Selection on Unobservables in Discrete Choice Models

Alden Cheng

April 14, 2023

Abstract

Selection on unobservables is an important concern for causal inference in observational studies, and accordingly, previous papers have developed methods for sensitivity analysis for OLS, binary choice models, instrumental variables, and movers designs. In this paper, I develop methods for sensitivity analysis for a setting that has not been previously studied — discrete choice models. In particular, I derive bounds for the omitted variables bias under an assumption about how much the consumer values the omitted variable(s) relative to the included control variables, and about the relationship between the omitted variable and the variable of interest. I provide theoretical results for my bounding procedure, and demonstrate its performance in simulations. Finally, I show in several empirical applications that my procedure produces economically meaningful bounds.

1 Introduction

Concerns over selection on unobservables are common in most observational studies in economics. While the credibility revolution has led to the development of many quasi-experimental methods aimed at overcoming these challenges (Angrist and Pischke 2010), there remains settings with important economic questions without quasi-experimental variation in the variable of interest, so researchers have little choice other than to rely on observational methods. Moreover,

Accordingly, a number of papers have developed methods for sensitivity analysis for OLS (Imbens 2003; Altonji, Elder, and Taber 2005; Oster 2019; Cinelli and Hazlett 2020; Diegert, Masten and Poirier 2022; Masten and Poirier 2023), instrumental variables (Conley, Hansen, and Rossi 2012; Nevo and Rosen 2012), movers designs (Finkelstein, Gentzkow, and Williams 2021), binary choice models (Rosenbaum and Rubin 1983; Ding and VanderWeele 2016; VanderWeele and Ding 2017), and structural methods (Andrews, Gentzkow, and Shapiro 2020). With the exception of Andrews,

I am very grateful to Alberto Abadie and Whitney Newey for their valuable comments.

Gentzkow, and Shapiro (who focus on a somewhat different problem), these papers typically bound the omitted variables bias based on two assumptions: (i) the relationship between the omitted variable with the outcome, and (ii) the relationship between the omitted variable and the variable of interest.

In this paper, I develop methods sensitivity analysis for a commonly used model that is more "structural" than those considered in the aforementioned papers that is nonetheless commonly used in economic analysis: discrete choice models (Nevo and Whinston 2010). Specifically, I consider a random utility model, where consumers choose between different products. The utility for consumer i if she chooses product j is given by:

$$u_{ij} = \beta x_{ij} + \alpha w_{ij} + \sum_{k=1}^{K} \delta_k z_{ij}^k + \epsilon_{ij},\ \epsilon \perp (x_{ij}, w_{ij}, z_{ij}')',$$

and she chooses the product that gives her the highest utility, $j^* = argmin_j\{u_{ij}\}$.

Similar to previous papers on sensitivity analysis, we are interested on the "effect" β of a variable of interest x_{ij}, but since a scale normalization is required for utility, it makes more sense to consider the marginal rate of subsitution (MRS) between x_{ij} and another variable w_{ij}. For example, if w_{ij} is price, then β/α represents the consumers' willingness-to-pay for the characteristic x_{ij}. We are able to estimate β and α consistently if we include all of the controls z_{ij}, but the worry is that some of the z_{ij}^k's are not observed, in which case we may have omitted variables bias (OVB).

There are several notable differences between this random utility model that the more "reduced form" models considered in previous studies on sensitivity analysis. First, the "left hand side variable" in the equation of interest — utility — is unobserved, unlike the outcome variable in OLS and IV.[1] This is problematic because sensitivity analysis in previous papers typically require an assumption related to the partial R^2 from a hypothetical regression of the outcome variable on the omitted variable, and such an assumption makes less sense when the outcome variable is latent (unobserved). To deal with this, I replace this hypothetical R^2 assumption with an assumption about how much more the consumer values the omitted variable, relative to the included covariates. An appealing feature about this assumption is that it has a clear economic interpretation, whereas the R^2 assumption only describes a statistical relationship.

Second, there is not always a convenient exact formula for the OVB (unlike OLS). So, although discussions regarding the OVB in discrete choice models often invoke intuitions from the OLS case, we

[1]This is similar to the latent index model for binary choice. However, there is only a single equation per observation in binary choice, whereas in discrete choice settings where a consumer is choosing between J products, there are either $J-1$ or J equations per observation (consumer), depending on whether or not there is an outside option.

need to formalize these intuitions when deriving precise results on the possible magnitudes of OVB. I do so by first considering a very specific parametric setting where the observed and unobserved covariates as well as the error terms come from a Gaussian distribtion, which allows me to derive exact formulas for the OVB. I then show that the OVB formula holds asymptotically under more general distributions for the covariates, and under arbitrary error distributions.

Third, as mentioned earlier, we are interested in the ratio of two coefficients, unlike in OLS and IV settings (where there is no scaling variable). We may potentially be concerned that the scaling variable w_{ij} may likewise suffer from omitted variables bias, so in my sensitivity analysis I also derive bounds that account for possible OVB for the scaling variable.

My sensitivity analysis derives bounds on the omitted variables bias for the coefficient of interest β/α based on two main assumptions. First, the researcher needs to make an assumption about how many times more the consumer values the omitted characteristic relative to the included control characteristics (M). As mentioned above, this is analogous to the assumption in sensitivity analysis in OLS of the partial R^2 from a hypothetical regression of the outcome on the omitted variable.

Second, the researcher can make an assumption on the maximum R^2 from a regression of the variable of interest x_{ij} on all of the controls z_{ij} (observe and unobserved) whihc we will call $R^2_{max,x}$, and similarly for w_{ij} if we allow for possible endogeneity in the scaling variable. This second assumption is "optional", in that we can still obtain bounds without it (by essentially assuming the "worst case scenario" of an R^2 infinitesimally close to one), although we can obtain a tighter bound if the researcher is willing to make the R^2 assumption. Variants of this assumption are typically also made in sensitivity analysis for OLS — for example, this is closely related to what Oster (2019) calls the "proportional selection relationship".

My sensitivity analysis can be used by researchers in at least two ways. First, the researcher may have strong priors about what M and $R^2_{x,max}$ are, in which case she can compute the identified set (i.e., bounds) for β/α under these assumed assumed values of M and $R^2_{x,max}$. Alternatively, the researcher can consider a "test" of whether β/α is equal to a certain value τ^*, e.g., zero, or some other value predicted by ecconomic theory (which we can think of as the null hypothesis). We can then derive the set of values for (M, R^2_{max}) for which OVB can explain the discrepancy between the estimate $\hat{\beta}/\hat{\alpha}$ and the value given by the null hypothesis τ^*. Based on contextual knowledge, the researcher can then determine whether these values of M and R^2_{max} are plausible.

This paper contributes to a large literature on sensitivity analysis, both in economics and outside of it. However, most of the literature in economics has focused on OLS (Imbens 2003; Altonji, Elder, and Taber 2005; Oster 2019; Cinelli and Hazlett 2020; Diegert, Masten and Poirier 2022; Masten and Poirier 2023), and instrumental variables settings (Conley, Hansen, and Rossi 2012; Nevo and Rosen 2012), so this paper is the first (to the best of my knowledge) to consider sensitivity analysis in discrete choice settings.

Outside of economics, Ding and VanderWeele (2016) and VanderWeele and Ding (2017) also consider sensitivity analysis for choice models. However, their assumptions are formulated in terms of relative risks, which may be natural in certain settings (e.g., epedemiology), but are typically harder to interpret in economics. Moreover, their bounding procedure is most readily applied to cases with binary outcomes and becomes unwieldy if one tries to extend it to many unordered categorical outcomes, which is the setting for discrete choice models. By contrast, the assumptions for my bounding procedure do not depend on the number of categories (products).

This paper proceeds as follows. In section 2, I consider a simplied version of the model to provide intuition, before providing more general theoretical results in section 3. In section 4, I present simulation evidence on the performance of my bounding procedure, and in section 5, I provide empirical applications which show that my procedure provides economically meaningful bounds. Section 6 concludes.

2 Simplified Model with Exogenous Scaling Variable

In this section, I consider a simplified model where we assume that the scaling variable is exogenous, in order to build intuition. The general version of the model is presented in section 3.

2.1 Setup

There are N consumers (indexed by i) are choosing between $J \geq 2$ products (indexed by j). We assume that there is no outside option, but the results all carry over to the case with an outside option, with slight notational changes. Suppose that consumer i's indirect utility from choosing product j is given by:

$$u_{ij} = \beta x_{ij} + \alpha w_{ij} + \sum_{k=1}^{K} \delta_k z_{ij}^k + \epsilon_{ij}, \tag{1}$$

$$\alpha \neq 0,\ \mathbb{E}[\epsilon_{ij}] = 0,\ Var(x_{ij}) = Var(w_{ij}) = Var(z_{ij}^k) = Var(\epsilon_{ij}) = 1\ \forall k,$$

$$\epsilon_{ij} \perp (x_{ij}, w_{ij}, z_{ij}^1, ..., z_{ij}^K)'.$$

The researcher wants to learn consumers' preferences for the variable x_{ij}, which could be a product characteristic, for example.[2] The magnitude of the coefficient on x_{ij}, β alone is difficult to interpret, since the scale normalization for the utility is arbitary. Hence, we assume there is a scaling variable w_{ij}, so the quantity of interest to the researcher is β/α, which tells us how much the consumer values x_{ij} relative to w_{ij} (i.e., the MRS between x_{ij} and w_{ij}). For example, if w_{ij} is the price of j, then β/α represents consumers' willing-to-pay for x_{ij}.

There are additional covariates $z_{ij}^1, ...z_{ij}^K$, that are not of economic interest, but may need to be controlled for in order for the estimate of β/α to be consistent, i.e., to avoid OVB; however, the researcher observes a subset of these controls $z_{ij}^1, ..., z_{ij}^L$, $L < K$. Without loss of generality, we assume that there is a single omitted variable, i.e., $K = L + 1$, because otherwise, we can always consider the "composite" omitted variable given by $\sum_{k=L+1}^{K} \delta_k z_{ij}^k$. Also without loss of generality, assume that $\delta_k \geq 0$ for all k (otherwise, we can multiply z_{ij}^k by -1), and that $\delta_l > 0$ for some $l \in \{1, ..., L\}$.

Let us denote the estimate of β/α obtained from controlling for none of the z_{ij}^l's, and only for $z_{ij}^1, ..., z_{ij}^L$ by $\check{\beta}/\check{\alpha}$ and $\hat{\beta}/\hat{\alpha}$ respectively. In general, both of these estimates may suffer from OVB, so in this paper I derive bounds for the true parameter β/α based only on parameters identified in the data. To derive these bounds, I need to make several assumptions, which I state and discuss below.

Assumption D (Preferences for Omitted Variables Relative to Included Controls). Assume that consumers value the omitted variable at most M times more than the included controls:

$$\delta_K \leq M \sum_{l=1}^{L} \delta_l. \tag{2}$$

Assumption R0 (R^2 from a regression of x_{ij} on all the z_{ij}^k's). Assume that the R^2 from a regression of x_{ij} on $z_{ij}^1, ..., .z_{ij}^K$ is at most $R^2_{x,max} \leq 1$.

Remark 1. Assumption D is the most important assumption for deriving bounds for the OVB, and the value of M should be chosen based on contextual knowledge (noting that all the z_{ij}^k are standardized). Alternatively, suppose that the researcher wants to test whether the discrepancy between the estimate $\hat{\beta}/\hat{\alpha}$ and some target value τ^* (which need not be zero) can be explained by OVB. Then, we can compute the minimum value of M required so that that OVB can completely explain this discrepancy.

Remark 2. Similar to M, the value of $R^2_{x,max}$ should be specified based on contextual knowledge. However, we can still derive bounds without the researcher making an assumption on $R^2_{x,max}$, since

[2] This variable must vary across products, since a variable that varies only across consumers affects the consumer's utility for all products in their choice set equally, and is thus not identified.

this corresponds to the most conservative assumption would be to assume that $R^2_{x,max} = 1$, although tighter bounds can be obtained if one assumes a smaller value of $R^2_{x,max}$.

Remark 3. Instead of assuming a value for $R^2_{x,max}$, we can equivalently assume that the R^2 from a regression including all controls (denoted $R^2_{x,all}$) explains is at most $M_{x,R}$ times the R^2 from a regression of x_{ij} on $z^1_{ij}, ..., z^L_{ij}$ (denoted $R^2_{x,observed}$), whether this latter R^2 is identified in the data. This choice of $M_{x,R}$ is related to $R^2_{x,max}$ by the inequality: $R^2_{x,max} = M_R R^2_{x,observed}$. or $M_R = R^2_{x,max}/R^2_{x,observed}$.

This is closely related to proportional selection relationship described in Oster (2019). Assume without loss that there is only one observed control z^1_{ij} (otherwise, we can define $Z^1_{ij} \equiv \sum_{l=1}^{L} \delta_l z^l_{ij}$ and normalize accordingly). Then, there exists a value κ (denoted δ in Oster 2019), where the following relationship holds:

$$\kappa \cdot \frac{Cov(x_{ij}, z^1_{ij})}{Var(z^1_{ij})} = \frac{Cov(x_{ij}, z^2_{ij})}{Var(z^2_{ij})},$$

which simplifies to:

$$\kappa \cdot Corr(x_{ij}, z^1_{ij}) \equiv \kappa \rho^2_{x,z^1} = Corr(x_{ij}, z^2_{ij}) = \rho^2_{x,z^2}.$$

The R^2 values from regressions of x_{ij} on z^1_{ij}, and both z^1_{ij} and z^2_{ij} are given by ρ^2_{x,z^1} and $\rho^2_{x,z^1} + \rho^2_{x,z^2} = (1+\kappa)\rho^2_{x^1}$ respectively. Hence, the equation $R^2_{x,max} = M_{x,R} R^2_{x,observed}$ is equivalent to $(1+\kappa)\rho^2_{x^1} = M_{x,R}\rho^2_{x^1}$, or $M_{x,R} = 1+\kappa$. Oster (2019) argues that in many empirical applications, a value of $\kappa = 1$ is reasonable. This corresponds in our framework to a choice of $M_{x,R} = 2$ or $R^2_{x,max} = \min\{2R^2_{x,observed}, 1\}$.

Remark 4. Since Assumption R0 restricts $R^2 < R^2_{x,max} \leq 1$, this implies that x_{ij} is not perfectly predicted by a linear combination of $z^1_{ij}, ..., z^K_{ij}$. This is necessary because otherwise, it will be impossible to distinguish between effects of the z^k_{ij}'s and the effect of x_{ij}.

In this section, we consider a simplified setting where the scaling variable is exogenous, given by Assumption E below. This mirrors the OLS case more closely, since we are only trying to bound the OVB for one variable. The general case where the scaling variable is not necessarily exogenous is considered in the next section.

