Identifying and Estimating Perceived Returns to Binary Investments*

Clint M. Harris†

June 9, 2020

Abstract

I introduce a new method for estimating agents' perceived returns to binary investments that exploits the mechanical relationship between perceived prices and perceived returns in binary choice settings. Identification is achieved using instruments for prices that are uncorrelated with both price misperceptions and unobserved components of perceived returns. The method provides estimates of perceived returns in terms of compensating variation, which naturally implies effects for subsidies and taxes. These estimates condition on observed characteristics, allowing for heterogeneity in predicted subsidy and tax effects across types of individuals. Because these estimates are of distributions instead of point elasticities, they imply effects that are nonlinear in policy magnitude. I demonstrate the advantages of the new method relative to two related alternatives in a series of data simulations.

*I thank Mary Kate Batistich, Trevor Gallen, Kendall Kennedy, Soojin Kim, Dan Millimet, Kevin Mumford, Victoria Prowse, and Miguel Sarzosa as well as seminar participants at Case Western Reserve University, The European Association of Labor Economists Meeting, Kansas State University, The Midwest Economics Association Meeting, The National Tax Association Meeting, Purdue University, The Southern Economic Association Meeting, and The US Census Bureau for helpful comments.

†Wisconsin Institute for Discovery, University of Wisconsin-Madison, 330 N Orchard Street, Madison, WI 53715 USA; email: clint.harris@wisc.edu

JEL Codes: C31, D84, D61

Keywords: Biased Beliefs, Moment Inequalities, Returns to Investments, Revealed Preference, Subsidies, Taxes

1 Introduction

In this paper I introduce a method for estimating distributions of perceived private returns to binary investments. These structural perceived returns estimates are of type-specific distributions of agents' compensating variation associated with a binary choice. This method complements program evaluation methods that estimate effects of specific policy shocks on binary choices by allowing for predictions of counterfactual policies that differ from past policies in magnitude or targeted population. For instance, Harris (2020) applies this method to estimate perceived returns to college, allowing for counterfactual predictions of targeted college attendance subsidies (and taxes) for diverse groups of individuals. The gain in the breadth of counterfactual implications is achieved by assuming common effects of perceived prices on perceived returns across agent types while using price instruments that are uncorrelated with both unobserved perceived returns and price misperceptions.

The method of this paper is applicable to any binary choice setting. It has particular value in settings with a high potential for information frictions. The inference as to which settings these are is a matter for researcher judgment, but may include choices in noncompetitive environments (where mistakes won't cause exit, making it hard to learn from other agents' outcomes), choices that are made infrequently (making it hard to learn from one's own past outcomes), or choices that past research suggests, for instance using surveys, may be subject to incomplete information. Settings in which pecuniary costs are unknown to agents are particularly attractive targets of the method because of its robustness to this challenge. Among others, healthcare, education, home purchases, R&D, and firm export decisions appear to be attractive targets for the method due to their large overlap with the above criteria.

This method introduced in this paper relies on revealed preference arguments in a binary choice framework. The scale of the latent variable in these models is not identified from the observed choice. The core identifying assumption throughout this paper is that prices perceived by agents have a known (to the researcher and to agents) causal effect on latent perceived returns.[1] This allows introduction of a constraint that adds an additional equation to a system that is otherwise underidentified by a single parameter, leading to complete identification. The empirical challenge addressed by the method is to validate this structural constraint.

Two salient alternatives to the identification strategy proposed in this paper exist. The first is to assume a mapping between elicited perceived returns and perceived returns, though economists typically stop short of such assumptions, assuming instead that elicited responses proxy for perceived returns (Jensen, 2010; Wiswall and Zafar, 2015; Bleemer and Zafar, 2018). The second is to assume a mapping between estimated actual returns and perceived returns, such as rational expectations. This sort of assumption is ubiquitous in economics, but Cunha, Heckman, and Navarro (2005) and related research surveyed by Cunha and Heckman (2007) are particularly interested in the estimation of perceived returns, rather than making these assumptions only as a means to some other end.

The validity of each of these three broadly defined methods in applications is a function of the validity of researcher assumptions on the mappings between objects in data and agents' beliefs. If these assumptions are correct, all three methods will provide consistent results. In practice, researchers should choose between methods based on application-specific intuition regarding plausibility of the identification assumptions, data availability, and the external validity of results as determined by consistency with those of existing experimental or quasi-experimental studies.

Applications using elicited perceived returns naturally require data containing survey responses of specific questions. Applications using assumptions on actual returns require longitudinal data, given that actual returns accrue over time. The present method requires cross-sectional data that contains observed choices and prices. It is thus useful not only in applications where the identifying assumptions are preferred to those of alternatives, but in those in which these alternative methods cannot be applied due to data limitations. As cross-sectional data on choices and prices is much more common than

[1]For simplicity, the exposition of this paper involves a monetary price that affects perceived returns at a one to one rate. Of course, many things can be thought of as prices. Generally, the estimated scale of perceived returns will be inherited from scale of the chosen price variable.

longitudinal data and surveys of subjective expectations, the present method has much wider applicability.

Though explicitly estimating agents' perceptions is uncommon, studies that estimate effects of prices on choices are ubiquitous and wide-ranging in both topic and methodology. These studies vary in their modelling of agent beliefs, most commonly either making stronger assumptions than those in this paper such as perfect information or rational expectations (Berry, Levinsohn, and Pakes, 1995; Cheng, 2014; Pakes, Porter, Ho, and Ishii, 2015) or avoiding any information assumptions altogether (Dynarski, 2003; Eriksen and Ross, 2015; Frean, Gruber, and Sommers, 2017).[2] The present study stakes out a middle ground between these extremes by developing a semi-structural framework that allows for strong counterfactual predictions while still allowing for systematic mistakes in agents' beliefs.

I show how to estimate perceived returns using, respectively, a single-stage maximum likelihood method, a moment inequality method, and a maximum likelihood control function method. The single-stage maximum likelihood method and moment inequality method are implemented as in Dickstein and Morales (2018) (henceforth, DM), but the identification assumptions are necessarily different here because the structural parameters of interest are different.[3] The three methods presented differ in the settings under which the chosen constraint is justified. Using simulated data, I show that the new control function method dominates the alternatives in consistency and precision across settings.

The plan of the rest of this paper is as follows. Section 2 introduces the empirical model. Section 3 describes the econometric strategy and the assumptions required for identification for each of the three methods. Section 4 evaluates the robustness of each

[2]While commonly-applied program evaluation methods estimate effects of policy shocks on behavior without taking a stance on agents' beliefs, these studies nonetheless commonly discuss information transmission as an important channel. For instance, Deming and Dynarski (2010) suggest information frictions as one explanation for the relatively small estimated effects of large price shifts on college attendance in the case of Pell Grants, while Dynarski, Libassi, Michelmore, and Owen (2018) propose the elimination of information frictions as an explanation for relatively large estimated effects of a policy that had a relatively small effect on actual prices.

[3]DM make the assumption that the relationship between revenues and profits in firm export decisions is defined by a known demand elasticity parameter. In short, this paper makes assumptions on the exogeneity of instrumental variables in place of assumptions on agents' beliefs about market structure.

method to various empirical challenges in a series of simulated data exercises. Section 5 concludes.

2 Model

The two-sector generalized Roy (1951) model provides a helpful framework for considering selection based on potential outcomes. Agents choose to select the investment, $S_i = 1$ or to not do so, $S_i = 0$. I define $Y_{1,i}$ as agent i's perceived discounted present value of lifetime earnings associated with choosing the investment and $Y_{0,i}$ as their perceived discounted present value of lifetime earnings associated with not doing so. I further define C_i as their perceived net present value cost of making the investment, which includes psychic costs, their preferences over nonpecuniary outcomes associated with selection, and prices paid all expressed in monetary values. Unlike common applications of the Roy model, $Y_{1,i}$, $Y_{0,i}$, and C_i are not observed by the researcher for any individual because they represent agent perceptions.

Given some explanatory variables X_i, I express the perceived potential outcomes and costs for individual i with the following linear-in-parameters production functions:

$$\begin{aligned} Y_{1,i} =& X_i\beta_1 + \epsilon_{1,i} \\ Y_{0,i} =& X_i\beta_0 + \epsilon_{0,i} \\ C_i =& X_i\beta_C + \widetilde{Price}_i\gamma + \epsilon_{Ci}, \end{aligned} \tag{1}$$

where agent i's belief about their price, $\widetilde{Price}_i$, contributes only to the perceived pecuniary cost of selection at known marginal rate γ and $\epsilon_{0,i}$, $\epsilon_{1,i}$, and ϵ_{Ci} are mean zero error terms.[4] Assuming perfect credit markets and either risk-neutral agents or complete markets with totally diversifiable risks, the perceived earnings and monetary-equivalent costs are sufficient to define the decision rule of the agents. Agents choose between alternatives

[4]Pricing may vary at the individual level due to price discrimination and geographic and temporal variation in supply and demand.

in order to maximize perceived returns such that:

$$S_i = \begin{cases} 1 \text{ if } Y_{1,i} - Y_{0,i} - C_i \geq 0, \\ 0 \text{ otherwise.} \end{cases} \tag{2}$$

Defining the perceived return $Y_i \equiv Y_{1,i} - Y_{0,i} - C_i$, the net marginal effects $\beta \equiv \beta_1 - \beta_0 - \beta_C$ and the net error $\epsilon_i \equiv \epsilon_{1,i} - \epsilon_{0,i} - \epsilon_{Ci}$, I write the perceived return to selection in terms of explanatory variables as

$$Y_i = X_i\beta - \widetilde{Price}_i\gamma + \epsilon_i. \tag{3}$$

Note that the under the assumptions above, the perceived return is linear in perceived price, identifying it as compensating variation. Relaxations of these assumptions are left to future research.

It follows from the model that β, γ, $\widetilde{Price}_i$, and the distribution of ϵ_i are sufficient to identify the distribution of latent perceived returns, Y_i. A standard probit or similar procedure would estimate a scale-invariant version of perceived returns by normalizing the scale of the distribution of ϵ_i, assuming perfect information on prices, and estimating scale-normalized β and γ. The methods of this paper depart from this by instead imposing the structural constraint $\gamma = 1$, such that prices are assumed to be fully borne by the decision maker, and estimating β and the scale of the distribution of ϵ_i while allowing for imperfect information on prices.[5] Note that an accurate assumption on any model parameter, such as an element within β, would identify the model; γ is chosen because of the strong intuition for the true effect of perceived prices on perceived returns. The assumption on γ is not an innocuous normalization, it is valid only when imposed in the context of a structural model that identifies the causal effect of perceived prices on

[5]In the presence of taxes or subsidies, the method can be applied with an alternative assumption on γ that encompasses the institutional details, or by conceiving of $\widetilde{Price}_i$ as the perceived net price actually faced by the agent. Because the relative effects of X_i and $\widetilde{Price}_i$ on S_i are fundamental to any binary choice model, bias in the assumed value of γ will produce proportionate bias in the estimates of all other model parameters.

perceived returns. I present three methods below that differ in the steps they take to validate this assumption.

3 Empirical Strategy

The methods employed in this paper address three main challenges. First, the scale of perceived returns, Y_i, is not identified from the observed choice, S_i. This is addressed by constraining γ to its true value in the context of a structural model. Second, because this model must admit a causal interpretation of γ as the effect of perceived prices on perceived returns, using observed prices in place of unobserved perceived prices in the model will threaten the validity of this assumption if observed prices measure unobserved perceived prices with error. Third, even if it were observed, correlation between $\widetilde{Price}_i$ and ϵ_i would threaten the validity of the causal assumption on γ.

It is helpful to the exposition to begin by presenting a common model of agents' beliefs about prices,

$$Price_i = \widetilde{Price}_i + X_i\alpha + \nu_i, \tag{4}$$

where prices observed by the researcher are composed of a component that is known to agents, $\widetilde{Price}_i$, an unknown component that is explained by the controls, $X_i\alpha$, (where α gives the price misperceptions attributable to each control), and an unknown mean-zero residual, ν_i. Rational expectations amounts to the assumption $\alpha = 0$ while perfect information adds the assumption that $\nu_i = 0 \ \forall i$.

This assumption allows us to present an empirically tractable version of perceived returns,

$$\begin{aligned} Y_i &= X_i\beta - \widetilde{Price}_i\gamma + \epsilon_i \\ &= X_i\beta - Price_i\gamma + X_i\alpha\gamma + \nu_i\gamma + \epsilon_i \\ &= X_i\theta - Price_i\gamma + \nu_i\gamma + \epsilon_i, \end{aligned} \tag{5}$$

by substituting in prices observed by the researcher for agents' unobserved perceived prices and defining $\theta = \beta + \alpha\gamma$.[6] This representation helpfully presents unexplained price misperceptions as omitted variables, which by construction will causes bias due to correlation between $Price$ and ν.

First, I consider a setting in which misperceptions on prices are fully explained by X_i ($\nu_i = 0\ \forall i$) and perceived prices are exogenous ($Cov(\widetilde{Price}, \epsilon) = 0$). In this model, the prices observed by the researcher can credibly stand in for perceived prices in the latent variable equation. In this setting simple binary choice methods consistently estimate perceived returns when imposing the constraint $\gamma = 1$.

Second, I consider a setting in which agents have misperceptions on prices that are not captured by X_i while maintaining that both $\widetilde{Price}_i$ and ν_i are uncorrelated with ϵ_i. In this setting, the minor adjustments to standard methods described above would provide consistent estimates if we could observe agents' unobserved perceived prices, but will provide biased estimates of the scale of perceived returns when using observed prices. Here, the moment inequalities developed by DM provide bounds on perceived returns using instruments, Z_i, which are independent of agents' misperceptions about prices.

Third, I consider a setting in which agents have misperceptions on prices that are not captured by X_i while allowing for both $\widetilde{Price}_i$ and ν_i to be correlated with ϵ_i. This allows for the error in perceived returns to vary systematically with prices faced by agents, and with how incorrect agents are about prices. In this setting, even if we observed perceived prices, their correlation with unobserved components of perceived returns, ϵ_i would produce bias in estimates of the effect of prices on selection in the standard binary choice framework, which will invalidate the assumed value for the coefficient on perceived prices, γ. I present a control function method that is robust to this setting in addition to those above which uses instruments, Z_i, that are uncorrelated with both agents' misperceptions about prices, ν_i, and unobserved components of perceived returns, ϵ_i.

[6]The distinction between the extent to which each control contributes to misperceptions in prices, α, and to perceived returns, β, is presented to emphasize that the methods in this paper are robust to systematic bias in perceptions, even though they are not separately identified.

3.1 Degenerate Misperceptions on Exogenous Prices

First, I will describe a baseline method for estimating perceived returns with a simple adjustment to common binary choice methods. This procedure will provide consistent estimates of the perceived returns distribution under two assumptions that are likely to be violated in most applications. First, this method assumes that prices and the perceived returns error term are uncorrelated. Second, it assumes that misperceptions about prices are fully explained by X_i. It is presented first because it illustrates the common logic of the methods of this paper in simple terms.

