\mathcal{Z}^M;\beta) = \{\mathcal{M}_s(z;\beta);\text{ for all } z \in \mathcal{Z}^M\}$. We denote the identified set defined by these inequalities as $\Omega(\mathcal{M}_s, \mathcal{Z}^M) = \{\beta \in \Gamma_\beta : \mathcal{M}_s(z;\beta) \geq 0 \text{ for all } z \in \mathcal{Z}^M\}$. The following theorem contains the properties of the set $\Omega(\mathcal{M}_s, \mathcal{Z}^M)$. We use β^* to denote the true value of β.

Theorem 1 *Properties of* $\Omega(\mathcal{M}_s, \mathcal{Z}^M)$.

1. *If Assumptions 1, 2 and 3 hold, then* $\beta^* \in \Omega(\mathcal{M}_s, \mathcal{Z}^M)$, *for any* $\mathcal{Z}^M \subseteq \mathcal{Z}$ *and any* $\beta^* \in \Gamma_\beta$.

2. *If Assumptions 1 and 2 hold and* $X_2^* = X = Z_2$, *then* $\beta^* = \Omega(\mathcal{M}_s, \mathcal{Z}^M)$, *for any* $\mathcal{Z}^M \subseteq \mathcal{Z}$ *such that the rank of the* $K \cdot M$ *matrix* $(z^1, \ldots, z^M) \in \mathcal{Z}^M$ *is larger than* K, *and any* $\beta^* \in \Gamma_\beta$.

3. *If Assumptions 1, 2 and 3 hold and* $\bar{\bar{\sigma}}_\varepsilon^2 \geq \bar{\sigma}_\varepsilon^2$, *then* $\Omega(\mathcal{M}_s, \mathcal{Z}^M|\bar{\sigma}_\varepsilon^2) \subseteq \Omega(\mathcal{M}_s, \mathcal{Z}^M|\bar{\bar{\sigma}}_\varepsilon^2)$, *for any* $\mathcal{Z}^M \subseteq \mathcal{Z}$.

The proofs of the three properties of $\Omega(\mathcal{M}_s, \mathcal{Z}^M)$ in Theorem 1 are contained in Sections A.3, A.4, and A.5, respectively.[20] The first property indicates that, if Assumptions 1, 2

[19] We describe its derivation in detail in Section A.3.

[20] As Sections A.3, A.4, and A.5 show, Theorem 1 does not require that the right-truncated expectation of ν is convex. However, it still requires the distribution of ν to be log concave. The reason is

and 3 hold, then the score function inequalities contain the true value of the parameter vector, for any β^* and any valid set of instruments. In general, the set $\Omega(\mathcal{M}_s, \mathcal{Z}^M)$ will also contain values of β other than its true value. However, as the second property in Theorem 1 states, in the particular case in which there is no expectational error (i.e. X_2^* $= X$) and we use the observed vector of choice characteristics as instruments (i.e. $Z_2 =$ X), the identified set defined by the score function inequalities is a singleton, containing only the true value of the parameter vector. Finally, in those cases in which there is measurement error (i.e. $X_2^* \neq X$) and, therefore, we use a vector of instruments different from the observed covariates (i.e. $Z_2 \neq X$), the third property in Theorem 1 states that the set $\Omega(\mathcal{M}_s, \mathcal{Z}^M)$ increases monotonically in the variance of the expectational error. Specifically, every value of the vector β contained in the identified set that corresponds to a given value of the variance of the measurement error will also be contained in the identified set generated by a larger variance.

Computing the moments in equation (13) is no harder than evaluating the likelihood function for a binary choice model that assumes there is no expectational error. We describe the formation of the moments below.

3.1.2 Conditional Revealed Preference Inequalities

For any $z \in \mathcal{Z}$, we define the conditional revealed preference moment inequality as

$$\mathcal{M}_r(z;\beta) = \mathbb{E}\left[\begin{array}{c} m_r^-(d, Z_1, X;\beta) \\ m_r^+(d, Z_1, X;\beta) \end{array} \middle| Z = z\right] \geq 0,$$

where the two moment functions are defined as

$$m_r^-(d, Z_1, X;\beta) = -(1-d)(\beta_1 Z_1 + \beta_2 X) + d\mathbb{E}\big[\nu | \nu \geq -(\beta_1 Z_1 + \beta_2 X), Z_1, X\big], \quad \text{(14a)}$$
$$m_r^+(d, Z_1, X;\beta) = d(\beta_1 Z_1 + \beta_2 X) + (1-d)\mathbb{E}\big[-\nu | -\nu \geq \beta_1 Z_1 + \beta_2 X, Z_1, X\big]. \quad \text{(14b)}$$

While m_r^+ is increasing in the index $\beta_1 Z_1 + \beta_2 X$, the opposite is true for m_r^-. We derive these inequalities from the individual revealed preference inequality of the version of the model described in Section 2.2 that does not contain expectational error (see equation (8)) –hence, the label *revealed preference* inequality.[21] Analogously to the case of the score function inequality, we define $\mathcal{M}_r(\mathcal{Z}^M;\beta) = \{\mathcal{M}_r(z;\beta);$ for all $z \in \mathcal{Z}^M\}$

that the log concavity of the distribution of ν is sufficient to guarantee that both $F_\nu(\cdot)/(1-F_\nu(\cdot))$ and $(1-F_\nu(\cdot))/F_\nu(\cdot)$ are convex.

[21] We describe its derivation in detail in Section A.6.

and $\Omega(\mathcal{M}_r, \mathcal{Z}^M) = \{\beta \in \Gamma_\beta : \mathcal{M}_r(z;\beta) \geq 0 \text{ for all } z \in \mathcal{Z}^M\}$. The following theorem contains the properties of the set $\Omega(\mathcal{M}_r, \mathcal{Z}^M)$.

Theorem 2 *Properties of* $\Omega(\mathcal{M}_r, \mathcal{Z}^M)$.

1. *If Assumptions 1, 2 and 3 hold, then* $\beta^* \in \Omega(\mathcal{M}_r, \mathcal{Z}^M)$, *for any* $\mathcal{Z}^M \subseteq \mathcal{Z}$ *and any* $\beta^* \in \Gamma_\beta$.

2. *If Assumptions 1 and 2 hold and* $X_2^* = X = Z_2$, *then* $\exists\ \beta \in \Gamma_\beta$ *such that* $\beta \neq \beta^*$ *and* $\beta \in \Omega(\mathcal{M}_r, \mathcal{Z}^M)$, *for any* $\mathcal{Z}^M \subseteq \mathcal{Z}$ *and any* $\beta^* \in \Gamma_\beta$.

3. *If Assumptions 1, 2 and 3 hold and* $\bar{\bar{\sigma}}_\varepsilon^2 \geq \bar{\sigma}_\varepsilon^2$, *then* $\Omega(\mathcal{M}_r, \mathcal{Z}^M|\bar{\sigma}_\varepsilon^2) \subseteq \Omega(\mathcal{M}_r, \mathcal{Z}^M|\bar{\bar{\sigma}}_\varepsilon^2)$, *for any* $\mathcal{Z}^M \subseteq \mathcal{Z}$.

The proof of the three properties of $\Omega(\mathcal{M}_r, \mathcal{Z}^M)$ in Theorem 2 are contained in Sections A.6, A.7, and A.8, respectively. The first property indicates that, if Assumptions 1, 2 and 3 hold, then $\Omega(\mathcal{M}_r, \mathcal{Z}^M)$ contains the true value of the parameter vector, for any β^* and instrument vector. As Theorem 2.2 indicates, the set $\Omega(\mathcal{M}_r, \mathcal{Z}^M)$ will generally also contain values of β other than its true value. In fact, contrary to the case of the score function inequalities, the set identified by the revealed preference inequalities is never a singleton. Even in cases in which the econometrician assumes that the observed covariate, X, is identical to the true covariate, X_2^*, and, therefore, uses X as an instrument, it is still true that the set $\Omega(\mathcal{M}_r, \mathcal{Z}^M)$ will contain values of the parameter vector that are different from β^*. A comparison of Theorem 1.2 and Theorem 2.2 leads to the conclusion that, for any $\mathcal{Z}^M$, if $X_2^* = X = Z_2$, then $\Omega(\mathcal{M}_s, z|\sigma_\varepsilon^2 = 0) \subset \Omega(\mathcal{M}_r, z|\sigma_\varepsilon^2 = 0)$. This does not imply that, for any $\mathcal{Z}^M$, the inequalities $\mathcal{M}_r(\mathcal{Z}^M;\beta)$ are irrelevant once we include the inequalities $\mathcal{M}_s(\mathcal{Z}^M;\beta)$. On the contrary, Theorems 1 and 2 provide little guidance on how $\Omega(\mathcal{M}_s, \mathcal{Z}^M)$ compares to $\Omega(\mathcal{M}_r, \mathcal{Z}^M)$ when the variance of the expectational error is different from zero. The third property in Theorem 2 is analogous to the third property in Theorem 1.

Computing the moment inequalities in equation (14) requires evaluating the right-truncated expectation of ν at different values in its support. For both probit and logit models, the right-truncated expectation of the structural error may be computed using standard statistical packages.[22]

[22] If $\nu \sim \mathbb{N}(0, \sigma^2)$, then $\mathbb{E}[\nu|\nu \geq y] = (\sigma\phi(y/\sigma))/(1 - \Phi(y/\sigma))$. If $\nu \sim$ *Logistic*, then $\mathbb{E}[\nu|\nu \geq y] = -y\exp(y) + (1 + \exp(y))\ln(1 + \exp(y))$.

3.2 Unconditional Moment Inequalities

The moment inequalities described in equations (13) and (14) condition on particular values of the instrument vector, Z. In empirical applications in which at least one of the variables in the vector Z is continuous, the sample analogue of these moment inequalities will likely involve an average over very few observations (if any). Therefore, for estimation, it will be more useful to work with unconditional moment inequalities.

Each of the unconditional moment inequalities is defined by an *instrument function*. Here, we propose a particular set of instrument functions.[23] We identify each of these functions by a $K \times 1$ vector q of 0s and 1s. We group all these vectors q into a matrix Q that forms the standard basis in R^K. For each q, we define an instrument function:

$$\Psi_q(Z) = \Big\{ \prod_{k=1}^{K} (\mathbb{1}\{Z_k \geq 0\})^{q_k} (\mathbb{1}\{Z_k < 0\})^{1-q_k}; q_k \in \{0,1\}, k = 1, \ldots, K \Big\}, \tag{15}$$

where Z_k is the kth element of Z and q_k is the kth element of q.[24]

Using the instrument functions Ψ_q, $q \in Q$, we define Q unconditional score function moment inequalities and Q unconditional revealed preference inequalities. We denote the unconditional score function inequality defined by the instrument Ψ_q as $\mathcal{M}_s^q(\beta)$. Analogously, we denote the corresponding revealed preference inequality as $\mathcal{M}_r^q(\beta)$.

3.2.1 Unconditional Score Function Moment Inequalities

For each $q \in Q$, the resulting unconditional score function moment inequality is

$$\mathcal{M}_s^q(\beta) = \mathbb{E}\Big[\Psi_q(Z) \cdot m_s^+(d, Z_1, X; \beta) + \Psi_q(-Z) \cdot m_s^-(d, Z_1, X; \beta)\Big] \geq 0. \tag{16}$$

We group the Q unconditional score function moment inequalities into the vector $\mathcal{M}_s(\beta)$ and define an identified set for this group of inequalities: $\Omega(\mathcal{M}_s) = \{\beta \in \Gamma_\beta : \mathcal{M}_s(\beta) \geq$

[23]The unconditional moment inequalities proposed here generate a larger identified set than that defined by the conditional moments described in Section 3.1. The main advantage of the moments proposed here is its computational simplicity, while they still generate a bounded set. Papers that define unconditional moments that imply no loss of information with respect to their conditional counterpart are Amstrong (2012) and Andrews and Shi (2013). The instrument functions used in these papers are computationally very intensive in our setting.

[24]Here we are implicitly assuming that the vector Z has dimensions $K \times 1$ (i.e. $T = K - P$). As an example, in the particular case in which $K = 2$, the matrix Q defines 4 instrument functions: $\Psi_1(Z) = \mathbb{1}\{Z_1 \geq 0\}\mathbb{1}\{Z_2 \geq 0\}$, $\Psi_2(Z) = \mathbb{1}\{Z_1 \geq 0\}\mathbb{1}\{Z_2 < 0\}$, $\Psi_3(Z) = \mathbb{1}\{Z_1 < 0\}\mathbb{1}\{Z_2 \geq 0\}$, $\Psi_4(Z) = \mathbb{1}\{Z_1 < 0\}\mathbb{1}\{Z_2 < 0\}$.

$0\}$. The following theorem contains the properties of the set $\Omega(\mathcal{M}_s)$.

Theorem 3 *Properties of* $\Omega(\mathcal{M}_s)$.

1. *If Assumptions 1, 2 and 3 hold, then* $\beta^* \in \Omega(\mathcal{M}_s)$, *for any* $\beta^* \in \Gamma_\beta$.

2. *If Assumptions 1 and 2 hold,* $X_2^* = X = Z_2$, *and, for every* $q \in Q$,

$$\mathbb{E}\Bigg[\Psi_q(Z) + \Psi_q(-Z)\Bigg] \neq 0, \tag{17}$$

   then $\beta^* = \Omega(\mathcal{M}_s)$.

3. *If Assumptions 1, 2 and 3 hold and* $\bar{\bar{\sigma}}_\varepsilon^2 \geq \bar{\sigma}_\varepsilon^2$, *then* $\Omega(\mathcal{M}_s|\bar{\sigma}_\varepsilon^2) \subseteq \Omega(\mathcal{M}_s|\bar{\bar{\sigma}}_\varepsilon^2)$.

4. *If Assumptions 1, 2 and 3 hold and, for every* $\beta \in \Gamma_\beta$ *and* $q \in Q$,

$$\frac{\partial \mathcal{M}_s^q(\beta)}{\partial \beta_k} > 0, \qquad \textit{if } q_k = 1, \tag{18a}$$

$$\frac{\partial \mathcal{M}_s^q(\beta)}{\partial \beta_k} < 0, \qquad \textit{if } q_k = 0, \tag{18b}$$

   then $\Omega(\mathcal{M}_s)$ *is bounded and closed.*

5. *If Assumptions 1, 2 and 3 hold, then* $\Omega(\mathcal{M}_s)$ *is convex.*

The first property indicates that the true value of the parameter vector is contained in the identified set defined by the Q unconditional score function moment inequalities in equation (16). This comes immediately from Theorem 1.1 and the fact that, for every $q \in Q$, we can rewrite $\mathcal{M}_s^q(\beta)$ as

$$\mathcal{M}_s^q(\beta) = \mathbb{E}\Big[\Psi_q(Z) \cdot \mathbb{E}\big[m_s^+(d, Z_1, X; \beta)\big|Z\big] + \Psi_q(-Z) \cdot \mathbb{E}\big[m_s^-(d, Z_1, X; \beta)\big|Z\big]\Big]. \tag{19}$$

For every $Z \in \mathcal{Z}$, $\mathbb{E}\big[m_s^+(d, Z_1, X; \beta^*)\big|Z\big] \geq 0$ and $\mathbb{E}\big[m_s^-(d, Z_1, X; \beta^*)\big|Z\big] \geq 0$. Therefore, $\mathcal{M}_s^q(\beta^*) \geq 0$.

The proof of Theorem 3.2 is contained in Section A.9. The condition in equation (17) establishes that, for every $q \in Q$, there is some value in the support of Z such that either $\Psi_q(Z) = 1$ or $\Psi_q(-Z) = 1$. Equation (17) guarantees that none of the Q moments described in equation (15) will be equal to 0 for lack of observations that verify the instrument function. If equation (17) holds, then Theorem 3.2 establishes that, in the particular case in which there is no expectational error and we use the observed covariate X as an instrument, Z_2, then the only value of the parameter vector β that is consistent with these Q unconditional inequalities is the true value of the parameter vector, β^*.

In cases in which $X \neq X_2^*$, the identified set $\Omega(\mathcal{M}_s)$ will generally contain values of the parameter vector β other than its true value. Theorem 3.3 indicates that the set of values that are included in $\Omega(\mathcal{M}_s)$ increases in the variance of the expectational error, σ_ε^2. This is a direct consequence of Theorem 1.3 and the unconditional inequalities being weighted averages of the conditional moment inequalities across different values of Z (equation (19)).

Equation (18) in Theorem 3.4 establishes some conditions under which the Q unconditional score function moment inequalities imply that the identified set is bounded. That is both, for every β_k, $k = 1, \ldots, K$, both $+\infty$ and $-\infty$ are excluded from $\Omega(\mathcal{M}_s)$. Equation (18) requires that the partial derivative of each moment function, $\mathcal{M}_s^q(\beta)$, with respect to each element of the parameter vector, β_k, is determined by the instrument function in equation (15). In order for equation (18) to hold, the correlation between the endogenous variable X, and its instrument, Z_2, must be sufficiently large. Therefore, this property of the identified set $\Omega(\mathcal{M}_s)$ depends on the instruments being sufficiently strong. The proof of Theorem 3.4 is contained in Section A.10.

