SPECIFICATION AND ESTIMATION OF THE PRICE RESPONSIVENESS OF ALCOHOL DEMAND: A POLICY ANALYTIC PERSPECTIVE

Srikant Devaraj

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Economics
Indiana University

February 2016

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Joseph V. Terza, PhD, Chair

_____

Yaa Akosa Antwi, PhD

Doctoral Committee

_____

Josette Jones, PhD

January 13, 2016

_____

Jisong Wu, PhD

ACKNOWLEDGEMENTS

It has been four and a half years since I started working on my PhD; finally my dream is becoming reality. When I started this journey I did not envisage what a tough task I had taken up on my shoulders. As it turned out, I surely needed some guidance and support, and I am very much indebted to everyone for guiding me in the right direction.

I am forever grateful to God for giving me an opportunity to choose Dr. Joseph V. Terza as my advisor and committee chair. As with most doctoral students, I also faced some challenges with the achievable research questions/objectives, research techniques and analytical processes. Dr. Terza helped me ease through all those challenges and motivated me to achieve my goals. He made innumerable comments on my work and graciously helped me with my writing style. He was very flexible to meet anytime, and worked with me with extreme patience. This dissertation would not have been possible without his invaluable guidance, inspirational advice, and tireless help. Thank You, Joe!!

I am extremely fortunate to have Dr. Yaa Akosa Antwi, Dr. Josette Jones, and Dr. Jisong Wu as my dissertation committee members, who come with great experience and knowledge. I really appreciate their constant support and encouragement. I thank Dr. Anne B. Royalty (Director of the PhD program), Dr. Mark Wilhelm, Dr. Wendy Morrison, Dr. Steve Russell, Dr. Richard Steinberg, Dr. Subir Chakrabarti, and Dr. Paul Carlin for providing great support to me in this endeavor. I would also like to thank Dr. Sumedha Gupta, Dr. Henry Mak, Dr. Jaesoo Kim, and all the Faculty members of Department of Economics at IUPUI for their lectures that enhanced my understanding on related topics and applications in the area. The entire faculty has always provided me with constructive feedback and thoughtful comments during my presentations that kept

me moving forward. I thank Dana Ward (Office Coordinator), who has always gone an extra mile to promptly help me with any paperwork needed during the process. I thank the Department of Economics at IUPUI for offering tuition remission for my doctoral degree program. I also would like to express my gratitude to my fellow cohort students – Ronia Hawash, Ausmita Ghosh, and Yan Yang in the Department of Economics at IUPUI for all the great times that we have shared over the years.

This doctoral journey would have never been a success without the unconditional support of Dr. Michael J. Hicks (Center Director), Dr. Dagney G. Faulk (Director of Research), and other staff members of Center for Business and Economic Research at Ball State University. I sincerely appreciate the thoughtfulness and generosity that they have showed me over my career span. Dr. Hicks and Dr. Faulk have not only been my mentors, but also provided invaluable suggestions and guidance over the years. I also thank Dr. Pankaj C. Patel for the support and encouragement throughout this endeavor.

Most importantly, none of this would have been possible without the love and blessings of my parents, my in-laws and my sister. Their prayers have kept me going this far. I thank all my friends for their support especially during difficult times. I dedicate this dissertation in memory of my grandmother, who immensely inspired me to dream bigger. Above all, I thank and additionally dedicate this dissertation and the last line of this section to my dear wife Anitha, for the unprecedented love and encouragement that she has provided to me during this long process.

Srikant Devaraj

SPECIFICATION AND ESTIMATION OF THE PRICE RESPONSIVENESS OF ALCOHOL DEMAND: A POLICY ANALYTIC PERSPECTIVE

Accurate estimation of alcohol price elasticity is important for policy analysis – e.g.., determining optimal taxes and projecting revenues generated from proposed tax changes. Several approaches to specifying and estimating the price elasticity of demand for alcohol can be found in the literature. There are two keys to policy-relevant specification and estimation of alcohol price elasticity. First, the underlying demand model should take account of alcohol consumption decisions at the *extensive margin* – i.e., individuals' decisions to drink or not – because the price of alcohol may impact the drinking initiation decision and one's decision to drink is likely to be structurally different from how much they drink if they decide to do so (the *intensive margin*). Secondly, the modeling of alcohol demand elasticity should yield both theoretical and empirical results that are *causally interpretable*.

The elasticity estimates obtained from the existing two-part model takes into account the extensive margin, but are not causally interpretable. The elasticity estimates obtained using aggregate-level models, however, are causally interpretable, but do not explicitly take into account the extensive margin. There currently exists no specification and estimation method for alcohol price elasticity that *both* accommodates the extensive margin *and* is causally interpretable. I explore additional sources of bias in the extant approaches to elasticity specification and estimation: 1) the use of logged (vs. nominal) alcohol prices; and 2) implementation of unnecessarily restrictive assumptions underlying the conventional two-part model. I propose a *new* approach to elasticity specification and

estimation that covers the two key requirements for policy relevance and remedies all such biases. I find evidence of substantial divergence between the new and extant methods using both simulated and the real data. Such differences are profound when placed in the context of alcohol tax revenue generation.

Joseph V. Terza, PhD, Chair

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1.

Background and Significance

1. Introduction

From principles of economics, the quantity demanded of any product is linked to its price. In the context of alcohol, an increase in alcohol prices would result in a decrease in alcohol consumption. Alcohol pricing policies are used to stem negative externalities associated with alcohol use and abuse (Elder et al., 2010; Leung & Phelps, 1993). Some example of such negative externalities include increases in: traffic fatalities resulting from drunk driving (Chaloupka, Saffer, & Grossman, 1993; Kenkel, 1993; Mullahy and Sindelar, 1994; Ruhm, 1996; Sloan, Reilly, & Schenzler, 1994); underage drinking (Grossman, Chaloupka, Saffer, & Laixuthai, 1994); utilization of publicly financed healthcare programs (Manning, Keeler, Newhouse, Sloss, & Wasserman, 1989); alcohol consumption among pregnant women (Patra et al., 2011); alcohol-related crime and domestic violence (Cook & Moore, 1993, 2002; Markowitz & Grossman, 1998, 2000); the incidence of sexually transmitted diseases (Chesson, Harrison, & Kassler, 2000); cirrhosis of the liver (Sloan, Reilly, & Schenzler, 1994); and adverse labor market outcomes (Mullahy & Sindelar, 1996; Terza, 2002).

Several state legislatures in the US have imposed, or are considering increasing, sumptuary taxes (often referred to as "sin taxes" or Pigouvian taxes) predominantly on alcohol and tobacco to reduce negative externalities. Studies have shown that such taxes increase the prices of alcohol, decrease the demand for alcohol, and thus lead to lower

negative externalities (Byrnes, Shakeshaft, Petrie, & Doran, 2013; Cook & Durrance, 2013).

There also has been growing interest in raising alcohol excise taxes to increase government revenues so as to reduce budget deficits or to fund various state and federal programs such as: alcohol and substance abuse treatment programs; drug courts; family court services; early childhood education programs; law enforcement; and health care. The Alcohol and Tobacco Tax and Trade Bureau (TTB) estimated the federal excise tax revenues from alcohol to be $13.9 billion in 2014. State and local government revenues from alcoholic beverages sales taxes amounted to $6.2 billion in 2014.[1]

The federal government increased excise taxes on all alcoholic beverages in 1991. Since then, the Congressional Budget Office has proposed additional increases in excise taxes on alcohol as a means of reducing budget deficits (Congressional Budget Office, 2013). States also levy additional taxes on alcoholic beverages. For example most states, in addition to general sale tax rates, levy alcohol excise taxes per gallon at the wholesale or retail level separately. A few states also levy ad valorem taxes on each alcohol type expressed as a percentage of its retail price.[2] Such ad valorem tax rates differ in on- and off-premise sales of alcohol. Any change in alcohol taxes would impact the revenues generated by the state/federal government as a result of it.

To summarize, one reason to raise alcohol taxes is to increase the government revenues to reduce budget deficits or fund various governmental programs. Another

---

[1] See http://ttb.gov/statistics/final15.pdf for details on federal excise tax revenues and http://www2.census.gov/govs/statetax/G14-STC-Final.pdf for details on state and local sale tax revenues.

[2] Some states that levy ad valorem taxes do not apply general sales tax on alcoholic products.

reason to raise alcohol taxes is to alleviate the negative consequences of alcohol use and abuse. Such Pigouvian taxes will have an impact on the quantity of alcohol consumed as a result of price increase. Elasticity is a pertinent measure that captures the change in consumption as a result of change in prices through taxes. Existing models that forecast revenue generation from an increase in alcohol taxes incorporate the price elasticity of alcohol demand in their models along with changes in alcohol taxes, the price of alcohol and current alcohol consumption (Alcohol Justice, 2014).

The alcohol price elasticity can also be used to study the effect of a varied set of proposed tax rates for curbing alcohol consumption to a certain level and, hence, impact public health issues (for example: reducing alcohol abuse among pregnant women, reducing a specific percentage of alcohol-related traffic accidents, bringing down under-age drinking, reducing crime and violence due to alcohol consumption, decreasing the number of sexually transmitted disease incidences, or achieving other policy goals). Furthermore, aside from the applications on the demand side, knowing the alcohol price elasticities is also valuable to the alcohol industry or the supply side. It could help the industry determine the changes in its sales and profits as a result of change in prices from tax changes and from government imposed policies (such as minimum legal drinking age, monetary penalties for underage drinking, blood alcohol concentration limits for driving, etc.).

Therefore, accurate estimation of the alcohol price elasticity is important for policymakers to forecast tax revenues from increases in alcohol taxes, and evaluate the optimal level of alcohol taxes intended to maximize revenues from taxes or restrict the growth in alcohol consumption for social welfare. It is also an equally essential measure

for the alcohol industry to estimate the effect of new (or proposed) federal and state taxes on the industry's sales and profits.

2. Role of Elasticity to Inform Alcohol Pricing Policy through Revenue Generation

Elasticities also play a key role in determining the direction of tax revenue changes due to tax rate changes. Ornstein (1980) and Levy and Sheflin (1985) show that if alcohol demand is inelastic, an increase in alcohol excise taxes will have a trivial shrinking effect on consumption; increase tax revenues; and increase disposable income spent on alcohol by consumers. Alternatively, when alcohol demand is elastic, an increase in excise taxes decreases consumption by a relatively larger amount and leads to a decline in tax revenues. Leung and Phelps (1993) extend the Ornstein (1980) model by relaxing some of the restrictions on the magnitude of the relevant demand elasticities. They show that revenues are maximized by setting the tax rates in such a way that the equilibrium alcohol consumption level is in the elastic part of the demand curve.

Alcohol Justice is an organization that monitors the alcohol industry and also leads campaigns for increasing alcohol taxes at a national and state-level to fund government programs on alcohol prevention and treatment. They derived a simple model to estimate revenue generation through alcohol tax increases (Alcohol Justice, 2014). The model incorporates the elasticity of alcohol demand, changes in alcohol taxes, the price of alcohol and current alcohol consumption. Their model shows that the more elastic is demand, the smaller the change in revenues. Also, different own price elasticities of alcohol demand yield different revenue generation values. As in the published studies cited above, when the price responsiveness of alcohol or any product is

elastic, increases in prices will reduce revenues. Alternatively, when demand is inelastic, tax revenues will be higher with increases in prices. In chapter 2 (section 5), I discuss the Alcohol Justice (2014) model in detail and demonstrate how different specifications of, and estimation methods for, the price elasticity of alcohol demand can lead to divergent revenue generation policies.

3. Existing Approaches to Alcohol Elasticity Specification and Estimation

Many different approaches to specifying and estimating the price elasticity of demand for alcohol can be found in the literature. Elder et al. (2010) does a systematic review of thirty-eight studies on alcohol elasticity. Gallet (2007) performs a meta-analysis of 132 studies that estimate the price elasticity of alcohol demand. Nelson (2014) meta-analyze 114 studies of beer elasticities. Wagenaar et al. (2009) find 112 studies that estimate the relationship between alcohol price/taxes and consumption. The studies found in these reviews implement different specifications, estimation methods and datasets. Approaches to alcohol demand regression modeling found in the literature include the: double-log, semi-log, Tobit, two-part, three-stage budgeting, and finite mixture. The systematic reviews, however, do not give attention to the methodological aspects of elasticity. Unfortunately, nearly all (if not all) extant estimates of alcohol price elasticity [including almost all of the studies meta analyzed by Elder et al. (2010), Gallet (2007), Nelson (2014), and Wagenaar et al. (2009)] and are of limited usefulness in the context of empirical policy analysis because they are subject to bias from one or more of a number of sources.

4. Exploring sources of bias in extant alcohol elasticity specification and estimation

As discussed in the section 2, accurate estimation of alcohol price elasticity is important for policy analysis. A complicating factor in the specification and estimation of the own price elasticity of alcohol demand is the typical abundance of zeros among the observed alcohol consumption values, which is the *first* source of bias. Such zero values present a challenge in econometric modeling and estimation because one's decision to drink [extensive margin] may be structurally different from his choice as to how much to drink (if he decides to drink) [intensive margin]. According to the American Medical Association, alcoholism is classified as illness. Alcohol consumption has negative effects with potential risk of addiction and alcohol abuse. Even light or moderate drinkers may show signs of slight dependency, which could be revealed by a strong craving to drink at certain occasions. The addictive and abusive potential of alcohol drinking takes a toll on a drinker's own health, inability to make rational decisions, impacts public health, and reduces the disposable income of drinkers. Individuals, who foresee these adverse effects of alcoholism or due to their cultural norms, may restrain from drinking.

Furthermore, it is quite plausible that the price of alcohol differentially impacts these two margins of the consumer's alcohol demand decision. Studies have shown that the drinking initiation decisions are negatively responsive to prices of alcohol (Chaloupka & Laixuthai, 1997; Cameron and Williams, 2001; Farrell et al., 2003; Manning et al., 1995; Ruhm et al., 2012). Youth when faced with higher alcohol prices were highly unlikely to switch from being abstainers to moderate drinkers (Williams, Chaloupka, & Wechsler, 2005). Therefore, it is essential to allow for this distinction in the specification and estimation of the price elasticity of alcohol demand. The two-part model developed

by Manning et al. (1995) [MBM henceforth] to estimate the own price elasticity of alcohol demand is indeed designed to account for the structural difference between the extensive and intensive margins. Of all the alcohol elasticity studies we surveyed, only three alcohol elasticity studies take explicit account of the extensive margin by implementing the MBM approach (Farrell et al., 2003; Manning et al., 1995; Ruhm et al., 2012).

The *second* source of bias in extant alcohol elasticity literature stems from the fact that the modeling of alcohol demand elasticity should yield both theoretical and empirical results that are *causally interpretable* and, therefore, useful for the analysis of potential changes in alcohol consumption that would result from exogenous (and ceteris paribus) changes in the price of alcohol (e.g., a change in tax policy). Terza, Jones, Devaraj et al. (2015) [TJD et al. henceforth], show that the elasticity measure suggested by MBM is not causally interpretable. Therefore, although the three aforementioned studies take explicit account of the extensive margin, they do not produce elasticity estimates that are causally interpretable. On the other hand, the remaining studies that we surveyed (the overwhelming majority of all studies surveyed) are designed to produce causally interpretable results.[3] Unfortunately, nearly all of these studies are based on aggregate-level (e.g. state-level) models and data and are, therefore, incapable of taking explicit account of individual alcohol demand decisions at the extensive margin. Most of these studies implement a log-log demand specification and use aggregate data (e.g., Goel & Morey, 1995; Lee & Tremblay, 1992; Levy & Sheflin, 1983; Wilkinson, 1987; Nelson, 1990; Young & Bielinska-Kwapisz, 2003). There are also a few studies that use

[3] See TJD et al. (2015) for details.

individual data to estimate alcohol elasticity, but do not allow for structural differences in the modeling of the extensive margin (Ayyagari, Deb, Fletcher, Gallo, & Sindelar, 2013; Kenkel, 1996). TJD et al. suggest an alternative elasticity specification (estimator) for the two-part context that is causally interpretable.

The *third* source of bias, among the studies that explicitly account for the extensive margin (Manning et al., 1995; Farrell et al., 2003; Ruhm et al., 2012; TJD et al.), is the imposition of unnecessary restrictions on the two-part model underlying elasticity specification and estimation. Such restrictions make simple ordinary least squares (OLS) estimation of the parameters of the intensive margin possible. This ease in estimation comes, however, at the cost of potential misspecification bias. Moreover, these restrictions are unnecessary because equally simple nonlinear least squares (NLS) estimators can be implemented. .

Nearly all of the conceptual and empirical treatments of alcohol demand elasticity found in the literature use log-price rather than nominal price. The origin of this practice traces to the convenience it affords via applying the ordinary least squares (OLS) method to a linear demand model with log consumption as the dependent variable and log price and other demand determinants as the independent variables. There is, however, no substantive reason for using log-price vs. nominal price and imposing this restriction on the model may lead to *fourth* source of bias.

In summary, there currently exist no specification and estimation method for alcohol price elasticity that accommodates the extensive margin, is causally interpretable, is less restrictive and uses the nominal price of alcohol. One of the primary goals of this dissertation is to detail and evaluate a new approach to the specification and estimation of

alcohol price elasticity [UPO henceforth] that takes into account these key aspects for policy relevance. Using simulated and real data, I compare the elasticities obtained using UPO to extant (biased) approaches. I also evaluate such differences in the context of revenue generation.

5. Goals of the Dissertation

The first objective of this dissertation is to develop a specification and estimation method for the own-price elasticity of alcohol demand that takes explicit treatment of the extensive margin in modeling and causal interpretability. In this chapter of the dissertation, I will first discuss the importance of accounting for the extensive margin in model specification. I will then detail the TJD et al. two-part model that is designed for this purpose. I will also address why the MBM model is not causally interpretable. Finally, I will detail the causally interpretable two-part-model-based alcohol elasticity specification and estimation approach of TJD et al.

The second objective of the dissertation is to compare the TJD et al. elasticity specification and estimation method to the extant approach that accounts for the extensive margin but is not causally interpretable (the MBM approach). I will compare the elasticities obtained by MBM and the TJD et al. with simulated and real data. I also demonstrate how the raw elasticity differences (TJD et al. vs. MBM approach) translate to policy differences in the revenue generation context. Such policy differences will be evaluated in an empirical context using data from the Ruhm et al. (2012) study.

The third objective of this dissertation is to develop a new elasticity specification and estimation method that take into account the extensive margin; is causally

interpretable; uses nominal prices of alcohol instead of logged price; and relaxes the unnecessarily restrictive assumptions underlying the conventional two-part model (the UPO approach).

The fourth objective of dissertation is to compare the UPO elasticity specification and estimation method to the aggregated log-log demand based approach which yields causally interpretable theoretical and empirical results but does not (cannot) account for individual drinking decisions at the extensive margin. First, I will create a state-level database by artificially aggregating data from the Ruhm et al. (2012) study. Secondly, I estimate alcohol price elasticities by applying the conventional log-log model to the artificially aggregated database. Third, I compare this aggregated elasticity estimate with that obtained using the UPO method. Finally, I will discuss how the raw elasticity differences obtained in this comparison translate to policy differences in the revenue generation contexts.

The final objective of this dissertation is to compare the elasticities obtained by using a version of the unrestricted causally interpretable two-part model with logged prices [UPOL henceforth] and UPO methods using simulated and real dataset. I will then evaluate how differences in elasticity estimates translate to differences in revenue generation.

6. Overview of the Dissertation

This dissertation will be organized as follows. The first and second objectives presented in section 5 of this chapter will be discussed in chapter 2 of the dissertation. The third and fourth objectives will be discussed in chapter 3 of this dissertation. Finally,

the last objective will be discussed in chapter 4 of this dissertation. Chapter 5 will provide a summary and discussion of the results obtained in the main chapters of the dissertation.

Chapter 2.

Specification and Estimation of Alcohol Price Elasticity in Individual-Level Demand Models with Zero-Valued Consumption Outcomes

1. Introduction

Numerous studies have estimated the own-price elasticity of alcohol demand using different data and methods (Elder et al., 2010; Gallet, 2007; Nelson, 2014; Wagenaar et al., 2009).[4] Previous studies have applied different models to estimate the alcohol demand elasticity using utility maximization theory, where consumers allocate their limited income towards activities and goods that maximize their utility (Ayyagari et al., 2013; Blake and Nied, 1997; Coate & Grossman, 1988; Farrell et al., 2003; Kenkel, 1996, 1993; Laixuthai & Chaloupka, 1993; Manning, Blumberg, & Moulton, 1995; Mullahy, 1998).

A complicating factor in the specification and estimation of the own price elasticity of alcohol demand is the typical abundance of zeros among the observed alcohol consumption values. According to the Center for Disease Control and Prevention's (CDC) National Health Interview Survey (NHIS), 51.6% of adults aged 18 and above were current regular drinkers in the year 2012.[5] Amongst the remaining share, 21.3% of adults were life-time abstainers, 12.8% adults were current infrequent drinkers,

---

[4] Elder et al. (2010) conduct a systematic review of 38 studies that specifically estimate price elasticities of alcohol demand. Gallet (2007) performs meta-analysis of 132 studies on alcohol demand elasticity. Wagenaar, Salois, and Komro (2009) find 112 studies that estimate the relationship between alcohol price/tax and consumption. Nelson (2014) conduct meta-analysis on 191 estimates of beer elasticities.

[5] Refer to page 75 of CDC's National Health Interview Survey (NHIS) report on http://www.cdc.gov/nchs/data/series/sr_10/sr10_260.pdf.

8.0% adults were former infrequent drinkers and 5.9% adults were former regular drinkers. Such zero values present a challenge in econometric modeling and estimation because one's decision to drink [*extensive margin*] may be structurally different from his choice as to how much to drink (if he decides to drink) [*intensive margin*]. According to the American Medical Association, alcoholism is classified as illness. Alcohol consumption, like cigarettes, substance use and illicit drugs, has negative effects with potential risk of addiction and alcohol abuse. Even light or moderate drinkers may show signs of slight dependency, which could be revealed by a strong craving to drink at certain occasions (for example: to overcome stress, excessive drinking during social events, etc.). The addictive and abusive potential of alcohol drinking takes a toll on a drinker's own health, inability to make rational decisions, impacts public health, and reduces the disposable income of drinkers. It is quite plausible that individuals, who foresee these adverse effects of alcoholism or due to their cultural norms, may restrain from drinking.

In particular, the price of alcohol may differentially impact these two margins of the consumer's alcohol demand decision.[6] Several empirical studies have shown that the drinking initiation decisions are negatively responsive to prices of alcohol (Manning et al., 1995; Chaloupka & Laixuthai, 1997; Cameron & Williams, 2001; Farrell et al., 2003; Ruhm et al., 2012). Also, youth when faced with higher alcohol prices were highly

---

[6] A total of 35.2% adults [i.e., 21.3% life-time abstainers + 8.0% former infrequent drinkers + 5.9% former regular drinkers] did not consume alcohol in 2012. Life-time abstainers had fewer than 12 drinks in his/her lifetime. The former (current) infrequent drinkers had at least 12 drinks in his/her lifetime and no drinks (fewer than 12 drinks) during the last year of NHIS survey period. The former regular drinkers had at least 12 drinks in his/her lifetime/1 year and had no drinks in the past year.

unlikely to switch from being abstainers to moderate drinkers (Williams, et al., 2005). Therefore, it is essential to allow for this distinction in the specification and estimation of the price elasticity of alcohol demand.

To date, only three studies (Farrell et al., 2003; Manning et al., 1995; Ruhm et al. 2012) have accounted for this important and essential two-part modeling aspect (i.e., differentiating extensive and intensive margins) of alcohol demand. Manning et al. (1995) [henceforth MBM] was the first study to suggest and apply a two-part-modeling-based estimator to account for the systematic difference between the extensive and intensive margins. This approach has also been applied by Farrell et al. (2003) and Ruhm et al. (2012). However, Terza, Jones, Devaraj et al. (2015) [henceforth TJD et al.] argue that the MBM approach produces elasticity results that are not causally interpretable because they are not founded in a potential outcomes framework that is causally interpretable.[7] They derive an elasticity measure and estimator that follow from well-defined potential outcomes based framework placed in the two-part modeling context and argue, therefore, that their approach does indeed produce causally interpretable elasticity estimates.

In order to assess whether the lack of causal interpretability of the MBM approach has empirical consequences (e.g. potential bias), in the present chapter I perform simulation analysis, re-estimate the Ruhm et al. (2012) model using the method of TJD et al. [henceforth, the potential outcomes (PO) *method*] and compare the resultant elasticity

---

[7] Refer TJD et al.; Pages 13 to 15 and pages 52 to 59 of Angrist & Pischke (2009); and Terza (2014). Health Policy Analysis from a Potential Outcomes Perspective: Smoking During Pregnancy and Birth Weight. Unpublished manuscript, Department of Economics, Indiana University Purdue University Indianapolis.

estimates to the those obtained via the MBM method. I replicate the study by Ruhm et al. (2012) using the consumption data from the second wave of National Epidemiological Survey of Alcohol and Related Conditions (NESARC) survey and price data from Uniform Product Code (UPC) barcode scanners collected by AC Nielsen. I find substantively different elasticity estimates from the MBM method vs. the PO method using both the simulated and real data.

The elasticity of alcohol demand is used in forecasting the changes in total tax revenues that may result from changes in the tax rate (Alcohol Justice, 2014; Leung & Phelps, 1993; Ornstein, 1980; Levy & Sheflin, 1985). Therefore, in this chapter, I will also evaluate how differences in the elasticity estimates (MBM vs. PO) translate to differences in empirical policy measures in the contexts of sin tax revenue generation.

Overall, applying the PO method of TJD et al. to the samples used in Ruhm et al. (2012) study, I find substantial divergence between the estimates of alcohol price elasticity. These differences in the raw elasticity estimates become even more evident when placed in the policy making (tax revenue generation) context. The discussion in TJD et al. supporting the PO approach, combined with the present comparison results, favor the use of the PO elasticity estimator.

This chapter is structured as follows. In the next section, I will outline the two-part model of alcohol demand, the elasticity measures proposed by MBM and TJD et al. (the PO specification), and the MBM and PO elasticity estimators.[8] In section 3, I will compare both MBM and PO estimates using simulated and real data from Ruhm et al.

[8] For details of these modeling aspects, see TJD et al.

(2012) study. I also discuss the data and results in detail. The comparison of elasticity estimates in the context of a change in beer tax revenues are given in the same section. The final section summarizes and concludes the chapter.

2. The Two-Part Model of Alcohol Demand and Relevant Elasticity Estimators

A complicating factor in the specification and estimation of the own price elasticity of alcohol demand is the typical abundance of zeros among the observed alcohol consumption values. Such zero values present a challenge in econometric modeling and estimation because one's decision to drink may be structurally different from his choice as to how much to drink (if he decides to drink). In particular, it is quite plausible that the price of alcohol differentially impacts these two margins of the consumer's alcohol demand decision. Therefore, it is essential to allow for this distinction in the specification and estimation of the price elasticity of alcohol demand. In this section, I detail the extant approaches that take into account the extensive and intensive margins (MBM and TJD et al.), and I also identify additional sources of bias in these and other extant approaches to the specification and estimation of the price elasticity of alcohol demand.

2.1 The MBM Elasticity Measure and Estimator

MBM was the first study to suggest and apply a two-part-modeling-based elasticity measure (η^{MBM}) and estimator ($\hat{\eta}^{MBM}$) to account for the systematic difference between the extensive and intensive margins. They model the *extensive margin* as

$$A > 0 \text{ iff } I(P\beta_{P1} + X\beta_{X1} + \varepsilon^{EM} > 0) \tag{2-1}$$

where A denotes the level of alcohol consumption, P is log-price, X is a vector of regression controls, $(\varepsilon^{EM} \mid P, X)$ is a logistically distributed random error term, $\beta_1' = [\beta_{P1} \quad \beta_{X1}']$ is the vector of parameters to be estimated and I(C) denotes the indicator function whose value is 1 if condition C holds and 0 if not. The *intensive margin* is modeled as

$$(A \mid A > 0) = \exp(P\beta_{P2} + X\beta_{X2} + \varepsilon^{IM}) \tag{2-2}$$

where $(\varepsilon^{IM} \mid P, X)$ is the random error term, with unspecified distribution, defined such that $E[\varepsilon^{IM} \mid P, X] = 0$ with $E[\exp(\varepsilon^{IM}) \mid P, X] = \psi$ (a constant); and $\beta_2' = [\beta_{P2} \quad \beta_{X2}']$ is the vector of parameters to be estimated. Consistent estimates of β_1 and β_2 are obtained using the following two-part protocol.

Part 1: Estimate β_1 by applying maximum likelihood logistic regression based on (1) to the full sample with $I(A > 0)$ as the dependent variable and [P X] as the vector of regressors.

Part 2: Estimate β_2 by applying OLS to

$$\ln(A \mid A > 0) = P\beta_{P2} + X\beta_{X2} + \varepsilon^{IM} \tag{2-3}$$

using the subsample of observations for whom $A > 0$.

In this modeling context, MBM define the own price elasticity of alcohol demand to be

$$\eta^{MBM} = \left(1 - E[\Lambda(P\beta_{P1} + X\beta_{X1})]\right)\beta_{P1} + \beta_{P2} \tag{2-4}$$

with corresponding consistent estimator

$$\hat{\eta}^{MBM} = \left(1 - \frac{\sum_{i=1}^{n} \Lambda(P_i\hat{\beta}_{P1} + X_i\hat{\beta}_{X1})}{n}\right)\hat{\beta}_{P1} + \hat{\beta}_{P2}\,. \tag{2-5}$$

where $\Lambda(\ )$ denotes the logistic cumulative distribution function (cdf), P_i and X_i are the observed values of P and X for the ith sampled individual (i = 1, ..., n), and the $\hat{\beta}$s are the parameter estimates from the above two-part protocol. The correct asymptotic standard error of (2-5) is derived in Appendix A.

2.2 The PO-based Elasticity Measure and Estimator of TJD et al.

TJD et al. argue that the MBM approach produces elasticity results that are not causally interpretable because they are not founded in a potential outcomes framework that is causally interpretable. They propose the following elasticity measure and estimator which follow from a well-defined PO framework placed in the two-part modeling context

$$\begin{aligned}\eta^{PO} = {} & E\big[\lambda(P\beta_{P1} + X\beta_{X1})\exp(P\beta_{P2} + X\beta_{X2})\beta_{P1} \\ & + \Lambda(P\beta_{P1} + X\beta_{X1})\exp(P\beta_{P2} + X\beta_{X2})\beta_{P2}\big] \\ & \times \frac{1}{E\Big[\Lambda(P^{exog}\beta_{P1} + X\beta_{X1})\exp(P^{exog}\beta_{P2} + X\beta_{X2})\Big]}\end{aligned} \tag{2-6}$$

and

$$\hat{\eta}^{PO} = \sum_{i=1}^{n} \frac{1}{n} \Big\{ \lambda(P_i\hat{\beta}_{P1} + X_i\hat{\beta}_{X1}) \exp(P_i\hat{\beta}_{P2} + X_i\hat{\beta}_{X2}) \hat{\beta}_{P1} + \Lambda(P_i\hat{\beta}_{P1} + X_i\hat{\beta}_{X1}) \exp(P_i\hat{\beta}_{P2} + X_i\hat{\beta}_{X2}) \hat{\beta}_{P2} \Big\} \times \left(\frac{1}{\sum_{i=1}^{n} \frac{1}{n} \Lambda(P_i\hat{\beta}_{P1} + X_i\hat{\beta}_{X1}) \exp(P_i\hat{\beta}_{P2} + X_i\hat{\beta}_{X2})} \right) . \quad (2\text{-}7)$$

where $\lambda(\ )$ denotes the logistic probability density function (pdf); and $\hat{\beta}_1' = [\hat{\beta}_{P1}' \quad \hat{\beta}_{X1}']$ and $\beta_2' = [\hat{\beta}_{P2}' \quad \hat{\beta}_{X2}']$ are the two-part estimates described above.[9] The correct asymptotic standard error of (2-7) is derived in Appendix B.

2.3 Causal Interpretability

TJD et al. argue that η^{PO} and $\hat{\eta}^{PO}$ are causally interpretable because they can be derived within a coherent potential outcomes framework. Because there is no apparent potential outcomes framework primitive for η^{MBM} (and, therefore, $\hat{\eta}^{MBM}$) TJD et al. conclude that it (and the MBM estimator) is not causally interpretable. As a result, they are not useful for empirical policy analysis.

3. Bias from using the MBM instead of TJD et al.

In the present section, as a follow-up to this conceptual argument favoring the PO specification and estimator [(2-6) and (2-7)] over the MBM approach [(2-4) and (2-5)], I

---

[9] See TJD et al. for details.

examine potential divergence between the two approaches in the two-part modeling context from both theoretical and practical perspectives.

3.1 A Simulation Study of the Bias

In Appendix C, I show that the difference between η^{MBM} and η^{PO} (the bias from implementing $\hat{\eta}^{MBM}$ instead of the causally interpretable $\hat{\eta}^{PO}$) can be formally expressed as

$$\eta^{MBM} - \eta^{PO} = \beta_{P1}\left(\frac{\omega}{\nu} - \zeta\right) \tag{2-8}$$

where

$$\nu \equiv E[\Lambda(W\beta_1)\exp(W\beta_2)]$$

$$\zeta \equiv E[\Lambda(W\beta_1)]$$

$$\omega \equiv E[\Lambda(W\beta_1)^2\exp(W\beta_2)]$$

$$W = [P \quad X]$$

and the expected values are with respect to W. To get a sense of the range of (2-8) and the extent of the influences of various factors on it, I simulated values of ν, ζ and ω using the following population design

$$P \sim U\{.5, .5\}$$

$$X = [U\{.5, .5\} \quad 1]$$

$$\beta_2 = [\beta_{P2} \quad \beta_{X2} \quad \beta_{C2}]'$$

$$\beta_1 = [h\times\beta_{P2} \quad \beta_{X1} \quad \beta_{C1}]' \tag{2-9}$$

where $U\{\mu, \sigma^2\}$ denotes the uniform random variable with mean μ and variance σ^2, β_{P2} is the coefficient of price in the second part of the model (intensive margin), β_{Xj} and β_{Ci} (j = 1 [extensive], 2 [intensive]) are the coefficient of the control variable and the constant term, respectively, for each of the parts of the model, and h is a factor representing the relative influence of log price on the extensive margin vs. the intensive margin ($0 \leq h \leq \infty$). The bigger is h, the greater the relative influence of log price on the extensive vs. the intensive margin. The "true" values of ν, ζ and ω for this simulated population design were obtained by generating a "super sample" of 2 million values for W based on (2-9) and then evaluating

$$\nu \equiv \sum_{t=1}^{T} \frac{1}{T}\left\{\Lambda(W_t \beta_1)\exp(W_t \beta_2)\right\}$$

$$\zeta \equiv \sum_{t=1}^{T} \frac{1}{T}\Lambda(W_t \beta_1)$$

$$\omega \equiv \sum_{t=1}^{T} \frac{1}{T}\left\{\Lambda(W_t \beta_1)^2 \exp(W_t \beta_2)\right\}.$$

To investigate the nature of the bias, I varied h and β_{C1} with β_{P2}, β_{X2} and β_{X1} all set equal to -1; and $\beta_{C2} = 1$. By increasing h, I increase the relative influence of log price on the extensive margin (vis-a-vis the intensive margin). Ceteris paribus increases in β_{C1} correspond to increases in the fraction of drinkers in the population. The values of the nominal bias (2-8) corresponding to the various (h, β_{C1}) pairs are given in Table 2.1 along with the bias as a percentage of η^{PO} (in parentheses) -- a measure of relative bias. In each cell I also report the fraction of drinkers in the population [in brackets].

I first note that the nominal values of the bias in Table 2.1 are uniformly negative. This follows from the law of demand as it applies to the extensive margin (i.e., the negativity of β_{P1}) and the apparent positivity of the bias factor $\left(\frac{\omega}{\nu} - \zeta\right)$ in (2-8). It is also clear from Table 2.1 that for any given value of β_{C1} (i.e., for a population with a given fraction of drinkers) the absolute values of both nominal and percentage bias monotonically increase as h increases (i.e., as log alcohol price becomes relatively more influential at the extensive margin). We can also see from Table 2.1 that for a given value of h, the bias appears to peak when the fraction of drinkers is in the low to mid-level range (i.e., from about 15% to 50%). Note that the bias can get quite large even for reasonable levels of h and the population proportion of drinkers – e.g. at h=3 and ζ = 62.66% ($\beta_{C1} = 3$) the bias is 67.64%.

3.2 Evaluating and Testing the Bias in a Real Data Context

As part of their examination of how estimates of the price elasticity of the demand for alcohol can vary depending on the researcher's choice of alcohol pricing database, Ruhm et al. (2012) consider a two-part model of alcohol demand in which

A = average daily volume of ethanol consumption from beer in ounces during the past year

P = log of price of beer in $ per ounce of ethanol

X = [gender, marital status, age, race, family size, education, census region, and occupation (blue collar, white collar, and service), household income]

where

gender = 1 if female, 0 otherwise

marital status = 1 if married, 0 otherwise

age = log of age

race = 1 if black, 0 otherwise; 1 if Hispanic origin, 0 otherwise; and 1 if other race,

0 otherwise

familysize = log of family size

education = 1 if no high school, 0 otherwise; 1 if some college, 0 otherwise; and 1 if completed college, 0 otherwise

region of residence = 1 if Midwest, 0 otherwise; 1 if Southern, 0 otherwise; and 1 if Western, 0 otherwise

occupation = 1 if blue collar, 0 otherwise; 1 if white collar, 0 otherwise; and 1 if service occupation, 0 otherwise

and

household income = log of household income.

The analysis sample for this model is drawn from the Uniform Product Code (UPC) barcode scanner dataset collected by AC Nielsen in grocery stores from 51 markets in the U.S. The UPC data contains accurate information on alcohol prices by type of beverage and packaging size. Ruhm et al. (2012) compared elasticity estimates obtained using UPC prices with those obtained via the commonly used American Chamber of Commerce Research Association (ACCRA) [now, Council for Community and Economic Research (C2ER)] prices. They were able to obtain both UPC and

ACCRA beer prices for only 35 states.[10] Data on beer consumption and the control variables comprising the vector X were drawn from the second wave of the National Epidemiological Survey of Alcohol and Related Conditions (NESARC) conducted in 2004-2005. The NESARC is a longitudinal survey that elicited information from respondents regarding alcohol consumption, alcohol use disorders and treatment services. The summary statistics of the Ruhm et al. (2012) analysis sample (size n = 23,743) are shown in Table 2.2.

As did Ruhm et al. (2012), I applied the conventional two-part estimation protocol culminating in (2-3) and obtained the estimates of β_1 and β_2 given in Table 2.3. Using these parameter estimates I calculated the price elasticity of demand for alcohol using both the causally interpretable PO-based estimator $\hat{\eta}^{PO}$ in (2-7) and the MBM estimator $\hat{\eta}^{MBM}$ in (2-5) proposed by MBM and implemented by Ruhm et al. (2012). The latter is not causally interpretable. The results are given in Table 2.4. Both estimates are statistically significant. The elasticity estimate of MBM is 0.089 higher in absolute value than $\hat{\eta}^{PO}$ and the difference is statistically significant.[11] In this case, estimated alcohol demand would be seen as price elastic if η^{MBM} and $\hat{\eta}^{MBM}$ were taken as the relevant measure and estimator. Whereas, using the causally interpretable η^{PO} and $\hat{\eta}^{PO}$ the opposite inference would be drawn.

---

[10]The 35 states from which the price data were taken are AL, AR, AZ, CA, CT, DC, FL, GA, ID, IL, IN, IA, KY, LA, MD, MA, MI, MS, MO, NE, NV, NH, NM, NC, NY,OH, OR, SC, VA, TN, TX WA, WV, WI, and WY.

[11] The correct asymptotic standard error of the difference is derived in Appendix D.

It is interesting that the results in the present real data example correspond closely with the case for the simulated population depicted in the first column, fifth row of Table 2.1 (cell 1-5). In that cell, h = 0.5 (measuring the relative influence of price in the extensive vs. the intensive margin) and the % of the population who are drinkers is 34%. As can be seen in Table 2.4, in the present real world example, the estimated value of h ($\hat{h}$) is .726 and the percentage of drinkers in the sample is 36%. Therefore, cell 1-5 in Table 2.1 is most closely relevant. For the hypothetical population represented therein, the model predicts an 8% bias from using the $\hat{\eta}^{MBM}$. The estimated bias as a percentage of $\hat{\eta}^{PO}$ is about 9%.

3.3 Revenue Generation from Tax Changes

Elasticities play a role in determining the revenues generated from taxes. Alcohol Justice is an organization that monitors the alcohol industry and also leads campaigns for increasing alcohol taxes at a national and state-level to fund government programs on alcohol prevention and treatment. They derive a simple model to estimate revenue generation through alcohol tax increases. The model incorporates the elasticity of alcohol demand, changes in alcohol taxes, the price of alcohol and current alcohol consumption. Their model shows that the more elastic is demand, the smaller the change in revenues. Alternatively, when demand is inelastic, tax revenues will be higher with increases in prices. Also, different own price elasticities of alcohol demand yield different revenue generation values.

Following Alcohol Justice (2014), the change in tax revenues as a result of the alcohol tax changes can be expressed as follows:

$$\Delta \text{Rev} = \left[(t+\delta) \times \left[1 + \left(\frac{\delta}{\mathcal{P}} \right) \eta \right] \times A_1 \right] - [t \times A_1] \tag{2-10}$$

where ΔRev is change in tax revenues due to change in the alcohol taxes; t is the current alcohol tax rate; δ is the change (increase or decrease) in the alcohol tax rate or the change (increase or decrease) in price of alcohol due to change in the alcohol tax rate (assuming 100% pass-through of excise tax rates to the retail price); A_1 is the current alcohol consumption; $\mathcal{P}$ is the current nominal price of alcohol and η is elasticity of alcohol demand.[12]

In order assess the substantive consequences of an estimation bias of this size, I placed it in the context of a $0.8533 per gallon increase in the federal excise tax on all alcoholic beverages that has been proposed by the Congressional Budget Office (CBO).[13] Using the calculator developed by Alcohol Justice (2014) [AJ], I projected the implied corresponding change in tax revenue based on each of the elasticity estimates ($\hat{\eta}^{PO} = -0.983$ and $\hat{\eta}^{MBM} = -1.073$). Aside from the elasticity value and the size of the

---

[12] When the excise tax is increased, the price of alcohol is also more likely to increase at least to the level of the tax increase. Studies have shown that the increase in the excise tax rate for each alcohol-type is more than fully passed-through to the price of the relevant alcohol-type. For a detailed discussion of alcohol tax pass-throughs see (Congressional Budget Office, 1990; Kenkel, 2005; Young and Bielińska-Kwapisz, 2002).

[13] With the aim of reducing the federal debt, the CBO frequently offers a number policy options. During the fiscal years 2014 and 2015, as one of many options suggested as means of raising revenues, was a proposed an increase in excise taxes on all alcoholic beverages (Congressional Budget Office, 2013 – see Option 32: https://www.cbo.gov/budget-options/2013/44854; Congressional Budget Office, 2014 – see Option 71: https://www.cbo.gov/budget-options/2014/49653).

proposed tax change, the AJ revenue calculator requires the following, which I held fixed for this example

current excise tax rate of beer = \$0.5867 per gallon[14]

total U.S. consumption of beer in 2011 = 6.303 billion gallons[15]

U.S. national average beer price in 2011 = \$15.20 per gallon.[16]

We also assume, for this illustration, that the tax increase is fully passed through to the retail price.[17] The estimated changes in tax revenue from the proposed tax change are \$4.9380 billion using $\hat{\eta}^{PO}$ and \$4.8920 billion using $\hat{\eta}^{MBM}$. The latter falls short of the former by \$46.082 million. To place this shortfall in perspective, I note that it is equal to 10.2% of the budget for National Institute of Alcohol Abuse and Alcoholism (NIAAA) for Fiscal Year 2015 (NIAAA, 2015).[18]

4. Summary and Conclusion

MBM propose and implement an estimator of the own-price elasticity of the demand for alcohol that is based on a conventional two-part model of alcohol consumption. This estimator has been implemented by Farrell et al. (2003) and Ruhm et al. (2012). Although the two-part modeling approach to alcohol demand is reasonable, TJD et al. argue that the elasticity estimator suggested by MBM, and its corresponding

---

14 Obtained from Congressional Budget Office (2013), (https://www.cbo.gov/budget-options/2013/44854)

15 Obtained from Brewers Almanac (Beer Institute, 2013)

16 Obtained from ACCRA (C2ER-COLI, 2015)

17 For discussion of tax pass-through rates see Congressional Budget Office (1990), Kenkel, (2005) and Young and Bielińska-Kwapisz (2002)

18 National Institute of Alcohol Abuse and Alcoholism (NIAAA) has an annual budget of \$447.4 million (in FY2015). See http://www.niaaa.nih.gov/grant-funding/management-reporting/financial-management-plan

implied elasticity measure, have no causal interpretation because they cannot be cast in a potential outcomes framework that is causally interpretable. TJD et al. develop an alternative elasticity specification [estimator] for the two-part context, which is causally interpretable (the PO approach).

To examine the determinants and extent of the divergence (bias) between MBM's stylized elasticity specification (and estimator) vs. PO approach, I conducted a simulation study in which I varied: 1) the level of the relative price influence at the extensive vs. intensive margins; and 2) the fraction of the population who are drinkers. I found that the former has a positive and monotonic effect on the bias, while the influence of the latter peaks when the fraction of drinkers is in the low to mid-level range.

As a follow-up to the conceptual discussion favoring the PO-based approach, and as a complement to the simulation study, I applied both methods to one of the models considered by Ruhm et al. (2012) using the same dataset as was analyzed by them. I found the elasticity estimates to be statistically significant from zero and from each other ($\hat{\eta}^{MBM} - \hat{\eta}^{PO} = -0.089$; p-value = .0286). To place this difference in a policy-relevant context, for each of the two elasticity estimates, I calculated the projected tax revenue change that would result from a proposed change in the federal excise tax on alcohol. I found the difference in tax revenue projections to be substantial; amounting to more than 10% of the yearly budget of the NIAAA. The discussion in TJD et al. supporting their potential outcomes framework approach, combined with the present comparison results, favor the use of the PO elasticity estimator.

Table 2.1: Simulation Analysis of the Bias

		h					
		0.5	**1**	**2**	**3**	**5**	**10**
β_{C1}	**-10**	-0.00 (0.00%) [0.00%]	-0.00 (0.00%) [0.00%]	-0.00 (0.01%) [0.00%]	-0.00 (0.02%) [0.00%]	-0.01 (0.10%) [0.01%]	-0.41 (3.83%) [0.19%]
	-5	-0.00 (0.20%) [0.43%]	-0.01 (0.44%) [0.40%]	-0.04 (1.25%) [0.44%]	-0.11 (2.89%) [0.60%]	-0.64 (12.18%) [1.40%]	-5.70 (126.41%) [7.91%]
	-3	-0.02 (1.31%) [3.02%]	-0.06 (2.92%) [2.78%]	-0.22 (7.90%) [2.97%]	-0.55 (16.52%) [3.76%]	-1.89 (49.94%) [6.66%]	-6.51 (220.38%) [15.41%]
	-1	-0.08 (5.72%) [17.40%]	-0.20 (12.39%) [15.63%]	-0.62 (29.52%) [14.68%]	-1.21 (51.68%) [15.49%]	-2.66 (109.58%) [18.45%]	-6.45 (292.24%) [23.47%]
	0	-0.10 (7.99%) [34.14%]	-0.25 (17.65%) [30.45%]	-0.72 (41.35%) [26.57%]	-1.33 (69.34%) [25.51%]	-2.70 (134.23%) [25.93%]	-6.26 (316.13%) [27.54%]
	1	-0.09 (7.76%) [55.43%]	-0.23 (18.50%) [50.00%]	-0.69 (46.49%) [41.78%]	-1.27 (79.16%) [37.35%]	-2.58 (149.49%) [33.83%]	-6.03 (332.32%) [31.62%]
	3	-0.03 (2.77%) [88.31%]	-0.09 (8.50%) [84.37%]	-0.37 (31.71%) [73.43%]	-0.86 (67.64%) [62.66%]	-2.10 (149.36%) [50.01%]	-5.44 (344.20%) [39.79%]
	5	-0.01 (0.47%) [98.13%]	-0.02 (1.68%) [97.22%]	-0.10 (9.84%) [92.98%]	-0.37 (33.18%) [84.51%]	-1.45 (117.41%) [66.18%]	-4.77 (333.50%) [47.96%]
	10	-0.00 (0.00%) [99.99%]	-0.00 (0.01%) [99.98%]	-0.00 (0.09%) [99.94%]	-0.01 (0.53%) [99.78%]	-0.14 (13.12%) [96.74%]	-2.94 (240.20%) [68.39%]

%-bias in parentheses, % of population for whom I(A > 0)=1in square brackets.