Assumption E (Exogenous Scaling Variable). Assume that $w_{ij} \perp (x_{ij}, z^1_{ij}, ..., z^K_{ij})'$.

2.2 Gaussian Regressors and Errors

In this section, we first make a strong assumption about the distribution of the covariates and error terms, before showing that the results apply more generally in subsequent analyses.

Assumption N0 (Gaussian Distribution and i.i.d. Errors). Assume that the covariates and structural error terms are jointly normally distributed:

$$(x_{ij}, w_{ij}, z^1_{ij}, ..., z^K_{ij})' \sim^{i.i.d.} N(0, \Sigma),$$

where Σ is a positive-definite correlation matrix. In addition, assume that $Cov(z^k_{ij}, z^{k'}_{ij}) = 0$ for all $k' \neq k$.

Remark 5. Diegert, Masten, and Poirier (2022) note the "exogenous controls" assumption, i.e., that $z^{L+1}_{ij} \perp (z^1_{ij}, ..., z^L_{ij})'$ is unrealistic in many empirical settings. Nonetheless, the part in assumption N0 stating that the z^k_{ij}'s are independent is without loss of generality. This is because the OVB depends only on the space spanned by the z^k_{ij}'s, and thus, if the z^k_{ij}'s were not initially independent, we can always orthogonalize them using Gram-Schimdt or by partialling out. Independence of the z^k_{ij}'s is only assumed here to simplify the formulae.

Next, I introduce some notation for population estimates from different multinomial probit specifications. Denote estimates from the specification without controls using "checks" (e.g., $\check{\beta}/\check{\alpha}$) and estimates from the specification including the observed controls $z^1_{ij}, ..., z^L_{ij}$ using "hats" (e.g., $\hat{\beta}/\hat{\alpha}$). Also, denote correlations between two random variables U and V by $\rho_{U,V}$.

Lemma 1. *Under Assumptions D, N0, R0, and E, we have:*

$$\frac{\beta}{\alpha} \in I \equiv \left(\frac{\hat{\beta}}{\hat{\alpha}} - \left| \left(M \sum_{l=1}^{L} \frac{\hat{\delta}_l}{\hat{\alpha}} \right) \left(\sqrt{R^2_{x,max} - \sum_{l=1}^{L} \hat{\rho}^2_{x,z^l}} \right) \right|, \frac{\hat{\beta}}{\hat{\alpha}} + \left| \left(M \sum_{l=1}^{L} \frac{\hat{\delta}_l}{\hat{\alpha}} \right) \left(\sqrt{R^2_{x,max} - \sum_{l=1}^{L} \hat{\rho}^2_{x,z^l}} \right) \right| \right). \tag{3}$$

When $L = 1$ and $\hat{\rho}_{x,z^1} \neq 0$, these bounds can also be written as:

$$\frac{\beta}{\alpha} \in I = \left(\frac{\hat{\beta}}{\hat{\alpha}} - \left| \frac{M}{\hat{\rho}_{x,z^1}} \left(\frac{\hat{\beta}}{\hat{\alpha}} - \frac{\check{\beta}}{\check{\alpha}} \right) \left(\sqrt{R^2_{x,max} - \hat{\rho}^2_{x,z^1}} \right) \right|, \frac{\hat{\beta}}{\hat{\alpha}} + \left| \frac{M}{\hat{\rho}_{x,z^1}} \left(\frac{\hat{\beta}}{\hat{\alpha}} - \frac{\check{\beta}}{\check{\alpha}} \right) \left(\sqrt{R^2_{x,max} - \hat{\rho}^2_{x,z^1}} \right) \right| \right). \tag{4}$$

Moreover, these bounds are sharp, in the sense that for any τ in I, there exists a data-generating process satisfying Assumptions D, N0, R0, and E that yields the estimates $(\check{\alpha}, \check{\beta}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}_{x,z^1}, ..., \hat{\gamma}_{x,z^L})'$, and such that $\beta/\alpha = \tau$.

The full proof of Lemma 1 is given in the Appendix, but the key is that we have an exact formula for the omitted variables bias in the probit model with Gaussian covariates. Specifically, suppose we estimate utility without any controls:

$$v_{ij} = \breve{\beta} x_{ij} + \breve{\alpha} w_{ij} + e_{ij}. \tag{5}$$

Assumption N0 implies that for each k:

$$z_{ij}^k = \rho_{x,z^k} x_{ij} + \nu_{ij}^k,\ \mathbb{E}[\nu_{ij}^k | x_{ij}] = 0,\ Var(\nu_{ij}^k | x_{ij}) = \sigma_{\nu^k}^2. \tag{6}$$

and that $(\nu_{ij}^1, ..., \nu_{ij}^K)'$ is jointly normally distributed with a diagonal covariance matrix. So, substituting equations (6) and (1) into equation (5), we obtain:

$$v_{ij} = \underbrace{\left(\beta + \sum_{k=1}^{K} \delta_k \rho_{x,z^k}\right)}_{=\breve{\beta}} x_{ij} + \alpha w_{ij} + \underbrace{\left(\epsilon_{ij} + \sum_{k=1}^{K} \rho_{x,z^k} \nu_{ij}^k\right)}_{=e_{ij}}.$$

The error term e_{ij} has a Gaussian distribution as a consequence of Assumption N0, and thus, equation (5) is still a multinomial probit specification. From this we deduce that the population estimates of the coefficients on x_{ij} and w_{ij} are given by:

$$\breve{\beta} = \frac{\beta + \sum_{k=1}^{K} \delta_k \rho_{x,z^k}}{\sqrt{1 + \sum_{k=1}^{K} \delta_k^2 \sigma_{\nu^k}^2}},\ \breve{\alpha} = \frac{\alpha}{\sqrt{1 + \sum_{k=1}^{K} \delta_k^2 \sigma_{\nu^k}^2}}. \tag{7}$$

Taking the ratio of $\breve{\beta}$ and $\breve{\alpha}$, the attenuation term cancels out, and the omitted variables bias is given by $\sum_{k=1}^{K} \delta_k \rho_{x,z^k} / \alpha$. Similarly, one can show that when including $z_{ij}^1, ..., z_{ij}^L$ in the utility equation, the omitted variables bias is $\sum_{k=L+1}^{K} \delta_k \rho_{x,z^k} / \alpha = \delta_{L+1} \rho_{x,z^{L+1}} / \alpha$. The remainder of the proof derives straightforward bounds for δ_{L+1} and $\rho_{x,z^{L+1}}$ based on the assumptions made in the Lemma.

Remark 6. Other than computing bounds, another way to use the results above is for "hypothesis testing". Specifically, we would like to test whether the discrepancy between the estimates $\hat{\beta}/\hat{\alpha}$ and τ^* can be explained by omitted variables bias, so we form the following null hypothesis:

$$H_0 : \beta/\alpha = \tau^*.$$

We can then compute that the minimal values of (M, R^2_{max}) that are consistent with the null hypothesis:

$$|\hat{\beta}/\hat{\alpha} - \tau^*| = |\left(M\sum_{l=1}^{L}\hat{\delta}_l/\hat{\alpha}\right)\left(\sqrt{R^2_{max} - \sum_{l=1}^{L}\hat{\rho}^2_{x,z^l}}\right)|$$

$$\implies M = |\hat{\beta}/\hat{\alpha} - \tau^*| \cdot \left(\sum_{l=1}^{L}\hat{\delta}_l/\hat{\alpha}\right)^{-1}\left(\sqrt{R^2_{max} - \sum_{l=1}^{L}\hat{\rho}^2_{x,z^l}}\right),$$

and the researcher can gauge whether these values of (M, R^2_{max}) are plausible.

Remark 7. The estimates in the proof are asymptotic, abstracting from statistical uncertainty, but in practice, we may want to account for this uncertainty. Suppose the researcher wants a coverage rate of at least $1-\gamma$. Suppose that $sgn(\hat{\alpha} - q \cdot se(\hat{\alpha})) = sgn(\hat{\alpha} + q \cdot se(\hat{\alpha})) \neq 0$, otherwise, the bounds would be $(-\infty, \infty)$ since the coefficient in the denominator, α can approach zero from either positive or negative direction. Assume also that $\hat{\alpha} - q_{1-\gamma/2} \cdot se(\hat{\alpha}) > 0$ to simplify notation (analogous formulae can be obtained for the case $\hat{\alpha} + q_{1-\gamma/2} \cdot se(\hat{\alpha}) < 0$. Then, we can write the bounds with coverage rate at least $1-\gamma$ as:

$$\frac{\hat{\beta} - q_{1-\gamma/2} \cdot se(\hat{\beta})}{\hat{\alpha} + q_{1-\gamma/2} \cdot sgn\left(\hat{\beta} - 1.96 se(\hat{\beta})\right) se(\hat{\alpha})} - \left|\left(M\sum_{l=1}^{L}\frac{|\hat{\delta}_l| + q_{1-\gamma/2} \cdot se(\hat{\delta}_l)}{\hat{\alpha} - q_{1-\gamma/2} \cdot se(\hat{\alpha})}\right)\left(\sqrt{R^2_{max} - \sum_{l=1}^{L}\left(\hat{\rho}_{x,z^l} + q_{1-\gamma/2} \cdot se(\hat{\rho}_{x,z^l})\right)^2}\right)\right|$$

for the lower bound, and:

$$\frac{\hat{\beta} + q_{1-\gamma/2} \cdot se(\hat{\beta})}{\hat{\alpha} - q_{1-\gamma/2} \cdot sgn\left(\hat{\beta} - q_{1-\gamma/2} \cdot se(\hat{\beta})\right) se(\hat{\alpha})} + \left|\left(M\sum_{l=1}^{L}\frac{|\hat{\delta}_l| + q_{1-\gamma/2} \cdot se(\hat{\delta}_l)}{\hat{\alpha} - q_{1-\gamma/2} \cdot se(\hat{\alpha})}\right)\left(\sqrt{R^2_{max} - \sum_{l=1}^{L}\left(\hat{\rho}_{x,z^l} + q_{1-\gamma/2} \cdot se(\hat{\rho}_{x,z^l})\right)^2}\right)\right|$$

for the upper bound, where $q_{1-\gamma/2}$ is the $(1-\gamma/2)$-th quantile of a standard Gaussian distribution.[3]

Note also that this bound will have coverage rate greater than $1-\gamma$ in general, since it considers the "worst-case" value in the $(1-\gamma/2) \times 100\%$ CI for coefficient estimates entering the lower and upper bounds. If (asymptotic) coverage rate of exactly $1-\gamma$ is desired, one can in principle use the entire

[3] In the denominator, there is a term corresponding to the $sgn(\cdot)$ function is evaluated at the value of numerator. This is to account for the fact that (if the the denominator is positive) increasing the denominator reduces the value when the numerator is positive, but increases it when the numerator is negative.

covariance matrix for the coefficient estimates in conjunction with the delta method.

Remark 8. In practice, for computing the bounds with more than one control variable, one can orthogonalize the z_{ij}^l's using Gram-Schmidt, or by residualizing recursively. However, a simpler approach may to reduce the z_{ij}^l's into a single-dimensional composite control based on a linear combination using utility weights. Specifically, one can estimate the specification including the z_{ij}^l's, and then consider $\tilde{z}_{ij}^1 = \sum_{l=1}^{L} \hat{\delta}_1 z_{ij}^l$ as the composite variable. One can then estimate the specification using $\tilde{z}_{ij}^1$ as the control variable (after standardizing it to have unit variance), and compute the bounds using either equation (3) or (4). [4]

2.3 Relaxing the Distributional Assumption

The assumption that the covariates and error terms follow a Gaussian distribution is stronger than necessary. Below, I introduce an assumption under which we can obtain the same results.

Assumption N (Linear Conditional Expectation). Define $\mu_{ij} \equiv bx_{ij} + aw_{ij} + \sum_{k=1}^{K} d_k z_{ij}^k$, *and assume that:*

$$\mathbb{E}\left[(x_{ij}, w_{ij}, ..., z_{ij}^1, ..., z_{ij}^K)' | (\mu_{i1}, ..., \mu_{iJ})'\right]$$

is linear in $\mu_{i1}, ..., \mu_{iJ}$. *Also, assume that* $(x_{ij}, w_{ij}, ..., z_{ij}^1, ..., z_{ij}^K)'$ *is distributed i.i.d. with positive definite covariance matrix* Σ, *and that* $Cov(z_{ij}^k, z_{ij}^{k'}) = 0$ *for all* $k' \neq k$.

Remark 9. Assumption N relaxes Assumption N0 in two ways. First, it allows for more general distributions for the covariates $(x_{ij}, w_{ij}, ..., z_{ij}^1, ..., z_{ij}^K)'$, as long as they have the linear conditional expectation property given in the assumption. For example, this assumption is satisfied if the distribution of $(x_{ij}, w_{ij}, ..., z_{ij}^1, ..., z_{ij}^K)'$ belongs to the class of spherically symmetrically distributions, which the multivariate Gaussian distribution is a special case of.

The second way in which Assumption N relaxes Assumption N0 is that there are no assumptions on the distribution of ϵ_i. Not only does ϵ_{ij} not have to follow a Gaussian distribution, Assumption N also allows for non-zero correlations between ϵ_{ij} and $\epsilon_{ij'}$ $(j \neq j')$.

The following proposition states that the bounds given in Lemma 1 still holds if Assumption N0 is replaced with Assumption N.

[4] One scenario where this approach of using a composite control variable may not be preferable is if estimating the random utility model is computationally intensive. This is because this approach requires estimating the specification with controls twice – once to obtain the utility weights on the control variables, and a second time with the composite control variable to obtain the coefficients to compute the bounds. In this case, it may be preferable to orthogonalize the control variables beforehand, since this requires estimating the random utility model only once.

Proposition 1. *(OVB Bound with Exogenous Scaling Variable). Under Assumptions D, N, and R0, and E, the bounds for β/α is given by the interval I in equations (3) and (4). Moreover, these bounds are sharp, in the sense that for any τ in I, there exists a data-generating process satisfying Assumptions D, N, R0, and E that yields the estimates $(\breve{\alpha}, \breve{\beta}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}_{x,z^1}, ..., \hat{\gamma}_{x,z^L})'$, and such that $\beta/\alpha = \tau$.*

Proof. Under Assumption N, the coefficient estimates on the covariates derived in the proof of Lemma 1 still hold up to a scale, as shown in Ruud (1983). Since the parameter of interest is a ratio, the scale cancels out, and so the proof of Lemma 1 goes through when we replace Assumption N0 with Assumption N. □

Remark 10. A practical implication of this result is that as long as the researcher believes the regressors have the linear conditional expectation property described in Assumption N, it does not matter what distribution she assumes for the error term. In particular, the researcher may prefer to estimate a conditional logit model instead of a conditional probit model (even if she believes that ϵ_{ij} follow a Gaussian distribution) since it is typically less computationally expensive to do so, and the bounds for the OVB given in equations (3) and (4) will still hold asymptotically.