Finally, Theorem 3.5 establishes convexity of the set $\Omega(\mathcal{M}_s)$. Convexity is an important property for for inference in partially-identified models (see Beresteanu and Molinari (2008), Kaido and Santos (2011), and references therein). The proof of Theorem 3.5 is contained in Section A.11.

3.2.2 Unconditional Revealed Preference Moment Inequalities

For each $q \in Q$, the resulting unconditional revealed preference moment inequality is

$$\mathcal{M}_r^q(\beta) = \mathbb{E}\Big[\Psi_q(Z) \cdot m_r^+(d, Z_1, X; \beta) + \Psi_q(-Z) \cdot m_r^-(d, Z_1, X; \beta)\Big] \geq 0. \tag{20}$$

We group the Q unconditional revealed preference moment inequalities into the vector $\mathcal{M}_r(\beta)$ and define an identified set for this group of inequalities: $\Omega(\mathcal{M}_r) = \{\beta \in \Gamma_\beta : \mathcal{M}_r(\beta) \geq 0\}$. The following theorem contains the properties of the set $\Omega(\mathcal{M}_r)$.

Theorem 4 *Properties of* $\Omega(\mathcal{M}_r)$.

1. *If Assumptions 1, 2 and 3 hold, then* $\beta^* \in \Omega(\mathcal{M}_r)$, *for any* $\beta^* \in \Gamma_\beta$.

2. *If Assumptions 1 and 2 hold and* $X_2^* = X = Z_2$*, then* $\exists\, \beta \in \Gamma_\beta$ *such that* $\beta \neq \beta^*$ *and* $\beta \in \Omega(\mathcal{M}_r)$*, for any* $\beta^* \in \Gamma_\beta$ *and* $\beta^* \neq 0$.

3. *If Assumptions 1, 2 and 3 hold and* $\bar{\bar{\sigma}}^2_\varepsilon \geq \bar{\sigma}^2_\varepsilon$*, then* $\Omega(\mathcal{M}_r|\bar{\sigma}^2_\varepsilon) \subseteq \Omega(\mathcal{M}_r|\bar{\bar{\sigma}}^2_\varepsilon)$.

4. *If Assumptions 1, 2 and 3 hold and, for every* $\beta \in \Gamma_\beta$ *and* $q \in Q$,

$$\frac{\partial \mathcal{M}_r^q(\beta)}{\partial \beta_k} > 0, \qquad \text{if } q_k = 1, \tag{21a}$$

$$\frac{\partial \mathcal{M}_r^q(\beta)}{\partial \beta_k} < 0, \qquad \text{if } q_k = 0, \tag{21b}$$

   then $\Omega(\mathcal{M}_r)$ *is bounded and closed.*

5. *If Assumptions 1, 2 and 3 hold, then* $\Omega(\mathcal{M}_r)$ *is convex.*

Properties 1, 3, 4, and 5 are identical to those described in Theorem 3 for the case of the set $\Omega(\mathcal{M}_s)$. We refer to the comments in Section 3.2.1 and proofs in Sections A.10 and A.11.

Among the list of characteristics listed in Theorems 3 and 4, the only difference between the sets $\Omega(\mathcal{M}_s)$ and $\Omega(\mathcal{M}_r)$ is the identification properties of these two sets in the case in which the econometrician assumes that there is no expectational error in covariates, $X_2^* = X$, and, accordingly, uses such observed covariates as arguments of the instrument function. While the set defined by the unconditional score function inequalities collapses to a singleton containing only β^* (see Theorem 3.4), the set defined by the unconditional revealed preference inequalities will include values of the parameter vector other than its true value (see Theorem 4.4). This is an immediate consequence of Theorem 2.2 and that each moment $\mathcal{M}_r^q(\beta)$ is a weighted average of a particular set of conditional moments, $\mathcal{M}_r(z;\beta)$.

3.2.3 Combining Score Function and Revealed Preference Inequalities

We define a vector $\mathcal{M}(\beta)$ that groups the Q unconditional score function moment inequalities and the Q unconditional revealed preference inequalities. We denote these $2Q$ inequalities as $\mathcal{M}(\beta)$; i.e. $\mathcal{M}(\beta) = (\mathcal{M}_s(\beta),\ \mathcal{M}_r(\beta))$. We define the identified set characterized by these moment inequalities: $\Omega(\mathcal{M}) = \{\beta \in \Gamma_\beta : \mathcal{M}(\beta) \geq 0\}$. By definition, the set $\Omega(\mathcal{M})$ will be weakly smaller than both $\Omega(\mathcal{M}_s)$ and $\Omega(\mathcal{M}_r)$; i.e. $\Omega(\mathcal{M}) \subseteq \Omega(\mathcal{M}_s)$ and $\Omega(\mathcal{M}) \subseteq \Omega(\mathcal{M}_r)$. The following theorem describes some properties of the set $\Omega(\mathcal{M})$.

Theorem 5 *Properties of* $\Omega(\mathcal{M})$.

1. *If Assumptions 1, 2 and 3 hold, then* $\beta^* \in \Omega(\mathcal{M})$, *for any* $\beta^* \in \Gamma_\beta$.

2. *If Assumptions 1 and 2 hold,* $X_2^* = X = Z_2$, *and, for every* $q \in Q$,

$$\mathbb{E}\Bigg[\Psi_q(Z) + \Psi_q(-Z)\Bigg] \neq 0, \tag{22}$$

then $\beta^* = \Omega(\mathcal{M})$.

3. *If Assumptions 1, 2 and 3 hold and* $\bar{\bar{\sigma}}_\varepsilon^2 \geq \bar{\sigma}_\varepsilon^2$, *then* $\Omega(\mathcal{M}|\bar{\sigma}_\varepsilon^2) \subseteq \Omega(\mathcal{M}|\bar{\bar{\sigma}}_\varepsilon^2)$.

4. *If Assumptions 1, 2 and 3 hold and, for every* $\beta \in \Gamma_\beta$ *and* $q \in Q$,

$$\max\left\{\frac{\partial \mathcal{M}_s^q(\beta)}{\partial \beta_k}, \frac{\partial \mathcal{M}_r^q(\beta)}{\partial \beta_k}\right\} > 0, \qquad \text{if } q_k = 1, \tag{23a}$$

$$\min\left\{\frac{\partial \mathcal{M}_s^q(\beta)}{\partial \beta_k}, \frac{\partial \mathcal{M}_r^q(\beta)}{\partial \beta_k}\right\} < 0, \qquad \text{if } q_k = 0, \tag{23b}$$

then $\Omega(\mathcal{M})$ *is bounded and closed.*

5. *If Assumptions 1, 2 and 3 hold, then* $\Omega(\mathcal{M})$ *is convex.*

Theorem 5 is a direct implication of Theorems 3 and 4 and the vector $\mathcal{M}(\beta)$ including both the unconditional score and revealed preference inequalities.

Given that both the unconditional score and revealed preference inequalities are satisfied at the true value of the parameter vector, it holds that this true value is contained in the set $\Omega(\mathcal{M})$ (Theorem 5.1). Given that $\Omega(\mathcal{M}) \subseteq \Omega(\mathcal{M}_s)$ and $\beta^* = \Omega(\mathcal{M}_s)$ in those cases in which $X_2^* = X = Z_2$, it is immediate that $\Omega(\mathcal{M})$ will also exclusively contain the true value of the parameter vector, β^*, in these same cases (Theorem 5.2). As Theorems 1.3 and 2.3 indicate, both the conditional score function and revealed preference moment inequalities become "weaker" as the variance of the expectational error increases. More specifically, each moment in the vector $\mathcal{M}(\beta)$ is a weighted average of a subset of conditional score function and revealed preference inequalities. Therefore, as the variance of the expectational error gets bigger, the set $\Omega(\mathcal{M}_s)$ will also include more values of the parameter vector β (Theorem 5.3). Theorem 5.4 is a direct consequence of Theorems 3.4 and 4.4; given that $\Omega(\mathcal{M}) \subseteq \Omega(\mathcal{M}_s)$ and $\Omega(\mathcal{M}) \subseteq \Omega(\mathcal{M}_r)$, it is enough that either $\Omega(\mathcal{M}_s)$ or $\Omega(\mathcal{M}_r)$ are closed and bounded for

$\Omega(\mathcal{M})$ to be closed and bounded. Finally, Theorem 5.5 is just an immediate implication of all the unconditional score function and revealed preference inequalities being convex.

3.3 Simulation Exercise

Table 1 contains the identified sets $\Omega(\mathcal{M}_s)$, $\Omega(\mathcal{M}_r)$ and $\Omega(\mathcal{M})$ for our simulation exercise. For all three sets, we assume that the econometrician knows the exact distribution of the structural error. We compute the maximum and minimum value of each of the three parameters that belongs to the three identified sets.

Table 1: Identified Set

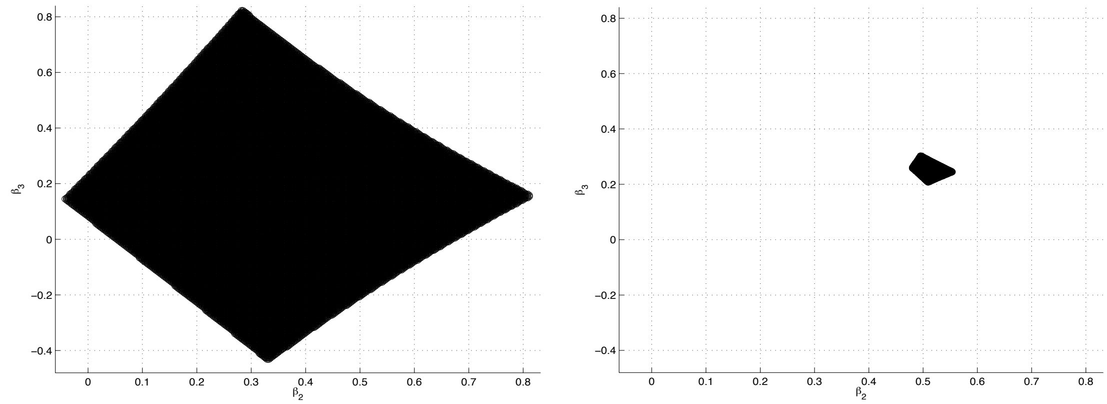
Set	$\min(\beta_1)$	$\max(\beta_1)$	$\min(\beta_2)$	$\max(\beta_2)$	$\min(\beta_3)$	$\max(\beta_3)$
$\Omega(\mathcal{M}_r)$	-0.47	1.17	-0.21	0.90	-0.61	0.95
$\Omega(\mathcal{M}_s)$	0.44	0.56	0.48	0.56	0.20	0.31
$\Omega(\mathcal{M})$	0.44	0.56	0.48	0.56	0.20	0.31

In Table 1, the true value of the parameter vector, $(0.5, 0.5, 0.25)$, lies within the smallest square circumscribing each of the three identified sets. The set $\Omega(\mathcal{M}_r)$ is much larger than the set $\Omega(\mathcal{M}_s)$; $\Omega(\mathcal{M})$ and $\Omega(\mathcal{M}_s)$ are virtually identical. For example, the set $\Omega(\mathcal{M}_r)$ indicates that the true value of β_1 lies on the interval $(-0.47,\ 1.17)$, $\Omega(\mathcal{M}_s)$ and $\Omega(\mathcal{M})$ indicate that it lies on the interval $(0.44, 0.56)$. In sum, the revealed preference moment inequalities are far less informative about the true value of the parameter vector than the score function inequalities. We illustrate the relative size of these two identified sets in a 2-dimensional space in Figure 1.[25]

We illustrate the predicted probabilities in Figure 2. In this figure, we report both the true probability as well as the minimum and maximum predicted probability from the model, using the "true" covariates for a particular agent and the set of parameter values contained in the identified sets $\Omega(\mathcal{M}_s)$, $\Omega(\mathcal{M}_r)$ and $\Omega(\mathcal{M})$. In Figure 2, we fix (X_1^*, X_3^*) to arbitrary values and plot the minimum, maximum, and the true probabilities in our simulation for a range of values of the X_2^* covariate. Combining the revealed preference and score function moments yields a range of predicted probabilities that closely follows the true probabilities.

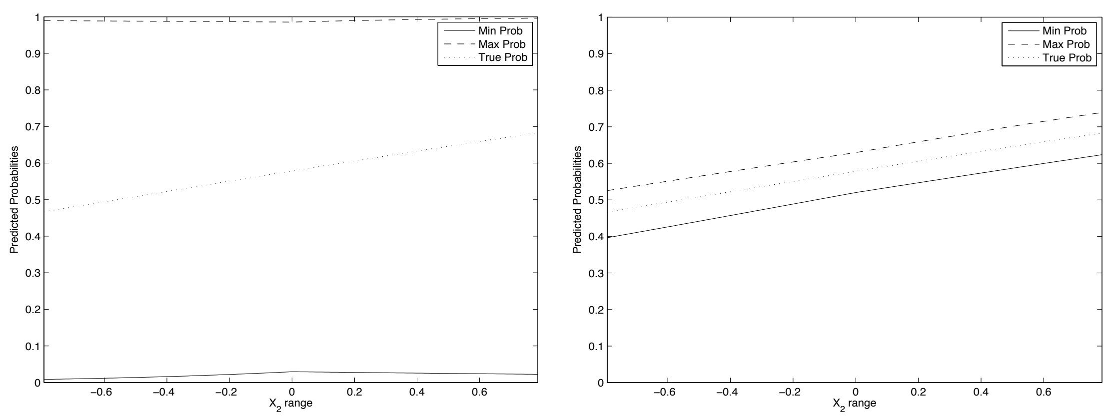
[25]These figures provide a slice of the 3-dimensional identified set, whose end points are reported in Table 1. We project this 3-dimensional object into the 2-dimensional (β_2, β_3) space, and draw the corresponding region.

Figure 1: Identified Set



(a) Revealed Preference (b) Score Function

Figure 2: Predicted Probability



(a) Revealed Preference (b) Score Function

Property 3 in Theorem 5 states that the size of the identified set $\Omega(\mathcal{M})$ increases as the variance of the expectational error increases. We show this in Table 2. The bounds in this table use both the score function and the revealed preference inequalities. In order to compute each row of this table, we have generated samples of $100,000$ observations using a statistical model identical to that described in Section 2.4, but with the variance of the expectational error, $\sigma^2_{\epsilon^x}$, set to equal the value indicated in the first column. Table 2 shows that, for each element of the parameter vector β, the lower bound diminishes and the upper bound increases as we set the variance to larger values in our simulation exercise.

Table 2: Identified Set for Different Variances

$\sigma^2_{\varepsilon^x}$	min(β_1)	max(β_1)	min(β_2)	max(β_2)	min(β_3)	max(β_3)
0.2	0.47	0.53	0.49	0.53	0.22	0.29
0.4	0.44	0.56	0.47	0.56	0.20	0.31
0.6	0.41	0.60	0.46	0.61	0.17	0.34
0.8	0.39	0.64	0.45	0.70	0.13	0.37
1.0	0.36	0.67	0.44	0.96	0.10	0.40
1.2	0.33	0.71	0.43	0.98	0.07	0.43
1.4	0.30	0.75	0.42	1.00	0.04	0.46
1.6	0.20	0.79	0.41	1.02	-0.07	0.49

4 Potential Misspecification of the Model

In spite of the empirical relevance of the statistical model described in Section 2.2, to our knowledge, the previous literature does not provide a valid estimator for such a model. In this section we consider three possible deviations from the model described in Section 2.2 that appear in the literature. These alternative specifications substitute one of the three assumptions of our binary choice model for a more restrictive alternative assumption. These additional restrictions lead to a different identification and estimation approach from the moment inequality estimator introduced in Section 3. In this section, we explore the properties of these alternative procedures as ways to identify the parameter vector β for the binary choice model described in Section 2.2. We show that, if the model described in Section 2.2 is the true model, assuming a simplified version of it may lead to large biases in the identification of the index coefficients and in the prediction of the choice probabilities.

Section 4.1 discusses the case in which the variable Z selected as an instrument by the researcher does not satisfy Assumption 3; i.e. $\mathbb{E}[\varepsilon|Z] \neq 0$. Section 4.2 considers the properties of those estimators that rely on the assumption that there is no difference in the structural error across choices that drives the individual's decision; i.e. $\nu = 0$. Section 4.3 studies the properties of the typical estimation procedure used in those cases in which Assumption 3 is substituted by the stronger assumption that there is no expectational error; i.e. $\varepsilon = 0$. We study each of these cases in turn.

4.1 Invalid Instrument

Even if the researcher believes that the statistical model described in Section 2.2 is adequate to capture the main features of her empirical setting, three issues may complicate the application of the inequality estimator described in Section 3. First, there may be no observed vector of excluded variables, Z_2, in the dataset available. Second, even if such a vector of excluded variables exist, the correlation between these variables and the mismeasured vector of covariates, X, might be so small such that the resulting identified set $\Omega(\mathcal{M})$ is unbounded (i.e. the restrictions in equation (23) fail to hold). Finally, there may be cases in which the econometrician wrongly assumes that X_2 has no expectational error (i.e. assumes that $X_2^* = X$). In these three cases, the econometrician may decide to include the endogenous variable X as an argument of the instrument function, $\{\Psi_q;\, q = 1, \ldots, Q\}$, in place of Z_2. Implicitly, the econometrician substitutes Assumption 3 by the following alternative assumption:

Assumption 3(b) *The distribution of ε conditional on (X, ν) has support equal to $(-\infty, \infty)$ and expectation equal to 0:* $\mathbb{E}[\varepsilon|X, \nu] = 0$.