Recalling the specification of perceived returns given in (5), these assumptions imply that

$$\begin{aligned} Y_i &= X_i\beta - Price_i\gamma + X_i\alpha\gamma + \epsilon_i \\ &= X_i\theta - Price_i\gamma + \epsilon_i. \end{aligned} \tag{6}$$

Here it is apparent that controlling for the explanatory variables X_i is sufficient to achieve conditional unconfoundedness for observed prices in this model. Assuming, for example, that the distribution of the error term is given by

$$\epsilon_i | X_i, Price_i \sim \mathcal{N}(0, \sigma^2), \tag{7}$$

and applying the decision rule described in (2) is sufficient to consistently estimate perceived returns by maximum likelihood.[7] Defining $(\beta^*, \theta^*, \gamma^*) = (\frac{\beta}{\sigma}, \frac{\theta}{\sigma}, \frac{\gamma}{\sigma})$ for notational convenience, the probability of selection is given by

$$Pr(S_i = 1 | X_i, Price_i) = \Phi(X_i\theta^* - Price_i\gamma^*), \tag{8}$$

where $\Phi(\cdot)$ denotes the standard normal CDF. The parameters (θ^*, γ^*) are the values

[7] The normality assumption is unnecessary for any of the methods in this paper. For the current method, any error distribution is acceptable.

that maximize the log-likelihood:

$$\mathcal{L}(\theta^*, \gamma^*|X_i, Price_i) = \sum_i S_i \log\left[\Phi\Big(X_i\theta^* - Price_i\gamma^*\Big)\right] + (1 - S_i)\log\left[1 - \Phi\Big(X_i\theta^* - Price_i\gamma^*\Big)\right]. \tag{9}$$

The estimates of perceived returns are then given by

$$\hat{Y}_i|X_i, Price_i \sim \mathcal{N}(X_i\hat{\theta} - Price_i\gamma, \hat{\sigma}^2), \tag{10}$$

where imposing the assumed value on γ (rather than the standard constraint $\sigma = 1$) is the only difference from a standard probit.

3.2 Non-Degenerate Misperceptions on Exogenous Prices

It is unlikely that a researcher would have access to all variables associated with misperceptions on prices. For example, individuals often make investment decisions before prices are fully realized, such as when college tuition changes while a student is attending. Grodsky and Jones (2007) find that individuals' elicited beliefs about prices for college were consistently inaccurate, with misperceptions unexplained by a wide variety of controls.

This section adapts the moment inequality method developed by DM to the current setting. The moment inequalities allow for information frictions on prices ($Var(\nu_i) \neq 0$) but maintain that $\widetilde{Price}_i$ and ν_i are independent of ϵ_i.[8] The misperceptions on prices thus produce problems analogous to those associated with classical measurement error, and probit estimates of the effect of $Price_i$ on selection will be biased.

This method makes use of instruments, Z_i that are independent of (ϵ_i, ν_i). For $X_i \subset$

[8] DM address endogeneity through a structural model in which export revenues forecasts export profits at a rate defined by the demand elasticity. In a model analogous to that of this paper, this rate would instead be 1.

Z_i, taking the expectation of (4) conditional on Z_i implies

$$\mathbb{E}[Price_i|Z_i] = \mathbb{E}[\widetilde{Price}_i|Z_i] + X_i\alpha, \tag{11}$$

such that observed prices can stand in for unobserved perceived prices. The method uses two types of moment inequalities to obtain bounds on the parameters of perceived returns, (θ, σ). I will present the inequalities and provide a brief discussion here. For the derivation and further discussion of the moment inequalities, see DM.

3.2.1 Revealed Preference Moment Inequalities

Maintaining that $(\beta^*, \theta^*, \gamma^*) = (\frac{\beta}{\sigma}, \frac{\theta}{\sigma}, \frac{\gamma}{\sigma})$ for notational convenience, the conditional revealed preference moment inequalities are

$$\begin{aligned} &\mathbb{E}\left[S_i(X_i\theta^* - Price_i\gamma^*) + (1 - S_i)\frac{\phi(X_i\theta^* - Price_i\gamma^*)}{1 - \Phi(X_i\theta^* - Price_i\gamma^*)}\middle|Z_i\right] \geq 0, \\ &\mathbb{E}\left[-(1 - S_i)(X_i\theta^* - Price_i\gamma^*) + S_i\frac{\phi(X_i\theta^* - Price_i\gamma^*)}{\Phi(X_i\theta^* - Price_i\gamma^*)}\middle|Z_i\right] \geq 0. \end{aligned} \tag{12}$$

These inequalities are consistent with the revealed preference argument that perceived returns are positive for those who select the investment and negative for those who do not. Here, I provide a brief overview of the intuition.

Regarding the first inequality, consider an agent that selects the investment such that $S_i = 1$. Following the revealed preference argument articulated in (2) and the representation of perceived returns in (3), it follows that this individual's perceived return is positive, i.e.,

$$S_i(X_i\beta - \widetilde{Price}_i\gamma + \epsilon_i) \geq 0. \tag{13}$$

This expression cannot be computed directly because researchers do not observe $\widetilde{Price}_i$ or ϵ_i. However, given the independence of ν_i and ϵ_i (and by extension S_i), we can substitute prices for perceived prices in the expectation conditional on Z_i using (11),

$$\mathbb{E}[S_i(X_i\theta - Price_i\gamma + \epsilon_i)|Z_i] \geq 0. \tag{14}$$

The second term in the first inequality in (12) is a biased approximation of $\mathbb{E}[S_i\epsilon_i|Z_i]$ that exploits the closed form for $\mathbb{E}[S_i\epsilon_i|X_i, \widetilde{Price}_i]$ under the normality assumption on ϵ_i.[9]

Heuristically, if observed prices are a mean-preserving spread of perceived prices, substituting them in place of perceived prices will mistakenly increase expected perceived returns unconditional on selection for some agents and decrease them for others. For the agents for whom this expectation increases, the expectation of the error conditional on selection approaches zero. For those for whom it decreases, the expectation of the error conditional on selection approaches positive infinity. In many cases, this second effect will dominate the overall expectation of the error conditional on selection.[10] The second inequality follows from similar intuition applied to individuals who do not select the investment.

3.2.2 Odds-Based Moment Inequalities

The conditional odds-based moment inequalities are

$$\begin{aligned}
&\mathbb{E}\left[\left(S_i\frac{1-\Phi(X_i\theta^* - Price_i\gamma^*)}{\Phi(X_i\theta^* - Price_i\gamma^*)} - (1-S_i)\right)\middle| Z_i\right] \geq 0,\\
&\mathbb{E}\left[\left((1-S_i)\frac{\Phi(X_i\theta^* - Price_i\gamma^*)}{1-\Phi(X_i\theta^* - Price_i\gamma^*)} - S_i\right)\middle| Z_i\right] \geq 0.
\end{aligned} \tag{15}$$

They are derived from the conditional score equation,

$$\mathbb{E}\left[S_i\frac{\phi(X_i\beta^* - \widetilde{Price}_i\gamma^*)}{\Phi(X_i\beta^* - \widetilde{Price}_i\gamma^*)} - (1-S_i)\frac{\phi(X_i\beta^* - \widetilde{Price}_i\gamma^*)}{1-\Phi(X_i\beta^* - \widetilde{Price}_i\gamma^*)}\middle| X_i, \widetilde{Price}_i\right] = 0. \tag{16}$$

[9]The bias makes substitution of prices for perceived prices nontrivial, and contributes to the inequality.

[10]Global convexity of $\mathbb{E}[\epsilon_i|\epsilon_i < \kappa]$ in κ is necessary for the inequalities to hold regardless of the value of κ and the variance of the misperception term. This condition is satisfied by both the normal and logistic distributions.

Considering the second inequality, the score function can be rearranged to be a function of the model-predicted odds of selecting the investment,

$$\mathbb{E}\left[\left((1-S_i)\frac{\Phi(X_i\beta^* - \widetilde{Price}_i\gamma^*)}{1-\Phi(X_i\beta^* - \widetilde{Price}_i\gamma^*)} - S_i\right)\middle| X_i, \widetilde{Price}_i\right] = 0. \tag{17}$$

The advantage of this transformation is that the odds-ratio is globally convex in its arguments. Replacing the unobserved $\widetilde{Price}_i$ with $Price_i$ changes the equation into an inequality by application of Jensen's inequality due to the global convexity of the odds ratio. As the index of the odds ratio increases, the model-predicted odds of a given outcome approach positive infinity, while the odds approach zero as the index decreases. When the index is replaced with a mean-preserving spread of itself (via replacing perceived prices with prices), this first effect will usually dominate the second regardless of the distributional assumption.[11] This inequality holds when taking its expectation conditional on Z_i by law of iterated expectations. The first inequality follows from similar intuition for those who do not select the investment.

3.2.3 Estimation Using Moment Inequalities

Under the information assumptions provided, the true parameters $\psi = (\theta, \sigma)$ will be contained within the set of parameters that satisfy the inequalities, which I define as Ψ_0. First, because it is computationally expensive to compute the inequalities conditional on Z_i, I will instead use unconditional inequalities that are consistent with the conditional inequalities described above. Additionally, in small samples it is possible that the true parameters will not strictly satisfy these inequalities, so it is necessary to construct a test of the hypothesis that a given value $\psi_p = (\theta_p, \sigma_p)$ is consistent with the inequalities. To do this I employ the modified method of moments procedure described by Andrews and Soares (2010), which yields a confidence set of parameters $\hat{\Psi}_0$ that I fail to reject are consistent with the inequalities, where an element of this set is given by $(\hat{\theta}_p, \hat{\sigma}_p)$. A

[11] Global convexity of the odds ratio is necessary for this condition to hold for all values of the index and for all magnitudes of mean-preserving spreads. This condition is satisfied by log-concave distributions, such as the normal and logistic.

description of the estimation procedure is provided in Appendix A.

To infer estimates of perceived returns from the estimated set of parameters that satisfy the moment inequalities, first note that, given the true (β, σ), perceived returns are given by

$$Y_i | X_i, \widetilde{Price}_i \sim \mathcal{N}(X_i\beta - \widetilde{Price}_i\gamma, \sigma^2). \tag{18}$$

Thus, even given the true (β, σ), the problem remains that we do not observe $\widetilde{Price}_i$ in the data. However, it is possible to bound perceived returns at the true parameter values using $Price_i$ and $\mathbb{E}[Price_i|Z_i]$, which we do have access to.

For valid Z_i, equation (4) implies that it is possible to approximate $\widetilde{Price}_i$ with $\varphi Price_i + (1-\varphi)\mathbb{E}[Price_i|Z_i]$, where φ minimizes $\mathbb{E}[\widetilde{Price}_i - (\varphi Price_i + (1-\varphi)\mathbb{E}[Price_i|Z_i])]^2$. It must be that $\varphi \in [0,1]$, though we cannot estimate it. Given $\varphi \in [0,1]$, bounds on perceived returns for a given $(\hat{\theta}_p, \hat{\sigma}_p)$ can be constructed using

$$Y_i | X_i, Price_i, Z_i \sim \mathcal{N}\big(X_i\hat{\theta}_p - \varphi Price_i\gamma - (1-\varphi)\mathbb{E}[Price_i|Z_i]\gamma, \hat{\sigma}_p^2\big). \tag{19}$$

Note that the PDF of this distribution is non-monotonic in φ, so setting $\varphi = 0$ and $\varphi = 1$ will not bound its PDF across its entire support. Computing the distribution for all $\varphi \in [0,1]$ for each $(\hat{\theta}_p, \hat{\sigma}_p) \in \hat{\Psi}_0$ is necessary to provide bounds for the perceived returns distribution.[12]

3.3 Non-Degenerate Misperceptions on Endogenous Prices

In this section I develop a control function approach that has several advantages over the moment inequalities for the current application. First, it is robust to price misperceptions that are correlated with perceived returns as well as endogeneity in perceived prices. Second, it provides point estimates of model parameters. Third, it is substantially less computationally costly, allowing for the inclusion of a richer set of explanatory

[12] In practice, choosing any set of values between zero and one, including zero and one, will approximate these bounds. DM describe an alternative method that can be used to bound the CDF of perceived returns.

variables. Fourth, it provides estimates of the association between the first stage error in prices and perceived returns. These last two provides for a broader set of heterogeneous policy predictions conditional on observed covariates. Fifth, it places no restrictions on distribution assumptions for the error term.[13]

The control function approach uses instruments Z_i, with $X_i \subset Z_i$, that are uncorrelated with both ν_i and ϵ_i in the following system of equations:

$$\begin{aligned} Y_i &= X_i\beta - \widetilde{Price_i}\gamma + \epsilon_i \\ \widetilde{Price_i} &= Z_i\pi + u_i \\ Price_i = \widetilde{Price_i} + X_i\alpha + \nu_i &= Z_i\delta + u_i + \nu_i, \end{aligned} \tag{20}$$

wherein π provides the mapping of the instruments to perceived prices, while δ provides the mapping to actual prices. The second line represents what would be the first stage in a two-step instrumental variables procedure if perceived prices were observable. The third line combines the assumption on beliefs given in (4) with the first stage on perceived prices to obtain an estimable first stage equation in terms of observed prices.[14]

Given the above, I derive a control function that can be used to estimate perceived returns as follows,[15]

$$\begin{aligned} Y_i =& X_i\beta - \widetilde{Price_i}\gamma + \epsilon_i \\ =& X_i\theta - Price_i\gamma + \nu_i\gamma + \epsilon_i \\ =& X_i\theta - Price_i\gamma + (u_i + \nu_i)\rho + \eta_i \\ =& X_i\theta - Price_i\gamma + (\widehat{u_i + \nu_i})\rho + \zeta_i. \end{aligned} \tag{21}$$

The second line follows directly from the representation of beliefs about prices given in

[13]This last advantage is quite small in practice, as the most commonly invoked distributional assumptions satisfy the conditions required for the moment inequalities.

[14]The presentation here assumes a parametric first stage. It is also possible to implement a nonparametric first stage using the expectation of $Price_i$ conditional on Z_i. The parametric specification is presented because it is likely more familiar to readers, but should not be viewed as a restriction of the control function method.

[15]As an closely-related alternative, we could perform a instrumental variables probit to obtain identical estimates of θ. The control function method has the advantage of conditioning on the first stage error, which produces more precise person-specific estimates of perceived returns.

(4). The third line substitutes in the linear projection of the composite error $\nu_i\gamma + \epsilon_i$ on the first stage error $u_i + \nu_i$, wherein $\rho = \mathbb{E}[(u_i + \nu_i)(\nu_i\gamma + \epsilon_i)]/\mathbb{E}[(u_i + \nu_i)^2]$. The fourth line substitutes the estimated residuals from the first stage regression of $Price_i$ on Z_i in for their unobserved true values, generating a new error, ζ_i. This new error will converge asymptotically to η_i, but will differ in small samples due to sampling error in the estimation of the residual from the first stage, $(\widehat{u_i + \nu_i})$. Note that it is unnecessary (and impossible) in this setting to distinguish between u_i and ν_i, and that (u_i, ν_i, ϵ_i) can be freely correlated if Z_i is a valid instrument for perceived prices.