Table 2.2: Descriptive Statistics of Ruhm et al. (2012) Sample

Variable Name	Description	Mean N=23,743	Std.Dev
Extensive Margin (Drinker/Non-Drinker)	Beer drinker during the past year	0.36	0.48
Intensive Margin (Alcohol Consumption if Drinker)	Daily ethanol from beer in ounces during the past year	0.42	1.11
Intensive Margin (log of Alcohol Consumption if Drinker)	Log of oz of daily ethanol from beer during the past year	-2.64	2.08
X Variables			
lnbeer	Logged price of beer per oz of ethanol from UPC barcode data	0.22	0.08
female	Female gender (1=yes)	0.58	0.49
married	Currently married (1=yes)	0.50	0.50
lnage	Ln(age)	3.82	0.36
black	Black race (1=yes)	0.21	0.40
hispanic	Hispanic origin (1=yes)	0.22	0.41
other	Other race (1=yes)	0.04	0.21
lnfamsize	ln(Family size)	0.82	0.58
nohs	No high school (1=yes)	0.16	0.36
somecllg	Some college attendance (1=yes)	0.32	0.47
college	Completed college (1=yes)	0.27	0.45
midwest	Midwest region	0.22	0.41
south	Southern region	0.39	0.49
west	Western region	0.25	0.43
bluecllr	Blue-collar occupation	0.15	0.36
whitcllr	White-collar occupation	0.54	0.50
servwrkr	Service occupation	0.15	0.36
lnincome	ln(household income)	10.58	0.92

Table 2.3: Two-part Model Parameter Estimates for Ruhm et al. (2012)

Independent Variable	Extensive Margin MLE Logit for $\hat{\beta}_2$ Dep. Variable: I(A > 0)	Intensive Margin OLS for $\hat{\beta}_2$ Dep. Variable: (ln(A) | A > 0)
lnbeer	-0.532**	-0.733**
	(-2.20)	(-2.15)
female	-1.091***	-1.258***
	(-35.73)	(-28.29)
married	-0.166***	-0.276***
	(-4.66)	(-5.28)
lnage	-0.780***	-1.132***
	(-16.33)	(-15.99)
black	-0.497***	-0.00949
	(-12.09)	(-0.15)
hispanic	-0.186***	-0.329***
	(-4.63)	(-5.78)
other	-0.508***	-0.249**
	(-6.89)	(-2.29)
lnincome	0.219***	-0.0411
	(10.86)	(-1.43)
lnfamsize	-0.0503	-0.150***
	(-1.56)	(-3.24)
nohs	-0.109**	0.0797
	(-2.12)	(1.03)
somecllg	0.0530	-0.238***
	(1.33)	(-4.14)
college	0.194***	-0.387***
	(4.46)	(-6.20)
midwest	0.0275	0.172**
	(0.45)	(2.01)
south	-0.0867*	0.271***
	(-1.79)	(3.90)
west	0.123**	0.182***
	(2.56)	(2.66)
bluecllr	0.490***	0.425***
	(8.18)	(4.55)
whitcllr	0.458***	0.0855
	(8.73)	(1.01)
servwrkr	0.435***	0.192**
	(7.24)	(2.01)
_cons	0.598**	2.914***
	(2.04)	(6.87)
sample size	23743	8543

t statistics in parentheses; * p<0.10, ** p<0.05, *** p<0.01

Table 2.4: Causal and Non-Causal Elasticity Estimates

$\hat{\eta}^{PO}$	$\hat{\eta}^{MBM}$	Difference $\hat{\eta}^{MBM} - \hat{\eta}^{PO}$	%-Difference $\frac{\hat{\eta}^{MBM} - \hat{\eta}^{PO}}{\hat{\eta}^{PO}} \times 100\%$	$\hat{h} = \hat{\beta}_{P1} / \hat{\beta}_{P2}$	% of Sample I(A > 0) = 1
-.983*** (-4.195)	-1.073*** (-4.181)	-.089** (-2.188)	9.1%	0.726	36%

T-statistics in parentheses; ** $p < 0.05$, *** $p < 0.01$

Chapter 3.

Specification and Estimation of Alcohol Price Elasticity in Aggregate-Level Demand Models: Consequences of Ignoring the Extensive Margin

1. Introduction

Recall, in Chapters 2, I discuss two of four keys to accurate estimation of elasticity: 1) accounting for the extensive margin; and 2) causal interpretability. In Chapter 2, I noted that the two-part modeling-based estimator of Manning et al. (1995) [the *MBM method*] accounts for the extensive margin, but is not causally interpretable. Therein, I also discuss the elasticity specification and estimator proposed by Terza, Jones, Devaraj et al. (2015) [henceforth TJD et al.] (the *PO method*) that *both* takes account of the extensive margin *and* produces causally interpretable estimates. Moreover, I produce evidence of potential bias that may result from applying the MBM method (which is not causally interpretable) by comparing it to the PO method in simulated and real data settings.

To complement the analysis in Chapter 2, in the present chapter I will consider the most common approach to elasticity specification and estimation – viz. the log-log demand model with aggregated data [henceforth the aggregated log-log (*AGG-LOG) method*]. The discussion here will complement the analysis in Chapter 2. As TJD et al. show, the AGG-LOG method produces simple elasticity estimates which can be cast in a potential outcomes framework that is causally interpretable. For obvious reasons,

however, aggregation precludes the modeling of individual consumption decisions at the extensive margin.[19]

In the present chapter I explore the following additional sources of bias in extant approaches to elasticity specification and estimation: 1) data aggregation; 2) the use of logged (vs. nominal) alcohol prices; and 3) implementation of an unnecessarily restrictive version of the two-part model. I introduce a new approach to elasticity specification and estimation that remedies all such biases [UPO method henceforth].

This chapter is structured as follows. In the next section, I will detail the three aforementioned sources of bias. In section 3, I will introduce a new approach to alcohol demand elasticity specification and estimation that is free of these biases and, using simulated and real data, compare it to extant (biased) approaches. In section 4, I will compare the most commonly used extant method (AGG-LOG) with the newly proposed UPO methods using simulated data and a real analysis sample from the Ruhm et al. (2012) study. I also discuss the data and results in detail. The comparison of elasticity estimates in the context of a change in beer tax revenues are also given therein. The final section summarizes and concludes the chapter.

---

[19] There are other individual-level approaches that produce elasticity estimates that are causally interpretable – viz. Kenkel (1996) [who implements a Tobit model] and Ayyagari et al. (2013) [who use a finite mixture specification]. Neither of these modeling approaches explicitly allows for systematic differences in demand decisions made at the extensive margin. Therefore, we expect that they, like the AGG-LOG method, are subject to potential bias when used to evaluate individual-level responsiveness to price. Empirical evaluation of the extent of this bias is, however, beyond the scope of this dissertation.

2. Additional Sources of Bias in Elasticity Estimation

Correct specification and accurate estimation of the own price elasticity of demand lies at the heart of effective formulation and evaluation of alcohol pricing policy. Policy analytic goals such as the determination of optimal alcohol taxes (Kenkel, 1996; Pogue & Sgontz, 1989) and the projection of revenues from alcohol tax changes (Alcohol Justice, 2014) are cases in point. Unfortunately, nearly all (if not all) extant estimates of alcohol price elasticity [including almost all of the studies meta analyzed by Gallet (2007), Wagenaar et al. (2009), Nelson (2014)] and are of limited usefulness in the context of empirical policy analysis because they are subject to bias from one or more of a number of sources.

Among all of the studies I surveyed, only three take explicit account of the fact that for many individuals in the population, their utility maximizing consumption bundles include zero alcohol use.[20] Aside from these three studies, the vast literature on this subject ignores the likely possibility that an individual's decision to drink at all [the *extensive margin*] and his decision regarding how much to drink (if one chooses to drink) [the *intensive margin*] structurally differ (from both behavioral and econometric modeling perspectives). Failure to incorporate this distinction in any model of alcohol demand will likely lead to a biased elasticity estimate. The three studies that do draw this distinction all implement the two-part modeling approach to elasticity estimation suggested by Manning et al. (1995) [henceforth MBM].

---

[20] These studies are: Manning et al. (1995) [henceforth MBM]; Farrell et al. (2003); and Ruhm et al (2012).

TJD et al. however, argue that the three aforementioned studies are themselves subject to bias because, although the MBM elasticity measure takes explicit account of both the extensive and intensive margins, it is not informative for alcohol pricing policy. This lack of policy relevance of the MBM approach follows from the fact that it cannot be placed in a potential outcomes framework and, as such, cannot be interpreted as representing the causal relationship between price and alcohol consumption. TJD et al. suggest an alternative elasticity specification (estimator) for the two-part context that is derived within a potential outcomes framework that is causally interpretable.

In this section, I identify and detail additional sources of bias in TJD et al. (PO method) and other extant approaches using two-part model (MBM method) to the specification and estimation of the price elasticity of alcohol demand: 1) implementing a restricted version of the two-part model (TJD et al.); 2) data aggregation that ignores the extensive margin (AGG-LOG); and 3) the use of logged (vs. nominal) alcohol prices as a matter of convenience (TJD et al. and AGG-LOG).

2.1 Restrictive Nature of TJD et al.

Recall from Chapter 2 that the MBM approach and the PO method of TJD et al. are based on the two-part model specified in equations (2-1) and (2-2). Specifically, they model the intensive margin as

$$(A \mid A > 0) = \exp(P\beta_{P2} + X\beta_{X2} + \varepsilon^{IM}) \qquad (3\text{-}1)$$

where A denotes the level of alcohol consumption, P is log-price, X is a vector of regression controls, $(\varepsilon^{IM} \mid P, X)$ is the random error term, with unspecified distribution,

defined such that $E[\varepsilon^{IM} \mid P, X] = 0$ with $E[\exp(\varepsilon^{IM}) \mid P, X] = \psi$ (a constant); and $\beta_2' = [\beta_{P2} \quad \beta_{X2}']$ is the vector of parameters to be estimated. Consistent estimates of β_1 and β_2 are obtained using the following two-part protocol.

The intensive margin as given in (3-1) is, however, unnecessarily restrictive. Neither the explicit inclusion of ε^{IM} therein nor its accompanying conditional mean assumptions are necessary. They are imposed merely as a matter of convenience – so that the regression parameters $\beta' = [\beta_P \quad \beta_X']$ can be estimated via OLS. Instead, we need only assume that

$$E[A \mid P, X, A > 0] = \exp(P\beta_{P2} + X\beta_{X2}). \tag{3-2}$$

and the relevant regression parameters can still be estimated by applying the nonlinear least squares (NLS) estimation method directly to (3-2).

This difference in assumptions is not trivial. Assumption (3-2) encompasses a broader class of models than the conventional intensive margin specification given in (3-1). Therefore, the conventional two-part model may be subject to misspecification bias. As a result, in general, causal effect estimators cast in the conventional two-part modeling framework, $\hat{\eta}^{PO}$ in particular, may be biased.

2.2 Log-Linear Models with Aggregated Data: Ignoring the Extensive Margin

The most widely implemented approach to estimation of the own-price elasticity of the demand for alcohol is applying the ordinary least squares (OLS) method to a linear demand model to an aggregated level dataset with log consumption as the dependent

variable and log price and other demand determinants as the independent variables (henceforth the *aggregate LOG method [AGG-LOG]*). This includes a majority of the aggregate-level studies covered by the meta analyses by Wagenaar et al. (2009), Nelson (2014) and Gallet (2007).[21] Here the OLS estimate of the regression coefficient of log price is taken as the elasticity estimate. Its popularity notwithstanding, I argue that AGG-LOG is potentially biased for analyzing pricing policies aimed at modifying alcohol demand behavior at the individual level because it ignores individual alcohol demand decisions at the extensive margin. The AGG-LOG model of the demand for alcohol is expressed as

$$\bar{A} = \exp(\bar{P}\pi_P + \bar{X}\pi_X + \xi)\,. \qquad (3\text{-}3)$$

where $\bar{A}$ denotes the observed level of alcohol consumption, $\bar{P}$ is logged alcohol price, and $\bar{X}$ is a vector of other alcohol demand determinants (controls); all of which are measured at an aggregated level (e.g. averages at the level of the county, state, etc.). The vector of regression parameters to be estimated is $\pi' = [\pi_P \quad \pi_X']$ and ξ is the random term defined such that $E\left[\xi \mid \bar{P}, \bar{X}\right] = 0$. In this case, it is easy to show that own-price elasticity of alcohol demand is π_P -- the coefficient of $\bar{P}$ in (3-3) [$\eta^{AL} = \pi_P$]. Under these assumptions π_P can be easily estimated by applying the OLS estimator to the following log-log version of (3-3)

---

[21] Almost all aggregate-level studies in Wagenaar et al. (2009) meta-analysis, at least 169 out of 191 beer elasticity estimates from meta-analysis by Nelson (2014), and 974 out of 1,172 alcohol elasticity estimates meta-analyzed by Gallet (2007) uses either Double-Log or System (Almost Ideal Demand System, Rotterdam) models, which ignore the extensive margins.

$$\ln(\bar{A}) = \bar{P}\pi_p + \bar{X}\pi_X + \xi\,. \qquad (3\text{-}4)$$

Owing to its simplicity, this approach is widely implemented and results obtained from it are often used to analyze pricing policies aimed at modifying alcohol demand behavior at the individual level. If, however, the population from which the aggregated quantities $\bar{A}$, $\bar{P}$ and $\bar{X}$ are drawn includes a nontrivial proportion of non-drinkers, and the distinction between the extensive and intensive margins is important, in which case individual level demand behavior may be more accurately characterized by a two-part model, then this AGG-LOG approach is likely to be biased because it ignores this distinction.

2.3 Using Log vs. Nominal Prices

Nearly all of the conceptual and empirical treatments of alcohol demand elasticity found in the literature use log-price rather than nominal price. This is also true of TJD et al. in the development of their causally interpretable elasticity measure, $\hat{\eta}^{PO}$.[22] The origin of this practice traces to the convenience it affords via the log-log OLS (AGG-LOG) model discussed in the previous section. There is, however, no substantive reason for using log-price vs. nominal price and imposing this restriction on the model may lead to bias.

[22] TJD et al. specified their model using log price to conform to the extant literature.

3. An Alternative Elasticity Specification and Estimator

In the previous section I point out the lack of causal interpretability of the MBM approach to elasticity specification and estimation (discussed in detail in TJD et al.) and identify three additional potential sources of bias in extant elasticity measures and estimators (including TJD et al.): 1) data aggregation that ignores the extensive margin (AGG-LOG); 2) implementing a restricted version of the two-part model (TJD et al.); and 3) the use of logged (vs. nominal) alcohol prices as a matter of convenience (TJD et al. and AGG-LOG). In this section, I introduce a new approach to alcohol demand elasticity specification and estimation that extends the PO model of TJD et al. so as to avoid the three aforementioned biases. Table 3.1 summarizes various extant alcohol elasticity estimators used by empirical researchers in the published literature and the potential bias arising from each of them. An overwhelming majority of the extant literature meta-analyzed by Wagenaar et al. (2009) [81.4% elasticity estimates], Nelson (2014) [88.5% elasticity estimates] and Gallet (2007) [86.35% elasticity estimates] uses the AGG-LOG model to estimate the price elasticity of alcohol demand in aggregated data settings.[23] Unfortunately, the extensive margin is ignored in all of these studies; therefore, the policies implemented using these AGG-LOG elasticity estimates are subject to bias.

---

[23] These AGG-LOG elasticity estimates from the extant literature include double-log model and system models (such as the Almost Ideal Demand System (AIDS) and Rotterdam modeling approaches) when applied to aggregated data. The studies using system models allocate consumer expenditure from disposable income to different categories (such as alcohol, cigarettes, food, etc.) or its sub-categories. Using the aggregate data, the (log of) share of the expenditure on alcohol relative to total expenditure is then regressed on the log of a retail price index and other covariates to compute the (semi-) elasticity of alcohol.

I begin by specifying the following *unrestricted* version of the two part model given in (1) and (2). This unrestricted model is cast in terms of nominal rather that log prices:

Extensive Margin

$$A > 0 \text{ iff } I(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1} + e^{EMNOM} > 0) \tag{3-5}$$

where $\mathcal{P}$ is nominal price, $(e^{EMNOM} \mid \mathcal{P}, X)$ is a logistically distributed random error term and $a_1' = [a_{\mathcal{P}1} \quad a_{X1}']$ is the vector of parameters to be estimated.

Intensive Margin

$$(A \mid A > 0) = \exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2} + e^{IMNOM}) \tag{3-6}$$

where $a_2' = [a_{\mathcal{P}2} \quad a_{X2}']$ is the vector of parameters to be estimated and $(e^{IMNOM} \mid \mathcal{P}, X)$ is the random error term, with unspecified distribution, defined such that

$$E[A \mid \mathcal{P}, X, A > 0] = \exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2}) . \tag{3-7}$$

Consistent estimates of a_1 and a_2 are obtained using the following unrestricted two-part protocol.

Part 1: Estimate a_1 by applying maximum likelihood logistic regression based on (3-5) to the full sample, with I(A > 0) as the dependent variable and [$\mathcal{P}$ X] as the vector of regressors.

Part 2: Estimate a_2 by applying NLS to

$$A = \exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2}) + e^{NOM} \tag{3-8}$$

using the subsample of observations for whom $A > 0$.

Following the logic of TJD et al. in their derivation of the elasticity of alcohol demand in terms of log price, based on (3-5) through (3-7), I can express the unrestricted elasticity measure and estimator in terms of nominal price [henceforth, *unrestricted PO* (UPO)] as:

$$\begin{aligned}\eta^{UPO} = {} & E\big[\lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})a_{\mathcal{P}1} \\ & + \Lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})a_{\mathcal{P}2}\big] \\ & \times \frac{E[\mathcal{P}]}{E\big[\Lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})\big]}\end{aligned} \tag{3-9}$$

and

$$\begin{aligned}\hat{\eta}^{UPO} = {} & \sum_{i=1}^{n}\frac{1}{n}\big\{\lambda(\mathcal{P}_i\,\hat{a}_{\mathcal{P}1} + X_i\hat{a}_{X1})\exp(\mathcal{P}_i\,\hat{a}_{\mathcal{P}2} + X_i\hat{a}_{X2})\hat{a}_{\mathcal{P}1} \\ & + \Lambda(\mathcal{P}_i\,\hat{a}_{\mathcal{P}1} + X_i\hat{a}_{X1})\exp(\mathcal{P}_i\,\hat{a}_{\mathcal{P}2} + X_i\hat{a}_{X2})\hat{a}_{\mathcal{P}2}\big\} \\ & \times \left(\frac{\sum_{i=1}^{n}\frac{1}{n}\mathcal{P}_i}{\sum_{i=1}^{n}\frac{1}{n}\Lambda(\mathcal{P}_i\,\hat{a}_{\mathcal{P}1} + X_i\hat{a}_{X1})\exp(\mathcal{P}_i\,\hat{a}_{\mathcal{P}2} + X_i\hat{a}_{X2})}\right)\end{aligned} \tag{3-10}$$

where $\hat{a}_1' = [\hat{a}_{\mathcal{P}1} \quad \hat{a}_{X1}']$ and $\hat{a}_2' = [\hat{a}_{\mathcal{P}2} \quad \hat{a}_{X2}']$ are the unrestricted two-part estimates described above. The correct asymptotic standard error of (3-10) is derived in Appendix E. Similar to the PO elasticity specification and estimator, η^{UPO} and $\hat{\eta}^{UPO}$ are founded in a potential outcomes framework that is causally interpretable (see TJD et al.). Therefore, (3-10) produces elasticity estimates that are causally interpretable and, as such, are useful for policy analysis. In the section 4, I use simulated and real data, to compare results obtained using $\hat{\eta}^{UPO}$ (our preferred estimator) to those from the most commonly used extant method discussed in section 2, viz., AGG-LOG ($\hat{\eta}^{AL}$).

4. Bias from using AGG-LOG method Ignoring the Extensive Margin and the UPO

In Appendix F, I show that the difference between η^{AL} and η^{UPO} (the bias from implementing $\hat{\eta}^{AL}$ instead of the $\hat{\eta}^{UPO}$) can be expressed as:

$$\eta^{AL} - \eta^{UPO} = \pi_P - \left[\left(1 - \frac{u}{v}\right)a_{\mathcal{P}1} + a_{\mathcal{P}2}\right]m_{\mathcal{P}} \tag{3-11}$$

where[24]

$$v \equiv E[\Lambda(Wa_1)\exp(Wa_2)] \tag{3-12}$$

$$u \equiv E[\Lambda(Wa_1)^2\exp(Wa_2)] \tag{3-13}$$

---

[24] The expected values are with respect to W.

$$W = [\mathcal{P} \quad X]$$

$$m_{\mathcal{P}} \equiv E[\mathcal{P}] \tag{3-14}$$

$\mathcal{P}$ is the nominal price of alcohol; $a_{\mathcal{P}1}$ and $a_{\mathcal{P}2}$ are the coefficients of nominal prices in the extensive and intensive margins, respectively, in the unrestricted two-part model with nominal prices defined in (3-5) and (3-6); and π_P is the coefficient of average log price of alcohol ($\overline{P}$) in the AGG-LOG demand model in (3-4).

4.1 A Simulation-Based Study of the Bias Between $\hat{\eta}^{AL}$ and $\hat{\eta}^{UPO}$

Similar to the approach taken in section 3.1 of chapter 2, I focus on two factors in my simulation study of the divergence of $\hat{\eta}^{AL}$ and $\hat{\eta}^{UPO}$ as given in (3-11): the relative influence of nominal price at the extensive margin; and the fraction of drinkers in the population. To get a sense of the range of (3-11) and the extent of the influences of these two important factors on it, I simulated values of u and v using the following pseudo population design:

$$\mathcal{P} \sim U\{.5, .5\}$$

$$X = [U\{.5, .5\} \quad 1]$$

$$a_2 = [a_{\mathcal{P}2} \quad a_{X2} \quad a_{C2}]'$$

$$a_1 = [h \times a_{\mathcal{P}2} \quad a_{X1} \quad a_{C1}]'$$

$$(A \mid \mathcal{P}, X) = I(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1} + e^{IMNOM} > 0) \times A^*$$

e^{IMNOM} is logistically distributed

$$(A^{*} \mid \mathcal{P}, X) \sim \text{gamma}(\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2}), 1) \tag{3-15}$$

where $U\{m, v\}$ denotes the uniform random variable with mean m and variance v, $\text{gamma}(sh, sc)$ denotes the gamma random variable with shape parameter sh and scale parameter sc, $a_{\mathcal{P}2}$ is the coefficient of nominal price in the second part of the unrestricted two-part model (intensive margin), a_{Xj} and a_{Cj} (j = 1 [extensive], 2 [intensive]) are the coefficient of the control variable and the constant term, respectively, for each of the parts of the model, and h is a factor representing the relative influence of nominal price on the extensive margin vs. the intensive margin ($0 \leq h \leq \infty$). A is assumed to be a gamma random variable so that it does not conform to the restrictions imposed by the restricted two-part model. The bigger is h, the greater the relative influence of nominal price on the extensive vs. the intensive margin. The "true" values of v, u and $m_{\mathcal{P}}$ for this simulated population design were obtained by generating a "super sample" of 2 million values (n* = 2000K) for W and A based on the sampling design in (3-15) and then evaluating

$$v \equiv \sum_{i^{*}=1}^{n^{*}} \frac{1}{n^{*}} \left\{ \Lambda(W_{i^{*}} a_1) \exp(W_{i^{*}} a_2) \right\}$$

$$u \equiv \sum_{i^{*}=1}^{n^{*}} \frac{1}{n^{*}} \left\{ \Lambda(W_{i^{*}} a_1)^2 \exp(W_{i^{*}} a_2) \right\}$$

and

$$m_{\mathcal{P}} \equiv \sum_{i^{*}=1}^{n^{*}} \frac{1}{n^{*}} \left\{ \mathcal{P}_i) \right\}$$

respectively. To get the true value of π_P I first equally split the super sample into 50 arbitrary states. I then aggregate the simulated data at to the state-level by taking the

means of A and X, represented by $\bar{A}$ and $\bar{X}$ respectively. The nominal prices, $\mathcal{P}$ are first averaged across states and that mean is logged to get $\bar{P}$. OLS is then applied to (3-4) using the super sample and the estimated coefficient of $\bar{P}$ ($\hat{\pi}_P$) is taken as the true value of π_P because

$$\pi_P = \text{p lim}(\hat{\pi}_P).^{25}$$

To investigate the nature of the bias, I varied h and a_{C1} with $a_{\mathcal{P}2}, a_{X2}$ and a_{X1} all set equal to -1; and a_{C2} is set equal to 1. By increasing h, I increased the relative influence of nominal price at the extensive margin (vis-a-vis the intensive margin). Ceteris paribus increases in a_{C1} correspond to increases in the fraction of drinkers in the population. The values of the nominal bias (3-11) corresponding to a variety of (h, a_{C1}) pairs are given in Table 3.2 along with the bias as a percentage of η^{UPO} (in parentheses) -- a measure of relative bias. In each cell I also report the fraction of drinkers in the population [in brackets].

I first note that the nominal values of bias in Table 3.2 are uniformly positive for almost all combinations of h and a_{C1}. It is also evident from the Table 3.2 that for any given value of a_{C1} the absolute values of both nominal and percentage bias monotonically increases as h increases. Also, for a given value of h, the absolute values of bias vary upwards and then downwards when the fraction of drinkers increases. Note

---

[25] plim is short for "probability limit" which is a large sample (n approaches infinity) desirable statistical property of an estimator analogous to unbiasedness, the desirable small (or finite) sample property.

that the bias can get quite large even for reasonable levels of h and the population proportion of drinkers – e.g. at h=1 and proportion of drinkers=58.53% ($a_{C1} = 3$) the bias is 30.54%.

4.2 Comparison of $\hat{\eta}^{AL}$ and $\hat{\eta}^{UPO}$ with Real Data

As a means of demonstrating the potential empirical consequences of ignoring the extensive margin, I estimate own price elasticity by applying both the $\hat{\eta}^{AL}$ and the $\hat{\eta}^{UPO}$ estimation protocols to the following demand specification

A = average daily volume of ethanol consumption from beer in ounces during the past year

P = Nominal price of beer in $ per ounce of ethanol

X = [gender, marital status, age, race, household income, family size, education, region of residence, and occupation]

where variables and datasets are defined in section 3.2 of Chapter 2.

I then construct the artificially aggregated data by first taking the mean of the dependent variable (average daily volume of ethanol from beer in ounces) and the observable confounders (such as gender, marital status, age, race, household income, family size, and education) across all observations in each state. The UPC price data was already available at a state-level, measured in weighted price per ounce of ethanol. The mean of nominal price and actual consumption across observations at the state level were then logged as a precursor to the implementation of the AGG-LOG method. This artificial state-level aggregated database was then used to estimate the price elasticity of alcohol demand using the AGG-LOG method detailed in the section 2.2 of this chapter.

The summary statistics of this artificially aggregated state-level database are presented in Table 3.3.

Our focus here is on the size and statistical significance of $\hat{\eta}^{AL}$, $\hat{\eta}^{UPO}$ and their difference $\hat{\eta}^{AL} - \hat{\eta}^{UPO}$. The standard error of $\hat{\eta}^{UPO}$ is derived in Appendix E and that of the difference statistic for the bias, $\hat{\eta}^{AL} - \hat{\eta}^{UPO}$, is derived in Appendix G. Table 3.4 shows the comparison of the AGG-LOG and UPO elasticity estimators for the price of beer. The first column presents the elasticity estimate $\hat{\eta}^{AL}$, obtained using the parameter estimation protocol in equation (3-4). The second column of Table 3.4 shows $\hat{\eta}^{UPO}$, the elasticity estimate in equation (3-10) obtained using the two-part protocol culminating in equation (3-8). For the demand specification with all controls, I find the AGG-LOG elasticity estimate is -0.7136 and insignificant, but the UPO estimate was -0.5057 and statistically significant at 5% level. The difference between AGG-LOG and UPO was -0.2079 and statistically insignificant for the demand specification with all controls. However, for an alternate demand specification without the region of residence and occupation as controls, I find the AGG-LOG estimate for beer is -0.8966 and was insignificant, whereas the UPO estimate is -0.4902 and was significant at 1% level. The difference between the AGG-LOG and the UPO estimates is -0.4064 and significant at 5% level. It should be noted that the elasticity estimate obtained using the UPO model is statistically significant in both demand specifications, implying that it is imperative to take into account the extensive margins, which AGG-LOG ignores. The percentage discrepancy between the AGG-LOG and UPO elasticity estimates is 82.91%.

In both the simulation and real data analyses, I find strong evidence of biased

estimates from ignoring the extensive margin in price elasticity of alcohol demand. In the following, I examine how such differences may translate to the assessment of potential revenue generation from changes in alcohol taxes.

4.3 Revenue Generation from Tax Changes

As discussed in section 3.3 of Chapter 2, alcohol elasticities can also be used to estimate the potential change in revenues that would result from an increase in taxes. Therein, I also detail the simple revenue generation model developed by Alcohol Justice (2014), an organization that leads campaigns for raising alcohol taxes to fund government programs. Combining the Alcohol Justice model with the elasticities obtained by applying the AGG-LOG method to the artificially aggregated version of the data from Ruhm et al. (2012), I forecast the change in revenue that would result from a proposal to increase the Federal excise tax on alcohol by Congressional Budget Office (CBO) [See section 3.3 of Chapter 2 for details]. Table 3.5 shows the results for the revenue generation forecasts based on both the AGG-LOG and the UPO elasticity estimates obtained in section 4.3 (viz., $\hat{\eta}^{UPO} = -0.4902$ and $\hat{\eta}^{AL} = -0.8966$). The difference is substantial, amounting to a $207.1 million per year shortfall for the former. Such forecasting mistakes could result in serious errors in budget appropriation. In this example, the discrepancy is equal to 46.3% of the budget for the National Institute of Alcohol Abuse and Alcoholism (NIAAA) in Fiscal Year 2015.

5. Summary and Conclusion

In this chapter, I introduced a specification of the price elasticity of alcohol demand [η^{UPO} in equation (3-9)] and its corresponding consistent estimator [$\hat{\eta}^{UPO}$ in equation (3-10)] that takes into account the extensive margin, is founded in a potential outcomes framework that is causally interpretable, uses nominal prices of alcohol instead of logged price and relaxes the unnecessarily restrictive assumptions underlying the conventional two-part model. To examine the extent of the bias between the widely used AGG-LOG method vs. the UPO method, I performed a simulation study, where I varied the relative influence of nominal price at the extensive margin and the fraction of drinkers in the population. I found uniformly positive and substantial bias for almost all combinations and for a given level of former, as latter increases the bias monotonically increases. I also applied both methods to the dataset analyzed by Ruhm et al. (2012) and found the elasticity estimates to be statistically different from each other ($\hat{\eta}^{AL} - \hat{\eta}^{UPO} = -0.4064$; p-value=0.0277) for a particular demand specification. Such differences are profound when placed in the context of revenue generation. The difference between the approaches in terms of projected revenue for a recently proposed federal excise tax increase for beer was $207.1 million per year, which is 46.3% of the yearly budget of the NIAAA.

Table 3.1: Summary of extant alcohol elasticity estimators and additional biases

Econometric Models	Accounts for extensive margin	Studies using Nominal prices instead of Log	Unrestricted version of the two-part model	Causally Interpretable	Percentage or # of elasticity studies
AGG-LOG Models	No	Mixed	NA	Yes	W=81.40% N=88.47% G=86.36%
Tobit	No	No	NA	Maybe	1 study
Finite Mixture Models	No	No	NA	Maybe	2 studies
2PM – MBM	Yes	No	No	No	3 studies
2PM – TJD et al.	Yes	No	No	Yes	1 study
2PM – This paper	Yes	Yes	Yes	Yes	-

NA =Not applicable

Wagenaar et al. (2009) [W], Nelson (2014) [N], Gallet (2007) [G]

Table 3.2: Simulation Analysis of the Bias: $\hat{\eta}^{AL}$ vs. $\hat{\eta}^{UPO}$

		h					
		0.5	**1**	**2**	**3**	**5**	**10**
$\mathbf{a_{C1}}$	**-10**	-336.873 (10727.96%) [0.00%]	-518.345 (12380.32%) [0.00%]	-126.565 (-2015.27%) [0.00%]	-109.704 (-1310.09%) [0.00%]	†	†
	-5	7.326 (-233.77%) [0.23%]	13.016 (-311.55%) [0.12%]	-2.730 (43.53%) [0.05%]	2.986 (-35.69%) [0.02%]	-17.351 (138.19%) [0.01%]	-157.021 (681.90%) [0.00%]
	-3	3.395 (-109.66%) [1.59%]	4.521 (-109.65%) [0.89%]	8.261 (-133.06%) [0.36%]	14.652 (-176.32%) [0.17%]	-5.458 (43.58%) [0.04%]	-30.161 (131.01%) [0.00%]
	-1	1.464 (-50.56%) [10.02%]	1.459 (-38.27%) [5.75%]	3.883 (-66.72%) [2.43%]	8.546 (-107.56%) [1.17%]	21.846 (-177.76%) [0.32%]	-25.452 (110.75%) [0.02%]
	0	1.309 (-48.66%) [21.57%]	2.372 (-68.69%) [13.04%]	3.343 (-63.18%) [5.86%]	4.608 (-62.40%) [2.93%]	15.278 (-128.68%) [0.82%]	19.018 (-83.04%) [0.04%]
	1	1.093 (-44.24%) [39.19%]	1.806 (-59.74%) [25.34%]	2.976 (-65.78%) [12.51%]	5.312 (-82.62%) [6.71%]	8.972 (-81.75%) [2.05%]	20.111 (-88.63%) [0.12%]
	3	0.555 (-25.34%) [77.11%]	0.728 (-30.54%) [58.53%]	1.574 (-51.03%) [35.20%]	2.785 (-66.59%) [22.57%]	5.725 (-75.44%) [9.42%]	21.639 (-1.035%) [0.79%]
	5	0.607 (-28.75%) [95.25%]	0.529 (-24.58%) [84.21%]	0.691 (-28.87%) [58.72%]	1.084 (-37.85%) [42.90%]	2.983 (-65.16%) [23.87%]	14.458 (-95.13%) [4.12%]
	10	0.571 (-27.29%) [99.96%]	0.575 (-27.46%) [99.76%]	0.554 (-26.30%) [91.58%]	0.568 (-26.37%) [75.68%]	0.645 (-26.52%) [54.86%]	2.632 (-58.23%) [26.67%]

%-bias in parentheses, % of population for whom I(A > 0)=1 in square brackets.
†Insufficient observations for the AGG-OLS simulation

Table 3.3: Summary Statistics of Artificially Aggregated Data

Variable Name	Description	Mean N=35	Std.Dev
Average Alcohol Consumption if Drinker	Average Daily ethanol from beer in ounces	0.1639	0.0430
log of Average Alcohol Consumption	Log of Average oz of daily ethanol from beer	-1.8418	0.2647
Price	Logged price of beer per oz of ethanol from UPC barcode data	0.1877	0.0808
X Variables			
female	Average share of females	0.5775	0.0282
married	Share of sample who are currently married	0.5043	0.0563
lnage	Log of (average age)	3.8117	0.0565
black	Share of Black population in the sample	0.2226	0.1705
hispanic	Share of sample who are Hispanic origin	0.1430	0.1397
other	Share of sample with other race	0.0395	0.0246
lnincome	Log of (average household income)	10.5434	0.1525
lnfamsize	Log of (Average Family size)	0.7953	0.0570
nohs	Share of sample with no high school	0.1447	0.0375
somecllg	Share of sample with some college attendance	0.3157	0.0374
college	Share of sample who have completed college	0.2748	0.0650

Table 3.4: Comparing $\hat{\eta}^{AL}$ vs. $\hat{\eta}^{UPO}$ Method Using Real Data

Demand Specification	$\hat{\eta}^{AL}$	$\hat{\eta}^{UPO}$	**Difference,** $\hat{\eta}^{AL} - \hat{\eta}^{UPO}$	**% Difference** $\frac{\hat{\eta}^{AL} - \hat{\eta}^{UPO}}{\hat{\eta}^{UPO}} \times 100\%$
With all controls	-0.7136 (-0.7623)	-0.5057** (-2.1309)	-0.2079 (-0.8663)	41.11%
Without region of residence and occupation as controls	-0.8966 (-1.0934)	-0.4902*** (-2.6907)	-0.4064** (-2.2017)	82.91%

T-statistics in parenthesis.
* p<0.10, ** p<0.05, *** p<0.01

Table 3.5: Results of Changes in Tax Revenues from Federal Excise Tax Increase

Description	**Change in tax revenues using $\eta = \hat{\eta}^{AL}$ (a)**	**Change in tax revenues using $\eta = \hat{\eta}^{UPO}$ (b)**	**Difference (a) – (b) nominal dollars**
Additional revenues from "proposed" federal excise tax increase	$4.921 billion	$5.128 billion	-$207.064 million

Chapter 4.

Specification and Estimation of Alcohol Price Elasticity in Individual-Level Demand Models Using Nominal vs. Log Prices

1. Introduction

Chapters 2 and 3, respectively, assess possible empirical consequences (bias) from implementing elasticity specifications and estimators that are not causally interpretable; or do not take account of individual alcohol demand decisions made at the extensive margin; or implements unnecessarily restrictive version of the two-part model. The results demonstrate that these modeling deficiencies can lead to substantial divergence from elasticity estimates obtained via an approach that is both causally interpretable, explicitly models the extensive margin, are least restrictive and uses nominal prices of alcohol. Moreover, the prior chapters show that these differences can translate to nontrivial differences in associated policy recommendations in the context of revenue generation. To remain consistent with existing literature, however, in Chapter 2, I cast the elasticity specifications, estimators, and comparisons thereof, in terms of log of prices instead of nominal prices. Recall that in Chapter 3, I develop the UPO approach to alcohol demand elasticity specification and estimation which is free of biases including the use of nominal prices. Majority of the extant literature on alcohol elasticity uses logged prices of alcohol instead of nominal prices. Other than the need to conform to extant literature, there is no clear reason to formulate alcohol demand models and their corresponding elasticity measures in terms of log vs. nominal prices.

In the present chapter, to investigate the empirical consequences of modeling alcohol demand in terms of log vs. nominal prices, I develop a version of the UPO

method for elasticity specification and estimation (detailed in Chapter 3) that is cast in terms of logged prices. I will henceforth refer to this approach as the *UPOL method*. Using simulated and real data from Ruhm et al. (2012), I apply the UPOL method and compare the results with the corresponding elasticity estimates obtained using the UPO method and reported in Chapters 3. I then evaluate how differences in the elasticity estimates (UPOL vs. UPO) translate to differences in revenue generation. I find reasonable differences in estimates of alcohol price elasticity. These differences are more pronounced when placed in the revenue generation policy making contexts.

This chapter is structured as follows. The next section will detail the UPOL method for elasticity specification and estimation. In section 3, I will compare the UPOL with the UPO methods using simulated data and a real analysis sample from the Ruhm et al. (2012) study. I also discuss the data and results in detail. The comparison of elasticity estimates in the context of a change in beer tax revenues are also given therein. The final section summarizes and concludes the chapter.

2. Unrestricted PO Elasticity Specification and Estimator with logged prices

In the previous section I point out that nearly all of the conceptual and empirical treatments of alcohol demand elasticity found in the literature use log-price rather than nominal price. This is also true of TJD et al. in the development of their causally interpretable elasticity measure, $\hat{\eta}^{PO}$, the MBM model, and most studies using AGG-LOG models.[26] The origin of this practice traces to the convenience it affords via the log-log OLS (AGG-LOG) model discussed in the previous section. There is, however,

[26] TJD et al. specified their model using log price to conform to the extant literature.

no substantive reason for using log-price vs. nominal price and imposing this restriction on the model may lead to bias. In Chapter 3, I develop the UPO model that takes explicit account of extensive margin, are causally interpretable, uses unrestricted version of two-part model, and uses nominal prices in the own price elasticity of alcohol demand specification and estimation. In this section, I introduce a version of UPO approach in which the prices are cast as logged prices (UPO approach). I begin by specifying the following *unrestricted* version of the two part model cast in nominal prices given in (1) and (2).

Extensive Margin

$$A > 0 \text{ iff } I(P\alpha_{P1} + X\alpha_{X1} + e^{EMLOG} > 0) \tag{4-1}$$

where P is logged price, $(e^{EMLOG} \mid P, X)$ is a logistically distributed random error term and $\alpha_1' = [\alpha_{P1} \quad \alpha_{X1}']$ is the vector of parameters to be estimated.

Intensive Margin

$$(A \mid A > 0) = \exp(P\alpha_{P2} + X\alpha_{X2} + e^{IMLOG}) \tag{4-2}$$

where $\alpha_2' = [\alpha_{P2} \quad \alpha_{X2}']$ is the vector of parameters to be estimated and $(e^{IMLOG} \mid P, X)$ is the random error term, with unspecified distribution, defined such that

$$E\left[A \mid P, X, A > 0\right] = \exp(P\alpha_{P2} + X\alpha_{X2}). \tag{4-3}$$

Consistent estimates of α_1 and α_2 are obtained using the following unrestricted two-part protocol.

Part 1: Estimate α_1 by applying maximum likelihood logistic regression based on (4-1) to the full sample, with I(A > 0) as the dependent variable and [P X] as the vector of regressors.

Part 2: Estimate α_2 by applying NLS to

$$A = \exp(P\alpha_{P2} + X\alpha_{X2}) + e^{LOG} \tag{4-4}$$

using the subsample of observations for whom A > 0.

Following the logic of TJD et al. in their derivation of the elasticity of alcohol demand in terms of log price, based on (4-1) through (4-3), I can express the unrestricted elasticity measure and estimator in terms of logged price [henceforth, *unrestricted PO with logged prices* (UPOL)] as:

$$\begin{aligned}\eta^{UPOL} = {} & E\big[\lambda(P\alpha_{P1} + X\alpha_{X1})\exp(P\alpha_{P2} + X\alpha_{X2})\alpha_{P1} \\ & + \Lambda(P\alpha_{P1} + X\alpha_{X1})\exp(P\alpha_{P2} + X\alpha_{X2})\alpha_{P2}\big] \\ & \times \frac{1}{E\big[\Lambda(P\alpha_{P1} + X\alpha_{X1})\exp(P\alpha_{P2} + X\alpha_{X2})\big]}\end{aligned} \tag{4-5}$$

and

$$\begin{aligned}\hat{\eta}^{UPOL} = {} & \sum_{i=1}^{n}\frac{1}{n}\big\{\lambda(P_i\ \hat{\alpha}_{P1} + X_i\hat{\alpha}_{X1})\exp(P_i\ \hat{\alpha}_{P2} + X_i\hat{\alpha}_{X2})\hat{\alpha}_{P1} \\ & + \Lambda(P_i\ \hat{\alpha}_{P1} + X_i\hat{\alpha}_{X1})\exp(P_i\ \hat{\alpha}_{P2} + X_i\hat{\alpha}_{X2})\hat{\alpha}_{P2}\big\}\end{aligned}$$

$$\times \left(\frac{1}{\sum_{i=1}^{n} \frac{1}{n} \Lambda(P_i \ \hat{\alpha}_{P1} + X_i \hat{\alpha}_{X1}) \exp(P_i \ \hat{\alpha}_{P2} + X_i \hat{\alpha}_{X2})} \right) \quad (4\text{-}6)$$

where $\hat{\alpha}_1' = [\hat{\alpha}_{P1} \quad \hat{\alpha}_{X1}']$ and $\hat{\alpha}_2' = [\hat{\alpha}_{P2} \quad \hat{\alpha}_{X2}']$ are the unrestricted two-part estimates described above. The correct asymptotic standard error of (4-6) is derived in Appendix H. Similar to the UPO elasticity specification and estimator, η^{UPOL} and $\hat{\eta}^{UPOL}$ are founded in a potential outcomes framework that is causally interpretable (see TJD et al. and Chapter 3), however, it uses logged prices instead of nominal prices. Therefore, (4-6) produces elasticity estimates that are causally interpretable, but may be subjected to bias, hence may not be useful for policy analysis. In the section 3, I use simulated and real data, to compare results obtained using $\hat{\eta}^{UPOL}$ to $\hat{\eta}^{UPO}$ (our preferred estimator obtained and discussed in Chapter 3).

3. Bias from using the UPOL method and the UPO

In Appendix I, I show that the difference between η^{UPOL} and η^{UPO} (the bias from implementing $\hat{\eta}^{UPOL}$ instead of the $\hat{\eta}^{UPO}$) can be expressed as:

$$\eta^{UPOL} - \eta^{UPO} = \left[\left(1 - \frac{\omega}{\nu} \right) \alpha_{P1} + \alpha_{P2} \right] - \left[\left(1 - \frac{u}{\nu} \right) a_{\mathcal{P}1} + a_{\mathcal{P}2} \right] m_{\mathcal{P}} \quad (4\text{-}7)$$

where[27]

[27] The expected values are with respect to W for nominal prices and $\vec{W}$ for logged prices.

$$\nu \equiv E[\Lambda(\vec{W}\alpha_1)\exp(\vec{W}\alpha_2)] \tag{4-8}$$

$$\omega \equiv E[\Lambda(\vec{W}\alpha_1)^2\exp(\vec{W}\alpha_2)] \tag{4-9}$$

$$v \equiv E[\Lambda(Wa_1)\exp(Wa_2)] \tag{4-10}$$

$$u \equiv E[\Lambda(Wa_1)^2\exp(Wa_2)] \tag{4-11}$$

$$\vec{W} = [P \quad X]$$

$$W = [\mathcal{P} \quad X]$$

$$m_{\mathcal{P}} \equiv E[\mathcal{P}] \tag{4-12}$$

$\mathcal{P}$ is the nominal price of alcohol; P is the logged price of alcohol; α_{P1} and α_{P2} are the coefficients of logged prices in the extensive and intensive margins, respectively, in the unrestricted two-part model with nominal prices defined in equations (4-1) and (4-2) of this chapter; and $a_{\mathcal{P}1}$ and $a_{\mathcal{P}2}$ are the coefficients of nominal prices in the extensive and intensive margins, respectively, in the unrestricted two-part model with nominal prices defined in equations (3-5) and (3-6) of Chapter 3.

3.1 A Simulation-Based Study of the Bias Between $\hat{\eta}^{UPOL}$ and $\hat{\eta}^{UPO}$

I focus on the relative influence of nominal price at the extensive margin in my simulation study of the divergence of $\hat{\eta}^{UPOL}$ and $\hat{\eta}^{UPO}$ as given in (4-7). To get a sense of the range of (4-7) and the extent of the influences of this important factor on it, I simulated values of u and v using the following pseudo population design:

$$\mathcal{P} \sim U\{.5,\ .5\}$$

$$X = [U\{.5,\ .5\} \quad 1]$$

$$a_2 = [a_{\mathcal{P}2} \quad a_{X2} \quad a_{C2}]'$$

$$a_1 = [h \times a_{\mathcal{P}2} \quad a_{X1} \quad a_{C1}]'$$

$$(A \mid \mathcal{P},\ X) = I(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1} + e^{IMNOM} > 0) \times A^*$$

e^{IMNOM} is logistically distributed

$$(A^* \mid \mathcal{P},\ X) \sim \text{gamma}(\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2}), 1) \qquad (4\text{-}13)$$

where $U\{m,\ v\}$ denotes the uniform random variable with mean m and variance v, $\text{gamma}(sh, sc)$ denotes the gamma random variable with shape parameter sh and scale parameter sc, $a_{\mathcal{P}2}$ is the coefficient of nominal price in the second part of the unrestricted two-part model (intensive margin), a_{Xj} and a_{Cj} (j = 1 [extensive], 2 [intensive]) are the coefficient of the control variable and the constant term, respectively, for each of the parts of the model, and h is a factor representing the relative influence of nominal price on the extensive margin vs. the intensive margin ($0 \leq h \leq \infty$). A is assumed to be a gamma random variable so that it does not conform to the restrictions imposed by the restricted two-part model. The bigger is h, the greater the relative influence of nominal price on the extensive vs. the intensive margin. The "true" values of v, u and $m_{\mathcal{P}}$ for this simulated population design were obtained by generating a "super sample" of 2 million values (n* = 2000K) for W and A based on the sampling design in (4-13) and then evaluating

$$\nu \equiv \sum_{i^*=1}^{n^*} \frac{1}{n^*}\{\Lambda(W_{i^*}\, a_1)\exp(W_{i^*}\, a_2)\}$$

$$u \equiv \sum_{i^*=1}^{n^*} \frac{1}{n^*}\left\{\Lambda(W_{i^*}\, a_1)^2\exp(W_{i^*}\, a_2)\right\}$$

and

$$m_{\mathcal{P}} \equiv \sum_{i^*=1}^{n^*} \frac{1}{n^*}\{\mathcal{P}_i)\}$$

respectively. To get the true value of ω and ν, I apply the two-part protocol discussed in (4-4) and obtain values for the unknown parameters α_1 and α_2. I then evaluate

$$\nu \equiv \sum_{i^*=1}^{n^*} \frac{1}{n^*}\left\{\Lambda(\vec{W}_{i^*}\, \alpha_1)\exp(\vec{W}_{i^*}\, \alpha_2)\right\}$$

and

$$\omega \equiv \sum_{i^*=1}^{n^*} \frac{1}{n^*}\left\{\Lambda(\vec{W}_{i^*}\, \alpha_1)^2\exp(\vec{W}_{i^*}\, \alpha_2)\right\}$$

To investigate the nature of the bias, I varied h and with $a_{\mathcal{P}2}, a_{X2}$ and a_{X1} all set equal to -1; a_{C2} is set equal to 1; and a_{C1} is set equal to 3. By increasing h, I increased the relative influence of nominal price at the extensive margin (vis-a-vis the intensive margin). The values of the nominal bias (4-7) corresponding to a different values of h are given in Table 4.1 along with the bias as a percentage of η^{UPO} (in parentheses) -- a measure of relative bias. In each cell I also report the fraction of drinkers in the population [in brackets].

I first note that the nominal values of bias in Table 4.1 are uniformly positive for different values of h. It is also evident from the Table 4.1 that the absolute values of both nominal and percentage bias monotonically increases as h increases. Note that the bias

can get quite large even for reasonable levels of h and the population proportion of drinkers – e.g. at h=1 and proportion of drinkers=58.53%, the bias is 44.87%.

