Online Appendix for "Patient Costs and Physicians' Information"

Michael J. Dickstein (New York University)
Jihye Jeon (Boston University)
Eduardo Morales (Princeton University)

November 2025

Contents

A Data: Additional Details

In Appendix A.1, we describe the steps we follow to build our sample. In Appendix A.2, we indicate how we obtain our measure of the out-of-pocket costs for all drugs and visits in the sample.

A.1 Sample Construction

The APAC data contain separate information on medical claims and drug claims. Here, we describe the steps we follow to build our estimation sample from this information. Table A.1 provides the number of observations included in the dataset at each sample restriction stage.

First, for each medical claim, which can include multiple "lines" with different dates and provider identifiers, we choose the earliest date and the most frequently listed provider.

Second, we select the subset of medical claims that list both claim and provider identifiers and a type 2 diabetes diagnosis, which we select using the ICD-10 diagnostic codes that include the words "type 2 diabetes mellitus" in the description. A list of these codes is available upon request.

Third, we link providers' information to the medical claims data. Based on the provider NPI number, we validate the provider information in the APAC data using the National Plan and Provider Enumeration System (NPPES) registry. From the NPPES registry, we collect information on providers' specialties and restrict the sample to medical claims for which the provider is in a specialty that often treats diabetes patients; i.e., family medicine, internal medicine, endocrinology, pediatrics, obstetrics and gynecology, clinical nurse specialists, and physician assistants.

Table A.1: Sample Restrictions - Number of Observations

Medical claims	
All medical claims (with non-mising claim ID)	89,921,304
Include only Oregon-based providers	82,558,080
Exclude missing NPIs	80,108,168
Apply specialty restrictions	34,238,516
Include only type 2 diabetes diagnosis	1,123,169
Exclude providers with max. yearly number of claims in the top or bottom 5%	802,801
Exclude providers with max. yearly number of diabetes claims in the bottom 10%	779,262
Matched claims	
All matched claims	600,044
Missing copay or rxdays	599,792
Restrict plan types/ markets	595,870
Restrict carriers	588,104
Exclude drugs filled before marketing date	588,104
Exclude refills	586,839
Restrict to specific drug classes	17,442

Note: Each line reports the number of observations that we preserve at each sample restriction stage.

Fourth, for each provider, we compute the maximum annual counts of all medical claims and all type 2 diabetes related claims. We then exclude claims corresponding to providers with maximum annual counts of all medical claims in the highest or lowest 5% of this distribution. We similarly exclude observations linked to providers with maximum yearly counts of type 2 diabetes claims in the lowest decile of this distribution: in practice, this excludes providers with a maximum yearly count of type 2 diabetes claims of zero or one.

Fifth, we merge drug claims with the medical claims left after the first four cleaning steps. The matching process first creates, for each patient, all combinations of a drug claim and a medical claim. We then match each drug claim to the medical claim whose date has the smallest distance to the drug fill date, and exclude matches whose distance is outside of the -7 to 180 day range.

Sixth, we exclude claims with missing information on plan type or carrier, or corresponding to plan types and carriers with a small number of observations. We further restrict claims to those corresponding to the following plan types: HMO, POS, PPO, SIF (Self-insured POS), SIP (Self-insured PPO), and EPO.

Seventh, we remove claims that reflect a refill, as well as claims for drugs outside of the two treatment classes we use in our analysis: DPP-4 inhibitors and SGLT2 inhibitors.

A.2 Constructing Copayment Measures

While the claims data report out-of-pocket costs only for the prescription filled by the patient, our analysis requires out-of-pocket costs for all drugs in the patient's choice set. To solve this missing data problem, we construct measures of the out-of-pocket costs that the patient would face in any given year for each drug in their choice set. We base this prediction on information on drug and year identifiers as well as the patient's plan type, carrier, and Metropolitan Statistical Area (MSA) of residence. Our baseline prediction model is a random forest model and, when estimating this model, we restrict the sample to prescriptions for 30-day supplies.

In Table A.2, we show different measures of the performance of our prediction model. We randomly split our sample into a training sample containing 75% of the observations and a test sample containing the remaining 25% of the observations. We then estimate the prediction model using alternately data from the training sample and from the full sample; we then compute the R^2 of a regression of observed prices on predicted prices for the observations in the two samples.

As shown in the first three rows of Table A.2, the R^2 equals 0.41 for the out-of-sample prediction, 0.51 for the in-sample prediction on the full sample, and 0.52 for the in-sample prediction on the training sample. The out-of-bag error is 0.46, as shown in the last row of Table A.2.

A regression model that includes drug-, plan type-, carrier-, patient MSA-, and year-specific fixed effects yields comparable goodness-of-fit measures. However, the comparison is not perfect, as the regression model cannot produce price predictions for observations corresponding to drug-plan type-carrier-patient MSA-year combinations for which the training sample contains no observations; 23% of the drug-plan type-carrier-patient MSA-year combinations have no observations in the

Table A.2: Copayment Predictions: Goodness-of-fit Measures of Random Forest Models

Excluded regressors:	None		Plan Type	Carrier
Model:	Rnd. Forest	Linear Reg.	Rnd. Forest	Rnd. Forest
Full Sample	0.51	0.50	0.47	0.36
Training Sample (75%)	0.52	0.51	0.48	0.37
Test Sample (25%)	0.41	0.42	0.39	0.31
Out-of-bag Error, Full Sample	0.46		0.48	0.52

Note: This table reports R^2 from regressing observed copayment on the copayment predicted using different methods. The first column uses a random forest model that incorporates indicators for carrier, plan type, year, drug, and patient MSA. The second column uses a linear regression model that includes the interaction of carrier, plan type, year, drug, and patient MSA fixed effects. In the third and fourth columns, we use a random forest model, but omit plan type and carrier, respectively, from the set of independent variables. In all specifications, we use observations that reflect a 30-day supply.

training sample.

In the third and fourth columns of Table A.2, we use the random forest method, but omit plan type and carrier indicators, respectively, from the set of prediction variables. We find the R^2 decreases more, and the out-of-bag error increases more, when we omit carrier indicators than when we omit plan type indicators, suggesting that carrier indicators are more important factors for predicting drug prices than plan type indicators.

B Moment Inequalities: Additional Details

We prove Theorems 1 and 2 in Appendix sections B.1 and B.2, respectively. In Appendix B.3, we show how to optimally choose $e_{jj'}(z_i, \theta)$ for every (z_i, θ) and drugs j and j', and introduce an alternative representation of the bounding inequality when the approximation point is chosen in this way. In Appendix B.4, we present an alternative derivation of the bounding inequality that mimics the steps described in Appendix B.1 to derive the odds-based inequality. The content of Appendix sections B.2 and B.3 reproduces results in Porcher et al. (2025).

B.1 Odds-based Moment Inequalities: Proof of Theorem 1

To prove Theorem 1, we show that, for any j and j' and $z_i \subseteq \mathcal{W}_i$, equation (7) holds for $\theta = \mu$; i.e.,

$$\mathbb{E}[d_{ij} \exp(-(u(x_{ij}, \mu) - u(x_{ij'}, \mu))) - d_{ij'} | z_i] \geqslant 0, \tag{B.1}$$

for any choices j and j' and any $z_i \subseteq \mathcal{W}_i$. We organize our proof in three steps, described below.

Step 1. Equation (2) implies that, for any observation i and any two choices j and j', it holds that

$$(d_{ij} + d_{ij'})(\mathbb{1}\{\mathbb{E}[\mathcal{U}_{ij} - \mathcal{U}_{ij'} | \mathcal{J}_i] \geqslant 0\} - d_{ij}) = 0. \tag{B.2}$$

Using equations (1a), (3), and (5), we can rewrite this expression as

$$(d_{ij} + d_{ij'})(\mathbb{1}\{\mathbb{E}[\Delta u_{ijj'} | \mathcal{W}_i] + \Delta \varepsilon_{ijj'} \geqslant 0\} - d_{ij}) = 0, \tag{B.3}$$

with $\Delta u_{ijj'} = u_{ij} - u_{ij'}$ and $\Delta \varepsilon_{ijj'} = \varepsilon_{ij} - \varepsilon_{ij'}$. As equation (B.3) holds for every observation i, it also holds on average across subsets of observations. Thus,

$$\mathbb{E}[\mathbb{1}\{\mathbb{E}[\Delta u_{ijj'} | \mathcal{W}_i] + \Delta \varepsilon_{ijj'} \geqslant 0\} - d_{ij} | \mathcal{W}_i, d_{ij} + d_{ij'} = 1] = 0,$$

and, given the distributional assumptions in equation (4), we can rewrite this equation as

$$\mathbb{E}\left[\frac{\exp(\mathbb{E}[\Delta u_{ijj'} | \mathcal{W}_i])}{1 + \exp(\mathbb{E}[\Delta u_{ijj'} | \mathcal{W}_i])} - d_{ij} \middle| \mathcal{W}_i, d_{ij} + d_{ij'} = 1\right] = 0.$$

Multiplying by $1 + \exp(\mathbb{E}[\Delta u_{ijj'} | \mathcal{W}_i])$ on both sides of the equality and grouping terms, we obtain

$$\mathbb{E}[(1 - d_{ij}) \exp(\mathbb{E}[\Delta u_{ijj'} | \mathcal{W}_i]) - d_{ij} | \mathcal{W}_i, d_{ij} + d_{ij'} = 1] = 0.$$

Conditional on the event $d_{ij} + d_{ij'} = 1$, the variable $1 - d_{ij}$ equals $d_{ij'}$, and we can thus write

$$\mathbb{E}[d_{ij'} \exp(\mathbb{E}[\Delta u_{ijj'} | \mathcal{W}_i]) - d_{ij} | \mathcal{W}_i, d_{ij} + d_{ij'} = 1] = 0.$$

Using the Law of Iterated Expectations (LIE), we eliminate the event $d_{ij} + d_{ij'} = 1$ from the conditioning set, obtaining

$$\mathbb{E}[d_{ij'} \exp(\mathbb{E}[\Delta u_{ijj'}|\mathcal{W}_i]) - d_{ij}|\mathcal{W}_i] = 0,$$

We divide both sides of this equality by $-\exp(\mathbb{E}[\Delta u_{ijj'}|\mathcal{W}_i])$ to further obtain

$$\mathbb{E}[d_{ij} \exp(-\mathbb{E}[\Delta u_{ijj'}|\mathcal{W}_i]) - d_{ij'}|\mathcal{W}_i] = 0. \tag{B.4}$$

Step 2. As d_{ij} is measurable with respect to $\mathcal{J}_i$, $\exp(x)$ is convex in x, and rational expectations implies $\Delta u_{ijj'}$ is a mean-preserving spread of $\mathbb{E}[\Delta u_{ijj'}|\mathcal{J}_i]$, Jensen's inequality applies and yields

$$\mathbb{E}[d_{ij} \exp(-\mathbb{E}[\Delta u_{ijj'}|\mathcal{J}_i])|\mathcal{J}_i] \leqslant \mathbb{E}[d_{ij} \exp(-\Delta u_{ijj'})|\mathcal{J}_i].$$

Using equation (5), we can rewrite this inequality as

$$\mathbb{E}[d_{ij} \exp(-\mathbb{E}[\Delta u_{ijj'}|\mathcal{W}_i])|\mathcal{J}_i] \leqslant \mathbb{E}[d_{ij} \exp(-\Delta u_{ijj'})|\mathcal{J}_i].$$

Furthermore, as $\mathcal{W}_i \subseteq \mathcal{J}_i$, applying the LIE, we obtain

$$\mathbb{E}[d_{ij} \exp(-\mathbb{E}[\Delta u_{ijj'}|\mathcal{W}_i])|\mathcal{W}_i] \leqslant \mathbb{E}[d_{ij} \exp(-\Delta u_{ijj'})|\mathcal{W}_i]. \tag{B.5}$$

Step 3. The equality in equation (B.4) and the inequality in equation (B.5) jointly imply

$$\mathbb{E}[d_{ij} \exp(-\Delta u_{ijj'}) - d_{ij'}|\mathcal{W}_i] \geqslant 0.$$

Finally, if $z_i \subseteq \mathcal{W}_i$, the LIE implies that

$$\mathbb{E}[d_{ij} \exp(-\Delta u_{ijj'}) - d_{ij'}|z_i] \geqslant 0,$$

and, given equation (1b), we can rewrite this inequality as in equation (B.1). □

B.2 Bounding Moment Inequalities: Proof of Theorem 2

The content of this section reproduces the results in Porcher et al. (2025), adjusted to the notation in our setting. To prove Theorem 2, we show that, for any choices j and j', any $z_i \subseteq \mathcal{W}_i$, and any $e_{jj'} \colon \mathcal{Z} \times \Theta \to \mathbb{R}$, equation (13) holds when $\theta = \mu$; i.e.,

$$\mathbb{E}[d_{ij} \exp(-e_{jj'}(z_i, \mu))(1 + e_{jj'}(z_i, \mu) - (u(x_{ij}, \mu) - u(x_{ij'}, \mu))) - d_{ij'}|z_i] \leqslant 0, \tag{B.6}$$

We organize our proof in four steps. Step 1 is the same as step 1 of the proof in Section B.1. We describe below steps 2 to 4.

Step 2. As $\exp(x)$ is convex in x, a first-order approximation to this function around any point bounds it from below. Denoting the approximation point for observation i as $e_{jj'}(z_i, \mu)$, we have

$$\mathbb{E}[d_{ij} \exp(-\mathbb{E}[\Delta u_{ijj'}|\mathcal{W}_i])|\mathcal{W}_i] \\ \geqslant \\ \mathbb{E}[d_{ij}(\exp(-e_{jj'}(z_i,\mu)) - \exp(-e_{jj'}(z_i,\mu))(\mathbb{E}[\Delta u_{ijj'}|\mathcal{W}_i] - e_{jj'}(z_i,\mu)))|\mathcal{W}_i]. \tag{B.7}$$

Combining the equality in equation (B.4) with the inequality in equation (B.7), we obtain

$$\mathbb{E}[d_{ij} \exp(-e_{jj'}(z_i,\mu))(1 + e_{jj'}(z_i,\mu) - \mathbb{E}[\Delta u_{ijj'}|\mathcal{W}_i]) - d_{ij'}|\mathcal{W}_i] \leqslant 0. \tag{B.8}$$

Step 3. As d_{ij} is measurable with respect to $\mathcal{J}_i$ and the assumption of rational expectations implies that $\Delta u_{ijj'}$ is a mean-preserving spread of $\mathbb{E}[\Delta u_{ijj'}|\mathcal{J}_i]$, it holds for any $z_i \subseteq \mathcal{W}_i$ that

$$\mathbb{E}[d_{ij} \exp(-e_{jj'}(z_i,\mu))(1 + e_{jj'}(z_i,\mu) - \mathbb{E}[\Delta u_{ijj'}|\mathcal{J}_i])|\mathcal{J}_i] = \\ \mathbb{E}[d_{ij} \exp(-e_{jj'}(z_i,\mu))(1 + e_{jj'}(z_i,\mu) - \Delta u_{ijj'})|\mathcal{J}_i].$$

Using equation (5), we can rewrite this equality as

$$\mathbb{E}[d_{ij} \exp(-e_{jj'}(z_i,\mu))(1 + e_{jj'}(z_i,\mu) - \mathbb{E}[\Delta u_{ijj'}|\mathcal{W}_i])|\mathcal{J}_i] = \\ \mathbb{E}[d_{ij} \exp(-e_{jj'}(z_i,\mu))(1 + e_{jj'}(z_i,\mu) - \Delta u_{ijj'})|\mathcal{J}_i].$$

As $\mathcal{W}_i \subseteq \mathcal{J}_i$, we can apply the LIE to further simplify this equality as

$$\mathbb{E}[d_{ij} \exp(-e_{jj'}(z_i,\mu))(1 + e_{jj'}(z_i,\mu) - \mathbb{E}[\Delta u_{ijj'}|\mathcal{W}_i])|\mathcal{W}_i] = \\ \mathbb{E}[d_{ij} \exp(-e_{jj'}(z_i,\mu))(1 + e_{jj'}(z_i,\mu) - \Delta u_{ijj'})|\mathcal{W}_i]. \tag{B.9}$$

Step 4. The inequality in equation (B.8) and the equality in equation (B.9) jointly imply

$$\mathbb{E}[d_{ij} \exp(-e_{jj'}(z_i,\mu))(1 + e_{jj'}(z_i,\mu) - \Delta u_{ijj'}) - d_{ij'}|\mathcal{W}_i] \leqslant 0.$$

Finally, we take an expectation on both sides of this inequality conditional on z_i. If $z_i \subseteq \mathcal{W}_i$, equation (1b) implies we can rewrite this inequality as in equation (B.6). □

B.3 Bounding Moment Inequalities: Approximation Point

In Section B.3.1, we derive the set e that minimizes $\Theta_0^b(e)$. In Section B.3.2, we introduce a simpler formulation of the inequality in equation (13) when $e_{jj'}(z, \theta)$ is chosen according to equation (18), or the "optimal" bounding inequality. The content of this section reproduces results in Porcher et al. (2025), with the notation adjusted to our setting.

B.3.1 Optimal Approximation Point

Given two choices j and j', we choose for every z_i in its support and every θ in the parameter space, the value $e_{jj'}(z_i, \theta)$ that maximizes the moment in equation (13b). To do so, we compute the value of $e_{jj'}(z_i, \theta)$ that sets the first derivative of $\mathbb{m}^b_{jj'}(z_i, \theta, e_{jj'}(\cdot))$ to zero; i.e., that is,

$$\frac{\partial \mathbb{m}^b_{jj'}(z_i, \theta, e_{jj'}(\cdot))}{\partial e_{jj'}(z_i, \theta)} = 0,$$

with $\mathbb{m}^b_{jj'}(z_i, \theta, e_{jj'}(\cdot))$ defined as in equation (13b). This first-order condition equals:

$$\mathbb{E}[-d_{ij} \exp(-e_{jj'}(z_i, \theta))(e_{jj'}(z_i, \theta) - (u(x_{ij}, \theta) - u(x_{ij'}, \theta)))|z_i] = 0.$$

Dividing by $-\exp(-e_{jj'}(z_i, \theta))$ on both sides of this expression, we obtain

$$\mathbb{E}[d_{ij}(e_{jj'}(z_i, \theta) - (u(x_{ij}, \theta) - u(x_{ij'}, \theta)))|z_i] = 0.$$

As, according to the model in Section 3, $\mathbb{E}[d_{ij}|z_i] \neq 0$ for every j, this equality holds if and only if

$$\mathbb{E}[e_{jj'}(z_i, \theta) - (u(x_{ij}, \theta) - u(x_{ij'}, \theta))|z_i, d_{ij} = 1] = 0.$$

The value of $e_{jj'}(z, \theta)$ that satisfies this equation is

$$e_{jj'}(z_i, \theta) = \mathbb{E}[u(x_{ij}, \theta) - u(x_{ij'}, \theta)|z_i, d_{ij} = 1],$$

which equals that in equation (18). To verify that $\mathbb{m}^b_{jj'}(z_i, \theta, e_{jj'}(\cdot))$ is maximized when equation (18) holds, we compute the second-order condition, which equals:

$$\mathbb{E}[d_{ij} \exp(-e_{jj'}(z_i, \theta))(-1 + e_{jj'}(z_i, \theta) - (u(x_{ij}, \theta) - u(x_{ij'}, \theta)))|z_i].$$

As $\exp(x) > 0$ for any x, the sign of this conditional expectation equals that of

$$\mathbb{E}[d_{ij}(-1 + e_{jj'}(z_i, \theta) - (u(x_{ij}, \theta) - u(x_{ij'}, \theta)))|z_i].$$

As $\mathbb{E}[d_{ij}|z_i] > 0$ for every j, the sign of this conditional expectation equals that of

$$\mathbb{E}[-1 + e_{jj'}(z_i, \theta) - (u(x_{ij}, \theta) - u(x_{ij'}, \theta))|z_i, d_{ij} = 1].$$

Plugging in this expression the value of $e_{jj'}(z_i, \theta)$ in equation (18), we find that it equals -1. Thus, the second-order condition is negative and, consequently, the value of $e_{jj'}(z_i, \theta)$ in equation (18) does indeed provide a maximum to the moment function $\mathbb{m}^b_{jj'}(z_i, \theta, e_{jj'}(\cdot))$ for every z_i and θ.