To estimate perceived returns, I assume that the new error in the perceived returns control function expression is normally distributed,

$$\eta_i | X_i, Price_i, (u_i + \nu_i) \sim \mathcal{N}(0, \sigma^2_\eta), \tag{22}$$

noting that the variance of the error of η_i will differ from that of ϵ_i if $\rho \neq 0$. The log-likelihood for the second stage of the control function approach is then given by

$$\begin{aligned}
&\mathcal{L}\Big(\theta^*, \gamma^*, \rho^* | X_i, (\widehat{u_i + \nu_i})\Big) = \\
&\sum_i S_i \log\left[\Phi\Big(X_i\theta^* - Price_i\gamma^* + (\widehat{u_i + \nu_i})\rho^*\Big)\right] \\
&+(1 - S_i)\log\left[1 - \Phi\Big(X_i\theta^* - Price_i\gamma^* + (\widehat{u_i + \nu_i})\rho^*\Big)\right].
\end{aligned} \tag{23}$$

I present the likelihood of only the second stage for simplicity, and because the insights of this paper are invariant to the estimator used for the first stage. In the simulations, I assume normality on the first stage error and use two-stage conditional maximum likelihood to estimate perceived returns, following Rivers and Vuong (1988), correcting for the inclusion of estimated regressors following Murphy and Topel (1985). Estimates of perceived returns are obtained by plugging the estimated parameters and the assumed coefficient on perceived prices into the latent variable equation:

$$Y_i | X_i, (\widehat{u_i + \nu_i}) \sim \mathcal{N}\Big(X_i\hat{\theta} - Price_i\gamma + (\widehat{u_i + \nu_i})\hat{\rho}, \hat{\sigma}^2_\eta\Big). \tag{24}$$

4 Simulated Data

In this section I apply the methods described above to simulated datasets to compare their performance. Here, I examine three cases that demonstrate the limitations of the methods, and in Appendix B I provide additional simulations and a computational comparison of the three methods. In summary, the control function method performs well across simulations, the moment inequalities are remarkably resilient to settings that differ from that described in Section 3.2 but nonetheless provide less precise estimates than the control function method at greater computational cost, while the probit method performs well only in one particularly well-behaved data-generating process (DGP). Due to the computational costs of the moment inequality method, I present a single simulation for each DGP rather than Monte Carlos with multiple simulations.

For each DGP, I assume rational expectations on prices such that $\alpha = 0$ and $\delta = \pi$. The only implication of this choice is to simplify the interpretation of the results; under these conditions $\theta = \beta$. I use the following DGP,

$$\begin{aligned} Y_i &= X_i\beta - \widetilde{Price}_i\gamma + \epsilon_i \\ Price_i &= \widetilde{Price}_i + \nu_i = Z_i\delta + u_i + \nu_i, \end{aligned} \tag{25}$$

where Z_i is uncorrelated with ϵ_i, u_i, and ν_i, and the nature of the covariance structure on these error terms will determine which methods will and will not provide consistent estimates of perceived returns. The control function method will also obtain estimates of

$$\begin{aligned} \rho &= \mathbb{E}[(u_i + \nu_i)(\nu_i\gamma + \epsilon_i)]/\mathbb{E}[(u_i + \nu_i)^2], \\ \sigma_\eta &= \sqrt{Var(\eta_i)} = \sqrt{Var(\nu_i\gamma + \epsilon_i - (u_i + \nu_i)\rho)}. \end{aligned} \tag{26}$$

Each DGP is comprised of $N = 10,000$ observations of agents whose decisions are governed by their perceived returns to selection. I construct the instrument vector as $Z_i = [X_i \; z_i]$ where X_i always includes only a constant, and z_i is a single instrument. Finally, I assume $\gamma = 1$, the constant $\beta_0 = 1$, and $\delta = [0 \; 1]'$ for all DGPs.

4.1 Perfect Information on Exogenous Prices

I begin with an unrealistically well-behaved benchmark DGP that corresponds to the setting described in Section 3.1. I generate data according to

$$\begin{bmatrix} z_i \\ u_i \\ \nu_i \\ \epsilon_i \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \Sigma); \quad \Sigma = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}, \tag{27}$$

where I include ν with a variance of zero for clarity.

Table 1 shows perceived returns estimates for one simulation of this DGP using all three methods. Figure 1 shows the distributions implied by the estimates for each method. Because this DGP is particularly well-behaved, all three methods' estimates are very close to the data-generating parameters. Additionally, the moment inequalities provide very tight bounds here because the first-stage error has relatively low variance such that making use of $\mathbb{E}[Price_i|Z_i]$ in place of $\widetilde{Price_i}$ introduces little uncertainty into the estimated perceived returns.

Table 1: Simulation 1, Perceived Returns Estimates

	Target	(1) Probit	(2) Moment Inequalities	(3) Control Function
Constant	1	0.986	[0.906, 1.066]	0.997
		(0.033)	N/A	(0.034)
σ	2	2.071	[1.976, 2.236]	.
		(0.040)	N/A	
σ_η	2	.	.	2.092
				(0.043)
$\rho_{u\nu}$	0	.	.	-0.051
				(0.036)
Observations		10000	10000	10000

Notes: Standard errors in parentheses, corrected for the inclusion of estimated regressors following Murphy and Topel (1985) in the case of the control function. Parameters are in monetary units. Estimates relate to expressions (10), (19), and (24), respectively. All data is generated in Stata using random seed 1234.

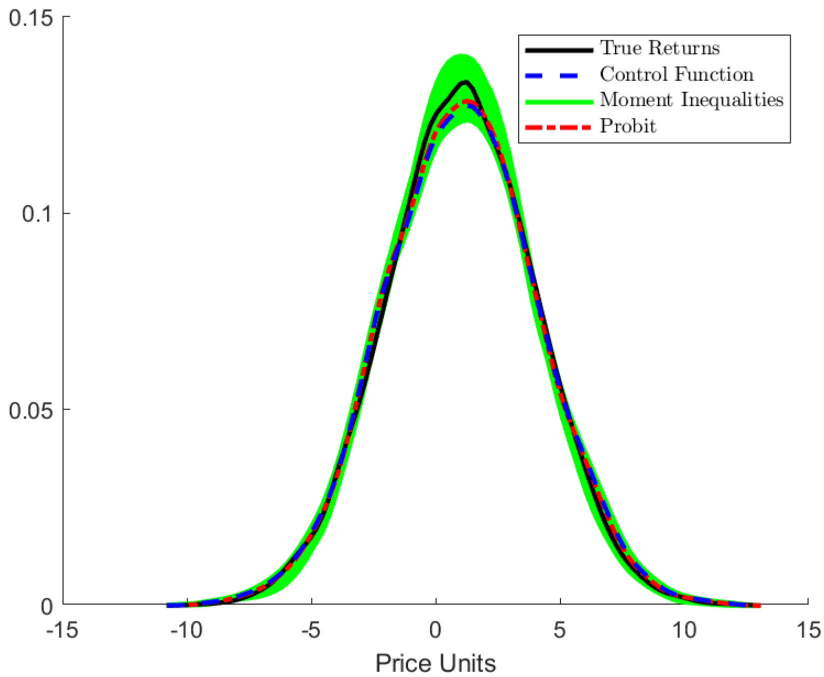

Figure 1: Simulation 1, Implied Perceived Returns Distributions

Notes: Estimated densities of perceived returns given by each method. Densities for each parameter vector in the moment inequalities' 95% confidence set are shown using $\varphi = [0, 1]$ with steps of 1/4.

4.2 Imperfect Information on Exogenous Prices

In this simulation, I consider a DGP that corresponds to the setting described in Section 3.2 in which agents do not precisely forecast prices such that $Price_i \neq \widetilde{Price}_i$. I generate data according to

$$\begin{bmatrix} z_i \\ u_i \\ \nu_i \\ \epsilon_i \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \Sigma); \quad \Sigma = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}. \tag{28}$$

Table 2 shows the estimates for one simulation of this DGP using all three methods. Figure 2 shows the distributions implied by the estimates for each method. The control function and moment inequality estimates are close to the true parameters. The control function estimates are significantly more precise than those of the moment inequalities. The probit's estimates are biased upward as predicted, given the normality assumptions on $\widetilde{Price}_i$ and ν_i (Yatchew and Griliches, 1985).

Table 2: Simulation 2, Perceived Returns Estimates

	Target	(1) Probit	(2) Moment Inequalities	(3) Control Function
Constant	1	1.657	[-1.757, 3.621]	0.932
		(0.079)	N/A	(0.067)
σ	2	5.294	[0.038, 4.173]	.
		(0.106)	N/A	
σ_η	3	.	.	2.932
				(0.103)
$\rho_{u\nu}$	.5	.	.	0.501
				(0.020)
Observations		10000	10000	10000

Notes: Standard errors in parentheses, corrected for the inclusion of estimated regressors following Murphy and Topel (1985) in the case of the control function. Parameters are in monetary units. Estimates relate to expressions (10), (19), and (24), respectively. All data is generated in Stata using random seed 1234.

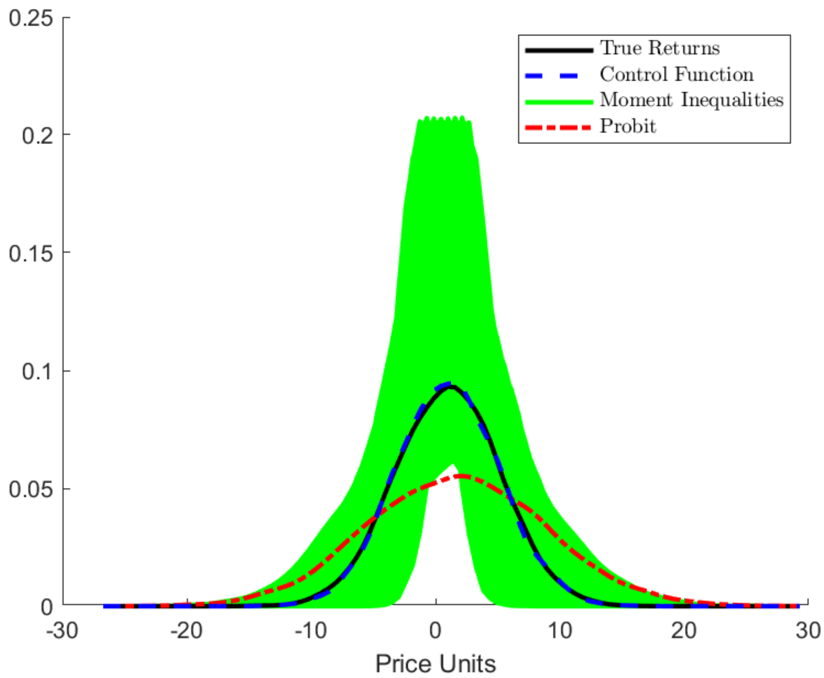

Figure 2: Simulation 2, Implied Perceived Returns Distributions

Notes: Estimated densities of perceived returns given by each method. Densities for each parameter vector in the moment inequalities' 95% confidence set are shown using $\varphi = [0, 1]$ with steps of 1/4.

4.3 Imperfect Information with Negatively Correlated Perceptions

In this simulation, I consider a DGP in which ν_i and ϵ_i are negatively correlated. In words, this provides that agents who incorrectly perceive high prices also (correctly or otherwise) perceive high returns. This setting is one case of that described in Section 3.3. I generate data according to

$$\begin{bmatrix} z_i \\ u_i \\ \nu_i \\ \epsilon_i \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \Sigma); \quad \Sigma = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 1 \\ 0 & 0 & 4 & -9 \\ 0 & 1 & -9 & 25 \end{bmatrix}. \tag{29}$$

Table 3 shows the estimates for one simulation of this DGP using all three methods. Figure 3 shows the distributions implied by the estimates for each method. The control function method's estimates are close to the true parameter values, while the the other methods' estimates are not.

In this case $\rho < 0$, which is relevant for the performance of the moment inequalities (see Appendix C). In short, this produces bias in probit estimates that is the opposite of that predicted by independent measurement error as described in Section 3.2, which suggests that the arguments provided there will not apply. This is because the positive selection on perceived prices indicated by $Cov(u, \epsilon) > 0$ is dominated by negative selection from misperceived prices indicated by $Cov(\nu, \epsilon) < 0$.

5 Conclusions

I introduce a novel control function method for estimating perceived returns to binary investments. I compare this method to two related alternatives adapted from previously established methods and show that the control function method dominates these alternatives in precision and consistency in a variety of simulations. This method is directly

Table 3: Simulation 3, Perceived Returns Estimates

	Target	(1) Probit	(2) Moment Inequalities	(3) Control Function
Constant	1	0.711	[-0.160, 1.144]	0.963
		(0.038)	N/A	(0.065)
σ	5	2.387	[0.049, 1.773]	.
		(0.043)	N/A	
σ_η	3	.	.	3.138
				(0.108)
$\rho_{u\nu}$	-.5	.	.	-0.544
				(0.051)
Observations		10000	10000	10000

Notes: Standard errors in parentheses, corrected for the inclusion of estimated regressors following Murphy and Topel (1985) in the case of the control function. Parameters are in monetary units. Estimates relate to expressions (10), (19), and (24), respectively. All data is generated in Stata using random seed 1234.

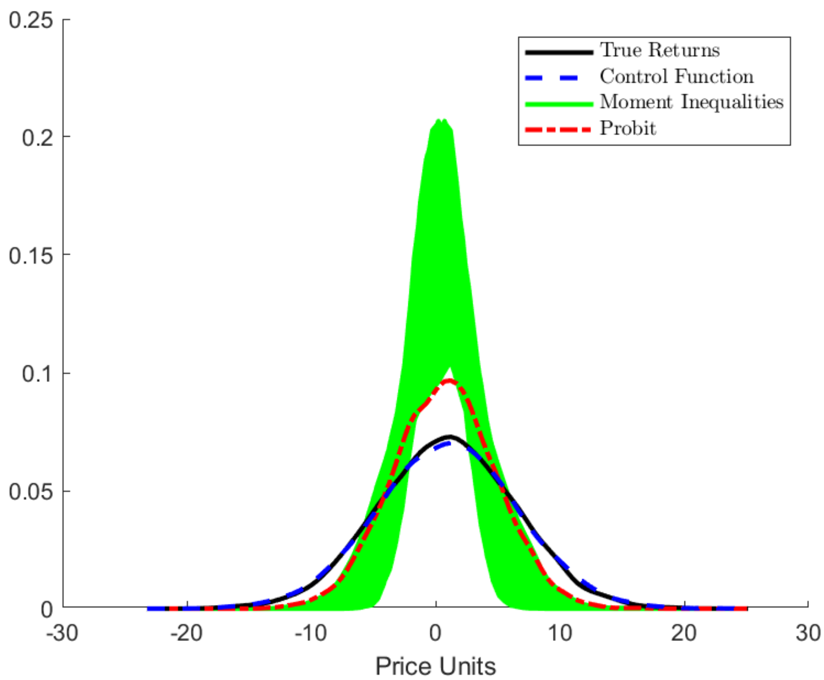

Figure 3: Simulation 3, Implied Perceived Returns Distributions

Notes: Estimated densities of perceived returns given by each method. Densities for each parameter vector in the moment inequalities' 95% confidence set are shown using $\varphi = [0, 1]$ with steps of 1/4.

applicable to many empirical questions, especially those subject to substantial information frictions such as such as college attendance, firm R&D, and automobile, home, or health insurance purchases.

