In this analysis, I will be exploring procedures for estimating demand functions. A demand function shows the causal relationship between the quantity demanded for a product and various independent variables (i.e. factors which are believed to influence demand). More specifically, I will be exploring the causal relationship between price changes for a given product and its impact on demand.

Economist theorize that holding everything else constant, when the price of a product falls, the quantity demanded of the product will increase. While, when the price of a product rises, the quantity demanded of the product will decrease. This means that there is an inverse relationship between price and quantity demanded;

*P* ↓ ⇒ *QD* ↑

*P* ↑ ⇒ *QD* ↓

Why might demand be downward sloping? There are a few reasons: (1) Intuitive reasoning; people would like to buy more units at lower prices and vice versa. (2) Diminishing marginal utility; consumer gets less and less additional satisfaction from consuming equal additional units of the same good. (3) Substitution effect; if the price of one good falls relative to its substitutes, then consumers will buy more of it. (4) Income effect; as the price of a good falls, the consumer can afford more units of that good, therefore their *purchasing power* increases.

Economists use *price elasticity* of demand to measure the responsiveness, or elasticity, of the quantity demanded of a good or service to a change in its price when nothing but the price changes. Therefore, the *price elasticity* of demand can be measured by;

*[1.0]*

Since the *price elasticity* of demand is the percentage change in quantity demanded for a percentage change in price, above can equivalently be expressed as the slope of the relationship between the natural logs of *QD* and *P*. That is;

*[1.1]*

*Price elasticity* of demand will always be a negative number. As price increases, quantity demanded decreases and as price decreases, quantity demanded increases.

*Elastic* results occur when the percentage change in quantity demanded exceeds the percentage change in price. so that the *price elasticity* is greater than 1 in absolute terms;

*%*Δ*QD* > *%*Δ*P*  ⇒ || > 1

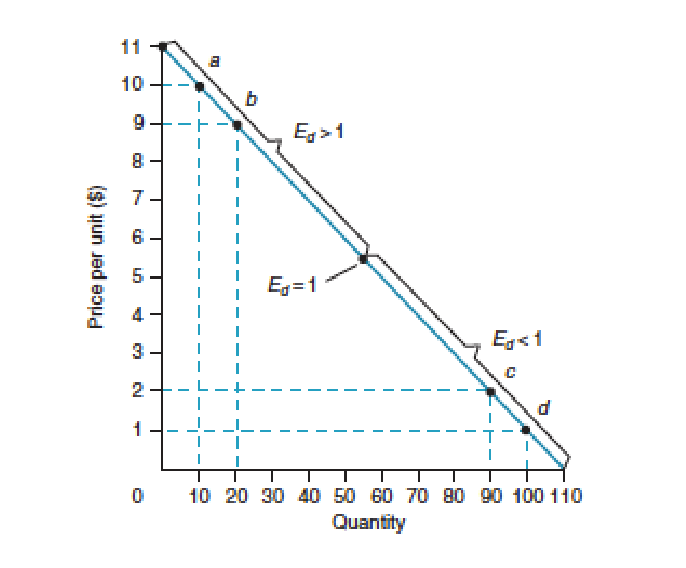
*Inelastic* results occur when the percentage change in quantity demanded is less than the percentage change in prices so that the *price elasticity* is less than 1 in absolute value;

*%*Δ*QD* < *%*Δ*P*  ⇒ || < 1

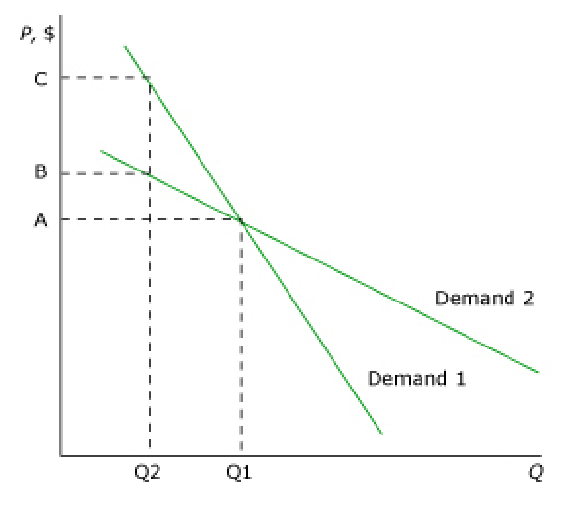
Some examples of *inelastic* goods are gas, salt, bread, tap water (necessary items), or cigarettes (addictive goods). Examples of *elastic* goods are shell gas (because consumers will shop at another gas stations), Apple (specific brands of some good), or a vacation plane ticket.

