

Web Appendix for Drug Innovations and Welfare Measures Computed from Market Demand: The Case of Anti-Cholesterol Drugs

Abe Dunn

January 26, 2012

I Summary

Section II of the appendix presents a more detailed description of the data and variables used in the paper. Section III provides additional estimation and robustness checks. Section IV shows estimates from the hedonic regression. Section V provides additional discussion of the instrumental variable strategy.

II Data (continued)

Information about medical conditions are obtained from individuals in the survey who are asked to write about their current medical condition and health history, including when their medical problems began. For each medical event (e.g., doctor visit or prescription drug purchase), individuals are asked about the medical conditions that gave rise to the event. Professional coders take the information provided by the individual and assign one of 5 digit ICD-9 codes (International Classification of Disease Code, Ninth Revision), which describe the individual's medical condition. To protect the identity of individuals in the sample, the 5 digit ICD-9 code is aggregated into 3 digit ICD-9 codes. The 5 digit ICD-9 codes are also aggregated into 260 clinically meaningful categories using Clinical Classification Software. In this paper, both the 3 digit ICD-9 codes and clinical classification codes are used to describe an individual's medical condition. After reviewing

risk factors mentioned in the National Cholesterol Education Program (2001) report, the 3 digit ICD-9 and clinical classification codes are placed into four categories: cholesterol disorders, heart disease, diabetes and hypertension.¹ While most of the categories consist of only a single code, I define heart disease very broadly to include many severe conditions, such as stroke, heart attacks and other related conditions associated with the thickening or hardening of arteries. All of these problems listed are chronic conditions, so that once an individual is observed as having the condition, she is assumed to continue to have the condition.²

The MEPS reports whether individuals have medical or prescription drug insurance and the type of plans. If the insurance plan is public, then the data identifies the insurance provider, that is, whether it is Medicare, Medicaid, or some other public agency.³ Unfortunately, the MEPS data

¹The three digit ICD9 and clinical classification codes used for each disease category are the following: Cholesterol Disorder (ICD9: 272), Diabetes (ICD9: 250), Hypertension (ICD9: 401), and Heart Disease (ICD9: 410-414, 433-437, 440, 444, and classification codes: 101, 104, 108, 109, 113, 114, and 116). Dr. Karen Rasmussen assisted in the assignment of these categories.

²There was one change in the survey in 2007 that had a noticeable impact on the reporting of chronic conditions like high cholesterol, heart disease, hypertension, and diabetes. Prior to 2007 the MEPS survey asked individuals to list their conditions, but in 2007 the survey was changed by specifically asking individuals if a physician told them whether they had any of these chronic conditions. That is, they were specifically asked whether a doctor has told them that they have high cholesterol, hypertension, heart disease or diabetes. This change had a measurable impact on the reporting of chronic conditions like high cholesterol, jumping 4 percent to 17.2 percent of the population. This caused an increase in the fraction of individuals reporting high cholesterol and a decrease in the fraction of individuals reporting high cholesterol that were using anti-cholesterol drugs. Although this change does not directly impact the model, it requires a slight modification in the construction of the quality-adjusted price index for 2007. The change in the questioning of respondents in 2007, increased responses related to chronic conditions which causes ξ_{jt} to fall for all drugs, since more individuals report a condition that previously did not. To account for this issue, I fix unobserved quality to 2006 levels by adding 0.46 to ξ_{jt} in 2007, which was the average drop in ξ_{jt} from 2006 to 2007. Similar results are found if a dummy variable for 2007 is included in the second-stage demand model and that dummy value is set to zero for the price index calculation. This has no effect on Model (4) where ξ_{jt} is held fixed.

³Medicare provides medical insurance but no prescription drug insurance until the passage of part D in 2006, whereas Medicaid provides both. The data on private plans includes whether the plan covers doctors visits, prescription drugs, or other services. Additional information about the individual's insurance coverage can be inferred from the individual's medical expenditures. Each time a consumer visits a doctor or purchases medical services such as prescription drugs, the survey records the amount charged and who pays, whether the payment is paid directly by the consumer or paid by a third party. The third party payments are classified as private, Medicare, Medicaid or various other types of public insurances.

does not provide detailed information on the structure of the individual's drug insurance plan. The MEPS only contains information on payments for drugs purchased by an individual, but not on drugs that are not purchased.⁴ For example, if an individual purchases Zocor, the data shows the out-of-pocket cost for Zocor, but not for the other anti-cholesterol drugs that could have been purchased.

The prescription drug transaction data provided in the MEPS includes the quantity of pills, the strength, and the National Drug Code (NDC) of each drug purchased. The NDC code is a number that uniquely identifies a drug and can be used to link the drug to the manufacturer and a specific product. Conversion from the NDC code to a specific product is done using Redbook data that links NDC codes to the products and manufacturer.⁵ In cases where the NDC code of the drug is not listed, I used the name of the drugs active molecule as listed by the pharmacist, and whether the drug is indicated as branded or generic to determine the identity of the anti-cholesterol drug.

The MEPS study supplements the survey data by contacting the individual's medical providers and pharmacies to obtain billing information. For instance, if a patient reports purchasing Zocor from a specific pharmacy, the pharmacy is contacted to provide a payment history for all purchases of Zocor for that individual.

I performed three checks on the MEPS data to determine whether the sample is representative. I compared the number of uninsured in the MEPS to that reported in the Census for 2002 and found that they matched. The Census estimate of the number of uninsured is 45.8 million, while the number from the MEPS is 43.6 million. I also computed the annual estimated national revenue shares in the MEPS for the top three sellers - Lipitor, Pravachol and Zocor - and compared them to the those reported by IMS health, which is a pharmaceutical market research firm that monitors drug sales from pharmacies, for the period 1999-2002. I found the differences between the samples to be relatively small. Finally, I compared the market shares from the MEPS to those from the MarketScan data. The MarketScan data is a convenience sample from commercial insurers and

⁴There is a wide variety of features that an insurance plan might have such as formulary restrictions, deductibles, and copayments that may be fixed or vary across drugs.