3 General Model

In this section, we will drop Assumption E, so that now both the the variable of interest and the scaling variable potentially both suffer from OVB. We modify Assumption R0 slightly, to consider additionally, the regression of w_{ij} on $z_{ij}^1, ..., z_{ij}^K$.

Assumption R (R^2 from regressions of x_{ij} and w_{ij} on all the z_{ij}^k's). Assume that the values of the R^2 from regressions of x_{ij} and w_{ij} on $z_{ij}^1, ..., .z_{ij}^K$ are strictly smaller than $R^2_{x,max} \leq 1$ and $R^2_{w,max} \leq 1$ respectively.

Remark 11. Again, the values of $R^2_{x,max}$ and $R^2_{w,max}$ should be specified by the researcher based on contextual knowledge.

Assumption C (Non-Zero Denominator). Assume that the following condition holds:

$$sgn\left(\hat{\alpha} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)\sqrt{R^2_{w,max} - \sum_{l=1}^{K-1} \hat{\rho}^2_{w,z^l}}\right) = sgn\left(\hat{\alpha} + (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)\sqrt{R^2_{w,max} - \sum_{l=1}^{K-1} \hat{\rho}^2_{w,z^l}}\right) \neq 0.$$

Remark 12. Assumption C guarantees that the bounds for the scaling variable does not include zero.

If this assumption is violated, the bound for β/α will be $(-\infty, \infty)$.[5]

Proposition 2. *(OVB Bound with Endogenous Scaling Variable). Under Assumptions D, N, R, and C, $\beta/\alpha \in I = (I_{min}, I_{max})$, where I_{min} and I_{max} are defined by:*

$$\begin{aligned} I_{min} \equiv \min_{\rho_{x,z^K}, \rho_{w,z^K}} & \frac{\hat{\beta} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)\rho_{x,z^K}}{\hat{\alpha} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)\rho_{w,z^K}} \\ s.t. & \left(1 - \sum_{l=1}^{K-1} \hat{\rho}^2_{w,z^l}\right) \rho^2_{x,z^K} + \left(1 - \sum_{l=1}^{K-1} \hat{\rho}^2_{x,z^l}\right) \rho^2_{w,z^K} \\ & + 2\left(-\hat{\rho}_{x,w} + \sum_{l=1}^{K-1} \hat{\rho}_{x,z^l}\hat{\rho}_{w,z^l}\right) \rho_{x,z^K}\rho_{w,z^K} \leq C_{K-1}, \qquad (8) \\ & \rho^2_{x,z^K} \leq R^2_{x,max} - \sum_{l=1}^{K-1} \hat{\rho}^2_{x,z^l}, \\ & \rho^2_{w,z^K} \leq R^2_{w,max} - \sum_{l=1}^{K-1} \hat{\rho}^2_{w,z^l}, \qquad (9) \end{aligned}$$

$$\begin{aligned} I_{max} \equiv \max_{\rho_{x,z^K}, \rho_{w,z^K}} & \frac{\hat{\beta} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)\rho_{x,z^K}}{\hat{\alpha} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)\rho_{w,z^K}} \\ s.t. & \left(1 - \sum_{l=1}^{K-1} \hat{\rho}^2_{w,z^l}\right) \rho^2_{x,z^K} + \left(1 - \sum_{l=1}^{K-1} \hat{\rho}^2_{x,z^l}\right) \rho^2_{w,z^K} \\ & + \quad 2\left(-\hat{\rho}_{x,w} + \sum_{l=1}^{K-1} \hat{\rho}_{x,z^l}\hat{\rho}_{w,z^l}\right) \rho_{x,z^K}\rho_{w,z^K} \leq C_{K-1}, \qquad (10) \\ & \rho^2_{x,z^K} \leq R^2_{x,max} - \sum_{l=1}^{K-1} \hat{\rho}^2_{x,z^l}, \\ & \rho^2_{w,z^K} \leq R^2_{w,max} - \sum_{l=1}^{K-1} \hat{\rho}^2_{w,z^l}, \qquad (11) \end{aligned}$$

and where $C_{\bar{k}}$ is defined by:

$$1 - \hat{\rho}^2_{x,w} + \sum_{k=1}^{\bar{k}} \left(2\hat{\rho}_{x,w}\hat{\rho}_{x,z^k}\hat{\rho}_{w,z^k} - \hat{\rho}^2_{x,z^k} - \hat{\rho}^2_{w,z^k}\right) + \sum_{k=1}^{\bar{k}} \sum_{k' \neq k} \left(\hat{\rho}_{x,z^k}\hat{\rho}_{w,z^{k'}}(\hat{\rho}_{x,z^k}\hat{\rho}_{w,z^{k'}} - \hat{\rho}_{x,z^{k'}}\hat{\rho}_{w,z^k})\right)$$

for any integer $\bar{k} \geq 0$. Moreover, when $R^2_{x,max} = R^2_{w,max} = 1$, this bound is sharp, in the sense that for any τ in I, there exists a data-generating process satisfying Assumptions D, N, R, and C, that yields the estimates $(\check{\alpha}, \check{\beta}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}_{x,z^1}, ..., \hat{\gamma}_{x,z^L})'$, and such that $\beta/\alpha = \tau$.

[5] Even though the setup of the model assumes $\alpha \neq 0$, β/α can still tend to positive and negative infinity as α approaches zero from both sides, hence the reason Assumption C is needed.

The full proof is given in the Appendix, but here I briefly explain the roles played by various quantities in the statement of the proposition. The constrained optimization problems (9) and (11) define the largest and smallest possible values of the true parameter based on the omitted variables bias formula, and the restrictions imposed by the assumptions. The term C_k is equal to the determinant of the kth leading principal minor of the covariance matrix Σ, and the first constraint in the optimization problems ensures that Σ is positive (semi-) definite.

Remark 13. The constrained optimization problems (9) and (11) are non-convex, and the solution generally involves finding zeroes of quartic polynomials and checking boundary conditions. The formulas are rather complicated, and are given in Appendix Section C. However, in practice, one can simply do a grid search over $(\gamma_{x,z^K}, \gamma_{w,z^K}) \in [-1,1]^2$ by evaluating the objective functions for each possible value and checking whether the constraints are satisfied.

Remark 14. Alternatively, one can simply ignore the first constraint and use only the second and third constraints. This will produce a looser bound in general, but there is a simple closed form solution for the bounds, given by:

$$I^{(2)}_{min} = \frac{\hat{\beta} - \left(M\sum_{l=1}^{L}\hat{\delta}_l\right)\left(\sqrt{R^2_{x,max} - \sum_{l=1}^{L}\hat{\rho}^2_{x,z^l}}\right)}{\hat{\alpha} + sgn\left(\hat{\beta} - \left(M\sum_{l=1}^{L}\hat{\delta}_l\right)\left(\sqrt{R^2_{x,max} - \sum_{l=1}^{L}\hat{\rho}^2_{x,z^l}}\right)\right)\left(M\sum_{l=1}^{L}\hat{\delta}_l\right)\left(\sqrt{R^2_{w,max} - \sum_{l=1}^{L}\hat{\rho}^2_{w,z^l}}\right)},$$

$$I^{(2)}_{max} = \frac{\hat{\beta} + \left(M\sum_{l=1}^{L}\hat{\delta}_l\right)\left(\sqrt{R^2_{x,max} - \sum_{l=1}^{L}\hat{\rho}^2_{x,z^l}}\right)}{\hat{\alpha} - sgn\left(\hat{\beta} - \left(M\sum_{l=1}^{L}\hat{\delta}_l\right)\left(\sqrt{R^2_{x,max} - \sum_{l=1}^{L}\hat{\rho}^2_{x,z^l}}\right)\right)\left(M\sum_{l=1}^{L}\hat{\delta}_l\right)\left(\sqrt{R^2_{w,max} - \sum_{l=1}^{L}\hat{\rho}^2_{w,z^l}}\right)},$$

assuming that $\hat{\alpha} - \left(M\sum_{l=1}^{L}\hat{\delta}_l\right)\left(\sqrt{R^2_{w,max} - \sum_{l=1}^{L}\hat{\rho}^2_{w,z^l}}\right) > 0$.

Remark 15. Intuitively, using the first constraint may produce a tighter bound since it utilizes information on the covariances between variables to rule out some values of γ_{x,z^K} and γ_{w,z^K}. For a simple example, consider the case where there is one control variable ($L = 1$), and parameter values:

$$\alpha = \beta = 10,\ \delta_1 = 1,\ \rho_{x,w} = 0.5,\ \rho_{x,z^1} = \rho_{w,z^1} = 0.5,$$

and suppose we assume $M = 1$, $R^2_{x,max} = R^2_{w,max} = 0.9$. Then, without using the first constraint, the objective function takes its minimum when:

$$\rho_{x,z^2} = \sqrt{R^2_{x,max} - \rho^2_{x,z^1}} \approx 0.75,$$
$$\rho_{w,z^2} = -\sqrt{R^2_{x,max} - \rho^2_{x,z^1}} \approx -0.75.$$

However, the first constraint tells us this is not possible. In particular, these values would imply that the following matrix:

$$\begin{bmatrix} 1 & \rho_{x,w} & \rho_{x,z^1} & \rho_{x,z^2} \\ \rho_{x,w} & 1 & \rho_{w,z^1} & \rho_{w,z^2} \\ \rho_{x,z^1} & \rho_{w,z^1} & 1 & 0 \\ \rho_{x,z^2} & \rho_{w,z^2} & 0 & 1 \end{bmatrix}$$

has a determinant of -0.62, so it is not positive semidefinite and cannot be a covariance matrix.[6] Hence, the true magnitude of either ρ_{x,z^2} or ρ_{w,z^2} must be smaller than the assumed values above, so that using the first constraint results in a tighter bound.

Remark 16. Accounting for sampling uncertainty is more tricky in this general case, since it is not immediately obvious whether a given direction of movement in the covariance terms ρ_{x,z^l} and ρ_{w,z^l} increases or decreases lower and upper bounds. In principle, one can consider whether the first constraint binds at the estimated values, and take the partial derivative with respect to each covariance term (to figure out if increasing the covariance tightens or loosens the constraint). However, this is tedious in practice (especially if L is large), and it still does not cover some cases: for example, even if the first constraint does not bind at the estimated values, it is possible that one cannot reject the null hypothesis that it binds at the true values of the parameters due to statistical uncertainty.

Therefore, at the cost of settling for a looser bound, one can use the bounds in Remark 14 that do not use information on the covariance terms ρ_{x,z^l} and ρ_{w,z^l} as a starting point. Specfically, suppose we want a coverage rate of at least $1-\gamma$. Assume also that we can reject the null hypothesis that $\alpha = 0$, and to simplify notation, that:

$$\hat{\alpha} - q_{1-\gamma/2} \cdot se(\hat{\alpha}) - \left(M \sum_{l=1}^{L} (\hat{\delta}_l + q_{1-\gamma/2} \cdot se(\hat{\delta}_l)) \right) \left(\sqrt{R^2_{w,max} - \sum_{l=1}^{L} \left(|\hat{\rho}_{w,z^l}| - q_{1-\gamma/2} \cdot se(\hat{\rho}_{w,z^l}) \right)^2} \right) > 0.$$

[6] One can check that the first three leading principal minors are positive, so that the covariance matrices for the $(x_{ij}, w_{ij}, z^1_{ij})'$ are valid.

Then, we can use the bounds:

$$I(q_{1-\gamma/2}) = \left(I^{(2)}_{min}(q_{1-\gamma/2}), I^{(2)}_{max}(q_{1-\gamma/2})\right) = \left(\frac{I^{(2)}_{min,num}(q_{1-\gamma/2})}{I^{(2)}_{min,den}(q_{1-\gamma/2})}, \frac{I^{(2)}_{max,num}(q_{1-\gamma/2})}{I^{(2)}_{max,den}(q_{1-\gamma/2})}\right),$$

where:

$$I^{(2)}_{min,num}(q_{1-\gamma/2}) \equiv \hat{\beta} - q_{1-\gamma/2} \cdot se(\hat{\beta}) - \left(M \sum_{l=1}^{L} (\hat{\delta}_l + q_{1-\gamma/2} \cdot se(\hat{\delta}_l))\right) \left(\sqrt{R^2_{x,max} - \sum_{l=1}^{L} \left(|\hat{\rho}_{x,z^l}| - q_{1-\gamma/2} \cdot se(\hat{\rho}_{x,z^l})\right)^2}\right),$$

$$I^{(2)}_{min,den}(q_{1-\gamma/2}) = \hat{\alpha} + q_{1-\gamma/2} \cdot se(\hat{\alpha})$$
$$+ sgn\left(I^{(2)}_{min,num}\right) \left(M \sum_{l=1}^{L} (\hat{\delta}_l + q_{1-\gamma/2} \cdot se(\hat{\delta}_l))\right) \left(\sqrt{R^2_{w,max} - \sum_{l=1}^{L} \left(|\hat{\rho}_{w,z^l}| - q_{1-\gamma/2} \cdot se(\hat{\rho}_{w,z^l})\right)^2}\right),$$

$$I^{(2)}_{max,num}(q_{1-\gamma/2}) \equiv \hat{\beta} + q_{1-\gamma/2} \cdot se(\hat{\beta}) + \left(M \sum_{l=1}^{L} (\hat{\delta}_l + q_{1-\gamma/2} \cdot se(\hat{\delta}_l))\right) \left(\sqrt{R^2_{x,max} - \sum_{l=1}^{L} \left(|\hat{\rho}_{x,z^l}| - q_{1-\gamma/2} \cdot se(\hat{\rho}_{x,z^l})\right)^2}\right),$$

$$I^{(2)}_{max,den}(q_{1-\gamma/2}) = \hat{\alpha} - q_{1-\gamma/2} \cdot se(\hat{\alpha})$$
$$- sgn\left(I^{(2)}_{max,num}\right) \left(M \sum_{l=1}^{L} (\hat{\delta}_l + q_{1-\gamma/2} \cdot se(\hat{\delta}_l))\right) \left(\sqrt{R^2_{w,max} - \sum_{l=1}^{L} \left(|\hat{\rho}_{w,z^l}| - q_{1-\gamma/2} \cdot se(\hat{\rho}_{w,z^l})\right)^2}\right).$$

While the distributional assumptions in Assumption N are more general than in N0, it still does not cover certain regressors used in practice such as indicator variables, so we may wonder about the consequences of misspecifying the error distribution in these cases. A well-known interpretation of the estimates from such a misspecified MLE is that they minimize the Kullback-Leibler divergence between the true and assumed distributions. However, from an applied perspective, this still does not provide the researcher with guidance on how far the coefficient estimates are from their true values due to the misspecification.