The determinants of demand elasticity are: (1) substitutability; products with close substitutes have more elastic demand (e.g. price of orange juice goes up, consumers substitute apple juice, cranberry juice, etc.). (2) generality; price elasticity for narrowly defined specific brands of products is greater than that for broadly defined product categories (e.g. demand elasticity of coke > demand elasticity of colas > demand elasticity of soft drinks). (3) proportion of income; the higher the price of the good relative to income, the greater the elasticity and vice versa. (4) degree of need; price elasticity of luxury goods is greater than that of necessity goods (e.g. demand elasticity for vitamins > demand elasticity for insulin). (5) complementarity; products with many complements have less elastic demand. (6) time; the more time that passes, the greater the demand elasticity of the product (e.g. demand elasticity for gas in the long run > demand elasticity for gas in the short run).

Something that might not be as obvious is that price elasticity of demand must be measured at a particular point on the demand curve;



Looking at a linear demand curve, as we move along the curve, is constant, but *P* and *Q* will change. Note, that the steeper the demand curve, the more inelastic the demand for the good. The flatter the demand curve, the more elastic the demand for the good. Thus, demand 2 (from below) is relatively more elastic than demand 1;



One reason why *elasticity* is important is that it tells us how much revenue changes as you change price. Suppose there is elastic demand (*|| > 1*); Therefore, when price increases revenue decreases (decrease in *Q* is bigger than increase in *P*). However, when price decreases revenue increases (increase in *Q* is bigger than decrease in *P*). Suppose now that there is inelastic demand (*|| < 1*); Therefore, when price increases revenue increases (decrease in *Q* is smaller than increase in *P*). However, when price decreases revenue decreases (increase in *Q* is smaller than decrease in *P*).

For the remainder of this analysis I will be covering the *logit model* for estimating demand. I will then be discussing the use of *Instrumental variables* (*IV*) to deal with the endogeneity problems in our demand function– namely, when the price regressor is correlated with the error term (*ε*).

The reason why I have decided to use the *logit model* in estimating demand is because it is the most popular model by economic researchers. The *logit model* is the most popular model because the formula for the choice probabilities takes a closed form and is easily interpretable.

I will begin by first expressing that a consumer’s indirect utility function is assumed to have the form;

*[2.0]*  – + +

Where each consumer *i* can choose from *j* products and where markets are indexed by *t* and *pjt* denotes the price of the product. Thus, each consumer chooses one of the following *J + 1* mutually exclusive alternative products indexed by *j = 0, 1, …, J*. The *0* choice is called the *outside option*—it means that you choose none of the above. Each *J* alternative is associated with a *K x 1* vector of observed characteristics (*xjt* = [*xjt1… xjtK*]). Each product-market *jt* is also associated with some unobserved product characteristics, *ξjt*. The idiosyncratic heterogeneity vector, *εit*, is distributed *iid* across *i* and *t*.

The *ε’*s are distributed according to the extreme value type 2 distribution;

*F(ε)*

This is a very helpful assumption, as it allows for the aggregate shares to have an analytical form. Thus, the type 2 extreme value distribution gives me the market share for product *j* in market *t*;

*[2.1] sjt*()

Where *sjt* is the market share for each product-market *jt*. I can derive cross price derivatives as *sj \* sk*, and own price derivative as *sj \** (*1 - sj*).

