CAUSAL ANALYSIS USING TWO-PART MODELS: A GENERAL FRAMEWORK FOR SPECIFICATION, ESTIMATION AND INFERENCE

Zhuang Hao

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

August 2018

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

Doctoral Committee

____
Joseph V. Terza, Ph.D., Chair

____
Srikant Devaraj, Ph.D.

June 22, 2018

____
Ziyue Liu, Ph.D.

____
Henry Mak, Ph.D.

____
Mark Ottoni-Wilhelm, Ph.D.

DEDICATION

To my beloved parents, Chunxiang Hao and Jie Liu

ACKNOWLEDGEMENT

First and foremost, I would like to express my sincerest gratitude to my advisor Professor Joseph V. Terza for his continuous support of my research, for his patience and motivation, for his career advice, and for his sense of humor. I have been learning how generous, caring, and hard-working an advisor can be from Professor Terza since I met him. I could not have imagined a better academic advisor for my Ph.D. study. Besides my advisor, I am indebted to the rest of my thesis committee: Professor Srikant Devaraj, Professor Ziyue Liu, Professor Henry Mak, and Professor Mark Ottoni-Wilhelm for their helpful comments and discussions during my dissertation work. Thank you.

My sincere thanks also go to Professor Vidhura Tennekoon who has always been there for me with abundant guidance and expertise. He was and remains my best role model for a researcher. I also thank Professor Steven Russell for his enthusiasm for teaching and providing me countless tips on what works and what doesn’t in the classroom, which is exceptionally beneficial to my three-year teaching experience. Besides the research, teaching has been my favorite part on the Ph.D. journey. I am definitely grateful to each and every faculty member in our department for their scientific knowledge and insightful suggestions during these past five years. Thank you.

My time in Indiana was made enjoyable essentially by friends and classmates who have already become a significant part of my life. I feel blessed to have them around in many memorable moments. While names are too numerous to list here, I want to at least mention Taul Cheong, Sanghoon Lee, Suitao Yang, Shichao Tang, Oindrila Bhattacharya, Qiaoni Shi, and Kai Zhu. I have known each for quite a while and I have really enjoyed spending time with them. Thank you.

Last but not least, I would like to thank my parents and other family members for showing me all their true love and reminding me of their strong beliefs that nothing is impossible. It would not have been possible for me to keep chasing dreams without their persistent support and encouragement. Thank you.

Zhuang Hao

CAUSAL ANALYSIS USING TWO-PART MODELS: A GENERAL FRAMEWORK FOR SPECIFICATION, ESTIMATION AND INFERENCE

The two-part model (2PM) is the most widely applied modeling and estimation framework in empirical health economics. By design, the two-part model allows the process governing observation at zero to systematically differ from that which determines non-zero observations. The former is commonly referred to as the extensive margin (EM) and the latter is called the intensive margin (IM). The analytic focus of my dissertation is on the development of a general framework for specifying, estimating and drawing inference regarding causally interpretable (CI) effect parameters in the 2PM context. Our proposed fully parametric 2PM (FP2PM) framework comprises very flexible versions of the EM and IM for both continuous and count-valued outcome models and encompasses all implementations of the 2PM found in the literature. Because our modeling approach is potential outcomes (PO) based, it provides a context for clear definition of targeted counterfactual CI parameters of interest. This PO basis also provides a context for identifying the conditions under which such parameters can be consistently estimated using the observable data (via the appropriately specified data generating process). These conditions also ensure that the estimation results are CI. There is substantial literature on statistical testing for model selection in the 2PM context, yet there has been virtually no attention paid to testing the "one-part" null hypothesis. Within our general modeling and estimation framework, we devise a relatively simple test of that null for both continuous and count-valued outcomes. We illustrate our

proposed model, method and testing protocol in the context of estimating price effects on the demand for alcohol.