B.3.2 Alternative Formulation of Optimal Bounding Moment Inequality

We present here an alternative formulation of the bounding inequality in equation (B.6) when $e_{jj'}(z_i, \mu)$ is determined as in equation (18). Combining equations (B.6) and (18), we obtain

$$\mathbb{E}[d_{ij} \exp(-\mathbb{E}[\Delta u_{ijj'}|z_i, d_{ij} = 1])(1 + \mathbb{E}[\Delta u_{ijj'}|z_i, d_{ij} = 1] - \Delta u_{ijj'}) - d_{ij'}|z_i] \leqslant 0, \tag{B.10}$$

where $\Delta u_{ijj'} = u(x_{ij}, \mu) - u(x_{ij'}, \mu)$. Note that

$$\mathbb{E}[d_{ij} \exp(-\mathbb{E}[\Delta u_{ijj'}|z_i, d_{ij} = 1])(1 + \mathbb{E}[\Delta u_{ijj'}|z_i, d_{ij} = 1] - \Delta u_{ijj'})|z_i] =$$
$$\mathbb{E}[d_{ij}|z_i] \exp(-\mathbb{E}[\Delta u_{ijj'}|z_i, d_{ij} = 1])(1 + \underbrace{(\mathbb{E}[\Delta u_{ijj'}|z_i, d_{ij} = 1] - \mathbb{E}[\Delta u_{ijj'}|z_i, d_{ij} = 1])}_{=0}).$$

Therefore, we can simplify the expression in equation (B.10) as

$$\mathbb{E}[d_{ij} \exp(-\mathbb{E}[\Delta u_{ijj'}|z_i, d_{ij} = 1]) - d_{ij'}|z_i] \leqslant 0,$$

which is equivalent to

$$\mathbb{E}[d_{ij} \exp(-\mathbb{E}[\Delta u_{ijj'}|z_i, d_{ij}]) - d_{ij'}|z_i] \leqslant 0.$$

Finally, using again the expression $\Delta u_{ijj'} = u(x_{ij}, \mu) - u(x_{ij'}, \mu)$, we obtain

$$\mathbb{E}[d_{ij} \exp(-\mathbb{E}[u(x_{ij}, \mu) - u(x_{ij'}, \mu)|z_i, d_{ij}]) - d_{ij'}|z_i] \leqslant 0. \tag{B.11}$$

As this bounding moment inequality uses the "optimal" approximation point introduced in equation (18), we refer to the bounding inequality in equation (B.11) as the "optimal" bounding inequality.

B.4 Alternative Derivation of Optimal Bounding Inequality

Building on a comment of an anonymous referee, we describe here a different derivation of the bounding inequality in equation (19). While the derivation of the general bounding inequality in Appendix B.2 uses a first-order approximation to the probability ratio of choosing two alternatives in the choice set, the derivation of the optimal bounding inequality we describe here uses Jensen's inequality instead. By using Jensen's inequality, we can more easily compare the bounding inequality in equation (19) to the odds-based inequality in equation (12). We organize our proof in four steps. Step 1 is the same as that in Appendix B.1. We describe here steps 2 to 4.

Step 2. For any $z_i \subseteq \mathcal{W}_i$, equation (9) and the LIE implies the following equality:

$$\mathbb{E}[d_{ij} \exp(-\mathbb{E}[\Delta u_{ijj'}|\mathcal{W}_i]) - d_{ij'}|z_i] = 0. \tag{B.12}$$

Step 3. As $\exp(x)$ is convex in x and $\mathbb{E}[\Delta u_{ijj'}|\mathcal{J}_i]$ is a mean-preserving spread of $\mathbb{E}[\Delta u_{ijj'}|z_i, d_{ij}]$ for any $z_i \subseteq \mathcal{W}_i$, Jensen's inequality applies and we have

$$\mathbb{E}[d_{ij}\exp(-\mathbb{E}[\Delta u_{ijj'}|z_i, d_{ij}])|z_i, d_{ij}] \leqslant \mathbb{E}[d_{ij}\exp(-\mathbb{E}[\Delta u_{ijj'}|\mathcal{J}_i])|z_i, d_{ij}]. \tag{B.13}$$

Using equation (5), we can rewrite this inequality as:

$$\mathbb{E}[d_{ij}\exp(-\mathbb{E}[\Delta u_{ijj'}|z_i, d_{ij}])|z_i, d_{ij}] \leqslant \mathbb{E}[d_{ij}\exp(-\mathbb{E}[\Delta u_{ijj'}|\mathcal{W}_i])|z_i, d_{ij}].$$

Applying the LIE, we further obtain

$$\mathbb{E}[d_{ij}\exp(-\mathbb{E}[\Delta u_{ijj'}|z_i, d_{ij}])|z_i] \leqslant \mathbb{E}[d_{ij}\exp(-\mathbb{E}[\Delta u_{ijj'}|\mathcal{W}_i])|z_i]. \tag{B.14}$$

Step 4. The equality in equation (B.12) and the inequality in equation (B.14) jointly imply

$$\mathbb{E}[d_{ij}\exp(-\mathbb{E}[\Delta u_{ijj'}|z_i, d_{ij}]) - d_{ij'}|z_i] \leqslant 0,$$

which coincides with equation (19). □

B.5 Moment Inequalities with Measurement Error in Covariates

Here, we derive valid inequalities that differ from those derived in Appendix sections B.1 and B.2 in two dimensions. First, we impose the form for utility in equation (23): $u_{ij} = \kappa_j + \alpha p_{ij}$. Second, we allow for the measured price difference $\Delta\hat{p}_{ijj'}$ to differ from the true price difference $\Delta p_{ijj'}$ by an unknown quantity that is mean zero conditional on the physician's information set; that is,

$$\mathbb{E}[\Delta\hat{p}_{ijj'} - \Delta p_{ijj'}|\mathcal{J}_i] = 0. \tag{B.15}$$

We then prove here that, given equations (1) to (5), (23), and (B.15), we have

$$\mathbb{E}[d_{ij}\exp(-(\Delta\kappa_{jj'} + \alpha\Delta\hat{p}_{ijj'})) - d_{ij'}|z_i] \geqslant 0, \tag{B.16}$$

for any choices j and j' and any $z_i \subseteq \mathcal{W}_i$; and, we have

$$\mathbb{E}[d_{ij'} - d_{ij}\exp(-e_{jj'}(z_i,\mu))(1 + e_{jj'}(z_i,\mu) - (\Delta\kappa_{jj'} + \alpha\Delta\hat{p}_{ijj'}))|z_i] \geqslant 0, \tag{B.17}$$

for any choices j and j', any $z_i \subseteq \mathcal{W}_i$, and any $e_{jj'}\colon \mathcal{Z}\times\Theta \to \mathbb{R}$.

Proof of equation (B.16). Combining equations (1) to (5) and equation (23), we have

$$\mathbb{E}[d_{ij}\exp(-(\Delta\kappa_{ijj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i])) - d_{ij'}|\mathcal{W}_i] = 0, \tag{B.18}$$

which equals equation (9) once we impose on that equation the restriction on utility in equation (23). As d_{ij} is measurable with respect to $\mathcal{J}_i$, $\exp(x)$ is convex in x, and equation (B.15) implies that $\Delta\hat{p}_{ijj'}$ is a mean-preserving spread of $\mathbb{E}[\Delta p_{ijj'}|\mathcal{J}_i]$, Jensen's inequality applies and yields

$$\mathbb{E}[d_{ij}\exp(-(\Delta\kappa_{ijj'}+\alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{J}_i]))|\mathcal{J}_i]\leqslant\mathbb{E}[d_{ij}\exp(-(\Delta\kappa_{ijj'}+\alpha\Delta\hat{p}_{ijj'}))|\mathcal{J}_i].$$

Using equation (5), we can rewrite this inequality as

$$\mathbb{E}[d_{ij}\exp(-(\Delta\kappa_{ijj'}+\alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))|\mathcal{J}_i]\leqslant\mathbb{E}[d_{ij}\exp(-(\Delta\kappa_{ijj'}+\alpha\Delta\hat{p}_{ijj'}))|\mathcal{J}_i].$$

Furthermore, as $\mathcal{W}_i\subseteq\mathcal{J}_i$, applying the LIE, we obtain

$$\mathbb{E}[d_{ij}\exp(-(\Delta\kappa_{ijj'}+\alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))|\mathcal{W}_i]\leqslant\mathbb{E}[d_{ij}\exp(-(\Delta\kappa_{ijj'}+\alpha\Delta\hat{p}_{ijj'}))|\mathcal{W}_i].\tag{B.19}$$

The equality in equation (B.18) and the inequality in equation (B.19) jointly imply

$$\mathbb{E}[d_{ij}\exp(-(\Delta\kappa_{ijj'}+\alpha\Delta\hat{p}_{ijj'}))-d_{ij'}|\mathcal{W}_i]\geqslant 0.$$

Finally, if $z_i\subseteq\mathcal{W}_i$, the LIE implies equation (B.16). □

Proof of equation (B.17)*.* We start the derivation from the equality in equation (B.18). As $\exp(x)$ is convex in x, a first-order approximation to this function around any point bounds it from below. Denoting the approximation point for observation i as $e_{jj'}(z_i,\mu)$, we then have

$$\mathbb{E}[d_{ij}\exp(-e_{jj'}(z_i,\mu))(1+e_{jj'}(z_i,\mu)-(\Delta\kappa_{jj'}+\alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))-d_{ij'}|\mathcal{W}_i]\leqslant 0.\tag{B.20}$$

As d_{ij} is measurable with respect to $\mathcal{J}_i$ and equation (B.15) implies that $\Delta\hat{p}_{ijj'}$ is a mean-preserving spread of $\mathbb{E}[\Delta p_{ijj'}|\mathcal{J}_i]$, it holds for any $z_i\subseteq\mathcal{W}_i$ that

$$\begin{gathered}\mathbb{E}[d_{ij}\exp(-e_{jj'}(z_i,\mu))(1+e_{jj'}(z_i,\mu)-(\Delta\kappa_{jj'}+\alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{J}_i]))|\mathcal{J}_i]=\\ \mathbb{E}[d_{ij}\exp(-e_{jj'}(z_i,\mu))(1+e_{jj'}(z_i,\mu)-(\Delta\kappa_{jj'}+\alpha\Delta\hat{p}_{ijj'}))|\mathcal{J}_i].\end{gathered}$$

Using equation (5), we can rewrite this equality as

$$\begin{gathered}\mathbb{E}[d_{ij}\exp(-e_{jj'}(z_i,\mu))(1+e_{jj'}(z_i,\mu)-(\Delta\kappa_{jj'}+\alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))|\mathcal{J}_i]=\\ \mathbb{E}[d_{ij}\exp(-e_{jj'}(z_i,\mu))(1+e_{jj'}(z_i,\mu)-(\Delta\kappa_{jj'}+\alpha\Delta\hat{p}_{ijj'}))|\mathcal{J}_i].\end{gathered}$$

As $\mathcal{W}_i\subseteq\mathcal{J}_i$, we can apply the LIE to further simplify this equality as

$$\begin{gathered}\mathbb{E}[d_{ij}\exp(-e_{jj'}(z_i,\mu))(1+e_{jj'}(z_i,\mu)-(\Delta\kappa_{jj'}+\alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))|\mathcal{W}_i]=\\ \mathbb{E}[d_{ij}\exp(-e_{jj'}(z_i,\mu))(1+e_{jj'}(z_i,\mu)-(\Delta\kappa_{jj'}+\alpha\Delta\hat{p}_{ijj'}))|\mathcal{W}_i].\end{gathered}\tag{B.21}$$

The inequality in equation (B.20) and the equality in equation (B.21) jointly imply

$$\mathbb{E}[d_{ij}\exp(-e_{jj'}(z_i,\mu))(1+e_{jj'}(z_i,\mu)-(\Delta\kappa_{jj'}+\alpha\Delta\hat{p}_{ijj'}))-d_{ij'}|\mathcal{W}_i]\leqslant 0.$$

Finally, as $z_i \subseteq \mathcal{W}_i$, the LIE implies equation (B.17). □

B.6 Instrument Functions

Given a scalar z_i, to compute the results in Section 6.2 and Appendix C, we use the instrument functions $g^{(1)}_{jj'}(z_i) = \mathbb{1}\{\Delta z_{ijj'} \leqslant 0\}$ and $g^{(2)}_{jj'}(z_i) = \mathbb{1}\{\Delta z_{ijj'} > 0\}$ with $\Delta z_{ijj'} = z_{ij} - z_{ij'}$. To compute the results in sections 7.2 and 7.3, we use the instrument functions

$$\begin{aligned}
g^{(1)}_{jj'}(z_i) &= \mathbb{1}\{\Delta z_{ijj'} \leqslant p^-_{25}(\Delta z_{ijj'})\},\\
g^{(2)}_{jj'}(z_i) &= \mathbb{1}\{p^-_{25}(\Delta z_{ijj'}) < \Delta z_{ijj'} \leqslant p^-_{50}(\Delta z_{ijj'})\},\\
g^{(3)}_{jj'}(z_i) &= \mathbb{1}\{p^-_{50}(\Delta z_{ijj'}) < \Delta z_{ijj'} \leqslant p^-_{75}(\Delta z_{ijj'})\},\\
g^{(4)}_{jj'}(z_i) &= \mathbb{1}\{p^-_{75}(\Delta z_{ijj'}) < \Delta z_{ijj'} < 0\},\\
g^{(5)}_{jj'}(z_i) &= \mathbb{1}\{0 \leqslant \Delta z_{ijj'} < p^+_{25}(\Delta z_{ijj'})\},\\
g^{(6)}_{jj'}(z_i) &= \mathbb{1}\{p^+_{25}(\Delta z_{ijj'}) \leqslant \Delta z_{ijj'} < p^+_{50}(\Delta z_{ijj'})\},\\
g^{(7)}_{jj'}(z_i) &= \mathbb{1}\{p^+_{50}(\Delta z_{ijj'}) \leqslant \Delta z_{ijj'} < p^+_{75}(\Delta z_{ijj'})\},\\
g^{(8)}_{jj'}(z_i) &= \mathbb{1}\{p^+_{75}(\Delta z_{ijj'}) \leqslant \Delta z_{ijj'}\},
\end{aligned}$$

where, for all $q \in [0, 100]$, $p^-_q(\Delta z_{ijj'})$ and $p^+_q(\Delta z_{ijj'})$ denote the qth percentile of the distribution of negative and positive values of $\Delta z_{ijj'}$, respectively.

B.7 Inference Procedures

In Section B.7.1 below, we describe how we implement the conditional chi-square (CC) test described in Section 3 of Cox and Shi (2023). In Section B.7.2, we describe how we implement the procedure in Holm (1979) to adjust p-values to account for the multiple hypotheses we test.

B.7.1 Cox and Shi (2023)

We denote the set of L moment inequalities that we use in estimation as $\bar{m}(\theta) \geqslant 0$ with $\bar{m}(\theta) = (\bar{m}_1(\theta), \ldots, \bar{m}_L(\theta))'$. We denote as μ the true value of the parameter vector θ. Our implementation of the inference procedure in Cox and Shi (2023) then includes the following steps.

Step 1: define a grid Θ_g that will contain the confidence set for μ.
Step 2: choose a point $\theta_p \in \Theta_g$. For a significance level α, we test $H_0 : \mu = \theta_p$ vs. $H_1 : \mu \neq \theta_p$.

Step 3: evaluate the quasi-likelihood ratio statistic at θ_p:

$$T(\theta_p) = \min_{\lambda:\lambda \geqslant 0} N(\bar{m}(\theta_p) - \lambda)'\tilde{\Sigma}(\theta_p)^{-1}(\bar{m}(\theta_p) - \lambda). \tag{B.22}$$

with

$$\tilde{\Sigma}(\theta_p) = \hat{\Sigma}(\theta_p) + \max\{0.012 - \det(\hat{\Omega}(\theta_p), 0\} Diag(\hat{\Sigma}(\theta_p)). \tag{B.23}$$

The matrix $\hat{\Sigma}(\theta_p)$ is

$$\hat{\Sigma}(\theta_p) = \frac{1}{N}\sum_{i=1}^{N}(\bar{m}_i(\theta) - \bar{m}(\theta))(\bar{m}_i(\theta) - \bar{m}(\theta))' \tag{B.24}$$

and the correlation matrix $\hat{\Omega}(\theta_p)$ is computed as $\hat{\Omega}(\theta_p) = Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p))\hat{\Sigma}(\theta_p)Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p))$, where $Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p))$ is a matrix such that $Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p))Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p)) = Diag^{-1}(\hat{\Sigma}(\theta_p))$, and $Diag(\hat{\Sigma}(\theta_p))$ is the diagonal matrix whose diagonal elements are equal to those of $\hat{\Sigma}(\theta_p)$.

The estimator of the variance-covariance matrix of $\bar{m}(\theta_p)$ in equation (B.23) is introduced in page 2808 in Andrews and Barwick (2012). The reason for using this estimator instead of that in equation (B.24) is that the expression in equation (B.22) requires the estimator of the variance-covariance matrix of $\bar{m}(\theta_p)$ to be invertible. By correcting the estimator in equation (B.24) according to equation (B.23), we ensure the resulting matrix is always non-singular. When using the expression in equation (B.23) as an estimator of the covariance matrix of $\bar{m}(\theta_p)$ in our empirical application (see sections 7.2 and 7.3), a limitation is that this estimator ignores the fact that the moments in $\bar{m}(\theta_p)$ depend on a vector of generated regressors; that is, the vector of predicted prices generated as discussed in Appendix A.2.

Step 4: count how many values of λ in equation (B.22) equal 0. We denote this number as $\hat{r}$.

Step 5: include θ_p in the $(1-\alpha)\%$ confidence set for μ if $T(\theta_p) \leqslant \chi^2_{\hat{r},1-\alpha}$, where $\chi^2_{\hat{r},1-\alpha}$ is the $100(1-\alpha)\%$ quantile of the chi-squared distribution with $\hat{r}$ degrees of freedom.

Step 6: repeat steps 2 to 5 for every θ_p in the grid Θ_g.

B.7.2 Adjusting P-values Following Holm (1979)

We describe our implementation of the procedure in Holm (1979) to adjust p-values when testing multiple hypotheses. This section's content follows that of Online Appendix A.8.2 in Dickstein and Morales (2018). Given tests $H_1, \ldots, H_S$ with individual p-values $p_1, \ldots, p_S$, we proceed as follows:

Step 1: rank hypotheses. Rank the S hypotheses in increasing order of their individual p-values. Denote this index as (i).

Step 2: adjust individual p-values. Denoting as $\tilde{p}_{(i)}$ the adjusted p-value for the (i)-th smallest individual p-value, we compute $\tilde{p}_{(i)} = \max_{j \leqslant i} \big\{ \min\{(S-j+1)p_{(j)}, 1\}\big\}$.

C Additional Simulation Results

In Appendix C.1, we describe the maximum likelihood estimator (MLE). In Appendix C.2, we consider additional cases beyond those presented in Table 2 in Section 6.2. In Appendix C.3, we present results analogous to those in Table 2, but with confidence sets computed following Andrews and Soares (2010). In Appendix C.4, we use simulated samples whose size is comparable to that in our empirical application to illustrate the statistical properties of confidence sets computed following the procedures in Cox and Shi (2023) and Andrews and Soares (2010).

C.1 Maximum Likelihood Estimator

To compute the maximum likelihood estimates for cases 1 to 4 in Table 2 as well as in the different tables presented below in this Appendix section, we solve the following maximization problem:

$$\underset{(\theta_\alpha,\theta_{\kappa_2},\theta_{\kappa_3})}{\operatorname{argmax}} \left\{ \sum_{i=1}^{N}\sum_{j=1}^{3} \mathbb{1}\{d_{ij}=1\} \ln\left(\frac{\exp(\theta_{\kappa_j}+\theta_\alpha x_{2ij})}{\sum_{j'=1}^{3}\exp(\theta_{\kappa_{j'}}+\theta_\alpha x_{2ij'})}\right)\right\}, \qquad \text{with } \theta_{\kappa_1}=0.$$

To compute the maximum likelihood estimates for case 5 in Table 2 as well as in the different tables presented below in this Appendix section, we solve the following maximization problem:

$$\underset{(\theta_\alpha,\theta_{\kappa_2},\theta_{\kappa_3})}{\operatorname{argmax}} \left\{ \sum_{i=1}^{N}\sum_{j=1}^{3} \mathbb{1}\{d_{ij}=1\} \ln\left(\frac{\exp(\theta_{\kappa_j}+\theta_\alpha p_{ij})}{\sum_{j'=1}^{3}\exp(\theta_{\kappa_{j'}}+\theta_\alpha p_{ij'})}\right)\right\}, \qquad \text{with } \theta_{\kappa_1}=0.$$

C.2 Additional Cases

Comparing the results in case 2(b) in Table C.1 to the results in case 2 in Table 2, we notice that the confidence set defined by the bounding inequalities is a singleton in both cases, and this point is equal to the true parameter value. The confidence set defined by the odds-based inequalities includes parameter values other than the true value. This confidence set is larger in case 2(b), when the dispersion parameter of the expectational error equals $\sigma_3 = 2$, than it is in case 2, when $\sigma_3 = 1$. A comparison of the results for cases 2 and 2(b) thus shows that, while the confidence set defined by the bounding inequalities is invariant to the dispersion in the expectational error, the confidence set defined by the odds-based inequalities increases in the dispersion in the expectational error.