In the final part of this section, I state a result showing that if the number of controls gets large, the main results hold asymptotically with probability one, with essentially no restrictions on the distributions of the covariates. Before stating the result, I introduce additional notation, and a new assumption that replaces the distributional assumptions in Assumption N. Denote by $Y \equiv (x_i', w_i', z_i^1, ..., z_i^K)'$ a random vector taking values in $\mathbb{R}^d$ where $d = KJ + 2$, and note that we can write $Y = \mu_Y + \Sigma^{1/2} Z$, for a random variable Z with $\mathbb{E}[Z] = 0$, $\mathbb{E}[ZZ'] = I_d$. Also, the matrix norm in this section, denoted by $||\cdot||$ corresponds to the spectral norm.

Assumption S (Steinberger and Leeb 2018). Let $\mathcal{V}_{d,J}$ be the collection of $d \times J$ matrices with orthonormal columns (i.e., the Stiefel manifold), equipped with the Haar measure $\nu_{d,J}(\cdot)$ (i.e., uniform distribution on the Stiefield manifold). Let S_k be the $m \times m = (Z_i Z_j'/d)_{i,j=1}^m$ Gram-matrix for m i.i.d. copies of Z, $Z_1, ..., Z_m$. For $g \geq 1$, let $G = G(S_m - I_m) = \Pi_{l=1}^g (S_m - I_m)_{i_l, j_l}$ for $(i_l, j_l) \in \{1, ..., m\}^2$, $i_l \leq j_l$, $1 \leq l \leq g$, and let $G = 1$ for $g = 0$ (so that G is the monomial of order g). Suppose that the following conditions hold for $m = 2$:

(S1)(a) There are constants $\bar{\epsilon} \in [0, 1/2]$ and $\bar{\alpha} \geq 1$ so that $\mathbb{E}||\sqrt{d}(S_m - I_m)||^{2m+1+\epsilon} \leq \bar{\alpha}$.

(S1)(b) There are constants $\bar{\beta} > 0$ and $\xi \in (0, 1/2]$ that satisfy the following: For any monomial $G = G(S_m - I_m)$ with degree $g \leq 2m$, we have $|d^{g/2}\mathbb{E}[G] - 1| \leq \bar{\beta}/d^{\xi}$ if G consists only of quadratic factors in elements above the diagonal, and $|d^{g/2}\mathbb{E}[G]| \leq \bar{\beta}/d^{\xi}$ if G contains a linear factor.

(S2) There is a constant $D \geq 1$ such that the following is true: if R is an orthogonal $d \times d$ matrix, then the marginal densities of the first $d - m + 1$ components of RZ are bounded by $\binom{d}{m-1}^{1/2} D^{d-k+1}$. Also, assume that $(x_{ij}, w_{ij}, ..., z_{ij}^1, ..., z_{ij}^K)'$ is distributed i.i.d. with positive definite covariance matrix Σ, and that $Cov(z_{ij}^k, z_{ij}^{k'}) = 0$ for all $k' \neq k$.

Note that we are considering a sequences of data-generating processes as d tends to infinity, but for notational simplicity, I will keep the dependence of the parameters and distributions on the sequence implicit.

Proposition 3. *(OVB Bound with Endogenous Scaling Variable, with many Covariates and Products). Suppose Assumptions S, N, R, and C, hold, and that $d \to \infty$, $N/d \to \infty$ and J remains finite or goes to infinity at a rate slower than $log(d)$. Then, there are subsets $\mathbb{J}(\Sigma)$ and $\mathbb{U}(\Lambda)$ from the Stiefel manifolds of dimensions $d \times J$ and $d \times d$ respectively, such that for any values of A satisfying $B \equiv \Sigma^{1/2} A (A'\Sigma A)^{-1/2} \in \mathbb{J}(\Sigma)$, and $\Sigma \in \mathbb{S} \equiv \{U \Lambda U' | \Lambda = diag(\lambda_l) > 0, U \in \mathbb{U}(\Lambda)\}$, the (non-inclusive) upper and lower bounds for β/α approach I_{min} and I_{max}, as given by equations (9) and (11) respectively, and the Haar measures of $\mathbb{J}^c(\Sigma)$ and $\mathbb{U}^c(\Lambda)$ tend to zero.*

Remark 17. It is worth mentioning a key difference in the asymptotics I consider here, compared to papers studying the properties of MLE in high dimensions, e.g., Sur and Candes (2019), and Zhao, Sur, and Candes (2022) who study high-dimensional logistic regressions. These papers typically assume that the number of covariates tend to infinity at the same rate as the number of observations (i.e., where

N/d remains bounded), whereas I assume that number of covariates grow at a slower rate (so that $N/d \to \infty$. On the other hand, these papers require much stronger assumptions on the distribution of the covariates and error terms. For example, Sur and Candes (2019) and Zhao, Sur, and Candes (2022) require that the covariates are jointly normally distributed and that the error term's distribution is correctly specified, whereas my result above requires neither of these assumptions.

Remark 18. As an example of an empirical setting where the number of controls gets large, consider demand estimation for consumers across many different markets. In these cases, it is often sensible to include market fixed effects in the utility equation, and potentially, to allow some of the coefficients to vary by market. A sampling scheme that mirrors the asymptotics assumed in Proposition 3 is as follows:

1. Suppose that there is a superpopulation of markets, from which we draw a random sample of markets.

2. Suppose that within each market, there is a superpopulation of individuals, from which we also draw a random sample.

The first step ensures that $d \to \infty$ since we include market fixed effects in the utility equation, while the second step ensures that $N/d \to \infty$.

4 Simulations

In this section, I present simulation results on the performance of the bounds I derived in the previous sections. Panel A shows specifications with exogenous scaling variable, whereas panel B shows specifications with endogenous scaling variable. For each specification, I run 100 simulations, with 10,000 observations in each simulation. All models are estimated using conditional logit, although the true error distribution is normally distributed. These bounds account for standard errors, and are computed to ensure a coverage rate of at least 95%.

In computing the bounds, I use the true value of $M = 1$. This is unknown to the researcher in practice, but as long as she chooses a larger value of M, she will obtain an even more conservative bound. I also derive bounds under alternative assumptions for $R^2_{x,max}$ and $R^2_{w,max}$, since these values are also unknown to the researcher in practice. For the conservative bounds, I use the most conservative value of 1 for the maximum R^2 values, whereas for the exact bounds, I use the true values of the R^2.

The results in Table 1 shows that the coverage rate is greater than 95%, as we would expect, and that making (valid) assumptions about the values of $R^2_{x,max}$ and $R^2_{w,max}$ can appreciably tighten these bounds.

Table 1: Simulation Results

True Parameters: $\beta/\alpha = 1/10$, $\delta_1 = 1$, $\delta_2 = 1$, M=1

Panel A: Exogenous Scaling Variable

Specification					Coverage Rate		Lower Bound (Mean)		Upper Bound (Mean)	
$\rho_{x,z1}$	$\rho_{x,z2}$	$\rho_{w,z1}$	$\rho_{w,z2}$	$\rho_{x,w}$	Conservative	Exact	Conservative	Exact	Conservative	Exact
0.25	0.25	0	0	0	1.00	1.00	0.007	0.083	0.248	0.177
-0.25	0.25	0	0	0	1.00	1.00	-0.008	0.079	0.262	0.181
0.25	-0.25	0	0	0	1.00	1.00	-0.056	0.031	0.204	0.122
-0.25	-0.25	0	0	0	1.00	1.00	-0.042	0.035	0.191	0.118

Panel B: Endogenous Scaling Variable

Specification					Coverage Rate		Lower Bound (Mean)		Upper Bound (Mean)	
$\rho_{x,z1}$	$\rho_{x,z2}$	$\rho_{w,z1}$	$\rho_{w,z2}$	$\rho_{x,w}$	Conservative	Exact	Conservative	Exact	Conservative	Exact
0.25	0.25	-0.1	0.1	0.1	1.00	1.00	0.013	0.082	0.277	0.175
-0.25	0.25	-0.1	0.1	0.1	1.00	1.00	0.002	0.078	0.294	0.179
0.25	-0.25	-0.1	0.1	0.1	1.00	1.00	-0.061	0.030	0.230	0.119
-0.25	-0.25	-0.1	0.1	0.1	1.00	1.00	-0.043	0.034	0.210	0.115
0.25	0.25	-0.1	-0.1	0.1	1.00	1.00	0.018	0.087	0.281	0.180
-0.25	0.25	-0.1	-0.1	0.1	1.00	1.00	0.006	0.083	0.302	0.185
0.25	-0.25	-0.1	-0.1	0.1	1.00	1.00	-0.055	0.033	0.234	0.123
-0.25	-0.25	-0.1	-0.1	0.1	1.00	1.00	-0.039	0.037	0.214	0.119

Notes: This table shows results from simulations described in the main text. Each 100 simulations are run for each specification, the number of observations in each simulation is 10,000, and all models are estimated using conditional logit. All bounds are formed using the true value of M=1, and the conservative bounds do not employ any assumption on the maximum R-squared from regressions of the endogenous and scaling variables on the controls, whereas the exact bounds use the actual R-squared values for the maximum R-squared. All bounds are formed accounting for standard errors in the parameter estimates, in order to ensure at least 95% coverage rate. The bounds in panel A are formed using the formula for exogenous scaling variables, whereas the boudns in panel B are formed using the formula for endogenous scaling variables.

5 Empirical Applications

In this section, I use two empirical applications to illustrate how the sensitivity analysis can be applied in practice. A general description of each empirical application, and the results of the sensitivity analysis is given here. Details concerning the assumptions and calculations underlying the sensitivity analysis are given in theppendix.

5.1 Chevalier and Goolsbee (2009): Are Consumers Forward Looking?

Chevalier and Goolsbee (2009) test whether consumers are as forward-looking as typical IO models about durable goods assume. They study this in the setting of college textbooks, where the publication of a new version of a textbook results in a dramatic fall in the value of the old version. At a high level, Chevalier and Goolsbee tests whether college students are more price-sensitive when a revision of the textbook is imminent, and to do so they model demand as a function of current price, its interaction with the probability of revision in the near future, as well as textbook characteristics.

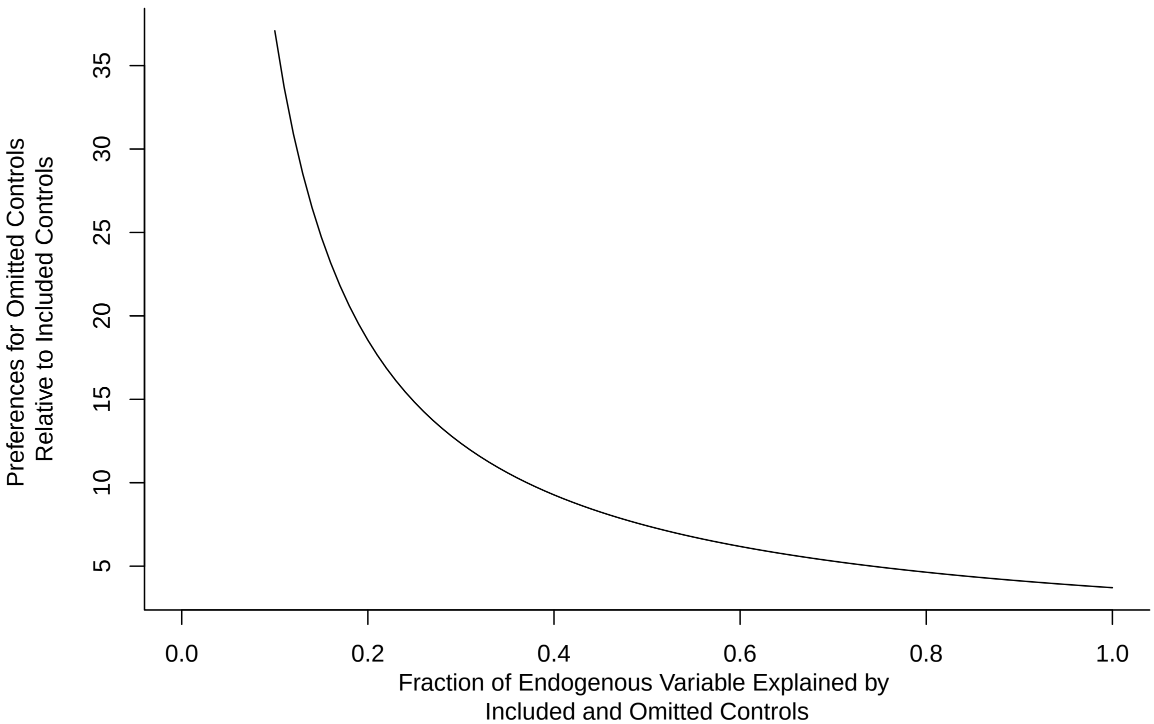
A concern with estimation of this demand model without using instruments is that unobserved quality of a textbook may be positively correlated with price, probability of revision, and demand. This corresponds to the "observational estimate" in column 1 of of Table 4 in Chevalier and Goolsbee (2005), which I use for my sensitivity analysis (although their preferred specifications in the latter columns and subsequent tables address this using IV).[7] Specifically, I derive conditions under which OVB can completely explain the finding from the observational estimate that textbook consumers are forward-looking.[8]

Figure 1 shows the values of M and $R^2_{x,max}$ under which the finding from observational estimates that consumers are forward-looking is not robust to OVB. For example, if we believe that about half of the variation in the interaction between price and revision probability can be explained by textbook characteristics (the included controls) and unobserved quality (the omitted variable), then OVB can completely explain the result if consumers value unobserved quality at least 7.5 times more than they do textbook characteristics. To the extent that such a condition is plausible, this highlights the importance of the IV specifications in Chevalier and Goolsbee (2009), which use current and expected future prices (as well as their interactions) as instruments in their preferred specifications.

[7] I use the results from the NBER working paper version, since the published version unfortunately contains a typographical error in the table that shows the main results.

[8] Some required information is not available from the results presented in the paper (specifically, the covariances between the regressors), so in these cases I need to make some assumptions. This is discussed in greater detail in the Appendix.

Figure 1: Can OVB Explain Results from Observational Estimates Showing that Textbook Consumers are Forward-Looking?



5.2 Cheng (2023): Are Consumers Responsive to Nursing Home Quality?

In my second empirical application, I consider estimates from Cheng (2023) of nursing home residents' demand for quality. Cheng estimates that demand for quality in the nursing home setting is an order of magnitude smaller than previous estimates from studies in hospital settings. In the following sensitivity analysis, I derive conditions under which OVB can completely explain the discrepancy between estimates of demand for quality in the nursing home and hospital settings.