If I normalize the mean utility associated with the outside option to 0, I get;

Then the probability of choosing the outside good is;

*[2.2] sjt*()

And the probability of choosing any other good *j* is;

*[2.3] sjt*()

By taking the log odds ratio, I can get the following linear equation;

*[2.4] –*  – +

Therefore, I can estimate this via linear regression.

As mentioned earlier, *ξjt* are unobserved product-market level characteristics in the model. This means that there could be a list of product attributes that are unobserved in the estimation of my model (e.g. quality or prestige characteristics). For example, there can be a positive association found between price and quality that could easily be a result of unobserved self-selection. In other words, individuals who are particularly prone to quality of healthcare may be especially likely to enroll in a health plan (regardless of the price). *ξjt* can also be in the form of measurement error in prices (often prices are averages). Since the unobserved characteristics are not independent from price, most often prices are endogenous.

The following regression example clearly identify where the endogeneity in *Price* can exist. Suppose;

*[3.0] 0 + 1X +*

where *X* is the predictor variable *Price* and *Y* is demand.If I manipulate equation [3.0] by using covariance algebra (take covariance *X* throughout the equation), I get the following;

*[3.1] cov(, X) cov(0 + 1X + , X)*

*. 1cov(X, X) + cov(, X)*

*. 1var(X) + cov(, X)*

*σyx 1 + σεx*

Now divide the population variance of the question predictor *X* throughout equation [3.1];

*[3.2] 1 +*

The ratio in equation [3.2] will give me the size of the residuals that leads to bias induced upon the *1* coefficient. The larger the ratio, the larger the bias (leading to the endogeneity issue). The reduction in the residual term related to bias is only possible when the second term on the right-hand side of the equation is zero;

*1 ,when σεx 0*

Statistical patterns detected in observational data do not, on their own, provide evidence of casual relationships (as illustrated in equations [3.0] – [3.3]. In this analysis, I am not relying on investigator-designed experiment’s or natural experiments to provide the required exogenous variation—instead I will be leveraging other methods to solve the endogeneity problem in *Price*. For instance, it might be possible to locate and carve out an exogenous part of the variability in the potentially endogenous predictor *Price* and use it to estimate the impact it has on demand. This can be achieved using an innovative and flexible technique called *Instrument Variables estimation* (*IVE*).

Although the application of *IVE* to real data can be complex, the key idea is straightforward. First, note that observed differences among participants in the values of the question predictor (such as *Price*) may conceal an unknown mixture of endogenous and exogenous variation. Sometimes it is possible to carve out a part of this variation that is arguably exogenous and use only it in our estimation of causal impacts on an outcome. Success at this task requires information beyond a simple knowledge of the values of the outcome and question predictor.

In addition to these two variables, you must also have data for each participant on a special kind of background variable that is called an *instrument*. By integrating this *instrument* in a particular way onto the analysis, I can identify exogenous variation that is present in the question predictor, (e.g. *Price*) and use only it to obtain an asymptotically unbiased estimate of the causal impact of the question predictor on an outcome (e.g. demand).

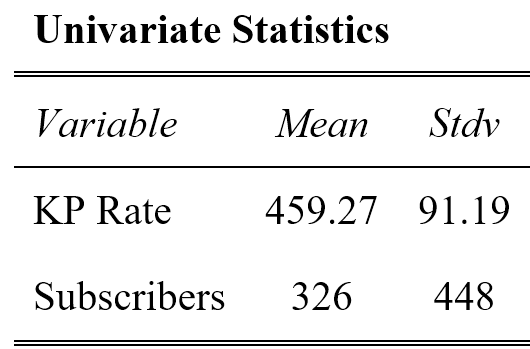
The dataset I will be using to estimate a demand function will be based on the Kaiser Permanente dataset provided by the Predictive Analytics team within the Membership, Market, Sales, and Analytics team (MMSA). The variables of interest are the question predictor, *Price* (*KP Rate*) and the outcome predictor, *Subscribers*.