Comparing the results in case 4(b) in Table C.1 to the results in case 4 in Table 2, we observe that, as we increase the value of σ_1 and σ_3, the confidence sets defined by the odds-based inequalities and by the bounding inequalities both increase. The downward bias in the MLE also increases.

Finally, comparing the results in case 5(b) in Table C.1 to the results in case 5 in Table 2, we observe that, as we increase the variance of the expectational error (i.e., as we increase σ_3 from 1 to 2), the downward bias in the MLE increases, while the confidence sets defined by the odds-based moment inequalities and by the combination of both types of inequalities remain empty.

Table C.1: Simulation Results - MLE and Confidence Intervals

Case	σ_1	σ_3	z_i	Estimator	MLE & Confidence Sets		
					α	κ_2	κ_3
2(b)	0	2	x_{2i}	MLE	1	0	1
				Odds-based*	[0.75 , 1.50]	[-0.50 , 0.50]	[0.50 , 1.50]
				Bounding	[1 , 1]	[0 , 0]	[1 , 1]
				Both	[1 , 1]	[0 , 0]	[1 , 1]
4(b)	2	2	x_{2i}	MLE	0.75	0	0.74
				Odds-based*	[0.80 , 2.50]	[-1.50 , 1.50]	[-0.50 , 2.50]
				Bounding	[0.55 , 1.45]	[-1.00 , 1.00]	[0.05 , 2.00]
				Both	[0.80 , 1.45]	[-1.00 , 1.00]	[0.05 , 2.00]
5(b)	0	2	p_i	MLE	0.64	-0.07	0.64
				Odds-based	∅	∅	∅
				Bounding	[0.65 , 0.65]	[-0.10 , -0.05]	[0.60 , 0.65]
				Both	∅	∅	∅

Note: See note to Table 2 in the main text.

C.3 Confidence Sets Following Andrews and Soares (2010)

Table C.2 is analogous to Table 2 in the main text, except here we compute the confidence sets following the procedure in Andrews and Soares (2010). For a detailed description of our implementation of the inference procedure in Andrews and Soares (2010), see Appendix A.7. in Dickstein and Morales (2018). A comparison of the results in Table 2 and Table C.2 shows that the confidence sets are nearly identical; the projected end points of the sets differ only in the second decimal place.

Table C.2: Simulation Results - MLE and Confidence Intervals

Case	σ_1	σ_3	z_i	Estimator	MLE & Confidence Sets		
					α	κ_2	κ_3
1	0	0	x_{2i}	Odds-based	[1 , 1]	[0 , 0]	[1 , 1]
				Bounding	[1 , 1]	[0 , 0]	[1 , 1]
				Both	[1 , 1]	[0 , 0]	[1 , 1]
2	0	1	x_{2i}	Odds-based*	[0.92 , 1.50]	[-0.34 , 0.34]	[0.66 , 1.32]
				Bounding	[1 , 1]	[0 , 0]	[1 , 1]
				Both	[1 , 1]	[0 , 0]	[1 , 1]
3	1	0	x_{2i}	Odds-based*	[1 , 1]	[0 , 0]	[1 , 1]
				Bounding	[0.80 , 1.10]	[-0.32 , 0.32]	[0.70 , 1.30]
				Both	[1 , 1]	[0 , 0]	[1 , 1]
4	1	1	x_{2i}	Odds-based*	[0.92 , 1.50]	[-0.48 , 0.50]	[0.65 , 1.50]
				Bounding	[0.80 , 1.10]	[-0.30 , 0.30]	[0.70 , 1.30]
				Both	[0.92 , 1.10]	[-0.33 , 0.30]	[0.70 , 1.30]
5	0	1	p_i	Odds-based	∅	∅	∅
				Bounding	∅	∅	∅
				Both	∅	∅	∅

Note: See note to Table 2 in the main text.

C.4 Finite-Sample Properties of Inference Procedures

In Appendix C.4.1, we present simulation results that explore the finite-sample properties of confidence sets computed following Cox and Shi (2023). In Appendix C.4.2, we compare the inference procedures in Andrews and Soares (2010) and Cox and Shi (2023) along two dimensions: their finite-sample properties and their computational convenience.

C.4.1 Properties of Inference Procedure in Cox and Shi (2023)

We rely on the same simulation setup described in Section 6.1 and consider settings that differ in the sample size N, the value of the parameters σ_1 and σ_3, and the variable used as instrument z_i. In each setting, we draw 250 simulated samples and, for each sample, we compute a 95% confidence set using both odds-based and bounding inequalities simultaneously. Table C.3 reports the following statistics: (i) the median boundaries of the projections of the confidence sets for each parameter, in the columns labeled *Confidence Sets* (*Median*); (ii) the share of the simulated samples that yield an empty confidence set, in the column labeled $P(\hat{\Theta} = \varnothing)$; and (iii) the share of the simulated samples in which the confidence set includes the true parameter value, in the column labeled $P(\mu \in \hat{\Theta})$.

Table C.3: Simulation Results - Confidence Sets Based on Cox and Shi (2023)

Case	σ_1	σ_3	z_i	Confidence Sets (Median) α	κ_2	κ_3	$P(\hat{\Theta} = \varnothing)$	$P(\mu \in \hat{\Theta})$
				Number of observations $= 10,000$				
1	0	0	x_{2i}	$[0.98, 1.10]$	$[-0.11, 0.12]$	$[0.91, 1.15]$	0.18	0.63
2	0.1	0.1	x_{2i}	$[0.98, 1.10]$	$[-0.12, 0.13]$	$[0.90, 1.16]$	0.15	0.67
3	0.2	0.2	x_{2i}	$[0.97, 1.11]$	$[-0.14, 0.15]$	$[0.88, 1.17]$	0.09	0.74
4	0.5	0.5	x_{2i}	$[0.96, 1.20]$	$[-0.22, 0.24]$	$[0.80, 1.28]$	< 0.01	0.93
5	0.6	0.6	x_{2i}	$[0.94, 1.26]$	$[-0.28, 0.28]$	$[0.76, 1.32]$	0	0.96
6	0	1	p_i	$[0.86, 0.95]$	$[-0.13, 0.08]$	$[0.80, 0.99]$	0.32	< 0.01
7	0	1.5	p_i	$[0.80, 0.82]$	$[-0.10, 0.04]$	$[0.80, 0.87]$	0.56	0
8	0	2	p_i	–	–	–	1	0
				Number of observations $= 50,000$				
1	0	0	x_{2i}	$[0.98, 1.03]$	$[-0.05, 0.05]$	$[0.95, 1.06]$	0.09	0.83
2	0.1	0.1	x_{2i}	$[0.98, 1.04]$	$[-0.06, 0.06]$	$[0.95, 1.06]$	0.04	0.87
3	0.2	0.2	x_{2i}	$[0.98, 1.05]$	$[-0.07, 0.07]$	$[0.93, 1.08]$	0.02	0.92
4	0.5	0.5	x_{2i}	$[0.96, 1.10]$	$[-0.14, 0.14]$	$[0.86, 1.16]$	0	> 0.99
5	0.6	0.6	x_{2i}	$[0.96, 1.12]$	$[-0.18, 0.18]$	$[0.82, 1.20]$	0	> 0.99
6	0	1	p_i	$[0.87, 0.90]$	$[-0.06, 0.01]$	$[0.84, 0.91]$	0.50	0
7	0	1.5	p_i	–	–	–	1	0
8	0	2	p_i	–	–	–	1	0

Note: Results are based on 250 simulated samples and on 95% confidence sets computed following the procedure in Cox and Shi (2023). The confidence sets are computed combining the odds-based inequalities in equation (25) and the bounding inequalities in equation (26). *Confidence Sets (Median)* contain the median across samples of the projections of the confidence set on each parameter. In the last two columns, we report the fraction of the samples for which the confidence set is empty or contains the true parameter, respectively.

In cases 1 to 5 in Table C.3, the instrument z_i is equal to the variable x_{2i}, which is a valid instrument in this setting. We emphasize three patterns in the results. First, consistent with the results in Table 2, the size of the confidence sets increases in σ_1 and σ_3. E.g., for $N = 10,000$, the median of the projection on α is $[0.98, 1.10]$ when $\sigma_1 = \sigma_3 = 0$, and $[0.96, 1.20]$ when $\sigma_1 = \sigma_3 = 0.5$. Second, the size of the confidence set decreases in N: confidence sets for $N = 10,000$ are larger than for $N = 50,000$, and both are larger than the confidence sets reported in Table 2 for $N = 4,000,000$. E.g., when $\sigma_1 = \sigma_3 = 0$, the confidence set for $N = 4,000,000$ only includes the true parameter value, but the confidence sets for $N = 10,000$ and $N = 40,000$ include parameter values other than the true value. Third, as we increase N or the value of σ_1 and σ_3, the share of empty confidence sets decreases, and the share of the confidence sets that include the true parameter value increases. E.g., fixing $N = 10,000$ and comparing the results for $\sigma_1 = \sigma_3 = 0$ to those for $\sigma_1 = \sigma_3 = 0.5$, the share of empty confidence sets decreases from 18% to less than 1%, and the share of confidence sets that include the true parameter increases from 63% to 93%.

In cases 5 to 8 in Table C.3, we set the instrument z_i equal to the covariate p_i, which is an invalid instrument in this setting. We emphasize two patterns in the results. First, the share of empty confidence sets increases with the sample size N and σ_3. E.g., fixing $N = 10,000$ and comparing the results for $\sigma_3 = 1$ to those for $\sigma_3 = 1.5$, the share of empty confidence sets increases from 32% to 56%. Second, the share of confidence sets that include the true parameter value is very close to zero for the different values of N and σ_3 considered in Table C.3.

C.4.2 Comparing Andrews and Soares (2010) and Cox and Shi (2023)

To illustrate the properties of the inference procedures in Andrews and Soares (2010) and Cox and Shi (2023), we compare the results reported in the top panel of Table C.3 with the results in Table C.4. Both sets of results apply to a sample size of $N = 10,000$. We observe three patterns.

First, the median projected confidence sets are similar regardless of whether we use the procedure in Andrews and Soares (2010) or the procedure in Cox and Shi (2023). E.g., in case 1, the median confidence set for α is $[0.98, 1.10]$ when we use the procedure in Cox and Shi (2023), and $[0.98, 1.08]$ under the procedure in Andrews and Soares (2010).

Second, for low values of σ_1 and σ_2, the probability that the confidence set computed following the procedure in Andrews and Soares (2010) contains the true parameter value is closer to the nominal confidence level than the analogous probability when the confidence set is computed following the procedure in Cox and Shi (2023). E.g., for $\sigma_1 = \sigma_3 = 0$, the probability that the true parameter value is included in a 95% confidence set is 0.63 for confidence sets built following Cox and Shi (2023) and 0.78 for confidence sets built following Andrews and Soares (2010). For larger values of σ_1 and σ_3, both procedures yield confidence sets where the probability that the set contains the true parameter value is similar to the nominal confidence level. E.g., for $\sigma_1 = \sigma_3 = 0.5$, the coverage probabilities are 0.93 and 0.96 for confidence sets computed following Cox and Shi (2023) and Andrews and Soares (2010), respectively.

Table C.4: Simulation Results - Confidence Sets Based on Andrews and Soares (2010)

Case	σ_1	σ_3	z_i	Confidence Sets (Median) α	κ_2	κ_3	$P(\hat{\Theta} = \varnothing)$	$P(\mu \in \hat{\Theta})$
Number of observations $= 10,000$								
1	0	0	x_{2i}	$[0.98, 1.08]$	$[-0.10, 0.10]$	$[0.92, 1.13]$	0.04	0.78
2	0.1	0.1	x_{2i}	$[0.98, 1.09]$	$[-0.11, 0.11]$	$[0.91, 1.13]$	0.05	0.79
3	0.2	0.2	x_{2i}	$[0.97, 1.10]$	$[-0.12, 0.13]$	$[0.90, 1.16]$	0.02	0.82
4	0.5	0.5	x_{2i}	$[0.95, 1.17]$	$[-0.21, 0.22]$	$[0.81, 1.26]$	0	0.96
5	0.6	0.6	x_{2i}	$[0.94, 1.23]$	$[-0.26, 0.26]$	$[0.77, 1.31]$	0	0.99
6	0	1	p_i	$[0.87, 0.94]$	$[-0.10, 0.07]$	$[0.85, 0.98]$	0.08	0
7	0	1.5	p_i	$[0.85, 0.85]$	$[-0.04, 0.04]$	$[0.90, 0.98]$	>0.99	0
8	0	2	p_i	–	–	–	1	0

Note: Results are based on 250 simulated samples and on 95% confidence sets computed following the procedure in Andrews and Soares (2010), as described in Appendix A.7. in Dickstein and Morales (2018). The confidence sets are computed combining the odds-based inequalities in equation (25) and the bounding inequalities in equation (26). *Confidence Sets (Median)* contain the median across samples of the projections of the confidence set on each parameter. In the last two columns, we report the fraction of the samples for which the confidence set is empty or contains the true parameter, respectively.

Third, when equating the instrument z_i to the covariate p_i, which is an invalid instrument when $\sigma_3 > 0$, we observe that the procedure in Cox and Shi (2023) yields an empty confidence set in a fraction of samples that is sometimes larger and sometimes smaller than the corresponding fraction when we use the procedure in Andrews and Soares (2010). For example, in case 6, with $\sigma_3 = 1$, the probability of obtaining an empty confidence set is 0.32 when using the procedure in Cox and Shi (2023) and 0.08 when using the procedure in Andrews and Soares (2010). In contrast, in case 7, with $\sigma_3 = 1.5$, the fractions of simulated samples that are empty are 0.56 when using the procedure in Cox and Shi (2023) and greater than 0.99 when using the procedure in Andrews and Soares (2010). When $\sigma_3 = 2$, both procedures yield an empty confidence set for all simulated samples.

In Table C.5, we compare the time it takes to compute one confidence set according to the procedures in Andrews and Soares (2010) and Cox and Shi (2023). Both inference procedures require the researcher to perform a hypothesis test at each point in a grid covering the parameter space. As a result, the time it takes to compute both confidence sets depends on the number of hypothesis tests the researcher must compute as well as the time it takes to perform each test.

The number of hypothesis tests the researcher needs to compute is equal to the number of points in the grid the researcher uses. We consider grids in which all unknown parameters are allowed to take the same number of possible values. In the simulations considered in Table C.5, we compute a confidence set for the 3-dimensional parameter vector $(\alpha, \kappa_1, \kappa_2)$ and, as a result, the number of points in the grid will be equal to the cube of the number of points that we allow each parameter to take. Specifically, as indicated in the first column in Table C.5, we consider grids with a number of points that varies between $132,651$ (that is, 51^3) and $531,441$ (that is, 81^3).

The computational cost of performing a hypothesis test at a given point in the grid depends on the number of observations in the sample and on the number of moment inequalities used

Table C.5: Simulation Results - Computing Time

Number of Points	Number of Moments	Confidence Set as in CS(2023)	Confidence Set as in AS(2010)
	Number of observations $= 10,000$		
132,651	24	185 s.	945 s.
226,981	24	298 s.	1,615 s.
357,911	24	451 s.	2,550 s.
531,441	24	672 s.	3,756 s.
132,651	48	340 s.	1,884 s.
132,651	72	521 s.	2,970 s.
132,651	96	711 s.	3,795 s.
132,651	120	916 s.	4,739 s.

Note: Results are based on one simulated sample and on 95% confidence sets computed following the procedures in Cox and Shi (2023) (denoted as CS(2023)) and in Andrews and Soares (2010) (denoted as AS(2010)). The steps we follow to implement the inference procedure in Cox and Shi (2023) are described in Appendix B.7.1; the steps we follow to implement the the inference procedure in Andrews and Soares (2010) are described in Appendix A.7. in Dickstein and Morales (2018). The confidence sets are computed by combining both the odds-based inequalities in equation (25) and the bounding inequalities in equation (26). The column labeled *Number of Points* indicates the number of possible values of the parameter vector included in the grid used to compute the confidence set; that is, the number of points in the grid Θ_g according to the notation in Section B.7.1. The column labeled *Number of Moments* indicates the number of moment inequalities used in estimation.

in the estimation. This is true for the inference procedures in both Andrews and Soares (2010) and Cox and Shi (2023). Given that the estimation sample in our empirical application includes approximately 10,000 observations, the simulations reported in Table C.5 take this number as fixed and focus on exploring the effect of altering the number of moment inequalities used in estimation. In our setting, the number of moment inequalities used in the estimation depends on two elements. First, the number of possible ordered pairs of choices we can build from all elements in the choice set; in our simulation setup, this count equals six, as the choice set includes $J = 3$ elements. Second, the number of instrument functions; that is, the constant K we introduce in Section 5.2 in the main draft. The results reported in Table C.5 explore values of K between $K = 2$ and $K = 10$.

As the results reported in Table C.5 show, computing a confidence set by implementing the procedure in Andrews and Soares (2010) takes between five and six times more time than implementing the procedure in Cox and Shi (2023). This ratio is relatively constant for the cases we consider in Table C.5 and thus appears to be robust to the number of points included in the grid and the number of moment inequalities used in the estimation.

D Estimation Results: Additional Details

In Appendix D.1, we provide additional details on the implementation of the testing procedure in Vuong (1989), and present the corresponding results. In Appendix D.2, we present additional results that complement the content of Section 7 in the main draft. In Appendix D.3, we present estimates from a nested logit analysis that incorporates choices from an additional drug class.

D.1 Results From Vuong (1989) Test

To determine which of the models reported in Table 3 fits the data better, we implement tests à la Vuong (1989). We compare all possible pairs of models among those listed in Table 3. The conclusion of these tests is that the model that assumes that physicians' information sets equal the average of current out-of-pocket costs by drug-carrier dominates all other models. We report in Table D.1 the value of the test statistic for the tests that compare each model to the model that assumes that physicians' information sets equal the average of current out-of-pocket costs by the drug-carrier. That the test statistics are always positive and far from zero indicates that this is indeed the preferred model at commonly used statistical significance levels.

Table D.1: Vuong (1989) Tests

Alternative Assumption on Information Set	Test Statistic
Perfect Information	8.56
Average Current Prices By Drug-Plan Type-Carrier	3.87
Average Current Prices By Drug-Plan Type	9.39
Average Current Prices By Drug	8.85
Lagged Prices	8.55
Average Lagged Prices By Drug-Plan Type-Carrier	4.76
Average Lagged Prices By Drug-Carrier	2.37
Average Lagged Prices By Drug-Plan Type	8.79
Average Lagged Prices By Drug	8.87

Note: In the column labeled "Test Statistic," we report the value of the test statistic for Vuong (1989) tests that compare the model that assumes that physicians know average current prices by drug-carrier to models that impose the alternative informational assumption listed in the column labeled "Alternative Assumption on Information Set."

Specifically, for any two models 1 and 2, the test statistic in Table D.1 equals

$$\text{test statistic} = \frac{L_N^1 - L_N^2 - 0.5(K_1 - K_2)\log(N)}{\sqrt{N}\omega_N} \tag{D.1}$$

where N is the sample size and, for $m = \{1, 2\}$, L_N^m denotes the log-likelihood evaluated at the corresponding maximum likelihood estimate, and K_m denotes the number of parameters. The variable ω_N equals the square root of $\mathbb{V}[\log(f_{1i}/f_{2i})]$, where, for $m = \{1, 2\}$, f_{mi} denotes the log-likelihood function for observation i and model m. To implement the test, we compare the statistic in equation (D.1) to the appropriate quantile from the standard normal distribution.

D.2 Additional Results

Maximum likelihood estimates in periods with large price reductions. In Table D.2, we examine whether large price decreases, which physicians are more likely to observe and use to update their information sets, affect the maximum likelihood estimates of the price coefficient α. The negative "interaction" terms reported in the table imply that price sensitivity is greater in periods in which we expect physicians to have more information on relative prices.

Table D.2: Estimation Results - Large Price Reduction

Information Set	κ_2	κ_3	α	Interaction
Current Prices	1.40 (0.03)	-0.27 (0.05)	-0.45 (0.04)	-0.07 (0.08)
Average Current Prices By Drug-Plan Type-Carrier	1.42 (0.03)	-0.34 (0.04)	-0.83 (0.06)	-0.41 (0.11)
Average Current Prices By Drug-Carrier	1.43 (0.03)	-0.32 (0.05)	-1.15 (0.07)	-0.07 (0.12)
Average Current Prices By Drug-Plan Type	1.43 (0.03)	-0.25 (0.05)	-1.19 (0.13)	-1.15 (0.32)

Note: Columns labeled α, κ_2 and κ_3 present maximum likelihood estimates of the corresponding parameter computed following equation (27). The column labeled *Information Set* indicates the vector of observed covariates z_i used to build the log-likelihood function in equation (27). The column labeled *Interaction* contains the coefficient on patient out-of-pocket costs interacted with an indicator that the change in price from $t-1$ to t is in the lowest decile—i.e. the largest decreases in price—among all price changes for that information set. For example, under the assumption that the information set includes current prices, the cutoff for that decile is a price reduction of at least \$0.60 per day or \$18 per month. We report standard errors using a bootstrap procedure that accounts for the two-step nature of our estimation.