⁵One case where a unique drug cannot be assigned to a round is when multiple anti-cholesterol medications are purchased. Since multiple medications are typically not prescribed, it is likely that a patient has switched drugs. So in cases where two drugs are purchased in a round, I assign the last drug taken in the round. I use information on the drug taken in the previous or following round to determine the drug that a person is switching to. If that information is not available I assign the drug with the greatest quantity purchased in the round.

large employers that is not representative, but is much larger, including several million individuals in most years. I found them to have similar trends, despite having different samples.

A Variables (continued)

An individual is classified as having prescription drug coverage (i.e. $DrugIns_{it} = 1$) if she has a private prescription drug insurance (including Medicare Part D) or is on Medicaid. This definition of drug coverage should account for nearly all individuals with drug insurance. To account for the possibility of misreporting, I use prescription drug expenditure information provided by the MEPS to mark individuals as covered if a third party pays for a significant amount of their drug coverage for the year. Specifically, I broaden the definition of those with prescription drug insurance by counting individuals as insured if 70 percent of their drug expenditures are covered by another party. In this data, I find that those with drug insurance pay 28.9 percent of prescription drug expenses out-of-pocket.

The price variable is calculated on an annual basis from transactions involving drug j . The task is complicated by the fact that what is observed in the data are transaction prices that vary by the strength of the drug per tablet, which is measured in milligrams, and the size of the bottle, which is measured in number of tablets. For example, Lipitor is available in strengths of 10mg, 20mg, 40mg and 80mg per tablet, and bottle sizes are typically 30, 60 or 90 tablets. Therefore, the number of prices for Lipitor is the number of strengths available times the number of bottle sizes, which in this case, is 12. In order to compare prices across different drugs and strengths, I choose a single price for each drug-strength category. To calculate price, for each drug-strength combination I run a regression that includes the different quantities of the drug purchased along with year dummies. The price of the drug-strength combination j in period t is the predicted value from these regressions for a bottle containing 30 tablets (the most frequently purchased quantity).

The MEPS data includes price information, but appears to contain imputed prices based on average wholesale prices, especially in the later years of the data. This is problematic since average wholesale prices are typically much higher than the actual price of the drug. As an alternative to using MEPS price information, I estimate prices using MarketScan claims data to impute the price of prescription drugs for the period 1999 to 2007. One advantage of the MarketScan data is that it has a much larger sample size, but more importantly, the MarketScan claims data set is based on adjudicated claims that have more precise information on price and do not contain imputations.

The identical estimation procedure was conducted on the MEPS data and MarketScan claims data for estimating price. I found a systematic bias in the later years of the MEPS data that appears to be related to how they impute prices.⁶ Therefore, for the period from 1999 to 2007 I use estimated prices from MarketScan. The prices from the computations for a selected set of drugs and strengths is reported in Figure 2, presented previously. To match the time frame of the price variables, I estimate the mean utility, δ_{jt} , on an annual basis. An additional reason to estimate these variables on an annual basis is that many insurance plans have open enrollment once a year.

The key dependent variable in the analysis is the choice of drug treatment. I assume that if the individual takes any medication in a period, then she is considered to be using medication in that period. The model does not explicitly address the issue of the quantity of medication consumed during each period. I assume that conditional on purchasing an anti-cholesterol drug in a period, individuals have inelastic demand and are, therefore, unresponsive to price. There are a number of reasons to hold quantity fixed within each period. First, the model already accounts for quantity because individuals are observed over three periods within a year, so I observe whether or not they purchase in each of the three periods. Quantity is also accounted for because I treat each strength of a drug as a distinct product. Second, to the extent that compliance does not vary over time for individuals, this assumption will have little effect on the price index calculation. Finally, one appeal of the proposed approach is that it incorporates a large amount of individual data and is relatively simple to implement; addressing the issue of the quantity of medications purchased may

⁶In the later years of the MEPS data the difference between the price of the branded and generic versions of some drugs was minimal, although the price difference reported by MarketScan for the same two drugs was large. One potential reason for the difference may be that the MEPS uses the average wholesale price of the drug to identify outliers, where outliers are defined as those observations where prices are more than 20 percent below the average wholesale price. However, in many cases drug prices are much different than the average wholesale price, especially for generic drugs. For instance, using the MEPS data I observe that the average wholesale price for 10 mg of generic Zocor in 2007 is \$2.70, but the average transaction price in the MarketScan data for 10 mg of generic Zocor is \$1.81. Therefore, prior to imputation, the generic Zocor prices reported in the MEPS survey would be identified as “outliers”, even when their values are close to the transaction prices reported in the MarketScan data. These “outliers” in the MEPS data would then be replaced with the average wholesale price, which would introduce a very large upward bias on price. As a robustness check, the model has also been estimated excluding the years that incorporate MEPS pricing data from 1996-1998. I obtain qualitatively similar results to those reported for the full sample.

unnecessarily complicate the model.⁷ Within each period I assume that individuals are 75 percent compliant.⁸

Table A1 provides some descriptive statistics on the drugs included in the sample and Table A2 presents the probit estimates that are used to compute the $RiskScore_{it}$ variable.

[Table A1. Drug Characteristics]

[Table A2. Probit of Risk Factors on the Decision to Take A Drug]

III Additional Estimation Results and Robustness Checks

Table A3 presents the estimates of additional coefficients from the first-stage maximum likelihood model that are not reported in Table 3 in the paper. These estimates include the interactions of the different molecules with the key health variables, $High\ Cholesterol_{it}$, $Heart\ Disease_{it}$, and Age_{it} . There are also interactions with these health variables and whether the drug is generic. The left out molecule is Zocor. These interactions indicate that relative to other treatments, it appears that Zocor is preferred by patients for the treatment of heart disease. Also, younger individuals prefer both Lipitor and Crestor, which are the most effective medications for lowering LDL cholesterol on average.

[Table A3. (Table 3. Continued). First-Stage Results from Conditional Logit Estimation]

The first-stage of the IV regression is reported in Table A4. As one would expect, there is a positive and significant relationship between the predicted price and the markup term, and a negative and significant relationship between the markup term and the price for generic drugs. The predicted demand appears to be a less important explanatory variable. Although for branded drugs, higher predicted demand indicates a higher price.

[Table A4. IV Estimation First-Stage]

⁷One might also even view the level of compliance as a distinct decision model. While the doctor talks with individuals periodically about their condition and are involved in the overall decision to prescribe medication, they are not typically reminding individuals to take medication on a daily basis.