A major advantage of the general empirical framework implemented here is that it provides estimates of distributions of perceived returns conditional on observed characteristics. This allows for differential predictions on the effects of group-specific subsidies or taxes on uptake for the investment. Importantly, the estimates are obtained in terms of compensating variation, which means they are directly applicable to analysis of the effects of subsidies and taxes. Finally, it also allows for differential predictions depending on the magnitude of such subsidies, something that point elasticities do not provide. Each method presented in this paper has these advantages, while the control function method is robust to the most empirical challenges.

It is worthwhile to discuss the scope for welfare gains from the type of policies described above, given that the methods presented in this paper have social value primarily insofar as they inform such policies. One source for welfare gains is from wedges between actual private returns and actual social returns (externalities). The second is from wedges between perceived private returns and actual private returns (information frictions). Investigation of these wedges is fundamental to policy solutions to socially inefficient choices. This paper develops methods that can be used in tandem with estimates of benefits of various decisions to inform policies that can potentially induce selection on gains.[16]

The control function method developed here shows promise for theoretical extensions to more complicated choice environments, such as multinomial or ordered decision processes. Additionally, it may be possible to nest it within other techniques, such as that of BLP (1995) or its extensions. Finally, applying the approach with constraints on variables other than prices (such as loans) can provide estimates of perceptions of objects

[16] Joint estimation of actual returns and perceived returns using the methods in this paper is left to future research. Such estimates have obvious policy relevance, but should be interpreted with caution due to the infeasibility of separately identifying misperceptions and subjective preferences using observed choices.

other than returns (such as credit constraints). Such extensions are left to future work.

References

Andrews, D. W., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica: Journal of the Econometric Society*, pp. 841–890.

Bleemer, Z., and B. Zafar (2018): "Intended College Attendance: Evidence from an Experiment on College Returns and Costs," *Journal of Public Economics*, 157, 184–211.

Cheng, T. C. (2014): "Measuring the Effects of Reducing Subsidies for Private Insurance on Public Expenditure for Health Care," *Journal of Health Economics*, 33, 159–179.

Cunha, F., J. Heckman, and S. Navarro (2005): "Separating Uncertainty from Heterogeneity in Life Cycle Earnings," *Oxford Economic Papers*, 57(2), 191–261.

Cunha, F., and J. J. Heckman (2007): "Identifying and estimating the distributions of ex post and ex ante returns to schooling," *Labour Economics*, 14(6), 870–893.

Deming, D., and S. Dynarski (2010): "College Aid," in *Targeting Investments in Children: Fighting Poverty When Resources Are Limited*, pp. 283–302. University of Chicago Press.

Dickstein, M. J., and E. Morales (2018): "What Do Exporters Know?," *The Quarterly Journal of Economics*, 133(4), 1753–1801.

Dynarski, S., C. Libassi, K. Michelmore, and S. Owen (2018): "Closing the gap: The effect of a targeted, tuition-free promise on college choices of high-achieving, low-income students," Discussion paper, National Bureau of Economic Research.

Dynarski, S. M. (2003): "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion," *American Economic Review*, 93(1), 279–288.

Eriksen, M. D., and A. Ross (2015): "Housing vouchers and the price of rental housing," *American Economic Journal: Economic Policy*, 7(3), 154–76.

Frean, M., J. Gruber, and B. D. Sommers (2017): "Premium subsidies, the mandate, and Medicaid expansion: Coverage effects of the Affordable Care Act," *Journal of Health Economics*, 53, 72–86.

Grodsky, E., and M. T. Jones (2007): "Real and Imagined Barriers to College Entry: Perceptions of Cost," *Social Science Research*, 36(2), 745–766.

Harris, C. M. (2020): "Estimating the perceived returns to college," *Available at SSRN 3577816*.

Jensen, R. (2010): "The (Perceived) Returns to Education and the Demand for Schooling," *The Quarterly Journal of Economics*, 125(2), 515–548.

Murphy, K. M., and R. H. Topel (1985): "Least Squares with Estimated Regressors," *Journal of Business and Economic Statistics*.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment Inequalities and Their Application," *Econometrica*, 83(1), 315–334.

Rivers, D., and Q. H. Vuong (1988): "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics*, 39(3), 347–366.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135–146.

Wiswall, M., and B. Zafar (2015): "How Do College Students Respond to Public Information about Earnings?," *Journal of Human Capital*, 9(2), 117–169.

Yatchew, A., and Z. Griliches (1985): "Specification error in probit models," *The Review of Economics and Statistics*, pp. 134–139.

Appendix A: Moment Inequality Estimation

I closely follow appendices A.5 and A.7 in Dickstein and Morales (2018) to estimate the moment inequalities' confidence set for the true parameter ψ. Adopting DM's procedure to the current setting would account for imputation of prices. I use a simplified version of their procedure, because I assume that prices are observed for all individuals, regardless of whether they select the investment. This assumption is irrelevant to the contributions of this paper, as each method admits imputation. I also deviate from DM in how I conduct the grid search over potential parameters in order to speed computation in the absence of parallelization.

The confidence set is based on the Andrews and Soares (2010) modified method of moments (MMM) test statistic. I index the moment inequalities used in estimation by $\ell = 1, ..., L$ and denote them

$$\bar{m}_\ell(\psi) \equiv \frac{1}{N}\sum_{i=1}^{N} m_\ell(Z_i, \psi), \qquad \ell = 1, ..., L,$$

where N is the sample size. The MMM test statistic

$$Q(\psi) = \sum_{\ell}^{L}[\min(\sqrt{N}\frac{\bar{m}_\ell(\psi)}{\hat{\sigma}_\ell(\psi)}, 0)]^2, \tag{A.1}$$

gives the sum of squared inequality violations, where

$$\hat{\sigma}_\ell(\psi) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(m_\ell(Z_i, \psi) - \bar{m}_\ell(\psi))^2}.$$

Note that as in Section 3.2, $X_i \subset Z_i$. $m_\ell(\cdot)$ is a conditional revealed preference or odds-based moment inequality constructed as described in DM, Appendix A.5. I compute a confidence set for the true parameter ψ using the following steps, closely following DM.

Step 1: define a grid Ψ_g that overlaps with the confidence set. I define this grid as a K-dimensional orthotope where K is the number of scalars indexed by $k = 1, ..., K$ within the parameter vector ψ. To define this grid, I choose ψ_{min} to minimize $Q(\psi)$, initializing the minimization with the control function estimates $\hat{\psi}_{CF} \equiv (\hat{\theta}_{CF}, \hat{\sigma}_{\eta,CF})$, which in simulations is typically near a minimum (zero) of $Q(\psi)$. The moment inequality confidence set encompass the control function estimates in simulations when they provide consistent bounds (see Appendix B), and there is good reason to believe that this will be the case generally (see Appendix C). Because $Q(\hat{\psi}_{min})$ will be close to zero, it is likely to be within the 95% confidence set, $\hat{\Psi}_0^{95}$, if this set is nonempty. I create boundaries in dimension k by multiplying the standard error of the kth parameter by a large number, and adding and subtracting this value from the parameter to form bounds in the kth dimension.[17] I repeat this for each of the K parameters to obtain bounds on a K-dimensional initial grid Ψ_g. I fill this grid with 10^K equidistant points.

Step 2: choose a point $\psi_p \in \Psi_g$. For speed, I test points in ascending order of their euclidean distance from $\hat{\psi}_{min}$. With ψ_p, I test the hypothesis that $\psi_p = \psi$:

$$H_0 : \psi = \psi_p \qquad vs. \qquad H_0 : \psi \neq \psi_p.$$

Step 3: evaluate the MMM test statistic at ψ_p:

$$Q(\psi_p) = \sum_{\ell}^{L} \left[\min(\sqrt{N}\frac{\bar{m}_\ell(\psi_p)}{\hat{\sigma}_\ell(\psi_p)}, 0) \right]^2, \tag{A.2}$$

Step 4: compute the correlation matrix of the moments evaluated at ψ_p:

$$\hat{\Omega}(\psi_p) = Diag^{-\frac{1}{2}}(\hat{\Sigma}(\psi_p))\hat{\Sigma}(\psi_p)Diag^{-\frac{1}{2}}(\hat{\Sigma}(\psi_p)),$$

where $Diag^{-\frac{1}{2}}(\hat{\Sigma}(\psi_p))$ is the $L \times L$ diagonal matrix that shares diagonal elements with $\hat{\Sigma}(\psi_p)$. $Diag^{-\frac{1}{2}}(\hat{\Sigma}(\psi_p))$ satisfies $Diag^{-\frac{1}{2}}(\hat{\Sigma}(\psi_p))Diag^{-\frac{1}{2}}(\hat{\Sigma}(\psi_p)) = Diag^{-1}(\hat{\Sigma}(\psi_p))$ where

$$\hat{\Sigma}(\psi_p) = \frac{1}{N}\sum_{i=1}^{N}(m(Z_i, \psi_p) - \bar{m}(\psi_p))(m(Z_i, \psi_p) - \bar{m}(\psi_p))',$$

[17] As there are negligible computational disadvantages from having a very large initial grid, I multiply the standard errors by 20.

$m(Z_i, \psi_p) = (m_1(Z_i, \psi_p), ..., m_L(Z_i, \psi_p))$, and $\bar{m}(\psi_p) = (\bar{m}_1(\psi_p), ..., \bar{m}_L(\psi_p))$, where

$$\bar{m}_\ell(\psi_p) \equiv \frac{1}{N}\sum_{i=1}^{N} m_\ell(Z_i, \psi_p), \qquad \forall \ell = 1, ..., L.$$

Step 5: simulate the asymptotic distribution of $Q(\psi_p)$. Take $R = 1000$ draws from the multivariate normal distribution $\mathcal{N}(0_L, I_L)$ where 0_L is a vector of zeros and I_L is an L-dimensional identity matrix. Denote each of these draws as χ_r. Define the criterion function $Q_{N,r}^{AA}(\psi_p)$ as

$$Q_{N,r}^{AA}(\psi_p) = \sum_{\ell=1}^{L}\left[\left(\min\left([\hat{\Omega}^{\frac{1}{2}}(\psi_p)\chi_r]_\ell, 0\right)\right)^2 \times \mathbb{1}\left(\sqrt{N}\frac{\bar{m}_\ell(\psi_p)}{\hat{\sigma}_\ell(\psi_p)} \leq \sqrt{\ln N}\right)\right],$$

where $[\hat{\Omega}^{\frac{1}{2}}(\psi_p)\chi_r]_\ell$ is the ℓth element of the vector $\hat{\Omega}^{\frac{1}{2}}(\psi_p)\chi_r$.

Step 6: compute the critical value. The critical value $\hat{c}_N^{AA}(\psi_p, 1-\alpha)$ is the $(1-\alpha)$-quantile distribution of the distribution of $Q_{N,r}^{AA}(\psi_p)$ across the R draws taken in step 5.

Step 7: reject or fail to reject ψ_p. If $Q(\psi_p) \leq \hat{c}_N^{AA}(\psi_p, 1-\alpha)$, include ψ_p in the estimated $(1-\alpha)\%$ confidence set, $\hat{\Psi}^{1-\alpha}$ and the (initially empty) grid $\Psi_{g'}$ that will contain the confidence set.

Step 8: repeat steps 2 through 7 until a ψ_p is not rejected. This will likely occur at the first point checked, ψ_{min}, as this parameter minimizes $Q(\psi_p)$ - though it does not maximize $\hat{c}_N^{AA}(\psi_p, 1-\alpha)$.

Step 9: form a small grid around each ψ_p in $\hat{\Psi}^{1-\alpha}$. Form $\Psi_{g,p}$, a local K-dimensional orthotope with 3 equidistant points in each dimension (with distance between points defined as in step 1), centered around ψ_p for each ψ_p in $\hat{\Psi}^{1-\alpha}$. Add $\Psi_{g,p}$ to the grid $\Psi_{g'}$ that will contain the confidence set.

Step 10: repeat steps 3 through 7 for every point in $\Psi_{g'}$ that has not yet been checked.

Step 11: iterate on steps 9 and 10 until all points in $\Psi_{g'}$ have been checked.

Step 12: ensure desired grid fineness. If the number of elements of the set $\hat{\Psi}^{1-\alpha}$ is below the desired minimum number, set the distance between grid points at one-half of the current value and repeat step 11. Repeat this step until the number of elements of $\hat{\Psi}^{1-\alpha}$ exceeds the desired number of such elements.

Appendix B: Additional Simulations

This section presents additional simulations. First, I present a series of variations on the setting described in Section 3.3, where the magnitudes and directions of selection and misperception biases vary. These simulations demonstrate the robustness of the control function method to a wide variety of empirical settings, while also demonstrating robustness of the moment inequalities in settings other than that described in Section 3.2 as well. Second, I present simulations with additional explanatory variables in order to demonstrate the computational advantages of the control function method.

As in the body of the paper, for each DGP, I assume rational expectations on prices such that $\alpha = 0$ and $\delta = \pi$. The only implication of this choice is to simplify the interpretation of the results; under these conditions $\theta = \beta$. I use the following DGP,

$$\begin{aligned} Y_i &= X_i\beta - \widetilde{Price}_i\gamma + \epsilon_i \\ Price_i &= \widetilde{Price}_i + \nu_i = Z_i\delta + u_i + \nu_i, \end{aligned} \tag{B.1}$$

where Z_i is uncorrelated with ϵ_i, u_i, and ν_i, and the nature of the covariance structure on these error terms will determine which methods will and will not provide consistent estimates of perceived returns. The control function method will also obtain estimates of

$$\begin{aligned} \rho &= \mathbb{E}[(u_i + \nu_i)(\nu_i\gamma + \epsilon_i)]/\mathbb{E}[(u_i + \nu_i)^2], \\ \sigma_\eta &= \sqrt{Var(\eta_i)} = \sqrt{Var(\nu_i\gamma + \epsilon_i - (u_i + \nu_i))}. \end{aligned} \tag{B.2}$$

Each DGP is comprised of $N = 10,000$ observations of agents whose decisions are governed by their perceived returns to selection. I construct the instrument vector as $Z_i = [X_i \ z_i]$ where X_i always includes only a constant unless otherwise stated, and z_i is a single instrument. Finally, I assume $\gamma = 1$, the constant $\beta_0 = 1$, and $\delta = [0 \ 1]'$ for all DGPs.

B.1 Endogeneity Exploration

B.1.1 Perfect Information on Positively Selected Prices

In this simulation, I consider a DGP in which u_i and ϵ_i are positively correlated. This setting is a case of the one described in Section 3.3. I generate data according to

$$\begin{bmatrix} z_i \\ u_i \\ \nu_i \\ \epsilon_i \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \Sigma); \quad \Sigma = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 12 & 0 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 6 & 0 & 4 \end{bmatrix}. \tag{B.3}$$

where I include ν with a variance of zero for clarity.