Predictive power of several price predictors. In Table D.3, we present the R^2 for different linear regressions of our measure of the out-of-pocket costs for every drug and medical visit in the sample on a constant and the covariate listed in the column labeled *Price Predictor*. The results show that there is strong serial correlation in out-of-pocket costs. In particular, the R^2 when using lagged prices as a covariate is 0.566 and 0.741 for the DPP-4 and SGLT2 drug classes, respectively. There is also a large dispersion in out-of-pocket costs between carriers within a plan type. For

Table D.3: R^2 in Regressions of Prices on Price Predictors

Price Predictor	DPP-4	SGLT2
Average Current Prices by Drug-Plan Type-Carrier	0.783	0.810
Average Current Prices by Drug-Carrier	0.632	0.659
Average Current Prices by Drug-Plan Type	0.200	0.248
Average Current Prices by Drug	0.034	0.020
Lagged Prices	0.566	0.741
Average Lagged Prices by Drug-Plan Type-Carrier	0.557	0.675
Average Lagged Prices by Drug-Carrier	0.482	0.579
Average Lagged Prices by Drug-Plan Type	0.080	0.168
Average Lagged Prices by Drug	0.012	0.003

Note: The R^2 values reported in the table reflect the output of a regression of patient price on a constant and the alternative price predictors listed in the column *Price Predictor*.

example, for the case of DPP-4 drugs, the R^2 decreases from 0.783 to 0.200 when using average current prices by drug and plan type instead of average current prices by drug, plan type, and carrier. The results in Table D.3 also reveal that drug-specific average out-of-pocket costs—that is, an average price that ignores information on plan types and carriers—are generally weak predictors of the out-of-pocket costs. For example, the R^2 when using average current prices by drug as the regression covariate is 0.034 and 0.020 for the DPP-4 and SGLT2 drugs, respectively.

Moment inequality estimates for patients with stable insurance plan enrollment. According to equation (4) in the main text, the vector of idiosyncratic preferences ε_i is independent of all other elements of the physician's information set, as included in $\mathcal{W}_i$. In particular, this assumption implies ε_i is independent of the physician's expected prices. However, because these expected prices can incorporate patient input, the statistical independence of the vector of idiosyncratic preferences ε_i and the vector of expected prices will be violated if ε_i is not independent of the patient's expected prices. For example, anticipating the need for consumption of a particular drug j, patients can choose an insurance plan they know offers generous coverage for that drug—that is, for the drug j for which their idiosyncratic preference shock ε_{ij} is the highest. This insurance plan choice pattern would generate a negative correlation between drug j's expected price and ε_{ij}, the idiosyncratic preference shock. This negative correlation would in turn generate a potential bias in both the maximum likelihood and the moment inequality estimates of the preference parameters.

To investigate the importance of the potential correlation between the vector of idiosyncratic preferences and expected drug prices that may arise from the endogenous choice of insurance plan, we re-estimate our model on a subsample that includes only observations for patients who are less likely to have switched plans due to drug coverage generosity. In particular, we exclude two sets of patient visits from our estimation sample. First, for those patients who we observe switching insurers during the sample period, we exclude from the sample their visits in all periods after switching insurers. Second, since we cannot determine whether patients switched insurers

Table D.4: Estimation Results for Selected Patients with Stable Insurer Affiliation

Price Predictor	κ_2	κ_3	α
Average Current Prices By Drug	[1.40, 1.95]	[-0.40 , 0.70]	[-5.80 ,-1.10]
Average Lagged Prices By Drug-Carrier	[1.55, 1.65]	[-0.10 , 0.05]	[-1.50 ,-1.35]
Average Lagged Prices By Drug-Plan Type	[1.50, 2.10]	[-0.40 , 0.45]	[-5.80 ,-1.20]
Average Lagged Prices By Drug	[1.40, 2.20]	[-0.45 , 1.00]	[-6.60 ,-1.10]

Note: Columns labeled α, κ_2, and κ_3 present projected 95% confidence sets computed using the moment inequalities described in Section 5 and the inference procedure in Cox and Shi (2023). The column labeled *Price Predictor* indicates the vector of observed covariates z_i that we use as instruments in our moment inequalities.

immediately before 2011, the first sample year, we also exclude all visits for patients who filled a DPP-4 drug prescription in 2011. Together, these two sample restrictions eliminate 41% of the observations in the original estimation sample. The remaining sample includes 5,072 office visits.

In Table D.4, we report the moment inequality estimates we obtain when using this restricted sample. Consistent with the results presented in Table 5 in the main text, we again fail to reject the null hypotheses that physicians know average current prices by drug and average lagged prices by drug and by drug-plan type. In addition, we also fail to reject the null hypothesis that physicians form their price expectations using information on the average lagged prices by drug and carrier.

Comparing the results in tables 5 and D.4, we observe that, for the three common price predictors, the projected confidence sets are similar regardless of whether we compute them using the full estimation sample or the restricted sample that excludes observations for patients whose idiosyncratic preferences are most likely to be correlated with their expected prices. The main difference between these confidence sets is that the sets computed using the restricted sample are slightly larger. As shown in the simulation results discussed in Appendix C.4, this difference may be explained by the smaller sample size we use to compute the confidence sets in Table D.4 relative to the sample size we use to compute the confidence sets reported in Table 5.

The projected confidence sets for the additional price predictor that we fail to reject in this restricted sample—the average lagged price by drug and carrier—are much tighter than the projected confidence sets for the other three price predictors reported in Table D.4. A possible explanation for this result is that, in fact, the average lagged prices by drug and carrier may not belong to the physician's information set (as we find in Section 7.2 when we use the full sample in estimation), but we may not be able to detect this when we estimate our moment inequalities on the restricted sample. More generally, our finding could be an example of the general feature that confidence sets for the true parameter may exhibit spurious precision when the identified set is empty; see Andrews and Kwon (2024) and Kaido and Molinari (2024) for details.

Table D.5: Estimation Results by Physician Group

Set of physicians	Price Predictor	α	κ_2	κ_3
Endocrinologist	Average Current Prices By Drug-Carrier-Plan Type	[-0.60 ,-0.40]	[1.25, 1.45]	[-0.35 , -0.10]
Primary Care Physicians	Average Lagged Prices By Drug-Plan Type	[-5.90 ,-1.40]	[1.40, 1.90]	[-0.80 , 0.10]
Graduated before 1996	Average Lagged Prices By Drug-Plan Type	[-3.85 ,-2.50]	[1.45, 1.60]	[-0.45 , -0.15]
Graduated after 1996*	Average Lagged Prices By Drug-Plan Type	[-7.00 ,-1.00]	[1.35, 2.50]	[-1.00 , 0.60]
Female*	Average Lagged Prices By Drug-Plan Type	[-6.75 ,-1.05]	[1.35, 2.25]	[-1.00 , -0.20]
Male	Average Lagged Prices By Drug-Plan Type	[-3.05 ,-2.40]	[1.60, 1.75]	[-0.40 , -0.05]
Lower tier medical school	Average Current Prices By Drug	[-4.00 ,-1.30]	[1.35, 1.75]	[-0.65 , 0.30]
Top tier medical school*	Average Current Prices By Drug-Plan Type	[-7.00 ,-2.35]	[1.15, 2.50]	[-1.00 , 0.75]
High plan type HHI	Average Lagged Prices By Drug-Plan Type	[-5.85 ,-1.50]	[1.30, 1.95]	[-0.85 , 0.20]
Low plan type HHI	Average Lagged Prices By Drug	[-6.60 ,-2.65]	[1.60, 1.95]	[-0.50 , 0.10]

Note: Columns labeled α, κ_2 and κ_3 present projected 95% confidence sets computed using the moment inequalities described in Section 5.2 and the inference procedure in Cox and Shi (2023). The confidence sets are computed by testing points in a 3-dimensional grid whose sides are $[-7.00, -0.05]$ (for α), $[1, 2.5]$ (for κ_2) and $[-1, 1]$ (for κ_3). We mark cases with an asterisk when the confidence set includes points outside the grid. The minimum distance between any two points in the grid is 0.05.

Heterogeneity in physician information and preferences. In Table D.5, we report projected confidence sets for parameters κ_2, κ_3 and α for each subsample of physicians we consider in Section 7.3 in the main text. To compute these confidence sets, we use as instruments for each group of physicians the most detailed price predictor for which we obtain a non-empty 95% confidence set; this is the predictor reported in Table D.5 in the column labeled *Information Set.*

We find that the main source of heterogeneity in preferences appears when comparing between physician specialties. For our estimate for α, the physician's price sensitivity, we find that primary care physicians are significantly more sensitive to price than endocrinologists. Conversely, the projected confidence sets for κ_2 and κ_3 show more similarity and overlap between the two specialties.

D.3 Estimation Results for Additional Classes

In this section, we extend our analysis to include an additional class of dual-therapy drugs. In particular, in addition to the DPP-4 inhibitors considered in our main analysis, we add information on SGLT2 inhibitors. For the SGLT2 class, Invokana is the first drug in the class, entering in 2013, and Farxiga and Jardiance enter in 2014. Our sample for this analysis thus includes only the years 2013-2016. In our data, we observe $9,356$ visits in which a drug in the DPP-4 and SGLT2 inhibitor classes is prescribed. That count drops to $7,538$ when we require that lagged prices be available.

Table D.6: Nested Logit Estimates

Information Set	κ_2	κ_3	κ_4	κ_5	κ_6	α	λ	N
Current Prices	1.41 (0.17)	0.02 (0.05)	-0.12 (0.06)	1.15 (0.19)	-0.41 (0.06)	-0.20 (0.02)	1.00 (0.12)	9356
Average Current Prices By Drug-Plan Type-Carrier	1.44 (0.17)	-0.04 (0.05)	0.01 (0.07)	1.21 (0.18)	-0.38 (0.06)	-0.49 (0.04)	1.00 (0.11)	9356
Average Current Prices By Drug-Carrier	1.44 (0.17)	-0.09 (0.05)	0.08 (0.07)	1.23 (0.19)	-0.35 (0.06)	-0.63 (0.05)	1.00 (0.12)	9356
Average Current Prices By Drug-Plan Type	1.39 (0.18)	0.04 (0.05)	-0.16 (0.07)	1.13 (0.20)	-0.42 (0.06)	-0.07 (0.06)	1.00 (0.13)	9356
Lagged Prices	1.41 (0.22)	0.13 (0.06)	-0.03 (0.07)	1.30 (0.25)	-0.34 (0.07)	-0.24 (0.03)	1.00 (0.16)	7538
Average Lagged Prices By Drug-Plan Type-Carrier	1.28 (0.18)	0.07 (0.05)	0.05 (0.06)	1.22 (0.20)	-0.32 (0.06)	-0.50 (0.06)	0.89 (0.12)	7538
Average Lagged Prices By Drug-Carrier	0.84 (0.14)	0.02 (0.03)	-0.00 (0.05)	0.73 (0.16)	-0.30 (0.04)	-0.49 (0.07)	0.56 (0.09)	7538
Average Lagged Prices By Drug-Plan Type	1.43 (0.25)	0.17 (0.06)	0.04 (0.07)	1.35 (0.27)	-0.29 (0.07)	-0.30 (0.07)	1.00 (0.18)	7538

Note: We use data on visits in which physicians prescribe a drug in the DPP-4 or SGLT2 inhibitor classes in the years 2013-2016. When using lagged information sets, we lose observations for years in which we do not observe a lagged price in the SGLT2 class. In the specification, we include (a) fixed effects for the drugs available in the two classes, which we denote κ, and (b) the sensitivity of the drug choice to the patient's out-of-pocket cost, denoted α. We normalize the fixed effect κ_1, corresponding to the effect for one specific DPP-4 inhibitor, to zero. If $\lambda = 1$, the nested logit model is equivalent to a multinomial logit model.

We compute maximum likelihood estimates of a nested logit model across the DPP-4 and SGLT2 inhibitor classes, modeling each class as a separate nest. We report the results in Table D.6, with each row in the table including the parameter estimates we find when placing different assumptions on physicians' information sets. As described in Section 7.1, we implement a two-step estimation procedure. First, we run a regression of realized prices on the variable we assume physicians use to form their price expectations. Second, we use the predicted price from this regression as our expected price measure to compute the second-stage nested logit maximum likelihood estimates.

Two key results appear in Table D.6. First, for most specifications, we find $\lambda = 1$ in our sample, which implies that idiosyncratic preferences are independent between drugs within a nest. In this case, the nested logit model is equivalent to a multinomial logit model. For the information sets of average lagged prices by drug-plan type-carrier and drug-carrier, we find $\lambda = 0.89$ or $\lambda = 0.56$, respectively. These estimates are robust to multiple starting values for the parameter search. The simulation results in Appendix E.1.4 suggest a possible explanation for why we find that the maximum likelihood estimate of λ equals one or is close to one in most cases we consider in Table D.6. As the results for cases 3 and 5 in Table E.1 illustrate, the maximum likelihood estimate of λ is biased upwards whenever the researcher assumes that the physician's information set is either too large or too small.

Second, the pattern we observe in the parameter estimates of the multinomial logit model for DPP-4 inhibitors—namely, that price coefficients are sensitive to the specification of expected prices—appears in the nested logit parameter estimates as well. In particular, as we vary physician information sets, starting from the assumption of perfect information to price expectations based on more aggregate variables, we find distinct price coefficients that tend to grow more negative, implying a more elastic demand. For example, when moving from the assumption that physicians know current prices to the assumption that they instead only know average current prices by drug and plan type, the estimate of α decreases from -0.20 to -0.63.[1]

In Section 9, we show how to specify moment inequalities to identify the parameters of a nested logit model. We do not implement those inequalities in the data that combine the SGLT2 and DPP-4 drug classes due to the large number of parameters to estimate. As illustrated in Table D.6, the nested logit model for SGLT2 and DPP-4 inhibitors incorporates seven parameters. Computing a confidence set over a seven-dimensional parameter space is computationally costly, as it requires using a very large grid Θ_g, following the notation in Appendix B.7.1. To avoid this computational cost, and because the diabetes guidelines suggest that the choice of treatment across dual-therapy classes is mostly determined by therapeutic motives, we focus only on the choice of treatment among DPP-4 inhibitors when computing the moment inequality estimates in sections 7.2 and 7.3.

D.4 Bounding Consumer Surplus

In this section, we first derive a lower bound on the expected consumer surplus implied by a distribution of physicians' information sets, $\mathcal{W}_i$. We use this result to compute an upper bound on the expected welfare gain from an intervention that provides physicians with perfect price information.

Given equations (1) and (2), the consumer surplus for an agent i with information set $\mathcal{J}_i$ is

$$\mathbb{S}_i = \sum_{j=1}^{J} \left\{ \mathbb{1}\{\mathbb{E}[\mathcal{U}_{ij}|\mathcal{J}_i] \geqslant \max_{j'=1,\ldots,J} \mathbb{E}[\mathcal{U}_{ij'}|\mathcal{J}_i]\}\mathcal{U}_{ij} \right\},$$

which, after adding and subtracting $\mathbb{E}[\mathcal{U}_{ij}|\mathcal{J}_i]$, we can rewrite as

$$\mathbb{S}_i = \sum_{j=1}^{J} \left\{ \mathbb{1}\{\mathbb{E}[\mathcal{U}_{ij}|\mathcal{J}_i] \geqslant \max_{j'=1,\ldots,J} \mathbb{E}[\mathcal{U}_{ij'}|\mathcal{J}_i]\}(\mathbb{E}[\mathcal{U}_{ij}|\mathcal{J}_i] + \mathcal{U}_{ij} - \mathbb{E}[\mathcal{U}_{ij}|\mathcal{J}_i]) \right\}.$$

To simplify notation, define $p^e_{ij} \equiv \mathbb{E}[p_{ij}|\mathcal{W}_i]$ and $p^e_i = \{p^e_{ij}\}_{j=1}^J$. Then, given equations (3), (5), (23) and (24), we can write the consumer surplus for an agent i with information set $\mathcal{J}_i$ in monetary

[1]In Table D.6, we do not report results that assume that every physician's information set equals either the average current prices by drug or the average lagged prices by drug. As competition in the SGLT2 class begins only in 2014, the number of distinct values that these potential information sets take in the sample for SGLT2 drugs is very limited.

units as:

$$S_i = \frac{1}{\alpha}\sum_{j=1}^{J}\big\{\mathbb{1}\{\kappa_j + \alpha p^e_{ij} + \varepsilon_{ij} \geqslant \max_{j'=1,\dots,J}\{\kappa_{j'} + \alpha p^e_{ij'} + \varepsilon_{ij'}\}\}(\kappa_j + \alpha p^e_{ij} + \varepsilon_{ij})\big\}$$
$$+ \frac{1}{\alpha}\sum_{j=1}^{J}\big\{\mathbb{1}\{\kappa_j + \alpha p^e_{ij} + \varepsilon_{ij} \geqslant \max_{j'=1,\dots,J}\{\kappa_{j'} + \alpha p^e_{ij'} + \varepsilon_{ij'}\}\}(\alpha p_{ij} - \alpha p^e_{ij})\big\}. \quad \text{(D.2)}$$

As the value of S_i depends only on the vector $(p^e_i, \varepsilon_i, p_i)$, we can write $S_i = S(p^e_i, \varepsilon_i, p_i)$. Assume here that: (a) ε_{ij} is independent of (p^e_i, p_i) for any alternative j; (b) ε_{ij} and $\varepsilon_{ij'}$ are independent for any two alternatives j and j'; and, (c) ε_{ij} follows a type I extreme value distribution for any alternative j.[2] Then, taking the expectation of S_i conditional on (p^e_i, p_i), we have:

$$S(p^e_i, p_i) = \frac{1}{\alpha}\log(\sum_{j=1}^{J}\exp(\kappa_j + \alpha p^e_{ij})) + \sum_{j=1}^{J}\mathcal{P}(d_{ij} = 1|p^e_i)(p_{ij} - p^e_{ij}) + C, \quad \text{(D.3)}$$

where C is a constant and $\mathcal{P}(d_{ij} = 1|p^e_i)$ equals the probability in equation (6), written as a function of the vector of expected prices p^e_i. Equation (D.3) matches equation (1) in Train (2015).

We can then take the expectation of $S(p^e_i, p_i)$ conditional on p^e_i to obtain the expression:

$$S(p^e_i) = \frac{1}{\alpha}\log(\sum_{j=1}^{J}\exp(\kappa_j + \alpha p^e_{ij})) + C, \quad \text{(D.4)}$$

Here, we abuse notation by using the same symbol S to denote the random variable in equation (D.2) as well as the functions in equations (D.3) and (D.4).

The function $S(x)$ is jointly convex in $x \in \mathbb{R}^J$. Thus, Jensen's inequality implies $S(\mathbb{E}[p^e_i|z_i]) \leqslant \mathbb{E}[S(p^e_i)|z_i]$ for any vector z_i. In addition, if $z_i \subseteq \mathcal{W}_i$, the Law of Iterated Expectations implies the following equality $\mathbb{E}[p^e_i|z_i] = \mathbb{E}[p_i|z_i]$. As a result, if we define $\hat{p}^e_i = \{\hat{p}^e_{ij}\}^J_{j=1}$ with $\hat{p}^e_{ij} = \mathbb{E}[p_{ij}|z_i]$, we can conclude that, if $z_i \subseteq \mathcal{W}_i$, we have

$$\mathbb{E}[S(p^e_i)|z_i] \geqslant S(\hat{p}^e_i). \quad \text{(D.5)}$$

Using the Law of Iterated Expectations again, we can further conclude that

$$\mathbb{E}[S(p^e_i)] \geqslant \mathbb{E}[S(\hat{p}^e_i)]. \quad \text{(D.6)}$$

According to this expression, we can compute a *lower* bound on the average consumer welfare in a population of interest by computing the consumer welfare of each observation i using $\hat{p}^e_i$, a proxy for i's true price expectation that relies on a subset z_i of i's true information set, $\mathcal{W}_i$.