3.2 Comparison of $\hat{\eta}^{UPOL}$ and $\hat{\eta}^{UPO}$ with Real Data

As a means of demonstrating the potential empirical consequences of ignoring the extensive margin, I estimate own price elasticity by applying both the $\hat{\eta}^{UPOL}$ and the $\hat{\eta}^{UPO}$ estimation protocols to the following demand specification

A = average daily volume of ethanol consumption from beer in ounces during the past year

P = Nominal price of beer in $ per ounce of ethanol

X = [gender, marital status, age, race, household income, family size, education, region of residence, and occupation]

where variables and datasets are defined in section 3.2 of Chapter 2.

Our focus here is on the size and statistical significance of $\hat{\eta}^{UPOL}$, $\hat{\eta}^{UPO}$ and their difference $\hat{\eta}^{UPOL} - \hat{\eta}^{UPO}$. The standard error of $\hat{\eta}^{UPO}$ is derived in Appendix H and that of the difference statistic for the bias, $\hat{\eta}^{UPOL} - \hat{\eta}^{UPO}$, is derived in Appendix J. Table 4.2 shows the comparison of the UPOL and UPO elasticity estimators for the price of beer. The first column presents the elasticity estimate $\hat{\eta}^{UPOL}$, obtained using the parameter estimation protocol in equation (4-4). The second column of Table 4.2 shows $\hat{\eta}^{UPO}$, the elasticity estimate in equation (3-10) of chapter 3 obtained using the two-part protocol culminating in equation (3-8) of chapter 3. I find the UPOL elasticity estimate is -0.4892 and statistically significant at 5% level. Recall from Chapter 3 that the UPO estimate was

-0.5057 and statistically significant at 5% level. The difference between the UPOL and the UPO estimates is 0.0165, but statistically insignificant. Nevertheless, the absolute percentage discrepancy between the UPOL and UPO elasticity estimates is 3.26%.

In both the simulation data analyses, I find strong evidence of biased estimates from using logged prices instead of nominal prices. However, in real data settings, with the demand specification used in Ruhm et al. (2012), I find suggestive evidence of bias. In the following section, I examine how such differences may translate to the assessment of potential revenue generation from changes in alcohol taxes.

3.3 Revenue Generation from Tax Changes

Combining the Alcohol Justice model (discussed in section 3.3 of chapter 2) with the elasticities obtained by applying the UPOL method to the data from Ruhm et al. (2012), I forecast the change in revenue that would result from a proposal to increase the Federal excise tax on alcohol by Congressional Budget Office (CBO) [See section 3.3 of Chapter 2 for details]. Table 4.3 shows the results for the revenue generation forecasts based on both the UPOL and the UPO elasticity estimates obtained in section 4.2 (viz., $\hat{\eta}^{UPO} = -0.4902$ and $\hat{\eta}^{UPOL} = -0.5057$). The difference amount to a $8.4 million per year shortfall for the former. Such forecasting mistakes could result in reasonable errors in budget appropriation. In this example, the discrepancy is equal to 2% of the budget for the National Institute of Alcohol Abuse and Alcoholism (NIAAA) in Fiscal Year 2015.

4. Summary and Conclusion

In this chapter, I compare the estimator of the price elasticity of alcohol demand [$\hat{\eta}^{UPO}$ in equation (3-10) of chapter 3] that uses nominal prices of alcohol instead of logged price (and is free from other biases discussed in Chapter 3) with an estimator that uses logged prices of alcohol in the demand specification [$\hat{\eta}^{UPOL}$ in equation (4-6) of this chapter]. To examine the extent of the bias between the UPOL vs. the UPO method, I performed a simulation study, where I varied the relative influence of nominal price at the extensive margin. I found uniformly positive and substantial bias for almost all variations. I also applied both methods to the dataset analyzed by Ruhm et al. (2012) and found suggestive evidence of smaller bias in the elasticity estimates ($\hat{\eta}^{UPOL} - \hat{\eta}^{UPO} = 0.0165$, but insignificant). Such differences are reasonable when placed in the context of revenue generation. The difference between the approaches in terms of projected revenue for a recently proposed federal excise tax increase for beer was $8.4 million per year, which is 2% of the yearly budget of the NIAAA.

Table 4.1: Simulation Analysis of the Bias: $\hat{\eta}^{UPOL}$ vs. $\hat{\eta}^{UPO}$

h					
0.5	**1**	**2**	**3**	**5**	**10**
1.002 (-45.75%) [77.11%]	1.069 (-44.87%) [58.53%]	1.512 (-49.04%) [35.20%]	2.278 (-54.47%) [22.57%]	4.708 (-62.04%) [9.42%]	14.51 (-69.38%) [0.79%]

%-bias in parentheses, % of population for whom I(A > 0)=1 in square brackets.

Table 4.2: Comparing $\hat{\eta}^{UPOL}$ vs. $\hat{\eta}^{UPO}$ Method Using Real Data

$\hat{\eta}^{UPOL}$	$\hat{\eta}^{UPO}$	**Difference,** $\hat{\eta}^{UPOL} - \hat{\eta}^{UPO}$	**% Difference** $\frac{\hat{\eta}^{UPOL} - \hat{\eta}^{UPO}}{\hat{\eta}^{UPO}} \times 100\%$
-0.4892**	-0.5057**	0.0165	-3.26%
(-2.1284)	(-2.1309)	(0.0498)	

T-statistics in parenthesis.
* p<0.10, ** p<0.05, *** p<0.01

Table 4.3: Results of Changes in Tax Revenues from Federal Excise Tax Increase

Description	**Change in tax revenues using $\eta = \hat{\eta}^{UPOL}$ (a)**	**Change in tax revenues using $\eta = \hat{\eta}^{UPO}$ (b)**	**Difference (a) – (b) nominal dollars**
Additional revenues from "proposed" federal excise tax increase	$5.129 billion	$5.121 billion	$8.41 million

Chapter 5.

Summary and Conclusion

1. Introduction

Several state legislatures in the US have imposed, or are considering increasing, sumptuary or Pigouvian taxes (e.g., "sin taxes") predominantly on alcohol and tobacco. Such pricing policies are set so as to reduce the consumption of alcohol abusers so as to minimize external costs to those harmed by them. There also has been growing interest in raising alcohol excise taxes to increase government revenues to reduce budget deficits or to fund various state and federal programs. The own-price elasticity of the demand for alcohol plays a vital role in determining optimal Pigouvian taxes (to reduce externalities) and in projecting revenues generated from proposed tax changes.[28] Therefore, accurate estimation of alcohol price elasticity is important for policy analysis.

2. Key Aspects of Policy Relevant Alcohol Elasticity Specification and Estimation

Many different approaches to specifying and estimating the price elasticity of demand for alcohol can be found in the literature. There are four keys to policy-relevant specification and estimation of alcohol price elasticity. *First*, the underlying demand model should take account of alcohol consumption decisions at the *extensive margin* – i.e., individuals' decisions to drink or not – because the price of alcohol may impact the drinking initiation decision. This is important because one's decision to drink is likely to be structurally different from how much they drink if they decide to do so (the *intensive*

[28] See Chapter 1 of dissertation for details on the role of elasticities in the context of revenue generation.

margin). *Secondly*, the modeling of alcohol demand elasticity should yield both theoretical and empirical results that are *causally interpretable* and, therefore, useful for the analysis of potential changes in alcohol consumption that would result from exogenous (and ceteris paribus) changes in the price of alcohol (e.g., a change in tax policy). *Thirdly*, the extant models that explicitly take extensive margins into account are unnecessarily restrictive and are merely imposed as a matter of convenience so that the parameters can be estimated using the Ordinary Least Squares (OLS) method. *Finally*, almost all the conceptual and empirical treatments of alcohol demand elasticity found in the literature use log-price rather than nominal price.

There currently exists no specification and estimation method for alcohol price elasticity that accommodates the extensive margin, is causally interpretable, is less restrictive and uses nominal prices of alcohol. One of the primary goals of this dissertation is to detail and evaluate a *new* approach to the specification and estimation of alcohol price elasticity that covers these four key requirements for policy relevance.

2.1. Models that Account for the Extensive Margin

The vast majority of extant studies that I surveyed do not accommodate the possibility that one's decision to drink or not [extensive margin] may require special attention in model design. If the relevant population includes a nontrivial proportion of non-drinkers and the extensive margin is ignored in modeling then the resultant elasticity measure is likely to be biased. The most widely implemented approach to estimation of the own-price elasticity of the demand for alcohol is applying the ordinary least squares (OLS) method to a linear demand model with log consumption as the dependent variable

and log price and other demand determinants as the independent variables [henceforth the *aggregate LOG method (AGG-LOG)*] on aggregate-level (e.g. state-level) models and data. Unfortunately, nearly all of these studies incapable of taking explicit account of individual alcohol demand decisions at the extensive margin, therefore, elasticity estimates obtained using AGG-LOG is potentially biased for analyzing pricing policies aimed at modifying alcohol demand behavior at the individual level.[29] The two-part model developed by Manning, Blumberg, Moulton (1995) [MBM henceforth] to estimate the own price elasticity of alcohol demand is indeed designed to account for the structural difference between the extensive and intensive margins. Of all the alcohol elasticity studies I surveyed, only three alcohol elasticity studies take explicit account of the extensive margin by implementing the MBM approach (Manning et al., 1995; Farrell et al., 2003; Ruhm et al., 2012).

2.2. Models that are Causally Interpretable

Terza, Jones, Devaraj et al. (2015) [TJD et al. henceforth] show that the two-part model elasticity measure suggested by MBM is not causally interpretable. TJD et al. propose an elasticity measure and estimator that follow from a well-defined potential outcomes (PO) framework placed in the two-part modeling context and argue, therefore, that their approach does indeed produce causally interpretable elasticity estimates.

In order to assess whether the lack of causal interpretability of the MBM approach has empirical consequences (e.g. potential bias), in Chapter 2 of my dissertation, I

[29] Other models such as Tobit (Kenkel, 1996) and Finite Mixture Models (Ayyagari et al., 2013) estimate own price elasticity of alcohol demand at an individual-level, however, they do not explicitly take extensive margins into account.

performed simulation analysis, estimate elasticities on a real world application using the TJD et al. method and compare the resultant elasticity estimates to those obtained via the MBM method. In the simulation study, I varied: 1) the level of the relative price influence at the extensive vs. intensive margins; and 2) the fraction of the population who are drinkers. I found that the former has a positive and monotonic effect on the bias, while the influence of the latter peaks when the fraction of drinkers is in the low to mid-level range. As a complement to the simulation study, I applied both methods to one of the models considered by Ruhm et al. (2012) using the same dataset as was analyzed by them. I found the elasticity estimates to be statistically significant from zero and from each other ($\hat{\eta}^{MBM} - \hat{\eta}^{PO} = -0.089$; p-value = .0286). These differences in the raw elasticity estimates could become even more evident when placed in the policy making context of revenue generation. Drawing from a proposed change in federal excise tax on alcohol, I found the difference in tax revenue projections to be substantial; amounting to more than 10% of yearly budget of the NIAAA.[30]

2.3. Models that Do Not Impose Unnecessary Modeling Restrictions

The two-part model underlying the MBM and TJD et al. is unnecessarily restrictive. In the intensive margin of the two-part model, MBM and TJD et al. explicitly include an error term and also make conditional mean assumptions, neither of them are necessary. The restrictions are imposed merely as a matter of convenience – so that the regression parameters can be estimated via OLS. In Chapter 3 of this dissertation, I develop a new specification and estimation of alcohol price elasticity [henceforth UPO

[30] See Chapter 2 for more details

method], which is an unrestricted version of two-part model, causally interpretable, and uses nominal prices of alcohol in their models. To examine the extent of the bias between the widely used AGG-LOG method vs. the UPO method, in Chapter 3, I performed a simulation study, where I varied the relative influence of nominal price at the extensive margin and the fraction of drinkers in the population. I found uniformly positive and substantial bias for almost all combinations and. for a given level of former, as latter increases the bias monotonically increases. I also applied both methods to the dataset analyzed by Ruhm et al. (2012) and found the elasticity estimates to be statistically different from each other ($\hat{\eta}^{AL} - \hat{\eta}^{UPO} = -0.4064$; p-value=0.0277) for a particular demand specification. This difference when placed in the context of revenue forecast from a proposed federal excise tax increase for beer was $207.1 million per year (46.7% of the yearly budget of NIAAA).[31]

2.4. Models that are Cast in terms of Nominal vs. Log Price

The existing studies using aggregate-level models/data and those studies applying the two-part model cast elasticity specifications and estimators in terms of log of prices instead of nominal prices. Apart from the need to adhere to the extant literature, there appears to be no valid reasons to use log prices for specifying and estimating the own price elasticity of alcohol demand. In Chapter 4, I investigate the empirical consequences of modeling alcohol demand in terms of log vs. nominal prices. In that chapter, I developed a version of the UPO method for elasticity specification and estimation (detailed in Chapter 3) that is cast in terms of logged prices [henceforth UPOL method].

[31] See Chapter 3 for more details.

I performed simulation and real data analysis by applying the UPOL method and comparing the results with the corresponding elasticity estimates obtained using the UPO method. In the simulation study, I varied the relative influence of nominal price at the extensive margin and found uniformly positive and substantial bias. However, using a dataset analyzed by Ruhm et al. (2012), I applied both methods and found suggestive evidence of smaller bias in the elasticity estimates ($\hat{\eta}^{UPOL} - \hat{\eta}^{UPO} = 0.0165$, but insignificant). This difference when placed in the context of revenue forecast from a proposed federal excise tax increases for beer was $8.4 million (a 2% yearly budget of NIAAA).[32]

3. Future Research

My dissertation points to interesting avenues for future research. The conceptual, empirical and policy differences in elasticity estimates between the newly introduced unbiased UPO method in my dissertation with other infrequently used alcohol demand models designed for individual-level data (such as Tobit models and Finite Mixture Models) is left for future research. In my dissertation, I focused on just one context of the many policy relevant implications (viz., revenue generation) on using accurate elasticities. I would like to investigate the policy implication of differences in elasticities obtained using the UPO vs. extant methods in the context of optimal alcohol taxation and use of fines as a means of modifying behavior.

Further, alcohol prices can be increased in 'chunks,' and alcohol is also purchased in 'chunks'. Yet, almost all existing literature on own price elasticity of alcohol demand

[32] See Chapter 4 for more details.

uses only point-instantaneous elasticity measure to estimate the response of infinitesimally small change in price of alcohol to change in alcohol consumption. This point-instantaneous elasticity measure may not be able to identify the effects of specific incremental change in alcohol price (through excise taxes) on alcohol consumption. Other two forms of alcohol elasticity measures such as point-incremental elasticity and arc elasticity using nominal price of alcohol could be helpful for policy purposes. Specifying and estimating elasticity that is policy relevant with specific incremental change in alcohol price on consumption of alcohol (point-incremental elasticity) or demand response to change in price between two points (arc elasticity) in an unrestricted causally interpretable two-part modeling context using nominal prices is left for future research.

Appendix A:

Asymptotic Distribution (and Standard Error) of $\hat{\eta}^{MBM}$ in Eqn (2-5)

We may write $\hat{\eta}^{MBM}$ as

$$\hat{\eta}^{MBM} = (1 - \hat{\zeta})\hat{\beta}_{P1} + \hat{\beta}_{P2}$$

where

$$\hat{\zeta} = \sum_{i=1}^{n} \frac{1}{n} \mathcal{Z}(\hat{\beta}_1, W_i)$$

$$\mathcal{Z}(\beta, W) = \Lambda(W\beta_1)$$

$\hat{\beta} = [\hat{\beta}_1' \quad \hat{\beta}_2']'$ (with $\hat{\beta}_1' = [\hat{\beta}_{P1} \quad \hat{\beta}_{X1}']$ and $\hat{\beta}_2' = [\hat{\beta}_{P2} \quad \hat{\beta}_{X2}']$) is the consistent estimate of the parameter vector $\beta = [\beta_1' \quad \beta_2']'$ (with $\beta_1' = [\beta_{P1} \quad \beta_{X1}']$ and $\beta_2' = [\beta_{P2} \quad \beta_{X2}']$)) [the parameters of equations (2-1) and (2-2)] obtained via the two-part protocol culminating in (2-3)

and

$W_i = [P_i \quad X_i]$ denotes the observation on $W = [P \quad X]$ for the ith individual in the sample (i = 1, ..., n). Let $\hat{\tau} = [\hat{\beta}_{P1} \quad \hat{\beta}_{P2} \quad \hat{\zeta}]'$ and $\tau = [\beta_{P1} \quad \beta_{P2} \quad \zeta]'$, where $\text{plim}[\hat{\tau}] = \tau$.

If we could show that

$$\text{AVAR}(\hat{\tau})^{-\frac{1}{2}} \sqrt{n}(\hat{\tau} - \tau) \xrightarrow{d} N(0, I)$$

where the formulation of $\text{AVAR}(\hat{\tau})$ is known, then we could apply the δ-method to obtain the asymptotic variance of $\hat{\eta}^{MBM}$ as

$$\text{avar}(\hat{\eta}^{MBM} - \eta^{MBM}) = c(\tau)\ \text{AVAR}(\hat{\tau})\ c(\tau)'$$

where $c(\tau) = [(1-\zeta) \quad 1 \quad -\beta_{P1}]$. Moreover, if we have a consistent estimator for $\text{AVAR}(\hat{\tau})$, say $\widehat{\text{AVAR}(\hat{\tau})}$ [i.e. $\text{plim}\left[\widehat{\text{AVAR}(\hat{\tau})}\right] = \text{AVAR}(\hat{\tau})$], then we could consistently estimate $\text{avar}(\hat{\eta}^{MBM})$ as

$$\widehat{\text{avar}(\hat{\eta}^{MBM})} = c(\hat{\tau})\ \widehat{\text{AVAR}(\hat{\tau})}\ c(\hat{\tau})'. \tag{A-1}$$

We focus, therefore, on finding the asymptotic distribution of $\hat{\tau}$ and, in particular, the formulation of its asymptotic covariance matrix.

First note that we can write τ as

$$\tau = \Xi\ \theta \tag{A-2}$$

where $\theta' = [\delta' \quad \gamma]$, $\delta = [\beta_1' \quad \beta_2']'$, $\gamma = \zeta$ (recall, $\beta_1' = [\beta_{P1} \quad \beta_{X1}']$ and $\beta_2' = [\beta_{P2} \quad \beta_{X2}']$)

$$\Xi = \begin{bmatrix} \ell_{\beta_{P1}} \\ \hline \ell_{\beta_{P2}} \\ \hline \ell_{\zeta} \end{bmatrix}$$

and ℓ_a is the unit row vector with the value "1" in the element position corresponding to the element position of a in the vector θ. Clearly then

$$\text{AVAR}(\hat{\tau}) = \Xi \; \text{AVAR}(\hat{\theta}) \; \Xi' \tag{A-3}$$

where $\hat{\theta}$ is the estimator of θ obtained from the following two-stage protocol.

First Stage

Consistently estimate $\delta = \beta$ via the following optimization estimator

$$\hat{\delta} = \arg\max_{\tilde{\delta}} \frac{\sum_{i=1}^{n} q_1(\tilde{\delta}, S_i)}{n} \tag{A-4}$$

where

$$q_1(\tilde{\delta}, S_i) = q_{11}(\tilde{\beta}_1, S_i) + q_{12}(\tilde{\beta}_2, S_i)$$

$$q_{11}(\tilde{\beta}_1, S_i) = I(A_i > 0) \ln[\Lambda(W_i \tilde{\beta}_1)] + [1 - I(A_i > 0)] \ln[1 - \Lambda(W_i \tilde{\beta}_1)]$$

$$q_{12}(\tilde{\beta}_2, S_i) = - I(A_i > 0)(\ln(A_i) - W_i \tilde{\beta}_2)^2$$

$$S_i = [A_i \quad X_i \quad P_i]$$

$\tilde{\delta} = [\tilde{\beta}_1' \quad \tilde{\beta}_2']'$, $\tilde{\beta}_1' = [\tilde{\beta}_{P1} \quad \tilde{\beta}_{X1}']$ and $\tilde{\beta}_2' = [\tilde{\beta}_{P2} \quad \tilde{\beta}_{X2}']$ and $\hat{\delta} = [\hat{\beta}_1' \quad \hat{\beta}_2']'$.

Second Stage

Consistently estimate γ via the following optimization estimator

$$\hat{\gamma} = \arg\max_{\tilde{\gamma}} \frac{\sum_{i=1}^{n} q(\hat{\delta}, \tilde{\gamma}, S_i)}{n} \tag{A-5}$$

where

$$q(\hat{\delta}, \tilde{\gamma}, S_i) = -(\mathcal{Z}(\hat{\beta}, W_i) - \tilde{\zeta})^2$$

$\hat{\delta} = [\hat{\beta}_1' \quad \hat{\beta}_2']'$ is the first stage estimator of β, $\hat{\beta}_1' = [\hat{\beta}_{P1} \quad \hat{\beta}_{X1}']$ and $\hat{\beta}_2' = [\hat{\beta}_{P2} \quad \hat{\beta}_{X2}']$. Use q_1 as shorthand notation for

$$q_1(\delta, S) = q_{11}(\beta_1, S) + q_{12}(\beta_2, S)$$

with

$$q_{11}(\beta_1, S) = I(A > 0)\ln[\Lambda(W\beta_1)] + [1 - I(A > 0)]\ln[1 - \Lambda(W\beta_1)]$$

$$q_{12}(\beta_2, S) = -I(A > 0)(\ln(A) - W\beta_2)^2$$

$$S = [A \quad X \quad P]$$

and use q as shorthand notation for

$$q(\delta, \tilde{\gamma}, S) = -(\mathcal{Z}(\beta_2, W) - \zeta)^2$$

and let $AVAR(\hat{\delta})$ denote the asymptotic covariance matrix of the first stage estimator.[33] Using well known results from asymptotic theory for two-stage estimators, we can show that[34]

$$AVAR(\hat{\theta})^{-\frac{1}{2}}\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I) \qquad \text{(A-6)}$$

---

[33] Note that a consistent estimate $AVAR(\hat{\delta})$ can be obtained from the packaged output for the first stage estimator because the first stage estimator is unaffected by the fact that it is a component of a two-stage estimator.

where $\hat{\theta}' = [\hat{\delta}' \quad \hat{\gamma}]$, $\text{plim}(\hat{\theta}) = \theta$

$$\text{AVAR}(\hat{\theta}) = \begin{bmatrix} D_{11} & D_{12} \\ D_{12}' & D_{22} \end{bmatrix} \tag{A-7}$$

$$D_{11} = \underset{2K\times 2K}{\text{AVAR}(\hat{\delta})} \tag{A-8}$$

$$D_{12} = E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\delta} q_1 \, '\nabla_{\gamma} q\right] E\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

$$- E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\delta} q_1 \, '\nabla_{\delta} q_1\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma\delta} q\right]' E\left[\nabla_{\gamma\gamma} q\right]^{-1} \tag{A-9}$$

and

$$D_{22} = \text{AVAR}(\hat{\gamma}) = E\left[\nabla_{\gamma\gamma} q\right]^{-1} \left[E\left[\nabla_{\gamma\delta} q\right] \text{AVAR}(\hat{\beta}) E\left[\nabla_{\gamma\delta} q\right]' \right.$$

$$- E\left[\nabla_{\gamma} q \nabla_{\delta} q_1\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma\delta} q\right]'$$

$$\left. - E\left[\nabla_{\gamma\delta} q\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma} q \nabla_{\delta} q_1\right]' \right] E\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

$$+ E\left[\nabla_{\gamma\gamma} q\right]^{-1} E\left[\nabla_{\gamma} q^2\right] E\left[\nabla_{\gamma\gamma} q\right]^{-1} \tag{A-10}$$

Fortunately, (A-9) and (A-10) can be simplified in a number of ways. First note that we can write

$$E\left[\nabla_{\gamma} q \nabla_{\delta} q_1\right] = E\left[\nabla_{\gamma} q E[\nabla_{\delta} q_1 \mid W]\right]$$

but

---

[34]Discussions of asymptotic theory for two-stage optimization estimators can be found in Newey & McFadden (1994), White (1994, Chapter 6), and Wooldridge (2010, Chapter 12).

$$\nabla_\delta q_1 = \nabla_\delta q_{11} + \nabla_\delta q_{12}$$

with

$$\nabla_\delta q_{11} = [\nabla_{\beta_1} \ln f(I(A > 0) \mid W) \quad 0 \quad 0]$$

$$\nabla_\delta q_{12} = \left[0 \quad 2\, I(A > 0)\left(\ln(A) - W\beta_2\right) W \quad 0\right]$$

where

$$f(I(A > 0) \mid W) = \Lambda(W\beta_1)^{I(A>0)} [1 - \Lambda(W\beta_1)]^{[1 - I(A>0)]} .$$

Therefore

$$E[\nabla_\delta q_1 \mid W] = \left[E[\nabla_{\beta_1} \ln f(I(A > 0) \mid W)] \qquad 2E\left[I(A > 0)\left(\ln(A) - W\beta_2\right) \mid W\right] W\right]$$

$$= [\,0 \quad 0]$$

because $E[\nabla_{\beta_1} \ln f(I(A > 0) \mid W)] = 0$ [see (13.20) on p. 477 of Wooldridge (2010)] and,

$E\left[I(A > 0)\left(\ln(A) - W\beta_2\right) \mid W\right] = 0$ by design. Finally, then we get

$$E\left[\nabla_\gamma q \nabla_\delta q_1\right] = 0$$

so

$$D_{12} = \underset{2K\times 2K}{-\,E\left[\nabla_{\delta\delta} q_1\right]^{-1}} \underset{2K\times 2K}{E\left[\nabla_\delta q_1 \,' \nabla_\delta q_1\right]} \underset{2K\times 2K}{E\left[\nabla_{\delta\delta} q_1\right]^{-1}} \underset{2K\times 1}{E\left[\nabla_{\gamma\delta} q\right]'} \underset{1\times 1}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \qquad \text{(A-11)}$$

and

$$D_{22} = AVAR(\hat{\gamma}) = \underset{1\times 1}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \underset{1\times 2K}{E\left[\nabla_{\gamma\delta} q\right]} \underset{2K\times 2K}{AVAR(\hat{\beta})} \underset{2K\times 1}{E\left[\nabla_{\gamma\delta} q\right]'} \underset{1\times 1}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}}$$

$$+ \underset{1\times 1}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \underset{1\times 1}{E\left[\nabla_\gamma q^2\right]} \underset{1\times 1}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} . \qquad \text{(A-12)}$$

so

$$\underset{(2K+1)\times(2K+1)}{\mathrm{AVAR}(\hat{\theta})} = \begin{bmatrix} \underset{2K\times 2K}{D_{11}} & \underset{2K\times 1}{D_{12}} \\ \underset{1\times 2K}{D_{12}'} & \underset{1\times 1}{D_{22}} \end{bmatrix}.$$

Let's consider each of the individual components of (A-11) and (A-12) in turn.

$\underset{2K\times 2K}{\mathbf{E[\nabla_{\delta\delta} q_1]^{-1}}}$

Written out explicitly we have

$$E[\nabla_{\beta\beta} q_1]^{-1} = \begin{bmatrix} E[\nabla_{\beta_1\beta_1} q_{11}]^{-1} & 0 \\ 0 & E[\nabla_{\beta_2\beta_2} q_{12}]^{-1} \end{bmatrix}. \tag{A-13}$$

Now

$$E[\nabla_{\beta_1\beta_1} q_{11}]^{-1} = -\mathrm{AVAR}(\hat{\beta}_1) \tag{A-14}$$

= the negative of the asymptotic covariance matrix for first stage, first part, logit estimation in the two-stage estimation protocol for θ. We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{\beta_1\beta_1} q_{11}]^{-1}$ is

$$\hat{E}[\nabla_{\beta_1\beta_1} q_{11}]^{-1} = -n\widehat{\mathrm{AVAR}}^*(\hat{\beta}_1) \tag{A-15}$$

where $\widehat{\mathrm{AVAR}}^*(\hat{\beta}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure. Also

$$\nabla_{\beta_2} q_{12} = 2\,I(A > 0)(\ln(A) - W\beta_2)W$$

and

$$\nabla_{\beta_2\beta_2} q_{12} = -2\,\mathrm{I}(A > 0)W'W\,.$$

Therefore

$$\mathrm{E}[\nabla_{\beta_2\beta_2} q_{12}] = -2\mathrm{E}\left[\mathrm{I}(A > 0)W'W\right].$$

A consistent estimator of $\mathrm{E}[\nabla_{\beta_2\beta_2} q_{12}]^{-1}$ is

$$\hat{\mathrm{E}}[\nabla_{\beta_2\beta_2} q_{12}]^{-1} = n_1\left[-2\sum_{i=1}^{n}\left\{\mathrm{I}(A_i > 0)W_i'W_i\right\}\right]^{-1} \qquad \text{(A-17)}$$

where n_1 is the size of the subsample for whom $\mathrm{I}(A > 0) = 1$, so

$$\hat{\mathrm{E}}[\nabla_{\delta\delta} q_1]^{-1} = \hat{\mathrm{E}}[\nabla_{\beta\beta} q_1]^{-1} = \begin{bmatrix} \hat{\mathrm{E}}[\nabla_{\beta_1\beta_1} q_{11}]^{-1} & 0 \\ 0 & \hat{\mathrm{E}}[\nabla_{\beta_2\beta_2} q_{12}]^{-1} \end{bmatrix}. \qquad \text{(A-18)}$$

$\mathbf{E}\left[\nabla_{\delta} \mathbf{q_1}' \nabla_{\delta} \mathbf{q_1}\right]$
2K×2K

Written out explicitly we have

$$\mathrm{E}\left[\nabla_{\beta} q_1' \nabla_{\beta} q_1\right] = \begin{bmatrix} \mathrm{E}[\nabla_{\beta_1} q_{11}' \nabla_{\beta_1} q_{11}] & \mathrm{E}[\nabla_{\beta_1} q_{11}' \nabla_{\beta_2} q_{12}] \\ \mathrm{E}[\nabla_{\beta_2} q_{12}' \nabla_{\beta_1} q_{11}] & \mathrm{E}[\nabla_{\beta_2} q_{12}' \nabla_{\beta_2} q_{12}] \end{bmatrix}. \qquad \text{(A-19)}$$

Because the first stage, first part, estimator of β_1 is MLE we can write

$$\mathrm{E}[\nabla_{\beta_1} q_{11}' \nabla_{\beta_1} q_{11}] = -\mathrm{E}[\nabla_{\beta_1\beta_1} q_{11}] = \left[\mathrm{AVAR}(\hat{\beta}_1)\right]^{-1}$$

= the inverse of the asymptotic covariance matrix for first stage,

first part, logit estimation in the two-stage estimation protocol for θ. We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{\beta_1} q_{11}' \nabla_{\beta_1} q_{11}]$ is

$$\hat{E}[\nabla_{\beta_1} q_{11}' \nabla_{\beta_1} q_{11}] = \frac{1}{n}\left[\widehat{AVAR} * (\hat{\beta}_1)\right]^{-1} \tag{A-20}$$

where $\widehat{AVAR} * (\hat{\beta}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure. The remaining block elements follow from

$$\nabla_{\beta_1} q_{11} = \nabla_{\beta_1} \ln f(I(A > 0) \mid W) = \left[I(A > 0)[1 - \Lambda(W\beta_1)] - [1 - I(A > 0)]\Lambda(W\beta_1)\right]W \tag{A-21}$$

$$\nabla_{\beta_2} q_{12} = 2\, I(A > 0)\left(\ln(A \mid A > 0) - W\beta_2\right)W\,. \tag{A-22}$$

where the formulation of $\nabla_{\beta_1} q_{11}$ comes from equation (16.4.8) on p. 350 of Fomby et al. (1984) and $\Lambda(\ \ )$ denotes the logistic cdf. The remaining required consistent matrix estimators are

$$\hat{E}[\nabla_{\beta_1} q_{11}' \nabla_{\beta_2} q_{12}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{\beta_1} \hat{q}_{11i}' \nabla_{\beta_2} \hat{q}_{12i} \tag{A-23}$$

$$\hat{E}[\nabla_{\beta_2} q_{12}' \nabla_{\beta_2} q_{12}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{\beta_2} \hat{q}_{12i}' \nabla_{\beta_2} \hat{q}_{12i} \tag{A-24}$$

where

$$\nabla_{\beta_1} \hat{q}_{11i} = \left[I(A_i > 0)[1 - \Lambda(W_i\hat{\beta}_1)] - [1 - I(A_i > 0)]\Lambda(W_i\hat{\beta}_1)\right]W_i \tag{A-25}$$

and

$$\nabla_{\beta_2}\hat{q}_{12i} = 2\,I(A_i > 0)\left(\ln(A_i \mid A_i > 0) - W_i\hat{\beta}_2\right)W_i \tag{A-26}$$

so

$$\hat{E}\left[\nabla_\delta q_1 '\nabla_\delta q_1\right] = \hat{E}\left[\nabla_\beta q_1 '\nabla_\beta q_1\right] = \begin{bmatrix} \hat{E}[\nabla_{\beta_1} q_{11} '\nabla_{\beta_1} q_{11}] & \hat{E}[\nabla_{\beta_1} q_{11} '\nabla_{\beta_2} q_{12}] \\ \hat{E}[\nabla_{\beta_2} q_{12} '\nabla_{\beta_1} q_{11}] & \hat{E}[\nabla_{\beta_2} q_{12} '\nabla_{\beta_2} q_{12}] \end{bmatrix}. \tag{A-27}$$

$$\underset{1\times 2K}{\mathbf{E}\left[\boldsymbol{\nabla}_{\gamma\delta}\mathbf{q}\right]}$$

Written out explicitly we have

$$\nabla_\gamma q = \nabla_\zeta q_c = 2\left(\mathcal{Z}(\beta, S) - \zeta\right) \tag{A-28}$$

and

$$E[\nabla_{\gamma\delta} q] = \left[E[\nabla_{\zeta\beta_1} q] \quad 0\right]$$

$$= 2\left[E[\nabla_{\beta_1}\mathcal{Z}] \quad 0\right] \tag{A-29}$$

where

$$\nabla_{\beta_1}\mathcal{Z} = \lambda(W\beta_1)\,W\,. \tag{A-30}$$

The requisite consistent matrix estimator is

$$\hat{E}[\nabla_{\beta_1}\mathcal{Z}] = \sum_{i=1}^{n}\widehat{\nabla_{\beta_1}\mathcal{Z}}_i \tag{A-31}$$

where

$$\widehat{\nabla_{\beta_1}\mathcal{Z}}_i = \lambda(W_i\hat{\beta}_1)\,W_i \tag{A-32}$$

so

$$\hat{E}[\nabla_{\gamma\delta} q] = 2\left[\hat{E}[\nabla_{\beta_1} \mathcal{Z}] \quad 0\right]. \tag{A-33}$$

$\underset{1\times 1}{\mathbf{E}[\nabla_{\gamma\gamma}\mathbf{q}]^{-1}}$

$$E\left[\nabla_{\gamma\gamma} q\right]^{-1} = E\left[\nabla_{\zeta\zeta} q_c\right]^{-1} = -\frac{1}{2}. \tag{A-34}$$

$\underset{1\times 1}{\mathbf{E}\left[\nabla_{\gamma}\mathbf{q}^{2}\right]}$

Given that

$$\nabla_{\gamma} q = 2\left(\mathcal{Z}(\beta,\, W) - \zeta\right). \tag{A-35}$$

we have

$$E\left[\nabla_{\gamma} q^2\right] = 4E[(\mathcal{Z}(\beta,\, S) - \zeta)^2] \tag{A-36}$$

The corresponding consistent estimator is

$$\hat{E}\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right] = 4\sum_{i=1}^{n} \frac{1}{n}(\hat{\mathcal{Z}}_i - \hat{\zeta})^2. \tag{A-37}$$

Based on (A-8), (A-11) and (A-12) and using the two-stage estimator $\hat{\theta}$ we can consistently estimate (A-7) as

$$\widehat{\text{AVAR}(\hat{\theta})} = \begin{bmatrix} \hat{D}_{11} & \hat{D}_{12} \\ \hat{D}_{12}' & \hat{D}_{22} \end{bmatrix}$$

where

$$\hat{D}_{11} = \widehat{\text{AVAR}}(\hat{\delta}) = \widehat{\text{AVAR}}(\hat{\beta}) = \begin{bmatrix} n\widehat{\text{AVAR}} * (\hat{\beta}_1) & 0 \\ 0 & n_1\widehat{\text{AVAR}} * (\hat{\beta}_2) \end{bmatrix}$$

$$D_{12} = -\hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\delta} q_1 ' \nabla_{\delta} q_1\right] \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma\delta} q\right]' \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

$$D_{22} = \widehat{\text{AVAR}(\hat{\gamma})} = \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma\delta} q\right] \widehat{\text{AVAR}(\hat{\beta})} \hat{E}\left[\nabla_{\gamma\delta} q\right]' \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

$$+ \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma} q^2\right] \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

and using well known results from asymptotic theory for two-stage estimators, we can show that[35]

$$\widehat{\text{AVAR}(\hat{\theta})}^{-\frac{1}{2}} \begin{bmatrix} \sqrt{n}(\hat{\beta}_1 - \beta_1) \\ \sqrt{n_1}(\hat{\beta}_2 - \hat{\beta}_2) \\ \sqrt{n}(\hat{\zeta} - \zeta) \end{bmatrix} \xrightarrow{d} N(0, I). \qquad \text{(A-38)}$$

****

ASIDE:

---

[35]Discussions of asymptotic theory for two-stage optimization estimators can be found in Newey & McFadden (1994), White (1994, Chapter 6), and Wooldridge (2010, Chapter 12).

Notice that the "$\sqrt{n}$ blow up" is a bit tricky here. It implements $\sqrt{n}$ for $\hat{\beta}_1$ and $\hat{\zeta}$; but uses $\sqrt{n_1}$ for $\hat{\beta}_2$. We had to do this because we had to use the correct sample size (viz., n_1) for a number of the components of $\widehat{AVAR(\hat{\theta})}$ [viz., those that pertained to the estimation of $\hat{\beta}_2$]; in particular (A-17), (A-23) and (A-24). For this reason we had to be explicit about the denominators in all of the averages for the components of $\widehat{AVAR(\hat{\theta})}$. This meant that in the construction of the requisite asymptotic t-stats we had to explicitly include the "blow-up" in the numerator (i.e., we had to multiply by the square-root of the appropriate sample size). I refer to this as "tricky" because one typically does not have to do this. In the usual asymptotic t-stat construction the denominators of the averages ("n") need not be included in the construction of the asymptotic covariance matrix because it typically manifests as a multiplicative factor and, after pulling the diagonal and taking the square root to get the standard errors, this multiplicative $\sqrt{n}$ cancels with the "blow-up" factor in the numerator. For example, the asymptotic t-stat of the OLS estimator is

$$\frac{\sqrt{n}(\hat{\rho}_k - \rho_k)}{\sqrt{\widehat{AVAR(\hat{\rho})}}} = \frac{\sqrt{n}(\hat{\rho}_k - \rho_k)}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n}\mathcal{X}'\mathcal{X}_{kk}\right)^{-1}}} = \frac{\sqrt{n}(\hat{\rho}_k - \rho_k)}{\sqrt{n}\sqrt{\hat{\sigma}^2(\mathcal{X}'\mathcal{X}_{kk})^{-1}}} = \frac{(\hat{\rho}_k - \rho_k)}{\sqrt{\hat{\sigma}^2(\mathcal{X}'\mathcal{X}_{kk})^{-1}}}$$

where

n is the sample size

ρ_k is the coefficient of the kth regressor in the linear regression

$\hat{\rho}_k$ is its OLS estimator

σ^2 is the regression error variance estimator

$\mathcal{X}$ is the matrix of regressors

and $\mathcal{X}'\mathcal{X}_{kk}$ is the kth diagonal element of $\mathcal{X}'\mathcal{X}$. Note how the "$\sqrt{n}$ s" simply cancel. Note also that what we typically refer to as the "asymptotic standard error" can actually be written as the square root of the diagonal element of the consistent estimator of the asymptotic covariance matrix divided by n; in other words

$$\text{asy std err} = \sqrt{\frac{\widehat{\text{AVAR}(\hat{\rho})}}{n}}.$$

****

Now back to the issue at hand. Moreover

$$\widehat{\text{AVAR}(\hat{\tau})}^{-\frac{1}{2}}\sqrt{n}\begin{bmatrix}\sqrt{n}(\hat{\beta}_{P1} - \beta_{P1}) \\ \sqrt{n_1}(\hat{\beta}_{P2} - \hat{\beta}_{P2}) \\ \sqrt{n}(\hat{\zeta} - \zeta)\end{bmatrix} \xrightarrow{d} N(0, I). \tag{A-39}$$

where

$$\widehat{\text{AVAR}(\hat{\tau})} = \Xi\,\widehat{\text{AVAR}(\hat{\theta})}\,\Xi' \tag{A-40}$$

and τ and Ξ are defined as in (A-2). Now combining (A-1) with (A-38) and (A-39) we get

$$\widehat{\text{avar}(\hat{\eta}^{MBM})}^{-\frac{1}{2}}\sqrt{n}(\hat{\eta}^{MBM} - \eta^{MBM}) \xrightarrow{d} N(0, I) \tag{A-41}$$

where

$$\widehat{\text{avar}(\hat{\eta}^{\text{MBM}})} = c(\hat{\tau})\ \widehat{\text{AVAR}(\hat{\tau})}\ c(\hat{\tau})'$$

and

$$c(\tau) = [(1-\zeta) \quad 1 \quad -\beta_{P1}].$$

Appendix B:

Asymptotic Distribution (and Standard Error) of $\hat{\eta}^{PO}$ in Eqn (2-7)

We may write $\hat{\eta}^{PO}$ as

$$\hat{\eta}^{PO} = \frac{\hat{\kappa}}{\hat{\nu}}$$

where

$$\hat{\kappa} = \sum_{i=1}^{n} \frac{1}{n} \mathcal{K}(\hat{\beta}, W_i)$$

$$\hat{\nu} = \sum_{i=1}^{n} \frac{1}{n} \mathcal{V}(\hat{\beta}, W_i)$$

$$\mathcal{K}(\beta, W) = \lambda(W\beta_1)\exp(W\beta_2)\beta_{P1} + \Lambda(W\beta_1)\exp(W\beta_2)\beta_{P2}$$

$$\mathcal{V}(\beta, W) = \Lambda(W\beta_1)\exp(W\beta_2)$$

$\hat{\beta} = [\hat{\beta}_1' \quad \hat{\beta}_2']'$ (with $\hat{\beta}_1' = [\hat{\beta}_{P1} \quad \hat{\beta}_{X1}']$ and $\hat{\beta}_2' = [\hat{\beta}_{P2} \quad \hat{\beta}_{X2}']$) is the consistent estimate of the parameter vector $\beta = [\beta_1' \quad \beta_2']'$ (with $\beta_1' = [\beta_{P1} \quad \beta_{X1}']$ and $\beta_2' = [\beta_{P2} \quad \beta_{X2}']$)) [the parameters of equations (2-1) and (2-2)] obtained via the two-part protocol culminating in (2-3)

and

$W_i = [P_i \quad X_i]$ denotes the observation on $W = [P \quad X]$ for the ith individual in the sample (i = 1, ..., n). Let $\hat{\gamma} = [\hat{\kappa} \quad \hat{\nu}]'$ and $\gamma = [\kappa \quad \nu]'$, where $\text{plim}[\hat{\gamma}] = \gamma$. If we could show that

$$\text{AVAR}(\hat{\gamma})^{-\frac{1}{2}}\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, I)$$

where the formulation of $\text{AVAR}(\hat{\gamma})$ is known, then we could apply the δ-method to obtain the asymptotic variance of $\hat{\eta}^{PO}$ as

$$\text{avar}(\hat{\eta}^{PO}) = c(\gamma)\ \text{AVAR}(\hat{\gamma})\ c(\gamma)'$$

where $c(\tau) = [1/\nu \quad -\kappa/\nu^2]$. Moreover, if we have a consistent estimator for $\text{AVAR}(\hat{\gamma})$, say $\widehat{\text{AVAR}(\hat{\gamma})}$ [i.e. $\text{plim}\left[\widehat{\text{AVAR}(\hat{\gamma})}\right] = \text{AVAR}(\hat{\gamma})$], then we could consistently estimate $\text{avar}(\hat{\eta}^{PO})$ as

$$\widehat{\text{avar}(\hat{\eta}^{PO})} = c(\hat{\gamma})\ \widehat{\text{AVAR}(\hat{\gamma})}\ c(\hat{\gamma})'. \tag{B-1}$$

We focus, therefore, on finding the asymptotic distribution of $\hat{\gamma}$ and, in particular, the formulation of its asymptotic covariance matrix.

First, let $\theta' = [\delta' \quad \gamma']$ where $\delta = [\beta_1' \quad \beta_2']'$, $\gamma' = [\kappa \quad \nu]$ (recall, $\beta_1' = [\beta_{P1} \quad \beta_{X1}']$ and $\beta_2' = [\beta_{P2} \quad \beta_{X2}']$), and note that $\hat{\gamma}$ can be viewed as the second stage estimator in the following two-stage protocol

First Stage

Consistently estimate δ via the following optimization estimator

$$\hat{\delta} = \arg\max_{\tilde{\delta}} \frac{\sum_{i=1}^{n} q_1(\tilde{\delta}, S_i)}{n} \tag{B-2}$$

where

$$q_1(\tilde{\delta}, S_i) = q_{11}(\tilde{\beta}_1, S_i) + q_{12}(\tilde{\beta}_2, S_i)$$

$$q_{11}(\tilde{\beta}_1, S_i) = I(A_i > 0)\ln[\Lambda(W_i \tilde{\beta}_1)] + [1 - I(A_i > 0)]\ln[1 - \Lambda(W_i \tilde{\beta}_1)]$$

$$q_{12}(\tilde{\beta}_2, S_i) = -I(A_i > 0)(\ln(A_i) - W_i \tilde{\beta}_2)^2$$

$$S_i = [A_i \quad X_i \quad P_i]$$

$\tilde{\delta} = [\tilde{\beta}_1' \quad \tilde{\beta}_2']'$, $\tilde{\beta}_1' = [\tilde{\beta}_{P1} \quad \tilde{\beta}_{X1}']$ and $\tilde{\beta}_2' = [\tilde{\beta}_{P2} \quad \tilde{\beta}_{X2}']$ and $\hat{\delta} = [\hat{\beta}_1' \quad \hat{\beta}_2']'$

Second Stage

Consistently estimate γ via the following optimization estimator

$$\hat{\gamma} = \arg\max_{\tilde{\gamma}} \frac{\sum_{i=1}^{n} q(\hat{\delta}, \tilde{\gamma}, S_i)}{n} \tag{B-3}$$

where

$$q(\hat{\delta}, \tilde{\gamma}, S_i) = q_a(\hat{\delta}, \tilde{\kappa}, S_i) + q_b(\hat{\delta}, \tilde{\nu}, S_i)$$

$$q_a(\hat{\delta}, \tilde{\kappa}, S_i) = -(\mathcal{K}(\hat{\beta}, W_i) - \tilde{\kappa})^2$$

$$q_b(\hat{\delta}, \tilde{\nu}, S_i) = -(\mathcal{V}(\hat{\beta}, W_i) - \tilde{\nu})^2$$

$\hat{\delta} = [\hat{\beta}_1' \quad \hat{\beta}_2']'$ is the first stage estimator of δ, $\hat{\beta}_1' = [\hat{\beta}_{P1} \quad \hat{\beta}_{X1}']$ and $\hat{\beta}_2' = [\hat{\beta}_{P2} \quad \hat{\beta}_{X2}']$. Use q_l as shorthand notation for

$$q_l(\delta, S) = q_{l1}(\beta_1, S) + q_{l2}(\beta_2, S)$$

with

$$q_{11}(\beta_1, S) = I(A > 0)\ln[\Lambda(W\beta_1)] + [1 - I(A > 0)]\ln[1 - \Lambda(W\beta_1)]$$

$$q_{12}(\beta_2, S) = -I(A > 0)(\ln(A) - W\beta_2)^2$$

$$S = [A \quad X \quad P]$$

and use q as shorthand notation for

$$q(\delta, \tilde{\gamma}, S) = q_a(\delta, \kappa, S) + q_b(\delta, \nu, S)$$

with

$$q_a(\delta, \kappa, S) = -(\mathcal{K}(\beta, W) - \kappa)^2$$

$$q_b(\delta, \nu, S) = -(\mathcal{V}(\beta, W) - \nu)^2$$

and let $AVAR(\hat{\delta})$ denote the asymptotic covariance matrix of the first stage estimator.[36] Using well known results from asymptotic theory for two-stage estimators, we can show that[37]

$$AVAR(\hat{\gamma})^{-\frac{1}{2}}\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, I) \qquad \text{(B-4)}$$

where, $plim(\hat{\gamma}) = \gamma$

$$AVAR(\hat{\gamma}) = E\left[\nabla_{\gamma\gamma} q\right]^{-1}\left[E\left[\nabla_{\gamma\delta} q\right] AVAR(\hat{\beta}) E\left[\nabla_{\gamma\delta} q\right]'\right.$$

---

[36] Note that a consistent estimate $AVAR(\hat{\delta})$ can be obtained from the packaged output for the first stage estimator because the first stage estimator is unaffected by the fact that it is a component of a two-stage estimator.