⁸The compliance rate observed in my data is around 74 percent on average (measured by number of pills/days in the round). Wosinska (2002) reports that out of a 43 day period, consumers typically miss 13 days of treatment, which is a compliance rate of around 70 percent. I choose a figure slightly higher than these because these compliance rates exclude free samples.

A Robustness Checks

To determine the robustness of the results presented in this paper, Table A5 reports the estimates from four additional models which are compared to the baseline results (Model 1). One concern with the baseline model is that the instrumental variables rely on a markup term that is motivated by a strategic pricing model, although firms could deviate from this pricing behavior for a variety of reasons.⁹ Model 2 explores the robustness of the baseline result by excluding the markup terms and only using the predicted demand and the predicted demand interacted with the generic indicator variable. The results from Model 2 are similar to those of Model 1, although the coefficient on price increases to -1.06. This actually results in the quality-adjusted prices falling more rapidly than the main estimates reported in the paper, as will be shown below. Another potential concern is that using predicted demand based on the demographics of the population may be problematic if demographics are systematically correlated with market prices due to price discrimination. To determine if this is a problem, I employ an alternative set of instruments that does not rely on the demographics of the population. The instrument set attempts to capture factors impacting the competitive environment such as the number of competitors and number of generic competitors in the market along with interactions of these competitive variables and the number of years the product has been in the market and whether the product is generic.¹⁰ Model 3 shows the results using an alternative instrument set, which show a negative and statistically significant coefficient on price that is also smaller in absolute value than the baseline model.

Another potential concern is that the results may be driven by the expansion of the characteristics space due to the inclusion of random tastes for products (i.e. the logit error term). This general problem is outlined in great detail in Akerberg and Rysman (2005) where they discuss how the logit errors may lead to undesirable identification of parameters due to restrictions placed on the model by the logit functional form. The restrictiveness of the logit error term is already addressed in two ways in this paper. First, detailed consumer level information is used to capture the heterogeneity of market demand, which limits the importance of the idiosyncratic logit error term. Second, one

⁹For example, firms may adjust prices due to anticipated entry by competing drugs or by generics. Firms may adjust prices over the life cycle of a product. There is also the possibility of second degree price discrimination across formulations and strengths. There is also the possibility that prices are affected by the countervailing power of insurers.

¹⁰A similar approach is applied in Stern (1996).

of the key problems of the logit framework is that the coefficient on price may be identified without ever observing changes in the price of the products. The inclusion of drug-strength fixed effects, already included in the demand model, partially addresses this issue by ensuring that the price coefficient may only be identified through changes in price and not through a change in the number of products.

To further check the robustness of the results, I follow the suggestion of Akerberg and Rysman (2005) by including the number of product choices on the right hand side of the regression equation. Including the number of products allows for demand to shift among products due to “crowding” of the logit errors. A negative sign on this coefficient indicates that the demand for a product is “crowded out” by the introduction of new products. Since statin drugs may be more innovative than drugs from other classes, I allow for statin drugs to have a different effect on “crowding out” than other drugs. To incorporate this feature in the Akerberg and Rysman framework, I include both the number of drug products and the number of statin drug products on the right hand side. The results show that the introduction of non-statin drug products result in “crowding out” of other products, as indicated by the negative coefficient. However, to calculate the “crowding out” effect from the introduction of statin products, one must add the two coefficients, which sum to zero, indicating that there is no “crowding out” from the introduction of statin drug products.¹¹ Another effect of introducing these terms is that the absolute value on the price coefficient declines. As mentioned previously, a higher value on the price coefficient results in the quality-adjusted prices falling more rapidly.

I’ve left the results with the Akerberg-Rysman correction term as a robustness check, rather than my main specification, for three reasons. First, the inclusion of the drug-strength fixed effects and consumer level data partially address this issue. Second, the inclusion of the Akerberg-Rysman correction term reduces the degrees of freedom for identifying the price coefficient. Third, the baseline model reported in the paper offers more conservative estimates of the changes in the quality-adjusted prices over time.

A potentially important variable that is omitted in the above analysis is advertising to physicians.

¹¹I’ve estimated a similar model including both the number of products and number of products squared for all products and statin drug products. As one would expect, the estimates show that the effects of “crowding out” for non-statin products decline with the number of products, although I still find no evidence of “crowding out” for statins.

While this paper is not focused on the effect of advertising, one may be concerned that its exclusion may cause an omitted variable bias. To determine if this is a potential problem, an alternative model is estimated that uses advertising information available in the MEPS that indicates whether an individual received a free drug sample in a period. (Free drug samples are often provided to physicians to give to their patients and this is an important marketing tool.) Using the MEPS data, an aggregate measure of free samples is constructed for each drug in each year. Including this variable of advertising in the IV regression is highly significant (see Table A5, Model 5); but the price coefficient remains negative, highly significant and similar in magnitude, implying that the inclusion of the advertising variable does not affect the main result of the paper. The advertising variable is potentially endogenous, so I focus on the analysis that excludes advertising information.

Direct-to-Consumer-Advertising (DTCA) data is also not available and may impact consumer demand. However, the inclusion of the trend variable in the analysis should capture the growth in DTCA over the sample period, since research suggests that DTCA has primarily expansionary effects on the use of prescription drugs and has a limited impact on the substitution among drugs (See Zhe Jin and Iizuka (2003) and Zhe Jin and Iizuka (2007)).¹²

[Table A5. Alternative of Estimates of Mean Utility on Price]

The main quality-adjusted price indexes reported in the paper are based on the baseline demand estimates. To explore the robustness of the results to alternative estimates, Table A6 reports the quality-adjusted prices for a range of estimates for the price coefficient α . For both the 95th percentile and 5th percentile estimates for α , the quality-adjusted price falls below the average price and the Laspeyres price index. In addition, the quality-adjusted price falls more rapidly than the baseline model when the Akerberg-Rysman correction term is used, which suggests that the results presented in the paper are relatively conservative and may understate the drop in the quality-adjusted price over the study period.