Table B.1 shows the estimates for one simulation of this DGP using all three methods. Figure B.1 shows the distributions implied by the estimates for each method. The control function method estimates are close to the true parameter values, while the true β and σ_η are also encompassed by the moment inequality 95% confidence set. Note that the moment inequalities' 95% confidence set does not include σ because the assumptions of Section 3.2 are violated.

Note that although this setting deviates from that described in Section 3.2, the moment inequalities still bound perceived returns in this simulation. This appears to be due to the value of ρ. In short, $\rho \in [0, 1]$ restricts the endogeneity bias to be similar in magnitude and sign to what we expect from the information friction bias in the setting described in Section 3.2. A formal investigation of the moment inequalities' performance in this setting is provided in Appendix B.

Finally, it is interesting to note that the probit estimate of σ happens to be close to the true value in this simulation, despite the probit failing badly to estimate perceived returns, as is evident from Figure B.1. The probit assigns all variance in $Price_i$ to perceived returns, while the moment inequality and control function methods do not. The probit thus estimates a distribution of perceived returns with too high of variance, despite the estimated variance of the error term being close to the true value in this particular DGP.

Table B.1: Simulation B.1.1, Perceived Returns Estimates

	Target	(1) Probit	(2) Moment Inequalities	(3) Control Function
Constant	1	1.611	[-1.518, 3.739]	0.996
		(0.039)	N/A	(0.057)
σ	2	2.066	[0.121, 1.200]	.
		(0.039)	N/A	
σ_η	1	.	.	1.020
				(0.045)
$\rho_{u\nu}$	.5	.	.	0.482
				(0.018)
Observations		10000	10000	10000

Notes: Standard errors in parentheses, corrected for the inclusion of estimated regressors following Murphy and Topel (1985) in the case of the control function. Parameters are in monetary units. Estimates relate to expressions (10), (19), and (24), respectively. All data is generated in Stata using random seed 1234.

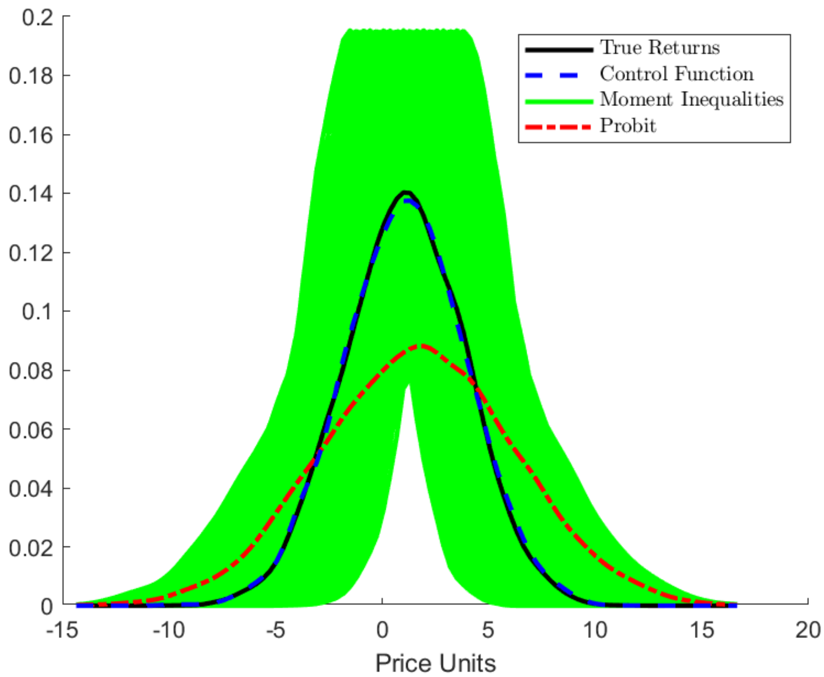

Figure B.1: Simulation B.1.1, Implied Perceived Returns Distributions

Notes: Estimated densities of perceived returns given by each method. Densities for each parameter vector in the moment inequalities' 95% confidence set are shown using $\varphi = [0, 1]$ with steps of 1/4.

B.1.2 Perfect Information on Negatively Selected Prices

In this simulation, I consider a DGP in which u and ϵ are negatively correlated. I generate data according to

$$\begin{bmatrix} z_i \\ u_i \\ \nu_i \\ \epsilon_i \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \Sigma); \quad \Sigma = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 12 & 0 & -6 \\ 0 & 0 & 0 & 0 \\ 0 & -6 & 0 & 4 \end{bmatrix}. \tag{B.4}$$

where I include ν with a variance of zero for clarity. This setting corresponds to the one described in Section 3.3. This simulation is substantively similar to the preceding one. It is included to provide a demonstration of the moment inequalities' inconsistent robustness to endogeneity, despite their performance in the preceding simulation.

Table B.2 shows the estimates for one simulation of this DGP using all three methods. Figure B.2 shows the distributions implied by the estimates for each method. The control function method estimates are close to the true parameter values, while the other methods' estimates are biased.

Table B.2: Simulation B.1.2, Perceived Returns Estimates

	Target	(1) Probit	(2) Moment Inequalities	(3) Control Function
Constant	1	0.731	[-0.713, 1.634]	1.062
		(0.024)	N/A	(0.078)
σ	2	0.994	[0.082, 0.337]	.
		(0.023)	N/A	
σ_η	1	.	.	1.030
				(0.064)
$\rho_{u\nu}$	-.5	.	.	-0.517
				(0.054)
Observations		10000	10000	10000

Notes: Standard errors in parentheses, corrected for the inclusion of estimated regressors following Murphy and Topel (1985) in the case of the control function. Parameters are in monetary units. Estimates relate to expressions (10), (19), and (24), respectively. All data is generated in Stata using random seed 1234.

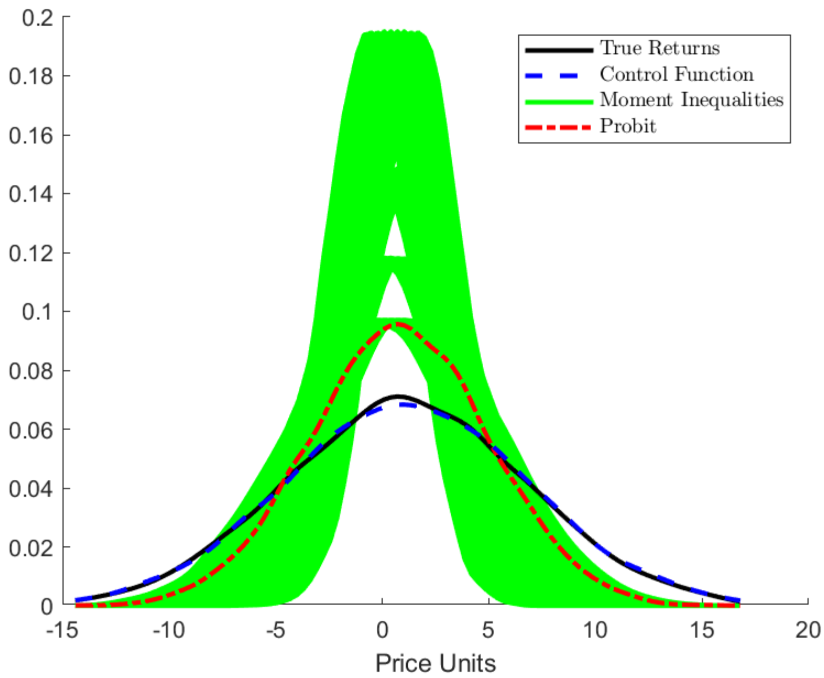

Figure B.2: Simulation B.1.2, Implied Perceived Returns Distributions

Notes: Estimated densities of perceived returns given by each method. Densities for each parameter vector in the moment inequalities' 95% confidence set are shown using $\varphi = [0, 1]$ with steps of 1/4.

B.1.3 Imperfect Information with Selection on Prices

In this simulation, I consider a DGP in which perceived prices are correlated with the error, and there are information frictions in prices. This setting is likely similar to those which occur naturally when there are substantial information frictions on prices. Correlation between the information friction ν_i and the potentially endogenous component of perceived returns, u_i, could be added but will create no complications for the method as only the composite error $u_i + \nu_i$ affects the estimation. I generate data according to

$$\begin{bmatrix} z_i \\ u_i \\ \nu_i \\ \epsilon_i \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \Sigma); \quad \Sigma = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 12 & 0 & 4 \\ 0 & 0 & 4 & 0 \\ 0 & 4 & 0 & 4 \end{bmatrix}. \tag{B.5}$$

Table B.3 shows the estimates for one simulation of this DGP using all three methods. Figure B.3 shows the distributions implied by the estimates for each method. The control function method estimates are close to the true parameter values, while the moment inequality 95% confidence set also encompasses them. The probit method's estimates are biased.

Table B.3: Simulation B.1.3, Perceived Returns Estimates

	Target	(1) Probit	(2) Moment Inequalities	(3) Control Function
Constant	1	1.591	[-1.691, 3.511]	0.910
		(0.057)	N/A	(0.059)
σ	2	3.569	[0.042, 2.990]	.
		(0.066)	N/A	
σ_η	2	.	.	1.918
				(0.067)
$\rho_{u\nu}$	.5	.	.	0.505
				(0.018)
Observations		10000	10000	10000

Notes: Standard errors in parentheses, corrected for the inclusion of estimated regressors following Murphy and Topel (1985) in the case of the control function. Parameters are in monetary units. Estimates relate to expressions (10), (19), and (24), respectively. All data is generated in Stata using random seed 1234.

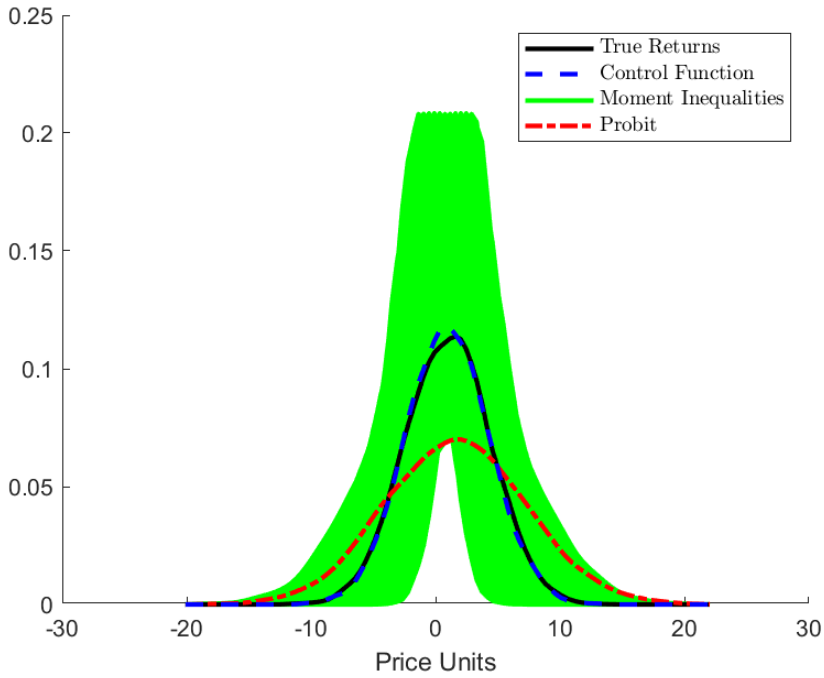

Figure B.3: Simulation B.1.3, Implied Perceived Returns Distributions

Notes: Estimated densities of perceived returns given by each method. Densities for each parameter vector in the moment inequalities' 95% confidence set are shown using $\varphi = [0, 1]$ with steps of 1/4.

B.2 Computational Comparison

Next, I present two simulations which include additional explanatory variables. This exercise is intended to demonstrate the computational advantages of the control function method over the moment inequality method, so they include the computation time taken to complete each procedure. These simulations use the DGP described in Section B.1.3 with the addition of the variables x_1 in the first simulation, and x_1 and x_2 in the second, where both coefficients are zero. I set $x_1 \sim \mathcal{N}(0, 4)$ and $x_2 \sim \mathcal{N}(0, 4)$. The results are shown in Table B.4 and Table B.5, respectively, where graphs of implied perceived returns are omitted because they are the same as figure B.3. All simulations are performed on a Linux server with two Intel Xeon X5550 CPUs and 48GB of RAM. Note that the run times in seconds for the moment inequalities are orders of magnitude higher than the control function for both simulations, and that this difference is increasing in the number of variables.

Table B.4: Simulation B.2.1, Perceived Returns Estimates

	Target	(1) Probit	(2) Moment Inequalities	(3) Control Function
Constant	1	1.591	[-10.202, 11.550]	0.911
		(0.057)	N/A	(0.059)
x_1	0	0.032	[-1.158, 1.196]	0.019
		(0.028)	N/A	(0.027)
σ	2	3.568	[0.042, 2.990]	
		(0.066)	N/A	
σ_η	2		[0.042, 2.990]	1.918
			N/A	(0.067)
$\rho_{u\nu}$	.5			0.505
				(0.018)
Observations		10000	10000	10000
Computation Time		0	3443	3

Notes: Standard errors in parentheses, corrected for the inclusion of estimated regressors following Murphy and Topel (1985) in the case of the control function. Parameters are in monetary units. Estimates relate to expressions (10), (19), and (24), respectively. All data is generated in Stata using random seed 1234. Computation time is rounded to the nearest whole second.

Table B.5: Simulation B.2.2, Perceived Returns Estimates

	Target	(1) Probit	(2) Moment Inequalities	(3) Control Function
Constant	1	1.591	[-23.443, 25.500]	0.911
		(0.057)	N/A	(0.059)
x_1	0	0.032	[-1.693, 1.732]	0.019
		(0.028)	N/A	(0.027)
x_2	0	0.002	[-1.517, 1.738]	0.002
		(0.028)	N/A	(0.027)
σ	2	3.568	[.0420, 4.0614]	
		(0.066)	N/A	
σ_η	2		[.0420, 4.0614]	1.918
			N/A	(0.067)
$\rho_{u\nu}$	.5			0.505
				(0.018)
Observations		10000	10000	10000
Computation Time		1	104485	2

Notes: Standard errors in parentheses, corrected for the inclusion of estimated regressors following Murphy and Topel (1985) in the case of the control function. Parameters are in monetary units. Estimates relate to expressions (10), (19), and (24), respectively. All data is generated in Stata using random seed 1234. Computation time is rounded to the nearest whole second.

Appendix C: Moment Inequalities and Endogeneity

For proofs of the validity of the moment inequalities for providing a confidence set that cannot be rejected as containing the true parameter vector, $\{\theta, \sigma\}$, in the context of information frictions on prices presented in Section 3.2, see DM. The inequalities also appear to consistently bound perceived returns in simulations when there is correlation between perceived prices and the unobserved error in perceived returns and correlation between information frictions and the unobserved error in perceived returns under the assumption $\frac{\rho}{\gamma-\rho} \geq 0$, which is weaker than the assumption described in Section 3.2. I provide proofs of consistency here for the revealed preference moment inequalities, and arguments for consistency for the odds-based moment inequalities, borrowing from the proofs provided by DM. Note that the parameters relevant to this section are those associated with the control function approach, (θ, σ_η), not those used in Section 3.2. I use the notation $(\theta^*, \gamma^*, \rho^*) = \left(\frac{\theta}{\sigma_\eta}, \frac{\gamma}{\sigma_\eta}, \frac{\rho}{\sigma_\eta}\right)$ throughout the following.