Using the lower bound on expected consumer welfare in equation (D.6), we can compute an

[2] Relative to the assumptions in equations (4), (5), and (23), the assumptions we impose here add that ε_i and p_i are not just mean independent but fully independent.

upper bound on the expected change in consumer welfare that results when we provide all physicians in the population of interest with perfect information on price:

$$\Delta\bar{S}(\mu) = \mathbb{E}[S(p_i) - S(\hat{p}_i^e)], \tag{D.7}$$

where we write this expression as an explicit function of the true parameter value, μ, to emphasize that our measure of the expected change in consumer welfare from the information intervention depends on the value of the physician's preference parameters.

To compute a confidence set for $\Delta\bar{S}(\mu)$, we use information on the confidence set for μ computed using our moment inequalities and the assumption that $z_i \subseteq \mathcal{W}_i$ where, as a reminder, z_i denotes the covariate vector used to compute the price prediction $\hat{p}_i^e$ entering equation (D.7). Formally, we compute a confidence interval for an upper bound on the expected change in consumer welfare that results from providing physicians in the population of interest with perfect price information as

$$[\min_{\theta\in\hat{\Theta}} \Delta\bar{S}(\theta), \max_{\theta\in\hat{\Theta}} \Delta\bar{S}(\theta)], \tag{D.8}$$

where $\hat{\Theta}$ is the moment inequality 95% confidence set for μ computed under the assumption that z_i belongs to the agent's true information set, $\mathcal{W}_i$.

When computing the bounds in equation (D.8) in our sample, we approximate the expectation in equation (D.7) using an average over all observations in the sample of interest. More specifically, we compute the sample analogue of the bounds in equation (D.8) as

$$[\min_{\theta\in\hat{\Theta}} \Delta\hat{\bar{S}}(\theta), \max_{\theta\in\hat{\Theta}} \Delta\hat{\bar{S}}(\theta)], \tag{D.9}$$

with the function $\hat{\bar{S}}(\theta)$ evaluated at $\theta = \mu$ defined as

$$\Delta\hat{\bar{S}}(\mu) = \frac{1}{\alpha}\frac{1}{n}\sum_{i=1}^{n}\Big\{\log(\sum_{j=1}^{J}\exp(\kappa_j + \alpha p_{ij})) - \log(\sum_{j=1}^{J}\exp(\kappa_j + \alpha\hat{\mathbb{E}}[p_{ij}|z_{ij}]))\Big\} \tag{D.10}$$

with $\hat{\mathbb{E}}[p_{ij}|z_i]$ the predicted value of p_{ij} computed using information on z_{ij} alone. When computing the bounds on the consumer surplus gain for endocrinologists reported in Table 7, we use as z_{ij} the value of drug j's average current price by drug, carrier, and plan type, and the confidence set described in the first row of Table D.5. When computing the bounds on the consumer surplus gain for primary care physicians reported in Table 7, we use as z_{ij} the value of drug j's average lagged price by drug and plan type, and the confidence set described in the second row of Table D.5. Thus, in practice, for each physician type, we specify the vector z_i in equation (D.10) as the most informative price predictor that we fail to reject using the tests reported in Table 6.

D.5 Observed vs. Predicted Market Shares

In Section 8, we describe our calculation of an upper bound on the change in consumer surplus after an information intervention. For this calculation, we must specify a price predictor that we assume belongs to physicians' information sets in the pre-intervention scenario. Here, we compare the observed market shares to the model-implied market shares computed under the assumption that this price predictor equals physicians' information sets.

In panels A and B in Table D.7, we report the predicted market shares and observed market shares for endocrinologists and primary care physicians. We have three products and two samples, and thus we can compare six ranges of predicted market shares to their observed values. We find that, for five out of six comparisons, the observed share is contained in our range of predicted probabilities. An exception is the prediction for Tradjenta in the primary care physician sample: the predicted share ranges from 6.12% to 12.96%, and the observed value is 13.19%. In Panel C, we report similar predicted and observed market shares using the full sample and equating all physicians' information sets to each of the price predictors listed in Table 5. In this case, for seven out of the nine comparisons, the observed share is contained in the corresponding confidence set.

Table D.7: Predicted Probability vs. Observed Shares, by Price Predictor and Drug

Price Predictor	Drug	Conf. Interval	Obs. Share
Panel A: Endocrinologists			
	Janumet	[16.86%, 19.85%]	19.01%
Avg. Current Prices by Drug-Carrier-Plan Type	Januvia	[65.85%, 69.64%]	66.17%
	Tradjenta	[12.60%, 14.90%]	14.82%
Panel B: Primary Care Physicians			
	Janumet	[15.57%, 21.56%]	17.67%
Avg. Lagged Prices by Drug-Plan Type	Januvia	[67.65%, 76.70%]	69.14%
	Tradjenta	[6.12%, 12.96%]	13.19%
Panel C: All Physicians			
	Janumet	[15.18%, 20.24%]	17.88%
Avg. Current Prices by Drug	Januvia	[64.77%, 74.58%]	68.68%
	Tradjenta	[8.29%, 17.62%]	13.44%
	Janumet	[17.09%, 20.10%]	17.88%
Avg. Lagged Prices by Drug-Plan Type	Januvia	[70.27%, 73.52%]	68.68%
	Tradjenta	[8.19%, 11.16%]	13.44%
	Janumet	[15.25%, 20.02%]	17.88%
Avg. Lagged Prices by Drug	Januvia	[67.77%, 75.23%]	68.68%
	Tradjenta	[7.29%, 14.05%]	13.44%

Notes: In the column *Obs. Share*, we report the observed aggregate market share for each of the three DPP-4 drugs in the sample. We do so for the subset of observations indicated in the corresponding panel; that is, Panel A reports market shares using information on visits to endocrinologists, Panel B uses information on visits to primary care physicians, and Panel C uses information on all visits in the sample. In the column *Conf. Interval*, we report the corresponding model-implied confidence intervals. To compute these, we assume the physician information set coincides with the predictor listed in the column *Price Predictor*, and set the parameters to the values included in the corresponding confidence set. That is, we use the confidence sets described in the first two rows in Table D.5 to compute the predicted shares reported in panels A and B, respectively, and those described in Table 5 to compute the predicted shares reported in Panel C.

E Alternative Discrete Choice Models

In Section 4, we present moment inequalities that partially identify the parameters of the multinomial logit model introduced in Section 3. Here, we show how to derive moment inequalities for alternative discrete choice models. In Appendix E.1, we introduce inequalities that partially identify the parameters of a nested logit model. In Appendix E.2, we introduce inequalities that partially identify the parameters of a multinomial probit model in which the normally distributed idiosyncratic components of utility may be correlated across choices for a given agent.

E.1 Moment Inequalities for Nested Logit Models

In Appendix E.1.1, we describe the nested logit model. Appendix E.1.2 describes a new set of moment inequalities, and Appendix E.1.3 shows formally that these inequalities can be used to partially identify the parameters of the nested logit model. Appendix E.1.4 presents simulation results that illustrate certain properties of these inequalities.

E.1.1 Model

We partition the set of J drugs in the physician's choice set into $k = 1, \ldots, K$ nests, and let $J_k \subset J$ denote the set of drugs included in nest k. We assume physicians choose drugs at each visit according to the restrictions imposed in equations (1) to (5) and (23), except now we impose a distributional assumption distinct from the assumption in equation (4). Here, we impose the vector of shocks ε_i follows the distribution

$$F_\varepsilon(\varepsilon_i|\mathcal{W}_i) = F_\varepsilon(\varepsilon_i) = \exp(-\sum_{k=1}^{K}(\sum_{j \in J_k} \exp(-\varepsilon_{ij}/\lambda))^{\lambda}), \qquad \text{for } \lambda \in (0,1], \tag{E.1}$$

where $F_\varepsilon(\cdot)$ denotes the cumulative distribution function of ε_i.

We define a dummy variable d_{ik} that equals one if physician i prescribes a drug that belongs to nest k, and zero otherwise; i.e., $d_{ik} = \sum_{j \in J_k} d_{ij}$. We also define the transformed utility function $\check{u}_{ij} \equiv u_{ij}/\lambda$. Thus, given equation (23), $\check{u}_{ij} = \check{\kappa}_j + \check{\alpha} p_{ij}$ with $\check{\kappa}_j \equiv \kappa_j/\lambda$ for $j = 1, \ldots, J$ and $\check{\alpha} \equiv \alpha/\lambda$. Equations (1), (2), (3), (5), (23), and (E.1) then imply that we can write the probability that drug $j \in J_k$ is prescribed, given $\mathcal{W}_i$ and that a drug in nest k has been prescribed, as

$$\mathcal{P}(d_{ij} = 1|\mathcal{W}_i, d_{ik} = 1) = \frac{\exp(\check{\kappa}_j + \check{\alpha}\mathbb{E}[p_{ij}|\mathcal{W}_i])}{\sum_{j' \in J_k} \exp(\check{\kappa}_{j'} + \check{\alpha}\mathbb{E}[p_{ij'}|\mathcal{W}_i])} \qquad \text{for all } j \in J_k. \tag{E.2}$$

We can write the probability that a drug in a nest k is prescribed given $\mathcal{W}_i$ as

$$\mathcal{P}(d_{ik} = 1|\mathcal{W}_i) = \frac{\exp(\lambda \ln(\sum_{j \in J_k} \exp(\check{\kappa}_j + \check{\alpha}\mathbb{E}[p_{ij}|\mathcal{W}_i])))}{\sum_{k'=1}^{K} \exp(\lambda \ln(\sum_{j \in J_{k'}} \exp(\check{\kappa}_j + \check{\alpha}\mathbb{E}[p_{ij}|\mathcal{W}_i])))} \qquad \text{for } k = 1, \ldots, K. \tag{E.3}$$

Finally, using equations (E.2) and (E.3), we write the probability that a drug j is prescribed as

$$\mathcal{P}(d_{ij} = 1|\mathcal{W}_i) = \mathcal{P}(d_{ij} = 1|\mathcal{W}_i, d_{ik} = 1)\mathcal{P}(d_{ik} = 1|\mathcal{W}_i). \tag{E.4}$$

Given a normalization $\breve{\kappa}_1 = 0$, the goal of estimation is to recover the value of the parameters $\{\breve{\kappa}_j\}_{j=2}^J$, $\breve{\alpha}$, and λ, and to learn about the content of the information sets $\{\mathcal{W}_i\}_{i=1}^N$. We use $\gamma \equiv (\gamma_\lambda, \gamma_{\breve{\alpha}}, \gamma_{\breve{\kappa}_2}, \ldots, \gamma_{\breve{\kappa}_J})$ to denote the unknown parameter vector, Γ to denote the parameter space, and $\gamma^* \equiv (\lambda, \breve{\alpha}, \breve{\kappa}_2, \ldots, \breve{\kappa}_J)$ to denote the true parameter value, as determined by equation (E.4).

E.1.2 Moment Inequalities

In this section, we show how to partially identify the vector γ^*. We use three types of moment inequalities: odds-based, bounding, and nested logit inequalities. The first two types are analogous to those used for multinomial logit models; the third type is specific to the nested logit model.

Odds-based Moment Inequality. For any two drugs j and j' in the nest k, any value of z_i in its support $\mathcal{Z}$, and any $\gamma \in \Gamma$, we define the following odds-based moment inequality

$$\mathbb{m}^o_{jj}(z_i, \gamma) \geqslant 0 \tag{E.5a}$$

with

$$\mathbb{m}^o_{jj'}(z_i, \gamma) \equiv \mathbb{E}[d_{ij} \exp(-(\gamma_{\breve{\kappa}_j} - \gamma_{\breve{\kappa}_{j'}} + \gamma_{\breve{\alpha}}\Delta p_{ijj'})) - d_{ij'}|z_i, d_{ik} = 1]. \tag{E.5b}$$

We also define the sets

$$\Gamma^o_{0,k} \equiv \{\gamma \in \Gamma \colon \mathbb{m}^o_{jj'}(z, \gamma) \geqslant 0 \text{ for all } z \in \mathcal{Z},\ j \in J_k, \text{ and } j' \in J_k\}, \tag{E.6}$$

and

$$\Gamma^o_0 \equiv \Gamma^o_{0,1} \cap \cdots \cap \Gamma^o_{0,K}. \tag{E.7}$$

Theorem 1 establishes a sufficient condition for the true parameter value γ^* to belong to Γ^o_0.

Theorem 1 *Let $\gamma^* \equiv (\lambda, \breve{\alpha}, \breve{\kappa}_2, \ldots, \breve{\kappa}_J)$ be defined by equation (E.4) and the normalization $\breve{\kappa}_1 = 0$. If $z_i \subseteq \mathcal{W}_i$ for every visit i, then $\gamma^* \in \Gamma^o_0$.*

Given the similarity between the probability in equation (6) and that in equation (E.2), the formal proof of Theorem 1 is analogous to the proof of Theorem 1 included in Appendix B.1.

Bounding Moment Inequality. For any two drugs j and j' in nest k, any value of $z_i \in \mathcal{Z}$, any $\gamma \in \Gamma$, and any function $e_{jj'} \colon \mathcal{Z} \times \Gamma \to \mathbb{R}$, we define the following bounding moment inequality

$$\mathbb{m}^b_{jj'}(z_i, \gamma, e_{jj'}(\cdot)) \leqslant 0 \tag{E.8a}$$

with

$$\mathbb{m}^b_{jj'}(z_i, \gamma, e_{jj'}(\cdot)) \equiv$$
$$\mathbb{E}[d_{ij} \exp(-e_{jj'}(z_i, \gamma))(1 + e_{jj'}(z_i, \gamma) - (\gamma_{\breve{\kappa}_j} - \gamma_{\breve{\kappa}_{j'}} + \gamma_{\breve{\alpha}} \Delta p_{ijj'})) - d_{ij'} | z_i, d_{ik} = 1]. \tag{E.8b}$$

The moment $\mathbb{m}^b_{jj'}(\cdot)$ depends on $e_{jj'}(z_i, \gamma)$, which is a deterministic function of the observed vector z_i and the unknown parameter vector γ, and may vary by pair of drugs j and j'. Defining e_k as the set of functions that includes $e_{jj'}(\cdot)$ for all drug pairs in nest k (i.e., $e_k = \{e_{jj'}(\cdot)\}_{j \in J_k, j' \in J_k}$), we define the set

$$\Gamma^b_{0,k}(e_k) \equiv \{\gamma \in \Gamma \colon \mathbb{m}^b_{jj'}(z, \gamma, e_{jj'}(\cdot)) \leqslant 0 \text{ for all } z \in \mathcal{Z},\ j \in J_k, \text{ and } j' \in J_k\}. \tag{E.9}$$

and

$$\Gamma^b_0(e) \equiv \Gamma^o_{0,1}(e_1) \cap \cdots \cap \Gamma^o_{0,K}(e_K). \tag{E.10}$$

where e is the set of functions including $e_{jj'}(\cdot)$ for all drug pairs; i.e., $e = \{e_{jj'}(\cdot)\}^{J,J}_{j=1,j'=1}$.

Regardless of the set of functions e, the following theorem establishes a sufficient condition for the true parameter value γ^* to belong to $\Gamma^b_0(e)$

Theorem 2 *Let $\gamma^* \equiv (\lambda, \breve{\alpha}, \breve{\kappa}_2, \ldots, \breve{\kappa}_J)$ be defined by equation (E.4) and $\breve{\kappa}_1 = 0$. If $z_i \subseteq \mathcal{W}_i$ for all i, then $\gamma^* \in \Gamma^b_0(e)$ for any set e of functions $e_{jj'} \colon \mathcal{Z} \times \Gamma \to \mathbb{R}$.*

Given the similarity between the probability in equation (6) and that in equation (E.2), the formal proof of Theorem 2 is analogous to the proof of Theorem 2 included in Appendix B.2, which itself reproduces results in Porcher et al. (2025).

Nested Logit Moment Inequality. To simplify the notation, define for any $k = 1, \ldots, K$ the set of prices $p_{ik} \equiv \{p_{ij}\}_{j \in J_k}$ and a set of functions $s_k(z_i) \equiv \{s_j(z_i)\}_{j \in J_k}$. Then, for any two nests k and k', any value of $z_i \in \mathcal{Z}$, any $\gamma \in \Gamma$, and any function $s_{k'} \colon \mathcal{Z} \to \mathbb{R}^{J_{k'}}$ we define the moment inequality

$$\mathbb{m}^n_{kk'}(z_i, \gamma, s_{k'}(\cdot)) \geqslant 0 \tag{E.11a}$$

with

$$\mathbb{m}^n_{kk'}(z_i, \gamma, s_{k'}(\cdot)) \equiv \mathbb{E}[d_{ik'} \exp(I_k(p_{ik}, \gamma) - \underline{I}_{k'}(p_{ik'}, \gamma, s_{k'}(z_i))) - d_{ik} | z_i] \tag{E.11b}$$

such that

$$I_k(p_{ik}, \gamma) = \gamma_\lambda \ln(\sum_{j \in J_k} \exp(\gamma_{\breve{\kappa}_j} + \gamma_{\breve{\alpha}} p_{ij})) \tag{E.11c}$$

and

$$\underline{I}_{k'}(p_{ik'}, \gamma, s_{k'}(z_i)) =$$
$$\gamma_\lambda \ln(\sum_{j \in J_{k'}} \exp(\gamma_{\breve{\kappa}_j} + \gamma_{\breve{\alpha}} s_j(z_i))) + \gamma_\lambda \gamma_{\breve{\alpha}} \sum_{j \in J_{k'}} \frac{\exp(\gamma_{\breve{\kappa}_j} + \gamma_{\breve{\alpha}} s_j(z_i))}{\sum_{j' \in J_{k'}} \exp(\gamma_{\breve{\kappa}_{j'}} + \gamma_{\breve{\alpha}} s_{j'}(z_i))} (p_{ij} - s_j(z_i)). \tag{E.11d}$$

The moment $\mathfrak{m}^n_{kk'}(\cdot)$ depends on $s_{k'}(z_i)$, which is a deterministic function of z_i and varies by nest. Defining s as the set of functions including $s_k(\cdot)$ for all nests (i.e., $s = \{s_j(\cdot)\}_{j=1}^J$), we define

$$\Gamma_0^n(s) \equiv \{\gamma \in \Gamma \colon \mathfrak{m}^n_{kk'}(z, \gamma, s_{k'}(\cdot)) \geqslant 0 \text{ for all } z \in \mathcal{Z},\ k = 1, \ldots, K, \text{ and } k' = 1, \ldots, K\}. \tag{E.12}$$

Regardless of set s, the following theorem establishes a sufficient condition for γ^* to belong to $\Gamma_0^n(s)$

Theorem 3 *Let $\gamma^* \equiv (\lambda, \check{\alpha}, \check{\kappa}_2, \ldots, \check{\kappa}_J)$ be defined by equation (E.4) and $\check{\kappa}_1 = 0$. If $z_i \subseteq \mathcal{W}_i$ for all i, then $\gamma^* \in \Gamma_0^n(s)$ for any set s of functions $s_j \colon \mathcal{Z} \to \mathbb{R}$.*

We prove Theorem 3 in Appendix E.1.3. To use the moment inequality in equation (E.11) for all pairs of nests k and k', one must specify the functions $s_j \colon \mathcal{Z} \to \mathbb{R}$ for all $j = 1, \ldots, J$. We choose the function $s_j(z_i)$ entering equation (E.11d) as

$$s_j(z_i) = \mathbb{E}[p_{ij}|z_i]. \tag{E.13}$$

Combining Moment Inequalities. The following corollary indicates that the odds-based moment inequality in equation (E.5), the bounding moment inequality in equation (E.8), and the nested logit moment inequality in equation (E.11) may be combined to identify the vector γ^*.

Corollary 1 *Let $\gamma^* \equiv (\lambda, \check{\alpha}, \check{\kappa}_2, \ldots, \check{\kappa}_J)$ be defined by equation (E.4) and $\check{\kappa}_1 = 0$. If $z_i \subseteq \mathcal{W}_i$ for all i, then $\gamma^* \in \Gamma_0^o \cap \Gamma_0^b(e) \cap \Gamma_0^n(s)$ for any set e of functions $e_{jj'} \colon \mathcal{Z} \times \Gamma \to \mathbb{R}$ and any set s of functions $s_j \colon \mathcal{Z} \to \mathbb{R}$.*

This corollary is an immediate implication of Theorems 1, 2, and 3.

E.1.3 Proof of Theorem 3

To prove Theorem 3, we show that, for any nests k and k', any $z_i \subseteq \mathcal{W}_i$, and any $s_{k'} \colon \mathcal{Z} \to \mathbb{R}^{J_{k'}}$ equation (E.11) holds for $\gamma = \gamma^*$; i.e.,

$$\mathfrak{m}^n_{kk'}(z_i, \gamma^*, s_{k'}(\cdot)) \geqslant 0. \tag{E.14}$$

We organize our proof in five steps, described below.