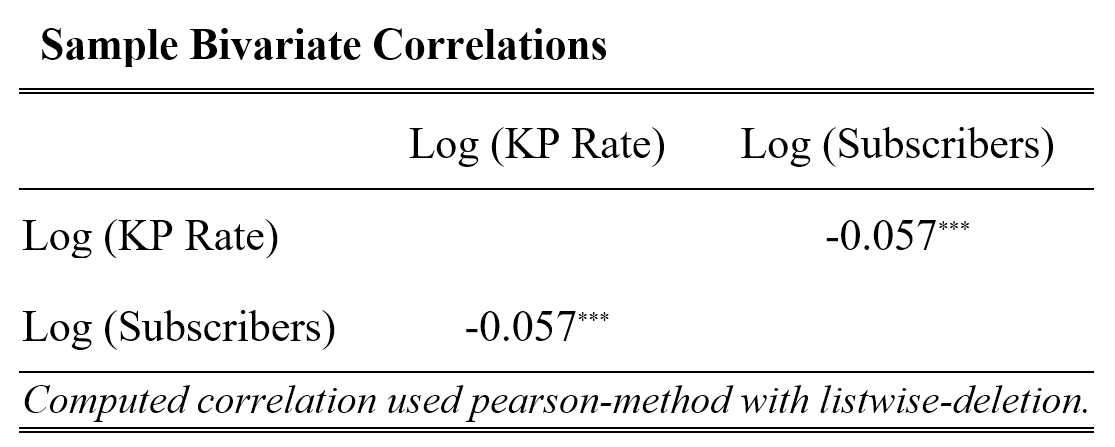
In the bottom table of figure [3], univariate statistics, I have presented univariate statistics on the two key variables in this analysis. The outcome variable *Subscribers* measures to count of subscribers in each group at a point in time. My principal question predictor, *Price* (*KP Rate*) measures the average rate charged to each employer per subscriber. The average *Price* and *Subscribers* per each group in my dataset is $459 and 326, respectively, per each group in the dataset.

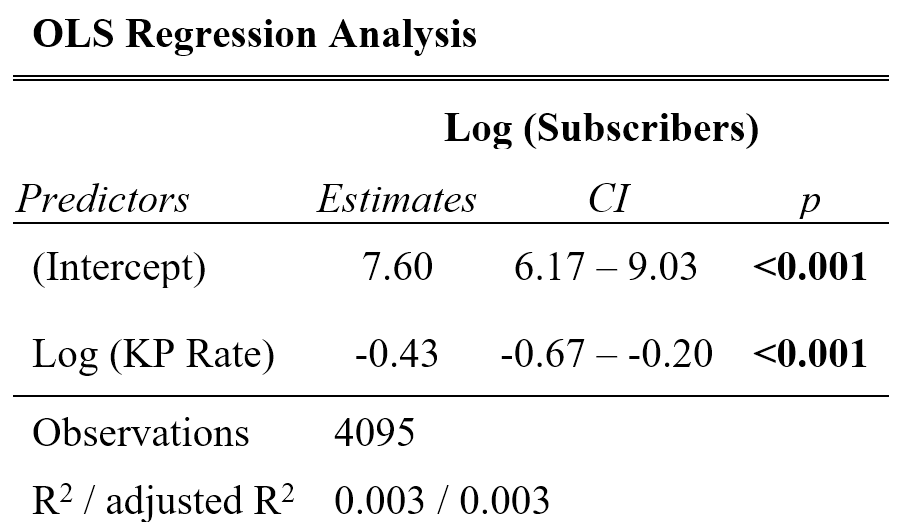
In the middle table (sample bivariate correlations) of figure [3], I have displayed a sample correlation matrix that summarizes the bivariate relationship between *Price* and *Subscribers*. Although the magnitude of the sample bivariate correlation between these variables is quite small, it is both statistically significant and negative (precisely the relationship direction I was anticipating), indicating that groups who had a higher *Price* tend to have a lower count of *Subscribers*.