The condition $\frac{\rho}{\gamma-\rho} \geq 0$ entails a setting in which both endogeneity and measurement error work against the causal effect of price on selection, yet the causal effect dominates, producing attenuation bias in estimates of the effect of perceived prices on selection if misperceptions and endogeneity are ignored. Heuristically, this restriction suggests that including $(u_i + \nu_i)$ with a multiplier of $-\gamma$ (recalling the definition of prices given in (20)) will strengthen the overall effect of prices on inequalities derived from (21) relative to multiplying this error by $(-\gamma + \rho)$.

Given that a positive γ is suggested by the law of demand, a slightly less general statement of this condition is that $\rho \in [0, \gamma]$. The assumption that $\rho \geq 0$ seems plausible, prices may be higher for individuals for whom perceived returns for selection are higher due to higher demand. The additional assumption that $\rho \leq \gamma$ has no such obvious theoretical support. I further note that under the assumptions of Section 3.2, $Cov(u_i, \nu_i) = Cov(u_i, \epsilon_i) = Cov(\nu_i, \epsilon_i) = 0$, which will obviously satisfy this condition, recalling that $\rho = \mathbb{E}[(u_i + \nu_i)(\nu_i \gamma + \epsilon_i)]/\mathbb{E}[(u_i + \nu_i)^2]$. Under these conditions the moment inequalities appear to provide consistent bounds for σ_η, but not necessarily σ. As these parameters serve the same function, this has no effect on the predictive capacity of any resulting estimates of perceived returns.

I begin by presenting a lemma that will be useful in the subsequent proofs. It also serves as the main point of departure from the proofs provided by DM.

Lemma 1 *If equations* (2), (3), (20), *and* (21) *hold and* $\frac{\rho}{\gamma-\rho} \geq 0$, *then*

$$\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|S_i=0,Z_i] \geq 0 \geq \mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|S_i=1,Z_i]. \tag{C.1}$$

Proof: From the definition of S_i given in (2) and (3), substituting in the expression of perceived returns in (21) implies

$$\begin{aligned} &\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|S_i=0,Z_i] \\ =\ &\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|X_i\theta - Price_i\gamma + (u_i+\nu_i)\rho + \eta_i \leq 0, Z_i]. \end{aligned} \tag{C.2}$$

Substituting in the definition of $Price_i$ provided in (20) and rearranging the conditioning inequality implies

$$\begin{aligned} &\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|X_i\theta - Price_i\gamma + (u_i+\nu_i)\rho + \eta_i \leq 0, Z_i] \\ =&\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|X_i\theta - Z_i\delta\gamma - (u_i+\nu_i)(\gamma-\rho) + \eta_i \leq 0, Z_i] \\ =&\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|(u_i+\nu_i)(\gamma-\rho) \geq (X_i\theta - Z_i\delta\gamma + \eta_i), Z_i]. \end{aligned} \tag{C.3}$$

Given the property of expectations of truncated variables that $\mathbb{E}[X|X \geq Y] \geq \mathbb{E}[X]$, it follows that

$$\begin{aligned} &\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|(u_i+\nu_i)(\gamma-\rho) \geq (X_i\theta - Z_i\delta\gamma + \eta_i), Z_i] \\ \geq&\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|Z_i] \\ =&0, \end{aligned} \tag{C.4}$$

where the last equality follows from the definition of $(u_i+\nu_i)$ given in (20). The definition of S_i given in (2) and (3), substituting in the expression of perceived returns in (21), also implies

$$
\begin{aligned}
&\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|S_i=1,Z_i] \\
&=\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|X_i\theta-Price_i\gamma+(u_i+\nu_i)\rho+\eta_i\geq 0,Z_i].
\end{aligned}
\tag{C.5}
$$

Substituting in the definition of $Price_i$ provided in (20) and rearranging the conditioning inequality implies

$$
\begin{aligned}
&\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|X_i\theta-Price_i\gamma+(u_i+\nu_i)\rho+\eta_i\geq 0,Z_i] \\
=&\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|X_i\theta-Z_i\delta\gamma-(u_i+\nu_i)(\gamma-\rho)+\eta_i\geq 0,Z_i] \\
=&\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|(u_i+\nu_i)(\gamma-\rho)\leq(X_i\theta-Z_i\delta\gamma+\eta_i),Z_i].
\end{aligned}
\tag{C.6}
$$

Given the property of expectations of truncated variables that $\mathbb{E}[X|X\leq Y]\leq\mathbb{E}[X]$, it follows that

$$
\begin{aligned}
&\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|(u_i+\nu_i)(\gamma-\rho)\leq(X_i\theta-Z_i\delta\gamma+\eta_i),Z_i] \\
\leq&\mathbb{E}[(u_i+\nu_i)(\gamma-\rho)|Z_i] \\
=&0,
\end{aligned}
\tag{C.7}
$$

where the last equality follows from the definition of $(u_i+\nu_i)$ given in (20). Substituting (C.4) into (C.2) and (C.7) into (C.5) implies (C.1). ■

C.1 Proof of Revealed Preference Inequality Robustness to Endogeneity

Lemma 2 *Suppose equations* (2), (3), *and* (21) *hold. Then*

$$
\mathbb{E}\left[S_i\big(X_i\theta-Price_i\gamma+(u_i+\nu_i)\rho+\eta_i\big)\Big|Z_i\right]\geq 0. \tag{C.8}
$$

Proof: From equations (2), (3), and (21),

$$
S_i=\mathbb{1}\{X_i\theta-Price_i\gamma+(u_i+\nu_i)\rho+\eta_i\geq 0\}. \tag{C.9}
$$

This implies

$$S_i\big(X_i\theta - Price_i\gamma + (u_i + \nu_i)\rho + \eta_i\big) \geq 0. \tag{C.10}$$

This inequality holds for every individual i, therefore it will hold in expectation conditional on Z_i. ■

Lemma 3 *Equations* (2)*,* (3)*,* (20)*,* (21)*, and* (22) *imply that*

$$\mathbb{E}\Bigg[S_i\big(X_i\theta^* - Z_i\delta\gamma^*\big) \\ +(1-S_i)\left((u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}{1-\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}\right)\Bigg|Z_i\Bigg] \geq 0 \tag{C.11}$$

Proof: Equation (C.8) and the definition of $Price_i$ from equation (20) imply

$$\mathbb{E}[S_i(X_i\theta - Z_i\delta\gamma)|Z_i] - \mathbb{E}[S_i(u_i+\nu_i)(\gamma-\rho)|Z_i\big] + \mathbb{E}[S_i\eta_i|Z_i] \geq 0. \tag{C.12}$$

The assumption in (20) implies that $\mathbb{E}[u_i + \nu_i|Z_i] = 0$, so it follows that

$$\mathbb{E}[S_i(u_i+\nu_i)(\gamma-\rho) + (1-S_i)(u_i+\nu_i)(\gamma-\rho)|Z_i] = 0.$$

Equation (22) implies that $\mathbb{E}[\eta_i|X_i, Price_i, (u_i+\nu_i)] = 0$, which implies

$$\mathbb{E}[S_i\eta_i + (1-S_i)\eta_i|X_i, Price_i, (u_i+\nu_i)] = 0.$$

Assuming the distribution of Z_i conditional on $X_i, Price_i, (u_i + \nu_i)$ is degenerate and applying the law of iterated expectations, the preceding two equations allow us to rewrite equation (C.12) as

$$\mathbb{E}[S_i(X_i\theta - Z_i\delta\gamma)|Z_i] + \mathbb{E}[(1-S_i)(u_i+\nu_i)(\gamma-\rho)|Z_i] - \mathbb{E}\big[(1-S_i)\eta_i\big|Z_i\big] \geq 0. \tag{C.13}$$

Assuming the distribution of Z_i conditional on $X_i, Price_i, (u_i + \nu_i)$ is degenerate and applying the law of iterated expectations also implies

$$
\begin{aligned}
\mathbb{E}[(1-S_i)\eta_i|Z_i] &= \mathbb{E}\big[\mathbb{E}[(1-S_i)\eta_i|S_i, X_i, Price_i, (u_i+\nu_i)]\big|Z_i\big] \\
&= \mathbb{E}\big[\mathbb{E}[(1-S_i)|X_i, Price_i, (u_i+\nu_i)]\mathbb{E}[\eta_i|S_i, X_i, Price_i, (u_i+\nu_i)]\big|Z_i\big] \\
&= \mathbb{E}\big[P\big(S_i=1|X_i, Price_i, (u_i+\nu_i)\big)\times 0\times\mathbb{E}[\eta_i|S_i=1, X_i, Price_i, (u_i+\nu_i)] \\
&\quad + P\big(S_i=0|X_i, Price_i, (u_i+\nu_i)\big)\times 1\times\mathbb{E}[\eta_i|S_i=0, X_i, Price_i, (u_i+\nu_i)]\big|Z_i\big] \\
&= \mathbb{E}\big[P\big(S_i=0|X_i, Price_i, (u_i+\nu_i)\big)\mathbb{E}[\eta_i|S_i=0, X_i, Price_i, (u_i+\nu_i)]\big|Z_i\big] \\
&= \mathbb{E}\big[\mathbb{E}[(1-S_i)|X_i, Price_i, (u_i+\nu_i)]\mathbb{E}[\eta_i|S_i=0, X_i, Price_i, (u_i+\nu_i)]\big|Z_i\big] \\
&= \mathbb{E}\Big[\mathbb{E}\big[(1-S_i)\mathbb{E}[\eta_i|S_i=0, X_i, Price_i, (u_i+\nu_i)]\big|X_i, Price_i, (u_i+\nu_i)\big]\Big|Z_i\Big] \\
&= \mathbb{E}\big[(1-S_i)\mathbb{E}[\eta_i|S_i=0, X_i, Price_i, (u_i+\nu_i)]\big|Z_i\big].
\end{aligned}
$$

This allows us to rewrite equation (C.13) as

$$
\mathbb{E}\big[S_i(X_i\theta - Z_i\delta\gamma) + (1-S_i)\big((u_i+\nu_i)(\gamma-\rho) - \mathbb{E}[\eta_i|S_i=0, X_i, Price_i, (u_i+\nu_i)]\big)\big|Z_i\big] \geq 0. \tag{C.14}
$$

Using the definition of S_i from equation (2) and substituting in equations (3) and (21), it follows that

$$
\begin{aligned}
\mathbb{E}[\eta_i|S_i=0, X_i, Price_i, (u_i+\nu_i)] &= \mathbb{E}\big[\eta_i\big|\big(-\eta_i \geq X_i\theta - Price_i\gamma + (u_i+\nu_i)\rho\big), X_i, Price_i, (u_i+\nu_i)\big] \\
&= -\mathbb{E}\big[-\eta_i\big|\big(-\eta_i \geq X_i\theta - Price_i\gamma + (u_i+\nu_i)\rho\big), X_i, Price_i, (u_i+\nu_i)\big],
\end{aligned}
$$

which allows us to rewrite

$$
\mathbb{E}[\eta_i|S_i=0, X_i, Price_i, (u_i+\nu_i)] = -\sigma_\eta \frac{\phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}{1-\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)} \tag{C.15}
$$

using equation (22) and applying the symmetry of the normal distribution. Equation (C.11) follows by applying this equality to (C.14) and dividing each side of the resulting inequality by σ_η. ■

Lemma 4 *Given* $\frac{\rho}{\gamma-\rho} \geq 0$, *equations* (2), (3), (20), *and* (21) *imply*

$$\mathbb{E}\left[(1-S_i)\left((u_i+\nu_i)\gamma^* + \frac{\phi\big(X_i\theta^* - Price_i\gamma^*\big)}{1-\Phi\big(X_i\theta^* - Price_i\gamma^*\big)}\right)\Bigg| Z_i\right]$$
$$\geq \tag{C.16}$$
$$\mathbb{E}\left[(1-S_i)\left((u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}{1-\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}\right)\Bigg| Z_i\right]$$

Proof: Using the definition of $Price_i$ from equation (20), it follows that

$$\mathbb{E}\left[(1-S_i)\left((u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}{1-\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}\right)\Bigg| Z_i\right]$$
$$=$$
$$\mathbb{E}\left[(1-S_i)\left((u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\right)\Bigg| Z_i\right]. \tag{C.17}$$

Law of iterated expectations and $S_i \in \{0,1\}$ implies

$$\mathbb{E}\left[(1-S_i)\left((u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\right)\Bigg| Z_i\right]$$
$$=$$
$$\mathbb{E}\left[(1-S_i)\mathbb{E}\left[(u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\Bigg)\Bigg| S_i=0, Z_i\right]\Bigg| Z_i\right]. \tag{C.18}$$

Because

$$\frac{\partial \frac{\phi(-x)}{1-\Phi(-x)}}{\partial x} \in (-1,0),$$

it follows that the expression

$$(u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)} \tag{C.19}$$

is monotonically increasing in $(u_i+\nu_i)(\gamma^*-\rho^*)$. It follows then that adding a positive value to this value will increase the value of the function. From (C.1) and the condition

$\frac{\rho^*}{\gamma^*-\rho^*} \geq 0$, $\frac{\rho^*}{\gamma^*-\rho^*}\mathbb{E}[(u_i+\nu_i)(\gamma^*-\rho^*)|S_i=0,Z_i] \geq 0\ \forall i$, so it follows that

$$\begin{aligned}
&\mathbb{E}\Bigg[(u_i+\nu_i)(\gamma^*-\rho^*)+\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i] \\
&+\frac{\phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)-\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i]\big)}{1-\Phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)-\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i]\big)}\Bigg|S_i=0,Z_i\Bigg] \\
&= \\
&\mathbb{E}\Bigg[\Big((u_i+\nu_i)(\gamma^*-\rho^*)+\frac{\rho^*}{\gamma^*-\rho^*}\mathbb{E}[(u_i+\nu_i)(\gamma^*-\rho^*)|S_i=0,Z_i]\Big) \\
&+\frac{\phi\Big(X_i\theta^*-Z_i\delta\gamma^*-\big((u_i+\nu_i)(\gamma^*-\rho^*)+\frac{\rho^*}{\gamma^*-\rho^*}\mathbb{E}[(u_i+\nu_i)(\gamma^*-\rho^*)|S_i=0,Z_i]\big)\Big)}{1-\Phi\Big(X_i\theta^*-Z_i\delta\gamma^*-\big((u_i+\nu_i)(\gamma^*-\rho^*)+\frac{\rho^*}{\gamma^*-\rho^*}\mathbb{E}[(u_i+\nu_i)(\gamma^*-\rho^*)|S_i=0,Z_i]\big)\Big)}\Bigg|S_i=0,Z_i\Bigg] \\
&\geq \\
&\mathbb{E}\Bigg[(u_i+\nu_i)(\gamma^*-\rho^*)+\frac{\phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)\big)}{1-\Phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)\big)}\Bigg|S_i=0,Z_i\Bigg],
\end{aligned} \tag{C.20}$$

where the second line relates to the third by this addition, and the first relates to the second by algebraic simplifications. Finally, because the term

$$(u_i+\nu_i)(\gamma^*-\rho^*)+\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i]$$

and the term

$$\frac{\phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)-\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i]\big)}{1-\Phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)-\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i]\big)}$$

are globally convex in $-(u_i+\nu_i)$, the entire function is globally convex in $-(u_i+\nu_i)$. It follows that

$$\begin{aligned}
&\mathbb{E}\Bigg[(u_i+\nu_i)\gamma^*+\frac{\phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)\gamma^*\big)}{1-\Phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)\gamma^*\big)}\Bigg|S_i=0,Z_i\Bigg] \\
&\geq \\
&\mathbb{E}\Bigg[(u_i+\nu_i)(\gamma^*-\rho^*)+\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i] \\
&+\frac{\phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)-\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i]\big)}{1-\Phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)-\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i]\big)}\Bigg|S_i=0,Z_i\Bigg]
\end{aligned} \tag{C.21}$$

by Jensen's inequality. Combining this inequality with that in (C.20) yields the result

$$\mathbb{E}\bigg[(u_i+\nu_i)\gamma^* + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)\gamma^*\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)\gamma^*\big)}\bigg|S_i=0, Z_i\bigg]$$
$$\geq \tag{C.22}$$
$$\mathbb{E}\bigg[(u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\bigg|S_i=0, Z_i\bigg].$$