Assumption 3(b) imposes that the expectational error is mean independent of the observed covariate instead of being mean independent of true (unobserved) covariate. Given this assumption, one can define an identified set that contains β^* without the instrument vector Z. We can write a vector instrument functions $\{\Psi_q;\, q = 1, \ldots, Q\}$ identical to that in equation (15) that takes the vector X as their argument. The resulting moment inequalities are:

$$\mathbb{M}_s^q(\beta) = \mathbb{E}\Big[\Psi_q((Z_1, X)) \cdot m_s^+(d, Z_1, X; \beta) + \Psi_q((-Z_1, -X)) \cdot m_s^-(d, Z_1, X; \beta)\Big] \geq 0,$$
$$\mathbb{M}_r^q(\beta) = \mathbb{E}\Big[\Psi_q((Z_1, X)) \cdot m_r^+(d, Z_1, X; \beta) + \Psi_q((-Z_1, -X)) \cdot m_r^-(d, Z_1, X; \beta)\Big] \geq 0.$$

for $q = \{1, \ldots, Q\}$. We use $\mathbb{M}(\beta)$ to denote the $2 \cdot Q$ moment inequalities whose instrument function depends on the observed vector of covariates (Z_1, X) (contrary to $\mathcal{M}(\beta)$, which identifies the inequalities described in Section 3 and whose instrument function depends on $Z = (Z_1, Z_2)$). We denote $\Omega(\mathbb{M})$ as the set of values of $\beta \in \Gamma_\beta$ that are consistent with the moments identified by $\mathbb{M}(\beta)$. Analogously, we can define $\Omega(\mathbb{M}_s)$ and $\Omega(\mathbb{M}_r)$.

If Assumptions 1, 2 and 3(b) hold, then $\Omega(\mathbb{M})$ contains the true value of the parameter vector and is convex, bounded, and closed. If Assumption 3 holds instead of

Assumption 3(b), then the non-zero correlation between the expectational error affecting X and the instrument vector, $\{\Psi_q((X,Z)); q = 1,\ldots,Q\}$, will affect the moments in the vector $\mathbb{M}(\beta)$. We use our simulation to explore the properties of the identified set, $\Omega(\mathbb{M})$, when Assumption 3 holds.

Table 3: Identified Set

Set	$\min(\beta_1)$	$\max(\beta_1)$	$\min(\beta_2)$	$\max(\beta_2)$	$\min(\beta_3)$	$\max(\beta_3)$
$\Omega(\mathbb{M}_r)$	-0.46	1.16	-0.21	0.81	-0.59	0.94

The results from the simulation appear in Table 3. In this example, $\Omega(\mathbb{M}_s) = \Omega(\mathbb{M}) = \emptyset$. At least two of the score function moments cross each other and, therefore, there is no value of the parameter vector that satisfies all score function moments. This shows that, given the statistical model described in Section 2.2, the true value of the parameter vector is not consistent with the vector of moments $\mathbb{M}(\beta)$.[26] In sum, if Assumption 3 holds instead of Assumption 3(b), we cannot guarantee that $\Omega(\mathbb{M}_s)$ or $\Omega(\mathbb{M}_r)$ contain the true value of the parameter β.

4.2 No Structural Error

The previous empirical literature that applies moment inequalities to identify the parameters of a discrete choice model often explicitly or implicitly rules out an idiosyncratic structural error. That is, the econometrician assumes that the unobserved components of the agent's utility function are identical across choices: $\nu = 0$. Examples of empirical applications using moment inequalities that impose this assumption include Ho (2009), Pakes (2010), Pakes et al. (2011), Holmes (2011), and Morales et al. (2011).[27]

[26] We also find that $\beta^* \in \Omega(\mathbb{M}_r)$. However, $\Omega(\mathbb{M}_r)$ is smaller than $\Omega(\mathcal{M}_r)$. The fact that $\Omega(\mathbb{M}_r)$ contains the true value of the parameter vector is due to the weak identification power of the revealed preference moment inequalities. That is, the revealed preference inequalities admit a wide range of β values in the identified set.

[27] Most of these papers consider empirical applications in which the choice set includes more than two options. Therefore, the results presented in Section 3 are not immediately applicable to the settings considered in these papers; we are working on extending our results to discrete choice models with more than two choices. The assumption imposing that $\nu = 0$ has been weakened in applications of moment inequalities to ordered choice models (see Katz (2007), Ishii (2008), Pakes et al. (2011)) and to the estimation of dynamic games (see Ciliberto and Tamer (2009)). Modelling a particular discrete choice decision as an ordered choice model implies that the error term ν can modeled as $j \cdot \nu$. In these models, the index j should have a quantitative interpretation; its applicability to general

In order to derive score function and revealed preference moment inequalities under the assumption of no structural error, we work with a statistical model that is equivalent to that in Section 2.2, except that we replace Assumption 2 with the following alternative assumption:

Assumption 2(b) *The marginal distribution function of ν has a degenerate distribution at $\nu = 0$.*

If Assumptions 1, 2(b) and 3 hold, then the score function moment inequality (equation (16)) has no identification power.[28] As a consequence, for each $q \in Q$, the only inequality with identification power is the following revealed preference inequality:

$$\mathfrak{M}_r^q(\beta) = \mathbb{E}\Big[-\Psi_q(Z)(1-d)(\beta_1 Z_1 + \beta_2 X) + \Psi_q(-Z)d(\beta_1 Z_1 + \beta_2 X)\Big] \geq 0. \tag{24}$$

and the identified set defined by these Q moment inequalities is $\Omega(\mathfrak{M}_r)$. If Assumptions 1, 2(b) and 3 hold, then the identified set $\Omega(\mathfrak{M}_r)$ contains the true value of the parameter vector and is convex, bounded, and closed.[29]

We use our simulation to explore the properties of the identified set $\Omega(\mathfrak{M}_r)$ in settings in which Assumption 2 holds and Assumption 2(b) does not. That is, we examine the coverage properties of the set $\Omega(\mathfrak{M}_r)$ in settings in which the structural error varies across choices but the econometrician wrongly assumes that it is identical across choices. Note that, for every $q \in Q$, we can write $\mathcal{M}_r^q(\beta) = \mathfrak{M}_r^q(\beta) + \gamma^q(\beta)$ and, for every value of $\beta \in \Gamma_\beta$ and every $q \in Q$, $\gamma^q(\beta) \geq 0$.[30] Therefore, $\Omega(\mathfrak{M}_r) \subset \Omega(\mathcal{M}_r)$ and we cannot prove that, if Assumptions 1, 2 and 3 hold (and Assumption 2(b) does not) β^* is contained in $\Omega(\mathfrak{M}_r)$. In words, when we wrongly assume that there is no structural error, the identified set defined by the resulting moment inequalities is biased inwards. This can lead to a false sense of precision. If Assumption 2(b) does not hold,

discrete choice problems is thus limited. In the application in Ciliberto and Tamer (2009), the moment inequalities estimator allows a structural error but contains no expectational error. We consider the effect of incorrectly assuming that there is no expectational error in Section 4.3.

[28] Proof in Section A.12.

[29] More precisely, the set $\Omega(\mathfrak{M}_r)$ is a K-dimensional cube. The "flat faces" in the boundary of the identified set is a direct implication of the linearity of the moment inequalities in the parameter vector β (see Kaido and Santos (2011)).

[30] The function $\gamma^q(\beta)$ is a weighted average of truncated expectations and works as a structural error correction term. Its exact expression is:

$$\mathbb{E}\Big[\Psi_q(Z)d\mathbb{E}[\nu|\nu \geq -(\beta_1 Z_1 + \beta_2 X), Z_1, X] + \Psi_q(-Z)(1-d)\mathbb{E}[-\nu| - \nu \geq \beta_1 Z_1 + \beta_2 X, Z_1, X]\Big].$$

the identified set $\Omega(\mathfrak{M}_r)$ may be so small that it may fail to contain the true value of the parameter vector. The results from the simulation exercise demonstrate this possibility: Table 4 shows that $\Omega(\mathfrak{M}_r)$ is much smaller than $\Omega(\mathcal{M}_r)$. While the resulting identified set still contains the true value of β_3, this is not true for the parameter β_2. The true value of β_2 is 0.5, while the identified set found for this parameter is $(-0.11, 0.24)$.[31]

Table 4: Identified Set

Set	$\min(\beta_1)$	$\max(\beta_1)$	$\min(\beta_2)$	$\max(\beta_2)$	$\min(\beta_3)$	$\max(\beta_3)$
$\Omega(\mathfrak{M}_r)$	0.50	0.50	-0.11	0.24	-0.21	0.28

The absence of structural error makes the resulting model fully deterministic. Therefore, for any given vector X^*, the predicted choice probability for any alternative is either 1 or 0.

4.3 No Expectational Error

Finally, we consider the potential misspecification in which the econometrician ignores expectational error. To our knowledge, no paper allows for classical errors-in-variables in a binary choice model.[32] The standard binary choice model allows for a single error term and assumes that its distribution is known up to a finite vector of parameters. Formally, the standard model estimated in the binary choice literature assumes $d \cdot (\beta X + \eta) \geq 0$, $\eta | X \sim f_\eta(\eta | X; \rho)$, where the distribution of the error term, $f_\eta(\eta | X; \rho)$, is unknown up to the parameter ρ. The probability that an individual with observable characteristics X selects choice j is equal to:

$$P(d = 1 | X) = \int_\eta \mathbb{1}\{\eta \geq -\beta X\} dF_\eta(\eta | X; \rho). \tag{25}$$

Only under very strong assumptions would this choice probability arise from a model that allows both for structural and measurement error, as in the model introduced in Section 2.2. In order to facilitate the comparison, assume the statistical model

[31] When we assume that the structural error equals 0, we first must define an alternative normalization by scale for the parameter β. For comparison purposes, we set the correct scale by fixing β_1 to be equal to its true value, 0.5.

[32] The moment inequality papers based on Pakes (2010) and Pakes et al. (2011) allow for measurement error but not for an individual and choice specific structural error (see Section (4.2)). In Berry et al. (2004), the authors allow only for measurement error in covariates that are constant across subsets of individuals.

described in equations (7) to (11) and impose the following distributional assumptions: $\nu \sim F_\nu(\nu|X^*;\rho_1)$, and $\varepsilon \sim F_\varepsilon(\varepsilon|X^*;\rho_2)$. The resulting conditional choice probabilities are

$$P(d=1|X) = \int_{x^*} \Big[\int_\nu \mathbb{1}\{\nu \geq -\beta X^*\} dF_\nu(\nu|X^*;\rho_1) \Big] dF_{x^*}(X^*|X;\rho_2) \tag{26}$$

where computing $F_{x^*}(X^*|X;\rho_2)$ requires knowledge of $F_\varepsilon(\varepsilon|X^*;\rho_2)$ and the marginal distribution of the vector of unobserved covariates, $F_{x^*}(X^*)$. Therefore, in the model in Section 2.2, in order to account for the measurement error in a maximum likelihood setting, one must make parametric assumptions on two additional distributions: (a) the distribution of the measurement error conditional on the true covariates, and, (b) the marginal distribution of the true covariates.[33]

The results of the simulation exercise show the asymptotic bias that would arise from ignoring expectational error. Table 5 presents the ML estimates for data generated following the statistical model described in Section 2.4, with the variance of the measurement error, $\sigma^2_{\epsilon^x}$, set equal to the value indicated in the first column. In order to make the results comparable, we rescale the ML estimates so that the estimates presented in every row are subject to the same normalization. The results show that as the variance of the measurement error in X_2 increases, the ML estimate of β_2 is biased downward and the ML estimate of β_3 has an upward bias.

Consider the particular case in which the variance of the expectational error is set to 0.2 (10% of the variance of the structural error). As shown in Table 5, at 0.4, the 95% confidence interval for β_2 and β_3 do not contain the true value of the parameter vector. Thus, even a relatively small variance of the measurement error might generate significant bias in the ML estimator.

[33]By redefining η as $\varepsilon + \nu$, one might be tempted to consider equation (25) as the reduced form analogue of equation (26). To do so, however, requires the researcher to impose strong independence assumptions on $F_\nu(\nu|X^*;\rho_1)$ and $F_\varepsilon(\varepsilon|X;\rho_2)$. As an example, if one assumes that $F_\nu(\nu|X^*;\rho_1) = F_\nu(\nu;\rho_1)$ and $F_\varepsilon(\varepsilon|X;\rho_2) = F_\varepsilon(\varepsilon;\rho_2)$, then equation (26) becomes

$$P(d=1|X_i) = \int_{\nu+\varepsilon} \mathbb{1}\{\nu+\varepsilon \geq -\beta X^*\} dF_\nu(\nu+\varepsilon;(\rho_1,\rho_2)), \tag{27}$$

which is equivalent to equation (25) with $F_\eta(\eta|X_i;\rho) = F_\nu(\nu+\varepsilon;(\rho_1,\rho_2))$. This approach requires the researcher to depart from the classical errors-in-variables assumption and impose that the measurement error is independent of the *observed* covariates. Only under these assumptions could we interpret the single unobserved component in standard discrete choice models as the sum of the structural and measurement error.

Table 5: Maximum Likelihood with Measurement Error

$\sigma^2_{\varepsilon^x}$	$\beta_1, 95\%\ CI$	$\beta_2, 95\%\ CI$	$\beta_3, 95\%\ CI$
0.2	[0.485,0.502]	[0.466,0.479]	[0.260,0.275]
0.4	[0.483,0.500]	[0.444,0.455]	[0.271,0.287]
0.6	[0.481,0.498]	[0.423,0.434]	[0.282,0.297]
0.8	[0.479,0.495]	[0.404,0.415]	[0.291,0.307]
1.0	[0.477,0.494]	[0.387,0.398]	[0.300,0.315]

5 Nonparametric Assumptions on Structural Errors

All the moment inequalities derived in Section 3 rely on the assumption that that the econometrician knows the marginal distribution of the structural error up to a scale parameter. In this section, we derive moment inequalities that do not require the econometrician to specify a distribution for ν. These moment inequalities hold at the true value of the parameter vector as long as Assumptions 1 and 3 hold and the structural error has zero mean. The lack of parametric assumptions on the distribution of ν impedes the derivation of the score function inequalities. However, we are still able to derive the revealed preference inequalities.

5.1 Revealed Preference Inequalities

If the econometrician does not assume a particular distribution function for ν, then she cannot compute the terms $\mathbb{E}[\nu|\nu \geq -(\beta_1 Z_1 + \beta_2 X), Z_1, X]$ and $\mathbb{E}[-\nu| - \nu \geq \beta_1 Z_1 + \beta_2 X, Z_1, X]$ in equation (14). Nevertheless, we can still find an inequality that will be satisfied at the true (normalized) value of the parameter vector. For every instrument function Ψ_q, $q = 1, \dots, Q$ (see equation (15)), we define the following inequality:

$$\mathsf{M}_r^q(\beta) = \mathbb{E}\Big[\Psi_q(Z) \cdot \mathsf{m}_r^+(d, Z_1, X; \beta) + \Psi_q(-Z) \cdot \mathsf{m}_r^-(d, Z_1, X; \beta)\Big] \geq 0, \tag{28}$$

with

$$\mathsf{m}_r^-(d, Z_1, X; \beta) = (1-d)(-(\beta_1 Z_1 + \beta_2 X)) + \mathbb{E}\big[\nu \mathbb{1}\{\nu \geq 0\}\big], \tag{29a}$$
$$\mathsf{m}_r^+(d, Z_1, X; \beta) = d(\beta_1 Z_1 + \beta_2 X) + \mathbb{E}\big[-\nu \mathbb{1}\{-\nu \geq 0\}\big]. \tag{29b}$$

We denote as $\Omega(\mathsf{M}_r)$ the identified set defined by these Q inequalities. In order to state the properties of this identified set, we first need to introduce the assumption that we substitute for Assumption 2:

Assumption 2(c) *The structural error ν has mean equal to 0; i.e.* $\mathbb{E}(\nu) = 0$.

Assumption 2(c) is more general than Assumption 2. Assumption 2(c) is compatible with any distribution for ν as long as it has zero mean, independently of whether it is log-concave or whether the truncated expectation is convex or concave in the truncation point.

Theorem 6 *If Assumptions 1, 2(c) and 3 hold, then, for any $\beta^* \in \Gamma_\beta$, $\beta^* \in \Omega(\mathsf{M}_r)$.*

The proof of Theorem 6 is contained in Appendix A.13. This theorem shows that one can derive a non-trivial identified set (i.e. a strict subset of the parameter space) based on linear moment inequalities that do not rely on parametric assumptions on the distribution of either the structural or the expectational error.