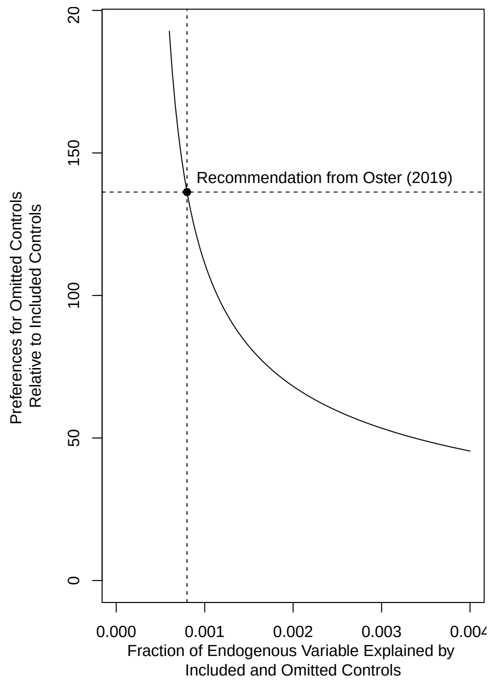
Using the estimates $\hat{\beta}$, $\hat{\alpha}$, $\hat{\delta}_1$, $\hat{\rho}_{x,w}$, $\hat{\rho}_{x,z^K}$, ρ_{w,z^K} from Appendix Table 2, Figure 2 plots the values of M and either $R^2_{x,max}$ or $M_{x,R}$ under which omitted variables bias can completely explain the discrepancy between the demand estimates. We observe in Figures 3a and 3b that under the suggested value of $M_{x,R}$ or $R^2_{x,max}$ from Oster (2019), indicated by the dashed vertical line, residents need to value the omitted variable $M > 100$ times more than observable quality measures in order for OVB to completely explain nursing home residents' low demand. On the other hand, Figures 3c and 3d show that if we assume that almost all of the variation in quality can be explained by the omitted variable, then residents need to value the omitted variable about $M = 2.7$ times more than observable quality measures in order for OVB to completely explain the low demand. Such a value of M may seem plausible, but recall that in order for OVB to explain the results, we also need residents to value the omitted variable positively even though it is strongly negatively correlated with quality, which seems unlikely.

When thinking about the importance of the omitted variable, it is also important to note that we should only consider the *portion of the omitted variable that is not predicted by the included controls.* For example, if we think that the level of comfort provided by the nursing home is an important omitted variable, it seems quite likely that comfort is correlated with staffing levels and cited deficiencies. So, although we may think that comfort is very important (suggesting a large value of M), after partialling out staffing levels, cited deficiencies, as well as for-profit and chain status, the importance of the remaining variation may matter much less for consumer utility (so that M can in fact be much smaller).

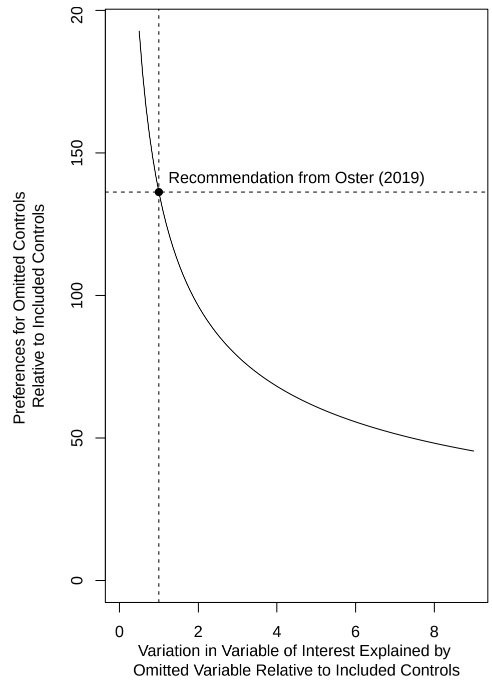
6 Conclusion

In this paper, I derive bounds for the OVB in discrete choice models under assumptions about the relative importance of the unobserved controls relative to the observed ones, as well as the R^2 from regressions of the variable of interest (and the scaling variable) on all controls. Simulation results confirm the validity of these bounds, and I also illustrate how these methods for sensitivity analysis can be used in practice in empirical applications studying whether textbook consumers are forward-looking, and residents' demand for nursing home quality.

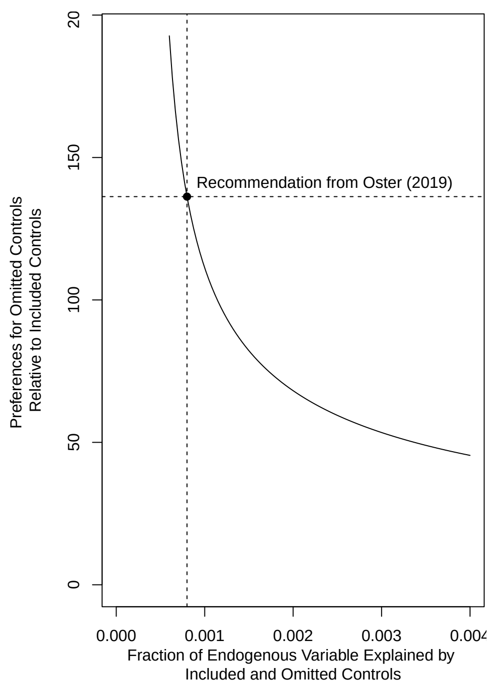
Figure 2: Conditions Under Which OVB Can Completely Explain Low Demand for Quality



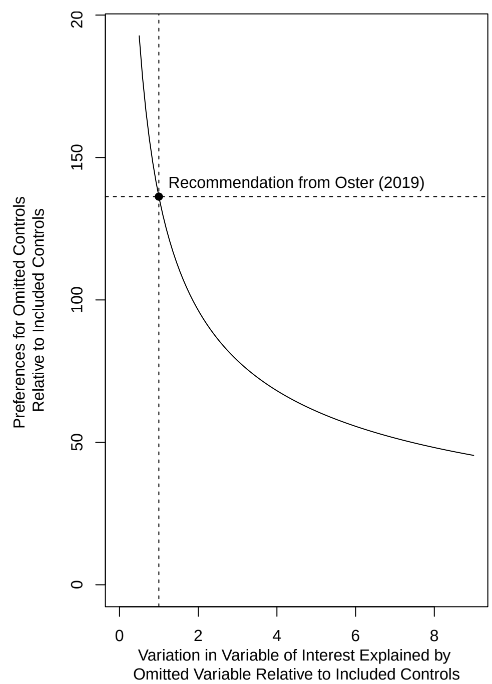
(a) Conditions for $(M, \bar{R}^2_{max})$



(b) Conditions for (M, M_R)



(c) Conditions for $(M, \bar{R}^2_{max})$



(d) Conditions for (M, M_R)

References

[1] Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy*, 113(1): 151–184.

[2] Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro, 2020. "On the Informativeness of Descriptive Statistics for Structural Estimates." *Econometrica*, 88(6): 2231–2258.

[3] Angrist, Joshua, and Jörn-Steffen Pischke, 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives*, 24(2): 3–30.

[4] Sur, Pragya, and Emmanuel Candes, 2019. "A modern maximum-likelihood theory for high-dimensional logistic regression." *Proceedings of the National Academy of Sciences*, 116(29): 14516–14525.

[5] Cheng, Alden, 2023. "Demand for Quality in the Presence of Information Frictions: Evidence from the Nursing Home Market." Working paper.

[6] Chevalier, Judith, and Austan Goolsbee, 2005. "Are durable goods consumers forward-looking? Evidence from college textbooks." NBER Working Paper 11421.

[7] Chevalier, Judith, and Austan Goolsbee, 2009. "Are durable goods consumers forward-looking? Evidence from college textbooks." *The Quarterly Journal of Economics,* 124(4): 1853–1884.

[8] Cinelli, Carlos, and Chad Hazlett, 2020. "Making Sense of Sensitivity: Extending Omitted Variable Bias." *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 82(1): 39–67.

[9] Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi, 2012. "Plausibly exogenous." *Review of Economics and Statistics*, 94(1): 260–272.

[10] Diegert, Paul, Matthew A. Masten, and Alexandre Poirier, 2022. "Assessing Omitted Variables Bias when the Controls are Endogenous." Working paper.

[11] Ding, Peng, and Tyler J. VanderWeele, 2016. "Sensitivity Analysis Without Assumptions." *Epidemiology*, 27(3): 368–377.

[12] Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams, 2021. "Place-Based Drivers of Mortality: Evidence from Migration." *American Economics Review*, 111(8): 2697–2735.

[13] Imbens, Guido W., 2003. "Sensitivity to Exogeneity Assumptions in Program Evaluation." *American Economic Review*, 93(2): 126–132.

[14] Oster, Emily, 2019. "Unobservables Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics*, 37(2): 187–204.

[15] Masten, Matthew A., and Alexandre Poirier, 2023. "The Effect of Omitted Variables on the Sign of Regression Coefficients." Working paper.

[16] Nevo, Aviv, and Adam M. Rosen, 2012. "Identification with imperfect instruments." *Review of Economics and Statistics*, 94(3): 659–671.

[17] Nevo, Aviv, and Michael D. Whinston, 2010. "Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference." *Journal of Economic Perspectives*, 24(2): 69–82.

[18] Rosenbaum, Paul R., and Donald B. Rubin. "Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome." *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(2): 212–218.

[19] Ruud, Paul A., 1983. "Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Choice Models." *Econometrica*, 51(1): 225–228.

[20] VanderWeele, Tyler J., and Peng Ding, 2017. "Sensitivity Analysis in Observational Research: Introducing the E-Value." *Annals of Internal Medicine*, 167: 268–274.

[21] Zhao, Qian, Pragya Sur, and Emmanuel Candes, 2022. "The Asymptotic Distribution of the MLE in High-dimensional Logistic Models: Arbitrary Covariance." *Bernoulli*, 28(3): 1835–1861.

Appendix

A Proofs

Proof of Lemma 1. Suppose we estimate utility without any controls:

$$v_{ij} = \check{\beta} x_{ij} + \check{\alpha} w_{ij} + e_{ij}. \tag{12}$$

Assumption N0 implies that for each k:

$$z_{ij}^k = \rho_{x,z^k} x_{ij} + \nu_{ij}^k,\ \mathbb{E}[\nu_{ij}^k | x_{ij}] = 0,\ Var(\nu_{ij}^k | x_{ij}) = \sigma_{\nu^k}^2. \tag{13}$$

and that $(\nu_{ij}^1, ..., \nu_{ij}^K)'$ is jointly normally distributed with a diagonal covariance matrix.

Substituting equations (13) and (1) into equation (12), we obtain:

$$v_{ij} = \underbrace{\left(\beta + \sum_{k=1}^{K} \delta_k \rho_{x,z^k}\right)}_{=\check{\beta}} x_{ij} + \alpha w_{ij} + \underbrace{\left(\epsilon_{ij} + \sum_{k=1}^{K} \rho_{x,z^k} \nu_{ij}^k\right)}_{=\check{\epsilon}_{ij}}.$$

The error term $\check{\epsilon}_{ij}$ is normally distributed as a consequence of Assumption N0, and thus, equation (12) is still a multinomial probit specification. From this we deduce that the population estimates of the coefficients on x_{ij} and w_{ij} are given by:

$$\check{\beta} = \frac{\beta + \sum_{k=1}^{K} \delta_k \rho_{x,z^k}}{\sqrt{1 + \sum_{k=1}^{K} \delta_k^2 \sigma_{\nu^k}^2}},\ \check{\alpha} = \frac{\alpha}{\sqrt{1 + \sum_{k=1}^{K} \delta_k^2 \sigma_{\nu^k}^2}}. \tag{14}$$

The denominator can be thought as a type of attentuation bias, arising from the fact the utility is scaled differently in the estimation of equations (12) and (1) — specifically, $\check{\epsilon}_{ij} = \epsilon_{ij} + \sum_{k=1}^{K} \rho_{x,z^k} \nu_{ij}^k$ is normalized to have unit variance in the estimation of equation (12), whereas in equation (1) ϵ_{ij} is normalized to have unit variance. Nonetheless, since the parameter of interest is the ratio of coefficients rather than the individual coefficients, this attenuation term cancels out in $\check{\beta}/\check{\alpha}$, so it does not affect the bias for the ratio.

From equation (14), we obtain:

$$\frac{\check{\beta}}{\check{\alpha}} \equiv \frac{\beta}{\alpha} + \underbrace{\frac{\sum_{k=1}^{K} \delta_k \rho_{x,z^k}}{\alpha}}_{\text{OVB (no controls)}},$$

and we see that the formula for the OVB closely mirrors the OLS case. Following the same steps as the case without any controls, we can obtain a similar equation for the OVB for the specification when $(z_{ij}^1, ..., z_{ij}^L)'$ are controlled for in the estimation:

$$\frac{\hat{\beta}}{\hat{\alpha}} \equiv \frac{\beta}{\alpha} + \underbrace{\frac{\delta_K \rho_{x,z^K}}{\alpha}}_{\text{OVB (observed controls)}}.$$

We would like to now consider a worst-case bound for the OVB for $\hat{\beta}/\hat{\alpha}$ under the assumptions we made. For this, it is useful to note that Assumption R0 implies that $\sum_{k=1}^{K} \rho_{x,z^k}^2 < R_{max}^2$, since:

$$\begin{aligned} 1 = Var(x_{ij}) &= Cov\left(\sum_{k=1}^{K} \rho_{x,z^k} z_{ij}^k + e_{ij}, \sum_{k=1}^{K} \rho_{x,z^k} z_{ij}^k + e_{ij}\right) \\ &= \sum_{k=1}^{K} \rho_{x,z^k}^2 + Var(e_{ij}) \end{aligned}$$

$$\implies R_{max}^2 > R^2 = 1 - \frac{Var(e_{ij})}{Var(x_{ij})} = 1 - (1 - \sum_{k=1}^{K} \rho_{x,z^k}^2) = \sum_{k=1}^{K} \rho_{x,z^k}^2.$$

Hence, the worst case bound for the OVB is given by:

$$\begin{aligned} &\max_{\delta_K, \rho_{x,z^K}} \delta_K \rho_{x,z^K} \\ &\qquad \text{s.t. } \delta_K \leq M \sum_{l=1}^{L} \delta_l, \delta_K \geq 0 \; \forall k \in \{L+1, ...K\} \\ &\qquad\quad \rho_{x,z^K}^2 \leq R_{max}^2 - \sum_{l=1}^{L} \rho_{x,z^l}^2, \end{aligned}$$

where the first constraint comes from Assumption D, and the last constraint comes from Assumption R0, and there is an analogous optimization problem for the minimum. [9] It is clear that the maximum is obtained when $\delta_K = M \sum_{l=1}^{L} \delta_l$ and $\rho_{x,z^K} = \sqrt{R_{max}^2 - \sum_{l=1}^{L} \rho_{x,z^l}^2}$, and in the solution for the

[9]More precisely, this last inequality should be strict, but that would imply that we are not optimization over a closed set and there is no solution in general. So, the OVB can be arbitrarily close to this maximum, but not be equal to it under our assumptions.

minimization problem, the sign of ρ_{x,z^K} at the minimum is flipped.