It follows immediately that

$$\mathbb{E}\bigg[(1-S_i)\mathbb{E}\bigg[(u_i+\nu_i)\gamma^* + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)\gamma^*\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)\gamma^*\big)}\bigg|S_i=0, Z_i\bigg]\bigg|Z_i\bigg]$$
$$\geq$$
$$\mathbb{E}\bigg[(1-S_i)\mathbb{E}\bigg[(u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\bigg|S_i=0, Z_i\bigg]\bigg|Z_i\bigg]. \tag{C.23}$$

Equation (C.16) follows from this by substituting in the definition of $Price_i$ from (20) and applying the law of iterated expectations. ■

Corollary 1 *Given (C.11), (C.16), and the definition of $Price_i$ given in* (20), *it follows that*

$$\mathbb{E}\bigg[S_i(X_i\theta^* - Price_i\gamma^*) + (1-S_i)\frac{\phi(X_i\theta^* - Price_i\gamma^*)}{1-\Phi(X_i\theta^* - Price_i\gamma^*)}\bigg|Z_i\bigg] \geq 0. \tag{C.24}$$

Proof: The result follows from equations (C.11), (C.16), substituting $\mathbb{E}[-S_i(u_i+\nu_i) = (1-S_i)(u_i+\nu_i)|Z_i]$, and substituting in the definition of $Price_i$ given in (20). ■

Lemma 5 *Suppose equations* (2), (3), *and* (21) *hold. Then*

$$\mathbb{E}\bigg[-(1-S_i)\big(X_i\theta - Price_i\gamma + (u_i+\nu_i)\rho + \eta_i\big)\bigg|Z_i\bigg] \geq 0. \tag{C.25}$$

Proof: From equations (2), (3), and (21),

$$S_i = \mathbb{1}\{X_i\theta - Price_i\gamma + (u_i+\nu_i)\rho + \eta_i \geq 0\}. \tag{C.26}$$

This implies

$$-(1 - S_i)\big(X_i\theta - Price_i\gamma + (u_i + \nu_i)\rho + \eta_i\big) \geq 0. \tag{C.27}$$

This inequality holds for every individual i, therefore it will hold in expectation conditional on Z_i. ■

Lemma 6 *Equations* (2), (3), (20), (21), *and* (22) *imply that*

$$\begin{aligned}\mathbb{E}\Bigg[& -(1 - S_i)\big(X_i\theta^* - Z_i\delta\gamma^*\big) \\ +S_i\Bigg(& -(u_i + \nu_i)(\gamma^* - \rho^*) + \frac{\phi\big(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*\big)}{\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*\big)}\Bigg)\Bigg|Z_i\Bigg] \geq 0\end{aligned} \tag{C.28}$$

Proof: Equation (C.25) and the definition of $Price_i$ from equation (20) imply

$$\mathbb{E}[-(1-S_i)(X_i\theta - Z_i\delta\gamma)|Z_i] + \mathbb{E}[(1-S_i)(u_i+\nu_i)(\gamma-\rho)|Z_i] - \mathbb{E}[(1-S_i)\eta_i|Z_i] \geq 0. \tag{C.29}$$

The assumption in (20) implies that $\mathbb{E}[u_i + \nu_i|Z_i] = 0$, so it follows that

$$\mathbb{E}[S_i(u_i + \nu_i)(\gamma - \rho) + (1 - S_i)(u_i + \nu_i)(\gamma - \rho)|Z_i] = 0.$$

Equation (22) implies that $\mathbb{E}[\eta_i|X_i, Price_i, (u_i + \nu_i)] = 0$, which implies

$$\mathbb{E}[S_i\eta_i + (1 - S_i)\eta_i|X_i, Price_i, (u_i + \nu_i)] = 0.$$

Assuming the distribution of Z_i conditional on $X_i, Price_i, (u_i + \nu_i)$ is degenerate and applying the law of iterated expectations, the preceding two equations allow us to rewrite equation (C.29) as

$$\mathbb{E}[-(1 - S_i)(X_i\theta - Z_i\delta\gamma)|Z_i] - \mathbb{E}[S_i(u_i + \nu_i)(\gamma - \rho)|Z_i] + \mathbb{E}\big[S_i\eta_i\big|Z_i\big] \geq 0. \tag{C.30}$$

Assuming the distribution of Z_i conditional on $X_i, Price_i, (u_i + \nu_i)$ is degenerate and applying the law of iterated expectations also implies

$$
\begin{aligned}
\mathbb{E}[S_i\eta_i|Z_i] &= \mathbb{E}\big[\mathbb{E}[S_i\eta_i|S_i, X_i, Price_i, (u_i+\nu_i)]\big|Z_i\big] \\
&= \mathbb{E}\big[\mathbb{E}[S_i|X_i, Price_i, (u_i+\nu_i)]\mathbb{E}[\eta_i|S_i, X_i, Price_i, (u_i+\nu_i)]\big|Z_i\big] \\
&= \mathbb{E}\big[P\big(S_i=1|X_i, Price_i, (u_i+\nu_i)\big)\times 1\times \mathbb{E}[\eta_i|S_i=1, X_i, Price_i, (u_i+\nu_i)] \\
&\quad + P\big(S_i=0|X_i, Price_i, (u_i+\nu_i)\big)\times 0\times \mathbb{E}[\eta_i|S_i=0, X_i, Price_i, (u_i+\nu_i)]\big|Z_i\big] \\
&= \mathbb{E}\big[P\big(S_i=1|X_i, Price_i, (u_i+\nu_i)\big)\mathbb{E}[\eta_i|S_i=1, X_i, Price_i, (u_i+\nu_i)]\big|Z_i\big] \\
&= \mathbb{E}\big[\mathbb{E}[S_i|X_i, Price_i, (u_i+\nu_i)]\mathbb{E}[\eta_i|S_i=1, X_i, Price_i, (u_i+\nu_i)]\big|Z_i\big] \\
&= \mathbb{E}\Big[\mathbb{E}\big[S_i\mathbb{E}[\eta_i|S_i=1, X_i, Price_i, (u_i+\nu_i)]\big|X_i, Price_i, (u_i+\nu_i)\big]\Big|Z_i\Big] \\
&= \mathbb{E}\big[S_i\mathbb{E}[\eta_i|S_i=1, X_i, Price_i, (u_i+\nu_i)]\big|Z_i\big].
\end{aligned}
$$

This allows us to rewrite equation (C.30) as

$$
\mathbb{E}\big[-(1-S_i)(X_i\theta-Z_i\delta\gamma)+S_i\big(-(u_i+\nu_i)(\gamma-\rho)+\mathbb{E}[\eta_i|S_i=1, X_i, Price_i, (u_i+\nu_i)]\big)\big|Z_i\big]\geq 0. \tag{C.31}
$$

Using the definition of S_i from equation (2) and substituting in equations (3) and (21), it follows that

$$
\begin{aligned}
\mathbb{E}[\eta_i|S_i=1, X_i, Price_i, (u_i+\nu_i)] &= \mathbb{E}\big[\eta_i\big|\big(-\eta_i\leq X_i\theta - Price_i\gamma+(u_i+\nu_i)\rho\big), X_i, Price_i, (u_i+\nu_i)\big] \\
&= -\mathbb{E}\big[-\eta_i\big|\big(-\eta_i\leq X_i\theta - Price_i\gamma+(u_i+\nu_i)\rho\big), X_i, Price_i, (u_i+\nu_i)\big],
\end{aligned}
$$

which allows us to rewrite

$$
\mathbb{E}[\eta_i|S_i=1, X_i, Price_i, (u_i+\nu_i)] = \sigma_\eta\frac{\phi\big(X_i\theta^* - Price_i\gamma^*+(u_i+\nu_i)\rho^*\big)}{\Phi\big(X_i\theta^* - Price_i\gamma^*+(u_i+\nu_i)\rho^*\big)} \tag{C.32}
$$

using equation (22) and applying the symmetry of the normal distribution. Equation (C.28) follows by applying this equality to (C.31) and dividing each side of the resulting inequality by σ_η. ■

Lemma 7 *Given* $\frac{\rho}{\gamma-\rho} \geq 0$*, equations* (2)*,* (3)*,* (20)*, and* (21) *imply*

$$\begin{gathered}
\mathbb{E}\left[S_i\left(-(u_i+\nu_i)\gamma^* + \frac{\phi\big(X_i\theta^* - Price_i\gamma^*\big)}{\Phi\big(X_i\theta^* - Price_i\gamma^*\big)}\right)\middle| Z_i\right] \\
\geq \\
\mathbb{E}\left[S_i\left(-(u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}{\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}\right)\middle| Z_i\right]
\end{gathered} \tag{C.33}$$

Proof: Using the definition of $Price_i$ from equation (20), it follows that

$$\begin{gathered}
\mathbb{E}\left[S_i\left(-(u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}{\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}\right)\middle| Z_i\right] \\
= \\
\mathbb{E}\left[S_i\left(-(u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\right)\middle| Z_i\right].
\end{gathered} \tag{C.34}$$

Law of iterated expectations and $S_i \in \{0,1\}$ implies

$$\begin{gathered}
\mathbb{E}\left[S_i\left(-(u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\right)\middle| Z_i\right] \\
= \\
\mathbb{E}\left[S_i\mathbb{E}\left[-(u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\right)\middle| S_i = 1, Z_i\right]\middle| Z_i\right].
\end{gathered} \tag{C.35}$$

Because

$$\frac{\partial \frac{\phi(-x)}{\Phi(-x)}}{\partial x} \in (0,1),$$

it follows that the expression

$$-(u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)} \tag{C.36}$$

is monotonically decreasing in $(u_i+\nu_i)(\gamma^*-\rho^*)$. It follows then that adding a negative value to this value will increase the value of the function. From (C.1) and the condition

$\frac{\rho^*}{\gamma^*-\rho^*} \geq 0$, $\frac{\rho^*}{\gamma^*-\rho^*}\mathbb{E}[(u_i+\nu_i)(\gamma^*-\rho^*)|S_i=1,Z_i] \leq 0 \; \forall i$, so it follows that

$$\begin{gathered}
\mathbb{E}\Bigg[-(u_i+\nu_i)(\gamma^*-\rho^*)-\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=1,Z_i] \\
+\frac{\phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)-\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=1,Z_i]\big)}{\Phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)-\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=1,Z_i]\big)}\Bigg|S_i=1,Z_i\Bigg] \\
= \\
\mathbb{E}\Bigg[-\big((u_i+\nu_i)(\gamma^*-\rho^*)+\frac{\rho^*}{\gamma^*-\rho^*}\mathbb{E}[(u_i+\nu_i)(\gamma^*-\rho^*)|S_i=1,Z_i]\big) \\
+\frac{\phi\Big(X_i\theta^*-Z_i\delta\gamma^*-\big((u_i+\nu_i)(\gamma^*-\rho^*)+\frac{\rho^*}{\gamma^*-\rho^*}\mathbb{E}[(u_i+\nu_i)(\gamma^*-\rho^*)|S_i=1,Z_i]\big)\Big)}{\Phi\Big(X_i\theta^*-Z_i\delta\gamma^*-\big((u_i+\nu_i)(\gamma^*-\rho^*)+\frac{\rho^*}{\gamma^*-\rho^*}\mathbb{E}[(u_i+\nu_i)(\gamma^*-\rho^*)|S_i=1,Z_i]\big)\Big)}\Bigg|S_i=1,Z_i\Bigg] \\
\geq \\
\mathbb{E}\Bigg[-(u_i+\nu_i)(\gamma^*-\rho^*)+\frac{\phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)\big)}{\Phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)\big)}\Bigg|S_i=1,Z_i\Bigg],
\end{gathered} \tag{C.37}$$

where the second line relates to the third by this addition, and the first relates to the second by algebraic simplifications. Finally, because the term

$$\frac{\phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)-\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=1,Z_i]\big)}{\Phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)-\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=1,Z_i]\big)}$$

is globally convex in $-(u_i+\nu_i)$, the function is globally convex in $-(u_i+\nu_i)$. It follows that

$$\begin{gathered}
\mathbb{E}\Bigg[-(u_i+\nu_i)\gamma^*+\frac{\phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)\gamma^*\big)}{\Phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)\gamma^*\big)}\Bigg|S_i=1,Z_i\Bigg] \\
\geq \\
\mathbb{E}\Bigg[-(u_i+\nu_i)(\gamma^*-\rho^*)-\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=1,Z_i] \\
+\frac{\phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)-\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=1,Z_i]\big)}{\Phi\big(X_i\theta^*-Z_i\delta\gamma^*-(u_i+\nu_i)(\gamma^*-\rho^*)-\mathbb{E}[(u_i+\nu_i)\rho^*|S_i=1,Z_i]\big)}\Bigg|S_i=1,Z_i\Bigg]
\end{gathered} \tag{C.38}$$

by Jensen's inequality. Combining this inequality with that in (C.37) yields the result

$$\mathbb{E}\bigg[-(u_i+\nu_i)\gamma^* + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)\gamma^*\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)\gamma^*\big)}\bigg|S_i=1,Z_i\bigg]$$
$$\geq \tag{C.39}$$
$$\mathbb{E}\bigg[-(u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\bigg|S_i=1,Z_i\bigg].$$

It follows immediately that

$$\mathbb{E}\Bigg[S_i\mathbb{E}\bigg[-(u_i+\nu_i)\gamma^* + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)\gamma^*\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)\gamma^*\big)}\bigg|S_i=1,Z_i\bigg]\Bigg|Z_i\Bigg]$$
$$\geq$$
$$\mathbb{E}\Bigg[S_i\mathbb{E}\bigg[-(u_i+\nu_i)(\gamma^*-\rho^*) + \frac{\phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\bigg|S_i=1,Z_i\bigg]\Bigg|Z_i\Bigg]. \tag{C.40}$$

Equation (C.33) follows from this by substituting the definition of $Price_i$ from (20) and applying the law of iterated expectations. ■

Corollary 2 *Given (C.28), (C.33), and the definition of $Price_i$ given in* (20), *it follows that*

$$\mathbb{E}\Bigg[-(1-S_i)(X_i\theta^* - Price_i\gamma^*) + S_i\frac{\phi(X_i\theta^* - Price_i\gamma^*)}{\Phi(X_i\theta^* - Price_i\gamma^*)}\Bigg|Z_i\Bigg] \geq 0. \tag{C.41}$$

Proof: The result follows from equations (C.28), (C.33), that $\mathbb{E}[S_i(u_i+\nu_i) = -(1-S_i)(u_i+\nu_i)|Z_i]$, and the definition of $Price_i$ given in (20). ■

Proof of Robustness of Revealed Preference Inequalities to Endogeneity: Combining equations (C.24) and (C.41) provides both inequalities defined in equation (12). ■

C.2 Argument for Odds-Based Inequality Robustness to Endogeneity

The following argument is constructed as a proof, where the components of the argument that do not meet the standards of a proof are discussed as they arise.