Finally, in the bottom table (OLS regression analysis) of figure [3], I have presented an ordinary least-squares (OLS) regression fit, for the same outcome/predictor relationship (*Price* and *Subscribers*). I took the log of both *Price* and *Subscribers*, to evaluate the elasticity relationship between the two variables. As I expected, the elasticity of *Price* on *Subscribers* is negative and statistically significant. The elasticity coefficient of *Price* is very small (abs(0.43)) indicating that *Price* is *inelastic*. However, I know that most likely *Price* is endogenous because *Price* was not administered randomly and exogenously to groups. Because we have different prices, for each group, in unobserved ways, the relationship between *Price* and *Subscribers* in this dataset could have easily been due to unobserved influences that are omitted in the OLS regression analysis. This means that the question predictor and residuals may be correlated, and the resulting OLS estimate regression slope many be bias. Nevertheless, having this naïve OLS estimated summary of the observed relationship is a good place to start.

Figure 3: Table results of Univariate statistics, Correlations, and Regression. Based on the endogenous variable, Price (KP Rate), and dependent variable, Subscribers







I do not need to conduct a full-blown regression analysis to obtain the OLS estimate of *Log*(*Price*) on *Log*(*Subscribers*) slope in the bottom panel of figure [3]. With a single predictor in the regression model, an OLS-estimate of slope can be obtained directly by dividing the sample covariance of the outcome and question predictor by the sample covariance of the question predictor;

*[4.0]*

– 0.433

Notice that the estimate in equation [4.0] is identical to that obtained in the OLS regression analysis in the bottom panel of figure [3]. However, taking this extra step of calculating the slope estimator in equation [4.0] will soon provide insight into the functioning of the OLS slope estimator itself and the *instrumental-variable estimation*.

I know from statistical theory that if predictor and residuals are uncorrelated, then an OLS estimate of the slope coefficient will be an unbiased estimate of the population relationship. On the other hand, if the predictors and residuals are correlated, then an OLS estimate of slope will be biased (see equations [3.0] – [3.3]). So, to accept a value of – 0.433 as an unbiased estimate of the impact of *Log*(*Price*) on *Log*(*Subscribers*) in my data example, I must be convinced that a groups *Price* is truly independent of the residuals in the statistical model. This would certainly be the case if a groups *Price* had been assigned randomly to groups. However, this cannot be true. In fact, *Price*’s are determined by several factors and far from random (e.g. risk profile of group, competitive positioning, broker relationship, etc.). In summary, when a question predictor like *Price* is potentially endogenous, I cannot rely on standard OLS methods of estimation to provide an unbiased estimate of its causal impact on an outcome. Instead, I need to use a different approach.

In the top graphic (panel a) of figure [4], I have presented a Venn diagram that is useful for thinking about variation in, and covariation between, variables in either the sample or the population. I use it to marshal arguments about the variation and covariation of outcome Y and question predictor X in the population. For instance, in panel a of figure [4], I display a pair of intersecting ellipses. The total area of the upper ellipse represents the population variability in the outcome . Similarly, the total area of the lower ellipse represents the population variability in the question predictor .