[Table A6. Alternative Price Index Comparison]

¹²A potential effect of DTCA is that individuals have their cholesterol checked, which changes the information set of individuals (i.e. prior to testing their cholesterol they were not aware of their condition, but after testing they became aware). This would not be reflected in my welfare analysis, since I condition on the health of the individuals in my study. In my analysis I assume that the learning of a health condition is information about the patient and does not affect the attributes of the prescription drug.

IV Hedonic Price Index

The hedonic regression is a frequency weighted regression of price on product characteristics and time dummies. The estimates of the hedonic price regression are reported in Table A7 below. The first column includes only a linear specification for LDL reduction, while Model 2 allows for a nonlinear relationship. Model 2 shows that not only are prices higher on drugs that are more effective at lowering LDL cholesterol, but that each additional unit of effectiveness has a smaller impact on price. Since the nonlinear effect on LDL levels appears to be important and the fit of this model is better, a price index is constructed using the model that allows for nonlinear effects. As an alternative to the hedonic price index, another hedonic index was constructed following the methodology of Pakes (2003) by letting LDL cholesterol have a different relationship with *price* in each year. Those results from the Pakes model are similar to those of Model 1 and Model 2.¹³

[Table A7. Hedonic Price Regression on Log(Price)]

The value of product characteristics may also be recovered from the demand estimates. To recover the effect of product characteristics on demand, I regress the product fixed-effects estimated in the demand model on the product characteristics of the different drugs in the data. Similar to the hedonic regression, these estimates show that the effect of the medication in lowering LDL cholesterol is the key explanatory characteristic, with the other product characteristics having a limited effect on demand. Model 1 suggests that a 10 percentage point reduction in LDL cholesterol would produce \$0.43 ($= 0.07 \cdot 10/1.61$) more in welfare per dose, on average.

[Table A8. Projection of Product Dummies on Product Characteristics]

A Welfare from Statin Drugs and Disease-Specific Price Indexes

To measure the impact of the introduction of statin drugs, we examine the hypothetical scenario of statin drugs entering the market in 1996 or in 2007. While this exercise is purely based on a hypothetical that has not occurred, the impact on welfare from the availability of statins provides a measure of the value of statins for those years. The results of the analysis are reported in Table

¹³To be more precise, the prices estimated from the Pakes hedonic approach are 1996: 1.000; 1997: 0.96; 1998: 0.87 1999: 0.88; 2000: 0.92; 2001: 0.89; 2002: 0.98; 2003: 0.99; 2004: 0.99; 2005: 0.99; 2006: 0.92; 2007: 0.72. One problem with applying this hedonic approach in this setting is that I observe a maximum of 30 products in a year, while Pakes (2003) examines around 200 products.

A9. I find that statins make up a significant fraction of consumer welfare, accounting for 72 percent of total welfare in 1996 and 85 percent of total welfare in 2007.¹⁴ Using the approach outlined in the text, these welfare changes may be translated into a quality-adjusted price reduction of 24 percent in 1996 and a price reduction of 31 percent in 2007.¹⁵ It is also interesting to note that one may find measurable differences in price changes depending on an individual’s health condition. To determine the implied price reduction by disease type, the reduction in price is calculated for groups of individuals with heart disease and those without. The results show that the development of the statin class of drugs was more important for individuals with heart disease in both years, with a quality-adjusted price reduction that is 3.2 percentage points lower for individuals with heart disease in 1996 and 5.2 percentage points lower in 2007.

[Table A9. Contribution of the Statin Class to Welfare]

V Demographics as Instruments in a Linear Demand Model using Micro Data

Using econometric theory Kennan (1989) shows that in some specialized settings observing individual data may be helpful in identifying demand; but more generally, in many commonly observed market scenarios, he shows that researchers cannot assume that micro data solves the endogeneity problem. Many empirical papers have confirmed his result by providing several examples where micro-level information does not correct the price endogeneity problem (e.g. Villas-Boas and Winer (1999), Gaynor and Vogt (2003), Goolsbee and Petrin (2004), and Chintagunta et al (2005)). In cases where micro data does not solve the endogeneity problem, Kennan hints at another potential benefit of using micro data: that the aggregate demographics in the market may be used as an instruments for price when individual demographics are included in the model.¹⁶ However, he

¹⁴These figures are calculated by estimating welfare when the statin drugs are available compared to welfare estimates when the statin drugs are not available.

¹⁵Given the large fraction of welfare attributed to the statin class, one might have expected an even larger price reduction. However, recall that this price index is a conservative estimates using current period prices and product characteristics, rather than base period products and characteristics. The price change may be considerably larger if one were interested in the price change using the base period products consisting of only non-statins.

¹⁶Kennan suggests this in footnote 5 of his paper. Gaynor and Vogt exploit this basic idea, but they provide little additional discussion regarding when or how demographics may be used as an instrument.

does not formally show when or why aggregate demographics are valid instruments. This section presents a brief proof that shows what assumptions are necessary for aggregate demographics to be valid instruments and provides a short discussion of why these instruments are different from those commonly used in the literature.

Rather than focus on a discrete choice model, I focus on a simple linear demand model similar to that analyzed in Kennan. Consider the following demand model:

$$Y_{it} = \alpha p_t + \beta \mathbf{z}_{it} + \omega_{it} \quad (1)$$

where Y_{it} is the quantity demanded for individual i in market t where there are a total of T markets and N consumers in each market. Let p_t be the price of the product in market t that is set simultaneously by a single producer in the market and is an endogenous variable. Let z_{it} be the demographics for individual i in market t (e.g. age, sex, education, income, or size of household). The parameters to be estimated include α and β and the error term is ω_{it} . Finally let the aggregate demographic variable, Z_t , simply be the mean demographics across the N individuals in market t :

$$\mathbf{Z}_t = \frac{\sum_{i=1}^{N_t} \mathbf{z}_{it}}{N_t}$$

Alternatively, to construct an instrument analogous to the discrete choice estimation, in a linear

model one may use the aggregate average utility as an instrument, $Z_t = \frac{\sum_{i=1}^{N_t} \beta \mathbf{z}_{it}}{N_t}$.¹⁷