[37] Discussions of asymptotic theory for two-stage optimization estimators can be found in Newey & McFadden (1994), White (1994, Chapter 6), and Wooldridge (2010, Chapter 12).

$$- \; E\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma\delta} q\right]'$$

$$- \; E\left[\nabla_{\gamma\delta} q\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right]' \Big] E\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

$$+ \; E\left[\nabla_{\gamma\gamma} q\right]^{-1} E\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right] E\left[\nabla_{\gamma\gamma} q\right]^{-1} \qquad \text{(B-5)}$$

Fortunately, (B-5) can be simplified in a number of ways. Note that we can write

$$E\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right] = E\left[\nabla_{\gamma} q' E[\nabla_{\delta} q_1 \mid W]\right]$$

but

$$\nabla_{\delta} q_1 = \nabla_{\delta} q_{11} + \nabla_{\delta} q_{12}$$

with

$$\nabla_{\delta} q_{11} = [\nabla_{\beta_1} \ln f(I(A > 0) \mid W) \quad 0 \quad 0]$$

$$\nabla_{\delta} q_{12} = \left[0 \quad 2\, I(A > 0)\left(\ln(A) - W\beta_2\right) W \quad 0\right]$$

where

$$f(I(A > 0) \mid W) = \Lambda(W\beta_1)^{I(A>0)} [1 - \Lambda(W\beta_1)]^{[1 - I(A>0)]} \; .$$

Therefore

$$E[\nabla_{\delta} q_1 \mid W] = \left[E[\nabla_{\beta_1} \ln f(I(A > 0) \mid W)] \qquad 2E\left[I(A > 0)\left(\ln(A) - W\beta_2\right) \mid W\right] W\right]$$

$$= [\,0 \quad 0]$$

because $E[\nabla_{\beta_1} \ln f(I(A > 0) \mid W)] = 0$ [see (13.20) on p. 477 of Wooldridge (2010)] and,

$E\left[I(A > 0)\left(\ln(A) - W\beta_2\right) \mid W\right] = 0$ by design. Finally, then we get

$$E\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right] = 0$$

so

$$\underset{2\times2}{AVAR(\hat{\gamma})} = \underset{2\times2}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \underset{2\times2K}{E\left[\nabla_{\gamma\delta} q\right]} \underset{2K\times2K}{AVAR(\hat{\beta})} \underset{2K\times2}{E\left[\nabla_{\gamma\delta} q\right]'} \underset{2\times2}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}}$$

$$+ \underset{2\times2}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \underset{2\times2}{E\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right]} \underset{2\times2}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}}. \qquad \text{(B-6)}$$

Let's consider each of the individual components of (B-6) in turn.

$E[\nabla_{\delta\delta} q_1]^{-1}$
2K×2K

Written out explicitly we have

$$E[\nabla_{\beta\beta} q_1]^{-1} = \begin{bmatrix} E[\nabla_{\beta_1\beta_1} q_{11}]^{-1} & 0 \\ 0 & E[\nabla_{\beta_2\beta_2} q_{12}]^{-1} \end{bmatrix}. \qquad \text{(B-7)}$$

Now

$$E[\nabla_{\beta_1\beta_1} q_{11}]^{-1} = -AVAR(\hat{\beta}_1) \qquad \text{(B-8)}$$

= the negative of the asymptotic covariance matrix for first stage, first part, logit estimation in the two-stage estimation protocol for θ. We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{\beta_1\beta_1} q_{11}]^{-1}$ is

$$\hat{E}[\nabla_{\beta_1\beta_1} q_{11}]^{-1} = -n\widehat{AVAR}*(\hat{\beta}_1) \qquad \text{(B-9)}$$

where $\widehat{\mathrm{AVAR}}^*(\hat{\beta}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure. Also

$$\nabla_{\beta_2} q_{12} = 2\,\mathrm{I}(\mathrm{A} > 0)(\ln(\mathrm{A}) - \mathrm{W}\beta_2)\mathrm{W}$$

and

$$\nabla_{\beta_2\beta_2} q_{12} = -2\,\mathrm{I}(\mathrm{A} > 0)\mathrm{W}'\mathrm{W}\,.$$

Therefore

$$\mathrm{E}[\nabla_{\beta_2\beta_2} q_{12}] = -2\mathrm{E}\left[\mathrm{I}(\mathrm{A} > 0)\mathrm{W}'\mathrm{W}\right]. \tag{B-10}$$

A consistent estimator of $\mathrm{E}[\nabla_{\beta_2\beta_2} q_{12}]^{-1}$ is

$$\hat{\mathrm{E}}[\nabla_{\beta_2\beta_2} q_{12}]^{-1} = \mathrm{n}_1\left[-2\sum_{\mathrm{i}=1}^{\mathrm{n}}\left\{\mathrm{I}(\mathrm{A}_\mathrm{i} > 0)\mathrm{W}_\mathrm{i}'\mathrm{W}_\mathrm{i}\right\}\right]^{-1} \tag{B-11}$$

where n_1 is the size of the subsample for whom $\mathrm{I}(\mathrm{A} > 0) = 1$, so

$$\hat{\mathrm{E}}[\nabla_{\delta\delta} q_1]^{-1} = \hat{\mathrm{E}}[\nabla_{\beta\beta} q_1]^{-1} = \begin{bmatrix} \hat{\mathrm{E}}[\nabla_{\beta_1\beta_1} q_{11}]^{-1} & 0 \\ 0 & \hat{\mathrm{E}}[\nabla_{\beta_2\beta_2} q_{12}]^{-1} \end{bmatrix}. \tag{B-12}$$

$\mathbf{E}\left[\nabla_{\delta} q_1 \,'\nabla_{\delta} q_1\right]$
2K×2K

Written out explicitly we have

$$\mathrm{E}\left[\nabla_{\beta} q_1 \,'\nabla_{\beta} q_1\right] = \begin{bmatrix} \mathrm{E}[\nabla_{\beta_1} q_{11} \,'\nabla_{\beta_1} q_{11}] & \mathrm{E}[\nabla_{\beta_1} q_{11} \,'\nabla_{\beta_2} q_{12}] \\ \mathrm{E}[\nabla_{\beta_2} q_{12} \,'\nabla_{\beta_1} q_{11}] & \mathrm{E}[\nabla_{\beta_2} q_{12} \,'\nabla_{\beta_2} q_{12}] \end{bmatrix}. \tag{B-13}$$

Because the first stage, first part, estimator of β_1 is MLE we can write

$$E[\nabla_{\beta_1} q_{11} '\nabla_{\beta_1} q_{11}] = -E[\nabla_{\beta_1\beta_1} q_{11}] = \left[AVAR(\hat{\beta}_1) \right]^{-1}$$

= the inverse of the asymptotic covariance matrix for first stage, first part, logit estimation in the two-stage estimation protocol for θ. We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{\beta_1} q_{11} '\nabla_{\beta_1} q_{11}]$ is

$$\hat{E}[\nabla_{\beta_1} q_{11} '\nabla_{\beta_1} q_{11}] = \frac{1}{n}\left[\widehat{AVAR} * (\hat{\beta}_1) \right]^{-1} \qquad \text{(B-14)}$$

where $\widehat{AVAR} * (\hat{\beta}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure. The remainder of the block elements follow from

$$\nabla_{\beta_1} q_{11} = \nabla_{\beta_1} \ln f(I(A > 0) \mid W)$$

$$= \left[I(A > 0)[1 - \Lambda(W\beta_1)] - [1 - I(A > 0)]\Lambda(W\beta_1) \right] W \qquad \text{(B-15)}$$

$$\nabla_{\beta_2} q_{12} = 2\, I(A > 0)\left(\ln(A \mid A > 0) - W\beta_2 \right) W . \qquad \text{(B-16)}$$

where the formulation of $\nabla_{\beta_1} q_{11}$ comes from equation (16.4.8) on p. 350 of Fomby et al. (1984) and $\Lambda(\ \ )$ denotes the logistic cdf. The remaining required consistent matrix estimators are:

$$\hat{E}[\nabla_{\beta_1} q_{11}' \nabla_{\beta_2} q_{12}] = \frac{1}{n_1} \sum_{i=1}^{n} \nabla_{\beta_1} \hat{q}_{11i}' \nabla_{\beta_2} \hat{q}_{12i} \tag{B-17}$$

$$\hat{E}[\nabla_{\beta_2} q_{12}' \nabla_{\beta_2} q_{12}] = \frac{1}{n_1} \sum_{i=1}^{n} \nabla_{\beta_2} \hat{q}_{12i}' \nabla_{\beta_2} \hat{q}_{12i} \tag{B-18}$$

where

$$\nabla_{\beta_1} \hat{q}_{11i} = \left[I(A_i > 0)[1 - \Lambda(W_i \hat{\beta}_1)] - [1 - I(A_i > 0)] \Lambda(W_i \hat{\beta}_1) \right] W_i \tag{B-19}$$

and

$$\nabla_{\beta_2} \hat{q}_{12i} = 2 I(A_i > 0) \left(\ln(A_i \mid A_i > 0) - W_i \hat{\beta}_2 \right) W_i \tag{B-20}$$

so

$$\hat{E}\left[\nabla_{\delta} q_1' \nabla_{\delta} q_1\right] = \hat{E}\left[\nabla_{\beta} q_1' \nabla_{\beta} q_1\right] = \begin{bmatrix} \hat{E}[\nabla_{\beta_1} q_{11}' \nabla_{\beta_1} q_{11}] & \hat{E}[\nabla_{\beta_1} q_{11}' \nabla_{\beta_2} q_{12}] \\ \hat{E}[\nabla_{\beta_2} q_{12}' \nabla_{\beta_1} q_{11}] & \hat{E}[\nabla_{\beta_2} q_{12}' \nabla_{\beta_2} q_{12}] \end{bmatrix}. \tag{B-21}$$

$\mathbf{E}\left[\boldsymbol{\nabla}_{\gamma\delta}\mathbf{q}\right]$
2×2K

Written out explicitly we have

$$\nabla_{\gamma} q = [(\nabla_{\kappa} q_a + 0) \quad (0 + \nabla_{\nu} q_b)]$$

$$= [\nabla_{\kappa} q_a \quad \nabla_{\nu} q_b]$$

$$= 2\left[\left(\mathcal{K}(\beta, S) - \kappa \right) \ \left(\mathcal{V}(\beta, S) - \nu \right) \right] \tag{B-22}$$

and

$$E[\nabla_{\gamma\delta} q] = \begin{bmatrix} E[\nabla_{\kappa\beta_1} q_a] & E[\nabla_{\kappa\beta_2} q_a] \\ E[\nabla_{\nu\beta_1} q_b] & E[\nabla_{\nu\beta_2} q_b] \end{bmatrix}$$

$$= 2 \begin{bmatrix} E[\nabla_{\beta_1} \mathcal{K}] & E[\nabla_{\beta_2} \mathcal{K}] \\ E[\nabla_{\beta_1} \mathcal{V}] & E[\nabla_{\beta_2} \mathcal{V}] \end{bmatrix} \tag{B-23}$$

where

$$\nabla_{\beta_1}\mathcal{K} = \{\lambda(W\beta_1)[1-2\Lambda(W\beta_1)]\beta_{P1}W + \lambda(W\beta_1)[1 \quad 0 \ldots 0]\}\exp(W\beta_2) + \lambda(W\beta_1)\exp(W\beta_2)\beta_{P2}W$$
$$= \lambda(W\beta_1)\exp(W\beta_2)\left[\{[1-2\Lambda(W\beta_1)]\beta_{P1} + \beta_{P2}\}W + [1 \quad 0 \ldots 0]\right] \qquad \text{(B-24)}$$

$$\nabla_{\beta_2}\mathcal{K} = \lambda(W\beta_1)\exp(W\beta_2)\beta_{P1}W + \Lambda(W\beta_1)\{\exp(W\beta_2)\beta_{P2}W + \exp(W\beta_2)[1 \quad 0 \ldots 0]\}$$
$$= \exp(W\beta_2)\left[\{\lambda(W\beta_1)\beta_{P1} + \Lambda(W\beta_1)\beta_{P2}\}W + \Lambda(W\beta_1)[1 \quad 0 \ldots 0]\right]. \qquad \text{(B-25)}$$

$$\nabla_{\beta_1}\mathcal{V} = \lambda(W\beta_1)\exp(W\beta_2)W \qquad \text{(B-26)}$$

and

$$\nabla_{\beta_2}\mathcal{V} = \Lambda(W\beta_1)\exp(W\beta_2)W. \qquad \text{(B-27)}$$

The following equalities were used in deriving the above results

$$\nabla_a\Lambda(a) = \lambda(a) = \Lambda(a)[1-\Lambda(a)]$$

$$\nabla_a\lambda(a) = \lambda(a)[1-2\Lambda(a)].$$

The requisite consistent matrix estimators are

$$\hat{E}[\nabla_{\beta_1}\mathcal{K}] = \sum_{i=1}^{n}\widehat{\nabla_{\beta_1}\mathcal{K}}_i \qquad \text{(B-28)}$$

$$\hat{E}[\nabla_{\beta_2}\mathcal{K}] = \sum_{i=1}^{n}\widehat{\nabla_{\beta_2}\mathcal{K}}_i \qquad \text{(B-29)}$$

$$\hat{E}[\nabla_{\beta_1}\mathcal{V}] = \sum_{i=1}^{n}\widehat{\nabla_{\beta_1}\mathcal{V}}_i \qquad \text{(B-30)}$$

and

$$\hat{E}[\nabla_{\beta_2}\mathcal{V}] = \sum_{i=1}^{n}\widehat{\nabla_{\beta_2}\mathcal{V}}_i \qquad \text{(B-31)}$$

where

$$\widehat{\nabla_{\beta_1}\mathcal{K}}_i = \lambda(W_i\hat{\beta}_1)\exp(W_i\hat{\beta}_2)\Big[\{[1-2\Lambda(W_i\hat{\beta}_1)]\hat{\beta}_{P1}+\hat{\beta}_{P2}\}W_i+[1 \quad 0 \ldots 0]\Big] \quad \text{(B-32)}$$

$$\widehat{\nabla_{\beta_2}\mathcal{K}}_i = \exp(W_i\hat{\beta}_2)\Big[\{\lambda(W_i\hat{\beta}_1)\hat{\beta}_{P1}+\Lambda(W_i\hat{\beta}_1)\hat{\beta}_{P2}\}W_i+\Lambda(W_i\hat{\beta}_1)[1 \quad 0 \ldots 0]\Big] \quad \text{(B-33)}$$

$$\widehat{\nabla_{\beta_1}\mathcal{V}}_i = \lambda(W_i\hat{\beta}_1)\exp(W_i\hat{\beta}_2)W_i \quad \text{(B-34)}$$

and

$$\widehat{\nabla_{\beta_1}\mathcal{V}}_i = \Lambda(W_i\hat{\beta}_1)\exp(W_i\hat{\beta}_2)W_i\,. \quad \text{(B-35)}$$

so

$$\hat{E}[\nabla_{\gamma\delta}q] = 2\begin{bmatrix}\hat{E}[\nabla_{\beta_1}\mathcal{K}] & \hat{E}[\nabla_{\beta_2}\mathcal{K}]\\ \hat{E}[\nabla_{\beta_1}\mathcal{V}] & \hat{E}[\nabla_{\beta_2}\mathcal{V}]\end{bmatrix}. \quad \text{(B-37)}$$

$\mathbf{E[\nabla_{\gamma\gamma}q]^{-1}}$

2×2

$$E\Big[\nabla_{\gamma\gamma}q\Big]^{-1} = \begin{bmatrix}E\big[\nabla_{\kappa\kappa}q_a\big]^{-1} & 0\\ 0 & E\big[\nabla_{\nu\nu}q_b\big]^{-1}\end{bmatrix} \quad \text{(B-38)}$$

because $\nabla_{\kappa\nu}q_a = \nabla_{\nu\kappa}q_b = 0$. Now

$$\nabla_{\kappa\kappa}q_a = \nabla_{\nu\nu}q_b = -2$$

therefore

$$E\Big[\nabla_{\gamma\gamma}q\Big]^{-1} = \begin{bmatrix}-\frac{1}{2} & 0\\ 0 & -\frac{1}{2}\end{bmatrix}. \quad \text{(B-39)}$$

$\mathbf{E\Big[\nabla_{\gamma}q'\nabla_{\gamma}q\Big]}$

2×2

Given that

$$\nabla_{\gamma} q = 2\left[\left(\mathcal{K}(\beta, W) - \kappa\right)\ \left(\mathcal{V}(\beta, W) - \nu\right)\right]. \tag{B-40}$$

we have

$$\mathrm{E}\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right] = 4\begin{bmatrix} \mathrm{E}[(\mathcal{K}(\beta, S) - \kappa)^2] & \mathrm{E}[(\mathcal{K}(\beta, S) - \kappa)(\mathcal{V}(\beta, S) - \nu)] \\ \mathrm{E}[(\mathcal{K}(\beta, S) - \kappa)(\mathcal{V}(\beta, S) - \nu)] & \mathrm{E}[(\mathcal{V}(\beta, S) - \nu)^2] \end{bmatrix}. \tag{B-41}$$

The corresponding consistent estimator is

$$\hat{\mathrm{E}}\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right] = 4\begin{bmatrix} \sum_{i=1}^{n} \frac{1}{n}(\hat{\mathcal{K}}_i - \hat{\kappa})^2 & \sum_{i=1}^{n} \frac{1}{n}(\hat{\mathcal{K}}_i - \hat{\kappa})(\hat{\mathcal{V}}_i - \hat{\nu}) \\ \sum_{i=1}^{n} \frac{1}{n}(\hat{\mathcal{K}}_i - \hat{\kappa})(\hat{\mathcal{V}}_i - \hat{\nu}) & \sum_{i=1}^{n} \frac{1}{n}(\hat{\mathcal{V}}_i - \hat{\nu})^2 \end{bmatrix}. \tag{B-42}$$

Based on the above results, we can consistently estimate (B-5) as

$$\widehat{\mathrm{AVAR}(\hat{\gamma})} = \hat{\mathrm{E}}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{\mathrm{E}}\left[\nabla_{\gamma\delta} q\right] \widehat{\mathrm{AVAR}(\hat{\beta})}\ \hat{\mathrm{E}}\left[\nabla_{\gamma\delta} q\right]' \hat{\mathrm{E}}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$
$$+\ \hat{\mathrm{E}}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{\mathrm{E}}\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right] \hat{\mathrm{E}}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

and using well known results from asymptotic theory for two-stage estimators, we can show that[38]

[38]Discussions of asymptotic theory for two-stage optimization estimators can be found in Newey & McFadden (1994), White (1994, Chapter 6), and Wooldridge (2010, Chapter 12).

$$\widehat{\text{AVAR}(\hat{\gamma})}^{-\frac{1}{2}}\sqrt{n}(\hat{\gamma}-\gamma) \xrightarrow{d} N(0, I)\,. \tag{B-43}$$

Combining (B-43) with (B-1) we also have that

$$\widehat{\text{avar}(\hat{\eta}^{PO})}^{-\frac{1}{2}}\sqrt{n}(\hat{\eta}^{PO} - \eta^{PO}) \xrightarrow{d} N(0, I) \tag{B-44}$$

where $\widehat{\text{avar}(\hat{\eta}^{PO})}$ is given in (B-1).

Appendix C.

Bias from Using $\hat{\eta}^{MBM}$ Instead of $\hat{\eta}^{PO}$ in a Two-Part Model

If we define

$$\nu \equiv E[\Lambda(W\beta_1)\exp(W\beta_2)]$$

$$\zeta \equiv E[\Lambda(W\beta_1)]$$

and

$$\omega \equiv E[\Lambda(W\beta_1)^2\exp(W\beta_2)]$$

we can write the bias from using the MBM approach vs. the correct PO method as the following rendition of the difference between (2-5) and (2-7)

$$\eta^{MBM} - \eta^{PO} = [(1-\zeta)\beta_{P1} + \beta_{P2}] - \left(\frac{\beta_{P1}E[\lambda(W\beta_1)\exp(W\beta_2)] + \beta_{P2}\nu}{\nu}\right)$$

$$= [(1-\zeta)\beta_{P1} + \beta_{P2}] - \left(\frac{\beta_{P1}E[\Lambda(W\beta_1)[1-\Lambda(W\beta_1)]\exp(W\beta_2) + \beta_{P2}\nu}{\nu}\right)$$

$$= [(1-\zeta)\beta_{P1} + \beta_{P2}] - \left(\frac{\beta_{P1}\nu - \beta_{P1}E[\Lambda(W\beta_1)^2\exp(W\beta_2)] + \beta_{P2}\nu}{\hat{\nu}}\right)$$

$$= \beta_{P1} - \zeta\beta_{P1} + \beta_{P2} - \beta_{P1} + \beta_{P1}\frac{\omega}{\nu} - \beta_{P2}$$

$$= \beta_{P1}\frac{\omega}{\nu} - \hat{\beta}_{P1}\zeta$$

$$= \beta_{P1}\left(\frac{\omega}{\nu} - \zeta\right).$$

Appendix D.

Asymptotic Distribution (and Standard Error) of $\hat{\eta}^{MBM} - \hat{\eta}^{PO}$

In Appendix C we showed that

$$\eta^{MBM} - \eta^{PO} = \beta_{P1}\left(\frac{\omega}{\nu} - \zeta\right)$$

where

$$\nu \equiv E[\Lambda(W\beta_1)\exp(W\beta_2)]$$

$$\zeta \equiv E[\Lambda(W\beta_1)]$$

and

$$\omega \equiv E[\Lambda(W\beta_1)^2\exp(W\beta_2)].$$

Using the corresponding consistent estimators for ω, ν and ζ we can write

$$\hat{\eta}^{MBM} - \hat{\eta}^{PO} = \hat{\beta}_{P1}\left(\frac{\hat{\omega}}{\hat{\nu}} - \hat{\zeta}\right)$$

where

$$\hat{\omega} = \sum_{i=1}^{n}\frac{1}{n}\Omega(\hat{\beta}, W_i)$$

$$\hat{\nu} = \sum_{i=1}^{n}\frac{1}{n}\mathcal{V}(\hat{\beta}, W_i)$$

$$\hat{\zeta} = \sum_{i=1}^{n}\frac{1}{n}\mathcal{Z}(\hat{\beta}_1, W_i)$$

$$\Omega(\beta, W) = \Lambda(W\beta_1)^2\exp(W\beta_2)$$

$\mathcal{V}(\beta, W) = \Lambda(W\beta_1)\exp(W\beta_2)$

$\mathcal{Z}(\beta, W) = \Lambda(W\beta_1)$

$\hat{\beta} = [\hat{\beta}_1' \quad \hat{\beta}_2']'$ (with $\hat{\beta}_1' = [\hat{\beta}_{P1} \quad \hat{\beta}_{X1}']$ and $\hat{\beta}_2' = [\hat{\beta}_{P2} \quad \hat{\beta}_{X2}']$) is the consistent estimate of the parameter vector $\beta = [\beta_1' \quad \beta_2']'$ (with $\beta_1' = [\beta_{P1} \quad \beta_{X1}']$ and $\beta_2' = [\beta_{P2} \quad \beta_{X2}']$)) [the parameters of equations (2-1) and (2-2)] obtained via the two-part protocol culminating in (2-3)

and

$W_i = [P_i \quad X_i]$ denotes the observation on $W = [P \quad X]$ for the ith individual in the sample (i = 1, ..., n).

Let $\hat{\tau} = [\hat{\beta}_{P1} \quad \hat{\omega} \quad \hat{\nu} \quad \hat{\zeta}]'$ and $\tau = [\beta_{P1} \quad \omega \quad \nu \quad \zeta]'$, where $\text{plim}[\hat{\tau}] = \tau$. If we could show that

$$\text{AVAR}(\hat{\tau})^{-\frac{1}{2}}\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{d} N(0, I)$$

where the formulation of $\text{AVAR}(\hat{\tau})$ is known, then we could apply the δ-method to obtain the asymptotic variance of $\hat{\eta}^{MBM} - \hat{\eta}^{PO}$ as

$$\text{avar}(\hat{\eta}^{MBM} - \hat{\eta}^{PO}) = c(\tau)\ \text{AVAR}(\hat{\tau})\ c(\tau)'$$

where $c(\tau) = \left[\left(\frac{\omega}{\nu} - \zeta\right) \;\; 1/\nu \;\; -\omega/\nu^2 \;\; -\beta_{P1}\right]$. Moreover, if we have a consistent estimator for $AVAR(\hat{\tau})$, say $\widehat{AVAR(\hat{\tau})}$ [i.e. $p\lim\left[\widehat{AVAR(\hat{\tau})}\right] = AVAR(\hat{\tau})$], then we could consistently estimate $avar(\hat{\eta}^{MBM} - \hat{\eta}^{PO})$ as

$$\widehat{avar(\hat{\eta}^{MBM} - \hat{\eta}^{PO})} = c(\hat{\tau})\; \widehat{AVAR(\hat{\tau})}\; c(\hat{\tau})'. \qquad \text{(D-1)}$$

We focus, therefore, on finding the asymptotic distribution of $\hat{\tau}$ and, in particular, the formulation of its asymptotic covariance matrix.

First note that we can write τ as

$$\tau = \Xi\, \theta \qquad \text{(D-2)}$$

where $\theta' = [\delta' \;\; \gamma']$, $\delta = [\beta_1' \;\; \beta_2']'$, $\gamma' = [\omega \;\; \nu \;\; \zeta]$ (recall, $\beta_1' = [\beta_{P1} \;\; \beta_{X1}']$ and $\beta_2' = [\beta_{P2} \;\; \beta_{X2}']$)

$$\Xi = \begin{bmatrix} \ell_{\beta_{P1}} \\ \hline 0_{3,2K} \quad I_3 \end{bmatrix}$$

ℓ_a is the unit row vector with the value "1" in the element position corresponding to the element position of a in the vector θ, $0_{b,c}$ is the matrix of zeros whose row and column dimensions are b and c, respectively, I_d is the identity matrix of order d, and K is the

column dimension of W. For future reference, let's set the following vector/matrix dimensions:

β_1 is K×1

β_2 is K×1

W is 1×K

τ is 4×1

$c(\tau)$ is 1×4

δ is 2K×1

γ is 3×1

θ is (2K+3)×1

Ξ is 4×(2K+3)

$\ell_{\beta_{P1}}$ is 1×(2K+3)

Clearly then

$$\text{AVAR}(\hat{\tau}) = \Xi \, \text{AVAR}(\hat{\theta}) \, \Xi' \tag{D-3}$$

where $\hat{\theta}$ is the estimator of θ obtained from the following two-stage protocol.

First Stage

Consistently estimate β via the following optimization estimator

$$\hat{\delta} = \arg\max_{\tilde{\delta}} \frac{\sum_{i=1}^{n} q_1(\tilde{\delta}, S_i)}{n} \tag{D-4}$$

where

$$q_1(\tilde{\delta}, S_i) = q_{11}(\tilde{\beta}_1, S_i) + q_{12}(\tilde{\beta}_2, S_i)$$

$$q_{11}(\tilde{\beta}_1, S_i) = I(A_i > 0)\ln[\Lambda(W_i\tilde{\beta}_1)] + [1 - I(A_i > 0)]\ln[1 - \Lambda(W_i\tilde{\beta}_1)]$$

$$q_{12}(\tilde{\beta}_2, S_i) = -I(A_i > 0)(\ln(A_i) - W_i\tilde{\beta}_2)^2$$

$$S_i = [A_i \quad X_i \quad P_i]$$

$\tilde{\delta} = [\tilde{\beta}_1' \quad \tilde{\beta}_2']'$, $\tilde{\beta}_1' = [\tilde{\beta}_{P1} \quad \tilde{\beta}_{X1}']$ and $\tilde{\beta}_2' = [\tilde{\beta}_{P2} \quad \tilde{\beta}_{X2}']$ and $\hat{\delta} = [\hat{\beta}_1' \quad \hat{\beta}_2']'$

Second Stage

Consistently estimate γ via the following optimization estimator

$$\hat{\gamma} = \arg\max_{\tilde{\gamma}} \frac{\sum_{i=1}^{n} q(\hat{\delta}, \tilde{\gamma}, S_i)}{n} \tag{D-5}$$

where

$$q(\hat{\delta}, \tilde{\gamma}, S_i) = q_a(\hat{\delta}, \tilde{\omega}, S_i) + q_b(\hat{\delta}, \tilde{\nu}, S_i) + q_c(\hat{\delta}, \tilde{\zeta}, S_i)$$

$$q_a(\hat{\delta}, \tilde{\omega}, S_i) = -(\Omega(\hat{\beta}, W_i) - \tilde{\omega})^2$$

$$q_b(\hat{\delta}, \tilde{\nu}, S_i) = -(\mathcal{V}(\hat{\beta}, W_i) - \tilde{\nu})^2$$

$$q_c(\hat{\delta}, \tilde{\zeta}, S_i) = -(\mathcal{Z}(\hat{\beta}, W_i) - \tilde{\zeta})^2$$

$\hat{\delta} = [\hat{\beta}_1' \quad \hat{\beta}_2']'$ is the first stage estimator of β, $\hat{\beta}_1' = [\hat{\beta}_{P1} \quad \hat{\beta}_{X1}']$ and $\hat{\beta}_2' = [\hat{\beta}_{P2} \quad \hat{\beta}_{X2}']$. Use q_1 as shorthand notation for

$$q_1(\delta, S) = q_{11}(\beta_1, S) + q_{12}(\beta_2, S)$$

with

$$q_{11}(\beta_1, S) = I(A > 0)\ln[\Lambda(W\beta_1)] + [1 - I(A > 0)]\ln[1 - \Lambda(W\beta_1)]$$

$$q_{12}(\beta_2, S) = -I(A > 0)(\ln(A) - W\beta_2)^2$$

$$S = [A \quad X \quad P]$$

and use q as shorthand notation for

$$q(\delta, \tilde{\gamma}, S) = q_a(\delta, \omega, S) + q_b(\delta, \nu, S) + q_c(\delta, \zeta, S)$$

with

$$q_a(\delta, \omega, S) = -(\Omega(\beta, W) - \omega)^2$$

$$q_b(\delta, \nu, S) = -(\mathcal{V}(\beta, W) - \nu)^2$$

$$q_c(\delta, \zeta, S) = -(\mathcal{Z}(\beta_2, W) - \zeta)^2$$

and let $AVAR(\hat{\delta})$ denote the asymptotic covariance matrix of the first stage estimator.[39] Using well known results from asymptotic theory for two-stage estimators, we can show that[40]

$$AVAR(\hat{\theta})^{-\frac{1}{2}}\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I) \qquad \text{(D-6)}$$

where $\hat{\theta}' = [\hat{\delta}' \quad \hat{\gamma}']$, $plim(\hat{\theta}) = \theta$

---

[39] Note that a consistent estimate $AVAR(\hat{\delta})$ can be obtained from the packaged output for the first stage estimator because the first stage estimator is unaffected by the fact that it is a component of a two-stage estimator.

[40] Discussions of asymptotic theory for two-stage optimization estimators can be found in Newey & McFadden (1994), White (1994, Chapter 6), and Wooldridge (2010, Chapter 12).

$$\mathrm{AVAR}(\hat{\theta}) = \begin{bmatrix} D_{11} & D_{12} \\ D_{12}' & D_{22} \end{bmatrix} \tag{D-7}$$

$$D_{11} = \underset{2K\times 2K}{\mathrm{AVAR}(\hat{\delta})} \tag{D-8}$$

$$\begin{aligned} D_{12} = {} & E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\delta} q_1 ' \nabla_{\gamma} q\right] E\left[\nabla_{\gamma\gamma} q\right]^{-1} \\ & - E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\delta} q_1 ' \nabla_{\delta} q_1\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma\delta} q\right]' E\left[\nabla_{\gamma\gamma} q\right]^{-1} \end{aligned} \tag{D-9}$$

$$\begin{aligned} D_{22} = \mathrm{AVAR}(\hat{\gamma}) = {} & E\left[\nabla_{\gamma\gamma} q\right]^{-1} \Big[E\left[\nabla_{\gamma\delta} q\right] \mathrm{AVAR}(\hat{\beta}) E\left[\nabla_{\gamma\delta} q\right]' \\ & - E\left[\nabla_{\gamma} q ' \nabla_{\delta} q_1\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma\delta} q\right]' \\ & - E\left[\nabla_{\gamma\delta} q\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma} q ' \nabla_{\delta} q_1\right]' \Big] E\left[\nabla_{\gamma\gamma} q\right]^{-1} \\ & + E\left[\nabla_{\gamma\gamma} q\right]^{-1} E\left[\nabla_{\gamma} q ' \nabla_{\gamma} q\right] E\left[\nabla_{\gamma\gamma} q\right]^{-1}. \end{aligned} \tag{D-10}$$

Fortunately, (D-9) and (D-10) can be simplified in a number of ways. First note that we can write

$$E\left[\nabla_{\gamma} q ' \nabla_{\delta} q_1\right] = E\left[\nabla_{\gamma} q ' E[\nabla_{\delta} q_1 \mid W]\right]$$

but

$$\nabla_{\delta} q_1 = \nabla_{\delta} q_{11} + \nabla_{\delta} q_{12}$$

with

$$\nabla_{\delta} q_{11} = \left[\nabla_{\beta_1} \ln f\left(I(A > 0) \mid W\right) \quad 0 \quad 0\right]$$

$$\nabla_{\delta} q_{12} = \left[0 \qquad 2\, I(A > 0)(\ln(A) - W\beta_2) W \quad 0\right]$$

where

$$f(I(A>0) \mid W) = \Lambda(W\beta_1)^{I(A>0)}[1-\Lambda(W\beta_1)]^{[1-I(A>0)]} .$$

Therefore

$$E[\nabla_\delta q_1 \mid W] = \Big[E[\nabla_{\beta_1} \ln f(I(A>0) \mid W)] \qquad 2E\big[I(A>0)(\ln(A) - W\beta_2) \mid W\big]W\Big]$$

$$= [\,0 \quad 0]$$

because $E[\nabla_{\beta_1} \ln f(I(A>0) \mid W)] = 0$ [see (13.20) on p. 477 of Wooldridge (2010)] and, $E\big[I(A>0)(\ln(A) - W\beta_2) \mid W\big] = 0$ by design. Finally, then we get

$$E\Big[\nabla_\gamma q' \nabla_\delta q_1\Big] = 0$$

so

$$D_{12} = \underset{2K\times 2K}{-E\big[\nabla_{\delta\delta} q_1\big]^{-1}} \underset{2K\times 2K}{E\big[\nabla_\delta q_1' \nabla_\delta q_1\big]} \underset{2K\times 2K}{E\big[\nabla_{\delta\delta} q_1\big]^{-1}} \underset{2K\times 3}{E\Big[\nabla_{\gamma\delta} q\Big]'} \underset{3\times 3}{E\Big[\nabla_{\gamma\gamma} q\Big]^{-1}} \qquad \text{(D-11)}$$

and

$$D_{22} = AVAR(\hat{\gamma}) = \underset{3\times 3}{E\Big[\nabla_{\gamma\gamma} q\Big]^{-1}} \underset{3\times 2K}{E\Big[\nabla_{\gamma\delta} q\Big]} \underset{2K\times 2K}{AVAR(\hat{\beta})} \underset{2K\times 3}{E\Big[\nabla_{\gamma\delta} q\Big]'} \underset{3\times 3}{E\Big[\nabla_{\gamma\gamma} q\Big]^{-1}}$$

$$+ \underset{3\times 3}{E\Big[\nabla_{\gamma\gamma} q\Big]^{-1}} \underset{3\times 3}{E\Big[\nabla_\gamma q' \nabla_\gamma q\Big]} \underset{3\times 3}{E\Big[\nabla_{\gamma\gamma} q\Big]^{-1}} . \qquad \text{(D-12)}$$

so

$$\underset{(2K+3)\times(2K+3)}{AVAR(\hat{\theta})} = \begin{bmatrix} \underset{2K\times 2K}{D_{11}} & \underset{2K\times 3}{D_{12}} \\ \underset{3\times 2K}{D_{12}'} & \underset{3\times 3}{D_{22}} \end{bmatrix}.$$

Let's consider each of the individual components of (D-6) and (D-7) in turn.

$\mathbf{E[\nabla_{\delta\delta} q_1]^{-1}}$
2K×2K

Written out explicitly we have

$$E[\nabla_{\beta\beta} q_1]^{-1} = \begin{bmatrix} E[\nabla_{\beta_1\beta_1} q_{11}]^{-1} & 0 \\ 0 & E[\nabla_{\beta_2\beta_2} q_{12}]^{-1} \end{bmatrix}. \quad \text{(D-13)}$$

Now

$$E[\nabla_{\beta_1\beta_1} q_{11}]^{-1} = -\,AVAR(\hat{\beta}_1) \quad \text{(D-14)}$$

= the negative of the asymptotic covariance matrix for first stage, first part, logit estimation in the two-stage estimation protocol for θ. We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{\beta_1\beta_1} q_{11}]^{-1}$ is

$$\hat{E}[\nabla_{\beta_1\beta_1} q_{11}]^{-1} = -\,n\widehat{AVAR}^*(\hat{\beta}_1) \quad \text{(D-15)}$$

where $\widehat{AVAR}^*(\hat{\beta}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure. Also

$$\nabla_{\beta_2} q_{12} = 2\,I(A > 0)(\ln(A) - W\beta_2)W$$

and

$$\nabla_{\beta_2\beta_2} q_{12} = -2\,I(A > 0)W'W\,.$$

Therefore

$$E[\nabla_{\beta_2\beta_2} q_{12}] = -2E\left[I(A>0)W'W\right]. \tag{D-16}$$

A consistent estimator of $E[\nabla_{\beta_2\beta_2} q_{12}]^{-1}$ is

$$\hat{E}[\nabla_{\beta_2\beta_2} q_{12}]^{-1} = \left[-2\frac{1}{n_1}\sum_{i=1}^{n}\left\{I(A_i>0)W_i'W_i\right\}\right]^{-1}$$

$$= n_1\left[-2\sum_{i=1}^{n}\left\{I(A_i>0)W_i'W_i\right\}\right]^{-1}. \tag{D-17}$$

where $\hat{\beta}_2$ is the first stage, second part, estimator of β_2 and n_1 is the size of the subsample for whom $I(A>0)=1$, so

$$\hat{E}[\nabla_{\delta\delta} q_1]^{-1} = \hat{E}[\nabla_{\beta\beta} q_1]^{-1} = \begin{bmatrix} \hat{E}[\nabla_{\beta_1\beta_1} q_{11}]^{-1} & 0 \\ 0 & \hat{E}[\nabla_{\beta_2\beta_2} q_{12}]^{-1} \end{bmatrix}. \tag{D-18}$$

$\mathbf{E\left[\nabla_\delta q_1 '\nabla_\delta q_1\right]}$
2K×2K

Written out explicitly we have

$$E\left[\nabla_\beta q_1 '\nabla_\beta q_1\right] = \begin{bmatrix} E[\nabla_{\beta_1} q_{11} '\nabla_{\beta_1} q_{11}] & E[\nabla_{\beta_1} q_{11} '\nabla_{\beta_2} q_{12}] \\ E[\nabla_{\beta_2} q_{12} '\nabla_{\beta_1} q_{11}] & E[\nabla_{\beta_2} q_{12} '\nabla_{\beta_2} q_{12}] \end{bmatrix}. \tag{D-19}$$

Because the first stage, first part, estimator of β_1 is MLE we can write

$$E[\nabla_{\beta_1} q_{11} '\nabla_{\beta_1} q_{11}] = -E[\nabla_{\beta_1\beta_1} q_{11}] = \left[AVAR(\hat{\beta}_1)\right]^{-1}$$

= the inverse of the asymptotic covariance matrix for first stage,

first part, logit estimation in the two-stage estimation protocol for θ. We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{\beta_1} q_{11}' \nabla_{\beta_1} q_{11}]$ is

$$\hat{E}[\nabla_{\beta_1} q_{11}' \nabla_{\beta_1} q_{11}] = \frac{1}{n}\left[\widehat{AVAR} * (\hat{\beta}_1)\right]^{-1} \tag{D-20}$$

where $\widehat{AVAR} * (\hat{\beta}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure. The remainder of the block elements follow from

$$\nabla_{\beta_1} q_{11} = \nabla_{\beta_1} \ln f(I(A > 0) \mid W) = \left[I(A > 0)[1 - \Lambda(W\beta_1)] - [1 - I(A > 0)]\Lambda(W\beta_1)\right]W \tag{D-21}$$

$$\nabla_{\beta_2} q_{12} = 2\,I(A > 0)(\ln(A) - W\beta_2)W\,. \tag{D-22}$$

where the formulation of $\nabla_{\beta_1} q_{11}$ comes from equation (16.4.8) on p. 350 of Fomby et al. (1984) and $\Lambda(\ \ )$ denotes the logistic cdf. The remaining required consistent matrix estimators are

$$\hat{E}[\nabla_{\beta_1} q_{11}' \nabla_{\beta_2} q_{12}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{\beta_1} \hat{q}_{11i}' \nabla_{\beta_2} \hat{q}_{12i} \tag{D-23}$$

$$\hat{E}[\nabla_{\beta_2} q_{12}' \nabla_{\beta_2} q_{12}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{\beta_2} \hat{q}_{12i}' \nabla_{\beta_2} \hat{q}_{12i} \tag{D-24}$$

where

$$\nabla_{\beta_1} \hat{q}_{11i} = \left[I(A_i > 0)[1 - \Lambda(W_i\hat{\beta}_1)] - [1 - I(A_i > 0)]\Lambda(W_i\hat{\beta}_1)\right]W_i \tag{D-25}$$

and

$$\nabla_{\beta_2}\hat{q}_{12i} = 2\,I(A_i > 0)(\ln(A_i) - W_i\hat{\beta}_2)\,W_i \tag{D-26}$$

so

$$\hat{E}\left[\nabla_\delta q_1{}'\nabla_\delta q_1\right] = \hat{E}\left[\nabla_\beta q_1{}'\nabla_\beta q_1\right] = \begin{bmatrix} \hat{E}[\nabla_{\beta_1} q_{11}{}'\nabla_{\beta_1} q_{11}] & \hat{E}[\nabla_{\beta_1} q_{11}{}'\nabla_{\beta_2} q_{12}] \\ \hat{E}[\nabla_{\beta_2} q_{12}{}'\nabla_{\beta_1} q_{11}] & \hat{E}[\nabla_{\beta_2} q_{12}{}'\nabla_{\beta_2} q_{12}] \end{bmatrix}. \tag{D-27}$$

$\underset{3\times 2K}{\mathbf{E}\left[\nabla_{\gamma\delta}\mathbf{q}\right]}$

Written out explicitly we have

$$\begin{aligned} \nabla_\gamma q &= [(\nabla_\omega q_a + 0 + 0) \quad (0 + \nabla_\nu q_b + 0) \quad (0 + 0 + \nabla_\zeta q_c)] \\ &= [\nabla_\omega q_a \quad \nabla_\nu q_b \quad \nabla_\zeta q_c] \\ &= 2\left[\left(\Omega(\beta, W) - \omega\right) \; \left(\mathcal{V}(\beta, W) - \nu\right) \; \left(\mathcal{Z}(\beta, W) - \zeta\right)\right] \end{aligned} \tag{D-28}$$

and

$$\begin{aligned} E[\nabla_{\gamma\delta} q] &= \begin{bmatrix} E[\nabla_{\omega\beta_1} q_a] & E[\nabla_{\omega\beta_2} q_a] \\ E[\nabla_{\nu\beta_1} q_b] & E[\nabla_{\nu\beta_2} q_b] \\ E[\nabla_{\zeta\beta_1} q_c] & 0 \end{bmatrix} \\ &= 2\begin{bmatrix} E[\nabla_{\beta_1}\Omega] & E[\nabla_{\beta_2}\Omega] \\ E[\nabla_{\beta_1}\mathcal{V}] & E[\nabla_{\beta_2}\mathcal{V}] \\ E[\nabla_{\beta_1}\mathcal{Z}] & 0 \end{bmatrix} \end{aligned} \tag{D-29}$$

where

$$\nabla_{\beta_1}\Omega = 2\,\Lambda(W\beta_1)\lambda(W\beta_1)\exp(W\beta_2)\,W \tag{D-30}$$

$$\nabla_{\beta_2}\Omega = \Lambda(W\beta_1)^2 \exp(W\beta_2)W \tag{D-31}$$

$$\nabla_{\beta_1}\mathcal{V} = \lambda(W\beta_1)\exp(W\beta_2)W \tag{D-32}$$

$$\nabla_{\beta_2}\mathcal{V} = \Lambda(W\beta_1)\exp(W\beta_2)W \tag{D-33}$$

and

$$\nabla_{\beta_1}\mathcal{Z} = \lambda(W\beta_1)W\,. \tag{D-34}$$

Note that

$$\nabla_a \Lambda(a) = \lambda(a) = \Lambda(a)[1 - \Lambda(a)]$$

$$\nabla_a \lambda(a) = \lambda(a)[1 - 2\Lambda(a)]\,.$$

The requisite consistent matrix estimators are

$$\hat{E}[\nabla_{\beta_1}\Omega] = \sum_{i=1}^{n} \widehat{\nabla_{\beta_1}\Omega}_i \tag{D-35}$$

$$\hat{E}[\nabla_{\beta_2}\Omega] = \sum_{i=1}^{n} \widehat{\nabla_{\beta_2}\Omega}_i \tag{D-36}$$

$$\hat{E}[\nabla_{\beta_1}\mathcal{V}] = \sum_{i=1}^{n} \widehat{\nabla_{\beta_1}\mathcal{V}}_i \tag{D-37}$$

$$\hat{E}[\nabla_{\beta_2}\mathcal{V}] = \sum_{i=1}^{n} \widehat{\nabla_{\beta_2}\mathcal{V}}_i \tag{D-38}$$

and

$$\hat{E}[\nabla_{\beta_1}\mathcal{Z}] = \sum_{i=1}^{n} \widehat{\nabla_{\beta_1}\mathcal{Z}}_i \tag{D-39}$$

where

$$\widehat{\nabla_{\beta_1}\Omega}_i = 2\Lambda(W_i\hat{\beta}_1)\lambda(W_i\hat{\beta}_1)\exp(W_i\hat{\beta}_2)W_i \tag{D-40}$$

$$\widehat{\nabla_{\beta_2}\Omega}_i = \Lambda(W_i\hat{\beta}_1)^2 \exp(W_i\hat{\beta}_2)\, W \tag{D-41}$$

$$\widehat{\nabla_{\beta_1}\mathcal{V}}_i = \lambda(W_i\hat{\beta}_1)\exp(W_i\hat{\beta}_2)\, W_i \tag{D-42}$$

$$\widehat{\nabla_{\beta_1}\mathcal{V}}_i = \Lambda(W_i\hat{\beta}_1)\exp(W_i\hat{\beta}_2)\, W_i \tag{D-43}$$

and

$$\widehat{\nabla_{\beta_1}\mathcal{Z}}_i = \lambda(W_i\hat{\beta}_1)\, W_i \tag{D-44}$$

so

$$\hat{E}[\nabla_{\gamma\delta}q] = 2\begin{bmatrix} \hat{E}[\nabla_{\beta_1}\Omega] & \hat{E}[\nabla_{\beta_2}\Omega] \\ \hat{E}[\nabla_{\beta_1}\mathcal{V}] & \hat{E}[\nabla_{\beta_2}\mathcal{V}] \\ \hat{E}[\nabla_{\beta_1}\mathcal{Z}] & 0 \end{bmatrix}. \tag{D-45}$$

$\mathbf{E[\nabla_{\gamma\gamma}q]^{-1}}$

3×3

$$E\left[\nabla_{\gamma\gamma}q\right]^{-1} = \begin{bmatrix} E\left[\nabla_{\omega\omega}q_a\right]^{-1} & 0 & 0 \\ 0 & E\left[\nabla_{\nu\nu}q_b\right]^{-1} & 0 \\ 0 & 0 & E\left[\nabla_{\zeta\zeta}q_c\right]^{-1} \end{bmatrix} \tag{D-46}$$

because $\nabla_{\omega\nu}q_a = \nabla_{\omega\zeta}q_a = \nabla_{\nu\omega}q_b = \nabla_{\nu\zeta}q_b = \nabla_{\zeta\omega}q_c = \nabla_{\zeta\nu}q_c = 0$. Now

$$\nabla_{\omega\omega}q_a = \nabla_{\nu\nu}q_b = \nabla_{\zeta\zeta}q_c = -2$$

therefore

$$E\left[\nabla_{\gamma\gamma}q\right]^{-1} = \begin{bmatrix} -\frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & -\frac{1}{2} \end{bmatrix}. \tag{D-47}$$

$$\underset{3\times 3}{\mathbf{E}\left[\nabla_\gamma \mathbf{q}'\nabla_\gamma \mathbf{q}\right]}$$

Given that

$$\nabla_\gamma q = 2\left[\left(\Omega(\beta, W)-\omega\right)\ \left(\mathcal{V}(\beta, W)-\nu\right)\ \left(\mathcal{Z}(\beta, W)-\zeta\right)\right]. \tag{D-48}$$

we have

$$\mathrm{E}\left[\nabla_\gamma q'\nabla_\gamma q\right] = 4\begin{bmatrix} \mathrm{E}[(\Omega(\beta, W)-\omega)^2] & \mathrm{E}[(\Omega(\beta, W)-\omega)(\mathcal{V}(\beta, S)-\nu)] & \mathrm{E}[(\Omega(\beta, W)-\omega)(\mathcal{Z}(\beta, S)-\zeta)] \\ \mathrm{E}[(\Omega(\beta, W)-\omega)(\mathcal{V}(\beta, S)-\nu)] & \mathrm{E}[(\mathcal{V}(\beta, S)-\nu)^2] & \mathrm{E}[(\mathcal{V}(\beta, S)-\nu)(\mathcal{Z}(\beta, S)-\zeta)] \\ \mathrm{E}[(\Omega(\beta, W)-\omega)(\mathcal{Z}(\beta, S)-\zeta)] & \mathrm{E}[(\mathcal{V}(\beta, S)-\nu)(\mathcal{Z}(\beta, S)-\zeta)] & \mathrm{E}[(\mathcal{Z}(\beta, S)-\zeta)^2] \end{bmatrix}. \tag{D-49}$$

The corresponding consistent estimator is

$$\hat{\mathrm{E}}\left[\nabla_\gamma q'\nabla_\gamma q\right] = 4\begin{bmatrix} \sum_{i=1}^{n}\frac{1}{n}(\hat{\Omega}_i-\hat{\omega})^2 & \sum_{i=1}^{n}\frac{1}{n}(\hat{\Omega}_i-\hat{\omega})(\hat{\mathcal{V}}_i-\hat{\nu}) & \sum_{i=1}^{n}\frac{1}{n}(\hat{\Omega}_i-\hat{\omega})(\hat{\mathcal{Z}}_i-\hat{\zeta}) \\ \sum_{i=1}^{n}\frac{1}{n}(\hat{\Omega}_i-\hat{\omega})(\hat{\mathcal{V}}_i-\hat{\nu}) & \sum_{i=1}^{n}\frac{1}{n}(\hat{\mathcal{V}}_i-\hat{\nu})^2 & \sum_{i=1}^{n}\frac{1}{n}(\hat{\mathcal{V}}_i-\hat{\nu})(\hat{\mathcal{Z}}_i-\hat{\zeta}) \\ \sum_{i=1}^{n}\frac{1}{n}(\hat{\Omega}_i-\hat{\omega})(\hat{\mathcal{Z}}_i-\hat{\zeta}) & \sum_{i=1}^{n}\frac{1}{n}(\hat{\mathcal{V}}_i-\hat{\nu})(\hat{\mathcal{Z}}_i-\hat{\zeta}) & \sum_{i=1}^{n}\frac{1}{n}(\hat{\mathcal{Z}}_i-\hat{\zeta})^2 \end{bmatrix}. \tag{D-50}$$

Based on (D-8), (D-11) and (D-12) and using the two-stage estimator $\hat{\theta}$ we can consistently estimate (D-7) as

$$\widehat{\mathrm{AVAR}(\hat{\theta})} = \begin{bmatrix} \hat{\mathrm{D}}_{11} & \hat{\mathrm{D}}_{12} \\ \hat{\mathrm{D}}_{12}' & \hat{\mathrm{D}}_{22} \end{bmatrix}$$

where

$$\hat{D}_{11} = \widehat{AVAR}(\hat{\delta}) = \widehat{AVAR}(\hat{\beta}) = \begin{bmatrix} n\widehat{AVAR}^*(\hat{\beta}_1) & 0 \\ 0 & n_1\widehat{AVAR}^*(\hat{\beta}_2) \end{bmatrix}$$

$$D_{12} = -\hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\delta} q_1 ' \nabla_{\delta} q_1\right] \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma\delta} q\right]' \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

$$D_{22} = \widehat{AVAR(\hat{\gamma})} = \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma\delta} q\right] \widehat{AVAR(\hat{\beta})}\, \hat{E}\left[\nabla_{\gamma\delta} q\right]' \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$
$$+ \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma} q ' \nabla_{\gamma} q\right] \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

and using well known results from asymptotic theory for two-stage estimators, we can show that[41]

$$\widehat{AVAR(\hat{\theta})}^{-\frac{1}{2}} \begin{bmatrix} \sqrt{n}(\hat{\beta}_1 - \beta_1) \\ \sqrt{n_1}(\hat{\beta}_2 - \hat{\beta}_2) \\ \sqrt{n}(\hat{\omega} - \omega) \\ \sqrt{n}(\hat{\nu} - \nu) \\ \sqrt{n}(\hat{\zeta} - \zeta) \end{bmatrix} \xrightarrow{d} N(0, I). \qquad \text{(D-51)}$$

****

ASIDE:

Notice that the "$\sqrt{n}$ blow up" is a bit tricky here. It implements $\sqrt{n}$ for $\hat{\beta}_1$, $\hat{\omega}$, $\hat{\nu}$ and $\hat{\zeta}$; but uses $\sqrt{n_1}$ for $\hat{\beta}_2$. We had to do this because we had to use the correct sample size

[41]Discussions of asymptotic theory for two-stage optimization estimators can be found in Newey & McFadden (1994), White (1994, Chapter 6), and Wooldridge (2010, Chapter 12).