Since we do not know the marginal distribution of ν, one might question whether the inequality $\mathsf{M}_r(\beta)$ is applicable. After all, the value of $\mathbb{E}[\nu \mathbb{1}\{\nu \geq 0\}]$ is a constant that depends on this marginal distribution. However, if Assumption 2(c) holds, then $\mathbb{E}[\nu \mathbb{1}\{\nu \geq 0\}] > 0$.[34] For any distribution F_ν, the parameter vector β is identified only up to scale and, therefore, we can divide by the term $\mathbb{E}[\nu \mathbb{1}\{\nu \geq 0\}]$ on both sides of the inequality in equation (28). This is equivalent to setting the normalizing constant equal to this truncated expectation. The resulting normalized inequality is

$$\tilde{\mathsf{M}}_r^q(\beta) = \mathbb{E}\Big[\Psi_q(Z) \cdot \tilde{\mathsf{m}}_r^+(d, Z_1, X; \tilde{\beta}) + \Psi_q(-Z) \cdot \tilde{\mathsf{m}}_r^-(d, Z_1, X; \tilde{\beta})\Big] \geq 0, \tag{30}$$

with

$$\tilde{\mathsf{m}}_r^-(d, Z_1, X; \beta) = (1-d)(-(\tilde{\beta}_1 Z_1 + \tilde{\beta}_2 X)) + 1, \tag{31a}$$

$$\tilde{\mathsf{m}}_r^+(d, Z_1, X; \beta) = d(\tilde{\beta}_1 Z_1 + \tilde{\beta}_2 X) + 1. \tag{31b}$$

and the resulting identified set is $\Omega(\tilde{\mathsf{M}}_r)$. For any $\beta \in \Gamma_\beta$ such that $\beta \in \Omega(\mathsf{M}_r)$, $\tilde{\beta} = \beta / \mathbb{E}[\nu \mathbb{1}\{\nu \geq 0\}] \in \Omega(\tilde{\mathsf{M}})$. Therefore, if the assumptions in Theorem 6 hold, the true parameter vector scaled by $\mathbb{E}[\nu \mathbb{1}\{\nu \geq 0\}]$ will be in $\Omega(\tilde{\mathsf{M}})$.

In Table 6, we show that the identified set defined by the Q inequalities $\tilde{\mathsf{M}}_r^q(\beta)$, $q = 1, \ldots, Q$, contains the true value of the parameter vector, β^*.[35] A comparison of $\Omega(\tilde{\mathsf{M}}_r)$

[34] Assumption 2(c) also implies that $\mathbb{E}[\nu \mathbb{1}\{\nu \geq 0\}] = \mathbb{E}[-\nu \mathbb{1}\{-\nu \geq 0\}]$.

[35] In order to facilitate the comparison of the identified sets across the different identification procedures, Table 6 presents the results from the numerical simulation that correspond to a value of

Table 6: Identified Set

Set	min(β_1)	max(β_1)	min(β_2)	max(β_2)	min(β_3)	max(β_3)
$\Omega(\tilde{\mathsf{M}}_r)$	-1.65	5.37	-0.83	3.87	-2.28	4.14

and $\Omega(\mathcal{M}_r)$ shows that dropping the distributional assumption on the structural error results in an identified set that includes a much larger area of the parameter space. The identification power of the set $\Omega(\tilde{\mathsf{M}}_r)$ seems likely to be limited.

6 Application: Entry into Export Markets

In this empirical application, we apply the methods developed in Section 3 to revisit a question that is at the core of the international trade literature on the gravity equation: we study how a firm's entry and exit decisions vary across destination countries depending on the distance between the destination and the country of origin. We assume that firms maximize the sum of profits across all foreign markets simply by maximizing profits market by market. This allows us to treat each firms' decision to export to a given destination in each time period as a binary decision problem.

6.1 Structural Model of Entry into Export Markets

Let i index firms, c countries and t time periods. Each firm must choose whether to export to each country in each period. Let the dummy d_{ict} denote firm i's export choice with respect to country c and period t. We model the static profits of exporting as:

$$\pi_{ict} = \beta_1 R_{ict} - \beta_2 - \beta_3 D_c + \nu^{\pi}_{ict}, \tag{32}$$

where R_{ict} denotes the revenue for firm i of exporting to country c at period t, D_c denotes the physical distance between the consumers located in country c and the firm's country of origin, and $\beta_2 + \beta_3 D_c$ is the fixed cost. All of the other determinants of the static profits are captured by the unobserved term, ν^{π}_{ict}. When firms do not export, we normalize profits to zero. Because our data is limited to Chilean firms, we assume that firms share the same country of origin. Equation (32) implicitly assumes

$\mathbb{E}[\nu \mathbb{1}\{\nu \geq 0\}]$ such that the scale of the parameter vector β is the same across the different estimation methods.

that $\beta_1 R_{ict}$ yields the revenue net of variable costs. In other words, variable costs are assumed to be a constant fraction of revenue across firms, countries and time periods.[36] A firm pays the fixed costs in every country-period pair in which it sells a positive quantity.

Our estimation approach does not require us to specify precisely the content of the information set of firm i at the time it decides whether to export to a country c at period t, $\mathcal{J}_{it}$. The only restriction we impose is that $(R_{ict-1}, D_c, \nu^{\pi}_{ict}) \in \mathcal{J}_{it}$. In words, at the time the firm decides whether to export or not, it knows the distance to the destination market, some determinants captured in the variable ν^{π}_{ict}, and the revenue it would have obtained last period if it had exported, R_{ict-1}.[37]

Firms may have imperfect information about the actual revenue they would obtain if they were to export. Let R^*_{ict} be the expected value of R_{ict} conditional on the information set of the firm at the time it decides whether to export or not, $R^*_{ict} = \mathbb{E}[R_{ict}|\mathcal{J}_{it}]$. Using the notation introduced in Section 2, we define the state vector (X^*_{ict}, ν_{ict}), with $X^*_{ict} = (R^*_{ict}, D_c)$, and write the expected payoffs from exporting as

$$U_{ict} = U(X^*_{ict}, \nu_{ict}) = \mathbb{E}[\pi_{ict}|\mathcal{J}_{it}] = \beta_1 R^*_{ict} - \beta_2 - \beta_3 D_c + \nu^{\pi}_{ict}. \tag{33}$$

We denote the optimal action that each firm takes as $\mathsf{d}_{ict} = \mathsf{d}(X^*_{ict}, \nu_{ict})$. We assume that firms export to some country if and only if the expected value of exporting is higher than the expected value of not exporting. Accordingly, the policy function is defined by the following inequality:

$$(\mathsf{d}_{ict} - (1 - \mathsf{d}_{ict}))U_{ict} \geq 0. \tag{34}$$

Measurement model. The econometrician does not observe the agents' expectations, R^*_{ict}, and observes the realized revenue only for those observations with $\mathsf{d}_{ict} = 1$.[38]

[36] As Morales et al. (2011) shows, one can provide micro foundations for this relationship between revenues and variable costs in a model that assumes firms that are monopolistically competitive and face CES demands in each potential destination market. In this setting, β_1 is equal to the reciprocal of the elasticity of substitution.

[37] We will assume below that we can write R_{ict} as a function of firm and foreign market characteristics. Therefore, assuming that R_{ict-1} is known to firm i at period t is equivalent to assuming that those firm and country characteristics are revealed *ex post*.

[38] This example departs from the measurement model described in Section 2 in that the econometrician only observes the realized value of the variable whose expectation enters the objective function for a selected sample. We chose this modification of the measurement model for our application because applied researchers often face data structures of this form. The key properties of our estimator remain unchanged under this modification.

Therefore, the econometrician needs to predict the value of R_{ict} for those observations such that $\mathsf{d}_{ict} = 0$. The econometrician bases this prediction on a reduced form approximation to the realized revenue. We assume the following expression for the realized revenue from exporting:

$$R_{ict} = r(X^R_{ict}; \theta) + \nu^R_{ict} + \varepsilon^R_{ict}, \tag{35}$$

where X_{ict} is a vector of observed covariates, both ν^R_{ict} and ε^R_{ict} are unobserved to the econometrician, and $\mathbb{E}[\nu^R_{ict}|\mathcal{J}_{it}] = \nu^R_{ict}$, and $\mathbb{E}[\varepsilon^R_{ict}|\mathcal{J}_{it}] = 0$. We impose no assumption on the statistical relationship between the vector X^R_{ict} and the agent's information set, $\mathcal{J}_{it}$. Defining $\hat{\theta}$ as an estimate of θ, we will use

$$\hat{R}_{ict} = r(X^R_{ict}; \hat{\theta}) \tag{36}$$

as a proxy for R^*_{ict} in our moment inequalities.

Using the same notation as in Section 2, $X_{ict} = \hat{R}_{ict}$, and $Z_{1ict} = D_c$. Finally, the econometrician also observes an instrument: a variable that is correlated with the measurement $\hat{R}_{ict}$ and is independent of the structural error and mean independent of the expectational error. Given that X^R_{ict} is observed for every (i, c, t), we will use $\hat{R}_{ict-1}$ as an instrument for $\hat{R}_{ict}$. Therefore, using the notation in Section 2, $Z_{2ict} = \hat{R}_{ict-1}$.

6.2 Estimation

The parameter vector to estimate is (θ, β). The vector θ projects the actual revenue from exporting, R_{ict}, onto the vector of firm and country characteristics X^R_{ict}. The vector β transforms revenue from exporting into variable profits (β_1) and parameterizes the fixed costs from exporting (β_2 and β_3).

We follow a two-step estimation procedure. In the first stage, we apply panel data estimation techniques to obtain point estimates of θ that are independent of the value estimated for β. In the second stage, we use the moment inequality framework described in Section 3 in order to obtain set estimates for β that are conditional on the first stage estimates of θ.[39]

[39]We could have followed an alternative estimation procedure and combine equations (33), (34) and (35) with Assumptions 1 to 3 to derive moment inequalities that set identify the parameter vector (θ, β). However, this estimation procedure has two potential shortcomings: (a) it does not use the available data on revenue (i.e. for the subset of observations with positive exports); (b) it is likely to generate very large identified sets. The reason for expecting large identified sets from this procedure

Using data on observed export revenues for firms, countries and years with positive exports, we obtain an estimate for θ. Section 2 of the Online Appendix describes the estimation of θ in detail. The outcome of this first stage is the variable $\hat{R}_{ict} = r(X^R_{ict}; \hat{\theta})$ and we use it to proxy for R^*_{ict}:

$$U_{ict} = \beta_1 \hat{R}_{ict} - \beta_2 - \beta_3 D_c + \nu_{ict} + \varepsilon_{ict}, \tag{37}$$

where $\nu_{ict} = \nu^{\pi}_{ict} + \beta_1 \mathbb{E}\big[R_{ict} - \hat{R}_{ict} | \mathcal{J}_{ict}\big]$, and $\varepsilon_{ict} = \beta_1 (R^*_{ict} - R_{ict}) + \beta_1 \big[(R_{ict} - \hat{R}_{ict}) - \mathbb{E}\big[R_{ict} - \hat{R}_{ict} | \mathcal{J}_{ict}\big]\big]$. The structural error, ν_{ict}, incorporates both the determinants of the static profits the econometrician does not observe, ν^{π}_{ict}, as well as the part of the approximation error from the first stage, R_{ict} - $\hat{R}_{ict}$, that is in the information set of the exporter. The measurement error, ε_{ict}, incorporates both the expectational error, R^*_{ict} - R_{ict}, and approximation error from the first stage that is orthogonal to the information set of the exporter.

We impose Assumptions 1 and 3 to the distribution of (ν_{ict}, ε_{ict}) conditional on (X^*_{ict}, Z_{ict}) and assume that ν_{ict} is *iid* across firms, countries and time periods and normally distributed with mean 0 and variance equal to 1.[40] Assumption 1 imposes exogeneity restrictions that are common in the international trade literature: unobservable (to the econometrician) determininants of entry in a given country and time period are independently distributed from observable determinants of this decision. Assumption 2 is automatically satisfied once we impose normality on the distribution of the structural error term. Concerning the error term ε_{ict}, Assumption 3 is an immediate consequence of this error term being mean independent of the agent's information set, $\mathcal{J}_{it}$, as long as we assume that lagged determinants of export revenues are contained in firms' information sets (i.e. $\forall$ c, $X^R_{ict-1} \in \mathcal{J}_{it}$).

6.2.1 Moment Inequalities

In this section, we apply the general identification framework in Section 3 to the model described in Section 6.1 in order to derive moment inequalities that contain the true value of the parameter vector β.

is that the variable D_c is likely to be very correlated with X^R_{ict}.

[40] In order to interpret the assumption that the variance of ν_{ict} is equal to 1, it is important to note that the variable $\hat{R}_{ict}$ is expressed in hundreds of thousands of real 1995 dollars. Therefore, assuming that the structural error has a variance of 1 is equivalent to assuming that there are unobservable factors affecting firms' export decisions and that the standard deviation of these unobservable factors is equal to USD 100,000.

Instrument functions. If we apply here the instrument functions defined in equation (15), then $\Psi_q = 0$ for many $q \in Q$. The joint support of the vector Z is such that, for some $q \in Q$, there exists no observation (i, c, t) such that $\Psi_q(Z_{ict}) = 1$. For example, if the variable that multiplies the parameter β_1 (i.e. $\hat{R}_{ict}$) enters with positive sign for some observation, the variables that multiply β_2 (i.e. -1) and β_3 (i.e. $-D_c$) will always enter with negative sign. Therefore, at least one of the instrument functions will be zero for all observations. Since we cannot directly apply the instrument functions in equation (15), we define alternative instrument functions. We describe them in detail in Section 3 of the Online Appendix.

Moment Inequalities. We base our estimation on an application of the moments $\mathcal{M}_s$ and $\mathcal{M}_r$ defined in equations (16) and (20). We present results for the three type of instrument vectors described in Section 3 of the Online Appendix.[41]

6.3 Results

Table 7 presents the results of a probit in which we project the variable d_{ict} on predicted revenues, $\hat{R}_{ict}$, a constant, and distance to country c, D_c. The estimates from this probit model would be consistent estimates of the parameter vector β if: (a) our approximation $\hat{R}_{ict}$ perfectly captures firms' expectations at the time of deciding whether to export or not; and, (b) the variable capturing factors known to the agent and unobserved to the econometrician follow a standard normal distribution. Different assumptions on the variance of the structural error affect the value of the estimates but not their sign or their relative value.

Table 7: ML Estimation: Probit

Parameter	Estimate	Std. Error
β_1	0.045	0.002
β_2	0.883	0.015
β_3	1.166	0.018

The value of β_1 indicates that exporters keep 4.5% of export revenue as variable profits. The value of β_2 implies a constant component of fixed export costs equal to USD $88,300$. The value of β_3 predicts that, if we compare two countries whose distance

[41]In order to interpret the estimates presented in Section 6.3, note that the variable D_c is expressed in terms of thousands of kilometers.

to Chile differs in $10,000$ km, then fixed export costs for the further away country will be USD $116,600$ larger.

Table 8: Identified Set

Set	min(β_1)	max(β_1)	min(β_2)	max(β_2)	min(β_3)	max(β_3)
No Meas. Error	0.045	0.045	0.883	0.883	1.166	1.166
Set 1	0.031	0.176	0.615	1.465	0.749	1.697
Set 2	0.078	0.142	0.626	0.753	1.435	1.685

Table 9: Confidence Set

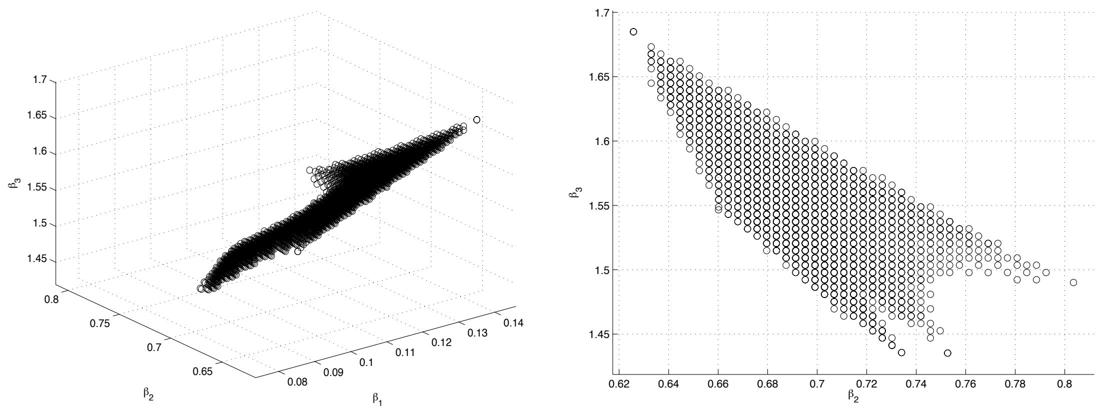
Set	min(β_1)	max(β_1)	min(β_2)	max(β_2)	min(β_3)	max(β_3)
Set 2	0.078	0.142	0.626	0.804	1.435	1.685

Our moment inequality estimator allows us to relax the assumption that our approximation $\hat{R}_{ict}$ is a perfect proxy for firms' expectations. For different sets of instruments, Table 8 shows the maximum and minimum numbers for each parameter contained in the estimated set. We focus here on the results generated by the set of instruments labeled Set 2 (see Section 3 of the Online Appendix for a detailed description of the three types of instruments considered in Table 8). Table 9 shows the confidence interval for the second set of instruments. We show the results with moment selection (see Andrews and Soares (2010)).