Now, we observe that $\hat{\delta}_l = \delta_l$ since $z^k_{ij} \perp z^{k'}_{ij}$ for $k \neq k'$ under Assumption N0, and under the same assumption, we have $\hat{\rho}_{x,z^l} = \rho_{x,z^l}$ where $\hat{\rho}_{x,z^l}$ is obtained from a regression of z^l_{ij} on x_{ij}. Hence we obtain a formula for the bounds for the true value β/α which depends only on quantities identified in the data:

$$\frac{\beta}{\alpha} \in I \equiv \left(\frac{\hat{\beta}}{\hat{\alpha}} - \left| \left(M \sum_{l=1}^{L} \frac{\hat{\delta}_l}{\hat{\alpha}} \right) \left(\sqrt{R^2_{max} - \sum_{l=1}^{L} \hat{\rho}^2_{x,z^l}} \right) \right|, \frac{\hat{\beta}}{\hat{\alpha}} + \left| \left(M \sum_{l=1}^{L} \frac{\hat{\delta}_l}{\hat{\alpha}} \right) \left(\sqrt{R^2_{max} - \sum_{l=1}^{L} \hat{\rho}^2_{x,z^l}} \right) \right| \right).$$

To obtain the alternative formula for the bound when $L = 1$ and $\hat{\rho}_{x,z^1} \neq 0$, note that the change in the estimate for β/α when we include the control z^1_{ij} is given by:

$$\frac{\breve{\beta}}{\breve{\alpha}} - \frac{\hat{\beta}}{\hat{\alpha}} = \frac{\hat{\delta}_1 \hat{\rho}_{x,z^1}}{\alpha}.$$

Setting $\delta_K = M\hat{\delta}_1$ and $\rho_{x,z^K} = \pm\sqrt{R^2_{max} - \hat{\rho}^2_{x,z^1}}$ and combining this with the equation above, we find that the worst-case OVB is:

$$\begin{aligned} \frac{\delta_K \rho_{x,z^K}}{\alpha} &= \pm M \cdot \frac{\hat{\delta}_1}{\alpha} \cdot \sqrt{R^2_{max} - \hat{\rho}^2_{x,z^1}} \\ &= \pm \frac{M}{\hat{\rho}_{x,z^1}} \left(\frac{\breve{\beta}}{\breve{\alpha}} - \frac{\hat{\beta}}{\hat{\alpha}} \right) \sqrt{R^2_{max} - \hat{\rho}^2_{x,z^1}}, \end{aligned}$$

and thus the bound for the true parameter is given by:

$$\frac{\beta}{\alpha} \in I = \left(\frac{\hat{\beta}}{\hat{\alpha}} - \left| \frac{M}{\hat{\rho}_{x,z^1}} \left(\frac{\hat{\beta}}{\hat{\alpha}} - \frac{\breve{\beta}}{\breve{\alpha}} \right) \left(\sqrt{R^2_{max} - \hat{\rho}^2_{x,z^1}} \right) \right|, \frac{\hat{\beta}}{\hat{\alpha}} + \left| \frac{M}{\hat{\rho}_{x,z^1}} \left(\frac{\hat{\beta}}{\hat{\alpha}} - \frac{\breve{\beta}}{\breve{\alpha}} \right) \left(\sqrt{R^2_{max} - \hat{\rho}^2_{x,z^1}} \right) \right| \right).$$

Finally, these bounds are sharp, since our proof for the worst-case bounds is constructive. In other words, for any values of $\tau \in I$ and identified quantities $(\hat{\beta}, \hat{\alpha}, \hat{\delta}_1, ..., \hat{\delta}_L, \hat{\rho}_{x,z^1}, ..., \hat{\rho}_{x,z^L})$, there are values of δ_K and ρ_{x,z^K} satisfying Assumptions D, N, R0, and E for which β/α satisfies:

$$\frac{\beta}{\alpha} = \frac{\hat{\beta}}{\hat{\alpha}} - \frac{\delta_K \rho_{x,z^K}}{\hat{\alpha}} \equiv \tau,$$

and is thus consistent with the data, e.g. by setting:

$$\delta_K = M \sum_{l=1}^{L} \hat{\delta}_l, \; \rho_{x,z^K} = \frac{(\hat{\beta}/\hat{\alpha} - \tau)(\hat{\alpha})}{M \sum_{l=1}^{L} \hat{\delta}_l}.$$

$\square$

Proof of Proposition 2. By Ruud (1983), we can assume without loss that the covariates and errors follow a joint normal distribution. In this case, we have:

$$\begin{aligned} z_{ij}^k &= \rho_{x,z^k} x_{ij} + \nu_{x,ij}^k,\ \mathbb{E}[\nu_{x,ij}^k | x_{ij}] = 0,\ Var(\nu_{x,ij}^k | x_{ij}) = \sigma_{\nu_x^k}^2. \\ &= \rho_{w,z^k} w_{ij} + \nu_{w,ij}^k,\ \mathbb{E}[\nu_{w,ij}^k | w_{ij}] = 0,\ Var(\nu_{w,ij}^k | w_{ij}) = \sigma_{\nu_w^k}^2. \end{aligned}$$

Substituting this into the probit specification that includes only observed controls, we obtain:

$$v_{ij} = \underbrace{(\beta + \delta_K \rho_{x,z^K})}_{=\hat{\beta}} x_{ij} + \underbrace{(\alpha + \delta_K \rho_{w,z^K})}_{=\hat{\alpha}} w_{ij} + \sum_{l=1}^{L} \delta_l z_{ij}^l + \underbrace{(\epsilon_{ij} + \rho_{x,z^K} \nu_{ij}^K)}_{=\hat{\epsilon}_{ij}}.$$

This implies that the estimated coeffients are related to the true parameter by:

$$\frac{\beta}{\alpha} = \frac{\hat{\beta} - \delta_K \rho_{x,z^K}}{\hat{\alpha} - \delta_K \rho_{w,z^K}}.$$

So, to obtain a bound for the true coefficient, we simply need to minimize and maximize this with respect to $(\delta_K, \rho_{x,z^K}, \rho_{w,z^K})$ under the constraints implied by Assumptions D, N, and R.

First, we note that at the maximum and minimum, the constraint on δ_K in Assumption D clearly binds, so we can set $\delta_K = M \sum_{l=1}^{K-1} \hat{\delta}_l$. The proof for Proposition 1 already derived the second constraint in the constrained optimization problems (9) and (11), and the third constraint is derived exactly the same way replacing x_{ij} with w_{ij}.

The first constraint is equivalent to the covariance matrix $\Sigma \in \mathbb{R}^{K+2,K+2}$ is positive semi-definite. This is because by Sylvester's criterion, Σ is positive semi-definite if and only if all of its leading principal minors are non-negative. The first $K+1$ leading principal minors do not involve the parameters we are optimizating over (i.e., they are given only by moments identified in the data), so we know that they must be positive (by Assumption N). The $(K+2)$-th leading principal minor is the determinant of Σ, which Lemma 2 in the Appendix shows, is equal to C_K. Rewriting the condition that $C_K \geq 0$ so that all the unknowns are on the left hand side, we obtain the first constraint in the constrained optimization problems (9) and (11).

We have shown that the values of ρ_{x,z^K} and ρ_{w,z^K} must satisfy the inequalities. Nonetheless, we must still check that the maximum and minimum exist. To do so, first, we note that the objective function is continuous, which is guaranteed by Assumption C. Next, the set of values $(\rho_{x,z^K}, \rho_{w,z^K})$ can take is clearly bounded, for example, by $[-1,1]^2$. To show that the set defined by the inequalities is closed, we note that the set can be written as the intersection of the inverse images of the intervals $(\infty, r_q]$ under the functions on the left hand side where r_q is the right hand side of the q-th inequality. Since the functions on the right hand side are continuous and $(\infty, r_q]$ are all closed, the inverse images are closed, and so is their intersection, and thus the set of allowed values is closed. Under the Heine-Borel theorem, the set of allowed values is thus compact, and combined with the fact that the objective function is continuous, this implies that the minimum and maximum exist.

Hence, we have proven that β/α lies in the well-defined interval (I_{min}, I_{max}). Finally, to show that the bound is tight when $R^2_{x,max} = R^2_{w,max} = 1$, we note that in this case, the second and third constraints in the constrained optimization problems (9) and (11) are implied by the first constraint.

$\square$

Lemma 2. *The determinant of the covariance matrix for $(x, w, z^1, ..., z^K)'$,Σ, is given by:*

$$C_K \equiv 1 - \hat{\rho}^2_{x,w} + \sum_{k=1}^{K} \left(2\hat{\rho}_{x,w}\hat{\rho}_{x,z^k}\hat{\rho}_{w,z^k} - \hat{\rho}^2_{x,z^k} - \hat{\rho}^2_{w,z^k} \right) + \sum_{k=1}^{K} \sum_{k' \neq k} \left(\hat{\rho}_{x,z^k}\hat{\rho}_{w,z^{k'}} \left(\hat{\rho}_{x,z^k}\hat{\rho}_{w,z^{k'}} - \hat{\rho}_{x,z^{k'}}\hat{\rho}_{w,z^k} \right) \right).$$

Proof of Lemma 2. The determinant for Σ can be calculated using the Leibniz rule. In particular, the determinant is given by:

$$\begin{aligned}
|\Sigma| &= \begin{vmatrix}
1 & \rho_{xw} & \rho_{x,z^1} & \cdots & \cdots & \cdots & \rho_{x,z^K} \\
\rho_{xw} & 1 & \rho_{w,z^1} & \cdots & \cdots & \cdots & \rho_{w,z^K} \\
\rho_{x,z^1} & \rho_{w,z^1} & 1 & 0 & \cdots & \cdots & 0 \\
\vdots & \vdots & 0 & \ddots & 0 & \cdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \ddots & 0 & \vdots \\
\vdots & \vdots & \vdots & \vdots & 0 & \ddots & \vdots \\
\rho_{x,z^K} & \rho_{w,z^K} & 0 & 0 & \cdots & \cdots & 1
\end{vmatrix} \\
&= \left(1 - \sum_{k=1}^{K} \rho^2_{w,z^k} \right) - \rho_{xw}|\Sigma_{\sim x,w}| + \sum_{k=1}^{K} (-1)^{k+1} \rho_{x,z^k}|\Sigma_{\sim x,z^k}|,
\end{aligned}$$

where $\Sigma_{\sim u,v}$ corresponds to the matrix after removing the first row, and the column which has $\rho_{u,v}$ as the first element, for any variables u and v. We can compute that:

$$\rho_{xw}|\Sigma_{\sim x,w}| = \rho_{xw} \begin{vmatrix} \rho_{xw} & \rho_{wz^1} & \cdots & \cdots & \cdots & \rho_{w,z^K} \\ \rho_{x,z^1} & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & 0 & \ddots & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 & \vdots \\ \vdots & \vdots & \vdots & 0 & \ddots & \vdots \\ \rho_{x,z^K} & 0 & \cdots & \cdots & 0 & 1 \end{vmatrix}$$

$$= \rho_{xw}\left(\rho_{xw} - \sum_{k=1}^{K} \rho_{w,z^k}\rho_{x,z^k}\right),$$

where the term $-\sum_{k=1}^{K} \rho_{w,z^k}\rho_{x,z^k}$ can be verified by induction. Similarly, we have:

$$\rho_{x,z^k}|\Sigma_{\sim x,z^k}| = \rho_{x,w}\rho_{w,z^k}\rho_{x,z^k} - \rho^2_{x,z^k} + \rho_{x,z^k}\sum_{k'\neq k} \rho_{w,z^{k'}}(\rho_{x,z^k}\rho_{w,z^{k'}} - \rho_{x,z^{k'}}\rho_{w,z^k}),$$

where the summation term can also be confirmed by induction. Combining these formulae, we obtain:

$$C_K \equiv 1-\hat{\rho}^2_{x,w}+\sum_{k=1}^{K}\left(2\hat{\rho}_{x,w}\hat{\rho}_{x,z^k}\hat{\rho}_{w,z^k}-\hat{\rho}^2_{x,z^k}-\hat{\rho}^2_{w,z^k}\right)+\sum_{k=1}^{K}\sum_{k'\neq k}\left(\hat{\rho}_{x,z^k}\hat{\rho}_{w,z^{k'}}(\hat{\rho}_{x,z^k}\hat{\rho}_{w,z^{k'}} - \hat{\rho}_{x,z^{k'}}\hat{\rho}_{w,z^k})\right),$$

as desired.

□

Proof of Proposition 3. Extending the proof in Ruud (1983) to discrete choice models, the MLE estimates will be consistent even if the distribution of ϵ_i is misspecified, as long as the expectation of each covariate conditional on $\bar{z}_i \equiv A'Y_i \equiv \beta x_i + \alpha w_i + \sum_{k=1}^{K} \delta_k z_i^k \in \mathbb{R}^J$ is linear in $\bar{z}_i$ where A is a $d \times J$ matrix. Denoting $B \equiv \Sigma^{1/2}A(A'\Sigma A)^{-1/2} \in \mathcal{V}_{d,J}$, we can write this condition as:

$$Pr(||\mathbb{E}[Z|B'Z] - BB'Z|| > t) = 0$$

for all $t > 0$, since

$$\mathbb{E}[Y|A'Y] - (\mu_Y + \Sigma^{1/2}P_{\Sigma^{1/2}A}\Sigma^{-1/2}(Y - \mu_Y)) = \Sigma^{1/2}(\mathbb{E}[Z|B'Z] - BB'Z),$$

where $P_{\Sigma^{1/2}A}$ denotes the projection matrix onto the subspace $\Sigma^{1/2}A$. This only holds exactly in finite dimensions for special distributions of Z (specifically, those described in Assumption D), but here I only need to show that as $d \to \infty$ and J remains finite or goes to infinity at a rate slower than $log(d)$, $Pr(||\mathbb{E}[Z|B'Z] - BB'Z|| > t)$ tends to zero for values of $A \in \mathbb{J}(\Sigma)$ and $\Sigma \in \mathbb{S} \equiv \{U\Lambda U' | \Lambda = diag(\lambda_l) > 0, U \in \mathbb{U}(\Lambda)\}$, for sets $\mathbb{J}(\Sigma)$ and $\mathbb{S}$ described in the Proposition.