(viz., n_1) for a number of the components of $\widehat{AVAR(\hat{\theta})}$ [viz., those that pertained to the estimation of $\hat{\beta}_2$]; in particular (D-17), (D-23) and (D-24). For this reason we had to be explicit about the denominators in all of the averages for the components of $\widehat{AVAR(\hat{\theta})}$. This meant that in the construction of the requisite asymptotic t-stats we had to explicitly include the "blow-up" in the numerator (i.e., we had to multiply by the square-root of the appropriate sample size). I refer to this as "tricky" because one typically does not have to do this. In the usual asymptotic t-stat construction the denominators of the averages ("n") need not be included in the construction of the asymptotic covariance matrix because it typically manifests as a multiplicative factor and, after pulling the diagonal and taking the square root to get the standard errors, this multiplicative $\sqrt{n}$ cancels with the "blow-up" factor in the numerator. For example, the asymptotic t-stat of the OLS estimator is

$$\frac{\sqrt{n}(\hat{\rho}_k - \rho_k)}{\sqrt{\widehat{AVAR(\hat{\rho})}}} = \frac{\sqrt{n}(\hat{\rho}_k - \rho_k)}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n}\mathcal{X}'\mathcal{X}_{kk}\right)^{-1}}} = \frac{\sqrt{n}(\hat{\rho}_k - \rho_k)}{\sqrt{n}\sqrt{\hat{\sigma}^2(\mathcal{X}'\mathcal{X}_{kk})^{-1}}} = \frac{(\hat{\rho}_k - \rho_k)}{\sqrt{\hat{\sigma}^2(\mathcal{X}'\mathcal{X}_{kk})^{-1}}}$$

where

n is the sample size

ρ_k is the coefficient of the kth regressor in the linear regression

$\hat{\rho}_k$ is its OLS estimator

σ^2 is the regression error variance estimator

$\mathcal{X}$ is the matrix of regressors

and $\mathcal{X}'\mathcal{X}_{kk}$ is the kth diagonal element of $\mathcal{X}'\mathcal{X}$. Note how the "$\sqrt{n}$ s" simply cancel. Note also that what we typically refer to as the "asymptotic standard error" can actually be written as the square root of the diagonal element of the consistent estimator of the asymptotic covariance matrix divided by n; in other words

$$\text{asy std err} = \sqrt{\frac{\widehat{\text{AVAR}(\hat{\rho})}}{n}}.$$

****

Now back to the issue at hand. Moreover

$$\widehat{\text{AVAR}(\hat{\tau})}^{-\frac{1}{2}}\sqrt{n}\begin{bmatrix} \sqrt{n}(\hat{\beta}_{P1} - \beta_{P1}) \\ \sqrt{n_1}(\hat{\beta}_{P2} - \hat{\beta}_{P2}) \\ \sqrt{n}(\hat{\omega} - \omega) \\ \sqrt{n}(\hat{\nu} - \nu) \\ \sqrt{n}(\hat{\zeta} - \zeta) \end{bmatrix} \xrightarrow{d} N(0, I). \tag{D-52}$$

where

$$\widehat{\text{AVAR}(\hat{\tau})} = \Xi\,\widehat{\text{AVAR}(\hat{\theta})}\,\Xi' \tag{D-53}$$

and τ and Ξ are defined as in (D-2). Now combining (D-1) with (D-52) and (D-53) we get

$$\widehat{\text{avar}(\hat{\eta}^{MBM} - \hat{\eta}^{PO})}^{-\frac{1}{2}}\sqrt{n}\left[\left(\hat{\eta}^{MBM} - \hat{\eta}^{PO}\right) - \left(\eta^{MBM} - \eta^{PO}\right)\right] \xrightarrow{d} N(0, I) \tag{D-54}$$

where

$$\widehat{\text{avar}(\hat{\eta}^{MBM} - \hat{\eta}^{PO})} = c(\hat{\tau})\ \widehat{\text{AVAR}(\hat{\tau})}\ c(\hat{\tau})'$$

and

$$c(\tau) = \left[\left(\frac{\omega}{\nu} - \zeta \right) \quad 1/\nu \quad -\omega/\nu^2 \quad -\beta_{P1} \right].$$

Appendix E.

Asymptotic Distribution (and Standard Error) of $\hat{\eta}^{UPO}$ in Eqn (3-10)

We may write $\hat{\eta}^{UPO}$ as

$$\hat{\eta}^{UPO} = \frac{\hat{k}}{\hat{v}} \times \hat{m}_{\mathcal{P}}$$

where

$$\hat{k} = \sum_{i=1}^{n} \frac{1}{n} K(\hat{a}, W_i)$$

$$\hat{v} = \sum_{i=1}^{n} \frac{1}{n} V(\hat{a}, W_i)$$

$$\hat{m}_{\mathcal{P}} = \sum_{i=1}^{n} \frac{1}{n} \mathcal{P}_i$$

Using the corresponding consistent estimators for k, v and $m_{\mathcal{P}}$ say

$$K(a, W) = \lambda(Wa_1)\exp(Wa_2)a_{\mathcal{P}1} + \Lambda(Wa_1)\exp(Wa_2)a_{\mathcal{P}2}$$

$$V(a, W) = \Lambda(Wa_1)\exp(Wa_2)$$

$$m_{\mathcal{P}} = E[\mathcal{P}]$$

$\mathcal{P}$ is the nominal prices of alcohol

$\hat{a} = [\hat{a}_1' \quad \hat{a}_2']'$ (with $\hat{a}_1' = [\hat{a}_{\mathcal{P}1} \quad \hat{a}_{X1}']$ and $\hat{a}_2' = [\hat{a}_{\mathcal{P}2} \quad \hat{a}_{X2}']$) is the consistent estimate of the parameter vector $a = [a_1' \quad a_2']'$ (with $a_1' = [a_{\mathcal{P}1} \quad a_{X1}']$ and $a_2' = [a_{\mathcal{P}2} \quad a_{X2}']$)) [the parameters of equation (3-10)] obtained via the two-part protocol culminating in (3-8) using nominal prices of alcohol]

and

$W_i = [\mathcal{P}_i \quad X_i]$ denotes the observation on $W = [\mathcal{P} \quad X]$ for the ith individual in the sample (i = 1, ..., n).

Let $\hat{\gamma} = [\hat{k} \quad \hat{v} \quad \hat{m}_{\mathcal{P}}]'$ and $\gamma = [k \quad v \quad m_{\mathcal{P}}]'$, where $\text{plim}[\hat{\gamma}] = \gamma$. If we could show that

$$\text{AVAR}(\hat{\gamma})^{-\frac{1}{2}} \sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, I)$$

where the formulation of $\text{AVAR}(\hat{\gamma})$ is known, then we could apply the δ-method to obtain the asymptotic variance of $\hat{\eta}^{UPO}$ as

$$\text{avar}(\hat{\eta}^{UPO}) = c(\gamma)\ \text{AVAR}(\hat{\gamma})\ c(\gamma)'$$

where $c(\gamma) = [m_{\mathcal{P}} / v \quad -k m_{\mathcal{P}} / v^2 \quad k / v\ ]$. Moreover, if we have a consistent estimator for $\text{AVAR}(\hat{\gamma})$, say $\widehat{\text{AVAR}(\hat{\gamma})}$ [i.e. $\text{plim}\left[\widehat{\text{AVAR}(\hat{\gamma})}\right] = \text{AVAR}(\hat{\gamma})$], then we could consistently estimate $\text{avar}(\hat{\eta}^{UPO})$ as

$$\widehat{\text{avar}(\hat{\eta}^{UPO})} = c(\hat{\gamma})\ \widehat{\text{AVAR}(\hat{\gamma})}\ c(\hat{\gamma})'. \qquad \text{(E-1)}$$

We focus, therefore, on finding the asymptotic distribution of $\hat{\gamma}$ and, in particular, the formulation of its asymptotic covariance matrix.

First, let $\theta' = [\delta' \quad \gamma']$ where $\delta = [a_1' \quad a_2']'$, $\gamma' = [k \quad v \quad m_{\mathcal{P}}]$ (recall, $a_1' = [a_{\mathcal{P}1} \quad a_{X1}']$ and $a_2' = [a_{\mathcal{P}2} \quad a_{X2}']$), and note that $\hat{\gamma}$ can be viewed as the second stage estimator in the following two-stage protocol

First Stage

Consistently estimate δ via the following optimization estimator

$$\hat{\delta} = \arg\max_{\tilde{\delta}} \frac{\sum_{i=1}^{n} q_1(\tilde{\delta}, S_i)}{n} \tag{E-2}$$

where

$$q_1(\tilde{\delta}, S_i) = q_{11}(\tilde{a}_1, S_i) + q_{12}(\tilde{a}_2, S_i)$$

$$q_{11}(\tilde{a}_1, S_i) = I(A_i > 0)\ln[\Lambda(W_i\tilde{a}_1)] + [1 - I(A_i > 0)]\ln[1 - \Lambda(W_i\tilde{a}_1)]$$

$$q_{12}(\tilde{a}_2, S_i) = -I(A_i > 0)\left(A_i - \exp\left(W_i\tilde{a}_2\right)\right)^2$$

$$S_i = [A_i \quad X_i \quad \mathcal{P}_i]$$

$\tilde{\delta} = [\tilde{a}_1' \quad \tilde{a}_2']'$, $\tilde{\delta}_1' = [\tilde{a}_{\mathcal{P}1} \quad \tilde{a}_{X1}']$ and $\tilde{a}_2' = [\tilde{a}_{\mathcal{P}2} \quad \tilde{a}_{X2}']$, $\hat{\delta} = [\hat{a}_1' \quad \hat{a}_2']'$, $\hat{a}_1' = [\hat{a}_{\mathcal{P}1} \quad \hat{a}_{X1}']$ and $\hat{a}_2' = [\hat{a}_{\mathcal{P}2} \quad \hat{a}_{X2}']$.

Second Stage

Consistently estimate γ via the following optimization estimator

$$\hat{\gamma} = \arg\max_{\tilde{\gamma}} \frac{\sum_{i=1}^{n} q(\hat{\delta}, \tilde{\gamma}, S_i)}{n} \tag{E-3}$$

where

$$q(\hat{\delta},\ \tilde{\gamma},\ S_i) = q_a(\hat{\delta},\ \tilde{k},\ S_i) + q_b(\hat{\delta},\ \tilde{v},\ S_i) + q_c(\hat{\delta},\ \tilde{m}_{\mathcal{P}},\ S_i)$$

$$q_a(\hat{\delta},\ \tilde{k},\ S_i) = -(K(\hat{a},\ W_i) - \tilde{k})^2$$

$$q_b(\hat{\delta},\ \tilde{v},\ S_i) = -(V(\hat{\alpha},\ W_i) - \tilde{v})^2$$

$$q_c(\hat{\delta},\ \tilde{m}_{\mathcal{P}},\ S_i) = -(\mathcal{P}_i - \tilde{m}_{\mathcal{P}})^2$$

and $\hat{\delta}$ is the first stage estimator of δ. Use q_l as shorthand notation for

$$q_l(\delta, S) = q_{l1}(a_1, S) + q_{l2}(a_2, S)$$

with

$$q_{l1}(a_1, S) = I(A > 0)\ln[\Lambda(W a_1)] + [1 - I(A > 0)]\ln[1 - \Lambda(W a_1)]$$

$$q_{l2}(a_2, S) = -I(A > 0)\left(A - \exp\left(W a_2\right)\right)^2$$

$$S = [A \quad X \quad \mathcal{P}].$$

Moreover, use q as shorthand notation for

$$q(\delta,\tilde{\gamma},\ S) = q_a(\delta,\ k,\ S) + q_b(\delta, v,\ S) + q_c(\delta, m_{\mathcal{P}},\ S)$$

with

$$q_a(\delta,\ k,\ S) = -(K(\alpha,\ W) - k)^2$$

$$q_b(\delta,\ v,\ S) = -(V(a\ ,\ W) - v)^2$$

$$q_c(\delta,\ m_{\mathcal{P}},\ S) = -(\mathcal{P} - m_{\mathcal{P}})^2$$

and let $AVAR(\hat{\delta})$ denote the asymptotic covariance matrix of the first stage estimator.[42] Using well known results from asymptotic theory for two-stage estimators, we can show that[43]

$$AVAR(\hat{\gamma})^{-\frac{1}{2}}\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, I) \tag{E-4}$$

where, $plim(\hat{\gamma}) = \gamma$

$$\begin{aligned} AVAR(\hat{\gamma}) = E\left[\nabla_{\gamma\gamma} q\right]^{-1} \Bigg[& E\left[\nabla_{\gamma\delta} q\right] AVAR(\hat{a}) E\left[\nabla_{\gamma\delta} q\right]' \\ & - E\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma\delta} q\right]' \\ & - E\left[\nabla_{\gamma\delta} q\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right]' \Bigg] E\left[\nabla_{\gamma\gamma} q\right]^{-1} \\ & + E\left[\nabla_{\gamma\gamma} q\right]^{-1} E\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right] E\left[\nabla_{\gamma\gamma} q\right]^{-1}. \end{aligned} \tag{E-5}$$

Fortunately, (E-5) can be simplified in a number of ways. Note that we can write

$$E\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right] = E\left[\nabla_{\gamma} q' E[\nabla_{\delta} q_1 \mid W]\right]$$

but

$$\nabla_{\delta} q_1 = \nabla_{\delta} q_{11} + \nabla_{\delta} q_{12}$$

with

---

[42] Note that a consistent estimate $AVAR(\hat{\delta})$ can be obtained from the packaged output for the first stage estimator because the first stage estimator is unaffected by the fact that it is a component of a two-stage estimator.

[43] Discussions of asymptotic theory for two-stage optimization estimators can be found in Newey & McFadden (1994), White (1994, Chapter 6), and Wooldridge (2010, Chapter 12).

$$\nabla_{\delta} q_{11} = [\nabla_{a_1} \ln f(I(A > 0) \mid W) \quad 0]$$

$$\nabla_{\delta} q_{12} = \left[0 \quad 2\,I(A > 0)\left(A - \exp\left(Wa_2\right)\right)\exp\left(Wa_2\right)W\right]$$

where

$$f(I(A > 0) \mid W) = \Lambda(Wa_1)^{I(A>0)}[1 - \Lambda(Wa_1)]^{[1 - I(A>0)]} .$$

Therefore

$$E[\nabla_{\delta} q_1 \mid W] = \left[E[\nabla_{a_1} \ln f(I(A > 0) \mid W)] \quad 2E\left[I(A > 0)\left(A - \exp\left(Wa_2\right)\right)\exp\left(Wa_2\right) \mid W\right]W\right]$$

$$= [\,0 \quad 0]$$

because $E[\nabla_{a_1} \ln f(I(A > 0) \mid W)] = 0$ [see (13.20) on p. 477 of Wooldridge (2010)] and,

$E\left[I(A > 0)\left(A - \exp\left(Wa_2\right)\right)\exp\left(Wa_2\right) \mid W\right] = 0$ by design. Finally, then we get

$$E\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right] = 0$$

so

$$\underset{3\times3}{AVAR(\hat{\gamma})} = \underset{3\times3}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \underset{3\times2K}{E\left[\nabla_{\gamma\delta} q\right]} \underset{2K\times2K}{AVAR(\hat{a})} \underset{2K\times3}{E\left[\nabla_{\gamma\delta} q\right]'} \underset{3\times3}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}}$$

$$+ \underset{3\times3}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \underset{3\times3}{E\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right]} \underset{3\times3}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} . \qquad \text{(E-6)}$$

Let's consider each of the individual components of (E-6) in turn.

$\underset{2K\times2K}{\mathbf{E[\nabla_{\delta\delta} q_1]^{-1}}}$

Written out explicitly we have

$$E[\nabla_{aa} q_1]^{-1} = \begin{bmatrix} E[\nabla_{a_1 a_1} q_{11}]^{-1} & 0 \\ 0 & E[\nabla_{a_2 a_2} q_{12}]^{-1} \end{bmatrix}. \quad \text{(E-7)}$$

Now

$$E[\nabla_{a_1 a_1} q_{11}]^{-1} = -\text{AVAR}(\hat{a}_1) \quad \text{(E-8)}$$

= the negative of the asymptotic covariance matrix for first stage, first part, logit estimation in the two-stage estimation protocol for θ. We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{a_1 a_1} q_{11}]^{-1}$ is

$$\hat{E}[\nabla_{a_1 a_1} q_{11}]^{-1} = -n\widehat{\text{AVAR}}^*(\hat{a}_1) \quad \text{(E-9)}$$

where $\widehat{\text{AVAR}}^*(\hat{a}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure. Also

$$\nabla_{a_2} q_{12} = 2\,I(A > 0)\left(A - \exp(Wa_2)\right)\exp(Wa_2)W$$

and

$$\nabla_{a_2 a_2} q_{12} = 2\,I(A > 0)\left[\left(A - \exp(Wa_2)\right) - \exp(Wa_2)\right]\exp(Wa_2)W'W.$$

Therefore

$$E[\nabla_{a_2 a_2} q_{12}] = 2E\left[I(A > 0)\left[\left(A - \exp(Wa_2)\right) - \exp(Wa_2)\right]\exp(Wa_2)W'W\right]. \quad \text{(E-10)}$$

A consistent estimator of $E[\nabla_{a_2 a_2} q_{12}]^{-1}$ is

$$\hat{E}[\nabla_{a_2 a_2} q_{12}]^{-1}$$

$$= n_1 \left[2 \sum_{i=1}^{n} \left\{ I(A_i > 0) \left[\left(A_i - \exp(W_i \hat{a}_2) \right) - \exp(W_i \hat{a}_2) \right] \exp(W_i \hat{a}_2) W_i' W_i \right\} \right]^{-1} \tag{E-11}$$

where n_1 is the size of the subsample for whom $I(A > 0) = 1$, so

$$\hat{E}[\nabla_{\delta\delta} q_1]^{-1} = \hat{E}[\nabla_{aa} q_1]^{-1} = \begin{bmatrix} \hat{E}[\nabla_{a_1 a_1} q_{11}]^{-1} & 0 \\ 0 & \hat{E}[\nabla_{a_2 a_2} q_{12}]^{-1} \end{bmatrix}. \tag{E-12}$$

$\mathbf{E}\left[\nabla_{\delta} q_1 ' \nabla_{\delta} q_1\right]$

2K×2K

Written out explicitly we have

$$E\left[\nabla_a q_1 ' \nabla_a q_1\right] = \begin{bmatrix} E[\nabla_{a_1} q_{11} ' \nabla_{a_1} q_{11}] & E[\nabla_{a_1} q_{11} ' \nabla_{a_2} q_{12}] \\ E[\nabla_{a_2} q_{12} ' \nabla_{a_1} q_{11}] & E[\nabla_{a_2} q_{12} ' \nabla_{a_2} q_{12}] \end{bmatrix}. \tag{E-13}$$

Because the first stage, first part, estimator of α_1 is MLE we can write

$$E[\nabla_{a_1} q_{11} ' \nabla_{a_1} q_{11}] = -E[\nabla_{a_1 a_1} q_{11}] = \left[AVAR(\hat{a}_1) \right]^{-1}$$

= the inverse of the asymptotic covariance matrix for first stage, first part, logit estimation in the two-stage estimation protocol for θ. We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{a_1} q_{11}' \nabla_{a_1} q_{11}]$ is

$$\hat{E}[\nabla_{a_1} q_{11}' \nabla_{a_1} q_{11}] = \frac{1}{n}\left[\widehat{AVAR}^*(\hat{a}_1)\right]^{-1} \tag{E-14}$$

where $\widehat{AVAR}^*(\hat{a}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure. The remainder of the block elements follow from

$$\nabla_{a_1} q_{11} = \nabla_{a_1} \ln f(I(A>0) \mid W) = \left[I(A>0)[1-\Lambda(Wa_1)] - [1-I(A>0)]\Lambda(Wa_1)\right]W \tag{E-15}$$

$$\nabla_{a_2} q_{12} = 2\,I(A>0)\left(A - \exp\left(Wa_2\right)\right)\exp\left(Wa_2\right)W\,. \tag{E-16}$$

where the formulation of $\nabla_{a_1} q_{11}$ comes from equation (16.4.8) on p. 350 of Fomby et al. (1984) and $\Lambda(\ \ )$ denotes the logistic cdf. The remaining required consistent matrix estimators are:

$$\hat{E}[\nabla_{a_1} q_{11}' \nabla_{a_2} q_{12}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{a_1}\hat{q}_{11i}' \nabla_{a_2}\hat{q}_{12i} \tag{E-17}$$

$$\hat{E}[\nabla_{a_2} q_{12}' \nabla_{a_2} q_{12}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{a_2}\hat{q}_{12i}' \nabla_{a_2}\hat{q}_{12i} \tag{E-18}$$

where

$$\nabla_{a_1}\hat{q}_{11i} = \left[I(A_i>0)[1-\Lambda(W_i\hat{a}_1)] - [1-I(A_i>0)]\Lambda(W_i\hat{a}_1)\right]W_i \tag{E-19}$$

and

$$\nabla_{a_2}\hat{q}_{12i} = 2\,I(A_i>0)\left(A_i - \exp\left(W_i\hat{a}_2\right)\right)\exp\left(W_i\hat{a}_2\right)W_i \tag{E-20}$$

so

$$\hat{E}\left[\nabla_{\delta} q_1 \,'\nabla_{\delta} q_1\right] = \hat{E}\left[\nabla_{a} q_1 \,'\nabla_{a} q_1\right] = \begin{bmatrix} \hat{E}[\nabla_{a_1} q_{11} \,'\nabla_{a_1} q_{11}] & \hat{E}[\nabla_{a_1} q_{11} \,'\nabla_{a_2} q_{12}] \\ \hat{E}[\nabla_{a_2} q_{12} \,'\nabla_{a_1} q_{11}] & \hat{E}[\nabla_{a_2} q_{12} \,'\nabla_{a_2} q_{12}] \end{bmatrix}. \quad \text{(E-21)}$$

$$\underset{3\times 2K}{\mathbf{E}\left[\boldsymbol{\nabla}_{\gamma\delta}\mathbf{q}\right]}$$

Written out explicitly we have

$$\begin{aligned} \nabla_{\gamma} q &= [(\nabla_{k} q_a + 0 + 0) \quad (0 + \nabla_{v} q_b + 0) \quad (0 + 0 + \nabla_{m_{\mathcal{P}}} q_c)] \\ &= [\nabla_{k} q_a \quad \nabla_{v} q_b \quad \nabla_{m_{\mathcal{P}}} q_c] \\ &= 2\left[\left(K(a,\, W) - k\right) \quad \left(V(a,\, W) - v\right) \quad (\mathcal{P}_i - m_{\mathcal{P}})\right] \end{aligned} \quad \text{(E-22)}$$

and

$$\begin{aligned} E[\nabla_{\gamma\delta} q] &= \begin{bmatrix} E[\nabla_{ka_1} q_a] & E[\nabla_{ka_2} q_a] \\ E[\nabla_{va_1} q_b] & E[\nabla_{va_2} q_b] \\ E[\nabla_{m_{\mathcal{P}} a_1} q_c] & E[\nabla_{m_{\mathcal{P}} a_2} q_c] \end{bmatrix} \\ &= 2\begin{bmatrix} E[\nabla_{a_1} K] & E[\nabla_{a_2} K] \\ E[\nabla_{a_1} V] & E[\nabla_{a_2} V] \\ 0 & 0 \end{bmatrix} \end{aligned} \quad \text{(E-23)}$$

where

$$\begin{aligned} \nabla_{a_1} K &= \{\lambda(Wa_1)[1 - 2\Lambda(Wa_1)]\, a_{\mathcal{P}1} W + \lambda(Wa_1)[1 \quad 0 \ldots 0]\} \exp(Wa_2) \\ &\qquad + \lambda(Wa_1) \exp(Wa_2)\, a_{\mathcal{P}2} W \\ &= \lambda(Wa_1) \exp(Wa_2)\left[\{[1 - 2\Lambda(Wa_1)]\, a_{\mathcal{P}1} + a_{\mathcal{P}2}\} W + [1 \quad 0 \ldots 0]\right] \end{aligned} \quad \text{(E-24)}$$

$$\begin{aligned} \nabla_{a_2} K &= \lambda(Wa_1) \exp(Wa_2)\, a_{\mathcal{P}1} W + \Lambda(Wa_1)\{\exp(Wa_2)\, a_{\mathcal{P}2} W + \exp(Wa_2)[1 \quad 0 \ldots 0]\} \\ &= \exp(Wa_2)\left[\{\lambda(Wa_1)\, a_{\mathcal{P}1} + \Lambda(Wa_1)\, a_{\mathcal{P}2}\} W + \Lambda(Wa_1)[1 \quad 0 \ldots 0]\right]. \end{aligned} \quad \text{(E-25)}$$

$$\nabla_{a_1} V = \lambda(Wa_1)\exp(Wa_2)\,W \tag{E-26}$$

and

$$\nabla_{a_2} V = \Lambda(Wa_1)\exp(Wa_2)\,W\,. \tag{E-27}$$

The following equalities were used in deriving the above results

$$\nabla_a \Lambda(a) = \lambda(a) = \Lambda(a)[1 - \Lambda(a)]$$

$$\nabla_a \lambda(a) = \lambda(a)[1 - 2\Lambda(a)]\,.$$

The requisite consistent matrix estimators are

$$\hat{E}[\nabla_{a_1} K] = \sum_{i=1}^{n} \widehat{\nabla_{a_1} K}_i \tag{E-28}$$

$$\hat{E}[\nabla_{a_2} K] = \sum_{i=1}^{n} \widehat{\nabla_{a_2} K}_i \tag{E-29}$$

$$\hat{E}[\nabla_{a_1} V] = \sum_{i=1}^{n} \widehat{\nabla_{a_1} V}_i \tag{E-30}$$

and

$$\hat{E}[\nabla_{a_2} V] = \sum_{i=1}^{n} \widehat{\nabla_{a_2} V}_i \tag{E-31}$$

where

$$\widehat{\nabla_{a_1} K}_i = \lambda(W_i \hat{a}_1)\exp(W_i \hat{a}_2)\left[\{[1 - 2\Lambda(W_i \hat{a}_1)]\hat{a}_{\mathcal{P}1} + \hat{a}_{\mathcal{P}2}\}W_i + [1 \quad 0 \ldots 0]\right] \tag{E-32}$$

$$\widehat{\nabla_{a_2} \mathcal{K}}_i = \exp(W_i \hat{a}_2)\left[\{\lambda(W_i \hat{a}_1)\hat{a}_{\mathcal{P}1} + \Lambda(W_i \hat{a}_1)\hat{a}_{\mathcal{P}2}\}W_i + \Lambda(W_i \hat{a}_1)[1 \quad 0 \ldots 0]\right] \tag{E-33}$$

$$\widehat{\nabla_{a_1} V}_i = \lambda(W_i \hat{a}_1)\exp(W_i \hat{a}_2)\,W_i \tag{E-34}$$

and

$$\widehat{\nabla_{a_1} V}_i = \Lambda(W_i \hat{a}_1)\exp(W_i \hat{a}_2)\,W_i \tag{E-35}$$

so

$$\hat{E}[\nabla_{\gamma\delta} q] = 2\begin{bmatrix} \hat{E}[\nabla_{a_1} K] & \hat{E}[\nabla_{a_2} K] \\ \hat{E}[\nabla_{a_1} V] & \hat{E}[\nabla_{a_2} V] \\ 0 & 0 \end{bmatrix}. \tag{E-36}$$

$\mathbf{E[\nabla_{\gamma\gamma} q]^{-1}}$
3×3

$$E\left[\nabla_{\gamma\gamma} q\right]^{-1} = \begin{bmatrix} E\left[\nabla_{kk} q_a\right]^{-1} & 0 & 0 \\ 0 & E\left[\nabla_{vv} q_b\right]^{-1} & 0 \\ 0 & 0 & E\left[\nabla_{m_{\mathcal{P}} m_{\mathcal{P}}} q_c\right]^{-1} \end{bmatrix} \tag{E-37}$$

because $\nabla_{kv} q_a = \nabla_{vk} q_b = \nabla_{km_{\mathcal{P}}} q_a = \nabla_{vm_{\mathcal{P}}} q_b = \nabla_{m_{\mathcal{P}} k} q_c = \nabla_{m_{\mathcal{P}} v} q_c = 0$. Now

$$\nabla_{kk} q_a = \nabla_{vv} q_b = \nabla_{m_{\mathcal{P}} m_{\mathcal{P}}} q_c = -2$$

therefore

$$E\left[\nabla_{\gamma\gamma} q\right]^{-1} = \begin{bmatrix} -\frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & -\frac{1}{2} \end{bmatrix}. \tag{E-38}$$

$\mathbf{E\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right]}$
3×3

Given that

$$\nabla_{\gamma} q = 2\left[(K(a, W) - k) \quad (V(a, W) - v) \quad (\mathcal{P}_i - m_{\mathcal{P}})\right]. \tag{E-39}$$

we have

$$E\left[\nabla_\gamma q' \nabla_\gamma q\right] = 4\begin{bmatrix} E[(K(a, W)-k)^2] & E[(K(a,W)-k)(V(a, W)-v)] & E[(K(a, W)-k)(\mathcal{P}-m_{\mathcal{P}}))] \\ E[(K(a, W)-k)(V(a, W)-v)] & E[(V(a, W)-v)^2] & E[(V(a, W)-v)(\mathcal{P}-m_{\mathcal{P}})] \\ E[(K(a, W)-k)(\mathcal{P}-m_{\mathcal{P}})] & E[(V(a, W)-v)(\mathcal{P}-m_{\mathcal{P}})] & E[(\mathcal{P}-m_{\mathcal{P}})^2] \end{bmatrix}. \tag{E-40}$$

The corresponding consistent estimator is

$$\hat{E}\left[\nabla_\gamma q' \nabla_\gamma q\right] = 4\begin{bmatrix} \sum_{i=1}^{n}\frac{1}{n}(\hat{K}_i-\hat{k})^2 & \sum_{i=1}^{n}\frac{1}{n}\begin{pmatrix}(\hat{K}_i-\hat{k})\\(\hat{V}_i-\hat{v})\end{pmatrix} & \sum_{i=1}^{n}\frac{1}{n}\begin{pmatrix}(\hat{K}_i-\hat{k})\\(\mathcal{P}_i-\hat{m}_{\mathcal{P}})\end{pmatrix} \\ \sum_{i=1}^{n}\frac{1}{n}\begin{pmatrix}(\hat{K}_i-\hat{k})\\(\hat{V}_i-\hat{v})\end{pmatrix} & \sum_{i=1}^{n}\frac{1}{n}(\hat{V}_i-\hat{v})^2 & \sum_{i=1}^{n}\frac{1}{n}\begin{pmatrix}(\hat{V}_i-\hat{v})\\(\mathcal{P}_i-\hat{m}_{\mathcal{P}})\end{pmatrix} \\ \sum_{i=1}^{n}\frac{1}{n}\begin{pmatrix}(\hat{K}_i-\hat{k})\\(\mathcal{P}_i-\hat{m}_{\mathcal{P}})\end{pmatrix} & \sum_{i=1}^{n}\frac{1}{n}\begin{pmatrix}(\hat{V}_i-\hat{v})\\(\mathcal{P}_i-\hat{m}_{\mathcal{P}})\end{pmatrix} & \sum_{i=1}^{n}\frac{1}{n}(\mathcal{P}_i-\hat{m}_{\mathcal{P}})^2 \end{bmatrix}. \tag{E-41}$$

Based on the above results, we can consistently estimate (E-5) as

$$\widehat{AVAR(\hat{\gamma})} = \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma\delta} q\right] \widehat{AVAR(\hat{a})}\, \hat{E}\left[\nabla_{\gamma\delta} q\right]' \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$
$$+\ \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_\gamma q' \nabla_\gamma q\right] \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

and using well known results from asymptotic theory for two-stage estimators, we can show that[44]

$$\widehat{\text{AVAR}(\hat{\gamma})}^{-\frac{1}{2}}\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, I)\,. \tag{E-42}$$

Combining (E-42) with (E-1) we also have that

$$\widehat{\text{avar}(\hat{\eta}^{UPO})}^{-\frac{1}{2}}\sqrt{n}(\hat{\eta}^{UPO} - \eta^{UPO}) \xrightarrow{d} N(0, I) \tag{E-43}$$

where $\widehat{\text{avar}(\hat{\eta}^{UPO})}$ is given in (E-1).

[44]Discussions of asymptotic theory for two-stage optimization estimators can be found in Newey & McFadden (1994), White (1994, Chapter 6), and Wooldridge (2010, Chapter 12).

Appendix F.

Bias from Using $\hat{\eta}^{AL}$ Instead of $\hat{\eta}^{UPO}$ in a Model of Alcohol Demand

As we saw in section 2.2 of Chapter 3, in the AGG-LOG framework the elasticity measure is defined as $\eta^{AL} = \pi_P$ in equation (3-4) rewritten here for convenience

$$\ln(\bar{A}) = \bar{P}\pi_P + \bar{X}\pi_X + \xi \tag{F-1}$$

where $\bar{A}$ is the average per-capita consumption of alcohol from aggregated data, $\bar{P}$ is log of average price of alcohol from aggregated data and $\bar{X}$ is the vector of average values of observable confounders. The corresponding consistent elasticity estimator, $\hat{\pi}_P$, is the OLS estimator of π_P in (F-1). Using (3-10), we can then write the difference between η^{AL} and η^{UPO} as

$$\begin{aligned}\eta^{AL} - \eta^{UPO} = \pi_P - E\big[&\lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})a_{\mathcal{P}1} \\ &+ \Lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})a_{\mathcal{P}2}\big] \\ &\times \frac{E[\mathcal{P}]}{E\big[\Lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})\big]}\end{aligned}$$

$$\begin{aligned}= \pi_P - E\big[&\Lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})[1 - \Lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})]\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})a_{\mathcal{P}1} \\ &+ \Lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})a_{\mathcal{P}2}\big] \\ &\times \frac{E[\mathcal{P}]}{E\big[\Lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})\big]}.\end{aligned} \tag{F-2}$$

If we define

$$v \equiv E[\Lambda(W a_1)\exp(W a_2)] \tag{F-3}$$

$$u \equiv E[\Lambda(W a_1)^2 \exp(W a_2)] \tag{F-4}$$

$$m_{\mathcal{P}} = E[\mathcal{P}]$$

and

$$W = [\mathcal{P} \quad X]$$

then (F-2) can be written

$$\eta^{AL} - \eta^{UPO} = \pi_P - \frac{[(v-u)a_{\mathcal{P}1} + va_{\mathcal{P}2}]m_{\mathcal{P}}}{v}$$

$$= \pi_P - \left[\left(1-\frac{u}{v}\right)a_{\mathcal{P}1} + a_{\mathcal{P}2}\right]m_{\mathcal{P}} \ .$$

Appendix G.

Asymptotic Distribution (and Standard Error) of $\hat{\eta}^{AL} - \hat{\eta}^{UPO}$

In Appendix F we showed that

$$\eta^{AL} - \eta^{UPO} = \pi_P - \left[\left(1-\frac{u}{v}\right)a_{\mathcal{P}1} + a_{\mathcal{P}2}\right]m_{\mathcal{P}}$$

$$v \equiv E[\Lambda(Wa_1)\exp(Wa_2)]$$

$$u \equiv E[\Lambda(Wa_1)^2\exp(Wa_2)]$$

$$m_{\mathcal{P}} = E[\mathcal{P}]$$

and

$$W = [\mathcal{P} \quad X].$$

Using the corresponding consistent estimators for u, v and $m_{\mathcal{P}}$ say

$$\hat{v} = \sum_{i=1}^{n}\frac{1}{n}V(\hat{a}, W_i)$$

$$\hat{u} = \sum_{i=1}^{n}\frac{1}{n}U(\hat{a}, W_i)$$

and

$$\hat{m}_{\mathcal{P}} = \sum_{i=1}^{n}\frac{1}{n}\mathcal{P}_i$$

where

$$V(a, W) = \Lambda(W a_1)\exp(W a_2)$$

$U(a, W) = \Lambda(Wa_1)^2 \exp(Wa_2)$

$\hat{a} = [\hat{a}_1' \quad \hat{a}_2']'$ (with $\hat{a}_1' = [\hat{a}_{\mathcal{P}1} \quad \hat{a}_{X1}']$ and $\hat{a}_2' = [\hat{a}_{\mathcal{P}2} \quad \hat{a}_{X2}']$) is the consistent estimate of the parameter vector $a = [a_1' \quad a_2']'$ (with $a_1' = [a_{\mathcal{P}1} \quad a_{X1}']$ and $a_2' = [a_{\mathcal{P}2} \quad a_{X2}']$)) [the parameters of equations (3-6) and (3-7), respectively] obtained via the unrestricted two-part protocol culminating in (3-8) .

and

$W_i = [\mathcal{P}_i \quad X_i]$ denotes the observation on $W = [\mathcal{P} \quad X]$ for the ith individual in the sample (i = 1, ..., n)

we can write

$$\hat{\eta}^{AL} - \hat{\eta}^{UPO} = \hat{\pi}_P - \left[\left(1 - \frac{\hat{u}}{\hat{v}}\right)\hat{a}_{\mathcal{P}1} + \hat{a}_{\mathcal{P}2}\right]\hat{m}_{\mathcal{P}}$$

with

$\hat{\pi} = [\hat{\pi}']'$ ($\hat{\pi}' = [\hat{\pi}_P \quad \hat{\pi}_X']$) is the OLS estimate of the parameter vector $\pi = [\pi']'$ ($\pi' = [\pi_P \quad \pi_X']$) – the vector of parameters in (3-3)].

Let $\hat{\tau} = [\hat{\pi}_P \quad \hat{a}_{\mathcal{P}1} \quad \hat{a}_{\mathcal{P}2} \quad \hat{u} \quad \hat{v} \quad \hat{m}_{\mathcal{P}}]'$ and $\tau = [\pi_P \quad a_{\mathcal{P}1} \quad a_{\mathcal{P}2} \quad u \quad v \quad m_{\mathcal{P}}]'$, where $\text{plim}[\hat{\tau}] = \tau$. If we could show that

$$\text{AVAR}(\hat{\tau})^{-\frac{1}{2}}\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{d} N(0, I)$$

where the formulation of $\text{AVAR}(\hat{\tau})$ is known, then we could apply the δ-method to obtain the asymptotic variance of $\hat{\eta}^{AL} - \hat{\eta}^{UPO}$ as

$$\text{avar}(\hat{\eta}^{AL} - \hat{\eta}^{UPO}) = c(\tau)\ \text{AVAR}(\hat{\tau})\ c(\tau)'$$

with

$$c(\tau) = \left[1 \quad -m_{\mathcal{P}}\left(1-\frac{u}{v}\right) - m_{\mathcal{P}} \quad m_{\mathcal{P}} a_{\mathcal{P}1}/v \quad -m_{\mathcal{P}} a_{\mathcal{P}1} u/v^2 \quad -\left[\left(1-\frac{u}{v}\right)a_{\mathcal{P}1} + a_{\mathcal{P}2}\right] \right].$$

Moreover, if we have a consistent estimator for $\text{AVAR}(\hat{\tau})$, say $\widehat{\text{AVAR}(\hat{\tau})}$ [i.e. $\text{plim}\left[\widehat{\text{AVAR}(\hat{\tau})}\right] = \text{AVAR}(\hat{\tau})$], then we could consistently estimate $\text{avar}(\hat{\eta}^{AL} - \hat{\eta}^{UPO})$ as

$$\widehat{\text{avar}(\hat{\eta}^{AL} - \hat{\eta}^{UPO})} = c(\hat{\tau})\ \widehat{\text{AVAR}(\hat{\tau})}\ c(\hat{\tau})'. \qquad \text{(G-1)}$$

We focus, therefore, on finding the asymptotic distribution of $\hat{\tau}$ and, in particular, the formulation of its asymptotic covariance matrix.

First note that we can write τ as

$$\tau = \Xi\,\theta \qquad \text{(G-2)}$$

where $\theta' = [\delta' \quad \gamma']$, $\delta = [\pi' \quad a_1' \quad a_2']'$, $\gamma' = [u \quad v \quad m_{\mathcal{P}}]$ (recall, $\pi' = [\pi_P \quad \pi_X']$, $a_1' = [a_{\mathcal{P}1} \quad a_{X1}']$, and $a_2' = [a_{\mathcal{P}2} \quad a_{X2}']$.

$$\Xi = \begin{bmatrix} \ell_{\pi_P} \\ \hline \ell_{a_{P1}} \\ \hline \ell_{a_{P2}} \\ \hline 0_{3,3K} \quad I_3 \end{bmatrix}$$

ℓ_a is the unit row vector with the value "1" in the element position corresponding to the element position of a in the vector θ, $0_{b,c}$ is the matrix of zeros whose row and column dimensions are b and c, respectively, I_d is the identity matrix of order d, and K is the column dimension of W. For future reference, let's set the following vector/matrix dimensions:

a_1 is K×1

a_2 is K×1

π is K×1

W is 1×K

τ is 6×1

$c(\tau)$ is 1×6

δ is 3K×1

γ is 3×1

θ is (3K+3)×1

Ξ is 6×(3K+3)

$\ell_{a_{\mathcal{P}1}}$ is 1×(3K+3)

Clearly then

$$\text{AVAR}(\hat{\tau}) = \Xi \, \text{AVAR}(\hat{\theta}) \, \Xi' \tag{G-3}$$

where $\hat{\theta}$ is the estimator of θ obtained from the following two-stage protocol.

First Stage

Consistently estimate δ via the following optimization estimator

$$\hat{\tilde{\delta}} = \arg\max_{\tilde{\delta}} \frac{\sum_{i=1}^{n} q_1(\tilde{\delta}, S_i)}{n} \tag{G-4}$$

where

$$q_1(\tilde{\delta}, S_i) = q_{11}(\tilde{\pi}, \bar{S}_i) + q_{12}(\tilde{a}_1, \vec{S}_i) + q_{13}(\tilde{a}_2, \vec{S}_i)$$

$$q_{11}(\tilde{\pi}, \bar{S}_i) = -(\ln(\bar{A}_i) - \bar{W}_i \tilde{\pi})^2$$

$$q_{12}(\tilde{a}_1, \vec{S}_i) = I(A_i > 0)\ln[\Lambda(W_i \tilde{a}_1)] + [1 - I(A_i > 0)]\ln[1 - \Lambda(W_i \tilde{a}_1)]$$

$$q_{13}(\tilde{a}_2, \vec{S}_i) = -I(A_i > 0)\left(A_i - \exp\left(W_i \tilde{a}_2\right)\right)^2$$

$$\bar{W} = \left[\bar{P} \quad \bar{X}\right]$$

$\bar{S}_i = [\bar{A}_i \quad \bar{X}_i \quad \bar{P}_i]$ for q_{11}, and $\vec{S}_i = [A_i \quad X_i \quad \mathcal{P}_i]$ for q_{12} and q_{13}

$\tilde{\delta} = [\tilde{\pi}' \quad \tilde{a}_1' \quad \tilde{a}_2']'$, $\tilde{a}_1' = [\tilde{a}_{\mathcal{P}1} \quad \tilde{a}_{X1}']$, $\tilde{a}_2' = [\tilde{a}_{\mathcal{P}2} \quad \tilde{a}_{X2}']$ and

$\tilde{\pi} = [\tilde{\pi}_P \quad \tilde{\pi}_X']$ and $\hat{\tilde{\delta}} = [\hat{\tilde{\pi}}' \quad \hat{\tilde{a}}_1' \quad \hat{\tilde{a}}_2']'$.