Lemma 8 *Equations* (2), (3), (21), (22), *and the assumption that the distribution of* Z_i *is degenerate conditional on* $(X_i, Price_i, (u_i + \nu_i))$ *imply that*

$$\mathbb{E}\left[S_i \frac{1 - \Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*\big)}{\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*\big)} - (1 - S_i)\middle| Z_i\right] \geq 0 \tag{C.42}$$

.

Proof: Equation (22) implies that

$$S_i - \mathbb{1}\{X_i\theta - Price_i\gamma + (u_i + \nu_i)\rho + \eta_i \geq 0\} \geq 0,$$

or, equivalently,

$$\begin{aligned} 1 - \mathbb{1}\{X_i\theta - Price_i\gamma + (u_i + \nu_i)\rho + \eta_i \geq 0\} - (1 - S_i) &\geq 0, \\ \mathbb{1}\{X_i\theta - Price_i\gamma + (u_i + \nu_i)\rho + \eta_i \leq 0\} - (1 - S_i) &\geq 0, \end{aligned}$$

for all i. Given that this inequality holds for all individuals, it will also hold in expectation, conditional on any set of variables, across individuals. It follows that

$$\mathbb{E}[\mathbb{1}\{X_i\theta - Price_i\gamma + (u_i + \nu_i)\rho + \eta_i \leq 0\} - (1 - S_i)|X_i, Price_i, (u_i + \nu_i)] \geq 0.$$

The distributional assumption in (22) implies

$$\mathbb{E}[1 - \Phi(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*) - (1 - S_i)|X_i, Price_i, (u_i + \nu_i)] \geq 0.$$

Dividing through by $\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)$ yields

$$\mathbb{E}\Bigg[\frac{1-\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)}{\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)} \\ - \frac{(1-S_i)}{\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)}\Bigg| X_i, Price_i, (u_i+\nu_i)\Bigg] \geq 0.$$

Adding and subtracting $1 - S_i$ gives

$$\mathbb{E}\Bigg[\frac{1-\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)}{\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)} \\ - \left(1 - 1 + \frac{1}{\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)}\right)(1-S_i)\Bigg| X_i, Price_i, (u_i+\nu_i)\Bigg] \geq 0,$$

which we can rearrange into

$$\mathbb{E}\Bigg[\frac{1-\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)}{\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)} \\ - \left(1 + \frac{1-\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)}{\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)}\right)(1-S_i)\Bigg| X_i, Price_i, (u_i+\nu_i)\Bigg] \geq 0,$$

which can then be rearranged into

$$\mathbb{E}\left[S_i\frac{1-\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)}{\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)} - (1-S_i)\middle| X_i, Price_i, (u_i+\nu_i)\right] \geq 0.$$

Equation (C.42) follows from the law of iterated expectations and the assumption that the distribution of Z_i conditional on $(X_i, Price_i, (u_i+\nu_i))$ is degenerate. ■

Lemma 9 *If equations* (2), (3), (20), *and* (21) *hold and* $\frac{\rho}{\gamma-\rho} \geq 0$, *then*

$$\begin{gathered}\mathbb{E}\left[S_i\frac{1-\Phi\big(X_i\theta^* - Price_i\gamma^*\big)}{\Phi\big(X_i\theta^* - Price_i\gamma^*\big)}\middle| Z_i\right] \\ \geq \\ \mathbb{E}\left[S_i\frac{1-\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}{\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}\middle| Z_i\right].\end{gathered} \tag{C.43}$$

Argument: Substituting the definition of $Price_i$ from equation (20), we have that

$$\mathbb{E}\left[S_i \frac{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\middle| Z_i\right]$$

$$=$$

$$\mathbb{E}\left[S_i \frac{1-\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}{\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}\middle| Z_i\right].$$

Because $S_i \in \{0,1\}$, it follows that

$$\mathbb{E}\left[S_i \frac{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\middle| Z_i\right] \geq 0$$

$$\Longleftrightarrow$$

$$\mathbb{E}\left[\frac{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\middle| S_i = 1, Z_i\right] \geq 0.$$

Proving the argument requires that

$$\mathbb{E}\left[\frac{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - (u_i+\nu_i)\rho^*\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - (u_i+\nu_i)\rho^*\big)}\middle| S_i = 1, Z_i\right]$$

$$\geq$$

$$\mathbb{E}\left[\frac{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\middle| S_i = 1, Z_i\right].$$

Given that

$$\frac{1-\Phi(x)}{\Phi(x)}$$

is globally convex in x, Jensen's inequality implies that

$$\mathbb{E}\left[\frac{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - (u_i+\nu_i)\rho^*\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - (u_i+\nu_i)\rho^*\big)}\middle| S_i = 1, Z_i\right]$$

$$\geq$$

$$\mathbb{E}\left[\frac{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - \mathbb{E}[(u_i+\nu_i)\rho^*|S_i=1,Z_i]\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - \mathbb{E}[(u_i+\nu_i)\rho^*|S_i=1,Z_i]\big)}\middle| S_i = 1, Z_i\right].$$

Meanwhile, (C.1) implies

$$\mathbb{E}\left[\frac{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - \mathbb{E}[(u_i+\nu_i)\rho^*|S_i=1,Z_i]\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - \mathbb{E}[(u_i+\nu_i)\rho^*|S_i=1,Z_i]\big)}\middle|S_i=1,Z_i\right]$$

$$\leq$$

$$\mathbb{E}\left[\frac{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\middle|S_i=1,Z_i\right].$$

Combining these inequalities yields

$$\mathbb{E}\left[\frac{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - (u_i+\nu_i)\rho^*\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - (u_i+\nu_i)\rho^*\big)}\middle|S_i=1,Z_i\right]$$

$$\geq$$

$$\mathbb{E}\left[\frac{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - \mathbb{E}[(u_i+\nu_i)\rho^*|S_i=1,Z_i]\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - \mathbb{E}[(u_i+\nu_i)\rho^*|S_i=1,Z_i]\big)}\middle|S_i=1,Z_i\right]$$

$$\leq$$

$$\mathbb{E}\left[\frac{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\middle|S_i=1,Z_i\right].$$

Thus, the argument holds if the first inequality dominates the second. There is good reason to believe that this will be the case. The first inequality arises from $Var(u_i+\nu_i|S_i=1,Z_i)$ (through Jensen's inequality), while the second arises from $\mathbb{E}[u_i+\nu_i|S_i=1,Z_i]$. Three points are salient here. First, given that $\mathbb{E}[u_i+\nu_i|Z_i]=0$, the value of $\mathbb{E}[u_i+\nu_i|S_i=1,Z_i]$ is a monotonic function of $Var(u_i+\nu_i|Z_i)$. Second, the function $(1-\Phi(x))/\Phi(x)$ has a very large second derivative for most of its support, such that the application of Jensen's inequality will have a large effect on the inequality. Thirdly, because ρ^* is constrained to be small relative to γ^*, $\mathbb{E}[u_i+\nu_i|S_i=1,Z_i]$ is likely to have a relatively small effect on the inequality.

Equation (C.43) follows from the preceding inequalities if the first inequality dominates the second by performing simple algebraic manipulations and applying the definition of $Price_i$ given in (20). ■

Lemma 10 *Equations* (2), (3), (21), (22), *and the assumption that the distribution of*

Z_i is degenerate conditional on $(X_i, Price_i, (u_i + \nu_i))$ imply that

$$\mathbb{E}\left[(1 - S_i)\frac{\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*\big)}{1 - \Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*\big)} - S_i\middle|Z_i\right] \geq 0 \tag{C.44}$$

.

Proof: Equation (22) implies that

$$\mathbb{1}\{X_i\theta - Price_i\gamma + (u_i + \nu_i)\rho + \eta_i \geq 0\} - S_i \geq 0,$$

Given that this inequality holds for all individuals, it will also hold in expectation, conditional on any set of variables, across individuals. It follows that

$$\mathbb{E}[\mathbb{1}\{X_i\theta - Price_i\gamma + (u_i + \nu_i)\rho + \eta_i \leq 0\} - S_i|X_i, Price_i, (u_i + \nu_i)] \geq 0.$$

The distributional assumption in (22) implies

$$\mathbb{E}[\Phi(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*) - S_i|X_i, Price_i, (u_i + \nu_i)] \geq 0.$$

Dividing through by $1 - \Phi(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*)$ yields

$$\mathbb{E}\Bigg[\frac{\Phi(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*)}{1 - \Phi(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*)} \\ - \frac{S_i}{1 - \Phi(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*)}\Bigg|X_i, Price_i, (u_i + \nu_i)\Bigg] \geq 0.$$

Adding and subtracting S_i gives

$$\mathbb{E}\Bigg[\frac{\Phi(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*)}{1 - \Phi(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*)} \\ - \left(1 - 1 + \frac{1}{1 - \Phi(X_i\theta^* - Price_i\gamma^* + (u_i + \nu_i)\rho^*)}\right)S_i\Bigg|X_i, Price_i, (u_i + \nu_i)\Bigg] \geq 0,$$

which we can rearrange into

$$\mathbb{E}\Bigg[\frac{\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)}{1-\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)} \\ - \left(1 + \frac{\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)}{1-\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)}\right)S_i\Bigg|X_i, Price_i, (u_i+\nu_i)\Bigg] \geq 0,$$

which is straightforward to rearrange into

$$\mathbb{E}\left[(1-S_i)\frac{\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)}{1-\Phi(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*)} - S_i\middle|X_i, Price_i, (u_i+\nu_i)\right] \geq 0.$$

Equation (C.44) follows from the law of iterated expectations and the assumption that the distribution of Z_i conditional on $(X_i, Price_i, (u_i+\nu_i))$ is degenerate. ■

Lemma 11 *If equations* (2), (3), (20), *and* (21) *hold and* $\frac{\rho}{\gamma-\rho} \geq 0$, *then*

$$\begin{gathered}\mathbb{E}\left[(1-S_i)\frac{\Phi\big(X_i\theta^* - Price_i\gamma^*\big)}{1-\Phi\big(X_i\theta^* - Price_i\gamma^*\big)}\middle|Z_i\right] \\ \geq \\ \mathbb{E}\left[(1-S_i)\frac{\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}{1-\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}\middle|Z_i\right].\end{gathered} \tag{C.45}$$

Argument: Substituting the definition of $Price_i$ from equation (20), we have that

$$\begin{gathered}\mathbb{E}\left[(1-S_i)\frac{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\middle|Z_i\right] \\ = \\ \mathbb{E}\left[(1-S_i)\frac{\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}{1-\Phi\big(X_i\theta^* - Price_i\gamma^* + (u_i+\nu_i)\rho^*\big)}\middle|Z_i\right].\end{gathered}$$

Because $S_i \in \{0,1\}$, it follows that

$$\mathbb{E}\left[(1-S_i)\frac{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\middle| Z_i\right] \geq 0$$

$$\iff$$

$$\mathbb{E}\left[\frac{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\middle| S_i = 0, Z_i\right] \geq 0.$$

Proving the argument requires that

$$\mathbb{E}\left[\frac{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - (u_i+\nu_i)\rho^*\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - (u_i+\nu_i)\rho^*\big)}\middle| S_i = 0, Z_i\right]$$

$$\geq$$

$$\mathbb{E}\left[\frac{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\middle| S_i = 0, Z_i\right].$$

Given that

$$\frac{\Phi(x)}{1-\Phi(x)}$$

is globally convex in x, Jensen's inequality implies that

$$\mathbb{E}\left[\frac{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - (u_i+\nu_i)\rho^*\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - (u_i+\nu_i)\rho^*\big)}\middle| S_i = 0, Z_i\right]$$

$$\geq$$

$$\mathbb{E}\left[\frac{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - \mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i]\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - \mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i]\big)}\middle| S_i = 0, Z_i\right].$$

Meanwhile, (C.1) implies

$$\mathbb{E}\left[\frac{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - \mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i]\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - \mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i]\big)}\middle| S_i = 0, Z_i\right]$$

$$\leq$$

$$\mathbb{E}\left[\frac{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\middle| S_i = 0, Z_i\right].$$

Combining these inequalities yields

$$\mathbb{E}\left[\frac{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - (u_i+\nu_i)\rho^*\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - (u_i+\nu_i)\rho^*\big)}\middle| S_i = 0, Z_i\right]$$

$$\geq$$

$$\mathbb{E}\left[\frac{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - \mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i]\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*) - \mathbb{E}[(u_i+\nu_i)\rho^*|S_i=0,Z_i]\big)}\middle| S_i = 0, Z_i\right]$$

$$\leq$$

$$\mathbb{E}\left[\frac{\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}{1-\Phi\big(X_i\theta^* - Z_i\delta\gamma^* - (u_i+\nu_i)(\gamma^*-\rho^*)\big)}\middle| S_i = 0, Z_i\right].$$

Thus, the argument holds if the first inequality dominates the second. There is good reason to believe that this will be the case. The first inequality arises from $Var(u_i+\nu_i|S_i = 0, Z_i)$ (through Jensen's inequality), while the second arises from $\mathbb{E}[u_i + \nu_i|S_i = 0, Z_i]$. Three points are salient here. First, given that $\mathbb{E}[u_i+\nu_i|Z_i] = 0$, the value of $\mathbb{E}[u_i+\nu_i|S_i = 0, Z_i]$ is a monotonic function of $Var(u_i+\nu_i|Z_i)$. Second, the function $\Phi(x)/(1-\Phi(x))$ has a very large second derivative for most of its support, such that the application of Jensen's inequality will have a large effect on the inequality. Thirdly, because ρ^* is constrained to be small relative to γ^*, $\mathbb{E}[u_i + \nu_i|S_i = 0, Z_i]$ is likely to have a relatively small effect on the inequality.

Equation (C.45) follows from the immediately preceding inequalities if the first inequality dominates the second by performing simple algebraic manipulations and applying the definition of $Price_i$ given in (20). ■

Argument for Odds Based Inequality Robustness to Endogeneity: Substituting equation (C.43) into (C.42) and equation (C.45) into (C.44) provides the inequalities defined in equation (15).