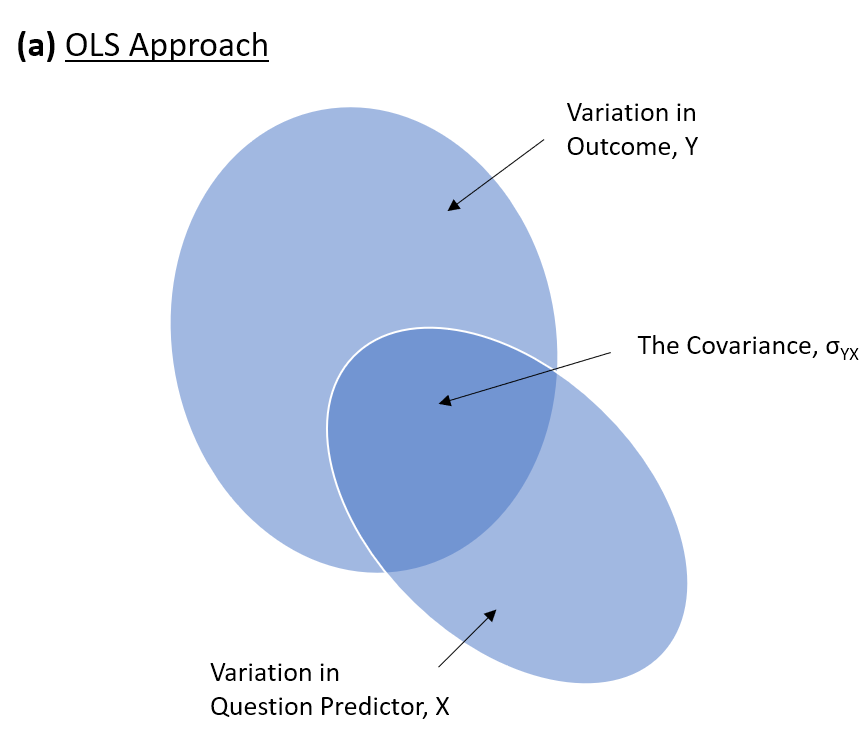
The intersection between the two ellipses symbolizes the population covariance of outcome variable and question predictor, *σyx*. In the analogy, when the outcome and predictor are strongly related, their covariance *σyx* is large. When the outcome and predictor are weakly related or not related at all, their covariance *σyx* will be small or zero. Conceptually, the ratio of the area of intersection to the total area of the upper ellipse then represents the portion of the total variability in outcome *Y* that has been successfully predicted by question predictor *X* (it is estimated, in the sample, by the R2 statistic). Consequently, on the Venn diagram, in the upper panel (panel a), the population regression slope, *β1*, is represented by the area of intersection of the two ellipses (representing *σyx*) divided by the total area of the lower ellipse (representing ). Working with this visual analogy for variation and covariation between outcome and predictor provides a useful explanatory tool here.

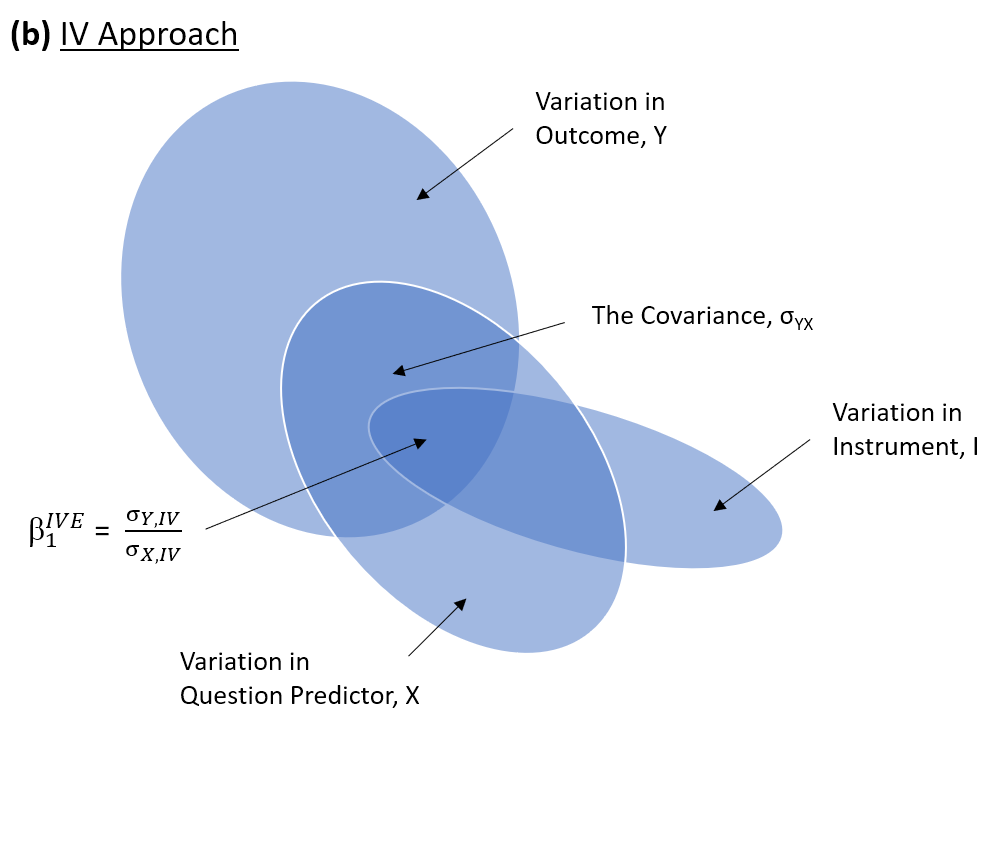
There is an obvious, yet important condition embedded in the algebraic statement that underpin the OLS estimator that must hold if OLS estimation is to succeed. I mention it here because it is the precursor to an analogous condition that must be satisfied for the *IVE* technique to be effective. In equation [3.2], notice that the population variance of the question predictor, , which appears in the denominators of both quotients on either side of the equal sign, cannot be zero. If it were zero, and I divided throughout the equation by it, the resulting quotients would be infinite and the value of the population regression slope, *β1*, inestimable. From a logical perspective, this makes sense. Why? Because it is just another way of stating that you can’t detect a relationship between an outcome and a question predictor if there is no variation in the predictor—in other words, if all observations in the sample have the same value for question predictor *X*. In the visual representation of figure [4], a no variation in *X*, would correspond to the area of the lower ellipse in the top panel (panel a) to shrivel up to nothing.