There are three necessary assumptions for the variable Z_t to be a valid instrument. First, the individual's own demographics need to be exogenous so that $E(\mathbf{z}_{it}\omega_{it}) = 0$. Note that this assumption is also necessary for the demographics to enter the above model and is a commonly made assumption in most demand models. Second, the individual's own demand for a product cannot be dependent on the demographics of other individuals in the market. So for individual $j \neq i$ it is assumed that $E(\mathbf{z}_{it}\omega_{jt}) = 0$. This second assumption should hold if there are no network effects that might cause the demographics of other individuals in the market to affect one's own utility. In addition to the above assumption, it is also necessary for the instrument to be correlated with the market price of the good, p_t , which should occur if the demographics have an impact on

¹⁷The effect of demographics on individual utility, $\beta \mathbf{z}_{it}$, may be estimated by using a first stage estimation of 1 that includes period t dummy variables, δ_t , instead of p_t .

demand and there is variation in demographics across markets. Given these three assumptions the consistency of an estimator that uses \mathbf{Z}_t as an instrument follows a simple proof, since it is only

necessary to show that $plim_{NT \rightarrow \infty} \frac{\sum_{t=1}^T \omega_{it} \mathbf{z}_t}{NT} \rightarrow 0$.

proof. To show that $plim_{NT \rightarrow \infty} \frac{\sum_{t=1}^T \omega_{it} \mathbf{z}_t}{NT} \rightarrow 0$, I begin by substituting the value of \mathbf{Z}_t and expanding the summation.

$$\begin{aligned}
 plim_{NT \rightarrow \infty} \frac{\sum_{t=1}^T \omega_{it} \mathbf{z}_t}{NT} &= plim_{NT \rightarrow \infty} \frac{\sum_{t=1}^T \omega_{it} \left(\frac{\sum_{i=1}^N \mathbf{z}_{it}}{N} \right)}{NT} = plim_{NT \rightarrow \infty} \frac{\sum_{t=1}^T \omega_{it} \left(\sum_{i=1}^N \mathbf{z}_{it} \right)}{N^2 T} \\
 &= plim_{NT \rightarrow \infty} \frac{\sum_{t=1}^T \omega_{it} \mathbf{z}_{it} + \omega_{it} \left(\sum_{j \neq i}^N \mathbf{z}_{jt} \right)}{N^2 T} \\
 &= plim_{NT \rightarrow \infty} \frac{\sum_{t=1}^T \omega_{it} \mathbf{z}_{it}}{N^2 T} + plim_{NT \rightarrow \infty} \frac{\sum_{t=1}^T \omega_{it} \left(\sum_{j \neq i}^N \mathbf{z}_{jt} \right)}{N^2 T}
 \end{aligned}$$

Given that it is assumed that $E(\mathbf{z}_{it} \omega_{it}) = 0$ and for individual $j \neq i$ it is also assumed $E(\mathbf{z}_{it} \omega_{jt}) = 0$,

then it follows that $plim \frac{\sum_{t=1}^T \omega_{it} \mathbf{z}_t}{NT} \rightarrow 0$.

This instrumenting strategy relies on an asymmetry that is present in many markets. That is, producers price to the entire market, so that only market wide demographics are relevant, not the demographics of a particular individual. In contrast, an individual's consumption decision is affected only by her characteristics and is unaffected by the characteristics of other consumers in the market.¹⁸

This approach may be compared to the IV strategy of using rival product characteristics to instrument for price because they are both related to consumer preferences rather than shifts in underlying costs. However, it is important to note that instruments using demographics are different in at least two important ways: First, demographics are a valid instrument if the unobserved

¹⁸Of course, this instrumenting strategy fails if producers can perfectly price discriminate, since each individual would have a unique price.

demand shock have a proportional effect on the population and firms are unable to alter products in a way to target a particular demographic (i.e. these instruments work well if firms have a limited ability to target certain demographic populations). Second, a bias typically arises because there is an unobserved attribute that the econometrician does not observe that is correlated with price. One advantage of having consumer level data is that the consumers do observe the unobserved demand changes, even if the econometrician does not. Therefore, the econometrician may allow the consumers to respond to unobserved attributes which may correct for a potential bias.

To see how consumer data may be used to correct for a potential bias, consider a hypothetical market for breakfast cereals where there is only a single producer and there are two versions of the cereal, one version that appeals to older adults and one that appeals to younger adults. Also suppose that a manufacturer only distributes one version of the cereal in each market and that the econometrician does not observe which version is introduced. In this case, one should expect the cereal manufacturer to distribute the version that appeals to older adults in markets where the average age is higher. Since the econometrician does not observe the version of the product, the assumption $E(\mathbf{z}_{it}\omega_{jt})$ will be violated.¹⁹ If individual information is available, one can allow the individual's value of the product to be different across markets. That is, estimate a model that allows for demographic information to have a unique effect in each market, so that individuals can flexibly respond to the quality of the product that is available in their market. This implies allowing age to have a unique effect in each market. In this case, even though the econometrician does not observe which version of the breakfast cereal was introduced in the market, it is sufficient that the individual does.

While this particular fix may address the problem that unobserved product characteristics may vary across markets and be correlated with the demographics, the data requirements are significantly greater and may also be computationally burdensome, depending on the number of markets. More practically, even if it is not feasible to allow demographics to have a unique effect in each market, the econometrician should allow for some flexibility in how consumers respond to product characteristics across markets to reduce the potential for endogeneity bias. For example, the econometrician may allow for a different parameter, β , in markets where the average age is older. The econometrician may also allow for greater flexibility in how consumer demographics are interacted with the product

¹⁹This problem is analogous to problems that arise when one assumes that rival product characteristics are exogenous, when they may, in fact, be chosen strategically.

characteristics. To allow for greater flexibility in the anti-cholesterol drug demand model in this paper, I interact each individual's health conditions with molecule dummies to allow for a distinct reaction to each molecule that depends on the health conditions of the individual.

Aggregating over individual demand predictions produces valid instruments in the drug choice model, but it is instructive to observe that population characteristics may produce bias demand estimates if individual information did not enter the demand model. Absent the inclusion of individual level data, any average change in utility caused by a change in the population's mean age or health condition could potentially enter the error term, ξ_{jt} . Therefore, if demographics are used as an instrument when the micro demographic information is not included in the model, it is possible that both ξ_{jt} and p_{jt} will be correlated with the instruments and the estimates would be bias. However, when micro data is included in the first stage of the estimation, then by construction, $\beta_1(\mathbf{z}_{it})\mathbf{x}_{jt}$, should be uncorrelated with the error term, ξ_{jt} .