Joseph V. Terza, Ph.D., Chair

TABLE OF CONTENTS

List of Tables xii
Chapter 1. Introduction, Background and Significance, Summary 1
Chapter 2. The Fully Parametric Two-Part Model (FP2PM) and Causal Inference in the Potential Outcomes (PO) Framework 5
2.1 Specifying the Parameter of Interest in the PO Framework 6
2.2 The General FP2PM PO Framework: The Relevant Conditional Probability Density/Mass Function (pdf/pmf) and Corresponding Conditional Mean Function 10
2.2.1 Extensive Margin (EM) 12
2.2.2 Intensive Margin (IM) 12
2.2.3 Pdf (Pmf) of the Continuous (Count-Valued) PO 13
2.2.4 Conditional Mean Function for the PO 14
2.2.5 Conventional Two-Part Model 15
2.2.6 Examples 16
Continuous Outcome Conventional Two-Part Model – Log-Normal IM 16
Count-Valued Outcome Conventional Two-Part Model – Poisson IM 17
2.2.7 The Generic Two-Part PO Specification Revisited 17
Continuous Outcome Models 18
Count-Valued Outcome Models 18
Chapter 3. Estimating Targeted Causal Parameters of Interest in the FP2PM PO Framework 20
3.1 Reconciling the PO with the Data Generating Process (DGP) in the FP2PM Framework 20
3.2 Full Information Maximum Likelihood Estimation of the Deep Parameters 22
3.2.1 Log-Likelihood Function and Conditional Mean Function – Continuous Model 22
Case I: Continuous Conventional Two-Part Model with Probit EM and Log-Normal IM 23
Case II: Continuous Conventional Two-Part Model with Probit EM and Generalized Gamma IM 24
3.2.2 Log-Likelihood Function and Conditional Mean Function – Count-Data

Model 26
Case III: Count-Valued Conventional Two-Part Model with Probit EM and Poisson IM 26
Case IV: Count-Valued Conventional Two-Part Model with Probit EM and Conway-Maxwell Poisson IM 27
Chapter 4. Testing the Non-Two-Part Model Null 30
4.1 No Structural Difference Between the EM and the IM (NSD) 31
4.1.1 Log-Likelihood Function and Conditioinal Mean Function – Continuous Model with NSD 31
Case V: Continuous Two-Part Model with Log-Normal EM and IM (NSD) 32
Case VI: Continuous Two-Part Model with Generalized Gamma EM and IM (NSD) 33
4.1.2 Log-Likelihood Function and Conditional Mean Function – Count-Data Model with NSD 34
Case VII: Count-Data Two-Part Model with Poisson EM and IM (NSD) 34
Case VIII: Count-Data Two-Part Model with Conway-Maxwell Poisson EM and IM (NSD) 35
4.2 Exploring the Robustness of Maintaining NSD 37
4.2.1 Robustness of NSD in a Continuous Two-Part Model: Case II vs. Case VI 37
Analytic Argument for Robustness of Case VI vis-a-vis Case II 38
Approximate MLE for Case VI 38
Simulating Case VI Data 41
Validating the Consistency of the Approximate MLE of the AIE Using Simulated Data 44
4.2.2 Robustness of NSD in a Count-Valued Two-Part Model: Case IV vs. Case VIII 47
Simulating Case IV Data 47
Investigating the Robustness of the NSD (Case IV vs. Case VIII) Using Simulated Data 49

4.3 A Test of No Parametric Difference between EM and IM with NSD Maintained 51
4.3.1 Generalized Gamma Model with NSD under One-Part Null 51
4.3.2 Conway-Maxwell Poisson Model with NSD under One-Part Null 53
Chapter 5. An Application: Estimating Price Effects on Alcohol Demand 60
5.1 Data 64
5.2 Empirical Results 65
5.3 Likelihood Ratio Test on One-Part Null 66
Chapter 6. Summary, Discussion and Conclusions 70
Appendices 72
Appendix I. A Brief Note on "Cake Debate" 72
Appendix II. Three Lemmas 74
References......78
Curriculum Vitae