Step 1. Given the definition of the function $I_k \colon \mathbb{R}^{J_k} \times \Gamma \to \mathbb{R}$ in equation (E.11c) for all $k = 1, \ldots, K$, and defining the vector of expected prices $p^e_{ik} \equiv \{p^e_{ij}\}_{j \in J_k}$ with $p^e_{ij} = \mathbb{E}[p_{ij}|\mathcal{W}_i]$ for every visit i and drug $j = 1, \ldots, J$, we can rewrite equation (E.3) as

$$\mathcal{P}(d_{ik} = 1|\mathcal{W}_i) = \frac{\exp(I_k(p^e_{ik}, \gamma^*))}{\sum_{k'=1}^K \exp(I_{k'}(p^e_{ik'}, \gamma^*))} \qquad \text{for } k = 1, \ldots, K. \tag{E.15}$$

Given this equation, we can write the following moment equality when comparing the relative probabilities of choosing alternatives in nests k and k':

$$\mathbb{E}\left[\frac{\exp(I_k(p^e_{ik}, \gamma^*))}{\exp(I_k(p^e_{ik}, \gamma^*)) + \exp(I_{k'}(p^e_{ik'}, \gamma^*))} - d_{ik}\middle|\mathcal{W}_i, d_{ik} + d_{ik'} = 1\right] = 0.$$

Defining $\Delta I^e_{ikk'}(\gamma^*) \equiv I_k(p^e_{ik}, \gamma^*) - I_{k'}(p^e_{ik'}, \gamma^*)$, we can rewrite this expression as

$$\mathbb{E}\left[\frac{\exp(\Delta I^e_{ikk'}(\gamma^*))}{1 + \exp(\Delta I^e_{ikk'}(\gamma^*))} - d_{ik}\middle|\mathcal{W}_i, d_{ik} + d_{ik'} = 1\right] = 0.$$

Multiplying by $1 + \exp(\Delta I^e_{ikk'}(\gamma^*))$ on both sides of this equality and grouping terms, we obtain

$$\mathbb{E}[(1 - d_{ik}) \exp(\Delta I^e_{ikk'}(\gamma^*)) - d_{ik}|\mathcal{W}_i, d_{ik} + d_{ik'} = 1] = 0.$$

Conditional on the event $d_{ik} + d_{ik'} = 1$, the variable $1 - d_{ik}$ equals $d_{ik'}$, and we can thus write

$$\mathbb{E}[d_{ik'} \exp(\Delta I^e_{ikk'}(\gamma^*)) - d_{ik}|\mathcal{W}_i, d_{ik} + d_{ik'} = 1] = 0.$$

Using the LIE, we eliminate the event $d_{ik} + d_{ik'} = 1$ from the conditioning set, obtaining

$$\mathbb{E}[d_{ik'} \exp(\Delta I^e_{ikk'}(\gamma^*)) - d_{ik}|\mathcal{W}_i] = 0. \tag{E.16}$$

This equality holds for any two nests k and k', any $\gamma^* \in \Gamma$ and any information set $\mathcal{W}_i$.

Step 2. Denote by $\underline{I}_{k'}(p^e_{ik'}, \gamma^*, s_{k'}(z_i))$ the first-order approximation to $I_{k'}(p^e_{ik'}, \gamma^*)$ around $s_{k'}(z_i) \equiv \{s_j(z_i)\}_{j \in J_{k'}}$. Then, $\underline{I}_{k'}(p^e_{ik'}, \gamma^*, s_{k'}(z_i))$ coincides with the expression in equation (E.11d) except that its first argument equals the vector of expected prices and its second argument equals the vector of true parameter values. That is,

$$\underline{I}_{k'}(p^e_{ik'}, \gamma^*, s_{k'}(z_i)) =$$
$$\lambda \ln(\sum_{j \in J_{k'}} \exp(\check{\kappa}_j + \check{\alpha} s_j(z_i))) + \lambda\check{\alpha} \sum_{j \in J_{k'}} \frac{\exp(\check{\kappa}_j + \check{\alpha} s_j(z_i))}{\sum_{j' \in J_{k'}} \exp(\check{\kappa}_{j'} + \check{\alpha} s_{j'}(z_i))}(p^e_{ij} - s_j(z_i)). \tag{E.17}$$

As $I_{k'}(p^e_{ik'}, \gamma^*)$ is convex in $p^e_{ik'} \in \mathbb{R}^{J_{k'}}$ for all $\gamma^* \in \Gamma$, it holds that $I_{k'}(p^e_{ik'}, \gamma^*) \geqslant \underline{I}_{k'}(p^e_{ik'}, \gamma^*, s_{k'}(z_i))$ for all $\gamma^* \in \Gamma$ and all $(p^e_{ik'}, s_{k'}(z_i)) \in \mathbb{R}^{2J_{k'}}$. Therefore, one can write

$$\Delta I^e_{ikk'}(\gamma^*) \leqslant I_k(p^e_{ik}, \gamma^*) - \underline{I}_{k'}(p^e_{ik'}, \gamma^*, s_{k'}(z_i)). \tag{E.18}$$

Step 3. Combining equations (E.16) and (E.18), we derive the following moment inequality

$$\mathbb{E}[d_{ik'} \exp(I_k(p^e_{ik}, \gamma^*) - \underline{I}_{k'}(p^e_{ik'}, \gamma^*, s_{k'}(z_i))) - d_{ik}|\mathcal{W}_i] \geqslant 0, \tag{E.19}$$

for any two nests k and k', any function $s_{k'}\colon \mathcal{Z} \to \mathbb{R}^{J_{k'}}$, any information set $\mathcal{W}_i$, and any $\gamma^* \in \Gamma$.

Step 4. As $d_{ik'}$ is measurable in $\mathcal{J}_i$, the function $\exp(I_k(p_{ik}, \gamma^*) - \underline{I}_{k'}(p_{ik'}, \gamma^*, s_{k'}(z_i)))$ is convex in $(p_{ik}, p_{ik'})$, and $(p_{ik}, p_{ik'})$ is a mean-preserving spread of $(p^e_{ik}, p^e_{ik'})$, Jensen's inequality implies

$$\begin{gathered}\mathbb{E}[d_{ik'} \exp(I_k(p^e_{ik}, \gamma^*) - \underline{I}_{k'}(p^e_{ik'}, \gamma^*, s_{k'}(z_i)))|\mathcal{J}_i] \\ \leqslant \\ \mathbb{E}[d_{ik'} \exp(I_k(p_{ik}, \gamma^*) - \underline{I}_{k'}(p_{ik'}, \gamma^*, s_{k'}(z_i)))|\mathcal{J}_i].\end{gathered}$$

Furthermore, as $\mathcal{W}_i \subseteq \mathcal{J}_i$, applying the LIE, we obtain

$$\begin{gathered}\mathbb{E}[d_{ik'} \exp(I_k(p^e_{ik}, \gamma^*) - \underline{I}_{k'}(p^e_{ik'}, \gamma^*, s_{k'}(z_i)))|\mathcal{W}_i] \\ \leqslant \\ \mathbb{E}[d_{ik'} \exp(I_k(p_{ik}, \gamma^*) - \underline{I}_{k'}(p_{ik'}, \gamma^*, s_{k'}(z_i)))|\mathcal{W}_i].\end{gathered} \tag{E.20}$$

Step 5. Combining equations (E.19) and (E.20), we can derive the moment inequality

$$\mathbb{E}[d_{ik'} \exp(I_k(p_{ik}, \gamma^*) - \underline{I}_{k'}(p_{ik'}, \gamma^*, s_{k'}(z_i)) - d_{ik}|\mathcal{W}_i] \geqslant 0,$$

for any two nests k and k' and any $\gamma^* \in \Gamma$. Finally, we take the expectation of both sides of this inequality conditional on z_i. If $z_i \subseteq \mathcal{W}_i$, the LIE implies that

$$\mathbb{E}[d_{ik'} \exp(I_k(p_{ik}, \gamma^*) - \underline{I}_{k'}(p_{ik'}, \gamma^*, s_{k'}(z_i)) - d_{ik}|z_i] \geqslant 0, \tag{E.21}$$

with $I_k(p_{ik}, \gamma^*)$ and $\underline{I}_{k'}(p_{ik'}, \gamma^*, s_{k'}(z_i))$ defined as in equations (E.11c) and (E.11d), but evaluated at $\gamma = \gamma^*$. □

E.1.4 Simulation

Setup. We consider a simulation set-up that is identical in every respect to that described in Section 6.1 except for the distribution of the vector $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3})$. For this simulation, we assume that ε_i is distributed according to the following distribution

$$F_\varepsilon(\varepsilon_i|\mathcal{W}_i) = \exp(-(\exp(-\varepsilon_{i1}/0.5) + \exp(-\varepsilon_{i2}/0.5))^{0.5} - \exp(-\varepsilon_{i3}), \tag{E.22}$$

which corresponds to the distribution in equation (E.1) with $\lambda = 0.5$, $K = 2$, $J_1 = \{1, 2\}$ and $J_2 = \{3\}$. That is, we group the three possible choices into two nests: the first two choices belong to the same nest, and the third one forms a nest in itself. We set $\alpha = 0.5$, and $(\kappa_1, \kappa_2, \kappa_3) = (0, 0, 0.5)$. As a result, $\check{\alpha} = 1$ and $(\check{\kappa}_1, \check{\kappa}_2, \check{\kappa}_3) = (0, 0, 1)$.

MLE: Implementation. We compute ML estimates of $(\lambda, \alpha, \kappa_2, \kappa_3)$ by solving the following optimization problem

$$\underset{(\gamma_\lambda,\gamma_\alpha,\gamma_{\kappa_2},\gamma_{\kappa_3})}{\operatorname{argmax}} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{3} \mathbb{1}\{d_{ij} = 1\} \ln \Big(\mathcal{P}\big(d_{ij} = 1 | z_i; (\gamma_\lambda, \gamma_\alpha, \gamma_{\kappa_2}, \gamma_{\kappa_3})\big) \Big) \right\}$$

with

$$\mathcal{P}(d_{ij} = 1 | z_i; (\gamma_\lambda, \gamma_\alpha, \gamma_{\kappa_2}, \gamma_{\kappa_3})) \equiv \frac{\exp((\gamma_{\kappa_j}/\gamma_\lambda) + (\gamma_\alpha/\gamma_\lambda)\mathbb{E}[p_{ij}|z_i])(\sum_{j' \in J_k} \exp((\gamma_{\kappa_{j'}}/\gamma_\lambda) + (\gamma_\alpha/\gamma_\lambda)\mathbb{E}[p_{ij'}|z_i]))^{\lambda-1}}{\sum_{k'=1}^{K}(\sum_{j' \in J_{k'}} \exp((\gamma_{\kappa_{j'}}/\gamma_\lambda) + (\gamma_\alpha/\gamma_\lambda)\mathbb{E}[p_{ij'}|z_i]))^{\lambda}}, \tag{E.23}$$

and the normalization $\gamma_{\kappa_1} = 1$. In equation (E.23), we use $(\gamma_\lambda, \gamma_\alpha, \gamma_{\kappa_2}, \gamma_{\kappa_3})$ to denote an unknown parameter vector whose true value is $(\lambda, \alpha, \kappa_2, \kappa_3)$. In cases 1 to 4 in Table E.1, we set $z_i = x_{2i}$ and, thus, $\mathbb{E}[p_{ij}|z_i] = x_{2ij}$. In case 5 in Table E.1, we set $z_i = p_i$ and, thus, $\mathbb{E}[p_{ij}|z_i] = p_{ij}$.

Moment Inequalities: Implementation. Following the approach described in Section 5, we compute confidence sets for $(\lambda, \alpha, \kappa_2, \kappa_3)$ using a finite number of unconditional moment inequalities implied by the conditional ones in equations (E.5), (E.8) and (E.11). Specifically, for each nest $k \in \{1, 2\}$, each ordered pair of drugs (j, j') belonging to nest k, and a set of S instrument functions $g^{(s)}_{jj'}(z_i)$ for $s = 1, \ldots, S$, we use the odds-based moment inequality

$$\mathbb{E}[(d_{ij} \exp(-((\gamma_{\kappa_j}/\gamma_\lambda) - (\gamma_{\kappa_{j'}}/\gamma_\lambda) + (\gamma_\alpha/\gamma_\lambda)\Delta p_{ijj'})) - d_{ij'})d_{ik} g^{(s)}_{jj'}(z_i)] \geqslant 0, \tag{E.24}$$

and the bounding moment inequality

$$\begin{gathered}\mathbb{E}[(d_{ij'} - d_{ij} \exp(-e_{jj'}(z_i, \gamma)) \times \\ (1 + e_{jj'}(z_i, \gamma) - ((\gamma_{\kappa_j}/\gamma_\lambda) - (\gamma_{\kappa_{j'}}/\gamma_\lambda) + (\gamma_\alpha/\gamma_\lambda)\Delta p_{ijj'})))d_{ik} g^{(s)}_{jj'}(z_i)] \geqslant 0,\end{gathered} \tag{E.25}$$

where, abusing notation, we now use γ to denote the parameter vector $\gamma \equiv (\gamma_\lambda, \gamma_\alpha, \gamma_{\kappa_2}, \gamma_{\kappa_3})$, whose true value is $\gamma^* \equiv (\lambda, \alpha, \kappa_2, \kappa_3)$. Equations (E.24) and (E.25) correspond to the unconditional version of the conditional moment inequalities in equations (E.5) and (E.8). In our particular simulation, since only the nest $k = 1$ includes more than one alternative and $J_1 = \{1, 2\}$, we implement the inequalities in equations (E.24) and (E.25) exclusively for $k = 1$ and $(j, j') \in \{(1, 2), (2, 1)\}$. For the function $e_{jj'}(\cdot)$ entering equation (E.25), we compute it as the predicted value of a spline regression of $\Delta p_{ijj'}$ on z_i estimated on the subset of observations with $d_{ij} = 1$. For the functions $g^{(s)}_{jj'}(z_i)$ for $s = 1, \ldots, S$ entering equations (E.24) and (E.25), unless otherwise noted, we set $z_i = x_{2i}$ and $S = 2$, and define them as

$$g^{(1)}_{jj'}(x_{2i}) = \mathbb{1}\{\Delta x_{2ijj'} \geqslant 0\} \qquad \text{and} \qquad g^{(2)}_{jj'} = \mathbb{1}\{\Delta x_{2ijj'} < 0\}, \tag{E.26}$$

with $\Delta x_{2ijj'} = x_{2ij} - x_{2ij'}$.

In addition to the moment inequalities in equations (E.24) and (E.25), we also use an unconditional version of the conditional moment inequality in equation (E.11). Specifically, for a set of S instrument functions $\tilde{g}^{(s)}_{kk'}(z_i)$ for $s = 1, \ldots, S$, we use the moment inequality

$$\mathbb{E}[(d_{ik'} \exp(I_k(p_{ik}, \gamma) - \underline{I}_{k'}(p_{ik'}, \gamma, s_{k'}(z_i))) - d_{ik})\tilde{g}^{(s)}_{kk'}(z_i)] \geqslant 0 \tag{E.27}$$

with $I_k(p_{ik}, \gamma)$ defined as in equation (E.11c) and $\underline{I}_{k'}(p_{ik'}, \gamma, s_{k'}(z_i))$ defined as in equation (E.11d), but written as a function of the parameter vector $\gamma \equiv (\gamma_\lambda, \gamma_\alpha, \gamma_{\kappa_2}, \gamma_{\kappa_3})$. That is, for example, the term $I_k(p_{ik}, \gamma)$ in equation (E.27) corresponds to $\gamma_\lambda \ln(\sum_{j \in J_k} \exp((\gamma_{\kappa_j}/\gamma_\lambda) + (\gamma_\alpha/\gamma_\lambda)p_{ij})$. For the functions $\tilde{g}^{(s)}_{kk'}(z_i)$ for $s = 1, \ldots, S$ entering equation (E.27), unless otherwise noted, we set $z_i = x_{2i}$ and $S = 2$, and define them as

$$\tilde{g}^{(1)}_{kk'}(x_{2i}) = \mathbb{1}\{\sum_{j \in J_k} \exp(s_j(x_{2i})) - \sum_{j \in J_{k'}} \exp(s_j(x_{2i})) \geqslant 0\}, \tag{E.28}$$

with $\tilde{g}^{(2)}_{kk'}(x_{2i}) = 1 - \tilde{g}^{(1)}_{kk'}(x_{2i})$ and the function $s_j(\cdot)$ defined in equation (E.13).

Results. Table E.1 presents simulation results. We first discuss the maximum likelihood (ML) estimates. Comparing the nested logit estimates of $\hat{\alpha} = \alpha/\lambda$, $\hat{\kappa}_2 = \kappa_2/\lambda$, and $\hat{\kappa}_3 = \kappa_3/\lambda$ to the multinomial logit estimates of α, κ_2, κ_3 (in Table 2), we find nearly identical patterns in the estimates for all five cases we consider. First, the ML estimates of the nested logit parameters $\hat{\alpha}$, $\hat{\kappa}_2$, and $\hat{\kappa}_3$ are consistent if the researcher's assumed information set coincides with the agent's true information set—i.e., if $z_i = x_{2i}$ and $\sigma_1 = 0$, as in cases 1 and 2. Second, the estimates are biased downward if the researcher's assumed information set is a strict subset of the agent's information set (i.e., if $z_i = x_{2i}$ and $\sigma_1 > 0$, as in cases 3 and 4) or if the agent's information set is a strict subset of the researcher's assumed one (i.e., if $z_i = p_i$ and $\sigma_3 > 0$, as in case 5). When bias arises, it does so due to the combination of a downward bias in the estimates of α, κ_2, κ_3 and an upward bias in the estimates of λ. Given the assumed distribution of the vector ε_i in equation (E.1), the upward bias in the estimate of λ makes the estimated version of the nested logit model closer to the multinomial logit model than what the true value of λ implies.

The confidence intervals (CIs) for the nested logit parameters $\hat{\alpha} = \alpha/\lambda$ and $\hat{\kappa}_2 = \kappa_2/\lambda$ presented in Table E.1 are identical in all cases to the CIs for the multinomial logit parameters α and κ_2 presented in Table 2. The reason is that the CIs for the nested logit parameters $\hat{\alpha}$ and $\hat{\kappa}_2$ are bounded by the odds-based and bounding moment inequalities in equations (E.24) and (E.25), respectively, when applied to the pairs of alternatives $(j, j') \in \{(1, 2), (2, 1)\}$. These moment inequalities are identical to the odds-based and bounding moment inequalities in equations (25) and (26) when applied to the same set of pairs of alternatives $(j, j') \in \{(1, 2), (2, 1)\}$, which are themselves the moment inequalities we use to compute the CIs reported in Table 2.

Table E.1: Simulation Results - MLE and Confidence Intervals

Case	σ_1	σ_3	z_i	Estimator	MLE & Confidence Sets			
					α	κ_2	κ_3	λ
1	0	0	x_{2i}	MLE	0.50	0	0.50	0.50
				Odds-based	[0.50 , 0.50]	[0 , 0]	[0.50 , 0.50]	[0.50 , 0.50]
				Bounding	[0.50 , 0.50]	[0 , 0]	[0.50 , 0.50]	[0.50 , 0.50]
				Both	[0.50 , 0.50]	[0 , 0]	[0.50 , 0.50]	[0.50 , 0.50]
2	0	1	x_{2i}	MLE	0.50	0	0.50	0.50
				Odds-based	[0.47 , 0.55]	[-0.18 , 0.15]	[0.32 , 0.65]	[0.22 , 0.60]
				Bounding	[0.47 , 0.55]	[0 , 0]	[0.42 , 0.58]	[0.47 , 0.55]
				Both	[0.47 , 0.55]	[0 , 0]	[0.42 , 0.58]	[0.47 , 0.55]
3	1	0	x_{2i}	MLE	0.49	0	0.48	0.53
				Odds-based	[0.50 , 0.50]	[0 , 0]	[0.48 , 0.50]	[0.50 , 0.50]
				Bounding	[0.50 , 0.50]	[-0.15 , 0.15]	[0.42 , 0.55]	[0.47 , 0.60]
				Both	[0.50 , 0.50]	[0 , 0]	[0.48 , 0.50]	[0.50 , 0.50]
4	1	1	x_{2i}	MLE	0.49	0	0.48	0.53
				Odds-based*	[0.47 , 0.60]	[-0.28 , 0.28]	[0.22 , 0.65]	[0.22 , 0.63]
				Bounding	[0.47 , 0.58]	[-0.18 , 0.18]	[0.35 , 0.65]	[0.45 , 0.73]
				Both	[0.47 , 0.58]	[-0.18 , 0.18]	[0.35 , 0.65]	[0.45 , 0.63]
5	0	1	p_i	MLE	0.46	-0.01	0.46	0.53
				Odds-based	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
				Bounding	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
				Both	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

Note: σ_1 and σ_3 are parameters of the distributions of x_{1ij} and x_{3ij}, as indicated in footnote 9 in the main text. *MLE* indicates the maximum likelihood estimate. *Odds-based*, *Bounding*, and *Both* contain the projections on each parameter of 95% confidence sets computed according to the procedure in Andrews and Soares (2010). *Odds-based* indicates the corresponding confidence set is computed using only inequalities of the type in equations (E.24) and (E.27); *Bounding* indicates the confidence set is computed using only inequalities of the type in equations (E.25) and (E.27); *Both* indicates the confidence set is computed using inequalities of the type in equations (E.24), (E.25) and (E.27). In cases 1 to 4, we build the moment inequalities using the instrument functions in equations (E.26) and (E.28). In case 5, we build the inequalities using analogous instrument functions, but with $z_i = p_i$. In all cases, confidence sets are computed by testing points in a 4-dimensional grid; we mark cases with an asterisk when the confidence set includes points outside the grid.