Given the scale implied by the assumption that the variance of the structural error term is equal to 1, our estimate of the identified set in Table 8 predicts a profit margin over export revenue between 7.8% and 14.2%, a constant component of fixed costs between USD $62,600$ and USD $75,300$, and an increase in these entry costs between USD $143,500$ and USD $168,500$ per additional $10,000$ km. While Table 8 indicates the limits of the smallest possible cube containing the actually estimated set, Figure 3 characterizes the actual set. From Figure 3a, it is immediate that the estimated set is much smaller than the smallest cube containing it. Therefore, our moment inequality estimates are much more informative than the numbers in Table 8 imply. Figure 3b makes this even more transparent, by projecting the estimated set on the space defined by the coordinates (β_2, β_3).

A comparison of Tables 7 and 8 shows that our estimation of the identified set does not include the Maximum Likelihood estimate. In particular, the MLE underpredicts

Figure 3: Confidence Set



(a) 3 Dimensions (b) 2 Dimensions

the share of export revenue that accrues to the exporter as well as the slope at which entry costs increase with distance. It overpredicts the constant term of entry costs. The downward bias in the estimate of β_1 implies that the MLE will tend to underpredict entry of those firms and countries that are expected to generate large export revenues: e.g. it underpredicts entry by large firms and in large countries. The upward bias in the estimates of the constant term on fixed costs, β_2, and the downward bias in the estimates of their slope with respect to distance, β_3, implies that the MLE will tend to underpredict entry in countries that are close to the country of origin of the firm and overpredict entry in countries that are more distant.

Table 10: Predictions Across Distance Quantiles

	[0,1)	[1,2)	[2,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,9)	[9,10)
Moment Inequality Predictions										
Min. number	1,857	759	704	244	147	140	41	35	15	15
Max. number	2,211	925	790	311	197	181	75	65	30	26
Min. sales	5,317	105	4,782	592	1,721	3,311	5	14	16	749
Max. sales	6,267	138	5,880	641	1,923	3,660	9	37	64	994
Probit Predictions										
Number	1,634	891	731	373	247	200	127	115	60	31
Sales	4,215	86	4,169	491	1,615	2,826	8	17	16	553

Table 10 shows that the biases in the MLE may have important implications for the total number of exporters and volume of exports.[42] As indicated in the previous paragraph, the MLE implies a lower number of entrants than the moment inequalities estimator for the 10% of countries that are closer to Chile, and a larger number of entrants than the moment inequalities estimator for the furthest 70% of countries (with the exception of the decile 0.9 to 1). However, because the MLE tends to underpredict entry by those firms that would export the largest quantities conditional on entry, the MLE significantly underpredicts the total volume of exports for each of the country groups in Table 10.

7 Conclusion

This paper shows how to identify and estimate the parameters of a binary choice model with covariates that are endogenous due to expectational error. We introduce a statistical model in which the error component determining agents' choices is the sum of a structural error and a expectational error term. We motivate this statistical model as a natural extension of the standard binary choice model for cases in which agents with rational expectations face uncertainty about the exact payoffs associated with their actions. Accordingly, the only restriction we impose on the distribution of the expectational error is that it is mean independent of the information set of the agent at the time she makes a decision.

We apply our moment inequality estimation approach to the analysis of a single-agent static entry model. Using firm-level export data, we estimate parameters that determine costs incurred by exporting firms upon entering new foreign markets. We allow firms' decisions to be driven both by expectational errors and components of their information set that are unobserved to the econometrician. Our results suggest that the typical assumption of perfect foresight can severely bias the parameter estimates and provide a distinct economic interpretation of the observed decisions.

[42] In order to compute Table 10, we fit the empirical distribution of D_c and assume a distribution of R^*_{ict} that coincides with the empirical distribution of $\hat{R}_{ict}$. This is not meant as a substantive assumption. It is only done for the sake of illustrating the implications of the bias in the MLE estimates.

References

Amstrong, Tim B., "Asymptotically Exact Inference in Conditional Moment Inequality Models," *mimeo*, December 2012.

An, Mark Yuying, "Log Concave Probability Distributions: Theory and Testing," *mimeo*, November 1995.

Andrews, Donald W.K. and Gustavo Soares, "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 2010, *78*, 119–157.

_ **and Xiaoxia Shi**, "Inference Based on Conditional Moment Inequalities," *mimeo*, July 2011.

_ **and** _ , "Nonparametric Inference Based on Conditional Moment Inequalities," *mimeo*, December 2011.

_ **and** _ , "Inference Based on Conditional Moment Inequalities," *Econometrica*, 2013, *81* (2), 609–666.

Bagnoli, Mark and Theodore C. Bergstrom, "Log-Concave Probability and Its Applications," *Economic Theory*, 2005, *26* (2), 445–469.

Beresteanu, Arie and Francesca Molinari, "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 2008, *76* (4), 763–814.

Berry, Steve, James Levinsohn, and Ariel Pakes, "Estimating Differentiated Product Demand Systems from a Combination of Micro and Macro Data: The Market for New Vehicles," *Journal of Political Economy*, 2004, *112* (1), 68–105.

Blundell, Richard W. and James L. Powell, "Endogeneity in Nonparametric and Semiparametric Regression Models," in Mathias F. Dewatripont, Lars P. Hansen, and Stephen J. Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications, Eight World Congress, Vol. II*, Cambridge: Cambridge University Press, 2003.

_ **and** _ , "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 2004, *71*, 655–679.

Chernozhukov, Victor, Han Hong, and Elie Tamer, "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 2007, *75*, 1243–1284.

Chesher, Andrew, "Identification in Nonseparable Models," *Econometrica*, 2003, *71* (2), 1401–1444.

— , “Nonparametric Identification under Discrete Variation,” *Econometrica*, 2005, *73* (2), 1525–1550.

— , “Instrumental Values,” *Journal of Econometrics*, 2007, *139* (1), 15–34.

— , “Instrumental Variable Models for Discrete Outcomes,” *Econometrica*, 2010, *75* (2), 575–601.

— , “Semiparametric Structural Models of Binary Response: Shape Restrictions and Partial Identification,” *mimeo*, 2011.

— **, Adam Rosen, and Konrad Smolinski**, “An Instrumental Variable Model of Multiple Discrete Choice,” *mimeo*, 2011.

— **and** — , “An Instrumental Variable Random Coefficients Model for Binary Outcomes,” *mimeo*, 2012.

— **and Konrad Smolinski**, “Sharp identied sets for discrete variable IV models,” *mimeo*, 2010.

Ciliberto, Federico and Elie Tamer, “Market Structure and Multiple Equilibria in the Airline Markets,” *Econometrica*, 2009, *77*, 1791–1828.

Florens, Jean-Pierre, James J. Heckman, Costas Meghir, and Edward Vytlacil, “Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects,” *Econometrica*, 2008, *76* (5), 1191–1206.

Heckman, James J. and Bo E. Honoré, “The Empirical Content of the Roy Model,” *Econometrica*, 1990, *58* (5), 1121–1149.

Ho, Katherine, “Insurer-Provider Networks in the Medical Care Market,” *American Economic Review*, 2009, *99* (1), 393–430.

Holmes, Thomas J., “The Diffusion of Wal-Mart and Economies of Density,” *Econometrica*, 2011, *79* (1), 253–302.

Imbens, Guido W. and Whitney K. Newey, “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” *Econometrica*, 2009, *77* (5), 1481–1512.

Ishii, Joy, “Compatibility, Competition, and Investment in Network Industries: ATM Networks in the Banking Industry,” *mimeo*, 2008.

Kaido, Hiroaki and Andrés Santos, “Asymptotically Efficient Estimation of Modes Defined by Convex Moment Inequalities,” *mimeo*, April 2011.

Katz, Michael, "Supermarkets and Zoning Laws," *Unpublished PhD Dissertation*, 2007, *Harvard University.*

Lewbel, Arthur, "Semiparametric Qualitative Response Model Estimation With Unknown Heteroskedasticity or Instrumental Variables," *Journal of Econometrics*, 2000, *97*, 145–177.

Magnac, Thierry and Eric Maurin, "Identification and Information in Monotone Binary Models," *Journal of Econometrics*, 2007, *139*, 76–104.

_ and _ , "Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data," *The Review of Economic Studies*, 2008, *75* (3), 835–864.

Morales, Eduardo, Gloria Sheu, and Andrés Zahler, "Gravity and Extended Gravity: Estimating a Structural Model of Export Entry," *mimeo*, 2011.

Pakes, Ariel, "Alternative Models for Moment Inequalities," *Econometrica*, 2010, *78* (6), 1783–1822.

_ , Jack R. Porter, Katherine Ho, and Joy Ishii, "Moment Inequalities and their Application," *mimeo*, December 2011.

Pratt, John W., "Concavity of the the Log Likelihood," *Journal of the American Statistical Association*, 1981, *76*, 103–106.

Shaikh, Azeem M. and Edward J. Vytlacil, "Partial Identification in Triangular Systems of Equations With Binary Dependent Variables," *Econometrica*, 2011, *79* (3), 949–955.

Vytlacil, Edward J. and Nese Yildiz, "Dummy Endogenous Variables in Weakly Separable Models," *Econometrica*, 2007, *75* (3), 757–779.

Appendix

A.1 Proof of Remark 1

Let f_ν be the density function of ν and f'_ν its derivative. We can write

$$\frac{\partial^2 \mathbb{E}[\nu|\nu \geq y]}{\partial y^2} = -\left(\frac{f'_\nu(y)}{f_\nu(y)} - 2\frac{f_\nu(y)}{1-F_\nu(y)}\right) \cdot \left(y - \mathbb{E}[\nu|\nu \geq y]\right) - 1.$$

For the case in which $\nu \sim \mathbb{N}(0, \sigma^2)$, this second derivative becomes

$$\left(\frac{\Phi(y/\sigma)}{1-\Phi(y/\sigma)} - \frac{y}{\sigma}\right)^2 + \frac{\phi(y/\sigma)}{1-\Phi(y/\sigma)}\left(\frac{\phi(y/\sigma)}{1-\Phi(y/\sigma)} - \frac{y}{\sigma}\right) - 1. \tag{38}$$

For the case in which $\nu \sim$ *Logistic*, the same second derivative becomes:

$$(1+\exp(y)) \cdot (y + \ln(1+\exp(y))) - 1. \tag{39}$$

By simulating equations (38) and (39) one can show that both expressions are positive for every $y \in (-\infty, \infty)$. ■

A.2 Partial Identification: Example

Here, we show that β is partially identified in a model that imposes restrictions that are stronger than those in Assumptions 1 to 3. Even under the assumptions of this stricter model, we can still find at least two structures

$$S^{a_1} \equiv \{\beta^{a_1}, f^{a_1}(X|X^*, Z_2) f^{a_1}(Z_2|X^*) f^{a_1}(X^*)\},$$
$$S^{a_2} \equiv \{\beta^{a_2}, f^{a_2}(X|X^*, Z_2) f^{a_2}(Z_2|X^*) f^{a_2}(X^*)\},$$

such that

$$\mathcal{P}(d, Z, X) = \int f(d|X^*, X; \beta^{a_i}) f^{a_i}(X|X^*, Z_2) f^{a_i}(Z_2|X^*) f^{a_i}(X^*) dX^*, \tag{40}$$

and $\beta^{a_1} \neq \beta^{a_2}$. If β is partially identified in this stricter model, it will also be partially identified in the more general model described in Section 2.2.

The distributional assumptions that characterize this stricter model are:

$$f(d|X^*, X; \beta) = 1 - \Phi(-\beta X^*), \tag{41a}$$
$$f(X_2|X_2^*, Z_2) = \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{X_2 - X_2^*}{\sigma_x}\right)^2\right], \tag{41b}$$
$$f(Z_2|X_2^*) = \frac{1}{\sigma_z\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{Z_2 - X_2^*}{(1-\rho_z)\sigma_z}\right)^2\right], \tag{41c}$$
$$f(X^*) = \frac{1}{\sigma_{x^*}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{X^*}{\sigma_{x^*}}\right)^2\right]. \tag{41d}$$

In words, these distributional assumptions are consistent with a discrete choice model in which:

1. The distribution of ν conditional on (X, X^*) is normal with mean 0 and variance 1.
2. We can write $X_2 = X_2^* + \epsilon_2$ and $Z_2 = X_2^* + \epsilon_2^z$ and
   (a) The distribution of ϵ_2 conditional on (X_2^*, Z_2) is normal with mean 0 and variance σ_x^2.
   (b) The distribution of ϵ_2^z conditional on X_2^* is normal with mean 0, variance σ_z^2, and the correlation coefficient between ϵ_2^z and X^* is equal to ρ_z.
3. The marginal distribution of X^* is normal with mean 0 and variance $\sigma_{x^*}^2$.

The distributional assumptions in equation (41) are stronger than Assumptions 1 to 3 in Section 2.2. In particular, they impose parametric functional forms on the marginal distribution of the true unobserved covariates, X_2^*, and on the distribution of both the mismeasured covariates, X_2, and instrument, Z_2, conditional on the true unobserved covariates. Equation (41) also imposes full independence between the the measurement error, ϵ_2, and the vector (X_2^*, ν).

Importantly, both Assumptions 1 to 3 and the assumptions implicitly contained in equation (41) are similar in that they do not specify the joint distribution of the instrument vector, Z_2, and the true unobserved covariates, X_2^*. In equation (41c), this joint distribution depends on the value of the parameter ρ_z, which captures the correlation between the true covariate and the measurement error affecting the instrument.

As the following result indicates, even in the statistical model described in equation (41), it is true that the parameter vector β is not point-identified.

Result A.2.1 *There exists empirical distributions of the vector of observable variables (d, Z, X), $\mathcal{P}(d, Z, X)$, such that there are at least two structures S^{a_1} and S^{a_2} for which*

1. *both S^{a_1} and S^{a_2} verify equation (40).*
2. *both S^{a_1} and S^{a_2} verify the distributional assumptions in equation (41).*
3. *$\beta^{a_1} \neq \beta^{a_2}$.*

This result can be proved by combining the following two lemmas.

Lemma A.2.1 *The parameter vector β is point-identified only if the parameter σ_{x^*} is point-identified.*

Proof: Define $X^* = \sigma_{x^*}\tilde{X}^*$, such that $var(\tilde{X}^*) = 1$. We can then rewrite equation (41a) as

$$f(d|X^*, X; \beta) = 1 - \Phi(-(\beta\sigma_{x^*})\tilde{X}^*).$$

The vector β only enters the likelihood function in equation (40) multiplied by σ_{x^*}. Therefore, we can only identify the value of β if we know the value of σ_{x^*}. ■

Lemma A.2.2 *The parameter vector σ_x is point-identified if and only if the parameter ρ_z is assumed to be equal to zero.*

Proof: Restrictions (41b), (41c), and (41d) imply that the variables (X_2, Z_2) are jointly normal and, therefore, their joint distribution is fully characterized by their variances and covariances. Consequently, the only restrictions that can identify the parameter vector (σ_x, σ_z, ρ_z, σ_x^*) are

$$\begin{aligned} var(X_2) &= \sigma_{x^*}^2 + \sigma_x^2, \\ var(Z_2) &= \sigma_{x^*}^2 + \sigma_z^2 + \rho_z\sigma_x\sigma_z, \\ cov(X_2, Z_2) &= \sigma_{x^*}^2 + \rho_z\sigma_x\sigma_z. \end{aligned}$$

From these restrictions, we obtain that

$$\sigma_z^2 = var(Z_2) - cov(X_2, Z_2).$$

Given that $\sigma_x^2 \geq 0$ and $-1 \leq \rho_z \leq 1$, we obtain that

$$max\{cov(X_2, Z_2) - \sigma_x\sigma_z, 0\} \leq \sigma_{x^*}^2 \leq min\{var(X_2), cov(X_2, Z_2) + \sigma_x\sigma_z\}.$$

Only in the particular case in which ρ_z is assumed to be equal to 0, we obtain that

$$\sigma_{x^*}^2 = cov(X_2, Z_2). \quad \blacksquare$$

In summary, we can conclude that: (a) the model in equation (41) is a special case of the model in Section 2.2; (b) the parameter vector β is generally set-identified in the model in equation (41), unless we assume that $\rho_z = 0$; (c) the model in Section 2.2 does not specify the distribution of the instrument conditional on the true covariate (i.e. does not specify the value of ρ_z); (d) therefore, the parameter vector β is set-identified in the model in Section 2.2.