In addition to needing variability in *X*, the presence of the population variance of the question predictor in the denominator of the second term to the right of the equal sign in equation [3.2] is also important. As I have implied earlier, this quotient represents the bias that will be introduced into an OLS estimate of the population slope if the question predictor and the residuals are correlated, in the population. Notice that the bias term is again a quotient: the covariance of predictor and residual, divided by the population variance of the predictor. This suggest that the magnitude of the bias that could be obtained in an OLS estimate of slope is sensitive not only to any covariation that maybe present between question predictor and residual but also to the amount of variation present in the predictor itself.

There will be a perfect storm when predictor and residuals are correlated (so that the sample estimate of numerator *σεx* is non-zero) and there is little variation in the predictor (so that the sample estimate also approaches zero). In this case, the impact of any bias present due to the covariation of predictor and residuals in the numerator will be inflated, in the quotient, by the presence of the very small quantity that is present in the denominator. Clearly, this is another good reason to design research to ensure that you have substantial variability in the question predictor.

Figure 4: The use of Venn diagrams to illustrate the population variation and covariation among outcome Y, potentially endogenous question predictor, X, and instrument I, used for distinguishing the OLS and IV approaches.





In the current example, using observational data, I am estimating a demand function. My outcome variable *Subscribers*, is a measure of the number of subscribers within a group account enrolled in a Kaiser Healthcare plan. The important question predictor, *Price*, is used to calculate the price elasticity of each group account. Theoretically, there should be a causal relationship between the latter and the former. Consequently, I would like to use the observational data to obtain a credible estimate of the causal impact of *Price* on *Subscribers*. However, I suspect that the question predictor *Price* is potentially endogenous because Kaiser Permanente has been able to choose predetermined rates for each group account. As a result, an OLS estimate of the relationship between *Subscribers* on *Price* may provide a biased view of the hypothesized underlying population relationship between the marked price and the count of subscribers within a group.

In statistics, as in life, it is usually the case that we can always do better if we have some way to incorporate additional useful information into our decision. Setting all skepticism aside, let’s imagine for a moment that I had information available on an additional and very special kind of variable—named *I*, for *instrument*—that has also been measured for all groups in the sample dataset. Let’s ask ourselves: What properties would such an *instrument* need to have, to be helpful? How could I incorporate it into the analysis, if I wanted to end up with an unbiased estimate of the critical relationship between *Subscribers* and *Price*?