Fix J and d. Then, under Assumption S, we know from Proposition 3.4 of Steinberger and Leeb (2018) that for each $\tau \in (0,1)$ and $J < d$, there is a Borel set $\mathbb{G} \in \mathcal{V}_{d,J}$ such that for each diagonal positive definite matrix Λ, there exists a collection $\mathbb{U}(\Lambda) = \mathbb{U}(\mathbb{G},\Lambda) \subseteq \mathcal{O}_d$ of orthogonal matrices, satisfying the condition that the sets

$$\mathbb{S} \equiv \mathbb{S}(\mathbb{G}) \equiv \{U\Lambda U' : diag(\lambda_i) > 0, U \in \mathbb{U}(\mathbb{G},\Lambda)\},$$

and

$$\mathbb{J}(\Sigma) \equiv \mathbb{J}(\Sigma,\mathbb{G}) \equiv \{A \in \mathcal{V}_{d,J} : \Sigma^{1/2}A(A'\Sigma A)^{-1/2} \in \mathbb{G}\},$$

have the properties that for $\Sigma \in \mathbb{S}$,

$$\sup_{\Lambda:\Lambda=diag(\lambda_i)>0} \nu_{d,d}(\mathbb{U}(\Lambda)) \leq \sqrt{\kappa_1 d^{-\tau\xi_1\left(1-\frac{\gamma_1}{\tau}\frac{J}{\xi_1 log(d)}\right)}}, \tag{15}$$

$$\sup_{\Sigma\in\mathbb{S}} \nu_{d,J}(\mathbb{J}^c(\Sigma)) \leq \sqrt{\kappa_1 d^{-\tau\xi_1\left(1-\frac{\gamma_1}{\tau}\frac{J}{\xi_1 log(d)}\right)}}, \tag{16}$$

and for any $A \in \mathbb{J}(\Sigma)$ and every $t > 0$,

$$\sup_{B\in\mathbb{G}} Pr(||\mathbb{E}[Z||B'Z] - B'BZ|| > t) \leq \frac{1}{t} d^{-\tau\xi_1} + \frac{\gamma_1}{1-\tau}\frac{J}{3\xi_1 log(d)}, \tag{17}$$

where $\xi_1 \equiv \min\{\xi, \epsilon/2 + 1/4, 1/2\}/3$, and $\gamma_1 = \max\{g_1, 6 + 2log(2D\sqrt{\pi e})\}$, with the constant κ_1 depending only on $\bar{\alpha}$ and $\bar{\beta}$, and g_1 being a global constant. The first two inequalities show that the measure of $\mathbb{J}^c(\Sigma)$ and $\mathbb{U}^c(\Lambda)$ tend to zero as $d \to \infty$ and J remains finite or tends to infinity at a rate slower than $log(d)$.

Finally, given that:

$$||\mathbb{E}[Y|A'Y] - (\mu_Y + \Sigma^{1/2}P_{\Sigma^{1/2}A}\Sigma^{-1/2}(Y-\mu_Y))|| \leq ||\Sigma||^{1/2}||(\mathbb{E}[Z|B'Z] - BB'Z)||,$$

we need to show that $||\Sigma||$ is bounded. Recall that we assumed $z_{ij}^k \perp z_{ij}^{'}$, so the only non-zero elements in Σ are the diagonals, and the entries corresponding to ρ_{x,z^k} and ρ_{w,z^k}. Writing the mth element of Y as Y_m, we have:

$$\begin{aligned}
||\Sigma|| &\leq \sqrt{||\Sigma||_1 ||\Sigma||_\infty} = ||\Sigma||_\infty \\
&= \max_{1\leq m\leq m'} \sum_{1\leq m\leq m'} |\mathbb{E}[Y_k' Y_{k'}]| \\
&= O(J/d),
\end{aligned}$$

where the inequality in the first line is due to Hölder's inequality, and $||\Sigma||_1 = ||\Sigma||_\infty$ since they are equal to the maximum of the column and row sums (respectively) of the absolute values of elements in Σ, which are equal since Σ is symmetric. This implies that as $d \to \infty$ and J remaining finite or tending to infinity at a rate slower than $log(d)$, we have, for any $t > 0$:

$$\begin{aligned}
Pr(||\mathbb{E}[Y|A'Y] - (\mu_Y + \Sigma^{1/2} P_{\Sigma^{1/2}A} \Sigma^{-1/2}(Y - \mu_Y))|| > t) &\leq Pr(||\Sigma||^{1/2} ||(\mathbb{E}[Z|B'Z] - BB'Z)|| > t) \\
\text{(for sufficiently large } d) &\leq Pr(||(\mathbb{E}[Z|B'Z] - BB'Z)|| > t) \\
&\leq \frac{1}{t} d^{-\tau\xi_1} + \frac{\gamma_1}{1-\tau} \frac{J}{3\xi_1 log(d)}
\end{aligned}$$

which tends to zero, as desired.

□

B Calculations for Empirical Applications

B.1 Chevalier and Goolsbee (2009)

The parameter of interest is λ, which is given by the ratio of the coefficient on the interaction between price and revision probability and the coefficient on price (multiplied by -1). First, using the theory developed in this paper, I will derive conditions under which OVB can fully explain the finding that consumers are forward looking when in fact consumers are fully myopic ($\lambda = 0$). Since the paper does not provide information about the covariance between the endogenous variable and the control variables (as well as the covariances between the control variables), some approximations have to be made in the calculation below.

Consider the normalized "composite" control variable:

$$\tilde{z}_{ij} \equiv \sum_{k=1}^{L} \delta_k z_{ij}^k,$$

and normalize it so that:

$$\tilde{z}_{ij} \equiv \frac{\sum_{k=1}^{L} \hat{\delta}_k z_{ij}^k,}{\sqrt{Var\left(\sum_{k=1}^{L} \hat{\delta}_k z_{ij}^k\right),}}.$$

If we replace the controls in the estimation equation with $\tilde{z}_{ij}$, the coefficient on $\tilde{z}$ will be $\sqrt{Var\left(\sum_{k=1}^{L} \hat{\delta}_k z_{ij}^k\right)}$, which is 0.31 if we ignore the correlations between the z_{ij}^k's (which are not given in the paper).

If we also normalize the coefficients on the variable of interest and scaling variable, we obtain $\hat{\beta} \geq 1.13$, $\hat{\alpha} \approx -1.28$. Assuming that the true value of α is negative (i.e., price elasticity of demand is negative), the test of whether consumers are fully myopic reduces to checking whether: $0 \in \left[\hat{\beta} - \left|M\sqrt{Var(\hat{z}_{ij}^*)}\left(\sqrt{R_{max}^2 - \rho_{x,z_{ij}^*}^2}\right)\right|, \hat{\beta} + \left|M\sqrt{Var(\hat{z}_{ij}^*)}\left(\sqrt{R_{max}^2 - \rho_{x,z_{ij}^*}^2}\right)\right|\right]$. Since we do not have information on $\rho_{x,z_{ij}^*}^2$, we simply assume that it is zero in our calculations.

B.2 Cheng (2023)

Quality is quantified in terms of risk-adjusted mortality rate (in percent), and demand for quality is measured in terms of the marginal rate of substitution (MRS) of quality with respect to distance (in miles). As a benchmark for the demand estimates from hospital settings, we use the estimate of 1.8 from Chandra, Finkelstein, Sacarny, and Syverson (2016). Estimating demand for nursing home quality in California between 2008–2010 using a conditional logit model,[10] Column 1 of Appendix Table 2 shows that when we do not include any controls, the estimated MRS is:

$$MRS = -\frac{0.020/0.015}{-1.249/3.882} \cdot 0.01 = 0.041,$$

which is more than 40 times smaller than 1.8, and column 2 shows that the estimate remains largely unchanged if we include controls for RN, LPN, and CNA staffing levels, as well as number of complaint deficiencies that nursing home cited for, and whether the nursing home is for-profit and/or part of a chain. Column 3 verifies that the estimates are unchanged when we use the "composite" control variable

[10] Cheng (2022) considers a more complicated structural demand model which takes unobserved choice set constraints due to selective admissions practices by nursing homes into account. However, for simplicity, we will ignore these choice set constraints in our application here.

$\tilde{z}_{ij}$ which is given by the weighted sum of the original controls with utility weights (normalized to have unit variance).

C Closed Form Expressions for I_{min} and I_{max} in Proposition 2

For I_{min}, consider the candidate values:

$$\rho_{x,z^K} = \sqrt{R^2_{x,max} - \sum_{l=1}^{K-1} \hat{\rho}^2_{x,z^l}},$$

$$\rho_{w,z^K} = -sgn\left(\hat{\beta} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)\sqrt{R^2_{x,max} - \sum_{l=1}^{K-1} \hat{\rho}^2_{x,z^l}}\sqrt{R^2_{x,max} - \sum_{l=1}^{K-1} \hat{\rho}^2_{x,z^l}}\right).$$

If this pair of values satisfy the first constraint, then substituting them into objective function gives us I_{min}. Similarly, flipping the signs of these two values, if the first constraint is satisfied, then they are the solution to the maximization problem or I_{max}.

Now, suppose that that the first inequality is not satisfied with these values. This implies that it holds with equality since the function on the left hand side is continuous (and thus the inequality still holds if we perturb ρ_{x,z^K} or ρ_{w,z^K} in the direction that increases their magnitude). Now, solving ρ_{x,z^K} in terms of ρ_{w,z^K} in this equation, we obtain:

$$\left(1 - \sum_{l=1}^{K-1} \hat{\rho}^2_{w,z^l}\right)\rho^2_{x,z^K} + \left(1 - \sum_{l=1}^{K-1} \hat{\rho}^2_{x,z^l}\right)\rho^2_{w,z^K} + 2\left(-\hat{\rho}_{x,w} + \sum_{l=1}^{K-1} \hat{\rho}_{x,z^l}\hat{\rho}_{w,z^l}\right)\rho_{x,z^K}\rho_{w,z^K} = C_{K-1},$$

$$\rho_{x,z^K} = Q(\rho_{w,z^K}) \equiv \frac{-B \pm \sqrt{B^2 - 4AC(\rho_{w,z^K})}}{2A}, \tag{18}$$

where:

$$A \equiv 1 - \sum_{l=1}^{K-1} \hat{\rho}^2_{w,z^l},$$

$$B(\rho_{w,z^K}) \equiv 2\left(-\hat{\rho}_{x,w} + \sum_{l=1}^{K-1} \hat{\rho}_{x,z^l}\hat{\rho}_{w,z^l}\right)\rho_{w,z^K},$$

$$C(\rho_{w,z^K}) \equiv \left(1 - \sum_{l=1}^{K-1} \hat{\rho}^2_{x,z^l}\right)\rho^2_{w,z^K} - C_{K-1}.$$

Substituting this into the objective function and simplifying, we obtain:

$$
\begin{aligned}
\frac{\hat{\beta}-(M\cdot\sum_{l=1}^{K-1}\hat{\delta}_l)\rho_{x,z^K}}{\hat{\alpha}-(M\cdot\sum_{l=1}^{K-1}\hat{\delta}_l)\rho_{w,z^K}} &= \frac{\hat{\beta}-(M\cdot\sum_{l=1}^{K-1}\hat{\delta}_l)\left(\frac{-B(\rho_{w,z^K})\pm\sqrt{B(\rho_{w,z^K})^2-4AC(\rho_{w,z^K})}}{2A}\right)}{\hat{\alpha}-(M\cdot\sum_{l=1}^{K-1}\hat{\delta}_l)\rho_{w,z^K}} \\
&= \frac{2A\hat{\beta}-(M\cdot\sum_{l=1}^{K-1}\hat{\delta}_l)\left(-B(\rho_{w,z^K})\pm\sqrt{B(\rho_{w,z^K})^2-4AC(\rho_{w,z^K})}\right)}{2A\left(\hat{\alpha}-(M\cdot\sum_{l=1}^{K-1}\hat{\delta}_l)\rho_{w,z^K}\right)}.
\end{aligned}
$$

The first-order conditions are given by:

$$
\begin{aligned}
&\frac{-(FH(\rho_{w,z^K}))\left[G\pm\frac{1}{2}E(\rho_{w,z^K})^{-1/2}\left(2G^2\rho_{w,z^K}-8A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\rho_{w,z^K}\right)\right]}{H(\rho_{w,z^K})^2} \\
&\quad -\frac{\left[2A\hat{\beta}-F\left(-B(\rho_{w,z^K})\pm E(\rho_{w,z^K})^{1/2}\right)\right](-2AF)}{H(\rho_{w,z^K})^2}=0,
\end{aligned}
$$

where:

$$
\begin{aligned}
E(\rho_{w,z^K}) &\equiv B(\rho_{w,z^K})^2-4AC(\rho_{w,z^K}), \\
F &\equiv M\cdot\sum_{l=1}^{K-1}\hat{\delta}_l, \\
G &\equiv 2\left(\hat{\rho}_{x,w}-\sum_{l=1}^{K-1}\hat{\rho}_{x,z^l}\hat{\rho}_{w,z^l}\right), \\
H(\rho_{w,z^K}) &\equiv 2A\left(\hat{\alpha}-(M\cdot\sum_{l=1}^{K-1}\hat{\delta}_l)\rho_{w,z^K}\right).
\end{aligned}
$$

From this, we obtain:

$$
\begin{aligned}
&G\pm E(\rho_{w,z^K})^{-1/2}\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)\rho_{w,z^K}=\frac{\left[2A\hat{\beta}-F\left(-B(\rho_{w,z^K})\pm E(\rho_{w,z^K})^{1/2}\right)\right]2A}{H(\rho_{w,z^K})} \\
&\implies \left[G\pm E(\rho_{w,z^K})^{-1/2}\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)\rho_{w,z^K}\right]\left(\hat{\alpha}-F\rho_{w,z^K}\right)=2A\hat{\beta}+FB(\rho_{w,z^K})\mp E(\rho_{w,z^K})^{1/2} \\
&\implies \left[GE(\rho_{w,z^K})^{1/2}\pm\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)\rho_{w,z^K}\right]\left(\hat{\alpha}-F\rho_{w,z^K}\right)=\left(2A\hat{\beta}+FB(\rho_{w,z^K})\right)E(\rho_{w,z^K})^{1/2}\mp E(\rho_{w,z^K}) \\
&\implies \pm\left[\left(\hat{\alpha}-F\rho_{w,z^K}\right)\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)\rho_{w,z^K}+E(\rho_{w,z^K})\right]=\left[2A\hat{\beta}+FB(\rho_{w,z^K})-\left(\hat{\alpha}-F\rho_{w,z^K}\right)G\right]E(\rho_{w,z^K})^{1/2} \\
&\implies \left[\left(\hat{\alpha}-F\rho_{w,z^K}\right)\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)\rho_{w,z^K}+E(\rho_{w,z^K})\right]^2=\left[2A\hat{\beta}+FB(\rho_{w,z^K})-\left(\hat{\alpha}-F\rho_{w,z^K}\right)G\right]^2E(\rho_{w,z^K}).
\end{aligned}
$$