A.3 Proof of Theorem 1.1

Lemma A.3.1 *For any distribution function F_ν, and any true value of the parameter vector, $\beta^* \in \Gamma^\beta$, the score function of the model defined in Section 2.2, conditional on the vector (X^*, Z_2), implies*

$$\mathbb{E}\left[d\frac{F_\nu\big(-\beta^*X^*\big)}{1-F_\nu\big(-\beta^*X^*\big)} - (1-d)\Big|X^*, Z_2\right] = 0, \tag{42a}$$

$$\mathbb{E}\left[(1-d)\frac{1-F_\nu\big(-\beta^*X^*\big)}{F_\nu\big(-\beta^*X^*\big)} - d\Big|X^*, Z_2\right] = 0. \tag{42b}$$

Proof: Let $L(d|X^*, Z_2)$ denote the log-likelihood conditional on (X^*, Z_2). Given Assumption 1:

$$L(d|X^*, Z_2) = \mathbb{E}\Big[d\log\big(1 - F_\nu(-\beta X^*)\big) + (1-d)\log\big(F_\nu(-\beta X^*)\big)|X^*, Z_2\Big].$$

The corresponding score function is:

$$\frac{\partial L(d|X^*, Z_2)}{\partial \beta} =$$
$$\mathbb{E}\left[d\frac{1}{1-F_\nu(-\beta X^*)}\frac{\partial\big(1-F_\nu(-\beta X^*)\big)}{\partial \beta} + (1-d)\frac{1}{F_\nu(-\beta X^*)}\frac{\partial F_\nu(-\beta X^*)}{\partial \beta}\middle| X^*, Z_2\right]. \tag{43}$$

Reordering terms:

$$\frac{\partial L(d|X^*, Z_2)}{\partial \beta} = \mathbb{E}\left[d\frac{F_\nu(-\beta X^*)}{1-F_\nu(-\beta X^*)}\frac{\frac{\partial(1-F_\nu(-\beta X^*))}{\partial \beta}}{\frac{\partial F_\nu(-\beta X^*)}{\partial \beta}} + (1-d)\middle| X^*, Z_2\right]$$

For any F_ν and any value of β,

$$\frac{\frac{\partial F_\nu(-\beta X^*)}{\partial \beta}}{\frac{\partial\big(1-F_\nu(-\beta X^*)\big)}{\partial \beta}} = -1. \tag{44}$$

Therefore, we can rewrite the score function as:

$$\mathbb{E}\left[d\frac{F_\nu\big(-\beta X^*\big)}{1-F_\nu\big(-\beta X^*\big)} - (1-d)\middle| X^*, Z_2\right]$$

and, for the true value of the parameter vector, β^*, it holds that

$$\mathbb{E}\left[d\frac{F_\nu\big(-\beta^* X^*\big)}{1-F_\nu\big(-\beta^* X^*\big)} - (1-d)\middle| X^*, Z_2\right] = 0,$$

which is identical to equation (42a). In order to derive equation (42b), we reorder the terms in equation (43) into

$$\frac{\partial L(d|X^*, Z_2)}{\partial \beta} = \mathbb{E}\left[d + (1-d)\frac{1-F_\nu(-\beta X^*)}{F_\nu(-\beta X^*)}\frac{\frac{\partial F_\nu(-\beta X^*)}{\partial \beta}}{\frac{\partial(1-F_\nu(-\beta X^*))}{\partial \beta}}\middle| X^*, Z_2\right].$$

Using again equation (44), we obtain

$$\mathbb{E}\left[-(1-d)\frac{1-F_\nu\big(-\beta X^*\big)}{F_\nu\big(-\beta X^*\big)} + d\middle| X^*, Z_2\right],$$

and, for the particular case in which $\beta = \beta^*$, it holds

$$\mathbb{E}\left[-(1-d)\frac{1-F_\nu\big(-\beta^* X^*\big)}{F_\nu\big(-\beta^* X^*\big)} + d\middle| X^*, Z_2\right] = 0.$$

Multiplying by -1 on both sides, we obtain equation (42b). ■

Lemma A.3.2 *For any distribution function F_ν, and any true value of the parameter vector, $\beta^* \in \Gamma^\beta$, the score function of the model defined in Section 2.2, conditional on the vector Z, implies*

$$\mathbb{E}\left[d\frac{F_\nu\big(-\beta^* X^*\big)}{1-F_\nu\big(-\beta^* X^*\big)} - (1-d)\Big|Z\right] = 0, \tag{45a}$$

$$\mathbb{E}\left[(1-d)\frac{1-F_\nu\big(-\beta^* X^*\big)}{F_\nu\big(-\beta^* X^*\big)} - d\Big|Z\right] = 0. \tag{45b}$$

Proof: From equation (9) and $Z = (Z_1, Z_2)$ note that Z is a subset of the vector (X^*, Z_2). Therefore, Lemma A.3.2 follows from equation (42) and the Law of Iterated Expectations (LIE). ■

Lemma A.3.3 *For any log concave distribution function F_ν,*

$$\frac{\partial^2(F_\nu(y)/(1-F_\nu(y)))}{\partial y^2} \geq 0, \tag{46a}$$

$$\frac{\partial^2(1-F_\nu(y))/F_\nu(y))}{\partial y^2} \geq 0, \tag{46b}$$

for any y in the support of F_ν.

Proof: The first derivative of $F_\nu(y)/(1-F_\nu(y))$ is:

$$\frac{\partial(F_\nu(y)/(1-F_\nu(y)))}{\partial y} = \frac{1}{1-F_\nu(y)}\frac{F'_\nu(y)}{1-F_\nu(y)},$$

and the second derivative of $F_\nu(y)/(1-F_\nu(y))$ is:

$$\frac{\partial^2(F_\nu(y)/(1-F_\nu(y)))}{\partial y^2} = \frac{\partial(1/(1-F_\nu(y)))}{\partial y}\frac{F'_\nu(y)}{1-F_\nu(y)} + \frac{1}{1-F_\nu(y)}\frac{\partial(F'_\nu(y)/(1-F_\nu(y)))}{\partial y}.$$

For any distribution function F_ν, it holds

$$\frac{1}{1-F_\nu(y)} \geq 0, \quad \frac{F'_\nu(y)}{1-F_\nu(y)} \geq 0, \quad \frac{\partial(1/(1-F_\nu(y)))}{\partial y} \geq 0.$$

Moreover, as Heckman and Honoré (1990) and Bagnoli and Bergstrom (2005) show, for any log concave distribution function F_ν, it holds

$$\frac{\partial(F'_\nu(y)/(1-F_\nu(y)))}{\partial y} \geq 0.$$

Therefore, we can conclude that

$$\frac{\partial^2(F_\nu(y)/(1-F_\nu(y)))}{\partial y^2} \geq 0.$$

Analogously, the first derivative of $(1-F_\nu(y))/F_\nu(y)$ is:

$$\frac{\partial((1-F_\nu(y))/F_\nu(y))}{\partial y}=\frac{1}{F_\nu(y)}\frac{-F'_\nu(y)}{F_\nu(y)},$$

and the second derivative of $(1-F_\nu(y))/F_\nu(y)$ is:

$$\frac{\partial^2((1-F_\nu(y))/F_\nu(y))}{\partial y^2}=\frac{\partial(1/F_\nu(y))}{\partial y}\frac{-F'_\nu(y)}{F_\nu(y)}+\frac{1}{F_\nu(y)}\frac{\partial(-F'_\nu(y)/F_\nu(y))}{\partial y}.$$

For any distribution function F_ν, it holds

$$\frac{1}{F_\nu(y)}\geq 0,\quad \frac{-F'_\nu(y)}{1-F_\nu(y)}\leq 0,\quad \frac{\partial(1/F_\nu(y))}{\partial y}\leq 0.$$

Moreover, as Yuying An (1995) shows, for any log concave distribution function F_ν, it holds

$$\frac{\partial(-F'_\nu(y)/F_\nu(y))}{\partial y}\geq 0.$$

Therefore, we can conclude that

$$\frac{\partial^2((1-F_\nu(y))/F_\nu(y))}{\partial y^2}\geq 0. \quad \blacksquare$$

Lemma A.3.4 *For any log concave distribution function F_ν and any random variable η such that*

$$\mathbb{E}[\eta|X^*,Z_2]=0,$$

it holds that

$$\begin{aligned}\mathbb{E}\Big[\frac{F_\nu(y+\eta)}{1-F_\nu(y+\eta)}\Big|X^*,Z_2\Big]&\geq\frac{F_\nu(y)}{1-F_\nu(y)},\\ \mathbb{E}\Big[\frac{1-F_\nu(y+\eta)}{F_\nu(y+\eta)}\Big|X^*,Z_2\Big]&\geq\frac{1-F_\nu(y)}{F_\nu(y)}.\end{aligned}$$

Proof: From Lemma A.3.3 and Jensen's Inequality. $\blacksquare$

Lemma A.3.5 *If Assumptions 2 and 3 hold, then, for every value of (d,X^*,Z_2) and for every value of $\beta\in\Gamma^\beta$:*

$$\begin{aligned}\mathbb{E}\Bigg[\frac{F_\nu\big(-(\beta_1Z_1+\beta_2X_2^*+\varepsilon)\big)}{1-F_\nu\big(-(\beta_1Z_1+\beta_2X_2^*+\varepsilon)\big)}-\frac{F_\nu\big(-\beta X^*\big)}{1-F_\nu\big(-\beta X^*\big)}\Bigg|d,X^*,Z_2\Bigg]&\geq 0,\\ \mathbb{E}\Bigg[\frac{1-F_\nu\big(-(\beta_1Z_1+\beta_2X_2^*+\varepsilon)\big)}{F_\nu\big(-(\beta_1Z_1+\beta_2X_2^*+\varepsilon)\big)}-\frac{1-F_\nu\big(-\beta X^*\big)}{F_\nu\big(-\beta X^*\big)}\Bigg|d,X^*,Z_2\Bigg]&\geq 0.\end{aligned}$$

Proof: From Lemmas A.3.3 and A.3.4. $\blacksquare$

Proof of Theorem 1.1 By the LIE, we can rewrite equation (42a) as

$$\mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[d\frac{F_\nu\big(-\beta^*X^*\big)}{1-F_\nu\big(-\beta^*X^*\big)}-(1-d)\middle|d,X^*,Z_2\right]\middle|X^*,Z_2\right]\middle|Z\right]=0,$$

or, equivalently,

$$\mathbb{E}\left[\mathbb{E}\left[d\mathbb{E}\left[\frac{F_\nu\big(-\beta^*X^*\big)}{1-F_\nu\big(-\beta^*X^*\big)}\middle|d,X^*,Z_2\right]-(1-d)\middle|X^*,Z_2\right]\middle|Z\right]=0.$$

From this expression and Lemma A.3.5,

$$\mathbb{E}\left[\mathbb{E}\left[d\mathbb{E}\left[\frac{F_\nu\big(-(\beta_1^*Z_1+\beta_2^*X_2^*+\varepsilon)\big)}{1-F_\nu\big(-(\beta_1^*Z_1+\beta_2^*X_2^*+\varepsilon)\big)}\middle|d,X^*,Z_2\right]-(1-d)\middle|X^*,Z_2\right]\middle|Z\right]\geq 0.$$

Simplifying via LIE, we obtain

$$\mathbb{E}\left[d\frac{F_\nu\big(-(\beta_1^*Z_1+\beta_2^*X_2^*+\varepsilon)\big)}{1-F_\nu\big(-(\beta_1^*Z_1+\beta_2^*X_2^*+\varepsilon)\big)}-(1-d)\middle|Z\right]\geq 0,$$

or, equivalentely,

$$\mathbb{E}\left[d\frac{F_\nu\big(-(\beta_1^*Z_1+\beta_2^*X)\big)}{1-F_\nu\big(-(\beta_1^*Z_1+\beta_2^*X)\big)}-(1-d)\middle|Z\right]\geq 0,$$

for every $Z\in\mathcal{Z}$. In conclusion, the true value of the parameter vector, β^*, satisfies the score function moment inequality in equation (13b). Following the exact same steps, we can immediately prove that β^* also satisfies the score function moment inequality in equation (13a). Therefore, the true value of the parameter vector is included in the set $\Omega(\mathcal{M}_s,\mathcal{Z}^M)$. ■

A.4 Proof of Theorem 1.2

Lemma A.4.1 *If Assumptions 1 and 2 hold and $X_2^*=X=Z_2$, for any given $z\in\mathcal{Z}^M$ and any $\beta^*\in\Gamma_\beta$, it holds*

$$\Omega(\mathcal{M}_s,\mathcal{Z}^M)=\{\beta\in\Gamma_\beta:\beta z=\beta^*z,\ \textit{for all } z\in\mathcal{Z}^M\}.$$

Proof: Given that $X_2^*=X=Z_2$, we rewrite the index function as:

$$\beta X^*=\beta_1Z_1+\beta_2Z_2=\beta Z.$$

In this case, the vectorial conditional moment $\mathcal{M}_s(z;\beta)$ becomes

$$\mathbb{E}\left[d\frac{F_\nu(-\beta Z)}{1-F_\nu(-\beta Z)} - (1-d)\Big|Z=z\right] \geq 0,$$
$$\mathbb{E}\left[(1-d)\frac{1-F_\nu(-\beta Z)}{F_\nu(-\beta Z)} - d\Big|Z=z\right] \geq 0.$$

Doing simple algebra, we can rewrite these expressions as

$$\mathbb{E}\big[d|Z=z\big]\frac{F_\nu(-\beta z)}{1-F_\nu(-\beta z)} - \mathbb{E}\big[(1-d)|Z=z\big] \geq 0,$$
$$\mathbb{E}\big[1-d|Z=z\big]\frac{1-F_\nu(-\beta z)}{F_\nu(-\beta z)} - \mathbb{E}\big[d|Z=z\big] \geq 0,$$

and, reorganizing terms,

$$\frac{F_\nu(-\beta z)}{1-F_\nu(-\beta z)} \geq \frac{\mathbb{E}\big[(1-d)|Z=z\big]}{\mathbb{E}\big[d|Z=z\big]} = \frac{F_\nu(-\beta^* z)}{1-F_\nu(-\beta^* z)}, \tag{49a}$$
$$\frac{F_\nu(-\beta z)}{1-F_\nu(-\beta z)} \leq \frac{\mathbb{E}\big[(1-d)|Z=z\big]}{\mathbb{E}\big[d|Z=z\big]} = \frac{F_\nu(-\beta^* z)}{1-F_\nu(-\beta^* z)}. \tag{49b}$$

For any value of the vector z, equations (49a) and (49b) are satisfied only if $\beta z = \beta^* z$. We prove this statement by contradiction: (a) if β is such that $\beta z > \beta^* z$, then equation (49a) is not satisfied; (b) if β is such that $\beta z < \beta^* z$, then equation (49b) is not satisfied. Therefore, from equations (49a) and (49b) we obtain the equality:

$$\beta z = \beta^* z. \quad \blacksquare$$

Proof of Theorem 1.2. If we derive equations (49a) and (49b) for a set of vectors $\{z^1, z^2, \dots,$ $z^M\}$, we can analogously derive the following linear system of equations:

$$\beta z^1 = \beta^* z^1,$$
$$\vdots$$
$$\beta z^M = \beta^* z^M.$$

As long as the matrix $(z^1, \dots, z^M)$ has rank larger than K (i.e. the number of elements in β), there is a single value of β that solves this linear system of equations and it is β^*. $\blacksquare$

A.5 Proof of Theorem 1.3

Lemma A.5.1 *Let $g(Y)$ be a convex function of Y and let f_Y denote the density function of a random variable Y with support$(Y) = (-\infty, \infty)$, $\mathbb{E}(Y) = 0$ and $Var(Y) = \sigma_Y^2$, then*

$$\frac{\partial \mathbb{E}[g(Y)]}{\partial \sigma_Y} \geq 0. \tag{50}$$

Proof: Let $f_{\tilde{Y}}$ be the density function of the random variable $\tilde{Y} = (1/\sigma_Y)Y$. Then

$$\mathbb{E}[g(Y)] = \int_{-\infty}^{\infty} g(y) f_Y(y) dy = \int_{-\infty}^{\infty} g(\sigma_Y \tilde{y}) f_{\tilde{Y}}(\tilde{y}) d\tilde{y}$$

and

$$\frac{\partial \mathbb{E}[g(Y)]}{\partial \sigma_Y} = \int_{-\infty}^{\infty} \tilde{y} g'(\sigma_Y \tilde{y}) f_{\tilde{Y}}(\tilde{y}) d\tilde{y} = \int_{-\infty}^{0} \tilde{y} g'(\sigma_Y \tilde{y}) f_{\tilde{Y}}(\tilde{y}) d\tilde{y} + \int_{0}^{\infty} \tilde{y} g'(\sigma_Y \tilde{y}) f_{\tilde{Y}}(\tilde{y}) d\tilde{y}.$$

Given that $\mathbb{E}(Y) = 0$,

$$\int_{-\infty}^{0} \tilde{y} f_{\tilde{Y}}(\tilde{y}) d\tilde{y} + \int_{0}^{\infty} \tilde{y} f_{\tilde{Y}}(\tilde{y}) d\tilde{y} = 0,$$

and, given that $g(Y)$ is convex,

$$g'(\bar{y}) \leq g'(\bar{\bar{y}}), \text{ if and only if } \bar{\bar{y}} \geq \bar{y}.$$

Consequently,

$$\frac{\partial \mathbb{E}[g(Y)]}{\partial \sigma_Y} \geq 0. \quad \blacksquare$$