References

- [1] Iizuka, Toshiaki and Ginger Jin, (2005), "The Effects of Prescription Drug Advertising on Doctor Visits" , *Journal of Economics & Management Strategy*, 14(3) pgs 701-727.
- [2] Iizuka, Toshiaki and Ginger Jin, (2007), "Direct to Consumer Advertising and Prescription Choice" , *Journal of Industrial Economics*, 55(4) pgs 771-771.
- [3] Kennan, John, (1989), "Simultaneous Equation Bias in Disaggregated Econometric Models" , *Review of Economic Studies*, 56(1) pgs 151-156.
- [4] Pakes, Ariel, (2003), "A Reconsideration of Hedonic Price Indexes with an Application to PC's" , *American Economic Review*, 93(5) pgs 1578-1596.

VI Tables

Table A1. Drug Characteristics

BRAND NAME	ACTIVE INGREDIENT	PATENT HOLDER	DRUG CLASS	STRENGTH IN MG	APPROVAL DATE	In Sample	TOT CHOL	LDL	HDL	TRI
LIPITOR	ATORVASTATIN	PFIZER	STATIN	10	12/17/1996	Yes	-29	-39	6	-19
LIPITOR	ATORVASTATIN	PFIZER	STATIN	20	12/17/1996	Yes	-33	-43	9	-26
LIPITOR	ATORVASTATIN	PFIZER	STATIN	40	12/17/1996	Yes	-37	-50	6	-29
LIPITOR	ATORVASTATIN	PFIZER	STATIN	80	12/17/1996	No	-45	-60	5	-37
BAYCOL	CERIVASTATIN	BAYER PHARMS	STATIN	0.3	6/26/1997	Yes	-22	-31	8	-16
BAYCOL	CERIVASTATIN	BAYER PHARMS	STATIN	0.4	6/26/1997	Yes	-24	-34	7	-16
BAYCOL	CERIVASTATIN	BAYER PHARMS	STATIN	0.8	6/26/1997	No	-30	-42	9	-22
LESCOL	FLUVASTATIN	NOVARTIS	STATIN	20	12/31/1993	Yes	-17	-22	3	-12
LESCOL	FLUVASTATIN	NOVARTIS	STATIN	40	12/31/1993	Yes	-19	-25	4	-14
LESCOL	FLUVASTATIN	NOVARTIS	STATIN	80	12/31/1993	No	-25	-36	6	-18
LOVASTATIN	LOVASTATIN	GENERIC	STATIN	10	12/17/2001	No	-16	-21	5	-10
LOVASTATIN	LOVASTATIN	GENERIC	STATIN	20	12/17/2001	Yes	-17	-24	7	-10
LOVASTATIN	LOVASTATIN	GENERIC	STATIN	40	12/17/2001	Yes	-22	-30	7	-14
MEVACOR	LOVASTATIN	MERCK	STATIN	10	8/31/1987	No	-16	-21	5	-10
MEVACOR	LOVASTATIN	MERCK	STATIN	20	8/31/1987	Yes	-17	-24	7	-10
MEVACOR	LOVASTATIN	MERCK	STATIN	40	8/31/1987	Yes	-22	-30	7	-14
MEVACOR	LOVASTATIN	MERCK	STATIN	60	8/31/1987	No	-29	-40	10	-19
PRAVOCOL	PRAVASTATIN	BRISTOL-MEYER SQUIB	STATIN	10	10/31/1991	No	-16	-22	7	-15
PRAVOCOL	PRAVASTATIN	BRISTOL-MEYER SQUIB	STATIN	20	10/31/1991	Yes	-24	-32	2	-11
PRAVOCOL	PRAVASTATIN	BRISTOL-MEYER SQUIB	STATIN	40	10/31/1991	Yes	-25	-34	12	-24
PRAVOCOL	PRAVASTATIN	BRISTOL-MEYER SQUIB	STATIN	80	10/31/1991	No	-27	-37	3	-19
PRAVASTATIN	PRAVASTATIN	GENERIC	STATIN	10	4/24/2006	No	-16	-22	7	-15
PRAVASTATIN	PRAVASTATIN	GENERIC	STATIN	20	4/24/2006	Yes	-24	-32	2	-11
PRAVASTATIN	PRAVASTATIN	GENERIC	STATIN	40	4/24/2006	Yes	-25	-34	12	-24
PRAVASTATIN	PRAVASTATIN	GENERIC	STATIN	80	4/24/2006	No	-27	-37	3	-19
CRESTOR	ROSUVASTATIN	ASTRAZENECA	STATIN	20	8/12/2003	Yes	-40	-55	8	-23
CRESTOR	ROSUVASTATIN	ASTRAZENECA	STATIN	40	8/12/2003	No	-46	-63	10	-28
CRESTOR	ROSUVASTATIN	ASTRAZENECA	STATIN	5	8/12/2003	No	-33	-44	13	-35
CRESTOR	ROSUVASTATIN	ASTRAZENECA	STATIN	10	8/12/2003	Yes	-36	-52	14	-10
ZOCOR	SIMVASTATIN	MERCK	STATIN	10	12/23/1991	Yes	-23	-30	12	-15
ZOCOR	SIMVASTATIN	MERCK	STATIN	20	12/23/1991	Yes	-28	-38	8	-19
ZOCOR	SIMVASTATIN	MERCK	STATIN	40	12/23/1991	Yes	-31	-41	9	-18
ZOCOR	SIMVASTATIN	MERCK	STATIN	5	12/23/1991	No	-19	-25	10	-12
ZOCOR	SIMVASTATIN	MERCK	STATIN	80	12/23/1991	No	-36	-47	8	-24
SIMVASTATIN	SIMVASTATIN	GENERIC	STATIN	10	12/20/2006	Yes	-23	-30	12	-15
SIMVASTATIN	SIMVASTATIN	GENERIC	STATIN	20	12/20/2006	Yes	-28	-38	8	-19
SIMVASTATIN	SIMVASTATIN	GENERIC	STATIN	40	12/20/2006	Yes	-31	-41	9	-18
SIMVASTATIN	SIMVASTATIN	GENERIC	STATIN	5	12/20/2006	No	-19	-25	10	-12
SIMVASTATIN	SIMVASTATIN	GENERIC	STATIN	80	12/20/2006	No	-36	-47	8	-24
VYTORIN	EZTIMBE/SIMVASTATIN	MSP SINGAPORE CO LLC	COMBO	10	7/23/2004	No	-31	-45	8	-23
VYTORIN	EZTIMBE/SIMVASTATIN	MSP SINGAPORE CO LLC	COMBO	20	7/23/2004	Yes	-36	-52	10	-24
VYTORIN	EZTIMBE/SIMVASTATIN	MSP SINGAPORE CO LLC	COMBO	40	7/23/2004	Yes	-39	-55	6	-23
VYTORIN	EZTIMBE/SIMVASTATIN	MSP SINGAPORE CO LLC	COMBO	80	7/23/2004	No	-43	-60	6	-31
ADVICOR	LOVASTATIN/NIACIN	ABBOTT LABORATORIES	COMBO	20	12/17/2001	Yes	-21	-30	20	-32
GENERIC*	FENOFIBRATE	GENERIC	FIBRIC ACID	145	12/31/1993	Yes	-18	-21	11	-29
GENERIC*	GEMFIBROZIL	GENERIC	FIBRIC ACID	600	9/29/1995	Yes	-18	-21	11	-29
LOPID*	GEMFIBROZIL	PFIZER	FIBRIC ACID	600	12/21/1981	Yes	-18	-21	11	-29
GENERIC	NIACIN	GENERIC	NICOTINIC ACID	500	1/1/1970	Yes	-5	-9	15	-11
NIASPAN	NIACIN	KOS	NICOTINIC ACID	500	7/28/1997	Yes	-5	-9	15	-11
ZETIA	EZTIMBE	MSP SINGAPORE CO LLC	OTHER	10	10/25/2002	Yes	-14	-20	4	-5
GENERIC	CHOLESTYRAMINE	GENERIC	SEQUESTRANTS	5.7	8/3/1973	Yes	-17	-23	4	0