Although this seems to be a completely inhospitable analytic situation, I can gain insight by again applying covariance algebra to the hypothesized population regression featured in equation [3.0]. This time, though, instead of taking the covariance throughout with predictor *X* (as I did in equation [3.1], I will take them with the new *instrument I*. This leads to the following result;

*[5.0] cov(Y, I) cov(0 + 1X + , I)*

*1cov(X, I) + cov(, I)*

Or, more parsimoniously;

*[5.1] 1 +*

Dividing through by the population covariance of *X* and *I*;

*[5.2] 1 +*

Here, surprisingly, notice a second interesting consequence of the specification of the population linear regression model. From equation [5.2], the population covariance of *Y* with *I* (*σYI*) divided by the population covariance of *X* with *I* (*σXI*) is again equal to the critical parameter representing the key population relationship of interest (*1*), provided that the second term on the right-hand side of the equation is zero. And it is zero when the new instrument is uncorrelated with the residual in the population regression model. That is;

*1 ,when 0*

In other words, if there were some way to locate an instrument *I* that is incontrovertibly uncorrelated with the population residuals in the question model, I would be home free. Then, the slope of the causal relationship between *Price* on *Subscribers* in the population would simply be equal to the ratio of two important population covariances: (a) the covariance of the outcome and instrument, *σYI*, and (b) the covariance of question predictor and instrument, *σXI*. So, if *I* were known and its values are measured in the sample, I could simply estimate each of these respective covariances by their corresponding sample statistics and replace them in the quotient by their sample equivalents. This would provide me with an (asymptotically) unbiased estimate of the causal impact of *Price* on *Subscribers* in the population. This alternative estimator is called the *instrument variable estimator* of *β1*;

*[5.3]*

Because statisticians refer to covariance as one of the *second moments* of a bivariate distribution, the expression on the right-hand side of equation [5.3] is referred to the *method-of-moments* *IVE* of *β1*.

In figure [4], below, I have reproduced the previous naïve findings (from figure [3]) and also have introduced the variable that I claim works well as an instrument in the investigation of the causal impact of *Price* on *Subscribers*. This variable is *Risk Scores*. *Risk Scores* is a continuous variable that calculates the risk profile of members (healthy vs sick, high utilizers vs low utilizers, costly members vs inexpensive members). Notice that, on average, groups had an average *Risk Score* of 1.59, but the sample standard deviation is almost as large as the mean, indicating that there is considerable variability in the values of the potential instrument.

I have to argue that the instrument variable *Risk Scores* is unlikely to be correlated with the residuals in the regression of outcome *Subscribers* on the potentially endogenous variable *Price*. This assumption might be difficult to hold. *Risk Scores* is based off the health profile of members. It could be argued that locations with higher unemployment rates and less access to healthier food can impact *Risk Scores*. Later, I include control variables to avoid this potential violation. However, for illustration purposes, suppose for now that *Risk Scores* is not correlated with the residuals in the regression of outcome *Subscribers* on the potentially endogenous variable *Price*. Thus, although some part of the variation in *Price* may have been determined endogenously, some other parts of it may be exogenous and related to *Risk Scores*. If this is truly the case, then I can employ *Risk Scores* as an instrument to obtain an asymptotically unbiased estimate of the causal relationship between *Price* and *Subscribers*.

In figure [4], notice that the endogenous question predictor *Price* and instrument *Risk Scores* are indeed related. This relationship is very important. The higher the *Risk Score* the higher the *Price* (*KP Rate*)—which make intuitive sense. In addition, the outcome *Subscribers* has a negative relationship with the instrument *Risk Score*. Thus, the greater the *Risk Score* the less *Subscribers* enroll within a group account. Substituting the corresponding sample covariances into equation [5.3], I obtain an asymptotically unbiased method-of-moments IVE of the impact of *Price* on *Subscribers*;

5.2595

Notice that this coefficient is still negative but much larger in magnitude (- 5.2595 vs - 0.433). This suggest that groups are much more price sensitive to price changes. Provided that the instrument, *Risk Score*, satisfies the critical assumption I described earlier, then this new value of - 5.2595 is an asymptotically unbiased estimate of the impact of *Price* on *Subscribers*.

Figure 4: Table results of Univariate statistics, Correlations, and Methods-of-Moments IVE Estimate. Based on the endogenous variable Price (KP Rate), IV Risk Score, and dependent variable, Subscribers