Second Stage

Consistently estimate γ via the following optimization estimator

$$\hat{\gamma} = \arg\max_{\tilde{\gamma}} \frac{\sum_{i=1}^{n} q(\hat{\delta}, \tilde{\gamma}, S_i)}{n} \tag{G-5}$$

where

$$q(\hat{\delta}, \tilde{\gamma}, S_i) = q_a(\hat{\delta}, \tilde{u}, S_i) + q_b(\hat{\delta}, \tilde{v}, S_i) + q_c(\hat{\delta}, \tilde{m}_{\mathcal{P}}, S_i)$$

$$q_a(\hat{\delta}, \tilde{u}, S_i) = -(U(\hat{\delta}, W_i) - \tilde{u})^2$$

$$q_b(\hat{\delta}, \tilde{v}, S_i) = -(V(\hat{\delta}, W_i) - \tilde{v})^2$$

$$q_c(\hat{\delta}, \tilde{m}_{\mathcal{P}}, S_i) = -(\mathcal{P}_i - \tilde{m}_{\mathcal{P}})^2$$

$\hat{\delta} = [\hat{\pi}' \quad \hat{a}_1' \quad \hat{a}_2']'$ is the first stage estimator of δ, $\hat{a}_1' = [\hat{a}_{\mathcal{P}1} \quad \hat{a}_{X1}']$ and $\hat{a}_2' = [\hat{a}_{\mathcal{P}2} \quad \hat{a}_{X2}']$. Use q_1 as shorthand notation for

$$q_1(\delta, S) = q_{11}(\pi, \bar{S}) + q_{12}(a_1, \vec{S}) + q_{13}(a_2, \vec{S})$$

where

$$q_{11}(\pi, \bar{S}) = -(\ln(\bar{A}) - \bar{W}\pi)^2$$

$$q_{12}(a_1, \vec{S}) = I(A > 0)\ln[\Lambda(W a_1)] + [1 - I(A > 0)]\ln[1 - \Lambda(W a_1)]$$

$$q_{13}(a_2, \vec{S}) = -I(A > 0)\left(A - \exp\left(W a_2\right)\right)^2$$

$\bar{S} = [\bar{A} \quad \bar{X} \quad \bar{P}]$ for q_{11}, and $\vec{S} = [A \quad X \quad \mathcal{P}]$ for q_{12} and q_{13},

and use q as shorthand notation for

$$q(\delta, \tilde{\gamma}, S) = q_a(\delta, u, S) + q_b(\delta, v, S) + q_c(\delta, m_{\mathcal{P}}, S)$$

where

$$q_a(\delta, u, S) = -(U(a, W) - u)^2$$

$$q_b(\delta, v, S) = -(V(a, W) - v)^2$$

$$q_c(\delta, m_{\mathcal{P}}, S) = -(\mathcal{P} - m_{\mathcal{P}})^2$$

with $\text{AVAR}(\hat{\delta})$ being the asymptotic covariance matrix of the first stage estimator.[45] Using well known results from asymptotic theory for two-stage estimators, we can show that[46]

$$\text{AVAR}(\hat{\theta})^{-\frac{1}{2}}\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I) \tag{G-6}$$

where $\hat{\theta}' = [\hat{\delta}' \quad \hat{\gamma}']$, $\text{plim}(\hat{\theta}) = \theta$

$$\text{AVAR}(\hat{\theta}) = \begin{bmatrix} D_{11} & D_{12} \\ D_{12}' & D_{22} \end{bmatrix} \tag{G-7}$$

$$D_{11} = \underset{3K\times 3K}{\text{AVAR}(\hat{\delta})} \tag{G-8}$$

$$D_{12} = E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\delta} q_1 ' \nabla_{\gamma} q\right] E\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

$$- E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\delta} q_1 ' \nabla_{\delta} q_1\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma\delta} q\right]' E\left[\nabla_{\gamma\gamma} q\right]^{-1} \tag{G-9}$$

---

[45] Note that a consistent estimate $\text{AVAR}(\hat{\delta})$ can be obtained from the packaged output for the first stage estimator because the first stage estimator is unaffected by the fact that it is a component of a two-stage estimator.

[46] Discussions of asymptotic theory for two-stage optimization estimators can be found in Newey & McFadden (1994), White (1994, Chapter 6), and Wooldridge (2010, Chapter 12).

$$D_{22} = AVAR(\hat{\gamma}) = E\left[\nabla_{\gamma\gamma} q\right]^{-1}\left[E\left[\nabla_{\gamma\delta} q\right] AVAR(\hat{\delta}) E\left[\nabla_{\gamma\delta} q\right]'\right.$$

$$- E\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma\delta} q\right]'$$

$$\left. - E\left[\nabla_{\gamma\delta} q\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right]'\right] E\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

$$+ E\left[\nabla_{\gamma\gamma} q\right]^{-1} E\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right] E\left[\nabla_{\gamma\gamma} q\right]^{-1}. \qquad \text{(G-10)}$$

Fortunately, (G-9) and (G-10) can be simplified in a number of ways. First note that we can write

$$E\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right] = E\left[\nabla_{\gamma} q' E[\nabla_{\delta} q_1 \mid W]\right]$$

but

$$\nabla_{\delta} q_1 = \nabla_{\delta} q_{11} + \nabla_{\delta} q_{12} + \nabla_{\delta} q_{13}$$

with

$$\nabla_{\delta} q_{11} = \left[2(\ln(\bar{A}) - \bar{W}\pi)\bar{W} \quad 0 \quad 0\right]$$

$$\nabla_{\delta} q_{12} = [0 \quad \nabla_{a_1} \ln f(I(A > 0) \mid W) \quad 0]$$

$$\nabla_{\delta} q_{13} = \left[0 \quad 0 \quad 2\,I(A > 0)\left(A - \exp\left(Wa_2\right)\right)\exp\left(Wa_2\right) W\right]$$

where

$$f(I(A > 0) \mid W) = \Lambda(W a_1)^{I(A>0)} [1 - \Lambda(Wa_1)]^{[1 - I(A>0)]} .$$

Therefore

$$E[\nabla_{\delta} q_1 \mid W,\bar{W}] = \begin{bmatrix} 2E\left[(\ln(\bar{A}) - \bar{W}\pi) \mid \bar{W}\right]\bar{W} \\ E[\nabla_{a_1} \ln f(I(A>0) \mid W)] \\ 2E\left[I(A>0)\left(A - \exp(Wa_2)\right)\exp(Wa_2) \mid W\right]W \end{bmatrix}'$$

$$= [\,0 \quad 0 \quad 0]$$

because $E[\nabla_{a_1} \ln f(I(A>0) \mid W)] = 0$ [see (13.20) on p. 477 of Wooldridge (2010)], $E\left[I(A>0)\left(A - \exp(Wa_2)\right)\exp(Wa_2) \mid W\right] = 0$, and $E\left[(\ln(\bar{A}) - \bar{W}\pi) \mid \bar{W}\right] = 0$ by design.[47] Finally, then we get

$$E\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right] = 0$$

so

$$D_{12} = -\underset{3K\times 3K}{E\left[\nabla_{\delta\delta} q_1\right]^{-1}} \underset{3K\times 3K}{E\left[\nabla_{\delta} q_1' \nabla_{\delta} q_1\right]} \underset{3K\times 3K}{E\left[\nabla_{\delta\delta} q_1\right]^{-1}} \underset{3K\times 3}{E\left[\nabla_{\gamma\delta} q\right]'} \underset{3\times 3}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \qquad \text{(G-11)}$$

and

$$D_{22} = AVAR(\hat{\gamma}) = \underset{3\times 3}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \underset{3\times 3K}{E\left[\nabla_{\gamma\delta} q\right]} \underset{3K\times 3K}{AVAR(\hat{\delta})} \underset{3K\times 3}{E\left[\nabla_{\gamma\delta} q\right]'} \underset{3\times 3}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}}$$

$$+ \underset{3\times 3}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \underset{3\times 3}{E\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right]} \underset{3\times 3}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}}. \qquad \text{(G-12)}$$

so

---

[47] The last result warrants some discussion. We are assuming here that the correct model specification is the unrestricted two-part model detailed in (3-5) through (3-7). We also assume that, although (3-4) is not correctly specified in a causal sense, it can, nonetheless be viewed as a "best predictor" model in which the best predictor (the conditional mean of $(\ln(\bar{A}) \mid \bar{P}, \bar{X})$ is assumed to be equal to $\bar{W}\pi = \bar{P}\pi_p + \bar{X}\pi_X$. This implies that $E\left[(\ln(\bar{A}) - \bar{W}\pi) \mid \bar{W}\right] = 0$.

$$\underset{(3K+3)\times(3K+3)}{\text{AVAR}(\hat{\theta})} = \begin{bmatrix} \underset{3K\times 3K}{D_{11}} & \underset{3K\times 3}{D_{12}} \\ \underset{3\times 3K}{D_{12}'} & \underset{3\times 3}{D_{22}} \end{bmatrix}.$$

Let's consider each of the individual components of (G-9) and (G-10) in turn.

$E[\nabla_{\delta\delta} q_1]^{-1}$
3K×3K

Written out explicitly we have

$$E[\nabla_{\delta\delta} q_1]^{-1} = \begin{bmatrix} E[\nabla_{\pi\pi} q_{11}]^{-1} & 0 & 0 \\ 0 & E[\nabla_{a_1 a_1} q_{12}]^{-1} & 0 \\ 0 & 0 & E[\nabla_{a_2 a_2} q_{13}]^{-1} \end{bmatrix}. \qquad \text{(G-13)}$$

Now

$$\nabla_{\pi} q_{11} = 2(\ln(\bar{A}) - \bar{W}\pi)\bar{W}$$

and

$$\nabla_{\pi\pi} q_{11} = -2\bar{W}'\bar{W}.$$

Therefore

$$E[\nabla_{\pi\pi} q_{11}] = -2E\left[\bar{W}'\bar{W}\right]. \qquad \text{(G-14)}$$

A consistent estimator of $E[\nabla_{\pi\pi} q_{11}]^{-1}$ is

$$\hat{E}[\nabla_{\pi\pi} q_{11}]^{-1} = \left[-2\frac{1}{n_2}\sum_{i=1}^{n}\left\{\bar{W}_i'\bar{W}_i\right\}\right]^{-1}$$

$$= n_2 \left[-2 \sum_{i=1}^{n} \left\{ \bar{W}_i' \bar{W}_i \right\} \right]^{-1}. \qquad \text{(G-15)}$$

where $\hat{\pi}$ is the OLS estimator of π in the AGG-LOG model and n_2 is the size of the aggregated sample. Similarly,

$$E[\nabla_{a_1 a_1} q_{12}]^{-1} = -AVAR(\hat{a}_1) \qquad \text{(G-16)}$$

= the negative of the asymptotic covariance matrix for first stage, first part, logit estimation of α_1 in the unrestricted two-part estimation protocol culminating in (14). We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{a_1 a_1} q_{12}]^{-1}$ is

$$\hat{E}[\nabla_{a_1 a_1} q_{12}]^{-1} = -n\widehat{AVAR}^*(\hat{a}_1) \qquad \text{(G-17)}$$

where $\widehat{AVAR}^*(\hat{a}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure. Also

$$\nabla_{a_2} q_{13} = 2\, I(A > 0)\left(A - \exp\left(Wa_2\right)\right)\exp\left(Wa_2\right)W$$

and

$$\nabla_{a_2 a_2} q_{13} = -2\, I(A > 0)\left[\left(A - \exp\left(Wa_2\right)\right) - \exp\left(Wa_2\right)\right]\exp\left(Wa_2\right)W'W .$$

Therefore

$$E[\nabla_{a_2a_2} q_{13}] = 2E\left[I(A>0)\left[\left(A-\exp\left(Wa_2\right)\right)-\exp\left(Wa_2\right)\right]\exp\left(Wa_2\right)W'W\right].$$

(G-18)

A consistent estimator of $E[\nabla_{a_2a_2} q_{13}]^{-1}$ is

$$\hat{E}[\nabla_{a_2a_2} q_{13}]^{-1} = \left[2\frac{1}{n_1}\sum_{i=1}^{n}\left\{\begin{matrix} I(A_i>0)\left[\left(A_i-\exp\left(W_i a_2\right)\right)-\exp\left(W_i a_2\right)\right] \\ \exp\left(W_i a_2\right)W_i'W_i \end{matrix}\right\}\right]^{-1}$$

$$= n_1\left[2\sum_{i=1}^{n}\left\{\begin{matrix} I(A_i>0)\left[\left(A_i-\exp\left(W_i a_2\right)\right)-\exp\left(W_i a_2\right)\right] \\ \exp\left(W_i a_2\right)W_i'W_i \end{matrix}\right\}\right]^{-1}.$$

(G-19)

where $\hat{a}_2$ is the first stage, second part, estimator of a_2 , and n_1 is the size of the subsample for whom $I(A > 0) = 1$

so

$$\hat{E}[\nabla_{\delta\delta} q_1]^{-1} = \begin{bmatrix} \hat{E}[\nabla_{\pi\pi} q_{11}]^{-1} & 0 & 0 \\ 0 & \hat{E}[\nabla_{a_1a_1} q_{12}]^{-1} & 0 \\ 0 & 0 & \hat{E}[\nabla_{a_2a_2} q_{13}]^{-1} \end{bmatrix}. \quad \text{(G-20)}$$

$\mathbf{E}\left[\nabla_{\delta}\mathbf{q_1}'\nabla_{\delta}\mathbf{q_1}\right]$
3K×3K

Written out explicitly we have

$$E\left[\nabla_{\delta} q_1'\nabla_{\delta} q_1\right] = \begin{bmatrix} E[\nabla_{\pi} q_{11}'\nabla_{\pi} q_{11}] & E[\nabla_{\pi} q_{11}'\nabla_{a_1} q_{12}] & E[\nabla_{\pi} q_{11}'\nabla_{a_2} q_{13}] \\ E[\nabla_{a_1} q_{12}'\nabla_{\pi} q_{11}] & E[\nabla_{a_1} q_{12}'\nabla_{a_1} q_{12}] & E[\nabla_{a_1} q_{12}'\nabla_{a_2} q_{13}] \\ E[\nabla_{a_2} q_{13}'\nabla_{\pi} q_{11}] & E[\nabla_{a_2} q_{13}'\nabla_{a_1} q_{12}] & E[\nabla_{a_2} q_{13}'\nabla_{a_2} q_{13}] \end{bmatrix}.$$

(G-21)

Because the first stage, first part, estimator of a_1 is MLE we can write

$$E[\nabla_{a_1} q_{12}' \nabla_{a_1} q_{12}] = -E[\nabla_{a_1 a_1} q_{12}] = \left[\mathrm{AVAR}(\hat{a}_1)\right]^{-1}$$

where $\mathrm{AVAR}(\hat{a}_1)$ is defined as in (G-16). We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{a_1} q_{12}' \nabla_{a_1} q_{12}]$ is

$$\hat{E}[\nabla_{a_1} q_{12}' \nabla_{a_1} q_{12}] = \frac{1}{n}\left[\widehat{\mathrm{AVAR}}^*(\hat{a}_1)\right]^{-1} \tag{G-22}$$

where $\widehat{\mathrm{AVAR}}^*(\hat{a}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure. The remaining block elements follow from

$$\nabla_{\pi} q_{11} = 2(\ln(\bar{A}) - \bar{W}\pi)\bar{W}. \tag{G-23}$$

$$\nabla_{a_1} q_{12} = \nabla_{a_1} \ln f(I(A>0) \mid W) = \left[I(A>0)[1-\Lambda(Wa_1)] - [1-I(A>0)]\Lambda(Wa_1)\right]W \tag{G-24}$$

and

$$\nabla_{a_2} q_{13} = 2\,I(A>0)\left(A - \exp\left(Wa_2\right)\right)\exp\left(Wa_2\right)W. \tag{G-25}$$

where the formulation of $\nabla_{a_1} q_{12}$ comes from equation (16.4.8) on p. 350 of Fomby et al. (1984) and $\Lambda(\ )$ denotes the logistic cdf. The remaining required consistent matrix estimators are

$$\hat{E}[\nabla_\pi q_{11} '\nabla_\pi q_{11}] = \frac{1}{n_2}\sum_{j=1}^{n_2} \nabla_\pi \hat{q}_{11j} '\nabla_\pi \hat{q}_{11j} \tag{G-26}$$

$$\hat{E}[\nabla_\pi q^*_{11} '\nabla_{a_1} q_{12}] = \frac{1}{n}\sum_{i=1}^{n} \nabla_\pi \hat{q}^*_{11i} ' \nabla_{a_1} \hat{q}_{12i} \tag{G-27}$$

$$\hat{E}[\nabla_\pi q^*_{11} '\nabla_{a_2} q_{13}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_\pi \hat{q}^*_{11i} '\nabla_{a_2} \hat{q}_{13i} \tag{G-28}$$

$$\hat{E}[\nabla_{a_1} q_{12} '\nabla_{a_2} q_{13}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{a_1} \hat{q}_{12i} '\nabla_{a_2} \hat{q}_{13i} \tag{G-29}$$

$$\hat{E}[\nabla_{a_2} q_{13} '\nabla_{a_2} q_{13}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{a_2} \hat{q}_{13i} '\nabla_{a_2} \hat{q}_{13i} \tag{G-30}$$

where

$$\nabla_\pi \hat{q}_{11i} = 2(\ln(\bar{A}_i) - \bar{W}_i \hat{\pi})\bar{W}_i \tag{G-31}$$

$$\nabla_{a_1} \hat{q}_{12i} = \left[I(A_i > 0)[1 - \Lambda(W_i \hat{a}_1)] - [1 - I(A_i > 0)]\Lambda(W_i \hat{a}_1)\right] W_i \tag{G-32}$$

and

$$\nabla_{a_2} \hat{q}_{13i} = 2\, I(A_i > 0)\left(A_i - \exp\left(W_i \hat{a}_2\right)\right)\exp\left(W_i \hat{a}_2\right) W_i \tag{G-33}$$

The components of equations (G-27) and (G-28) are obtained using

$$\nabla_\pi q^*_{11i} = 2(\ln(\bar{A}_i) - \bar{W}^*_i \pi)\bar{W}^*_i .$$

where

$\bar{W}^*_i = [\bar{P}_i \quad \bar{X}_i]$ denotes the observation on $\bar{W} = [\bar{P} \quad \bar{X}]$ that pertains to the ith individual in the sample (i = 1, ..., n) for the relevant aggregation unit (j = 1, ..., n_2)

so

$$\hat{E}\left[\nabla_{\delta} q_1 \, '\nabla_{\delta} q_1\right] = \begin{bmatrix} \hat{E}[\nabla_{\pi} q_{11} \, '\nabla_{\pi} q_{11}] & \hat{E}[\nabla_{\pi} q_{11} \, '\nabla_{a_1} q_{12}] & \hat{E}[\nabla_{\pi} q_{11} \, '\nabla_{a_2} q_{13}] \\ \hat{E}[\nabla_{a_1} q_{12} \, '\nabla_{\pi} q_{11}] & \hat{E}[\nabla_{a_1} q_{12} \, '\nabla_{a_1} q_{12}] & \hat{E}[\nabla_{a_1} q_{12} \, '\nabla_{a_2} q_{13}] \\ \hat{E}[\nabla_{a_2} q_{13} \, '\nabla_{\pi} q_{11}] & \hat{E}[\nabla_{a_2} q_{13} \, '\nabla_{a_1} q_{12}] & \hat{E}[\nabla_{a_2} q_{13} \, '\nabla_{a_2} q_{13}] \end{bmatrix}. \tag{G-34}$$

$\underset{3\times 3K}{\mathbf{E}\left[\nabla_{\gamma\delta}\mathbf{q}\right]}$

Written out explicitly we have

$$\nabla_{\gamma} q = [(\nabla_u q_a + 0 + 0) \quad (0 + \nabla_v q_b + 0) \quad (0 + 0 + \nabla_{m_{\mathcal{P}}} q_c)]$$

$$= [\nabla_u q_a \quad \nabla_v q_b \quad \nabla_{m_{\mathcal{P}}} q_c]$$

$$= 2\left[\left(U(a, W) - u\right) \quad \left(V(a, W) - v\right) \quad (\mathcal{P}_i - m_{\mathcal{P}})\right] \tag{G-35}$$

and

$$E[\nabla_{\gamma\delta} q] = \begin{bmatrix} E[\nabla_{u\pi} q_a] & E[\nabla_{u\alpha_1} q_a] & E[\nabla_{u\alpha_2} q_a] \\ E[\nabla_{v\pi} q_b] & E[\nabla_{v\alpha_1} q_b] & E[\nabla_{v\alpha_2} q_b] \\ E[\nabla_{m_{\mathcal{P}}\pi} q_c] & E[\nabla_{m_{\mathcal{P}}\alpha_1} q_c] & E[\nabla_{m_{\mathcal{P}}\alpha_2} q_c] \end{bmatrix}$$

$$= 2\begin{bmatrix} E[\nabla_{\pi} U] & E[\nabla_{a_1} U] & E[\nabla_{a_2} U] \\ E[\nabla_{\pi} V] & E[\nabla_{a_1} V] & E[\nabla_{a_2} V] \\ 0 & 0 & 0 \end{bmatrix} \tag{G-36}$$

where

$$\nabla_{\pi} U = \nabla_{\pi} V = 0. \tag{G-37}$$

$$\nabla_{a_1} U = 2\,\Lambda(Wa_1)\lambda(Wa_1)\exp(Wa_2)\,W \tag{G-38}$$

$$\nabla_{a_2} U = \Lambda(Wa_1)^2 \exp(Wa_2)\,W \quad . \tag{G-39}$$

$$\nabla_{a_1} V = \lambda(Wa_1)\exp(Wa_2)\,W \tag{G-40}$$

and

$$\nabla_{a_2} V = \Lambda(Wa_1)\exp(Wa_2)\,W\,. \tag{G-41}$$

Note that

$$\nabla_a \Lambda(a) = \lambda(a) = \Lambda(a)[1 - \Lambda(a)]$$

$$\nabla_a \lambda(a) = \lambda(a)[1 - 2\Lambda(a)].$$

The requisite consistent matrix estimators are

$$\hat{E}[\nabla_\pi U] = \hat{E}[\nabla_\pi V] = 0 \tag{G-42}$$

$$\hat{E}[\nabla_{a_1} U] = \sum_{i=1}^{n} \widehat{\nabla_{a_1} U}_i \tag{G-43}$$

$$\hat{E}[\nabla_{a_2} U] = \sum_{i=1}^{n} \widehat{\nabla_{a_2} U}_i \tag{G-44}$$

$$\hat{E}[\nabla_{a_1} V] = \sum_{i=1}^{n} \widehat{\nabla_{a_1} V}_i \tag{G-45}$$

and

$$\hat{E}[\nabla_{a_2} V] = \sum_{i=1}^{n} \widehat{\nabla_{a_2} V}_i \tag{G-46}$$

where

$$\widehat{\nabla_\pi U}_i = \widehat{\nabla_\pi V}_i = 0 \tag{G-47}$$

$$\widehat{\nabla_{a_1} U}_i = 2\,\Lambda(W_i\,\hat{a}_1)\lambda(W_i\,\hat{a}_1)\exp(W_i\,\hat{a}_2)\,W_i \tag{G-48}$$

$$\widehat{\nabla_{a_2} U}_i = \Lambda(W_i\,\hat{a}_1)^2\exp(W_i\,\hat{a}_2)\,W \tag{G-49}$$

$$\widehat{\nabla_{a_1} V}_i = \lambda(W_i\,\hat{a}_1)\exp(W_i\,\hat{a}_2)\,W_i \tag{G-50}$$

and

$$\widehat{\nabla_{a_2} V_i} = \Lambda(W_i \hat{a}_1) \exp(W_i \hat{a}_2) W_i \tag{G-51}$$

so

$$\hat{E}[\nabla_{\gamma\delta} q] = 2 \begin{bmatrix} 0 & \hat{E}[\nabla_{a_1} U] & \hat{E}[\nabla_{a_2} U] \\ 0 & \hat{E}[\nabla_{a_1} V] & \hat{E}[\nabla_{a_2} V] \\ 0 & 0 & 0 \end{bmatrix}. \tag{G-52}$$

$E[\nabla_{\gamma\gamma} q]^{-1}$

3×3

$$E\left[\nabla_{\gamma\gamma} q\right]^{-1} = \begin{bmatrix} E\left[\nabla_{uu} q_a\right]^{-1} & 0 & 0 \\ 0 & E\left[\nabla_{vv} q_b\right]^{-1} & 0 \\ 0 & 0 & E\left[\nabla_{m_p m_p} q_c\right]^{-1} \end{bmatrix} \tag{G-53}$$

because $\nabla_{uv} q_a = \nabla_{vu} q_b = \nabla_{um_p} q_a = \nabla_{vm_p} q_b = \nabla_{m_p u} q_c = \nabla_{m_p v} q_c = 0$. Now

$$\nabla_{uu} q_a = \nabla_{vv} q_b = \nabla_{m_p m_p} q_c = -2$$

therefore

$$E\left[\nabla_{\gamma\gamma} q\right]^{-1} = \begin{bmatrix} -\frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & -\frac{1}{2} \end{bmatrix}. \tag{G-54}$$

$E\left[\nabla_\gamma q' \nabla_\gamma q\right]$

3×3

Given that

$$\nabla_\gamma q = 2\left[\left(U(a, W) - u\right) \quad \left(V(a, W) - v\right) \quad (\mathcal{P}_i - m_{\mathcal{P}})\right]. \tag{G-55}$$

we have

$$E\left[\nabla_\gamma q' \nabla_\gamma q\right] = 4\begin{bmatrix} E[(U(a, W) - u)^2] & E[(U(a, W) - u)(V(a, S) - v)] & E[(U(a, W) - u)(\mathcal{P} - m_{\mathcal{P}})] \\ E[(U(a, W) - u)(V(a, S) - v)] & E[(V(a, S) - v)^2] & E[(V(a, S) - v)(\mathcal{P} - m_{\mathcal{P}})] \\ E[(U(a, W) - u)(\mathcal{P} - m_{\mathcal{P}})] & E[(V(a, S) - v)(\mathcal{P} - m_{\mathcal{P}})] & E[(\mathcal{P} - m_{\mathcal{P}})^2] \end{bmatrix}. \tag{G-56}$$

The corresponding consistent estimator is

$$\hat{E}\left[\nabla_\gamma q' \nabla_\gamma q\right] = 4\begin{bmatrix} \sum_{i=1}^{n} \frac{1}{n}(\hat{U}_i - \hat{u})^2 & \sum_{i=1}^{n} \frac{1}{n}\begin{pmatrix}(\hat{U}_i - \hat{u}) \\ (\hat{V}_i - \hat{v})\end{pmatrix} & \sum_{i=1}^{n} \frac{1}{n}\begin{pmatrix}(\hat{U}_i - \hat{u}) \\ (\mathcal{P}_i - \hat{m}_{\mathcal{P}})\end{pmatrix} \\ \sum_{i=1}^{n} \frac{1}{n}\begin{pmatrix}(\hat{U}_i - \hat{u}) \\ (\hat{V}_i - \hat{v})\end{pmatrix} & \sum_{i=1}^{n} \frac{1}{n}(\hat{V}_i - \hat{v})^2 & \sum_{i=1}^{n} \frac{1}{n}\begin{pmatrix}(\hat{V}_i - \hat{v}) \\ (\mathcal{P}_i - \hat{m}_{\mathcal{P}})\end{pmatrix} \\ \sum_{i=1}^{n} \frac{1}{n}\begin{pmatrix}(\hat{U}_i - \hat{u}) \\ (\mathcal{P}_i - \hat{m}_{\mathcal{P}})\end{pmatrix} & \sum_{i=1}^{n} \frac{1}{n}\begin{pmatrix}(\hat{V}_i - \hat{v}) \\ (\mathcal{P}_i - \hat{m}_{\mathcal{P}})\end{pmatrix} & \sum_{i=1}^{n} \frac{1}{n}(\mathcal{P}_i - \hat{m}_{\mathcal{P}})^2 \end{bmatrix}. \tag{G-57}$$

Based on (G-8), (G-11) and (G-12) and using the two-stage estimator $\hat{\theta}$ we can consistently estimate (G-7) as

$$\widehat{AVAR(\hat{\theta})} = \begin{bmatrix} \hat{D}_{11} & \hat{D}_{12} \\ \hat{D}_{12}' & \hat{D}_{22} \end{bmatrix}$$

where

$$\hat{D}_{11} = \widehat{AVAR}(\hat{\delta}) = \begin{bmatrix} n_2 \widehat{AVAR}*(\hat{\pi}) & 0 & 0 \\ 0 & n\widehat{AVAR}*(\hat{a}_1) & 0 \\ 0 & 0 & n_1 \widehat{AVAR}*(\hat{a}_2) \end{bmatrix}$$

$$D_{12} = -\hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\delta} q_1 ' \nabla_{\delta} q_1\right] \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma\delta} q\right]' \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

$$D_{22} = \widehat{AVAR(\hat{\gamma})} = \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma\delta} q\right] \widehat{AVAR(\hat{\delta})}\, \hat{E}\left[\nabla_{\gamma\delta} q\right]' \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

$$+\ \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma} q ' \nabla_{\gamma} q\right] \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

and using well known results from asymptotic theory for two-stage estimators, we can show that[48]

$$\widehat{AVAR(\hat{\theta})}^{-\frac{1}{2}} \begin{bmatrix} \sqrt{n_2}(\hat{\pi} - \pi) \\ \sqrt{n}(\hat{a}_1 - a_1) \\ \sqrt{n_1}(\hat{a}_2 - a_2) \\ \sqrt{n}(\hat{u} - u) \\ \sqrt{n}(\hat{v} - v) \\ \sqrt{n}(\hat{m}_{\mathcal{P}} - m_{\mathcal{P}}) \end{bmatrix} \xrightarrow{d} N(0, I)\,. \qquad \text{(G-58)}$$

****

ASIDE:

Notice that the "$\sqrt{n}$ blow up" is a bit tricky here. It implements $\sqrt{n}$ for $\hat{a}_1$, $\hat{u}$, $\hat{v}$ and $\hat{m}_{\mathcal{P}}$; but uses $\sqrt{n_1}$ for $\hat{a}_2$ and $\sqrt{n_2}$ for $\hat{\pi}$. We had to do this because we had to use the correct sample sizes (viz., n_2 and n_1) for a number of the components of $\widehat{AVAR(\hat{\theta})}$ [viz., those that pertained to the estimation of $\hat{\pi}$, and $\hat{a}_2$, respectively]; in particular (G-15), (G-19) (G-26) through (G-30). For this reason we had to be explicit about the denominators in all of the averages for the components of $\widehat{AVAR(\hat{\theta})}$. This meant that in

---

[48]Discussions of asymptotic theory for two-stage optimization estimators can be found in Newey & McFadden (1994), White (1994, Chapter 6), and Wooldridge (2010, Chapter 12).

the construction of the requisite asymptotic t-stats we had to explicitly include the "blow-up" in the numerator (i.e., we had to multiply by the square-root of the appropriate sample size). I refer to this as "tricky" because one typically does not have to do this. In the usual asymptotic t-stat construction the denominators of the averages ("n") need not be included in the construction of the asymptotic covariance matrix because it typically manifests as a multiplicative factor and, after pulling the diagonal and taking the square root to get the standard errors, this multiplicative $\sqrt{n}$ cancels with the "blow-up" factor in the numerator. For example, the asymptotic t-stat of the OLS estimator is

$$\frac{\sqrt{n}(\hat{\rho}_k - \rho_k)}{\sqrt{\widehat{AVAR(\hat{\rho})}}} = \frac{\sqrt{n}(\hat{\rho}_k - \rho_k)}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n}\mathcal{X}'\mathcal{X}_{kk}\right)^{-1}}} = \frac{\sqrt{n}(\hat{\rho}_k - \rho_k)}{\sqrt{n}\sqrt{\hat{\sigma}^2(\mathcal{X}'\mathcal{X}_{kk})^{-1}}} = \frac{(\hat{\rho}_k - \rho_k)}{\sqrt{\hat{\sigma}^2(\mathcal{X}'\mathcal{X}_{kk})^{-1}}}$$

where

n is the sample size

ρ_k is the coefficient of the kth regressor in the linear regression

$\hat{\rho}_k$ is its OLS estimator

σ^2 is the regression error variance estimator

$\mathcal{X}$ is the matrix of regressors

and $\mathcal{X}'\mathcal{X}_{kk}$ is the kth diagonal element of $\mathcal{X}'\mathcal{X}$. Note how the "$\sqrt{n}$ s" simply cancel. Note also that what we typically refer to as the "asymptotic standard error" can actually be written as the square root of the diagonal element of the consistent estimator of the asymptotic covariance matrix divided by n; in other words

$$\text{asy std err} = \sqrt{\frac{\widehat{\text{AVAR}(\hat{\rho})}}{n}}.$$

****

Now back to the issue at hand. Moreover

$$\widehat{\text{AVAR}(\hat{\tau})}^{-\frac{1}{2}} \begin{bmatrix} \sqrt{n_2}(\hat{\pi}_P - \pi_P) \\ \sqrt{n}(\hat{a}_{\mathcal{P}1} - a_{\mathcal{P}1}) \\ \sqrt{n_1}(\hat{a}_{\mathcal{P}2} - a_{\mathcal{P}2}) \\ \sqrt{n}(\hat{u} - u) \\ \sqrt{n}(\hat{v} - v) \\ \sqrt{n}(\hat{m}_{\mathcal{P}} - m_{\mathcal{P}}) \end{bmatrix} \xrightarrow{d} N(0, I). \tag{G-59}$$

where

$$\widehat{\text{AVAR}(\hat{\tau})} = \Xi \widehat{\text{AVAR}(\hat{\theta})} \Xi' \tag{G-60}$$

and τ and Ξ are defined as in (G-2). Now combining (G-1) with (G-59) and (G-60) we get

$$\widehat{\text{avar}(\hat{\eta}^{AL} - \hat{\eta}^{UPO})}^{-\frac{1}{2}} \sqrt{n} \left[\left(\hat{\eta}^{AL} - \hat{\eta}^{UPO} \right) - \left(\eta^{AL} - \eta^{UPO} \right) \right] \xrightarrow{d} N(0, I) \tag{G-61}$$

where

$$\widehat{\text{avar}(\hat{\eta}^{AL} - \hat{\eta}^{UPO})} = c(\hat{\tau}) \, \widehat{\text{AVAR}(\hat{\tau})} \, c(\hat{\tau})'$$

and

$$c(\hat{\tau}) = \left[1 \quad -\hat{m}_{\mathcal{P}} \left(1 - \frac{\hat{u}}{\hat{v}} \right) \quad -\hat{m}_{\mathcal{P}} \quad \hat{m}_{\mathcal{P}} \hat{a}_{\mathcal{P}1} / \hat{v} \quad -\hat{m}_{\mathcal{P}} \hat{a}_{\mathcal{P}1} \hat{u} / \hat{v}^2 \quad -\left[\left(1 - \frac{\hat{u}}{\hat{v}} \right) \hat{a}_{\mathcal{P}1} + \hat{a}_{\mathcal{P}2} \right] \right].$$

Appendix H.

Asymptotic Distribution (and Standard Error) of $\hat{\eta}^{UPOL}$ in Eqn (4-6)

We may write $\hat{\eta}^{UPOL}$ as

$$\hat{\eta}^{UPOL} = \frac{\hat{\kappa}}{\hat{v}}$$

where

$$\hat{\kappa} = \sum_{i=1}^{n} \frac{1}{n} \mathcal{K}(\hat{\alpha}, \vec{W}_i)$$

$$\hat{v} = \sum_{i=1}^{n} \frac{1}{n} \mathcal{V}(\hat{\alpha}, \vec{W}_i)$$

Using the corresponding consistent estimators for k, and v say

$$\mathcal{K}(\alpha, \vec{W}) = \lambda(\vec{W}\alpha_1)\exp(\vec{W}\alpha_2)\alpha_{P1} + \Lambda(\vec{W}\alpha_1)\exp(\vec{W}\alpha_2)\alpha_{P2}$$

$$\mathcal{V}(\alpha, \vec{W}) = \Lambda(\vec{W}\alpha_1)\exp(\vec{W}\alpha_2)$$

P is the logged prices of alcohol

$\hat{\alpha} = [\hat{\alpha}_1' \quad \hat{\alpha}_2']'$ (with $\hat{\alpha}_1' = [\hat{\alpha}_{P1} \quad \hat{\alpha}_{X1}']$ and $\hat{\alpha}_2' = [\hat{\alpha}_{P2} \quad \hat{\alpha}_{X2}']$) is the consistent estimate of the parameter vector $\alpha = [\alpha_1' \quad \alpha_2']'$ (with $\alpha_1' = [\alpha_{P1} \quad \alpha_{X1}']$ and $\alpha_2' = [\alpha_{P2} \quad \alpha_{X2}']$)) [the parameters of equation (4-6)] obtained via the two-part protocol culminating in (4-4) using logged prices of alcohol

and $\vec{W}_i = [P_i \quad X_i]$ denotes the observation on $\vec{W} = [P \quad X]$ for the ith individual in the sample

(i = 1, ..., n). Let $\hat{\gamma} = [\hat{\kappa} \quad \hat{\nu}]'$ and $\gamma = [\kappa \quad \nu]'$, where $\text{plim}[\hat{\gamma}] = \gamma$. If we could show that

$$\text{AVAR}(\hat{\gamma})^{-\frac{1}{2}}\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, I)$$

where the formulation of $\text{AVAR}(\hat{\gamma})$ is known, then we could apply the δ-method to obtain the asymptotic variance of $\hat{\eta}^{UPOL}$ as

$$\text{avar}(\hat{\eta}^{UPOL}) = c(\gamma)\ \text{AVAR}(\hat{\gamma})\ c(\gamma)'$$

where $c(\gamma) = [1/\nu \quad -\kappa/\nu^2]$. Moreover, if we have a consistent estimator for $\text{AVAR}(\hat{\gamma})$, say $\widehat{\text{AVAR}(\hat{\gamma})}$ [i.e. $\text{plim}\left[\widehat{\text{AVAR}(\hat{\gamma})}\right] = \text{AVAR}(\hat{\gamma})$], then we could consistently estimate $\text{avar}(\hat{\eta}^{UPOL})$ as

$$\widehat{\text{avar}(\hat{\eta}^{UPOL})} = c(\hat{\gamma})\ \widehat{\text{AVAR}(\hat{\gamma})}\ c(\hat{\gamma})'. \qquad \text{(H-1)}$$

We focus, therefore, on finding the asymptotic distribution of $\hat{\gamma}$ and, in particular, the formulation of its asymptotic covariance matrix.

First, let $\theta' = [\delta' \quad \gamma']$ where $\delta = [\alpha_1' \quad \alpha_2']'$, $\gamma' = [\kappa \quad \nu]$ (recall, $\alpha_1' = [\alpha_{P1} \quad \alpha_{X1}']$ and $\alpha_2' = [\alpha_{P2} \quad \alpha_{X2}']$), and note that $\hat{\gamma}$ can be viewed as the second stage estimator in the following two-stage protocol

First Stage

Consistently estimate δ via the following optimization estimator

$$\tilde{\delta} = \arg\max_{\tilde{\delta}} \frac{\sum_{i=1}^{n} q_1(\tilde{\delta}, S_i)}{n} \tag{H-2}$$

where

$$q_1(\tilde{\delta}, S_i) = q_{11}(\tilde{\alpha}_1, S_i) + q_{12}(\tilde{\alpha}_2, S_i)$$

$$q_{11}(\tilde{\alpha}_1, S_i) = I(A_i > 0)\ln[\Lambda(\vec{W}_i\,\tilde{\alpha}_1)] + [1 - I(A_i > 0)]\ln[1 - \Lambda(\vec{W}_i\,\tilde{\alpha}_1)]$$

$$q_{12}(\tilde{\alpha}_2, S_i) = -I(A_i > 0)\left(A_i - \exp\left(\vec{W}_i\,\tilde{\alpha}_2\right)\right)^2$$

$$S_i = [A_i \quad X_i]$$

$\tilde{\delta} = [\tilde{\alpha}_1' \quad \tilde{\alpha}_2']'$, $\tilde{\rho}_1' = [\tilde{\rho}_{P1} \quad \tilde{\rho}_{X1}']$ and $\tilde{\rho}_2' = [\tilde{\rho}_{P2} \quad \tilde{\rho}_{X2}']$ and $\hat{\tilde{\delta}} = [\hat{\tilde{\rho}}_1' \quad \hat{\tilde{\rho}}_2']'$

Second Stage

Consistently estimate γ via the following optimization estimator

$$\hat{\gamma} = \arg\max_{\tilde{\gamma}} \frac{\sum_{i=1}^{n} q(\hat{\delta}, \tilde{\gamma}, S_i)}{n} \tag{H-3}$$

where

$$q(\hat{\delta}, \tilde{\gamma}, S_i) = q_a(\hat{\delta}, \tilde{\kappa}, S_i) + q_b(\hat{\delta}, \tilde{\nu}, S_i)$$

$$q_a(\hat{\delta}, \tilde{\kappa}, S_i) = -(\mathcal{K}(\hat{\alpha}, \vec{W}_i) - \tilde{\kappa})^2$$

$$q_b(\hat{\delta}, \tilde{\nu}, S_i) = -(\mathcal{V}(\hat{\alpha}, \vec{W}_i) - \tilde{\nu})^2$$

$\hat{\delta}$ is the first stage estimator of δ. Use q_1 as shorthand notation for

$$q_1(\delta, S) = q_{11}(\alpha_1, S) + q_{12}(\alpha_2, S)$$

with

$$q_{11}(\alpha_1, S) = I(A > 0)\ln[\Lambda(\vec{W}\,\alpha_1)] + [1 - I(A > 0)]\ln[1 - \Lambda(\vec{W}\,\alpha_1)]$$

$$q_{12}(\alpha_2, S) = -I(A > 0)\left(A - \exp\left(W\,\alpha_2\right)\right)^2$$

$$S = [A \quad X]$$

and use q as shorthand notation for

$$q(\delta, \tilde{\gamma}, S) = q_a(\delta, \kappa, S) + q_b(\delta, \nu, S)$$

with

$$q_a(\delta, \kappa, S) = -(\mathcal{K}(\alpha, \vec{W}) - \kappa)^2$$

$$q_b(\delta, \nu, S) = -(\mathcal{V}(\alpha, \vec{W}) - \nu)^2$$

and let $AVAR(\hat{\delta})$ denote the asymptotic covariance matrix of the first stage estimator.[49] Using well known results from asymptotic theory for two-stage estimators, we can show that[50]

$$AVAR(\hat{\gamma})^{-\frac{1}{2}}\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, I) \qquad \text{(H-4)}$$

where, $plim(\hat{\gamma}) = \gamma$

---

[49] Note that a consistent estimate $AVAR(\hat{\rho})$ can be obtained from the packaged output for the first stage estimator because the first stage estimator is unaffected by the fact that it is a component of a two-stage estimator.

[50] Discussions of asymptotic theory for two-stage optimization estimators can be found in Newey & McFadden (1994), White (1994, Chapter 6), and Wooldridge (2010, Chapter 12).

$$\text{AVAR}(\hat{\gamma}) = \mathrm{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \Bigg[\mathrm{E}\left[\nabla_{\gamma\delta} q\right] \text{AVAR}(\hat{\alpha}) \mathrm{E}\left[\nabla_{\gamma\delta} q\right]'$$

$$- \mathrm{E}\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right] \mathrm{E}\left[\nabla_{\delta\delta} q_1\right]^{-1} \mathrm{E}\left[\nabla_{\gamma\delta} q\right]'$$

$$- \mathrm{E}\left[\nabla_{\gamma\delta} q\right] \mathrm{E}\left[\nabla_{\gamma\delta} q_1\right]^{-1} \mathrm{E}\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right]' \Bigg] \mathrm{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

$$+ \mathrm{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \mathrm{E}\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right] \mathrm{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \qquad \text{(H-5)}$$

Fortunately, (H-5) can be simplified in a number of ways. Note that we can write

$$\mathrm{E}\left[\nabla_{\gamma} q' \nabla_{\delta} q_1\right] = \mathrm{E}\left[\nabla_{\gamma} q' \mathrm{E}[\nabla_{\delta} q_1 \mid \vec{W}]\right]$$

but

$$\nabla_{\delta} q_1 = \nabla_{\delta} q_{11} + \nabla_{\delta} q_{12}$$

with

$$\nabla_{\delta} q_{11} = [\nabla_{\alpha_1} \ln f(\mathrm{I}(A > 0) \mid \vec{W}) \quad 0]$$

$$\nabla_{\delta} q_{12} = \left[0 \qquad 2\,\mathrm{I}(A > 0)\left(A - \exp\left(\vec{W}\alpha_2\right)\right)\exp\left(\vec{W}\alpha_2\right)\vec{W}\right]$$

where

$$f(\mathrm{I}(A > 0) \mid \vec{W}) = \Lambda(\vec{W}\alpha_1)^{\mathrm{I}(A>0)} [1 - \Lambda(\vec{W}\alpha_1)]^{[1 - \mathrm{I}(A>0)]} .$$

Therefore

$$\mathrm{E}[\nabla_{\delta} q_1 \mid \vec{W}] = \begin{bmatrix} \mathrm{E}[\nabla_{\alpha_1} \ln f(\mathrm{I}(A > 0) \mid \vec{W})] \\ 2\mathrm{E}\left[\mathrm{I}(A > 0)\left(A - \exp\left(\vec{W}\alpha_2\right)\right)\exp\left(\vec{W}\alpha_2\right) \mid \vec{W}\right]\vec{W} \end{bmatrix}'$$

$$= [0 \quad 0]$$

because $E[\nabla_{\alpha_1} \ln f(I(A>0) \mid \vec{W})] = 0$ [see (13.20) on p. 477 of Wooldridge (2010)] and,

$E\left[I(A>0)\left(A - \exp\left(\vec{W}\alpha_2\right)\right)\exp\left(\vec{W}\alpha_2\right) \mid \vec{W} \right] = 0$ by design. Finally, then we get

$$E\left[\nabla_\gamma q' \nabla_\delta q_1\right] = 0$$

so

$$\underset{3\times3}{AVAR(\hat{\gamma})} = \underset{3\times3}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \underset{3\times2K}{E\left[\nabla_{\gamma\delta} q\right]} \underset{2K\times2K}{AVAR(\hat{\alpha})} \underset{2K\times3}{E\left[\nabla_{\gamma\delta} q\right]'} \underset{3\times3}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}}$$

$$+ \underset{3\times3}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \underset{3\times3}{E\left[\nabla_\gamma q' \nabla_\gamma q\right]} \underset{3\times3}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}}. \qquad \text{(H-6)}$$

Let's consider each of the individual components of (H-6) in turn.