The CIs for $\hat{\kappa}_3 = \kappa_3/\lambda$ presented in Table E.1 are weakly larger than the CIs for κ_3 presented in Table 2. The reason is that, since $j = 3$ is the only choice in its nest, the parameter $\gamma_{\kappa_3}/\gamma_\lambda$ does not enter any odds-based or bounding moment inequality, and is thus identified only through the nested logit moment inequality in equation (E.27). As a result, the identification relies on the comparison of the relative preferences for the two nests $k \in \{1, 2\}$, and not on the within-nest comparison across different alternatives, resulting in a loss of identification power.

The CIs for λ presented in Table E.1 are exclusively determined by the nested logit moment inequality in equation (E.27). The derivation of this moment inequality requires both taking a first-order approximation to a convex function of the physician's price expectations around a proxy of such expectations (see *Step 2* in Section E.1.3) as well as substituting the physician's price expectation by the actual price inside a convex function (see *Step 4* in Section E.1.3). As a result, the parameter λ will generally be only partially identified.

Finally, case 5 in Table E.1 shows that, as in the multinomial logit case, the CI defined by our moment inequalities is empty when the researcher uses an instrument that does not belong to the agent's true information set. For example, when we use actual drug prices as instruments in a setting in which the agent has only imperfect expectations about them, as in case 5 in Table E.1, we find empty confidence sets.

E.2 Moment Inequalities for Multinomial Probit Models

In Appendix E.2.1, we describe the multinomial probit model. Appendix E.2.2 describes a new set of moment inequalities, and Appendix E.2.3 shows formally that these inequalities can be used to partially identify the parameters of the multinomial probit model. Appendix E.2.4 presents simulation results that illustrate certain properties of these inequalities.

E.2.1 Model

We assume physicians choose prescription drugs at each visit according to the restrictions imposed in equations (1) to (5) and (23) except we impose a distinct distributional assumption. Instead of the assumption in equation (4), we impose here that the vector of shocks ε_i follows the distribution

$$\varepsilon_i|\mathcal{W}_i \sim \mathbb{N}(0_J, \Sigma) \tag{E.29}$$

with 0_J a J-dimensional vector of 0s, and Σ a symmetric and positive semidefinite $J \times J$ matrix. For $j = 1, \ldots, J$, we denote $\sigma_j^2 = \text{Var}(\varepsilon_{ij}|\mathcal{W}_i)$ and $\sigma_{jj'} = \text{Cov}(\varepsilon_{ij}, \varepsilon_{ij'}|\mathcal{W}_i)$ for any $j' \neq j$.

The identification of the multinomial probit model requires us to impose additional restrictions on the elements of Σ. For simplicity, we impose that, for all j and j', $\sigma_j = 1$ and $\sigma_{jj'} = \eta q_{jj'}$, where $q_{jj'}$ is an observed measure of proximity between drugs j and j'. According to these restrictions, the variance of ε_{ij} equals one for all drugs, and the covariance between ε_{ij} and $\varepsilon_{ij'}$ for any two drugs j and j' is equal to a constant η multiplied by the measure $q_{jj'}$. We can then write Σ as

$$\Sigma = \begin{bmatrix} 1 & \eta q_{12} & \eta q_{13} & \ldots & \eta q_{1J} \\ \eta q_{21} & 1 & \eta q_{23} & \ldots & \eta q_{2J} \\ \eta q_{31} & \eta q_{32} & 1 & \ldots & \eta q_{3J} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \eta q_{J1} & \eta q_{J2} & 1 & \ldots & 1 \end{bmatrix}. \tag{E.30}$$

Equations (1), (2), (3), (5), (23), and (E.29) then imply that, for all $j = 1, \ldots, J$, the probability that drug j is prescribed given $\mathcal{W}_i$ is

$$\mathcal{P}(d_{ij} = 1|\mathcal{W}_i) =$$
$$\int_{\varepsilon_i} \mathbb{1}\{\kappa_j + \alpha\mathbb{E}[p_{ij}|\mathcal{W}_i] + \varepsilon_{ij} \geqslant \max_{j'=1,\ldots,J}\{\kappa_{j'} + \alpha\mathbb{E}[p_{ij'}|\mathcal{W}_i] + \varepsilon_{ij'}\}\}dF(\varepsilon_i|\mathcal{W}_i; \Sigma)d\varepsilon_i, \tag{E.31}$$

where $F(\varepsilon_i|\mathcal{W}_i;\Sigma)$ is the CDF of the multivariate normal distribution in equation (E.29).

Given the restrictions $\kappa_1 = 0$, $\sigma_j = 1$ for all j, and $\sigma_{jj'} = \eta q_{jj'}$ for any $j \neq j'$, the goal of estimation is to learn about the value of $\{\kappa_j\}_{j=2}^{J}$, α, and η, and about the content of $\{\mathcal{W}_i\}_{i=1}^{N}$. To simplify the notation, we use $\psi \equiv (\psi_\eta, \psi_\alpha, \psi_{\kappa_2}, \ldots, \psi_{\kappa_J})$ to denote the unknown parameter vector, Ψ to denote the parameter space, and $\psi^* \equiv (\eta, \alpha, \kappa_2, \ldots, \kappa_J)$ to denote the true parameter value.

E.2.2 Moment Inequalities

In this section, we show how to partially identify the parameter vector ψ^* defined by equations (E.30) and (E.31). We use two types of moment inequalities: odds-based inequalities similar to those described in Section 4.1, and bounding inequalities similar to those described in Section 4.2.

Odds-based Moment Inequalities. For any two drugs j and j', any value of $z_i \in \mathcal{Z}$, and any $\psi \in \Psi$, we define the following odds-based moment inequality

$$\mathbb{m}^{o}_{jj'}(z_i, \psi) \geqslant 0 \tag{E.32a}$$

with

$$\mathbb{m}^{o}_{jj'}(z_i, \psi) \equiv \mathbb{E}\left[(1-d_{ij})\frac{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'} + \alpha\Delta p_{ijj'}))}{1-\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'} + \alpha\Delta p_{ijj'}))} - d_{ij}\middle| z_i\right], \tag{E.32b}$$

where $\omega_{jj'}$ is the standard deviation of $\Delta\varepsilon_{ijj'}$. That is, in general $\omega_{jj'} = (\sigma_j^2 + \sigma_{j'}^2 - 2\sigma_{jj'})^{\frac{1}{2}}$ and, given the restrictions in equation (E.30),

$$\omega_{jj'} = (2(1-\eta q_{jj'}))^{\frac{1}{2}}. \tag{E.33}$$

We denote as Ψ_0^o the set of values of ψ that jointly satisfy the inequality in equation (E.32) for every value of z_i in its support and all pairs of drugs j and j' in the choice set. Formally, we define

$$\Psi_0^o \equiv \{\psi \in \Psi \colon \mathbb{m}^{o}_{jj'}(z, \psi) \geqslant 0 \text{ for all } z \in \mathcal{Z},\ j = 1, \ldots, J, \text{ and } j' = 1, \ldots, J\}. \tag{E.34}$$

Theorem 4 establishes a sufficient condition for the true parameter value ψ^* to belong to Ψ_0^o.

Theorem 4 *Let $\psi^* \equiv (\alpha, \eta, \kappa_2, \ldots, \kappa_J)$ be defined by equations* (E.30) *and* (E.31) *and the normalization $\kappa_1 = 0$. If $z_i \subseteq \mathcal{W}_i$, then $\psi^* \in \Psi_0^o$.*

We prove Theorem 4 in Appendix E.2.3.

Bounding Moment Inequalities. For any two drugs j and j' in the physician's choice set, any value of z_i in its support $\mathcal{Z}$, and any function $e_{jj'} \colon \mathcal{Z} \to \mathbb{R}$, we define the bounding moment inequality

$$\mathbb{m}^{b}_{jj'}(z_i, \psi, e_{jj'}(\cdot)) \geqslant 0 \tag{E.35a}$$

with

$$\mathbb{m}^b_{jj'}(z_i,\theta,e_{jj'}(\cdot)) \equiv \mathbb{E}\Bigg[(1-d_{ij}) + d_{ij}\Bigg(- \frac{1-\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha e_{jj'}(z_i)))}{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha e_{jj'}(z_i)))} + \frac{\alpha}{\omega_{jj'}}\frac{\phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha e_{jj'}(z_i)))}{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha e_{jj'}(z_i)))^2}(\Delta p_{ijj'} - e_{jj'}(z_i))\Bigg)\Bigg| z_i\Bigg], \tag{E.35b}$$

and $\omega_{jj'}$ equal to the expression in equation (E.33). The moment $\mathbb{m}^b_{jj'}(\cdot)$ depends on $e_{jj'}(z_i)$, which is a deterministic function of z_i and can vary by j and j'. Defining e as the set of functions including $e_{jj'}(\cdot)$ for all drug pairs (i.e., $e = \{e_{jj'}(\cdot)\}_{j,j'}$), we denote as $\Psi^b_0(e)$ the set of values of ψ that jointly satisfy the inequality in equation (E.35) for every value of z_i in its support and every pair of drugs j and j' in the physician's choice set. Formally,

$$\Psi^b_0(e) \equiv \{\psi \in \Psi \colon \mathbb{m}^b_{jj'}(z,\psi,e_{jj'}(\cdot)) \geqslant 0 \text{ for all } z \in \mathcal{Z},\ j=1,\ldots,J, \text{ and } j'=1,\ldots,J\}. \tag{E.36}$$

Regardless of the set e used to build the moment inequalities in equation (E.35), the following theorem establishes a sufficient condition for the true parameter value ψ^* to belong to $\Psi^b_0(e)$.

Theorem 5 *Let $\psi^* \equiv (\alpha,\eta,\kappa_2,\ldots,\kappa_J)$ be defined by equations (E.30) and (E.31) and the normalization $\kappa_1 = 0$. If $z_i \subseteq \mathcal{W}_i$, then $\psi^* \in \Psi^b_0(e)$ for any set e of functions $e_{jj'}\colon \mathcal{Z} \to \mathbb{R}$.*

We prove Theorem 5 in Appendix E.2.3. To use the moment inequality in equation (E.35) for any pair of drugs j and j', one must specify the function $e_{jj'}\colon \mathcal{Z}\to\mathbb{R}$. We choose the function $e_{jj'}(z_i)$ entering equation (E.35b) as

$$e_{jj'}(z_i) = \mathbb{E}[\Delta p_{ijj'} | z_i, d_{ij} = 1]. \tag{E.37}$$

E.2.3 Proof of Theorems 4 and 5

Proof of Theorem 4. To prove Theorem 4, we show that, for any drugs j and j' and any $z_i \subseteq \mathcal{W}_i$, equation (E.32) holds for $\psi = \psi^*$; i.e.,

$$\mathbb{m}^o_{jj'}(z_i,\psi^*) \geqslant 0. \tag{E.38}$$

We organize our proof in three steps. This proof has many similarities with the proof of Theorem 1 in Appendix B.1, with the differences highlighted below.

Step 1. Equation (2) implies that, for any i and any two drugs j and j' in i's choice set, the following inequality holds

$$\mathbb{1}\{\mathbb{E}[\mathcal{U}_{ij} - \mathcal{U}_{ij'} | \mathcal{J}_i] \geqslant 0\} - d_{ij} \geqslant 0. \tag{E.39}$$

This inequality indicates that physician i weakly prefers alternative j over alternative j' (i.e., $\mathbb{E}[\mathcal{U}_{ij} - \mathcal{U}_{ij'}|\mathcal{J}_i] \geqslant 0$) if she weakly prefers alternative j over all other alternatives (i.e., $d_{ij} = 1$). Relative to equation (B.2), the left-hand side of equation (E.39) is not multiplied by the dummy variable $d_{ij} + d_{ij'}$; that is, when comparing the expected utility of alternatives j and j', equation (E.39) does not condition on agents whose preferred choice is j or j'.

Given equations (5) and (23), we can rewrite equation (E.39) as

$$\mathbb{1}\{\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i] + \Delta\varepsilon_{ijj'} \geqslant 0\} - d_{ij} \geqslant 0, \tag{E.40}$$

where $\Delta\varepsilon_{ijj'} = \varepsilon_{ij} - \varepsilon_{ij'}$, $\Delta\kappa_{jj'} = \kappa_j - \kappa_{j'}$, and $\Delta p_{ijj'} = p_{ij} - p_{ij'}$. As equation (E.40) holds for every observation i, it also holds on average across subsets of observations. Thus,

$$\mathbb{E}[\mathbb{1}\{\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i] + \Delta\varepsilon_{ijj'} \geqslant 0\} - d_{ij}|\mathcal{W}_i] \geqslant 0. \tag{E.41}$$

Given the distributional assumption in equations (E.29), it holds that

$$\Delta\varepsilon_{ijj'}|\mathcal{W}_i \sim \mathbb{N}(0, \omega^2_{jj'}). \tag{E.42}$$

with $\omega_{jj'}$ as indicated in equation (E.33). Then, we can rewrite equation (E.41) as

$$\mathbb{E}[\Phi(\omega^{-1}_{jj'}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i])) - d_{ij}|\mathcal{W}_i] \geqslant 0.$$

Dividing by $1 - \Phi(\omega^{-1}_{jj'}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))$ and adding and subtracting d_{ij} inside the expectation, we obtain

$$\mathbb{E}\left[(1 - d_{ij})\frac{\Phi(\omega^{-1}_{jj'}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))}{1 - \Phi(\omega^{-1}_{jj'}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))} - d_{ij}\middle|\mathcal{W}_i\right] \geqslant 0, \tag{E.43}$$

for any two drugs j and j' in agent i's choice set. Equation (E.43) has two differences with respect to equation (B.4). First, equation (E.43) is an inequality, while the moment in equation (B.4) holds with equality. Second, the ratio of the probability of choosing drug j to choosing drug j' is multiplied by the variable $d_{ij'}$ in equation (B.4) and by the variable $1 - d_{ij}$ in equation (E.43). These two differences make the inequalities we derive for the multinomial probit model less informative about the true parameter value than the inequalities for the multinomial logit model.

Step 2. As d_{ij} is measurable in $\mathcal{J}_i$, the function $\Phi(x)/(1 - \Phi(x))$ is convex in $x \in \mathbb{R}$, and $\Delta p_{ijj'}$ is a mean-preserving spread of $\mathbb{E}[\Delta p_{ijj'}|\mathcal{J}_i]$, Jensen's inequality implies

$$\mathbb{E}\left[d_{ij}\frac{\Phi(\omega^{-1}_{jj'}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{J}_i]))}{1 - \Phi(\omega^{-1}_{jj'}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{J}_i]))}\middle|\mathcal{J}_i\right] \leqslant \mathbb{E}\left[d_{ij}\frac{\Phi(\omega^{-1}_{jj'}(\Delta\kappa_{jj'} + \alpha\Delta p_{ijj'}))}{1 - \Phi(\omega^{-1}_{jj'}(\Delta\kappa_{jj'} + \alpha\Delta p_{ijj'}))}\middle|\mathcal{J}_i\right].$$

Using equation (5), we can rewrite this inequality as:

$$\mathbb{E}\left[d_{ij}\frac{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))}{1-\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))}\middle|\mathcal{J}_i\right] \leqslant \mathbb{E}\left[d_{ij}\frac{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha\Delta p_{ijj'}))}{1-\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha\Delta p_{ijj'}))}\middle|\mathcal{J}_i\right].$$

Furthermore, as $\mathcal{W}_i \subseteq \mathcal{J}_i$, applying the Law of Iterated Expectations (LIE), we obtain

$$\mathbb{E}\left[d_{ij}\frac{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))}{1-\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))}\middle|\mathcal{W}_i\right] \leqslant \mathbb{E}\left[d_{ij}\frac{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha\Delta p_{ijj'}))}{1-\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha\Delta p_{ijj'}))}\middle|\mathcal{W}_i\right]. \tag{E.44}$$

Step 3. Combining equations (E.43) and (E.44), we obtain the following inequality

$$\mathbb{E}\left[(1-d_{ij})\frac{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha\Delta p_{ijj'}))}{1-\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha\Delta p_{ijj'}))}-d_{ij}\middle|\mathcal{W}_i\right] \geqslant 0.$$

Finally, we take the expectation of both sides of this inequality conditional on z_i. If $z_i \subseteq \mathcal{W}_i$, the LIE implies that

$$\mathbb{E}\left[(1-d_{ij})\frac{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha\Delta p_{ijj'}))}{1-\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha\Delta p_{ijj'}))}-d_{ij}\middle|z_i\right] \geqslant 0.$$

which coincides with equation (E.38). □

Proof of Theorem 5. To prove Theorem 5, we show that, for any choices j and j', any $z_i \subseteq \mathcal{W}_i$, and any $e_{jj'}\colon \mathcal{Z} \to \mathbb{R}$, equation (E.35) holds when $\psi = \psi^*$; i.e.,

$$\mathbb{m}_{jj'}^{b}(z_i, \psi^*, e_{jj'}(\cdot)) \geqslant 0, \tag{E.45}$$

for any choices j and j', any $z_i \subseteq \mathcal{W}_i$, and any function $e_{jj'}\colon \mathcal{Z} \to \mathbb{R}$. We organize our proof in four steps. This proof has many similarities with the proof of Theorem 2, which is analogous to Theorem 5 in Appendix B.2 for the multinomial logit model. We stress below the differences between the proof of Theorem 5 and the proof of Theorem 2, which itself reproduces results in Porcher et al. (2025).

Step 1. Equation (2) implies that, for any i and any two drugs j and j' in i's choice set, it holds

$$(1-d_{ij}) - \mathbb{1}\{\mathbb{E}[\mathcal{U}_{ij}-\mathcal{U}_{ij'}|\mathcal{J}_i] < 0\} \geqslant 0. \tag{E.46}$$

This inequality indicates that there will be some alternative in physician i's choice set that she prefers over alternative j (i.e., $1 - d_{ij} = 1$) if she strictly prefers alternative j' over alternative j (i.e., $\mathbb{E}[\mathcal{U}_{ij} - \mathcal{U}_{ij'}|\mathcal{J}_i] < 0$). Using equation (23), we can rewrite equation (E.46) as

$$(1 - d_{ij}) - \mathbb{1}\{\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i] + \Delta\varepsilon_{ijj'} < 0\} \geqslant 0. \tag{E.47}$$

As equation (E.47) holds for every i, it also holds on average across subsets of observations. Thus,

$$\mathbb{E}[(1 - d_{ij}) - \mathbb{1}\{\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i] + \Delta\varepsilon_{ijj'} < 0\}|\mathcal{W}_i] \geqslant 0. \tag{E.48}$$

Given equation (E.42), we can rewrite equation (E.48) as

$$\mathbb{E}[(1 - d_{ij}) - (1 - \Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i])))|\mathcal{W}_i] \geqslant 0.$$

Dividing by $\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))$, we obtain

$$\mathbb{E}\left[\frac{(1 - d_{ij})}{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))} - \frac{(1 - \Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i])))}{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))}\middle|\mathcal{W}_i\right] \geqslant 0.$$

Adding and subtracting $1 - d_{ij}$ inside the expectation, we can further obtain

$$\mathbb{E}\left[(1 - d_{ij}) + d_{ij}\left(-\frac{1 - \Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))}{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))}\right)\middle|\mathcal{W}_i\right] \geqslant 0. \tag{E.49}$$

This inequality holds for any two alternatives j and j' in agent i's choice set. We could have obtained the inequality in equation (E.49) directly by multiplying by $1 - \Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))$ and dividing by $\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))$ both sides of the inequality in equation (E.43).