Proof of Theorem 1.3 From Appendix A.3, we write equation (13a) as:

$$\mathbb{E}\left[\mathbb{E}\left[d\mathbb{E}\left[\frac{F_\nu\big(-(\beta_1^* Z_1 + \beta_2^* X_2^* + \varepsilon)\big)}{1 - F_\nu\big(-(\beta_1^* Z_1 + \beta_2^* X_2^* + \varepsilon)\big)}\middle| d, X^*, Z_2\right] - (1-d)\middle| X^*, Z_2\right]\middle| Z\right] \geq 0.$$

Given Assumptions 2 and 3 and Lemmas A.3.3 and A.5.1, for any (d, X^*, Z_2) and $\beta \in \Gamma_\beta$

$$\frac{\partial\left\{\mathbb{E}\left[\frac{F_\nu\left(-(\beta_1 Z_1 + \beta_2 X_2^* + \varepsilon)\right)}{1 - F_\nu\left(-(\beta_1 Z_1 + \beta_2 X_2^* + \varepsilon)\right)}\middle| d, X^*, Z_2\right]\right\}}{\partial \sigma_\varepsilon} \geq 0,$$

and, therefore, for any $z \in \mathcal{Z}$ and $\beta \in \Gamma_\beta$

$$\frac{\partial\left\{\mathbb{E}\left[\mathbb{E}\left[d\mathbb{E}\left[\frac{F_\nu\left(-(\beta_1 Z_1 + \beta_2 X_2^* + \varepsilon)\right)}{1 - F_\nu\left(-(\beta_1 Z_1 + \beta_2 X_2^* + \varepsilon)\right)}\middle| d, X^*, Z_2\right] - (1-d)\middle| X^*, Z_2\right]\middle| Z = z\right]\right\}}{\partial \sigma_\varepsilon} \geq 0.$$

Following the same steps for equation (13b), we can conclude that:

$$\frac{\partial\left\{\mathbb{E}\left[\mathbb{E}\left[(1-d)\mathbb{E}\left[\frac{1 - F_\nu\left(-(\beta_1 Z_1 + \beta_2 X_2^* + \varepsilon)\right)}{F_\nu\left(-(\beta_1 Z_1 + \beta_2 X_2^* + \varepsilon)\right)}\middle| d, X^*, Z_2\right] - d\middle| X^*, Z_2\right]\middle| Z = z\right]\right\}}{\partial \sigma_\varepsilon} \geq 0.$$

Therefore, if $\bar{\bar{\sigma}}_\varepsilon \geq \bar{\sigma}_\varepsilon$, then, for any given value of $\beta \in \Gamma_\beta$ such that $\mathcal{M}_s(z;\beta|\bar{\sigma}_\varepsilon) \geq 0$, it will be true that $\mathcal{M}_s(z;\beta|\bar{\bar{\sigma}}_\varepsilon) \geq 0$. This implies that, for any $z \in \mathcal{Z}$, $\Omega(\mathcal{M}_s, z|\bar{\sigma}^2_\varepsilon) \subseteq \Omega(\mathcal{M}_s, z|\bar{\bar{\sigma}}^2_\varepsilon)$. Therefore, for any $\mathcal{Z}^M$, $\Omega(\mathcal{M}_s, \mathcal{Z}^M|\bar{\sigma}^2_\varepsilon) \subseteq \Omega(\mathcal{M}_s, \mathcal{Z}^M|\bar{\bar{\sigma}}^2_\varepsilon)$. ■

A.6 Proof of Theorem 2.1

Lemma A.6.1 *For any distribution function F_ν such that $\mathbb{E}[\nu|, X^*, Z_2] = 0$, it holds*

$$\mathbb{E}\big[d\mathbb{E}[\nu|d=1, X^*, Z_2]\big|X^*, Z_2\big] = \mathbb{E}\big[(1-d)\mathbb{E}[-\nu|d=0, X^*, Z_2]\big|X^*, Z_2\big].$$

Proof: Using the LIE,

$$\begin{aligned}
\mathbb{E}\big[\nu\big|X^*, Z_2\big] &= 0,\\
\mathbb{E}\big[\mathbb{E}[\nu|d, X^*, Z_2]\big|X^*, Z_2\big] &= 0,\\
\mathbb{E}[d|X^*, Z_2]\mathbb{E}[\nu|d=1, X^*, Z_2] &= -\mathbb{E}[1-d|X^*, Z_2]\mathbb{E}[\nu|d=0, X^*, Z_2],\\
\mathbb{E}[d|X^*, Z_2]\mathbb{E}[\nu|d=1, X^*, Z_2] &= \mathbb{E}[1-d|X^*, Z_2]\mathbb{E}[-\nu|d=0, X^*, Z_2],\\
\mathbb{E}\big[d\mathbb{E}[\nu|d=1, X^*, Z_2]\big|X^*, Z_2\big] &= \mathbb{E}\big[(1-d)\mathbb{E}[-\nu|d=0, X^*, Z_2]\big|X^*, Z_2\big]. \quad ■
\end{aligned}$$

Lemma A.6.2 *For any distribution function F_ν such that $\mathbb{E}[\nu|, X^*, Z_2] = 0$ and any $\beta^* \in \Gamma^\beta$, it holds:*

$$\mathbb{E}\big[(1-d)(-(\beta^* X^*)) + d\mathbb{E}[\nu|\nu \geq -\beta^* X^*, X^*]\big|X^*, Z_2\big] \geq 0, \tag{51a}$$

$$\mathbb{E}\big[d\beta^* X^* + (1-d)\mathbb{E}[-\nu| - \nu \geq \beta^* X^*, X^*]\big|X^*, Z_2\big] \geq 0. \tag{51b}$$

Proof: For simplicity of notation, we show the proof for equation (51b). The proof for equation (51a) is completely equivalent. Applying the conditional expectation operator $\mathbb{E}[\cdot|X^*, Z_2]$ to equation (8) for the case in which $d = 1$,

$$\mathbb{E}[d\beta^* X^* + d\nu|X^*, Z_2] \geq 0,$$

and, by the LIE,

$$\begin{aligned}
\mathbb{E}\Big[\mathbb{E}\big[d\beta^* X^* + d\nu\big|d, X^*, Z_2\big]\Big|X^*, Z_2\Big] &\geq 0,\\
\mathbb{E}\Big[d\beta^* X^* + d\mathbb{E}\big[\nu\big|d=1, X^*, Z_2\big]\Big|X^*, Z_2\Big] &\geq 0.
\end{aligned}$$

From Lemma A.6.1,

$$\mathbb{E}\Big[d\beta^* X^* + (1-d)\mathbb{E}\big[-\nu\big|d=0, X^*, Z_2\big]\Big|X^*, Z_2\Big] \geq 0,$$

and, from Assumption 1,

$$\mathbb{E}\Big[d\beta^* X^* + (1-d)\mathbb{E}\big[-\nu\big|d=0, X^*\big]\Big|X^*, Z_2\Big] \geq 0.$$

Using again equation (8), we obtain equation (51b),

$$\mathbb{E}\Big[d\beta^*X^* + (1-d)\mathbb{E}\big[-\nu\big|-\nu \geq \beta^*X^*, X^*\big]\Big|X^*, Z_2\Big] \geq 0.$$

If we follow the same steps starting from equation (8) for the case in which $d = 0$, then we obtain equation (51a). ■

Lemma A.6.3 *If Assumptions 2 and 3 hold, then, for every value of* (d, X^*, Z_2) *and for every value of* $\beta \in \Gamma^\beta$*:*

$$\mathbb{E}\Bigg[\mathbb{E}\big[\nu|\nu \geq -(\beta_1 Z_1 + \beta_2 X_2^* - \varepsilon), Z_1, X_2^* - (1/\beta_2)\varepsilon\big] - \mathbb{E}\big[\nu|\nu \geq -\beta X^*, X^*\big]\Bigg|d, X^*, Z_2\Bigg] \geq 0,$$

$$\mathbb{E}\Bigg[\mathbb{E}\big[-\nu|-\nu \geq \beta_1 Z_1 + \beta_2 X_2^* - \varepsilon, Z_1, X_2^* - (1/\beta_2)\varepsilon\big] - \mathbb{E}\big[-\nu|-\nu \geq \beta X^*, X^*\big]\Bigg|d, X^*, Z_2\Bigg] \geq 0.$$

Proof: From Jensen's Inequality. ■

Proof of Theorem 2.1 For simplicity of notation, we show the proof for equation (14b). The proof for equation (14a) is completely equivalent.

By the LIE, we can rewrite equation (51b) as

$$\mathbb{E}\Bigg[\mathbb{E}\Bigg[\mathbb{E}\Big[d\beta^*X^* + (1-d)\mathbb{E}\big[-\nu|-\nu \geq \beta^*X^*, X^*\big]\Big|d, X^*, Z_2\Big]\Bigg|X^*, Z_2\Bigg]\Bigg|Z\Bigg] \geq 0.$$

From this expression and Lemma A.6.3, we obtain a weaker inequality

$$\mathbb{E}\Bigg[\mathbb{E}\Big[d(\beta_1^* Z_1 + \beta_2^* X) + (1-d)\mathbb{E}\big[-\nu|-\nu \geq \beta_1^* Z_1 + \beta_2^* X_2, Z_1, X_2\big]\Big|X^*, Z_2\Big]\Bigg|Z\Bigg] \geq 0,$$

and, simplifying this expression, we conclude that

$$\mathbb{E}\Bigg[d(\beta_1^* Z_1 + \beta_2^* X) + (1-d)\mathbb{E}\big[-\nu|-\nu \geq \beta_1^* Z_1 + \beta_2^* X, Z_1, X\big]\Bigg|Z\Bigg] \geq 0. \ ■$$

A.7 Proof of Theorem 2.2

Lemma A.7.1 *If Assumptions 1 and 2 hold and* $X_2^* = X = Z_2$*, for any given* $z \in \mathcal{Z}^M$ *and any* $\beta^* \in \Gamma_\beta$*,* $\exists\ \beta \in \Gamma_\beta$ *such that* $\beta \neq \beta^*$ *and* $\beta \in \Omega(\mathcal{M}_r, \mathcal{Z}^M)$*.*

Proof: Given that $X_2^* = X = Z_2$, we rewrite the index function as $\beta X^* = \beta_1 Z_1 + \beta_2 X = \beta Z$. The two revealed preference inequalities in equations (14a) and (14b) conditional on $Z = z$ become

$$\mathbb{E}\Big[(1-d)(-\beta Z) + d\mathbb{E}\big[\nu|\nu \geq -\beta Z, Z\big]\Big|Z = z\Big] \geq 0, \tag{52a}$$

$$\mathbb{E}\Big[d\beta Z + (1-d)\mathbb{E}\big[-\nu|-\nu \geq \beta Z, Z\big]\Big|Z = z\Big] \geq 0. \tag{52b}$$

Given Assumption 2, both $\mathbb{E}\big[\nu|\nu \geq -\beta z, z\big]$ and $\mathbb{E}\big[-\nu| -\nu \geq \beta z, z\big]$ are larger or equal to zero for any value of $\beta \in \Gamma_\beta$ and $z \in \mathbb{R}^K$. Therefore, both inequalities (52a) and (52b) hold for any value of $\beta \in \Gamma_\beta$ such that $\beta z = 0$.

Also, as long as $var(\nu) > \delta_1$, for any $\delta_1 > 0$, $\exists\ \delta_2 > 0$ such that $\|\beta\| < \delta_2$ and $\beta \in \Omega(\mathcal{M}_r, \mathcal{Z}^M)$. In words, by continuity of the functions $\mathbb{E}\big[\nu|\nu \geq -\beta Z, Z\big]$ and $\mathbb{E}\big[-\nu| -\nu \geq \beta Z, Z\big]$ in the variance of ν, as long as this variance is different from zero, we can define a small ball around $\beta = 0$ such that all the values of the parameter vector in that ball are included in $\Omega(\mathcal{M}_r, \mathcal{Z}^M)$. ■

Proof of Theorem 2.2. Lemma A.7.1 holds for any arbitrary vector $\mathcal{Z}^M$. Therefore, for any arbitrary vector $\mathcal{Z}^M$, the set $\Omega(\mathcal{M}_r, \mathcal{Z}^M)$ always includes the point $\beta = 0$. Also, or any arbitrary vector $\mathcal{Z}^M$, as long as $var(\nu) > \delta_1$, for any $\delta_1 > 0$, $\exists\ \delta_2 > 0$ such that $\|\beta\| < \delta_2$ and $\beta \in \Omega(\mathcal{M}_r, \mathcal{Z}^M)$. ■

A.8 Proof of Theorem 2.3

We show the proof for equation (14b). The proof for equation (14a) is completely equivalent. From Section A.6, we can write the revealed preference moment inequality in eq. (14b) as:

$$\mathbb{E}\Bigg[\mathbb{E}\Bigg[d(\beta_1 Z_1 + \beta_2 X^* - \varepsilon) + \\ (1-d)\mathbb{E}\big[-\nu| -\nu \geq \beta_1 Z_1 + \beta_2 X_2^* - \varepsilon, Z_1, X_2^* - (1/\beta_2)\varepsilon\big]\Bigg| X^*, Z_2\Bigg]\Bigg| Z\Bigg] \geq 0.$$

Assumption 3 implies that

$$\frac{\partial\Big\{\mathbb{E}\Big[d(\beta_1 Z_1 + \beta_2 X_2^* + \varepsilon)\Big| X^*, Z_2\Big]\Big\}}{\partial \sigma_\varepsilon} = 0.$$

Assumptions 2 and 3 and Lemma A.5.1 imply that, for any (X^*, Z_2)

$$\frac{\partial\Big\{\mathbb{E}\Big[(1-d)\mathbb{E}\big[-\nu| -\nu \geq \beta_1 Z_1 + \beta_2 X_2^* - \varepsilon), Z_1, X_2^* - (1/\beta_2)\varepsilon\big]\Big| X^*, Z_2\Big]\Big\}}{\partial \sigma_\varepsilon} \geq 0.$$

Therefore, for any $z \in \mathcal{Z}$ and $\beta \in \Gamma_\beta$,

$$\frac{\partial \mathcal{M}_r(z;\beta)}{\partial \sigma_\varepsilon} \geq 0.$$

In conclusion, if $\bar{\bar{\sigma}}_\varepsilon \geq \bar{\sigma}_\varepsilon$, then, for for any $z \in \mathcal{Z}$, and any given value of $\beta \in \Gamma_\beta$ such that $\mathcal{M}_r(z;\beta|\bar{\sigma}_\varepsilon) \geq 0$, it will be true that $\mathcal{M}_r(z;\beta|\bar{\bar{\sigma}}_\varepsilon) \geq 0$. Therefore, for any $\mathcal{Z}^M$, $\Omega(\mathcal{M}_r, \mathcal{Z}^M|\bar{\sigma}^2_\varepsilon) \subseteq \Omega(\mathcal{M}_r, \mathcal{Z}^M|\bar{\bar{\sigma}}^2_\varepsilon)$. ■

A.9 Proof of Theorem 3.2

Given that $X_2^* = X = Z_2$, we write the index function as $\beta X^* = \beta_1 Z_1 + \beta_2 X = \beta Z$. For any vector $q \in Q$ (see equation (15)), we can find some other vector $q' \in Q$ such that $q' = 1 - q$. In this way, we can partition the Q vectors into $Q/2$ pairs. Let us index these pairs of vectors by $p = 1, \ldots, Q/2$. We define each pair $p = \{q_0^p, q_1^p\}$, with $q_1^p = 1 - q_0^p$.