$E[\nabla_{\delta\delta} q_1]^{-1}$
2K×2K

Written out explicitly we have

$$E[\nabla_{\alpha\alpha} q_1]^{-1} = \begin{bmatrix} E[\nabla_{\alpha_1\alpha_1} q_{11}]^{-1} & 0 \\ 0 & E[\nabla_{\alpha_2\alpha_2} q_{12}]^{-1} \end{bmatrix}. \qquad \text{(H-7)}$$

Now

$$E[\nabla_{\alpha_1\alpha_1} q_{11}]^{-1} = -AVAR(\hat{\alpha}_1) \qquad \text{(H-8)}$$

= the negative of the asymptotic covariance matrix for first stage, first part, logit estimation in the two-stage estimation protocol for θ. We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{\alpha_1\alpha_1} q_{11}]^{-1}$ is

$$\hat{E}[\nabla_{\alpha_1\alpha_1} q_{11}]^{-1} = -n\widehat{AVAR} * (\hat{\alpha}_1) \tag{H-9}$$

where $\widehat{AVAR} * (\hat{\alpha}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure. Also

$$\nabla_{\alpha_2} q_{12} = 2\, I(A > 0)\left(A - \exp\left(\vec{W}\,\alpha_2\right)\right)\exp\left(\vec{W}\,\alpha_2\right)\vec{W}$$

and

$$\nabla_{\alpha_2\alpha_2} q_{12} = 2\, I(A > 0)\left[\left(A - \exp\left(\vec{W}\,\alpha_2\right)\right) - \exp\left(\vec{W}\,\alpha_2\right)\right]\exp\left(\vec{W}\,\alpha_2\right)\vec{W}'\vec{W}\,.$$

Therefore

$$E[\nabla_{\alpha_2\alpha_2} q_{12}] = 2E\left[\, I(A > 0)\left[\left(A - \exp\left(\vec{W}\,\alpha_2\right)\right) - \exp\left(\vec{W}\,\alpha_2\right)\right]\exp\left(\vec{W}\,\alpha_2\right)\vec{W}'\vec{W}\,\right]. \tag{H-10}$$

A consistent estimator of $E[\nabla_{\alpha_2\alpha_2} q_{12}]^{-1}$ is

$$\hat{E}[\nabla_{\alpha_2\alpha_2} q_{12}]^{-1} = n_1\left[\, 2\sum_{i=1}^{n}\left\{ I(A_i > 0)\left[\left(A_i - \exp\left(\vec{W}_i\,\alpha_2\right)\right) - \exp\left(\vec{W}_i\,\alpha_2\right)\right]\exp\left(\vec{W}_i\,\alpha_2\right)\vec{W}_i'\vec{W}_i \right\}\right]^{-1} \tag{H-11}$$

where n_1 is the size of the subsample for whom $I(A > 0) = 1$, so

$$\hat{E}[\nabla_{\delta\delta} q_1]^{-1} = \hat{E}[\nabla_{\alpha\alpha} q_1]^{-1} = \begin{bmatrix} \hat{E}[\nabla_{\alpha_1\alpha_1} q_{11}]^{-1} & 0 \\ 0 & \hat{E}[\nabla_{\alpha_2\alpha_2} q_{12}]^{-1} \end{bmatrix}. \qquad \text{(H-12)}$$

$\mathbf{E}\left[\nabla_{\delta} q_1 ' \nabla_{\delta} q_1\right]$
2K×2K

Written out explicitly we have

$$E\left[\nabla_{\alpha} q_1 ' \nabla_{\alpha} q_1\right] = \begin{bmatrix} E[\nabla_{\alpha_1} q_{11} ' \nabla_{\alpha_1} q_{11}] & E[\nabla_{\alpha_1} q_{11} ' \nabla_{\alpha_2} q_{12}] \\ E[\nabla_{\alpha_2} q_{12} ' \nabla_{\alpha_1} q_{11}] & E[\nabla_{\alpha_2} q_{12} ' \nabla_{\alpha_2} q_{12}] \end{bmatrix}. \qquad \text{(H-13)}$$

Because the first stage, first part, estimator of ρ_1 is MLE we can write

$$E[\nabla_{\alpha_1} q_{11} ' \nabla_{\alpha_1} q_{11}] = -E[\nabla_{\alpha_1\alpha_1} q_{11}] = \left[\text{AVAR}(\hat{\alpha}_1)\right]^{-1}$$

= the inverse of the asymptotic covariance matrix for first stage, first part, logit estimation in the two-stage estimation protocol for θ. We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{\alpha_1} q_{11} ' \nabla_{\alpha_1} q_{11}]$ is

$$\hat{E}[\nabla_{\alpha_1} q_{11} ' \nabla_{\alpha_1} q_{11}] = \frac{1}{n}\left[\widehat{\text{AVAR}} * (\hat{\alpha}_1)\right]^{-1} \qquad \text{(H-14)}$$

where $\widehat{\text{AVAR}} * (\hat{\alpha}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure. The remainder of the block elements follow from

$$\nabla_{\alpha_1} q_{11} = \nabla_{\alpha_1} \ln f(I(A>0) \mid \vec{W}) = \left[I(A>0)[1 - \Lambda(\vec{W}\alpha_1)] - [1 - I(A>0)]\Lambda(\vec{W}\alpha_1) \right] \vec{W} \tag{H-15}$$

$$\nabla_{\alpha_2} q_{12} = 2\,I(A>0)\left(A - \exp\left(\vec{W}\alpha_2\right)\right)\exp\left(\vec{W}\alpha_2\right)\vec{W}. \tag{H-16}$$

where the formulation of $\nabla_{\alpha_1} q_{11}$ comes from equation (16.4.8) on p. 350 of Fomby et al. (1984) and $\Lambda(\ \ )$ denotes the logistic cdf. The remaining required consistent matrix estimators are:

$$\hat{E}[\nabla_{\alpha_1} q_{11}\,'\nabla_{\alpha_2} q_{12}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{\alpha_1}\hat{q}_{11i}\,'\nabla_{\alpha_2}\hat{q}_{12i} \tag{H-17}$$

$$\hat{E}[\nabla_{\alpha_2} q_{12}\,'\nabla_{\alpha_2} q_{12}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{\alpha_2}\hat{q}_{12i}\,'\nabla_{\alpha_2}\hat{q}_{12i} \tag{H-18}$$

where

$$\nabla_{\alpha_1}\hat{q}_{11i} = \left[I(A_i>0)[1 - \Lambda(\vec{W}_i\hat{\alpha}_1)] - [1 - I(A_i>0)]\Lambda(\vec{W}_i\hat{\alpha}_1) \right]\vec{W}_i \tag{H-19}$$

and

$$\nabla_{\alpha_2}\hat{q}_{12i} = 2\,I(A_i>0)\left(A_i - \exp\left(\vec{W}_i\hat{\alpha}_2\right)\right)\exp\left(\vec{W}_i\hat{\alpha}_2\right)\vec{W}_i \tag{H-20}$$

so

$$\hat{E}\left[\nabla_{\ddot{\delta}} q_1\,'\nabla_{\ddot{\delta}} q_1\right] = \hat{E}\left[\nabla_{\rho} q_1\,'\nabla_{\rho} q_1\right] = \begin{bmatrix} \hat{E}[\nabla_{\rho_1} q_{11}\,'\nabla_{\rho_1} q_{11}] & \hat{E}[\nabla_{\rho_1} q_{11}\,'\nabla_{\rho_2} q_{12}] \\ \hat{E}[\nabla_{\rho_2} q_{12}\,'\nabla_{\rho_1} q_{11}] & \hat{E}[\nabla_{\rho_2} q_{12}\,'\nabla_{\rho_2} q_{12}] \end{bmatrix}. \tag{H-21}$$

$\mathbf{E}\left[\boldsymbol{\nabla}_{\gamma\delta}\mathbf{q}\right]$
2×2K

Written out explicitly we have

$$\nabla_\gamma q = [(\nabla_\kappa q_a + 0) \quad (0 + \nabla_\nu q_b)\ ]$$

$$= [\nabla_\kappa q_a \quad \nabla_\nu q_b\ ]$$

$$= 2\left[\left(\mathcal{K}(\alpha, \vec{W}) - \kappa\right) \quad \left(\mathcal{V}(\alpha, \vec{W}) - \nu\right)\ \right] \tag{H-22}$$

and

$$E[\nabla_{\gamma\delta} q] = \begin{bmatrix} E[\nabla_{\kappa\alpha_1} q_a] & E[\nabla_{\kappa\alpha_2} q_a] \\ E[\nabla_{\nu\alpha_1} q_b] & E[\nabla_{\nu\alpha_2} q_b] \end{bmatrix}$$

$$= 2\begin{bmatrix} E[\nabla_{\alpha_1}\mathcal{K}] & E[\nabla_{\alpha_2}\mathcal{K}] \\ E[\nabla_{\alpha_1}\mathcal{V}] & E[\nabla_{\alpha_2}\mathcal{V}] \end{bmatrix} \tag{H-23}$$

where

$$\nabla_{\alpha_1}\mathcal{K} = \{\lambda(\vec{W}\alpha_1)[1 - 2\Lambda(\vec{W}\alpha_1)]\alpha_{P1}\vec{W} + \lambda(\vec{W}\alpha_1)[1 \quad 0 \ldots 0]\}\exp(\vec{W}\alpha_2)$$
$$+\lambda(\vec{W}\alpha_1)\exp(\vec{W}\alpha_2)\alpha_{P2}\vec{W}$$
$$= \lambda(\vec{W}\alpha_1)\exp(\vec{W}\alpha_2)\left[\ \{[1 - 2\Lambda(\vec{W}\alpha_1)]\alpha_{P1} + \alpha_{P2}\}\vec{W} + [1 \quad 0 \ldots 0]\right] \tag{H-24}$$

$$\nabla_{\alpha_2}\mathcal{K} = \lambda(\vec{W}\alpha_1)\exp(\vec{W}\alpha_2)\alpha_{P1}\vec{W} + \Lambda(\vec{W}\alpha_1)\{\exp(\vec{W}\alpha_2)\alpha_{P2}\vec{W} + \exp(\vec{W}\alpha_2)[1 \quad 0 \ldots 0]\}$$

$$= \exp(\vec{W}\alpha_2)\left[\{\lambda(\vec{W}\alpha_1)\alpha_{P1} + \Lambda(\vec{W}\alpha_1)\alpha_{P2}\}\vec{W} + \Lambda(\vec{W}\alpha_1)[1 \quad 0 \ldots 0]\right].$$

(H-25)

$$\nabla_{\alpha_1}\mathcal{V} = \lambda(\vec{W}\alpha_1)\exp(\vec{W}\alpha_2)\vec{W} \tag{H-26}$$

and

$$\nabla_{\alpha_2}\mathcal{V} = \Lambda(\vec{W}\alpha_1)\exp(\vec{W}\alpha_2)\vec{W}\,. \tag{H-27}$$

The following equalities were used in deriving the above results

$$\nabla_a \Lambda(a) = \lambda(a) = \Lambda(a)[1 - \Lambda(a)]$$

$$\nabla_a \lambda(a) = \lambda(a)[1 - 2\Lambda(a)].$$

The requisite consistent matrix estimators are

$$\hat{E}[\nabla_{\alpha_1} \mathcal{K}] = \sum_{i=1}^{n} \widehat{\nabla_{\alpha_1} \mathcal{K}}_i \tag{H-28}$$

$$\hat{E}[\nabla_{\alpha_2} \mathcal{K}] = \sum_{i=1}^{n} \widehat{\nabla_{\alpha_2} \mathcal{K}}_i \tag{H-29}$$

$$\hat{E}[\nabla_{\alpha_1} \mathcal{V}] = \sum_{i=1}^{n} \widehat{\nabla_{\alpha_1} \mathcal{V}}_i \tag{H-30}$$

and

$$\hat{E}[\nabla_{\alpha_2} \mathcal{V}] = \sum_{i=1}^{n} \widehat{\nabla_{\alpha_2} \mathcal{V}}_i \tag{H-31}$$

where

$$\widehat{\nabla_{\alpha_1} \mathcal{K}}_i = \lambda(\vec{W}_i \hat{\alpha}_1) \exp(\vec{W}_i \hat{\alpha}_2) \left[\{[1 - 2\Lambda(\vec{W}_i \hat{\alpha}_1)] \hat{\alpha}_{P1} + \hat{\alpha}_{P2}\} \vec{W}_i + [1 \quad 0 \ldots 0] \right] \tag{H-32}$$

$$\widehat{\nabla_{\alpha_2} \mathcal{K}}_i = \exp(\vec{W}_i \hat{\alpha}_2) \left[\{\lambda(\vec{W}_i \hat{\alpha}_1) \hat{\alpha}_{P1} + \Lambda(\vec{W}_i \hat{\alpha}_1) \hat{\alpha}_{P2}\} \vec{W}_i + \Lambda(\vec{W}_i \hat{\alpha}_1)[1 \quad 0 \ldots 0] \right] \tag{H-33}$$

$$\widehat{\nabla_{\alpha_1} \mathcal{V}}_i = \lambda(\vec{W}_i \hat{\alpha}_1) \exp(\vec{W}_i \hat{\alpha}_2) \vec{W}_i \tag{H-34}$$

and

$$\widehat{\nabla_{\alpha_1} \mathcal{V}}_i = \Lambda(\vec{W}_i \hat{\alpha}_1) \exp(\vec{W}_i \hat{\alpha}_2) \vec{W}_i . \tag{H-35}$$

so

$$\hat{E}[\nabla_{\gamma\delta} q] = 2\begin{bmatrix} \hat{E}[\nabla_{\alpha_1}\mathcal{K}] & \hat{E}[\nabla_{\alpha_2}\mathcal{K}] \\ \hat{E}[\nabla_{\alpha_1}\mathcal{V}] & \hat{E}[\nabla_{\alpha_2}\mathcal{V}] \end{bmatrix}. \tag{H-36}$$

$\mathbf{E[\nabla_{\gamma\gamma} q]^{-1}}$
2×2

$$E\left[\nabla_{\gamma\gamma} q\right]^{-1} = \begin{bmatrix} E\left[\nabla_{\kappa\kappa} q_a\right]^{-1} & 0 \\ 0 & E\left[\nabla_{\nu\nu} q_b\right]^{-1} \end{bmatrix} \tag{H-37}$$

because $\nabla_{\kappa\nu} q_a = \nabla_{\nu\kappa} q_b = 0$. Now

$$\nabla_{\kappa\kappa} q_a = \nabla_{\nu\nu} q_b = -2$$

therefore

$$E\left[\nabla_{\gamma\gamma} q\right]^{-1} = \begin{bmatrix} -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{bmatrix}. \tag{H-38}$$

$\mathbf{E\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right]}$
2×2

Given that

$$\nabla_{\gamma} q = 2\left[\left(\mathcal{K}(\alpha, \vec{W}) - \kappa\right)\left(\mathcal{V}(\alpha, \vec{W}) - \nu\right)\right]. \tag{H-39}$$

we have

$$E\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right] = 4\begin{bmatrix} E[(\mathcal{K}(\alpha, \vec{W}) - \kappa)^2] & E[(\mathcal{K}(\alpha, \vec{W}) - \kappa)(\mathcal{V}(\alpha, \vec{W}) - \nu)] \\ E[(\mathcal{K}(\alpha, \vec{W}) - \kappa)(\mathcal{V}(\alpha, \vec{W}) - \nu)] & E[(\mathcal{V}(\alpha, \vec{W}) - \nu)^2] \end{bmatrix}.$$

(H-40)

The corresponding consistent estimator is

$$\hat{E}\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right] = 4\begin{bmatrix} \sum_{i=1}^{n} \frac{1}{n}(\hat{\mathcal{K}}_i - \hat{\kappa})^2 & \sum_{i=1}^{n} \frac{1}{n}(\hat{\mathcal{K}}_i - \hat{\kappa})(\hat{\mathcal{V}}_i - \hat{v}) \\ \sum_{i=1}^{n} \frac{1}{n}(\hat{\mathcal{K}}_i - \hat{\kappa})(\hat{\mathcal{V}}_i - \hat{v}) & \sum_{i=1}^{n} \frac{1}{n}(\hat{\mathcal{V}}_i - \hat{v})^2 \end{bmatrix}. \tag{H-41}$$

Based on the above results, we can consistently estimate (H-5) as

$$\widehat{AVAR(\hat{\gamma})} = \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma\delta} q\right] \widehat{AVAR(\hat{\alpha})} \hat{E}\left[\nabla_{\gamma\delta} q\right]' \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$
$$+ \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right] \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

and using well known results from asymptotic theory for two-stage estimators, we can show that[51]

$$\widehat{AVAR(\hat{\gamma})}^{-\frac{1}{2}} \sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, I). \tag{H-42}$$

Combining (H-42) with (H-1) we also have that

$$\widehat{avar(\hat{\eta}^{UPOL})}^{-\frac{1}{2}} \sqrt{n}(\hat{\eta}^{UPOL} - \eta^{UPOL}) \xrightarrow{d} N(0, I) \tag{H-43}$$

where $\widehat{avar(\hat{\eta}^{UPOL})}$ is given in (H-1).

---

[51]Discussions of asymptotic theory for two-stage optimization estimators can be found in Newey & McFadden (1994), White (1994, Chapter 6), and Wooldridge (2010, Chapter 12).

Appendix I.

Bias from Using $\hat{\eta}^{UPOL}$ Instead of $\hat{\eta}^{UPO}$ in a Model of Alcohol Demand

From equations (4-6) of chapter 4 and (3-10) of chapter 3, we can write the bias using the UPOL approach vs. the UPO method as the following rendition of the difference between elasticities obtained from unrestricted PO model with log prices (UPOL), i.e., $\hat{\eta}^{UPOL}$ and those obtained from unrestricted PO model with nominal prices (UPO), i.e., $\hat{\eta}^{UPO}$.

$$\begin{aligned}
\eta^{UPOL} - \eta^{UPO} = \{ & E\big[\lambda(P\alpha_{P1} + X\alpha_{X1})\exp(P\alpha_{P2} + X\alpha_{X2})\alpha_{P1} \\
& + \Lambda(P\alpha_{P1} + X\alpha_{X1})\exp(P\alpha_{P2} + X\alpha_{X2})\alpha_{P2}\big] \\
& \times \frac{1}{E\big[\Lambda(P\alpha_{P1} + X\alpha_{X1})\exp(P\alpha_{P2} + X\alpha_{X2})\big]}\} \\
& - E\big[\lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})\,a_{\mathcal{P}1} \\
& + \Lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})\,a_{\mathcal{P}2}\big] \\
& \times \frac{E[\mathcal{P}]}{E\big[\Lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})\big]} \\
= \{ & E\big[\Lambda(P\alpha_{P1} + X\alpha_{X1})[1 - \Lambda(P\alpha_{P1} + X\alpha_{X1})]\exp(P\alpha_{P2} + X\alpha_{X2})\alpha_{P1} \\
& + \Lambda(P\alpha_{P1} + X\alpha_{X1})\exp(P\alpha_{P2} + X\alpha_{X2})\alpha_{P2}\big] \\
& \times \frac{1}{E\big[\Lambda(P\alpha_{P1} + X\alpha_{X1})\exp(P\alpha_{P2} + X\alpha_{X2})\big]}\} \\
& - E\big[\Lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})[1 - \Lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})]\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})a_{\mathcal{P}1}
\end{aligned}$$

$$+ \Lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})a_{\mathcal{P}2}\Big]$$

$$\times \frac{E\big[\mathcal{P}\big]}{E\big[\Lambda(\mathcal{P}a_{\mathcal{P}1} + Xa_{X1})\exp(\mathcal{P}a_{\mathcal{P}2} + Xa_{X2})\big]}. \quad \text{(I-1)}$$

If we define

$$v \equiv E[\Lambda(W a_1)\exp(W a_2)]$$

$$u \equiv E[\Lambda(W a_1)^2 \exp(W a_2)]$$

$$m_{\mathcal{P}} = E\big[\mathcal{P}\big]$$

$$\nu \equiv E[\Lambda(\vec{W} \alpha_1)\exp(\vec{W} \alpha_2)]$$

and

$$\omega \equiv E[\Lambda(\vec{W} \alpha_1)^2 \exp(\vec{W} \alpha_2)]$$

where $\vec{W}_i = [P_i \quad X_i]$ denotes the observation on $\vec{W} = [P \quad X]$ for the ith individual in the sample (i = 1, ..., n), with P expressed as log of alcohol price, and

$W_i = [\mathcal{P}_i \quad X_i]$ denotes the observation on $W = [\mathcal{P} \quad X]$ for the ith individual in the sample (i = 1, ..., n), $\mathcal{P}$ with expressed as nominal price of alcohol.

$$\eta^{AL} - \eta^{UPO} = \frac{[(\nu - \omega)\alpha_{P1} + \nu\alpha_{P2}]}{\nu} - \frac{[(\nu - u)a_{\mathcal{P}1} + \nu a_{\mathcal{P}2}]m_{\mathcal{P}}}{\nu}$$

$$= \left[\left(1 - \frac{\omega}{\nu}\right)\alpha_{P1} + \alpha_{P2}\right] - \left[\left(1 - \frac{u}{\nu}\right)a_{\mathcal{P}1} + a_{\mathcal{P}2}\right]m_{\mathcal{P}} \ . \quad \text{(I-2)}$$

Appendix J.

Asymptotic Distribution (and Standard Error) of $\hat{\eta}^{UPOL} - \hat{\eta}^{UPO}$

In Appendix I we showed that

$$\eta^{UPOL} - \eta^{UPO} = \left[\left(1-\frac{\ddot{\omega}}{\ddot{v}}\right)\alpha_{P1} + \alpha_{P2}\right] - \left[\left(1-\frac{u}{v}\right)a_{\mathcal{P}1} + a_{\mathcal{P}2}\right]m_{\mathcal{P}}$$

where

$$\ddot{v} \equiv E[\Lambda(\ddot{W}\alpha_1)\exp(\ddot{W}\alpha_2)]$$

$$\ddot{\omega} \equiv E[\Lambda(\ddot{W}\alpha_1)^2\exp(\ddot{W}\alpha_2)]$$

$$v \equiv E[\Lambda(\vec{W}a_1)\exp(\vec{W}a_2)]$$

$$u \equiv E[\Lambda(\vec{W}a_1)^2\exp(\vec{W}a_2)]$$

$$\ddot{W} = [P \quad X]$$

$$\vec{W} = [\mathcal{P} \quad X]$$

and

$$m_{\mathcal{P}} \equiv E[\mathcal{P}]$$

Using the corresponding consistent estimators for ω, v, $\ddot{v}$ and $\ddot{\omega}$ say

$$\hat{\ddot{\omega}} = \sum_{i=1}^{n}\frac{1}{n}\Omega(\hat{\alpha}, \ddot{W}_i)$$

$$\vec{\ddot{v}} = \sum_{i=1}^{n}\frac{1}{n}\mathcal{V}(\hat{\alpha}, \ddot{W}_i)$$

$$\hat{v} = \sum_{i=1}^{n} \frac{1}{n} V(\hat{a}, \vec{W}_i)$$

$$\hat{u} = \sum_{i=1}^{n} \frac{1}{n} U(\hat{a}, \vec{W}_i)$$

and

$$\hat{m}_{\mathcal{P}} = \sum_{i=1}^{n} \frac{1}{n} \mathcal{P}_i$$

$$\ddot{\Omega}(\alpha, \ddot{W}) = \Lambda(\ddot{W}\alpha_1)^2 \exp(\ddot{W}\alpha_2)$$

$$\ddot{\mathcal{V}}(\alpha, \ddot{W}) = \Lambda(\ddot{W}\alpha_1)\exp(\ddot{W}\alpha_2)$$

$$V(a, \vec{W}) = \Lambda(\vec{W}a_1)\exp(\vec{W}a_2)$$

$$U(a, \vec{W}) = \Lambda(\vec{W}a_1)^2 \exp(\vec{W}a_2)$$

$\hat{\alpha} = [\hat{\alpha}_1' \quad \hat{\alpha}_2']'$ (with $\hat{\alpha}_1' = [\hat{\alpha}_{\mathcal{P}1} \quad \hat{\alpha}_{X1}']$ and $\hat{\alpha}_2' = [\hat{\alpha}_{\mathcal{P}2} \quad \hat{\alpha}_{X2}']$) is the consistent estimate of the parameter vector $\alpha = [\alpha_1' \quad \alpha_2']'$ (with $\alpha_1' = [\alpha_{\mathcal{P}1} \quad \alpha_{X1}']$ and $\alpha_2' = [\alpha_{\mathcal{P}2} \quad \alpha_{X2}']$) [the parameters of equation (4-6)] obtained via the two-part protocol culminating in (4-4) using the nominal prices of alcohol

$\hat{a} = [\hat{a}_1' \quad \hat{a}_2']'$ (with $\hat{a}_1' = [\hat{a}_{\mathcal{P}1} \quad \hat{a}_{X1}']$ and $\hat{a}_2' = [\hat{a}_{\mathcal{P}2} \quad \hat{a}_{X2}']$) is the consistent estimate of the parameter vector $a = [a_1' \quad a_2']'$ (with $a_1' = [a_{\mathcal{P}1} \quad a_{X1}']$ and $a_2' = [a_{\mathcal{P}2} \quad a_{X2}']$)) [the parameters of equations (3-6) and (3-7), respectively] obtained via the unrestricted two-part protocol culminating in (3-8) using the log prices of alcohol

and $\ddot{W}_i = [P_i \quad X_i]$ denotes the observation on $\ddot{W} = [P \quad X]$ for the ith individual in the sample

(i = 1, ..., n), and $\vec{W}_i = [\mathcal{P}_i \quad X_i]$ denotes the observation on $\vec{W} = [\mathcal{P} \quad X]$ for the ith individual in the sample (i = 1, ..., n).

Let

$\hat{\tau} = [\hat{\alpha}_{P1} \quad \hat{\alpha}_{P2} \; \hat{a}_{\mathcal{P}1} \quad \hat{a}_{\mathcal{P}2} \; \hat{\ddot{\omega}} \quad \hat{\ddot{v}} \quad \hat{u} \quad \hat{v} \quad m_{\mathcal{P}}]'$

and

$\tau = [\alpha_{P1} \quad \alpha_{P2} \; a_{\mathcal{P}1} \quad a_{\mathcal{P}2} \; \ddot{\omega} \quad \ddot{v} \quad u \quad v \quad m_{\mathcal{P}}]'$, where $\text{plim}[\hat{\tau}] = \tau$.

If we could show that

$$\text{AVAR}(\hat{\tau})^{-\frac{1}{2}} \sqrt{n}(\hat{\tau} - \tau) \xrightarrow{d} N(0, I)$$

where the formulation of $\text{AVAR}(\hat{\tau})$ is known, then we could apply the δ-method to obtain the asymptotic variance of $\hat{\eta}^{UPOL} - \hat{\eta}^{UPO}$ as

$$\text{avar}(\hat{\eta}^{UPOL} - \hat{\eta}^{UPO}) = c(\tau) \; \text{AVAR}(\hat{\tau}) \; c(\tau)'$$

where $c(\tau) = \begin{bmatrix} \left(1-\frac{\ddot{\omega}}{\ddot{\nu}}\right) \\ 1 \\ -m_{\mathcal{P}}\left(1-\frac{u}{v}\right) \\ -m_{\mathcal{P}} \\ -\alpha_{P1}/\ddot{\nu} \\ \alpha_{P1}\ddot{\omega}/\ddot{\nu}^2 \\ m_{\mathcal{P}}\, a_{\mathcal{P}1}/\nu \\ -m_{\mathcal{P}}\, a_{\mathcal{P}1}\, u/\nu^2 \\ -\left(1-\frac{u}{v}\right)a_{\mathcal{P}1} - a_{\mathcal{P}2} \end{bmatrix}'$.

Moreover, if we have a consistent estimator for $AVAR(\hat{\tau})$, say $\widehat{AVAR(\hat{\tau})}$ [i.e. $p\lim\left[\widehat{AVAR(\hat{\tau})}\right] = AVAR(\hat{\tau})$], then we could consistently estimate $avar(\hat{\eta}^{UPOL} - \hat{\eta}^{UPO})$ as

$$\widehat{avar(\hat{\eta}^{UPOL} - \hat{\eta}^{UPO})} = c(\hat{\tau})\ \widehat{AVAR(\hat{\tau})}\ c(\hat{\tau})'. \qquad \text{(J-1)}$$

We focus, therefore, on finding the asymptotic distribution of $\hat{\tau}$ and, in particular, the formulation of its asymptotic covariance matrix.

First note that we can write τ as

$$\tau = \Xi\, \theta \qquad \text{(J-2)}$$

where $\theta' = [\delta' \quad \gamma']$, $\delta = [\alpha_1' \quad \alpha_2' \quad a_1' \quad a_2']'$, $\gamma' = [\ddot{\omega} \quad \ddot{\nu} \quad u \quad \nu \quad m_{\mathcal{P}}]$ (recall, $\alpha_1' = [\alpha_{P1} \quad \alpha_{X1}']$, $\alpha_2' = [\alpha_{P2} \quad \alpha_{X2}']$, $\hat{a}_1' = [\hat{a}_{\mathcal{P}1} \quad \hat{a}_{X1}']$, and $\hat{a}_2' = [\hat{a}_{\mathcal{P}2} \quad \hat{a}_{X2}']$).

$$\Xi = \begin{bmatrix} \ell_{\alpha_{P1}} \\ \hline \ell_{\alpha_{P2}} \\ \hline \ell_{a_{P1}} \\ \hline \ell_{a_{P2}} \\ \hline 0_{5,4K} \quad I_5 \end{bmatrix}$$

ℓ_a is the unit row vector with the value "1" in the element position corresponding to the element position of a in the vector θ, $0_{b,c}$ is the matrix of zeros whose row and column dimensions are b and c, respectively, I_d is the identity matrix of order d, and K is the column dimension of W. For future reference, let's set the following vector/matrix dimensions:

α_1 is K×1

α_2 is K×1

a_1 is K×1

a_2 is K×1

$\ddot{W}$ is 1×K

$\vec{W}$ is 1×K

τ is 9×1

$c(\tau)$ is 1×9

δ is 4K×1

γ is 5×1

θ is (4K+5)×1

Ξ is 9×(4K+5)

$\ell_{\alpha_{P1}} = \ell_{\alpha_{P2}} = \ell_{a_{\mathcal{P}1}} = \ell_{a_{\mathcal{P}2}} = 1\times(4K+5)$

Clearly then

$$\text{AVAR}(\hat{\tau}) = \Xi \, \text{AVAR}(\hat{\theta}) \, \Xi' \tag{J-3}$$

where $\hat{\theta}$ is the estimator of θ obtained from the following two-stage protocol.

First Stage

Consistently estimate δ via the following optimization estimator

$$\hat{\delta} = \arg\max_{\tilde{\delta}} \frac{\sum_{i=1}^{n} q_1(\tilde{\delta}, S_i)}{n} \tag{J-4}$$

where

$$q_1(\tilde{\delta}, S_i) = q_{11}(\tilde{\alpha}_1, \ddot{S}_i) + q_{12}(\tilde{\alpha}_2, \ddot{S}_i) + q_{13}(\tilde{a}_1, \vec{S}_i) + q_{14}(\tilde{a}_2, \vec{S}_i)$$

$$q_{11}(\tilde{\alpha}_1, \ddot{S}_i) = I(A_i > 0)\ln[\Lambda(\ddot{W}_i \tilde{\alpha}_1)] + [1 - I(A_i > 0)]\ln[1 - \Lambda(\ddot{W}_i \tilde{\alpha}_1)]$$

$$q_{12}(\tilde{\alpha}_2, \ddot{S}_i) = -I(A_i > 0)\left(A_i - \exp\left(\ddot{W}_i \tilde{\alpha}_2\right)\right)^2$$

$$q_{13}(\tilde{a}_1, \vec{S}_i) = I(A_i > 0)\ln[\Lambda(\vec{W}_i \tilde{a}_1)] + [1 - I(A_i > 0)]\ln[1 - \Lambda(\vec{W}_i \tilde{a}_1)]$$

$$q_{14}(\tilde{a}_2, \vec{S}_i) = -I(A_i > 0)\left(A_i - \exp\left(\vec{W}_i \tilde{a}_2\right)\right)^2$$

$\ddot{S}_i = [A_i \quad X_i \quad P_i]$ and $\vec{S}_i = [A_i \quad X_i \quad \mathcal{P}_i]$

$\tilde{\delta} = [\tilde{\alpha}_1' \quad \tilde{\alpha}_2' \quad \tilde{a}_1' \quad \tilde{a}_2']'$, $\tilde{\alpha}_1' = [\tilde{\alpha}_{P1} \quad \tilde{\alpha}_{X1}']$, $\tilde{\alpha}_2' = [\tilde{\alpha}_{P2} \quad \tilde{\alpha}_{X2}']$, $\tilde{a}_1' = [\tilde{a}_{\mathcal{P}1} \quad \tilde{a}_{X1}']$,

$\tilde{a}_2' = [\tilde{a}_{\mathcal{P}2} \quad \tilde{a}_{X2}']$ and $\hat{\delta} = [\hat{\alpha}_1' \quad \hat{\alpha}_2' \quad \hat{a}_1' \quad \hat{a}_2']'$

Second Stage

Consistently estimate γ via the following optimization estimator

$$\hat{\gamma} = \arg\max_{\tilde{\gamma}} \frac{\sum_{i=1}^{n} q(\hat{\delta}, \tilde{\gamma}, S_i)}{n} \tag{J-5}$$

where

$$q(\hat{\delta}, \tilde{\gamma}, S_i) = q_a(\hat{\delta}, \tilde{\ddot{\omega}}, \ddot{S}_i) + q_b(\hat{\delta}, \tilde{\ddot{v}}, \ddot{S}_i) + q_c(\hat{\delta}, \tilde{u}, \vec{S}_i) + q_d(\hat{\delta}, \tilde{v}, \vec{S}_i) + q_e(\hat{\delta}, \tilde{m}_{\mathcal{P}}, \vec{S}_i)$$

$$q_a(\hat{\delta}, \tilde{\ddot{\omega}}, \ddot{S}_i) = -(\ddot{\Omega}(\hat{\alpha}, \ddot{W}_i) - \tilde{\ddot{\omega}})^2$$

$$q_b(\hat{\delta}, \tilde{\ddot{v}}, \ddot{S}_i) = -(\ddot{V}(\hat{\alpha}, \ddot{W}_i) - \tilde{\ddot{v}})^2$$

$$q_c(\hat{\delta}, u, \vec{S}_i) = -(U(\hat{a}, \vec{W}_i) - \tilde{u})^2$$

$$q_d(\hat{\delta}, \tilde{v}, \vec{S}_i) = -(V(\hat{a}, \vec{W}_i) - \tilde{v})^2$$

$$q_e(\hat{\delta}, \tilde{m}_{\mathcal{P}}, \vec{S}_i) = -(\mathcal{P}_i - \tilde{m}_{\mathcal{P}})^2$$

$\hat{\delta} = [\hat{\alpha}_1' \quad \hat{\alpha}_2' \quad \hat{a}_1' \quad \hat{a}_2']'$ is the first stage estimator of δ,

$\hat{\alpha}_1' = [\hat{\alpha}_{P1} \quad \hat{\alpha}_{X1}']$, $\hat{\alpha}_2' = [\hat{\alpha}_{P2} \quad \hat{\alpha}_{X2}']$, $\hat{a}_1' = [\hat{a}_{\mathcal{P}1} \quad \hat{a}_{X1}']$ and $\hat{a}_2' = [\hat{a}_{\mathcal{P}2} \quad \hat{a}_{X2}']$.

Use q_l as shorthand notation for

$$q_l(\delta, S) = q_{l1}(\alpha_1, \ddot{S}) + q_{l2}(\alpha_2, \ddot{S}) + q_{l3}(a_1, \vec{S}) + q_{l4}(a_2, \vec{S})$$

with

$$q_{11}(\alpha_1, \ddot{S}) = I(A > 0)\ln[\Lambda(\ddot{W}\alpha_1)] + [1 - I(A > 0)]\ln[1 - \Lambda(\ddot{W}\alpha_1)]$$

$$q_{12}(\alpha_2, \ddot{S}) = -I(A > 0)\left(A - \exp\left(\ddot{W}\alpha_2\right)\right)^2$$

$$q_{13}(a_1, \vec{S}) = I(A > 0)\ln[\Lambda(\vec{W}a_1)] + [1 - I(A > 0)]\ln[1 - \Lambda(\vec{W}a_1)]$$

$$q_{14}(a_2, \vec{S}) = -I(A > 0)\left(A_i - \exp\left(\vec{W}_i a_2\right)\right)^2$$

$$\ddot{S} = [A \quad X \quad P] \text{ and } \vec{S} = [A \quad X \quad \mathcal{P}].$$

and use q as shorthand notation for

$$q(\delta, \tilde{\gamma}, S) = q_a(\delta, \ddot{\omega}, \ddot{S}) + q_b(\delta, \ddot{v}, \ddot{S}) + q_c(\delta, \vec{\omega}, \vec{S}) + q_d(\delta, \vec{v}, \vec{S}) + q_e(\delta, \mathcal{P}, \vec{S})$$

with

$$q_a(\delta, \ddot{\omega}, \ddot{S}) = -(\ddot{\Omega}(\alpha, \ddot{W}) - \ddot{\omega})^2$$

$$q_b(\delta, \ddot{v}, \ddot{S}) = -(\ddot{\mathcal{V}}(\alpha, \ddot{W}) - \ddot{v})^2$$

$$q_c(\delta, u, \vec{S}) = -(U(a, \vec{W}) - u)^2$$

$$q_d(\delta, v, \vec{S}) = -(V(a, \vec{W}) - v)^2$$

$$q_e(\delta, m_{\mathcal{P}}, \vec{S}) = -(\mathcal{P} - m_{\mathcal{P}})^2$$

and let $AVAR(\hat{\delta})$ denote the asymptotic covariance matrix of the first stage estimator.[52] Using well known results from asymptotic theory for two-stage estimators, we can show that[53]

---

[52] Note that a consistent estimate $AVAR(\hat{\delta})$ can be obtained from the packaged output for the first stage estimator because the first stage estimator is unaffected by the fact that it is a component of a two-stage estimator.

$$\text{AVAR}(\hat{\theta})^{-\frac{1}{2}}\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I) \tag{J-6}$$

where $\hat{\theta}' = [\hat{\delta}' \quad \hat{\gamma}']$, $\text{plim}(\hat{\theta}) = \theta$

$$\text{AVAR}(\hat{\theta}) = \begin{bmatrix} D_{11} & D_{12} \\ D_{12}' & D_{22} \end{bmatrix} \tag{J-7}$$

$$D_{11} = \underset{4K\times 4K}{\text{AVAR}(\hat{\delta})} \tag{J-8}$$

$$\begin{aligned} D_{12} &= E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\delta} q_1 ' \nabla_{\gamma} q\right] E\left[\nabla_{\gamma\gamma} q\right]^{-1} \\ &\quad - E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\delta} q_1 ' \nabla_{\delta} q_1\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma\delta} q\right]' E\left[\nabla_{\gamma\gamma} q\right]^{-1} \end{aligned} \tag{J-9}$$

$$\begin{aligned} D_{22} &= \text{AVAR}(\hat{\gamma}) = E\left[\nabla_{\gamma\gamma} q\right]^{-1} \Big[E\left[\nabla_{\gamma\delta} q\right] \text{AVAR}(\hat{\delta}) E\left[\nabla_{\gamma\delta} q\right]' \\ &\quad - E\left[\nabla_{\gamma} q ' \nabla_{\delta} q_1\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma\delta} q\right]' \\ &\quad - E\left[\nabla_{\gamma\delta} q\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\gamma} q ' \nabla_{\delta} q_1\right]' \Big] E\left[\nabla_{\gamma\gamma} q\right]^{-1} \\ &\quad + E\left[\nabla_{\gamma\gamma} q\right]^{-1} E\left[\nabla_{\gamma} q ' \nabla_{\gamma} q\right] E\left[\nabla_{\gamma\gamma} q\right]^{-1}. \end{aligned} \tag{J-10}$$

Fortunately, (J-9) and (J-10) can be simplified in a number of ways. First note that we can write

$$E\left[\nabla_{\gamma} q ' \nabla_{\delta} q_1\right] = E\left[\nabla_{\gamma} q ' E[\nabla_{\delta} q_1 \mid \vec{W}]\right]$$

---

[53]Discussions of asymptotic theory for two-stage optimization estimators can be found in Newey & McFadden (1994), White (1994, Chapter 6), and Wooldridge (2010, Chapter 12).

but

$$\nabla_{\delta} q_1 = \nabla_{\delta} q_{11} + \nabla_{\delta} q_{12} + \nabla_{\delta} q_{13} + \nabla_{\delta} q_{14}$$

with

$$\nabla_{\delta} q_{11} = [\nabla_{\alpha_1} \ln f(I(A>0) \mid \ddot{W}) \quad 0 \quad 0 \quad 0]$$

$$\nabla_{\delta} q_{12} = \left[0 \quad 2\, I(A>0)\left(A - \exp\left(\ddot{W}\alpha_2\right)\right)\exp\left(\ddot{W}\alpha_2\right)\ddot{W} \quad 0 \quad 0 \right]$$

$$\nabla_{\delta} q_{13} = \left[0 \quad 0 \quad \nabla_{a_1} \ln f(I(A>0) \mid \vec{W}) \quad 0 \right]$$

$$\nabla_{\delta} q_{14} = \left[0 \quad 0 \quad 0 \quad 2\, I(A>0)\left(A - \exp\left(\vec{W}a_2\right)\right)\exp\left(\vec{W}a_2\right)\vec{W} \right]$$

where

$$f(I(A>0) \mid \ddot{W}) = \Lambda(\ddot{W}\alpha_1)^{I(A>0)}[1 - \Lambda(\ddot{W}\alpha_1)]^{[1 - I(A>0)]} .$$

$$f(I(A>0) \mid \vec{W}) = \Lambda(\vec{W}a_1)^{I(A>0)}[1 - \Lambda(\vec{W}a_1)]^{[1 - I(A>0)]} .$$

Therefore

$$E[\nabla_{\delta} q_1 \mid W] = \begin{bmatrix} E[\nabla_{\alpha_1} \ln f(I(A>0) \mid \ddot{W})] \\ 2E\left[I(A>0)\left(A - \exp\left(\ddot{W}\alpha_2\right)\right)\exp\left(\ddot{W}\alpha_2\right) \mid \ddot{W} \right]\ddot{W} \\ E[\nabla_{a_1} \ln f(I(A>0) \mid \vec{W})] \\ 2E\left[I(A>0)\left(A - \exp\left(\vec{W}a_2\right)\right)\exp\left(\vec{W}a_2\right) \mid \vec{W} \right]\vec{W} \end{bmatrix}'$$

$$= [\,0 \quad 0 \quad 0 \quad 0]$$

because $E[\nabla_{\alpha_1} \ln f(I(A>0) \mid \ddot{W})] = 0$ and $E[\nabla_{a_1} \ln f(I(A>0) \mid \vec{W})] = 0$ [see (13.20) on p. 477 of Wooldridge (2010)], $E\left[I(A>0)\left(A - \exp\left(\ddot{W}\alpha_2\right)\right)\exp\left(\ddot{W}\alpha_2\right) \mid \ddot{W} \right] = 0$ and $E\left[I(A>0)\left(A - \exp\left(\vec{W}a_2\right)\right)\exp\left(\vec{W}a_2\right) \mid \vec{W} \right] = 0$, by design. Finally, then we get

$$E\left[\nabla_\gamma q' \nabla_\delta q_1\right] = 0$$

so

$$D_{12} = -\underset{4K\times 4K}{E\left[\nabla_{\delta\delta} q_1\right]^{-1}} \underset{4K\times 4K}{E\left[\nabla_\delta q_1' \nabla_\delta q_1\right]} \underset{4K\times 4K}{E\left[\nabla_{\delta\delta} q_1\right]^{-1}} \underset{4K\times 5}{E\left[\nabla_{\gamma\delta} q\right]'} \underset{5\times 5}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \qquad \text{(J-11)}$$

$$\underset{4K\times 5}{\phantom{E\left[\nabla_{\gamma\delta} q\right]'}}$$

and

$$D_{22} = AVAR(\hat{\gamma}) = \underset{5\times 5}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \underset{5\times 4K}{E\left[\nabla_{\gamma\delta} q\right]} \underset{4K\times 4K}{AVAR(\hat{\delta})} \underset{4K\times 5}{E\left[\nabla_{\gamma\delta} q\right]'} \underset{5\times 5}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}}$$

$$+ \underset{5\times 5}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}} \underset{5\times 5}{E\left[\nabla_\gamma q' \nabla_\gamma q\right]} \underset{5\times 5}{E\left[\nabla_{\gamma\gamma} q\right]^{-1}}. \qquad \text{(J-12)}$$

so

$$\underset{(4K+5)\times(4K+5)}{AVAR(\hat{\theta})} = \begin{bmatrix} \underset{4K\times 4K}{D_{11}} & \underset{4K\times 5}{D_{12}} \\ \underset{5\times 4K}{D_{12}'} & \underset{5\times 5}{D_{22}} \end{bmatrix}.$$

Let's consider each of the individual components of (J-6) and (J-7) in turn.

$E[\nabla_{\delta\delta} q_1]^{-1}$
4K×4K

Written out explicitly we have

$$E[\nabla_{\delta\delta} q_1]^{-1} = \begin{bmatrix} E[\nabla_{\alpha_1\alpha_1} q_{11}]^{-1} & 0 & 0 & 0 \\ 0 & E[\nabla_{\alpha_2\alpha_2} q_{12}]^{-1} & 0 & 0 \\ 0 & 0 & E[\nabla_{a_1a_1} q_{13}]^{-1} & 0 \\ 0 & 0 & 0 & E[\nabla_{a_2a_2} q_{14}]^{-1} \end{bmatrix}.$$

(J-13)

Now

$$E[\nabla_{\alpha_1\alpha_1} q_{11}]^{-1} = -\mathrm{AVAR}(\hat{\alpha}_1) \tag{J-14}$$

= the negative of the asymptotic covariance matrix for first stage, first part, logit estimation in the two-stage estimation protocol for θ. We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{\alpha_1\alpha_1} q_{11}]^{-1}$ is

$$\hat{E}[\nabla_{\alpha_1\alpha_1} q_{11}]^{-1} = -n\widehat{\mathrm{AVAR}}^*(\hat{\alpha}_1) \tag{J-15}$$

where $\widehat{\mathrm{AVAR}}^*(\hat{\alpha}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure. Also

$$\nabla_{\alpha_2} q_{12} = 2\,I(A>0)\left(A - \exp\left(\ddot{W}\alpha_2\right)\right)\exp\left(\ddot{W}\alpha_2\right)\ddot{W}$$

and

$$\nabla_{\alpha_2\alpha_2} q_{12} = 2\,I(A>0)\left[\left(A - \exp\left(\ddot{W}\alpha_2\right)\right) - \exp\left(\ddot{W}\alpha_2\right)\right]\exp\left(\ddot{W}\alpha_2\right)\ddot{W}'\ddot{W}.$$

Therefore

$$E[\nabla_{\alpha_2\alpha_2} q_{12}] = 2E\left[I(A>0)\left[\left(A - \exp\left(\ddot{W}\alpha_2\right)\right) - \exp\left(\ddot{W}\alpha_2\right)\right]\exp\left(\ddot{W}\alpha_2\right)\ddot{W}'\ddot{W}\right]. \tag{J-16}$$

A consistent estimator of $E[\nabla_{\alpha_2\alpha_2} q_{12}]^{-1}$ is

$$\hat{E}[\nabla_{\alpha_2\alpha_2} q_{12}]^{-1} = n_1\left[2\sum_{i=1}^{n}\left\{I(A_i>0)\left[\left(A_i-\exp\left(\ddot{W}_i\,\alpha_2\right)\right)-\exp\left(\ddot{W}_i\,\alpha_2\right)\right]\exp\left(\ddot{W}_i\,\alpha_2\right)\ddot{W}_i'\ddot{W}_i\right\}\right]^{-1} \tag{J-17}$$

where $\hat{\alpha}_2$ is the first stage, second part, estimator of α_2 , and n_1 is the size of the subsample for whom $I(A>0)=1$.

Similarly,

$$E[\nabla_{a_1a_1} q_{13}]^{-1} = -AVAR(\hat{a}_1) \tag{J-18}$$

= the negative of the asymptotic covariance matrix for first stage, first part, logit estimation in the two-stage estimation protocol for θ. We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{a_1a_1} q_{13}]^{-1}$ is

$$\hat{E}[\nabla_{a_1a_1} q_{13}]^{-1} = -n\widehat{AVAR}^*(\hat{a}_1) \tag{J-19}$$

where $\widehat{AVAR}^*(\hat{a}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure. Also

$$\nabla_{a_2} q_{13} = 2\,I(A>0)\left(A-\exp\left(\vec{W}a_2\right)\right)\exp\left(\vec{W}a_2\right)\vec{W}$$

and

$$\nabla_{a_2a_2} q_{14} = 2\,I(A>0)\left[\left(A-\exp\left(\vec{W}a_2\right)\right)-\exp\left(\vec{W}a_2\right)\right]\exp\left(\vec{W}a_2\right)\vec{W}'\vec{W}.$$

Therefore

$$E[\nabla_{a_2a_2} q_{14}] = 2E\left[I(A>0)\left[\left(A - \exp\left(\vec{W}a_2\right)\right) - \exp\left(\vec{W}a_2\right)\right]\exp\left(\vec{W}a_2\right)\vec{W}'\vec{W}\right].$$

(J-20)

A consistent estimator of $E[\nabla_{a_2a_2} q_{14}]^{-1}$ is

$$\hat{E}[\nabla_{a_2a_2} q_{14}]^{-1} = n_1\left[2\sum_{i=1}^{n}\left\{ I(A_i>0)\left[\left(A_i - \exp\left(\vec{W}_i a_2\right)\right) - \exp\left(\vec{W}_i a_2\right)\right]\exp\left(\vec{W}_i a_2\right)\vec{W}_i'\vec{W}_i\right\}\right]^{-1}$$

(J-21)

where $\hat{a}_2$ is the first stage, second part, estimator of a_2 , and n_1 is the size of the subsample for whom $I(A>0) = 1$.

so

$$\hat{E}[\nabla_{\delta\delta} q_1]^{-1} = \begin{bmatrix} \hat{E}[\nabla_{\alpha_1\alpha_1} q_{11}]^{-1} & 0 & 0 & 0 \\ 0 & \hat{E}[\nabla_{\alpha_2\alpha_2} q_{12}]^{-1} & 0 & 0 \\ 0 & 0 & \hat{E}[\nabla_{a_1a_1} q_{13}]^{-1} & 0 \\ 0 & 0 & 0 & \hat{E}[\nabla_{a_2a_2} q_{14}]^{-1} \end{bmatrix}.$$

(J-22)

$$\underset{4K\times 4K}{\mathbf{E}\left[\nabla_{\delta}\mathbf{q_1}'\nabla_{\delta}\mathbf{q_1}\right]}$$

Written out explicitly we have

$$E\left[\nabla_{\delta} q_1 ' \nabla_{\delta} q_1\right] = \begin{bmatrix} E\begin{bmatrix}\nabla_{\alpha_1} q_{11} '\\ \nabla_{\alpha_1} q_{11}\end{bmatrix} & E\begin{bmatrix}\nabla_{\alpha_1} q_{11} '\\ \nabla_{\alpha_2} q_{12}\end{bmatrix} & E\begin{bmatrix}\nabla_{\alpha_1} q_{11} '\\ \nabla_{a_1} q_{13}\end{bmatrix} & E\begin{bmatrix}\nabla_{\alpha_1} q_{11} '\\ \nabla_{a_2} q_{14}\end{bmatrix} \\ E\begin{bmatrix}\nabla_{\alpha_2} q_{12} '\\ \nabla_{\alpha_1} q_{11}\end{bmatrix} & E\begin{bmatrix}\nabla_{\alpha_2} q_{12} '\\ \nabla_{\alpha_2} q_{12}\end{bmatrix} & E\begin{bmatrix}\nabla_{\alpha_2} q_{12} '\\ \nabla_{a_1} q_{13}\end{bmatrix} & E\begin{bmatrix}\nabla_{\alpha_2} q_{12} '\\ \nabla_{a_2} q_{14}\end{bmatrix} \\ E\begin{bmatrix}\nabla_{a_1} q_{13} '\\ \nabla_{\alpha_1} q_{11}\end{bmatrix} & E\begin{bmatrix}\nabla_{a_1} q_{13} '\\ \nabla_{\alpha_2} q_{12}\end{bmatrix} & E\begin{bmatrix}\nabla_{a_1} q_{13} '\\ \nabla_{a_1} q_{13}\end{bmatrix} & E\begin{bmatrix}\nabla_{a_1} q_{13} '\\ \nabla_{a_2} q_{14}\end{bmatrix} \\ E\begin{bmatrix}\nabla_{a_2} q_{14} '\\ \nabla_{\alpha_1} q_{11}\end{bmatrix} & E\begin{bmatrix}\nabla_{a_2} q_{14} '\\ \nabla_{\alpha_2} q_{12}\end{bmatrix} & E\begin{bmatrix}\nabla_{a_2} q_{14} '\\ \nabla_{a_1} q_{13}\end{bmatrix} & E\begin{bmatrix}\nabla_{a_2} q_{14} '\\ \nabla_{a_2} q_{14}\end{bmatrix} \end{bmatrix}. \tag{J-23}$$

Because the first stage, first part, estimator of α_1 is MLE we can write

$$E[\nabla_{\alpha_1} q_{11} ' \nabla_{\alpha_1} q_{11}] = -E[\nabla_{\alpha_1 \alpha_1} q_{11}] = \left[\text{AVAR}(\hat{\alpha}_1)\right]^{-1}$$

= the inverse of the asymptotic covariance matrix for first stage, first part, logit estimation in the two-stage estimation protocol for θ. We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{\alpha_1} q_{11} ' \nabla_{\alpha_1} q_{11}]$ is

$$\hat{E}[\nabla_{\alpha_1} q_{11} ' \nabla_{\alpha_1} q_{11}] = \frac{1}{n}\left[\widehat{\text{AVAR}} * (\hat{\alpha}_1)\right]^{-1} \tag{J-24}$$

where $\widehat{\text{AVAR}} * (\hat{\alpha}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure.