Notes: The drug effectiveness information comes from randomized clinical studies reported on the drug labels. The selection of the efficacy measure is challenging, since there is no study that measures these efficacy levels under similar conditions for all medication. Results may vary considerably across studies depending on the details of the study (e.g. the number of people in the study, the age of the participants, or the overall health of the participants). Information on Gemfibrozil was not available from drug labels, so I apply the same effectiveness measures as for the other fibric acid derivatives. I checked that the values reported here fall roughly in the range of effectiveness stated in the National Cholesterol Education Program (2001) for each of the respective drug classes. The main demand estimates and quality-adjusted prices do not rely on these measures, since the demand model includes product-specific fixed effects. However, the hedonic methodology does rely on these measures.

Table A2. Probit of the Decision to Take An Anti-Cholesterol Drug on Risk Factors

	dF/dx	z-stat
Has High Cholesterol	0.506	(65.85)
Atherosclerotic Condition	0.073	(10.02)
Has Diabetes	0.057	(8.63)
Has Hypertension	0.047	(8)
Age \geq 40	0.042	(2.24)
Age	0.020	(3.66)
Age ²	0.000	(0.89)
Age ²	0.000	(-4.1)
Male	0.130	(4.87)
Age*Male	-0.001	(-2.75)
Perceived Health is Good	-0.032	(-4.26)
Number of Observations	106,510	
Pseudo R ²	0.187	

Notes: Reported Z-statistics are based on robust standard errors clustered by individual.

Table A3. (Table 3. Continued). First-Stage Results from Conditional Logit Estimation

Variable	Coef.	z-stat
Heart Disease*Lipitor	-0.179	(-3.27)
Heart Disease*Baycol	-0.851	(-3.45)
Heart Disease*Lescol	-0.438	(-3.6)
Heart Disease*Mevacor (or Generic Mev)	-0.415	(-4.15)
Heart Disease*Pravachol (or Generic Prav)	-0.105	(-1.29)
Heart Disease*Crestor	-0.088	(-0.72)
Heart Disease*Advicor	-0.233	(-0.56)
Heart Disease*Fibric Acid Derivative	-0.349	(-3.37)
Heart Disease*Nictonic Acid	0.001	(0)
Heart Disease*Sequestrant	-0.775	(-3.25)
Heart Disease*Zetia	-0.278	(-1.75)
Heart Disease*Generic	-0.079	(-0.98)
High Chol*Lipitor	0.181	(2.12)
High Chol*Baycol	-0.511	(-1.59)
High Chol*Lescol	-0.003	(-0.02)
High Chol*Mevacor (or Generic Mev)	0.087	(0.57)
High Chol*Pravachol (or Generic Prav)	-0.003	(-0.03)
High Chol*Crestor	0.316	(1.47)
High Chol*Advicor	-0.392	(-0.74)
High Chol*Fibric Acid Derivative	-0.034	(-0.21)
High Chol*Nictonic Acid	-0.162	(-0.65)
High Chol*Sequestrant	0.138	(0.39)
High Chol*Zetia	0.055	(0.21)
High Chol*Generic	0.231	(1.53)
Age*Lipitor	-0.006	(-3.19)
Age*Baycol	0.010	(1.39)
Age*Lescol	0.013	(2.81)
Age*Mevacor (or Generic Mev)	-0.002	(-0.69)
Age*Pravachol (or Generic Prav)	-0.002	(-0.67)
Age*Crestor	-0.009	(-2.12)
Age*Advicor	-0.026	(-2.07)
Age*Fibric Acid Derivative	-0.025	(-7.09)
Age*Nictonic Acid	-0.008	(-1.18)
Age*Sequestrant	0.010	(1.31)
Age*Zetia	0.010	(1.97)
Age*Generic	0.006	(2.06)
Number of Observations	106,510	
Pseudo R ²	0.444	

Notes: Reported Z-statistics are based on robust standard errors clustered by individual. Each variable reported here is relative to the no-drug treatment option that has a utility normalized to zero.