Given that $B(\rho_{w,z^K})$ is linear in ρ_{w,z^K}, $E(\rho_{w,z^K})$ is quadratic in ρ_{w,z^K}, this is a quartic equation. In particular, we have:

$$\left[\left(\hat{\alpha}-F\rho_{w,z^K}\right)\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)\rho_{w,z^K}+E(\rho_{w,z^K})\right]^2=\left[2A\hat{\beta}+FB(\rho_{w,z^K})-\left(\hat{\alpha}-F\rho_{w,z^K}\right)G\right]^2E(\rho_{w,z^K}).$$

$$\begin{aligned}\implies & E(\rho_{w,z^K})^2+2\left(\hat{\alpha}-F\rho_{w,z^K}\right)\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)\rho_{w,z^K}+\left(\hat{\alpha}-F\rho_{w,z^K}\right)^2\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)^2\rho^2_{w,z^K}\\ &=(2A\hat{\beta}-\hat{\alpha}G)^2E(\rho_{w,z^K}).\end{aligned}$$

$$\begin{aligned}\implies & \left[B(\rho_{w,z^K})^2-4AC(\rho_{w,z^K})\right]^2+\\ &+\left(\hat{\alpha}-F\rho_{w,z^K}\right)^2\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)^2\rho^2_{w,z^K}\\ &-(2A\hat{\beta}-\hat{\alpha}G)^2\left[B(\rho_{w,z^K})^2-4AC(\rho_{w,z^K})\right]\\ &+2\left(\hat{\alpha}-F\rho_{w,z^K}\right)\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)\rho_{w,z^K}=0\end{aligned}$$

$$\begin{aligned}\implies & \left\{\left[G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right]\rho^2_{w,z^K}+4AC_{K-1}\right\}^2+\\ &+\left(\hat{\alpha}^2-2\hat{\alpha}F\rho_{w,z^K}+F^2\rho^2_{w,z^K}\right)\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)^2\rho^2_{w,z^K}\\ &-(2A\hat{\beta}-\hat{\alpha}G)^2\left\{\left[G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right]\rho^2_{w,z^K}+4AC_{K-1}\right\}\\ &+2\left(\hat{\alpha}-F\rho_{w,z^K}\right)\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)\rho_{w,z^K}=0\end{aligned}$$

$$\begin{aligned}\implies & \left[G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right]^2\rho^4_{w,z^K}+8AC_{K-1}\left[G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right]\rho^2_{w,z^K}+16A^2C^2_{K-1}+\\ &+\hat{\alpha}^2\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)^2\rho^2_{w,z^K}-2\hat{\alpha}F\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)^2\rho^3_{w,z^K}\\ &+F^2\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)^2\rho^4_{w,z^K}\\ &-(2A\hat{\beta}-\hat{\alpha}G)^2\left[G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right]\rho^2_{w,z^K}-(2A\hat{\beta}-\hat{\alpha}G)^2 4AC_{K-1}\\ &+2\hat{\alpha}\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)\rho_{w,z^K}-F\left(G^2-4A\left(1-\sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)\rho^2_{w,z^K}=0\end{aligned}$$

$$
\begin{aligned}
\implies & \left[G^2 - 4A\left(1 - \sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right]^2 \rho^4_{w,z^K} + 8AC_{K-1}\left[G^2 - 4A\left(1 - \sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right]\rho^2_{w,z^K} + 16A^2C^2_{K-1} + \\
& + \hat{\alpha}^2\left(G^2 - 4A\left(1 - \sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)^2 \rho^2_{w,z^K} - 2\hat{\alpha}F\left(G^2 - 4A\left(1 - \sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)^2 \rho^3_{w,z^K} \\
& + F^2\left(G^2 - 4A\left(1 - \sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)^2 \rho^4_{w,z^K} \\
& - (2A\hat{\beta} - \hat{\alpha}G)^2\left[G^2 - 4A\left(1 - \sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right]\rho^2_{w,z^K} - (2A\hat{\beta} - \hat{\alpha}G)^2 4AC_{K-1} \\
& + 2\hat{\alpha}\left[G^2 - 4A\left(1 - \sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right]\rho_{w,z^K} - F\left(G^2 - 4A\left(1 - \sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right)\rho^2_{w,z^K} = 0
\end{aligned}
$$

so we can write the quartic equation as:

$$a\rho^4_{w,z^K} + b\rho^3_{w,z^K} + c\rho^2_{w,z^K} + d\rho_{w,z^K} + e = 0,$$

where:

$$
\begin{aligned}
a &\equiv (1 + F^2)\left[G^2 - 4A\left(1 - \sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right]^2, \\
b &\equiv -2\hat{\alpha}F\left[G^2 - 4A\left(1 - \sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right]^2, \\
c &\equiv \left\{8AC_{K-1} + \hat{\alpha}^2\left[G^2 - 4A\left(1 - \sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right] - (2A\hat{\beta} - \hat{\alpha}G)^2 - F\right\}\left[G^2 - 4A\left(1 - \sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right] \\
d &\equiv 2\hat{\alpha}\left[G^2 - 4A\left(1 - \sum_{l=1}^{K-1}\hat{\rho}^2_{x,z^l}\right)\right] \\
e &\equiv 4AC_{K-1}\left(4AC_{K-1} - (2A\hat{\beta} - \hat{\alpha}G)^2\right)
\end{aligned}
$$

Denoting:

$$
\begin{aligned}
p &\equiv 2c^3 - 9bcd + 27ad^2 + 26b^2e - 72ace, \\
q &\equiv c^2 - 3bd + 12ae,
\end{aligned}
$$

we can write the roots as:

$$\rho_{w,z^K}^{(1)} = -\frac{b}{4a} - \frac{1}{2}\sqrt{\frac{b^2}{4a^2} - \frac{2c}{3a} + \frac{q\sqrt[3]{2}}{3a\sqrt[3]{p+\sqrt{-4q^3+p^2}}} + \frac{\sqrt[3]{p+\sqrt{-4q^3+p^2}}}{3a\sqrt[3]{2}}}$$

$$- \frac{1}{2}\sqrt{\frac{b^2}{2a^2} - \frac{4c}{3a} - \frac{q\sqrt[3]{2}}{3a\sqrt[3]{p+\sqrt{-4q^3+p^2}}} - \frac{\sqrt[3]{p+\sqrt{-4q^3+p^2}}}{3a\sqrt[3]{2}} - \frac{\left(-\frac{b^3}{a^3} + \frac{4bc}{a^2} - \frac{8d}{a}\right)}{4\sqrt{\frac{b^2}{4a^2} - \frac{2c}{3a} + \frac{q\sqrt[3]{2}}{3a\sqrt[3]{p+\sqrt{-4q^3+p^2}}} + \frac{\sqrt[3]{p+\sqrt{-4q^3+p^2}}}{3a\sqrt[3]{2}}}}}$$

$$\rho_{w,z^K}^{(2)} = -\frac{b}{4a} - \frac{1}{2}\sqrt{\frac{b^2}{4a^2} - \frac{2c}{3a} + \frac{q\sqrt[3]{2}}{3a\sqrt[3]{p+\sqrt{-4q^3+p^2}}} + \frac{\sqrt[3]{p+\sqrt{-4q^3+p^2}}}{3a\sqrt[3]{2}}}$$

$$+ \frac{1}{2}\sqrt{\frac{b^2}{2a^2} - \frac{4c}{3a} - \frac{q\sqrt[3]{2}}{3a\sqrt[3]{p+\sqrt{-4q^3+p^2}}} - \frac{\sqrt[3]{p+\sqrt{-4q^3+p^2}}}{3a\sqrt[3]{2}} - \frac{\left(-\frac{b^3}{a^3} + \frac{4bc}{a^2} - \frac{8d}{a}\right)}{4\sqrt{\frac{b^2}{4a^2} - \frac{2c}{3a} + \frac{q\sqrt[3]{2}}{3a\sqrt[3]{p+\sqrt{-4q^3+p^2}}} + \frac{\sqrt[3]{p+\sqrt{-4q^3+p^2}}}{3a\sqrt[3]{2}}}}}$$

$$\rho_{w,z^K}^{(3)} = -\frac{b}{4a} - \frac{1}{2}\sqrt{\frac{b^2}{4a^2} - \frac{2c}{3a} + \frac{q\sqrt[3]{2}}{3a\sqrt[3]{p+\sqrt{-4q^3+p^2}}} + \frac{\sqrt[3]{p+\sqrt{-4q^3+p^2}}}{3a\sqrt[3]{2}}}$$

$$- \frac{1}{2}\sqrt{\frac{b^2}{2a^2} - \frac{4c}{3a} - \frac{q\sqrt[3]{2}}{3a\sqrt[3]{p+\sqrt{-4q^3+p^2}}} - \frac{\sqrt[3]{p+\sqrt{-4q^3+p^2}}}{3a\sqrt[3]{2}} + \frac{\left(-\frac{b^3}{a^3} + \frac{4bc}{a^2} - \frac{8d}{a}\right)}{4\sqrt{\frac{b^2}{4a^2} - \frac{2c}{3a} + \frac{q\sqrt[3]{2}}{3a\sqrt[3]{p+\sqrt{-4q^3+p^2}}} + \frac{\sqrt[3]{p+\sqrt{-4q^3+p^2}}}{3a\sqrt[3]{2}}}}}$$

$$\rho_{w,z^K}^{(4)} = -\frac{b}{4a} - \frac{1}{2}\sqrt{\frac{b^2}{4a^2} - \frac{2c}{3a} + \frac{q\sqrt[3]{2}}{3a\sqrt[3]{p+\sqrt{-4q^3+p^2}}} + \frac{\sqrt[3]{p+\sqrt{-4q^3+p^2}}}{3a\sqrt[3]{2}}}$$

$$+ \frac{1}{2}\sqrt{\frac{b^2}{2a^2} - \frac{4c}{3a} - \frac{q\sqrt[3]{2}}{3a\sqrt[3]{p+\sqrt{-4q^3+p^2}}} - \frac{\sqrt[3]{p+\sqrt{-4q^3+p^2}}}{3a\sqrt[3]{2}} + \frac{\left(-\frac{b^3}{a^3} + \frac{4bc}{a^2} - \frac{8d}{a}\right)}{4\sqrt{\frac{b^2}{4a^2} - \frac{2c}{3a} + \frac{q\sqrt[3]{2}}{3a\sqrt[3]{p+\sqrt{-4q^3+p^2}}} + \frac{\sqrt[3]{p+\sqrt{-4q^3+p^2}}}{3a\sqrt[3]{2}}}}}$$

Finally, denoting:

$$\rho^*_{x,z^K} \equiv \sqrt{R^2_{x,max} - \sum_{l=1}^{K-1} \hat{\rho}^2_{x,z^l}},$$

$$\rho^*_{w,z^K} \equiv -sgn\left(\hat{\beta} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)\sqrt{R^2_{x,max} - \sum_{l=1}^{K-1} \hat{\rho}^2_{x,z^l}}\right)\sqrt{R^2_{x,max} - \sum_{l=1}^{K-1} \hat{\rho}^2_{x,z^l}},$$

$$h(\rho_{x,z^K}, \rho_{w,z^K}) \equiv \left(1 - \sum_{l=1}^{K-1} \hat{\rho}^2_{w,z^l}\right)\rho^2_{x,z^K} + \left(1 - \sum_{l=1}^{K-1} \hat{\rho}^2_{x,z^l}\right)\rho^2_{w,z^K} + 2\left(-\hat{\rho}_{x,w} + \sum_{l=1}^{K-1} \hat{\rho}_{x,z^l}\hat{\rho}_{w,z^l}\right)\rho_{x,z^K}\rho_{w,z^K},$$

$$(\rho^{*,min}_{x,z^K}, \rho^{*,min}_{x,z^K}) \equiv \min\left\{\frac{\hat{\beta} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)Q(\rho^{(m)}_{w,z^K})}{\hat{\alpha} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)\rho^{(m)}_{w,z^K}}\right\}_{(Q(\rho^{(m)}_{w,z^K}),\rho^{(m)}_{w,z^K}) \in S}$$

$$(\rho^{*,max}_{x,z^K}, \rho^{*,max}_{x,z^K}) \equiv \max\left\{\frac{\hat{\beta} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)Q(\rho^{(m)}_{w,z^K})}{\hat{\alpha} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)\rho^{(m)}_{w,z^K}}\right\}_{(Q(\rho^{(m)}_{w,z^K}),\rho^{(m)}_{w,z^K}) \in S}$$

$$S \equiv \left\{(r_1, r_2)|r_1 = Q(r_2), (r_1, r_2) \in \mathbb{R}^2, r_1^2 \leq R^2_{x,max} - \sum_{l=1}^{K-1} \hat{\rho}^2_{x,z^l}, r_2^2 \leq R^2_{w,max} - \sum_{l=1}^{K-1} \hat{\rho}^2_{w,z^l}\right\},$$

where $Q(\cdot)$ is given by equation (18), we can write the lower and upper bounds as:

$$I_{min} = \frac{\hat{\beta} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)\rho^{*,min}_{x,z^K}}{\hat{\alpha} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)\rho^{*,min}_{w,z^K}},\ I_{max} = \frac{\hat{\beta} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)\rho^{*,max}_{x,z^K}}{\hat{\alpha} - (M \cdot \sum_{l=1}^{K-1} \hat{\delta}_l)\rho^{*,max}_{w,z^K}},$$

where:

$$(\rho^{*,min}_{x,z^K}, \rho^{*,min}_{x,z^K}) = \begin{cases} (\rho^*_{x,z^K}, \rho^*_{w,z^K}) & \text{if } h(\rho^*_{x,z^K}, \rho^*_{w,z^K}) \leq C_{K-1} \\ (\rho^{min,quartic}_{x,z^K}, \rho^{min,quartic}_{w,z^K}) & \text{otherwise} \end{cases}$$

$$(\rho^{*,max}_{x,z^K}, \rho^{*,max}_{x,z^K}) = \begin{cases} (-\rho^*_{x,z^K}, -\rho^*_{w,z^K}) & \text{if } h(-\rho^*_{x,z^K}, -\rho^*_{w,z^K}) \leq C_{K-1} \\ (\rho^{max,quartic}_{x,z^K}, \rho^{max,quartic}_{w,z^K}) & \text{otherwise.} \end{cases}$$

Table 2: Conditional Logit Estimates of Nursing Home Residents' Demand for Quality

	(1)	(2)	(3)
Resident Preferences			
Quality (s.d.)	0.020***	0.018***	0.018***
	(0.003)	(0.003)	(0.003)
Distance to Nursing Home (s.d.)	-1.249***	-1.250***	-1.250***
	(0.003)	(0.004)	(0.004)
RN Hours Per Resident-Day		0.563***	
		(0.007)	
LPN Hours per Resident-Day		0.386***	
		(0.007)	
CNA Hours Per Resident-Day		-0.250***	
		(0.003)	
Deficiencies		0.011***	
		(0.001)	
Chain		0.254***	
		(0.006)	
For-Profit		0.150***	
		(0.010)	
Utility Index for Controls (s.d.)			0.319***
			(0.003)
Number of Observations	7,778,104	7,780,646	7,780,646

Notes: This table shows conditional logit estimates of nursing home residents' demand for quality. Standard deviation of quality and distance are 0.015 and 3.882 respectively. Standard errors in parentheses.