Step 2. As the function $-(1 - \Phi(x))/\Phi(x)$ is concave in $x \in \mathbb{R}$, a first-order linear approximation of this function around any point will bound it from above. Denote as $\bar{o}(x, \bar{x})$ the first-order approximation to $o(x) = -(1 - \Phi(x))/\Phi(x)$ at $\bar{x}$; i.e.,

$$\bar{o}(x, \bar{x}) = -\frac{1 - \Phi(\bar{x})}{\Phi(\bar{x})} - \frac{-\phi(\bar{x})\Phi(\bar{x}) - (1 - \Phi(\bar{x}))\phi(\bar{x})}{\Phi(\bar{x})^2}(x - \bar{x}) = -\frac{1 - \Phi(\bar{x})}{\Phi(\bar{x})} + \frac{\phi(\bar{x})}{\Phi(\bar{x})^2}(x - \bar{x}).$$

As $o(x)$ is concave in $x \in \mathbb{R}$, we have $\bar{o}(x, \bar{x}) \geqslant o(x)$ for all $(x, \bar{x}) \in \mathbb{R}^2$. Consider now the function

$$g(\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]) = -\frac{1 - \Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))}{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'} + \alpha\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]))}. \tag{E.50}$$

The first-order approximation to $g(\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i])$ around a point $e_{jj'}(z_i)$, for some $z_i \in \mathcal{Z}$, is

$$\bar{g}(\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i], e_{jj'}(z_i)) = \tag{E.51}$$

$$-\frac{1-\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha e_{jj'}(z_i)))}{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha e_{jj'}(z_i)))}+\frac{\alpha}{\omega_{jj'}}\frac{\phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha e_{jj'}(z_i)))}{\Phi(\omega_{jj'}^{-1}(\Delta\kappa_{jj'}+\alpha e_{jj'}(z_i)))^2}(\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]-e_{jj'}(z_i)).$$

As the function $g(x)$ in equation (E.50) is concave in $x \in \mathbb{R}$ for all feasible values of $\omega_{jj'}$, $\Delta\kappa_{jj'}$ and α, we can conclude that

$$\bar{g}(\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i], e_{jj'}(z_i)) \geqslant g(\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]) \tag{E.52}$$

for all values of $\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i]$ and $e_{jj'}(z_i)$. Combining equations (E.49), (E.50) and (E.52), we find

$$\mathbb{E}[(1-d_{ij}) + d_{ij}\bar{g}(\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i], e_{jj'}(z_i))|\mathcal{W}_i] \geqslant 0, \tag{E.53}$$

with $\bar{g}(\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i], e_{jj'}(z_i))$ defined as in equation (E.51).

Step 3. As d_{ij} is measurable in $\mathcal{J}_i$, the function $\bar{g}(\cdot)$ is linear in its first argument, and $\Delta p_{ijj'}$ is a mean-preserving spread of $\mathbb{E}[\Delta p_{ijj'}|\mathcal{J}_i]$, we can write

$$\mathbb{E}[d_{ij}\bar{g}(\mathbb{E}[\Delta p_{ijj'}|\mathcal{J}_i], e_{jj'}(z_i))|\mathcal{J}_i] = \mathbb{E}[d_{ij}\bar{g}(\Delta p_{ijj'}, e_{jj'}(z_i))|\mathcal{J}_i].$$

Using equation (5), we can rewrite this equality as:

$$\mathbb{E}[d_{ij}\bar{g}(\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i], e_{jj'}(z_i))|\mathcal{J}_i] = \mathbb{E}[d_{ij}\bar{g}(\Delta p_{ijj'}, e_{jj'}(z_i))|\mathcal{J}_i].$$

Furthermore, as $\mathcal{W}_i \subseteq \mathcal{J}_i$, applying LIE, we obtain

$$\mathbb{E}[d_{ij}\bar{g}(\mathbb{E}[\Delta p_{ijj'}|\mathcal{W}_i], e_{jj'}(z_i))|\mathcal{W}_i] = \mathbb{E}[d_{ij}\bar{g}(\Delta p_{ijj'}, e_{jj'}(z_i))|\mathcal{W}_i]. \tag{E.54}$$

Step 4. Combining equations (E.53) and (E.54), we rewrite the former as

$$\mathbb{E}[(1-d_{ij}) + d_{ij}\bar{g}(\Delta p_{ijj'}, e_{jj'}(z_i))|\mathcal{W}_i] \geqslant 0. \tag{E.55}$$

Finally, we take an expectation on both sides of this inequality conditional on z_i. If $z_i \subseteq \mathcal{W}_i$,

$$\mathbb{E}[(1-d_{ij}) + d_{ij}\bar{g}(\Delta p_{ijj'}, e_{jj'}(z_i))|z_i] \geqslant 0, \tag{E.56}$$

which coincides with equation (E.45). □

E.2.4 Simulation

Setup. We consider a simulation identical to Section 6.1 except for the distribution of $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3})$. For this simulation, we assume ε_i follows a multivariate normal distribution with mean equal to a 3×1 vector of zeros and variance matrix given by equation (E.30) with $q_{12} = q_{21} = 1$ and $q_{13} = q_{23} = q_{32} = q_{31} = 0$. Thus,

$$\Sigma(\eta) = \begin{bmatrix} 1 & \eta & 0 \\ \eta & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{E.57}$$

where the notation makes explicit the dependency of Σ on the parameter η. In our simulation, we fix $\eta = 0.5$ and, thus, denoting the correlation coefficient between ε_{ij} and $\varepsilon_{ij'}$ as $\rho_{jj'}$, we have $\rho_{12} = 0.5$ and $\rho_{13} = \rho_{23} = 0$. We fix the other model parameters to $\alpha = 0.5$, and $(\kappa_1, \kappa_2, \kappa_3) = (0, 0, -2)$.

MLE: Implementation. The Cholesky decomposition of $\Sigma(\eta)$ results in the matrix

$$L(\eta) = \begin{bmatrix} 1 & 0 & 0 \\ \eta & (1-\eta^2)^{\frac{1}{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{E.58}$$

We can then write $\varepsilon_i(\eta) = L(\eta)\zeta_i$ with $\zeta_i \sim \mathbb{N}(0, I_3)$ and I_3 the 3×3 identity matrix. Denoting as $\zeta_i^r = (\zeta_{i1}^r, \zeta_{i2}^r, \zeta_{i3}^r)$ a draw from the distribution $\mathbb{N}(0, I_3)$, we obtain a draw $\varepsilon_i^r(\eta) = (\varepsilon_{i1}^r(\eta), \varepsilon_{i2}^r(\eta), \varepsilon_{i3}^r(\eta))$ from the distribution $\mathbb{N}(0, \Sigma(\eta))$ with $\varepsilon_{i1}^r(\eta) = \zeta_{i1}^r$, $\varepsilon_{i2}^r(\eta) = \eta\zeta_{i1} + (1-\eta^2)^{\frac{1}{2}}\zeta_{i2}^r$ and $\varepsilon_{i3}^r(\eta) = \zeta_{i3}^r$. We then compute ML estimates of $(\eta, \alpha, \kappa_2, \kappa_3)$ by solving the following problem:

$$\underset{(\psi_\eta, \psi_\alpha, \psi_{\kappa_2}, \psi_{\kappa_3})}{\operatorname{argmax}} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{3} \mathbb{1}\{d_{ij} = 1\} \ln\left(\frac{1}{R} \sum_{r=1}^{R} \frac{\exp((1/\delta)(\psi_{\kappa_j} + \psi_\alpha \mathbb{E}[p_{ij}|z_i] + \varepsilon_{ij}^r(\eta))}{\sum_{j'=1}^{J} \exp((1/\delta)(\psi_{\kappa_{j'}} + \psi_\alpha \mathbb{E}[p_{ij'}|z_i] + \varepsilon_{ij'}^r(\eta))} \right) \right\}.$$

with the normalization $\psi_{\kappa_1} = 1$, a number of simulation draws $R = 1000$, and a smoothing parameter $\delta = 0.2$. In cases 1 to 4 in Table E.2, we set $z_i = x_{2i}$ and, thus, $\mathbb{E}[p_{ij}|z_i] = x_{2ij}$. In case 5 in Table E.2, we set $z_i = p_i$ and, thus, $\mathbb{E}[p_{ij}|z_i] = p_{ij}$.

Moment Inequalities: Implementation. Following the approach described in Section 5, we compute confidence sets for $(\eta, \alpha, \kappa_2, \kappa_3)$ using a finite number of unconditional moment inequalities implied by the conditional ones in equations (E.32) and (E.35). Specifically, for each ordered pair of drugs (j, j') and each instrument function $g_{jj'}^{(s)}(z_i)$ for $s = 1, \ldots, S$, we use the odds-based inequality

$$\mathbb{E}\left[\left((1-d_{ij})\frac{\Phi(\psi_{jj'}^{-1}(\psi_{\kappa_j} - \psi_{\kappa_{j'}} + \psi_\alpha \Delta p_{ijj'}))}{1 - \Phi(\psi_{jj'}^{-1}(\psi_{\kappa_j} - \psi_{\kappa_{j'}} + \psi_\alpha \Delta p_{ijj'}))} - d_{ij}\right) g_{jj'}^{(s)}(z_i)\right] \geqslant 0, \tag{E.59}$$

and the bounding inequality

$$\mathbb{E}\Bigg[\Bigg((1-d_{ij}) + d_{ij}\Bigg(- \frac{1 - \Phi(\psi_{jj'}^{-1}(\psi_{\kappa_j} - \psi_{\kappa_{j'}} + \psi_\alpha e_{jj'}(z_i)))}{\Phi(\psi_{jj'}^{-1}(\psi_{\kappa_j} - \psi_{\kappa_{j'}} + \psi_\alpha e_{jj'}(z_i)))}$$

$$+\frac{\psi_\alpha}{\psi_{jj'}}\frac{\phi(\psi_{jj'}^{-1}(\psi_{\kappa_j}-\psi_{\kappa_{j'}}+\psi_\alpha e_{jj'}(z_i)))}{\Phi(\psi_{jj'}^{-1}(\psi_{\kappa_j}-\psi_{\kappa_{j'}}+\psi_\alpha e_{jj'}(z_i)))^2}(\Delta p_{ijj'}-e_{jj'}(z_i))\Bigg)\Bigg)g_{jj'}^{(s)}(z_i)\Bigg]\geqslant 0, \quad \text{(E.60)}$$

with the normalization $\psi_{\kappa_1}=0$ and

$$\psi_{jj'}=(2(1-\psi_\eta q_{jj'}))^{\frac{1}{2}}. \quad \text{(E.61)}$$

Note that we use ψ to denote the parameter vector $\psi \equiv (\psi_\eta, \psi_\alpha, \psi_{\kappa_2}, \psi_{\kappa_3})$, whose true value is $\psi^* \equiv (\eta, \alpha, \kappa_2, \kappa_3)$. Equations (E.59) and (E.60) correspond to the unconditional version of the conditional moment inequalities in equations (E.32) and (E.35). In our particular simulation, we compute the function $e_{jj'}(\cdot)$ entering equation (E.60) as the predicted value of a spline regression of $\Delta p_{ijj'}$ on z_i estimated with the subset of observations where $d_{ij}=1$. Unless otherwise noted, we set $z_i = x_{2i}$ and, for each ordered pair of alternatives (j, j'), we define $g_{jj'}^{(s)}(z_i)$ for $s = 1, \ldots, S$ as the set of indicator functions that partition the support of $\Delta x_{2ijj'}$ into thirty intervals, with $\Delta x_{2ijj'} = x_{2ij} - x_{2ij'}$.

Results. We present the simulation results in Table E.2. We first discuss the ML estimates. Comparing the results in Table E.2 to those in Table 2 in the main draft, we find that the bias in the maximum likelihood estimates in the multinomial probit model is qualitatively the same as the bias in the maximum likelihood estimates of the same parameters in the multinomial logit model. That is, the ML estimates of the multinomial probit parameters α and κ_3 are consistent if the researcher's assumed information set coincides with the agent's true information set (i.e., if $z_i = x_{2i}$ and $\sigma_1 = 0$, as in cases 1 and 2), and they are biased towards zero if (a) the researcher's assumed information set is a strict subset of the agent's information set (i.e., if $z_i = x_{2i}$ and $\sigma_1 > 0$, as in cases 3 and 4) or if (b) the agent's information set is a strict subset of the researcher's assumed one (i.e., if $z_i = p_i$ and $\sigma_3 > 0$, as in case 5). For the estimate of the parameter η, which equals the correlation in the idiosyncratic preference shocks corresponding to choices 1 and 2, again we find that the maximum likelihood estimate is consistent only in cases 1 and 2, when the researcher's assumed information set equals the agent's true information set. In cases 3 and 4, we find a downward bias in η.[3]

Comparing the results in Table E.2 to those in Table 2 in the main draft, we find that the confidence intervals for the parameters α, κ_2, and κ_3 in the multinomial probit model are larger than the confidence intervals for the same parameters when we assume a multinomial logit model. To understand why we obtain larger confidence intervals in the multiniomial probit model than in

[3]The reason why the estimates of α and η in cases 1 and 2 in Table E.2 do not coincide exactly with their true values is the limited number of normal draws we use to simulate the maximum likelihood function. As discussed above, the estimates in Table E.2 are computed using 1,000 draws. Using a larger number of draws is computationally expensive. Using 750 draws instead, we find estimates of α and η in cases 1 and 2 that equal 0.49 and 0.53, respectively. Using 500 draws, these estimates equal 0.48 and 0.54. Thus, the larger the number of draws, the closer the ML estimates of these parameters are to their true value.

Table E.2: Simulation Results - MLE and Confidence Intervals

Case	σ_1	σ_3	z_i	Estimator	MLE & Confidence Sets α	κ_2	κ_3	η
1	0	0	x_{2i}	MLE	0.49	0	-2	0.52
				Odds-based	[0.35 , 0.50]	[0 , 0]	[-2.00 , -0.80]	[0.35 , 0.65]
				Bounding	[0.35 , 0.50]	[0 , 0]	[-2.00 , -0.80]	[0.35 , 0.65]
				Both	[0.35 , 0.50]	[0 , 0]	[-2.00 , -0.80]	[0.35 , 0.65]
2	0	1	x_{2i}	MLE	0.49	0	-2	0.52
				Odds-based*	[0.35 , 0.50]	[-0.15 , 0.15]	[-2.00 , -0.80]	[0.20 , 0.80]
				Bounding	[0.35 , 0.50]	[0 , 0]	[-2.00 , -0.80]	[0.35 , 0.65]
				Both	[0.35 , 0.50]	[0 , 0]	[-2.00 , -0.80]	[0.35 , 0.65]
3	1	0	x_{2i}	MLE	0.47	0	-1.92	0.48
				Odds-based	[0.35 , 0.50]	[0 , 0]	[-2.00 , -0.80]	[0.20 , 0.65]
				Bounding	[0.35 , 0.50]	[-0.15 , 0.15]	[-2.00 , -0.80]	[0.20 , 0.65]
				Both	[0.35 , 0.50]	[0 , 0]	[-2.00 , -0.80]	[0.20 , 0.65]
4	1	1	x_{2i}	MLE	0.47	0	-1.92	0.48
				Odds-based*	[0.35 , 0.50]	[-0.15 , 0.15]	[-2.00 , -0.80]	[0.05 , 0.80]
				Bounding	[0.35 , 0.50]	[-0.15 , 0.15]	[-2.00 , -0.80]	[0.20 , 0.80]
				Both	[0.35 , 0.50]	[-0.15 , 0.15]	[-2.00 , -0.80]	[0.20 , 0.80]
5	0	1	p_i	MLE	0.45	-0.01	-1.96	0.50
				Odds-based	[0.35 , 0.35]	[0 , 0]	[-2.00 , -0.80]	[0.65 , 0.65]
				Bounding	[0.35 , 0.35]	[0 , 0]	[-2.00 , -0.80]	[0.65 , 0.65]
				Both	[0.35 , 0.35]	[0 , 0]	[-2.00 , -0.80]	[0.65 , 0.65]

Note: σ_1 and σ_3 are parameters of the distributions of x_{1ij} and x_{3ij}, as indicated in footnote 9 in the main text. *MLE* indicates the maximum likelihood estimate. *Odds-based*, *Bounding*, and *Both* contain the projections on each parameter of 95% confidence sets computed according to the procedure in Andrews and Soares (2010). *Odds-based* indicates the corresponding confidence set is computed using only inequalities of the type in equations (E.59); *Bounding* indicates the confidence set is computed using only inequalities of the type in equations (E.60); *Both* indicates the confidence set is computed using inequalities of the type in equations (E.59) and (E.60). In all cases, confidence sets are computed by testing points in a 4-dimensional grid; we mark cases with an asterisk when the confidence set includes points outside the grid.

the multinomial logit model, it is useful to compare the moment inequalities in equations (E.59) and (E.60) to the moment inequalities in equations (25) and (26), respectively.

In this comparison, we can see differences in the functional form for the ratio of the probability of choosing alternative j to the probability of choosing alternative j', which equals $\Phi(\cdot)/(1-\Phi(\cdot))$ in the case of the multinomial probit and $\exp(\cdot)$ in the case of the multinomial logit, as illustrated in equations (E.59) and (25). We also see differences in the functional form of the first-order approximation to this ratio, as illustrated in equations (E.60) and (26). More importantly, a key difference between the moment inequalities used to identify the parameters of the multinomial probit and the inequalities used to identify the parameters of the multinomial logit model is that we replace the dummy variable $d_{ij'}$ entering the moment inequalities used to identify the multinomial logit model (those in equations (25) and (26)) with the dummy variable $1 - d_{ij}$ in the moment inequalities used to identify the multinomial probit model (those in equations (E.59) and (E.60)). Since $1 - d_{ij} \geqslant d_{ij'}$ for every observation i and any two distinct alternatives j and j', this particular difference between the moment inequalities we use in the multinomial probit model and those we use in the multinomial logit model will tend to make the former weaker, resulting in larger confidence sets. This is reflected in the larger confidence intervals reported in Table E.2 relative to those reported in Table 2.

Comparing the results reported in Table E.2 for cases 1 to 4, we observe that the confidence intervals for the parameters α, κ_2, and κ_3 exhibit patterns that are qualitatively similar to those discussed in Section 6 for the multinomial logit model. Specifically, when the moment inequalities we use for estimation are valid (that is, when they use the variable x_{2i} as an instrument, which indeed belongs to the agent's information set), we observe that the moment inequality confidence set, once projected on each of the four parameters to estimate, increases in the variance of the expectational error, σ_3, and in the variance of the variable that belongs to the agent's information set but is not observed by the researcher, σ_1. In particular, larger values of σ_3 affect the confidence set defined by the odds-based moment inequalities, while larger values of σ_1 affect the confidence sets defined by the bounding inequalities. This is illustrated in Table E.2 through the confidence interval for κ_2. The most likely reason why the confidence intervals for α and κ_3 do not vary across cases 1 to 3 is that, for computational reasons, the grid of points we use to compute the confidence sets characterized in Table E.2 is fairly coarse and, as a result, we cannot capture small changes in the confidence set. More specifically, to compute the results reported in Table E.2 we use a grid where the smallest distance between any two values of α, any two values of κ_2, or any two values of η is 0.15. For κ_3, the smallest distance between any two values in the grid is even larger, as it is equal to 0.6.

Comparing the results reported in Table E.2 for case 5, we observe an important difference in the confidence set relative to that reported in Table 2 for the multinomial logit model: when the moment inequalities we use for estimation are not valid (that is, they use the variable p_i as an instrument, which does not belong to the agent's information set when $\sigma_3 > 0$), we observe that the confidence set computed for the multinomial logit model is empty, while the one we obtain for the multinomial probit model is non-empty. Because the moment inequalities we use to identify the parameters of the multinomial probit model are weaker than those we use in the multinomial logit model, they are less useful for detecting cases in which the researcher incorrectly assumes that a variable belongs to the agent's information set.

References

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

_ **and Panle Jia Barwick**, "Inference for Parameter Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 2012, *80* (6), 2805–2826.

_ **and Soonwoo Kwon**, "Misspecified Moment Inequality Models: Inference and Diagnostics," *The Review of Economic Studies*, 2024, *91* (1), 45–76.

Cox, Gregory and Xiaoxia Shi, "Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models," *The Review of Economic Studies*, 2023, *90* (1), 201–228.

Dickstein, Michael J. and Eduardo Morales, "What do Exporters Know?," *The Quarterly Journal of Economics*, 07 2018, *133* (4), 1753–1801.

Holm, Sture, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 1979, *6* (2), 65–70.

Kaido, Hiroaki and Francesca Molinari, "Information Based Inference in Models with Set-Valued Predictions and Misspecification," *mimeo*, January 2024.

Porcher, Charly, Eduardo Morales, and Thomas Fujiwara, "Measuring Information Frictions in Migration Decisions: A Revealed-Preference Approach," *mimeo*, 2025.

Train, Kenneth, "Welfare Calculations in Discrete Choice Models When Anticipated and Experienced Attributes Differ: A Guide with Examples," *Journal of Choice Modelling*, 2015, *16*, 15–22.

Vuong, Quang H., "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 1989, *57* (2), 307–333.