Table A4. IV Estimation First-Stage

Variable	Coef.	t-stat
Markup	12.86	(2.31)
Markup*Generic	-256.09	(-2.93)
Demand	0.00	(2.7)
Demand*Generic	0.00	(0.67)
Number of Observations	266	
R ²	0.933	

Table A5. Alternative Second-Stage Demand Estimates

Variable	Model 1: Baseline		Model 2: Two Instruments		Model 3: Alternative Instruments		Model 4: Akerberg- Rysman		Model 5: Advertising	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Price	-1.613	(-4.01)	-1.068	(-2.53)	-0.994	(-2.1)	-0.937	(-2.39)	-1.369	(-3.83)
Number of Products							-0.167	(-2.07)		
Number of Statin Products							0.157	(1.53)		
Advertising Proxy									12.648	(6.76)
Product Fixed Effects	Yes		Yes		Yes		Yes		Yes	
Number of Observations	266		266		266		266		266	
R ²	0.46		0.55		0.46		0.59		0.57	
Instruments	Demand, Demand*Generic, Markup, Markup*Generic		Demand, Demand*Generic		# of Competitors (branded and generic) and Molecule Age Interactions w/ # of Competitors		Demand, Demand*Generic, Markup, Markup*Generic		Demand, Demand*Generic, Markup, Markup*Generic	

Notes: The estimates of mean utility on price are based on the 266 product-year observations. Model 1 is the main IV specification reported in Table 4 of the text.

Table A6. Alternative Price Index Comparison

Year	Avg Price	Laspeyres	Hedonic	Quality-Adjusted Price Indexes for Different Values of Alpha (π)			
				Baseline	5th percentile	95th percentile	Akerberg-Rysman Demand
				$\pi = -1.61$	$\pi = -2.40$	$\pi = -.82$	$\pi = -0.94$
1996	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1997	1.07	0.99	0.99	1.03	1.02	1.06	1.05
1998	1.04	0.98	0.92	1.03	1.02	1.07	1.06
1999	1.08	1.01	0.94	0.95	0.96	0.88	0.90
2000	1.18	1.01	0.96	0.93	0.96	0.85	0.87
2001	1.17	1.00	0.95	0.90	0.94	0.79	0.82
2002	1.29	1.04	1.01	0.92	0.95	0.82	0.84
2003	1.33	1.07	1.04	0.90	0.94	0.79	0.82
2004	1.34	1.07	1.04	0.90	0.93	0.77	0.80
2005	1.37	1.09	1.04	0.91	0.95	0.81	0.84
2006	1.29	1.08	0.98	0.90	0.93	0.77	0.81
2007	1.00	0.88	0.67	0.73	0.83	0.40	0.49

Notes: The quality-adjusted price index is calculated as described in the text. Prices are normalized to 2007 dollars using the CPI. All quality-adjusted prices included in this table remove the effects of drug insurance (i.e. no moral hazard). The hedonic regression estimate used to calculate the hedonic price index is reported in Table A7. The Laspeyres index follows the BLS methodology where generics and Branded versions of the same molecule are treated as an identical product and the price index is computed using a geometric mean.

Table A7. Hedonic Price Regression on Log(Price)

	Model 1		Model 2	
	Coef.	t-stat	Coef.	t-stat
LDL Cholesterol Reduction (% Amount)	0.032	(9.12)	0.117	(4.12)
LDL Cholesterol Reduction (% Amount) ²			-0.001	(-2.73)
HDL Cholesterol Reduction	-0.004	(-0.36)	-0.008	(-0.63)
Trigliciride Reduction Levels	0.000	(-0.03)	-0.004	(-1.07)
STATIN	0.252	(1.05)	-0.079	(-0.35)
YEAR 1997	0.008	(1.07)	-0.006	(-0.84)
YEAR 1998	-0.060	(-4.24)	-0.089	(-6.94)
YEAR 1999	-0.030	(-1.46)	-0.061	(-3.27)
YEAR 2000	0.000	(0.01)	-0.040	(-2.56)
YEAR 2001	-0.015	(-0.72)	-0.056	(-3.19)
YEAR 2002	0.054	(2.85)	0.014	(0.97)
YEAR 2003	0.080	(4.02)	0.042	(3.04)
YEAR 2004	0.072	(3.17)	0.041	(2.77)
YEAR 2005	0.053	(2.09)	0.035	(1.91)
YEAR 2006	-0.011	(-0.4)	-0.016	(-0.76)
YEAR 2007	-0.388	(-14.37)	-0.394	(-18.91)
Constant	-0.504	(-2.9)	-1.668	(-3.29)
		Frequency weighted		Frequency weighted
Number of Observations		266		266
Adj. R ²		0.419		0.455

Notes: The above estimates are a frequency weighted regression of price on product characteristics and time dummies. T-statistics are based on standard error estimates clustered by year.

Table A8. Regression of Product Dummies on Product Characteristics

	Model 1			Model 2		
	Coef.	t-stat	t	Coef.	t-stat	t
LDL Cholesterol Reduction (% Amount)	0.070	(1.87)	1.87	0.266	(1.61)	1.61
LDL Cholesterol Reduction (% Amount) ²				-0.003	(-1.2)	-1.2
HDL Cholesterol Reduction	0.058	(0.61)	0.61	0.034	(0.35)	0.35
Triglyceride Reduction Levels	0.093	(1.79)	1.79	0.073	(1.36)	1.36
STATIN	-0.166	(-0.16)	-0.16	-1.392	(-0.71)	-0.71
Number of Observations		33			33	
Adj. R-Squared		0.274			0.285	

Notes: The dependent variable is taken from the product fixed effects from Table 4.

Table A9. Contribution of the Statin Class to Welfare

	1996	2007
Welfare from the Availability of Anti-cholesterol Drugs (in billions 2007 \$)	\$1.2	\$7.7
Welfare from the Availability of Statin Drugs (in billions 2007 \$)	\$0.9	\$6.5
Fraction of Welfare From Statins	71.9%	84.5%
<u>Hypothetical Reduction in Quality-Adjusted Price From Introduction of Statins:</u>		
Overall Decrease	-23.9%	-31.4%
Decrease for Those with Heart Disease	-26.1%	-34.6%
Decrease for Those without Heart Disease	-22.9%	-29.4%

Notes: Welfare and price index changes follow the methodology applied in Model 2 of Table 6 and are calculated assuming that remove the effects of drug insurance on demand.