Both vectors in a single pair p have the property that, for every value of $z \in \mathcal{Z}$ such that $\Psi_{q_0^p}(z) = 1$, it will be true that $\Psi_{q_1^p}(-z) = 1$. Therefore, for each given p, using the assumption $X = Z_2$, we can rewrite the moment inequalities

$$\mathcal{M}_s^{q_0^p}(\beta) = \mathbb{E}\Big[\Psi_{q_0^p}(Z) \cdot m_s^+(d, Z_1, Z_2; \beta) + \Psi_{q_0^p}(-Z) \cdot m_s^-(d, Z_1, Z_2; \beta)\Big] \geq 0,$$
$$\mathcal{M}_s^{q_1^p}(\beta) = \mathbb{E}\Big[\Psi_{q_1^p}(Z) \cdot m_s^+(d, Z_1, Z_2; \beta) + \Psi_{q_1^p}(-Z) \cdot m_s^-(d, Z_1, Z_2; \beta)\Big] \geq 0,$$

as

$$\mathcal{M}_s^{q_0^p}(\beta) = \mathbb{E}\Big[\Psi_{q_0^p}(Z) \cdot m_s^+(d, Z_1, Z_2; \beta) + \Psi_{q_0^p}(-Z) \cdot m_s^-(d, Z_1, Z_2; \beta)\Big] \geq 0,$$
$$\mathcal{M}_s^{q_1^p}(\beta) = \mathbb{E}\Big[\Psi_{q_0^p}(-Z) \cdot m_s^+(d, Z_1, Z_2; \beta) + \Psi_{q_0^p}(Z) \cdot m_s^-(d, Z_1, Z_2; \beta)\Big] \geq 0.$$

or, using the LIE,

$$\mathcal{M}_s^{q_0^p}(\beta) = \mathbb{E}\Big[\Psi_{q_0^p}(Z) \cdot \mathbb{E}\big[m_s^+(d, Z_1, Z_2; \beta)\big|Z\big] + \Psi_{q_0^p}(-Z) \cdot \mathbb{E}\big[m_s^-(d, Z_1, Z_2; \beta)\big|Z\big]\Big] \geq 0,$$
$$\mathcal{M}_s^{q_1^p}(\beta) = \mathbb{E}\Big[\Psi_{q_0^p}(-Z) \cdot \mathbb{E}\big[m_s^+(d, Z_1, Z_2; \beta)\big|Z\big] + \Psi_{q_0^p}(Z) \cdot \mathbb{E}\big[m_s^-(d, Z_1, Z_2; \beta)\big|Z\big]\Big] \geq 0.$$

Using equation (13), we can write

$$\mathbb{E}\big[m_s^-(d, Z_1, Z_2; \beta)\big|Z\big] = (1 - F_\nu\big(-\beta^* Z\big))\frac{F_\nu\big(-\beta Z\big)}{1 - F_\nu\big(-\beta Z\big)} - F_\nu\big(-\beta^* Z\big),$$
$$\mathbb{E}\big[m_s^+(d, Z_1, Z_2; \beta)\big|Z\big] = F_\nu\big(-\beta^* Z\big)\frac{1 - F_\nu\big(-\beta Z\big)}{F_\nu\big(-\beta Z\big)} - (1 - F_\nu\big(-\beta^* Z\big)),$$

and, plugging these expressions into $\mathcal{M}_s^{q_0^p}(\beta)$ and $\mathcal{M}_s^{q_1^p}(\beta)$, we obtain

$$\mathcal{M}_s^{q_0^p}(\beta) =$$
$$\mathbb{E}\Big[\Psi_{q_0^p}(Z)(1 - F_\nu\big(-\beta^* Z\big))(G_\nu(\beta, Z) - 1) + \Psi_{q_0^p}(-Z)F_\nu\big(-\beta^* Z\big)\Big[\frac{1}{G_\nu(\beta, Z)} - 1\Big]\Big] \geq 0,$$
$$\mathcal{M}_s^{q_1^p}(\beta) =$$
$$\mathbb{E}\Big[\Psi_{q_0^p}(Z)F_\nu\big(-\beta^* Z\big)\Big[\frac{1}{G_\nu(\beta, Z)} - 1\Big] + \Psi_{q_0^p}(-Z)(1 - F_\nu\big(-\beta^* Z\big))(G_\nu(\beta, Z) - 1)\Big] \geq 0,$$

where

$$G_\nu(\beta, Z) = \frac{F_\nu\big(-\beta^* Z\big)}{F_\nu\big(-\beta Z\big)}\frac{1 - F_\nu\big(-\beta Z\big)}{1 - F_\nu\big(-\beta^* Z\big)}.$$

Given that $G(\beta^*, Z) = 0$ for all Z, $\mathcal{M}_s^{q_0^p}(\beta^*) = \mathcal{M}_s^{q_1^p}(\beta^*) = 0$. Also, for any $k \in K$,

$$sign\Big(\frac{\partial \mathcal{M}_s^{q_0^p}(\beta)}{\partial \beta^k}\Big) = -sign\Big(\frac{\partial \mathcal{M}_s^{q_1^p}(\beta)}{\partial \beta^k}\Big),$$

and, therefore, either $\mathcal{M}_s^{q_0^p}(\beta)$ or $\mathcal{M}_s^{q_1^p}(\beta)$ will become negative as we move away from $\beta = 0$ by moving each of the scalar components of the vector β one at a time. We can define inequalities analogous to $\mathcal{M}_s^{q_0^p}(\beta)$ or $\mathcal{M}_s^{q_1^p}(\beta)$ for every $p = 1, \ldots, Q/2$. Given that the matrix Q is the standard basis in R^K, for every $\beta \neq \beta^*$ there exists at least one p such that one of the inequalities in the pair p does not hold at that value of β. Therefore, the only value of $\beta \in \Gamma_\beta$ such that $\mathcal{M}_s^q(\beta) \geq 0$ for all $q = \{1, \ldots, Q\}$ is $\beta = \beta^*$. ■

A.10 Proof of Theorem 3.4 and Theorem 4.4

We prove boundedness and closedness for $\Omega(\mathcal{M}_s)$. The proof for $\Omega(\mathcal{M}_r)$ is analogous.

Boundedness.

Lemma A.10.1 *The set $\Omega(\mathcal{M}_s)$ is bounded if, for very element of $\mathbb{P}^K = \{p \in R^K\colon \|p\| = 1\}$, the unit sphere in R^K, there exists at least one $q \in \{1, \ldots, Q\}$ such that*

$$\lim_{c\to\infty} \left\{ c \cdot \sum_{k=1}^{K} p_k \frac{\partial \mathcal{M}_s^q(\beta)}{\partial \beta_k} \right\} = -\infty.$$

This condition means that the identified set $\Omega(\mathcal{M}_s)$ is bounded if, as we move far away from the origin in any possible direction in R^K, there is at least one inequality, $\mathcal{M}_s^q(\beta)$, that becomes negative. This moment q may be different for each $p \in \mathbb{P}^K$. A sufficient condition for Lemma A.10.1 to hold is that, for every p in $\mathbb{P}^K$, $\exists\ q \in Q$ such that, for all $\beta \in \Gamma_\beta$ and all $k \in K$,

$$sign(p_k) = -sign\left(\frac{\partial \mathcal{M}_s^q(\beta)}{\partial \beta_k}\right). \tag{53}$$

The matrix Q is the standard basis in R^K (see equation (15)). Therefore, as long as equation (18) holds, for any p in $\mathbb{P}$, $\exists\ q \in Q$ that verifies equation (53). ■

Closedness. Given that $\Omega(\mathcal{M}_s)$ is bounded, it will also be closed because all the inequalities in the vector $\mathcal{M}_s(\beta)$ are weak inequalities. ■

A.11 Proof of Theorem 3.5 and Theorem 4.5

Lemma A.11.1 *Let β and Y be two $K \cdot 1$ vectors, g be a strictly convex function, and $\mathbb{E}_Y(\cdot)$ denote the expectation with respect to the random vector Y, then $\mathbb{E}_Y(g(Y'\beta))$ is a convex* *function of the random vector β; i.e.*

$$\frac{\partial^2 \mathbb{E}_Y(g(Y'\beta))}{\partial\beta\beta'}$$

is positive definite.

Proof: Using basic algebra,

$$\frac{\partial^2 \mathbb{E}_Y(g(Y'\beta))}{\partial\beta\beta'} = \mathbb{E}_Y[YY'g''(Y'\beta)].$$

Given that g is strictly convex, $g''(Y'\beta) > 0$, $g''(Y'\beta) = \sqrt{g''(Y'\beta)}\ \sqrt{g''(Y'\beta)}$, and we can rewrite

$$\frac{\partial^2 \mathbb{E}_Y(g(Y'\beta))}{\partial\beta\beta'} = \mathbb{E}_Y[\tilde{Y}\tilde{Y}']$$

where $\tilde{Y} = Y\sqrt{g''(Y'\beta)}$. Using this equation and properties of quadratic forms, we can conclude that the Hessian of $\mathbb{E}_Y(g(Y'\beta))$ is positive definite.

Proof of Theorem 3.5 and Theorem 4.5 The set $\Omega(\mathcal{M}_s) = \{\beta \in \Gamma_\beta : \mathcal{M}_s(\beta) \geq 0\}$ is convex if all the moment inequalities in the vector $\mathcal{M}_s(\beta)$ are convex in β (see Kaido and Santos (2011)). Lemmas A.3.3 and A.11.1 show that all the elements of $\mathcal{M}_s(\beta)$ are convex in β. Therefore, we can conclude that $\Omega(\mathcal{M}_s)$ is a convex set. Analogously, since Assumption 2 and Lemma A.11.1 yield that all the elements of $\mathcal{M}_r(\beta)$ are convex in β, then $\Omega(\mathcal{M}_r)$ is also a convex set. ■

A.12 Score Function Inequalities when $\nu = 0$.

Lemma A.12.1 *The infinite support condition on Assumption 3 implies that, for any vector (d, X^*, Z_2), any $\beta \in \Gamma_\beta$,*

$$Pr(-\beta_1 Z_1 - \beta_2 X > 0 | d, X^*, Z_2) > 0.$$

Proof:

$$\begin{aligned} Pr(-\beta_1 Z_1 - \beta_2 X > 0 | d, X^*, Z_2) &= Pr(-\beta_1 Z_1 - \beta_2 X^* + \varepsilon > 0 | d, X^*, Z_2) \\ &= 1 - F_\varepsilon(\beta_1 Z_1 + \beta_2 X^* | d, X^*, Z_2) > 0. \ \blacksquare \end{aligned}$$

Proof of Lack of Identification Power of Score Function Inequalities when $\nu = 0$. Assumption 2(b) implies that, for any given real value y, $F_\nu(y) = 0$ if $y < 0$, and $F_\nu(y) = 1$ if $y \geq 0$. In Appendix A.3, we showed that we can rewrite the conditional score function inequality in equation equation (42a) as

$$\mathbb{E}\left[\mathbb{E}\left[d\mathbb{E}\left[\frac{F_\nu\big(-(\beta_1 Z_1 + \beta_2 X)\big)}{1 - F_\nu\big(-(\beta_1 Z_1 + \beta_2 X)\big)}\middle| d, X^*, Z_2\right] - (1-d)\middle| X^*, Z_2\right]\middle| Z\right] \geq 0.$$

Therefore, if ν has a degenerate distribution at 0, then

$$\frac{F_\nu\big(-(\beta_1 Z_1+\beta_2 X)\big)}{1-F_\nu\big(-(\beta_1 Z_1+\beta_2 X)\big)}=\infty,$$

for any vector (Z_1, X) such that $-(\beta_1 Z_1+\beta_2 X)\geq 0$. Given that, from Lemma A.12.1, for any $\beta\in\Gamma_\beta$

$$Pr(-\beta_1 Z_1-\beta_2 X>0|d,X^*,Z_2)\geq 0,$$

this implies that, for any $\beta\in\Gamma_\beta$,

$$\mathbb{E}\left[\frac{F_\nu\big(-(\beta_1 Z_1+\beta_2 X)\big)}{1-F_\nu\big(-(\beta_1 Z_1+\beta_2 X)\big)}\middle|d,X^*,Z_2\right]=\infty.$$

Using the LIE, we can conclude that, for any value of $\beta\in\Gamma_\beta$ and any value of the vector Z,

$$\mathcal{M}_r(z;\beta)=\mathbb{E}[m_r^-(d,Z_1,X;\beta)|Z]=\infty.$$

The proof is completely analogous for the moment function $m_r^+(d,Z_1,X;\beta)$. Therefore, if ν has a degenerate distribution at 0, then the score function moment inequalities are equal to ∞ for any value of the parameter vector β in the parameter space. Therefore, these moment inequalities have no identification power. ■

A.13 Proof of Theorem 6

Lemma A.13.1 *For any random variable Y and any constant $y\in\mathbb{R}$, it holds*

$$\mathbb{E}[Y\mathbb{1}\{Y\geq 0\}]\geq\mathbb{E}[Y\mathbb{1}\{Y\geq y\}].$$

Proof:

$$\mathbb{E}[Y\mathbb{1}\{Y\geq y\}]=\int_y Yf_Y(Y)dY,$$

where $f_Y(Y)$ is the density function of Y. Then

$$\frac{\partial\mathbb{E}[Y\mathbb{1}\{Y\geq y\}]}{\partial y}=-yf_Y(y),$$

which is positive if $y<0$ and negative if $y>0$. Therefore, $\mathbb{E}[Y\mathbb{1}\{Y\geq y\}]$ reaches it maximum at $y=0$. ■

Lemma A.13.2 *For any value of $\beta \in \Gamma_\beta$, any $d \in \{0,1\}$, and any value of $(Z_1, X) \in R^K$,*

$$\mathbb{E}[\mathsf{m}_r^-(d, Z_1, X; \beta)|Z_1, X] \geq \mathbb{E}[m_r^-(d, Z_1, X; \beta)|Z_1, X],$$
$$\mathbb{E}[\mathsf{m}_r^+(d, Z_1, X; \beta)|Z_1, X] \geq \mathbb{E}[m_r^+(d, Z_1, X; \beta)|Z_1, X],$$

where $(\mathsf{m}_r^-(d, Z_1, X; \beta), \mathsf{m}_r^+(d, Z_1, X; \beta))$ are defined in equation (29) *and $(m_r^-(d, Z_1, X; \beta)$, $m_r^+(d, Z_1, X; \beta))$ are defined in equation* (14).

Proof: We show the proof for the inequality

$$\mathbb{E}[\mathsf{m}_r^-(d, Z_1, X; \beta)|Z_1, X] \geq \mathbb{E}[m_r^-(d, Z_1, X; \beta)|Z_1, X].$$

The proof for $\mathbb{E}[\mathsf{m}_r^+(d, Z_1, X; \beta)|Z_1, X] \geq \mathbb{E}[m_r^+(d, Z_1, X; \beta)|Z_1, X]$ is completely equivalent. From equation (14a), we can rewrite $\mathbb{E}[m_r^-(d, Z_1, X; \beta)|Z_1, X]$ as

$$\begin{aligned}
&\mathbb{E}\big[(1-d)(-(\beta_1 Z_1 + \beta_2 X)) + d\mathbb{E}\big[\nu|\nu \geq -(\beta_1 Z_1 + \beta_2 X), Z_1, X\big]\big|Z_1, X\big] = \\
&\mathbb{E}[1-d|Z_1, X](-(\beta_1 Z_1 + \beta_2 X)) + \mathbb{E}[d|Z_1, X]\mathbb{E}\big[\nu|\nu \geq -(\beta_1 Z_1 + \beta_2 X), Z_1, X\big] = \\
&\mathbb{E}[1-d|Z_1, X](-(\beta_1 Z_1 + \beta_2 X)) + \mathbb{E}[d|Z_1, X]\frac{\mathbb{E}\big[\nu\mathbb{1}\{\nu \geq -(\beta_1 Z_1 + \beta_2 X)\}|Z_1, X\big]}{\mathbb{E}[d|Z_1, X]} = \\
&\mathbb{E}[1-d|Z_1, X](-(\beta_1 Z_1 + \beta_2 X)) + \mathbb{E}\big[\nu\mathbb{1}\{\nu \geq -(\beta_1 Z_1 + \beta_2 X)\}|Z_1, X\big].
\end{aligned}$$

From Lemma A.13.1,

$$\mathbb{E}\big[\nu\mathbb{1}\{\nu \geq -(\beta_1 Z_1 + \beta_2 X)\}|Z_1, X\big] \leq \mathbb{E}\big[\nu\mathbb{1}\{\nu \geq 0\}|Z_1, X\big],$$

and, therefore,

$$\begin{gathered}
\mathbb{E}[m_r^-(d, Z_1, X; \beta)|Z_1, X] \leq \\
\mathbb{E}[1-d|Z_1, X](-(\beta_1 Z_1 + \beta_2 X)) + \mathbb{E}\big[\nu\mathbb{1}\{\nu \geq 0\}|Z_1, X\big] = \mathbb{E}[\mathsf{m}_r^-(d, Z_1, X; \beta)|Z_1, X]. \quad \blacksquare
\end{gathered}$$

Lemma A.13.3 *For any value of $\beta \in \Gamma_\beta$ and any $q = 1, \ldots, Q$, $\mathsf{M}_r^q(\beta) \geq \mathcal{M}_r^q(\beta)$.*

Proof: Using the LIE, we can write $\mathsf{M}_r^q(\beta)$ as

$$\mathbb{E}\Big[\mathbb{E}[\Psi_q(Z)|Z_1, X]\mathbb{E}[\mathsf{m}_r^+(d, Z_1, X; \beta)|Z_1, X] + \mathbb{E}[\Psi_q(-Z)|Z_1, X]\mathbb{E}[\mathsf{m}_r^-(d, Z_1, X; \beta)|Z_1, X]\Big],$$

and $\mathcal{M}_r^q(\beta)$ as

$$\mathbb{E}\Big[\mathbb{E}[\Psi_q(Z)|Z_1, X]\mathbb{E}[m_r^+(d, Z_1, X; \beta)|Z_1, X] + \mathbb{E}[\Psi_q(-Z)|Z_1, X]\mathbb{E}[m_r^-(d, Z_1, X; \beta)|Z_1, X]\Big].$$

Given Lemma A.13.2, it is immediate that $\mathsf{M}_r^q(\beta) \geq \mathcal{M}_r^q(\beta)$. $\blacksquare$

Proof of Theorem 6 From Theorem 4.1, we know that, for every $q = 1, \ldots, Q$, $\mathcal{M}_r^q(\beta^*) \geq 0$. Therefore, Lemma A.13.3 implies, for every $q = 1, \ldots, Q$, $\mathsf{M}_r^q(\beta^*) \geq 0$. By definition, if for every $q = 1, \ldots, Q$, $\mathsf{M}_r^q(\beta^*) \geq 0$, then $\beta^* \in \Omega(\mathsf{M}_r)$. $\blacksquare$