Similarly, because the first stage, first part, estimator of a_1 is also MLE we can write

$$E[\nabla_{a_1} q_{13}' \nabla_{a_1} q_{13}] = -E[\nabla_{a_1 a_1} q_{13}] = \left[AVAR(\hat{a}_1)\right]^{-1}$$

= the inverse of the asymptotic covariance matrix for first stage, first part, logit estimation in the two-stage estimation protocol for θ. We get an estimate of this directly from the Stata output.

A consistent estimator of $E[\nabla_{a_1} q_{13}' \nabla_{a_1} q_{13}]$ is

$$\hat{E}[\nabla_{a_1} q_{13}' \nabla_{a_1} q_{13}] = \frac{1}{n}\left[\widehat{AVAR} * (\hat{a}_1)\right]^{-1} \qquad \text{(J-25)}$$

where $\widehat{AVAR} * (\hat{a}_1)$ is the estimated variance-covariance matrix output by the Stata logit procedure.

The remainder of the block elements follow from

$$\nabla_{\alpha_1} q_{11} = \nabla_{\alpha_1} \ln f(I(A > 0) \mid \ddot{W}) = \left[I(A > 0)[1 - \Lambda(\ddot{W} \alpha_1)] - [1 - I(A > 0)]\Lambda(\ddot{W} \alpha_1) \right] \ddot{W} \qquad \text{(J-26)}$$

$$\nabla_{\alpha_2} q_{12} = 2\, I(A > 0)\left(A - \exp\left(\ddot{W} \alpha_2\right)\right)\exp\left(\ddot{W} \alpha_2\right)\ddot{W}. \qquad \text{(J-27)}$$

$$\nabla_{a_1} q_{13} = \nabla_{a_1} \ln f(I(A > 0) \mid \vec{W}) = \left[I(A > 0)[1 - \Lambda(\vec{W} a_1)] - [1 - I(A > 0)]\Lambda(\vec{W} a_1) \right] \vec{W} \qquad \text{(J-28)}$$

$$\nabla_{a_2} q_{14} = 2\, I(A > 0)\left(A - \exp\left(\vec{W} a_2\right)\right)\exp\left(\vec{W} a_2\right)\vec{W}. \qquad \text{(J-29)}$$

where the formulation of $\nabla_{\alpha_1} q_{11}$ and $\nabla_{a_1} q_{13}$ comes from equation (16.4.8) on p. 350 of Fomby et al. (1984) and $\Lambda(\ )$ denotes the logistic cdf. The remaining required consistent matrix estimators are

$$\hat{E}[\nabla_{\alpha_1} q_{11}\,'\nabla_{\alpha_2} q_{12}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{\alpha_1}\hat{q}_{11i}\,'\nabla_{\alpha_2}\hat{q}_{12i} \tag{J-30}$$

$$\hat{E}[\nabla_{\alpha_2} q_{12}\,'\nabla_{\alpha_2} q_{12}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{\alpha_2}\hat{q}_{12i}\,'\nabla_{\alpha_2}\hat{q}_{12i} \tag{J-31}$$

$$\hat{E}[\nabla_{\alpha_1} q_{11}\,'\nabla_{a_1} q_{13}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{\alpha_1}\hat{q}_{11i}\,'\nabla_{a_1}\hat{q}_{13i} \tag{J-32}$$

$$\hat{E}[\nabla_{\alpha_2} q_{12}\,'\nabla_{a_1} q_{13}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{\alpha_2}\hat{q}_{12i}\,'\nabla_{a_1}\hat{q}_{13i} \tag{J-33}$$

$$\hat{E}[\nabla_{\alpha_1} q_{11}\,'\nabla_{a_2} q_{14}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{\alpha_1}\hat{q}_{11i}\,'\nabla_{a_2}\hat{q}_{14i} \tag{J-34}$$

$$\hat{E}[\nabla_{\alpha_2} q_{12}\,'\nabla_{a_2} q_{14}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{\alpha_2}\hat{q}_{12i}\,'\nabla_{a_2}\hat{q}_{14i} \tag{J-35}$$

$$\hat{E}[\nabla_{a_1} q_{13}\,'\nabla_{a_2} q_{14}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{a_1}\hat{q}_{13i}\,'\nabla_{a_2}\hat{q}_{14i} \tag{J-36}$$

$$\hat{E}[\nabla_{a_2} q_{14}\,'\nabla_{a_2} q_{14}] = \frac{1}{n_1}\sum_{i=1}^{n} \nabla_{a_2}\hat{q}_{14i}\,'\nabla_{a_2}\hat{q}_{14i} \tag{J-37}$$

where

$$\nabla_{\alpha_1}\hat{q}_{11i} = \left[\, I(A_i > 0)[1 - \Lambda(\ddot{W}_i\,\hat{\alpha}_1)] - [1 - I(A_i > 0)]\,\Lambda(\ddot{W}_i\,\hat{\alpha}_1)\right]\ddot{W}_i \tag{J-38}$$

$$\nabla_{\alpha_2}\hat{q}_{12i} = 2\,I(A_i > 0)\left(A_i - \exp\left(\ddot{W}_i\,\hat{\alpha}_2\right)\right)\exp\left(\ddot{W}_i\,\hat{\alpha}_2\right)\ddot{W}_i \tag{J-39}$$

$$\nabla_{a_1}\hat{q}_{13i} = \left[I(A_i > 0)[1 - \Lambda(\vec{W}_i\,\hat{a}_1)] - [1 - I(A_i > 0)]\,\Lambda(\vec{W}_i\,\hat{a}_1) \right]\vec{W}_i \qquad \text{(J-40)}$$

and

$$\nabla_{a_2}\hat{q}_{14i} = 2\,I(A_i > 0)\left(A_i - \exp\left(\vec{W}_i\,\hat{a}_2\right)\right)\exp\left(\vec{W}_i\,\hat{a}_2\right)\vec{W}_i \qquad \text{(J-41)}$$

so

$$\hat{E}\left[\nabla_\delta q_1\,'\nabla_\delta q_1\right] = \begin{bmatrix} \hat{E}\begin{bmatrix}\nabla_{\alpha_1}q_{11}\,'\\ \nabla_{\alpha_1}q_{11}\end{bmatrix} & \hat{E}\begin{bmatrix}\nabla_{\alpha_1}q_{11}\,'\\ \nabla_{\alpha_2}q_{12}\end{bmatrix} & \hat{E}\begin{bmatrix}\nabla_{\alpha_1}q_{11}\,'\\ \nabla_{a_1}q_{13}\end{bmatrix} & \hat{E}\begin{bmatrix}\nabla_{\alpha_1}q_{11}\,'\\ \nabla_{a_2}q_{14}\end{bmatrix} \\ \hat{E}\begin{bmatrix}\nabla_{\alpha_2}q_{12}\,'\\ \nabla_{\alpha_1}q_{11}\end{bmatrix} & \hat{E}\begin{bmatrix}\nabla_{\alpha_2}q_{12}\,'\\ \nabla_{\alpha_2}q_{12}\end{bmatrix} & \hat{E}\begin{bmatrix}\nabla_{\alpha_2}q_{12}\,'\\ \nabla_{a_1}q_{13}\end{bmatrix} & \hat{E}\begin{bmatrix}\nabla_{\alpha_2}q_{12}\,'\\ \nabla_{a_2}q_{14}\end{bmatrix} \\ \hat{E}\begin{bmatrix}\nabla_{a_1}q_{13}\,'\\ \nabla_{\alpha_1}q_{11}\end{bmatrix} & \hat{E}\begin{bmatrix}\nabla_{a_1}q_{13}\,'\\ \nabla_{\alpha_2}q_{12}\end{bmatrix} & \hat{E}\begin{bmatrix}\nabla_{a_1}q_{13}\,'\\ \nabla_{a_1}q_{13}\end{bmatrix} & \hat{E}\begin{bmatrix}\nabla_{a_1}q_{13}\,'\\ \nabla_{a_2}q_{14}\end{bmatrix} \\ \hat{E}\begin{bmatrix}\nabla_{a_2}q_{14}\,'\\ \nabla_{\alpha_1}q_{11}\end{bmatrix} & \hat{E}\begin{bmatrix}\nabla_{a_2}q_{14}\,'\\ \nabla_{\alpha_2}q_{12}\end{bmatrix} & \hat{E}\begin{bmatrix}\nabla_{a_2}q_{14}\,'\\ \nabla_{a_1}q_{13}\end{bmatrix} & \hat{E}\begin{bmatrix}\nabla_{a_2}q_{14}\,'\\ \nabla_{a_2}q_{14}\end{bmatrix} \end{bmatrix}. \qquad \text{(J-42)}$$

$\underset{5\times 4K}{\mathbf{E}\left[\boldsymbol{\nabla}_{\gamma\delta}\mathbf{q}\right]}$

Written out explicitly we have

$$\nabla_\gamma q = \begin{pmatrix} (\nabla_{\ddot{\omega}}q_a + 0 + 0 + 0 + 0) \\ (0 + \nabla_{\ddot{v}}q_b + 0 + 0 + 0) \\ (0 + 0 + \nabla_u q_c + 0 + 0) \\ (0 + 0 + 0 + \nabla_v q_d + 0) \\ (0 + 0 + 0 + 0 + \nabla_{m_{\mathcal{P}}} q_e) \end{pmatrix}'$$

$$= [\nabla_{\ddot{\omega}}q_a \quad \nabla_{\ddot{v}}q_b \quad \nabla_u q_c \quad \nabla_v q_d \quad \nabla_{m_{\mathcal{P}}} q_e]$$

$$= 2\left[\left(\ddot{\Omega}(\alpha, \ddot{W}) - \ddot{\omega}\right)\left(\ddot{\mathcal{V}}(\alpha, \ddot{W}) - \ddot{v}\right)\quad\left(U(a, \vec{W}) - \vec{\omega}\right)\left(V(a, \vec{W}) - \vec{v}\right)\quad\left(\mathcal{P} - m_{\mathcal{P}}\right)\right]$$

(J-43)

and

$$E[\nabla_{\gamma\delta} q] = \begin{bmatrix} E[\nabla_{\ddot{\omega}\alpha_1} q_a] & E[\nabla_{\ddot{\omega}\alpha_2} q_a] & E[\nabla_{\ddot{\omega}a_1} q_a] & E[\nabla_{\ddot{\omega}a_2} q_a] \\ E[\nabla_{\ddot{v}\alpha_1} q_b] & E[\nabla_{\ddot{v}\alpha_2} q_b] & E[\nabla_{\ddot{v}a_1} q_b] & E[\nabla_{\ddot{v}a_2} q_b] \\ E[\nabla_{u\alpha_1} q_c] & E[\nabla_{u\alpha_2} q_c] & E[\nabla_{ua_1} q_c] & E[\nabla_{ua_2} q_c] \\ E[\nabla_{v\alpha_1} q_d] & E[\nabla_{v\alpha_2} q_d] & E[\nabla_{va_1} q_d] & E[\nabla_{va_2} q_d] \\ E[\nabla_{m_{\mathcal{P}}\alpha_1} q_e] & E[\nabla_{m_{\mathcal{P}}\alpha_2} q_e] & E[\nabla_{m_{\mathcal{P}}a_1} q_e] & E[\nabla_{m_{\mathcal{P}}a_2} q_e] \end{bmatrix}$$

$$= 2\begin{bmatrix} E[\nabla_{\alpha_1}\ddot{\Omega}] & E[\nabla_{\alpha_2}\ddot{\Omega}] & E[\nabla_{a_1}\ddot{\Omega}] & E[\nabla_{a_2}\ddot{\Omega}] \\ E[\nabla_{\alpha_1}\ddot{\mathcal{V}}] & E[\nabla_{\alpha_2}\ddot{\mathcal{V}}] & E[\nabla_{a_1}\ddot{\mathcal{V}}] & E[\nabla_{a_2}\ddot{\mathcal{V}}] \\ E[\nabla_{\alpha_1} U] & E[\nabla_{\alpha_2} U] & E[\nabla_{a_1} U] & E[\nabla_{a_2} U] \\ E[\nabla_{\alpha_1} V] & E[\nabla_{\alpha_2} V] & E[\nabla_{a_1} V] & E[\nabla_{a_2} V] \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \text{(J-44)}$$

where

$$\nabla_{\alpha_1}\ddot{\Omega} = 2\,\Lambda(\ddot{W}\,\alpha_1)\lambda(\ddot{W}\,\alpha_1)\exp(\ddot{W}\,\alpha_2)\,\ddot{W} \qquad \text{(J-45)}$$

$$\nabla_{\alpha_2}\ddot{\Omega} = \Lambda(\ddot{W}\,\alpha_1)^2\exp(\ddot{W}\,\alpha_2)\ddot{W} \qquad \text{(J-46)}$$

$$\nabla_{\alpha_1}\ddot{\mathcal{V}} = \lambda(\ddot{W}\,\alpha_1)\exp(\ddot{W}\,\alpha_2)\,\ddot{W} \qquad \text{(J-47)}$$

$$\nabla_{\alpha_2}\ddot{\mathcal{V}} = \Lambda(\ddot{W}\,\alpha_1)\exp(\ddot{W}\,\alpha_2)\,\ddot{W} \qquad \text{(J-48)}$$

$$\nabla_{a_1} U = 2\,\Lambda(\vec{W}a_1)\lambda(\vec{W}a_1)\exp(\vec{W}a_2)\,\vec{W} \qquad \text{(J-49)}$$

$$\nabla_{a_2} U = \Lambda(\vec{W}a_1)^2\exp(\vec{W}a_2)\vec{W} \qquad \text{(J-50)}$$

$$\nabla_{a_1} V = \lambda(\vec{W}a_1)\exp(\vec{W}a_2)\,\vec{W} \qquad \text{(J-51)}$$

$$\nabla_{a_2} V = \Lambda(\vec{W}a_1)\exp(\vec{W}a_2)\,\vec{W} \qquad \text{(J-52)}$$

and

$$\nabla_{\ddot{\omega}a} q_a = \nabla_{\ddot{\omega}a} q_b = \nabla_{\ddot{\nu}a} q_a = \nabla_{\ddot{\nu}a} q_b = \nabla_{u\alpha} q_c = \nabla_{u\alpha} q_d = \nabla_{v\alpha} q_c = \nabla_{v\alpha} q_d = \nabla_{m_p\delta} q_e = 0 .$$

(J-53)

Note that

$$\nabla_a \Lambda(a) = \lambda(a) = \Lambda(a)[1 - \Lambda(a)]$$

$$\nabla_a \lambda(a) = \lambda(a)[1 - 2\Lambda(a)] .$$

The requisite consistent matrix estimators are

$$\hat{E}[\nabla_{\alpha_1} \ddot{\Omega}] = \sum_{i=1}^{n} \widehat{\nabla_{\alpha_1} \ddot{\Omega}}_i \tag{J-54}$$

$$\hat{E}[\nabla_{\alpha_2} \ddot{\Omega}] = \sum_{i=1}^{n} \widehat{\nabla_{\alpha_2} \ddot{\Omega}}_i \tag{J-55}$$

$$\hat{E}[\nabla_{\alpha_1} \ddot{\mathcal{V}}] = \sum_{i=1}^{n} \widehat{\nabla_{\alpha_1} \ddot{\mathcal{V}}}_i \tag{J-56}$$

$$\hat{E}[\nabla_{\alpha_2} \ddot{\mathcal{V}}] = \sum_{i=1}^{n} \widehat{\nabla_{\alpha_2} \ddot{\mathcal{V}}}_i \tag{J-57}$$

$$\hat{E}[\nabla_{a_1} U] = \sum_{i=1}^{n} \widehat{\nabla_{a_1} U}_i \tag{J-58}$$

$$\hat{E}[\nabla_{a_2} U] = \sum_{i=1}^{n} \widehat{\nabla_{a_2} U}_i \tag{J-59}$$

$$\hat{E}[\nabla_{a_1} V] = \sum_{i=1}^{n} \widehat{\nabla_{a_1} V}_i \tag{J-60}$$

and

$$\hat{E}[\nabla_{a_2} V] = \sum_{i=1}^{n} \widehat{\nabla_{a_2} V}_i \tag{J-61}$$

where

$$\widehat{\nabla_{\alpha_1}\ddot{\Omega}_i} = 2\,\Lambda(\ddot{W}_i\,\hat{\alpha}_1)\lambda(\ddot{W}_i\,\hat{\alpha}_1)\exp(\ddot{W}_i\,\hat{\alpha}_2)\,\ddot{W}_i \tag{J-62}$$

$$\widehat{\nabla_{\alpha_2}\ddot{\Omega}_i} = \Lambda(\ddot{W}_i\,\hat{\alpha}_1)^2\exp(\ddot{W}_i\,\hat{\alpha}_2)\,\ddot{W} \tag{J-63}$$

$$\widehat{\nabla_{\alpha_1}\ddot{\mathcal{V}}_i} = \lambda(\ddot{W}_i\,\hat{\alpha}_1)\exp(\ddot{W}_i\,\hat{\alpha}_2)\,\ddot{W}_i \tag{J-64}$$

$$\widehat{\nabla_{\alpha_2}\ddot{\mathcal{V}}_i} = \Lambda(\ddot{W}_i\,\hat{\alpha}_1)\exp(\ddot{W}_i\,\hat{\alpha}_2)\,\ddot{W}_i \tag{J-65}$$

$$\widehat{\nabla_{a_1}U_i} = 2\,\Lambda(\vec{W}_i\,\hat{a}_1)\lambda(\vec{W}_i\,\hat{a}_1)\exp(\vec{W}_i\,\hat{a}_2)\,\vec{W}_i \tag{J-66}$$

$$\widehat{\nabla_{a_2}U_i} = \Lambda(\vec{W}_i\,\hat{a}_1)^2\exp(\vec{W}_i\,\hat{a}_2)\,\vec{W} \tag{J-67}$$

$$\widehat{\nabla_{a_1}V_i} = \lambda(\vec{W}_i\,\hat{a}_1)\exp(\vec{W}_i\,\hat{a}_2)\,\vec{W}_i \tag{J-68}$$

and

$$\widehat{\nabla_{a_2}V_i} = \Lambda(\vec{W}_i\,\hat{a}_1)\exp(\vec{W}_i\,\hat{a}_2)\,\vec{W}_i \tag{J-69}$$

so

$$\hat{E}[\nabla_{\gamma\delta}q] = 2\begin{bmatrix} \hat{E}[\nabla_{\alpha_1}\ddot{\Omega}] & \hat{E}[\nabla_{\alpha_2}\ddot{\Omega}] & 0 & 0 \\ \hat{E}[\nabla_{\alpha_1}\ddot{\mathcal{V}}] & \hat{E}[\nabla_{\alpha_2}\ddot{\mathcal{V}}] & 0 & 0 \\ 0 & 0 & \hat{E}[\nabla_{a_1}U] & \hat{E}[\nabla_{a_2}U] \\ 0 & 0 & \hat{E}[\nabla_{a_1}V] & \hat{E}[\nabla_{a_2}V] \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{J-70}$$

$\mathbf{E[\nabla_{\gamma\gamma}q]^{-1}}$
5×5

$$E\left[\nabla_{\gamma\gamma} q\right]^{-1} = \begin{bmatrix} E\left[\nabla_{\ddot{\omega}\ddot{\omega}} q_a\right]^{-1} & 0 & 0 & 0 & 0 \\ 0 & E\left[\nabla_{\ddot{v}\ddot{v}} q_b\right]^{-1} & 0 & 0 & 0 \\ 0 & 0 & E\left[\nabla_{uu} q_c\right]^{-1} & 0 & 0 \\ 0 & 0 & 0 & E\left[\nabla_{vv} q_d\right]^{-1} & 0 \\ 0 & 0 & 0 & 0 & E\left[\nabla_{m_{\mathcal{P}} m_{\mathcal{P}}} q_e\right]^{-1} \end{bmatrix} \tag{J-71}$$

because $\nabla_{ij} q_k = 0$ where $i, j = \{\ddot{\omega}, \ddot{v}, u, v, m_{\mathcal{P}}\}$; $k = \{a, b, c, d, e\}$ and $i \neq j$.

Now

$$\nabla_{\ddot{\omega}\ddot{\omega}} q_a = \nabla_{\ddot{v}\ddot{v}} q_b = \nabla_{uu} q_c = \nabla_{vv} q_d = \nabla_{m_{\mathcal{P}} m_{\mathcal{P}}} q_e = -2$$

therefore

$$E\left[\nabla_{\gamma\gamma} q\right]^{-1} = \begin{bmatrix} -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{2} \end{bmatrix}. \tag{J-72}$$

$$\underset{5\times 5}{\mathbf{E}\left[\boldsymbol{\nabla}_{\gamma}\mathbf{q}'\boldsymbol{\nabla}_{\gamma}\mathbf{q}\right]}$$

Given that

$$\nabla_\gamma q = 2\left[\left(\ddot{\Omega}(\alpha, \ddot{W}) - \ddot{\omega}\right)\ \left(\ddot{\mathcal{V}}(\alpha, \ddot{W}) - \ddot{v}\right)\ \left(U(a, \vec{W}) - u\right)\ \left(V(a, \vec{W}) - v\right)\quad \left(\mathcal{P}_i - m_{\mathcal{P}}\right)\ \right].$$

(J-73)

we have

$$E\left[\nabla_\gamma q' \nabla_\gamma q\right] =$$

$$4\begin{bmatrix}
E\begin{bmatrix}(\ddot{\Omega}(\alpha, \ddot{W})\\ -\ddot{\omega})^2\end{bmatrix} & E\begin{bmatrix}(\ddot{\Omega}(\alpha, \ddot{W})\\ -\ddot{\omega})\\ (\ddot{\mathcal{V}}(\alpha, \ddot{W})\\ -\ddot{v})\end{bmatrix} & E\begin{bmatrix}(\ddot{\Omega}(\alpha, \ddot{W})\\ -\ddot{\omega})\\ (U(a, \vec{W})\\ -u)\end{bmatrix} & E\begin{bmatrix}(\ddot{\Omega}(\alpha, \ddot{W})\\ -\ddot{\omega})\\ (V(a, \vec{W})\\ -v)\end{bmatrix} & E\begin{bmatrix}(\ddot{\Omega}(\alpha, \ddot{W})\\ -\ddot{\omega})\\ (P_i - m_{\mathcal{P}})\end{bmatrix} \\
E\begin{bmatrix}(\ddot{\Omega}(\alpha, \ddot{W})\\ -\ddot{\omega})\\ (\ddot{\mathcal{V}}(\alpha, \ddot{W})\\ -\ddot{v})\end{bmatrix} & E\begin{bmatrix}(\ddot{\mathcal{V}}(\alpha, \ddot{W})\\ -\ddot{v})^2\end{bmatrix} & E\begin{bmatrix}(\ddot{\mathcal{V}}(\alpha, \ddot{W})\\ -\ddot{v})\\ (U(a, \vec{W})\\ -u)\end{bmatrix} & E\begin{bmatrix}(\ddot{\mathcal{V}}(\alpha, \ddot{W})\\ -\ddot{v})\\ (V(a, \vec{W})\\ -v)\end{bmatrix} & E\begin{bmatrix}(\ddot{\mathcal{V}}(\alpha, \ddot{W})\\ -\ddot{v})\\ (P_i - m_{\mathcal{P}})\end{bmatrix} \\
E\begin{bmatrix}(\ddot{\Omega}(\alpha, \ddot{W})\\ -\ddot{\omega})\\ (U(a, \vec{W})\\ -u)\end{bmatrix} & E\begin{bmatrix}(\ddot{\mathcal{V}}(\alpha, \ddot{W})\\ -\ddot{v})\\ (U(a, \vec{W})\\ -u)\end{bmatrix} & E\begin{bmatrix}(U(a, \vec{W})\\ -u)^2\end{bmatrix} & E\begin{bmatrix}(U(a, \vec{W})\\ -u)\\ (V(a, \vec{W})\\ -v)\end{bmatrix} & E\begin{bmatrix}(U(a, \vec{W})\\ -u)\\ (P_i - m_{\mathcal{P}})\end{bmatrix} \\
E\begin{bmatrix}(\ddot{\Omega}(\alpha, \ddot{W})\\ -\ddot{\omega})\\ (\vec{\mathcal{V}}(\alpha, \vec{W})\\ -\vec{v})\end{bmatrix} & E\begin{bmatrix}(\ddot{\mathcal{V}}(\alpha, \ddot{W})\\ -\ddot{v})\\ (\vec{\mathcal{V}}(\alpha, \vec{W})\\ -\vec{v})\end{bmatrix} & E\begin{bmatrix}(U(a, \vec{W})\\ -u)\\ (\vec{\mathcal{V}}(\alpha, \vec{W})\\ -\vec{v})\end{bmatrix} & E\begin{bmatrix}(V(a, \vec{W})\\ -v)^2\end{bmatrix} & E\begin{bmatrix}(V(a, \vec{W})\\ -v)\\ (P_i - m_{\mathcal{P}})\end{bmatrix} \\
E\begin{bmatrix}(\ddot{\Omega}(\alpha, \ddot{W})\\ -\ddot{\omega})\\ (P_i - m_{\mathcal{P}})\end{bmatrix} & E\begin{bmatrix}(\ddot{\mathcal{V}}(\alpha, \ddot{W})\\ -\ddot{v})\\ (P_i - m_{\mathcal{P}})\end{bmatrix} & E\begin{bmatrix}(U(a, \vec{W})\\ -u)\\ (P_i - m_{\mathcal{P}})\end{bmatrix} & E\begin{bmatrix}(V(a, \vec{W})\\ -v)\\ (P_i - m_{\mathcal{P}})\end{bmatrix} & E\begin{bmatrix}(P_i - m_{\mathcal{P}})\\ (P_i - m_{\mathcal{P}})\end{bmatrix}
\end{bmatrix}.$$

(J-74)

The corresponding consistent estimator is

$$\hat{E}\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right] = 4 \begin{bmatrix} \sum_{i=1}^{n} \frac{1}{n} (\hat{\ddot{\Omega}}_i - \hat{\ddot{\omega}})^2 & \sum_{i=1}^{n} \frac{1}{n} (\hat{\ddot{\Omega}}_i - \hat{\ddot{\omega}})(\hat{\ddot{V}}_i - \hat{\ddot{v}}) & \cdots \\ \sum_{i=1}^{n} \frac{1}{n} (\hat{\ddot{\Omega}}_i - \hat{\ddot{\omega}})(\hat{\ddot{V}}_i - \hat{\ddot{v}}) & \sum_{i=1}^{n} \frac{1}{n} (\hat{\ddot{V}}_i - \hat{\ddot{v}})^2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \tag{J-75}$$

Based on (J-8), (J-11) and (J-12) and using the two-stage estimator $\hat{\theta}$ we can consistently estimate (J-7) as

$$\widehat{AVAR(\hat{\theta})} = \begin{bmatrix} \hat{D}_{11} & \hat{D}_{12} \\ \hat{D}_{12}' & \hat{D}_{22} \end{bmatrix}$$

where

$$\hat{D}_{11} = \widehat{AVAR}(\hat{\delta}) = \begin{bmatrix} n\widehat{AVAR}*(\hat{\alpha}_1) & 0 & 0 & 0 \\ 0 & n_1\widehat{AVAR}*(\hat{\alpha}_2) & 0 & 0 \\ 0 & 0 & n\widehat{AVAR}*(\hat{a}_1) & 0 \\ 0 & 0 & 0 & n_1\widehat{AVAR}*(\hat{a}_2) \end{bmatrix}$$

$$D_{12} = -\hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\delta} q_1' \nabla_{\delta} q_1\right] \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma\delta} q\right]' \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

$$D_{22} = \widehat{AVAR(\hat{\gamma})} = \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma\delta} q\right] \widehat{AVAR(\hat{\delta})}\, \hat{E}\left[\nabla_{\gamma\delta} q\right]' \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

$$+ \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1} \hat{E}\left[\nabla_{\gamma} q' \nabla_{\gamma} q\right] \hat{E}\left[\nabla_{\gamma\gamma} q\right]^{-1}$$

and using well known results from asymptotic theory for two-stage estimators, we can show that[54]

$$\widehat{\text{AVAR}(\hat{\theta})}^{-\frac{1}{2}} \begin{bmatrix} \sqrt{n}(\hat{\alpha}_1 - \alpha_1) \\ \sqrt{n_1}(\hat{\alpha}_2 - \alpha_2) \\ \sqrt{n}(\hat{a}_1 - a_1) \\ \sqrt{n_1}(\hat{a}_2 - a_2) \\ \sqrt{n}(\hat{\ddot{\omega}} - \ddot{\omega}) \\ \sqrt{n}(\hat{\ddot{v}} - \ddot{v}) \\ \sqrt{n}(\hat{u} - u) \\ \sqrt{n}(\hat{v} - v) \\ \sqrt{n}(\hat{m}_{\mathcal{P}} - m_{\mathcal{P}}) \end{bmatrix} \xrightarrow{d} N(0, I) . \qquad \text{(J-76)}$$

****

ASIDE:

Notice that the "$\sqrt{n}$ blow up" is a bit tricky here. It implements $\sqrt{n}$ for $\hat{\alpha}_1$, $\hat{\omega}$, $\hat{v}$ and $\hat{a}_1$; but uses $\sqrt{n_1}$ for $\hat{\alpha}_2$ and $\hat{a}_2$. We had to do this because we had to use the correct sample size (viz., n_1) for a number of the components of $\widehat{\text{AVAR}(\hat{\theta})}$ [viz., those that pertained to the estimation of $\hat{\rho}_2$, and $\hat{\alpha}_2$]; in particular (J-17), (J-21), (J-30) through (J-37). For this reason we had to be explicit about the denominators in all of the averages for the components of $\widehat{\text{AVAR}(\hat{\theta})}$. This meant that in the construction of the requisite

[54]Discussions of asymptotic theory for two-stage optimization estimators can be found in Newey & McFadden (1994), White (1994, Chapter 6), and Wooldridge (2010, Chapter 12).

asymptotic t-stats we had to explicitly include the "blow-up" in the numerator (i.e., we had to multiply by the square-root of the appropriate sample size). I refer to this as "tricky" because one typically does not have to do this. In the usual asymptotic t-stat construction the denominators of the averages ("n") need not be included in the construction of the asymptotic covariance matrix because it typically manifests as a multiplicative factor and, after pulling the diagonal and taking the square root to get the standard errors, this multiplicative $\sqrt{n}$ cancels with the "blow-up" factor in the numerator. For example, the asymptotic t-stat of the OLS estimator is

$$\frac{\sqrt{n}(\hat{\rho}_k - \rho_k)}{\sqrt{\widehat{AVAR(\hat{\rho})}}} = \frac{\sqrt{n}(\hat{\rho}_k - \rho_k)}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n}\mathcal{X}'\mathcal{X}_{kk}\right)^{-1}}} = \frac{\sqrt{n}(\hat{\rho}_k - \rho_k)}{\sqrt{n}\sqrt{\hat{\sigma}^2(\mathcal{X}'\mathcal{X}_{kk})^{-1}}} = \frac{(\hat{\rho}_k - \rho_k)}{\sqrt{\hat{\sigma}^2(\mathcal{X}'\mathcal{X}_{kk})^{-1}}}$$

where

n is the sample size

ρ_k is the coefficient of the kth regressor in the linear regression

$\hat{\rho}_k$ is its OLS estimator

σ^2 is the regression error variance estimator

$\mathcal{X}$ is the matrix of regressors

and $\mathcal{X}'\mathcal{X}_{kk}$ is the kth diagonal element of $\mathcal{X}'\mathcal{X}$. Note how the "$\sqrt{n}$ s" simply cancel. Note also that what we typically refer to as the "asymptotic standard error" can actually be written as the square root of the diagonal element of the consistent estimator of the asymptotic covariance matrix divided by n; in other words

$$\text{asy std err} = \sqrt{\frac{\widehat{\text{AVAR}(\hat{\rho})}}{n}}\,.$$

****

Now back to the issue at hand. Moreover

$$\widehat{\text{AVAR}(\hat{\tau})}^{-\frac{1}{2}}\sqrt{n}\begin{bmatrix} \sqrt{n}(\hat{\alpha}_{P1} - \alpha_{P1}) \\ \sqrt{n_1}(\hat{\alpha}_{P2} - \alpha_{P2}) \\ \sqrt{n}(\hat{a}_{\mathcal{P}1} - a_{\mathcal{P}1}) \\ \sqrt{n_1}(\hat{a}_{\mathcal{P}2} - a_{\mathcal{P}2}) \\ \sqrt{n}(\hat{\ddot{\omega}} - \ddot{\omega}) \\ \sqrt{n}(\hat{\ddot{v}} - \ddot{v}) \\ \sqrt{n}(\hat{u} - u) \\ \sqrt{n}(\hat{v} - v) \\ \sqrt{n}(\hat{m}_{\mathcal{P}} - m_{\mathcal{P}}) \end{bmatrix} \xrightarrow{d} N(0, I)\,. \tag{J-77}$$

where

$$\widehat{\text{AVAR}(\hat{\tau})} = \Xi\,\widehat{\text{AVAR}(\hat{\theta})}\,\Xi' \tag{J-78}$$

and τ and Ξ are defined as in (J-2). Now combining (J-1) with (J-77) and (J-78) we get

$$\widehat{\text{avar}(\hat{\eta}^{\text{UPOL}} - \hat{\eta}^{\text{UPO}})}^{-\frac{1}{2}}\sqrt{n}\left[\left(\hat{\eta}^{\text{UPOL}} - \hat{\eta}^{\text{UPO}}\right) - \left(\eta^{\text{UPOL}} - \eta^{\text{UPO}}\right)\right] \xrightarrow{d} N(0, I) \tag{J-79}$$

where

$$\widehat{\text{avar}(\hat{\eta}^{\text{UPOL}} - \hat{\eta}^{\text{UPO}})} = c(\hat{\tau})\;\widehat{\text{AVAR}(\hat{\tau})}\;c(\hat{\tau})'$$

and

$$c(\tau) = \begin{bmatrix} \left(1 - \frac{\ddot{\omega}}{\ddot{\nu}}\right) \\ 1 \\ -m_{\mathcal{P}}\left(1 - \frac{u}{v}\right) \\ -m_{\mathcal{P}} \\ -\alpha_{P1} / \ddot{\nu} \\ \alpha_{P1}\ddot{\omega} / \ddot{\nu}^2 \\ m_{\mathcal{P}}\, a_{\mathcal{P}1} / \nu \\ -m_{\mathcal{P}}\, a_{\mathcal{P}1}\, u / \nu^2 \\ -\left(1 - \frac{u}{v}\right) a_{\mathcal{P}1} - a_{\mathcal{P}2} \end{bmatrix}'.$$

REFERENCES

Alcohol Justice. (2014). Methodology and Faq. *Alcohol Tax/Fee Revenue Calculator*

Angrist, J. D., and Pischke, J. S. (2009). Mostly Harmless Econometrics: An Empiricist's Companion, Princeton university press.

Ayyagari, P., Deb, P., Fletcher, J., Gallo, W., and Sindelar, J. L. (2013). Understanding Heterogeneity in Price Elasticities in the Demand for Alcohol for Older Individuals. *Health Economics,* 22(1), 89-105.

Beer Institute. (2013). Brewers Almanac. Available on http://www.beerinstitute.org/br/beer-statistics/brewers-almanac, retrieved on May 27, 2015.

Blake, D., and Nied, A. (1997). The Demand for Alcohol in the United Kingdom. *Applied Economics,* 29(12), 1655-1672.

Byrnes, J., Shakeshaft, A., Petrie, D., and Doran, C. (2013). Can Harms Associated with High-Intensity Drinking Be Reduced by Increasing the Price of Alcohol?. *Drug and Alcohol Review,* 32(1), 27-30.

C2ER-COLI. (2015). Cost of Living Index. *Council of Community and Econoic Research*

Cameron, L., and Williams, J. (2001). Cannabis, Alcohol and Cigarettes: Substitures or Complements?. *The Economic Record,* 77(236), 19-34.

Chaloupka, F.J., and Laixuthai, A. (1997). Do Youths Substitute Alcohol and Marijuana? Some Econometric Evidence. *Eastern Economic Journal,* 23(3), 253-275.

Chaloupka, F. J., Saffer, H., and Grossman, M. (1993). Alcohol-Control Policies and Motor-Vehicle Fatalities. *The Journal of Legal Studies,* 22(1), 161-186.

Chesson, H., Harrison, P., and Kassler, W. J. (2000). Sex under the Influence: The Effect of Alcohol Policy on Sexually Transmitted Disease Rates in the United States*. *The Journal of Law and Economics,* 43(1), 215-238.

Coate, D., and Grossman, M. (1988). Effects of Alcoholic Beverage Prices and Legal Drinking Ages on Youth Alcohol Use. *Journal of Law and Economics,* 31(1).

Congressional Budget Office. (1990). The Distributional Effects of Increases in Federal Taxes on Tobacco, Alcoholic Beverages, and Motor Fuels. *Federal Taxation of Tobacco, Alcohol Beverages , and Motor Fuels,* Chapter 5, 69-70

Congressional Budget Office. (2013). Option#32 - Options for Reducing the Deficit: 2014 to 2023. Available https://www.cbo.gov/budget-options/2013/44854, retrieved on May 27, 2015

Congressional Budget Office. (2014). Option#71 - Options for Reducing the Deficit: 2015 to 2024. Available https://www.cbo.gov/budget-options/2014/49653, retrieved on May 27, 2015

Cook, P. J., and Durrance, C. P. (2013). The Virtuous Tax: Lifesaving and Crime-Prevention Effects of the 1991 Federal Alcohol-Tax Increase. *Journal of Health Economics,* 32(1), 261-267.

Cook, P. J., and Moore, M. J. (1993). Economic Perspectives on Reducing Alcohol-Related Violence. *Alcohol and Interpersonal Violence: Fostering Multidisciplinary Perspectives,* 24, 193-211.

Cook, P. J., and Moore, M. J. (2002). The Economics of Alcohol Abuse and Alcohol-Control Policies. *Health Affairs,* 21(2), 120-133.

Elder, R. W., Lawrence, B., Ferguson, A., Naimi, T. S., Brewer, R. D., Chattopadhyay, S. K., . . . Fielding, J. E. (2010). The Effectiveness of Tax Policy Interventions for Reducing Excessive Alcohol Consumption and Related Harms. *American Journal of Preventive Medicine,* 38(2), 217-229.

Farrell, S., Manning, W. G., and Finch, M. D. (2003). Alcohol Dependence and the Price of Alcoholic Beverages. *Journal of Health Economics,* 22(1), 117-147.

Fomby, T., Hill, R.C., and S. Johnson (1984): *Advanced Econometrics*, Springer-Verlag.Goel, R. K., & Morey, M. J. (1995). "The interdependence of cigarette and liquor demand." *Southern Economic Journal*, 451-459.

Gallet, C. A. (2007). The Demand for Alcohol: A Meta-Analysis of Elasticities*. *Australian Journal of Agricultural and Resource Economics,* 51(2), 121-135.

Goel, R. K., and Morey, M. J. (1995). The Interdependence of Cigarette and Liquor Demand. *Southern Economic Journal,* 62(2), 451-459.

Grossman, M., Chaloupka, F. J., Saffer, H., and Laixuthai, A. (1994). Effects of Alcohol Price Policy on Youth: A Summary of Economic Research. *Journal of Research on Adolescence,* 4(2), 347-364.

Kenkel, D. S. (1993). Drinking, Driving, and Deterrence: The Effectiveness and Social Costs of Alternative Policies. *Journal of Law and Economics,* 36(2), 877-913.

Kenkel, D. S. (1996). New Estimates of the Optimal Tax on Alcohol. *Economic Inquiry,* 34(2), 296-319.

Kenkel, D. S. (2005). Are Alcohol Tax Hikes Fully Passed through to Prices? Evidence from Alaska. *American Economic Review,* 95(2), 273-277.

Laixuthai, A., and Chaloupka, F. J. (1993). Youth Alcohol Use and Public Policy. *Contemporary Economic Policy,* 11(4), 70-81.

Lee, B., and Tremblay, V. J. (1992). Advertising and the Us Market Demand for Beer. *Applied Economics,* 24(1), 69-76.

Leung, S. F., and Phelps, C. E. (1993). My Kingdom for a Drink...? A Review of Estimates of the Price Sensitivity of Demand for Alcoholic Beverages. *Economics and the Prevention of Alcohol-Related Problems. NIAAA Research Monograph*(25), 1-32.

Levy, D., and Sheflin, N. (1983). New Evidence on Controlling Alcohol Use through Price. *Journal of Studies on Alcohol and Drugs,* 44(06), 929.

Levy, D., and Sheflin, N. (1985). The Demand for Alcoholic Beverages: An Aggregate Time-Series Analysis. *Journal of Public Policy & Marketing,* 4, 47-54.

Manning, W. G., Blumberg, L., and Moulton, L. H. (1995). The Demand for Alcohol: The Differential Response to Price. *Journal of Health Economics,* 14(2), 123-148.

Manning, W. G., Keeler, E. B., Newhouse, J. P., Sloss, E. M., and Wasserman, J. (1989). The Taxes of Sin: Do Smokers and Drinkers Pay Their Way?. *JAMA,* 261(11), 1604-1609.

Markowitz, S., and Grossman, M. (1998). Alcohol Regulation and Domestic Violence Towards Children. *Contemporary Economic Policy,* 16(3), 309-320.

Markowitz, S., and Grossman, M. (2000). The Effects of Beer Taxes on Physical Child Abuse. *Journal of Health Economics,* 19(2), 271-282.

Mullahy, J., and Sindelar, J. L. (1994). Do Drinkers Know When to Say When? An Empirical Analysis of Drunk Driving. *Economic Inquiry,* 32(3), 383-394.

Mullahy, J., and Sindelar, J. (1996). Employment, Unemployment, and Problem Drinking. *Journal of Health Economics,* 15(4), 409-434.

Mullahy, J. (1998). Much Ado About Two: Reconsidering Retransformation and the Two-Part Model in Health Econometrics. *Journal of Health Economics,* 17(3), 247-281.

Nelson, J. P. (1990). State Monopolies and Alcoholic Beverage Consumption. *Journal of Regulatory Economics,* 2(1), 83-98.

Nelson, J. P. (2014). Estimating the Price Elasticity of Beer: Meta-Analysis of Data with Heterogeneity, Dependence, and Publication Bias. *Journal of Health Economics,* 33, 180-187.

Newey, W.K. and McFadden, D. (1994): Large Sample Estimation and Hypothesis Testing, *Handbook of Econometrics*, Engle, R.F., and McFadden, D.L., Amsterdam: Elsevier Science B.V., 2111-2245, Chapter 36.

NIAAA. (2015). "FY 2015 Financial Management Plan. *National Institute on Alcohol Abuse and Alcoholism*, Available on http://www.niaaa.nih.gov/grant-funding/management-reporting/financial-management-plan, retrieved on May 27, 2015

Ornstein, S. I. (1980). Control of Alcohol Consumption through Price Increases. *Journal of Studies on Alcohol and Drugs,* 41(09), 807.

Patra, J., Bakker, R., Irving, H., Jaddoe, V. W., Malini, S., & Rehm, J. (2011). Dose–response relationship between alcohol consumption before and during pregnancy and the risks of low birthweight, preterm birth and small for gestational age (SGA)—a systematic review and meta-analyses. *BJOG: An International Journal of Obstetrics & Gynaecology, 118*(12), 1411-1421.

Pogue, T. F., and Sgontz, L. G. (1989). Taxing to Control Social Costs: The Case of Alcohol. *The American Economic Review,* 79(1), 235-243.

Ruhm, C. J. (1996). Alcohol Policies and Highway Vehicle Fatalities. *Journal of Health Economics,* 15(4), 435-454.

Ruhm, C. J., Jones, A. S., McGeary, K. A., Kerr, W. C., Terza, J. V., Greenfield, T. K., and Pandian, R. S. (2012). What Us Data Should Be Used to Measure the Price Elasticity of Demand for Alcohol?. *Journal of Health Economics,* 31(6), 851-862.

Sloan, F. A., Reilly, B. A., and Schenzler, C. (1994). Effects of Prices, Civil and Criminal Sanctions, and Law Enforcement on Alcohol-Related Mortality. *Journal of Studies on Alcohol and Drugs,* 55(4), 454.

Terza, J. V. (2002). Alcohol Abuse and Employment: A Second Look. *Journal of Applied Econometrics,* 17(4), 393-404.

Terza, J. V. (2014). Health Policy Analysis from a Potential Outcomes Perspective: Smoking During Pregnancy and Birth Weight. *Unpublished manuscript, Department of Economics, Indiana University Purdue University Indianapolis*.

Terza, J. V., Jones, A. S., Devaraj, S., Greenfield, T. K., Kerr, W. C., McGeary, K. A., and Ruhm, C. J. (2015). Specification and Estimation of Own Price Elasticity in

Two-Part Models of Alcohol Demand. *Unpublished manuscript, Department of Economics, Indiana University Purdue University Indianapolis*.

Wagenaar, A. C., Salois, M. J., and Komro, K. A. (2009). Effects of Beverage Alcohol Price and Tax Levels on Drinking: A Meta-Analysis of 1003 Estimates from 112 Studies. *Addiction,* 104(2), 179-190.

White, H. (1994): *Estimation, Inference and Specification Analysis*, Cambridge: Cambridge University Press.

Wilkinson, J. T. (1987). Reducing Drunken Driving: Which Policies Are Most Effective?. *Southern Economic Journal,* 54(2), 322-334.

Williams, J.; Chaloupka, F.J.; and Wechsler, H. (2005). Are there differential effects of price and policy on college students' drinking intensity?. *Contemporary Economic Policy,*23, (78-90).

Wooldridge, J.M. (2010): *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.

Young, D. J., and Bielińska-Kwapisz, A. (2002). Alcohol Taxes and Beverage Prices. *National Tax Journal,* 55(1), 57-73.

Young, D. J., and Bielinska-Kwapisz, A. (2003). Alcohol Consumption, Beverage Prices and Measurement Error. *Journal of Studies on Alcohol and Drugs,* 64(2), 235.

.

CURRICULUM VITAE

SRIKANT DEVARAJ

Education

Ph.D.	Economics, Indiana University, Indianapolis, February 2016
M.B.A.	Concentration in Finance and Operations, Ball State University, Muncie, July 2009
M.S.	Information and Communication Sciences, Ball State University, Muncie, July 2008
B.E.	Mechanical Engineering, University of Madras, India, May 2004

Certifications

Certificate in Business and Life Sciences (CBLS), Kelly School of Business, IU Bloomington, 2014
Project Management Professional, Project Management Institute, 2009

Employment

Research Economist & Research Assistant Professor, Center for Business and Economic Research, Ball State University, Aug 2015 – present
Senior Research Associate and Project Manager, Center for Business and Economic Research, Ball State University, Jul 2015–Aug 2015
Research Associate, Center for Business and Economic Research, Ball State University, May 2009–Jun 2010
Graduate Research Assistant, Insight & Research, Center for Media Design, Ball State University, Jan 2008–May 2009
Senior Project Engineer, VA Tech Wabag Ltd, India, Apr 2007–Jul 2007
Project Engineer, VA Tech Wabag Ltd, India, Apr 2006–Mar 2007
Graduate Engineer Trainee, VA Tech Wabag Ltd, India, Jun 2004–Mar 2006

Publications

1. Hicks, M.J., LaFaive, M., & Devaraj, S. Right to work effects on total factor productivity and population growth. *Cato Journal* (forthcoming)
2. Devaraj, S., & Patel, P.C. (2016). Attributing responsibility: Hospitals account for 20% of variance in Acute Myocardial Infarction patient mortality. *Journal of Healthcare Quality*, 38(1), 52-61.
3. Devaraj, S., Sharma, S, Fausto,D., Viernes, S., &Kharrazi, H. (2014). Barriers and facilitators to clinical decision support systems adoption: A systematic review. *Journal of Business Administration Research*, 3(2), 36-53.

4. Devaraj, S., Sharma, S., Hicks, M.J., & Faulk, D,G. (2014). The human development index of Indiana counties – An exploratory study. *International Journal of Business and Economic Development*, 2(1), 1-15.
5. Bott, J., Faulk, D., Guntupalli, A., Devaraj, S., & Holmes, M. (2011). An examination of generational differences and media exposure. *The Journal of Applied Management and Entrepreneurship*, 16(4), 78-100.

Conferences/Presentations

1. Exploring sources of bias in the specification and estimation of own-price elasticity in parametric models of alcohol demand, IUPUI, Dec 2015.
2. Potential Outcome and Conventional Estimators of Own Price Alcohol Demand Elasticity in the Two-Part Modeling Context: A Comparison, IUPUI, Dec 2014.
3. International conference on business and economic development, New York, Mar 2014.

Awards and Fellowships

Graduate Professional Scholarship (Tuition/fees), IUPUI, 2011 to present.
Member of runner-up team for developing reasonable innovative solution proposal and implementation strategy for a life science business issue, CBLS, Indiana University, May 2014.
Member of winning team in developing a reasonable best case commercialization path for a life science company's intellectual property, CBLS, Indiana University, Dec 2013.
Certificate of Appreciation from US Army Corps of Engineers, Aug 2010.
Service Learning Award, Center for Information and Communication Sciences, Ball State University, Jul 2008.
Graduate Assistantship (Tuition/fees/stipend), Ball State University, Jan 2008 to May 2009.
Star Performer Award, VA Tech Wabag Ltd, India, Feb 2007.
Scholarship from college for securing top position in the undergraduate studies, May 2004.