LIST OF TABLES

Table 1: Simulation Results of Continuous 2PM with NSD 54

Table 1-(1): $\{\kappa_2, \sigma_2\} = \{1.5, 2\}$ 54

Table 1-(2): $\{\kappa_2, \sigma_2\} = \{2, 2\}$ 55

Table 1-(3): $\{\kappa_2, \sigma_2\} = \{1, 2\}$ 56

Table 1-(4): $\{\kappa_2, \sigma_2\} = \{0.5, 2\}$ 57

Table 2: Simulation Results of Count-Valued 2PM with NSD 58

Table 2-(1): Simulation Parameter Design of Count-Valued 2PM with NSD 58

Table 2-(2): Simulation Results of Count-Valued 2PM with NSD 59

Table 3: Descriptive Statistics 67

Table 4: Regression Results 68

Table 5: Estimated Price Effects on Alcohol Consumption 69

Chapter 1.

Introduction, Background and Significance, Summary

The two-part model (2PM), introduced by Cragg (1971), is the most widely applied empirical modeling and estimation framework in empirical health economics and is gaining in popularity in a variety of other fields.[1] It applies to cases in which the outcome of interest is nonnegative with a non-trivial probability of having an observed value of zero. By design, the 2PM allows the process governing observation at zero (e.g. whether or not the individual drinks alcohol; whether or not the individual decides to visit a health care facility) to systematically differ from that which determines non-zero observations (e.g. how much the individual drinks given that he has chosen to become a drinker; how much the individual spends on health care if he or she spends at all). The former is commonly referred to as the *extensive margin* (EM) and the latter is called the *intensive margin* (IM).

The objectives of the dissertation are four-fold. First, we place our analytic focus on specifying, estimating and drawing inference regarding causally interpretable (CI) policy effect parameters in the 2PM context. Empirical modeling often begins with specification of relevant aspects of the data generating process (DGP) – e.g. the

[1] Applications of the 2PM are too numerous to list. Since (Duan et al. 1983), 2PMs have been conducted in a wide range of health care and service research recently (Burney et al. 2016, Hyun et al. 2016, Li et al. 2016, Liu et al. 2010, Madden 2008, Morozumi and Ii 2006, Buntin and Zaslavsky 2004, Ross and Chaloupka 2003, and Bradford et al. 2002). There are also numerous applications of the 2PM in health related fields, like medical research and biostatistics (Fang et al. 2017, Taylor and Pollard 2009, Kim and Muthén 2009, and Han and Kronmal 2006). 2PM is also a commonly used framework in many other applied economics literature such as agricultural economics (Chang and Meyrhoefer 2016 and Hertz 2010), demographic economics (Gurmu and Trivedi 1996), tourism economics (Arulampalam and Booth 1997), and financial economics (see Brown et al. 2015 and J.S. Ramalho and Silva 2009).

probability density function (pdf) [or, the probability mass function (pmf)] of the observed outcome of interest conditional on a vector of observed covariates; or, the conditional mean of the outcome. Here we diverge from usual practice by commencing the modeling discussion at a deeper level – in the potential outcomes (PO) framework. By beginning the modeling at the PO level, we can clearly and rigorously specify the parameter of interest (the estimation objective of most applied economic research) in a way to ensure that it is CI.

Secondly, we propose a very general fully parametric 2PM (FP2PM) framework for the PO and DGP that encompasses all continuous and count-valued outcome studies found in the literature. Within this encompassing framework, we specify very general and flexible versions of the EM and IM for both the continuous and count-valued cases. Moreover, our proposed framework obviates the "cake debates"[2] of the early and mid-1980's.

Third, not only does our PO-based framework allow us to clearly specify the targeted counterfactual CI parameter of interest, but it also provides a context where to derive the requisite conditions under which this parameter can be consistently estimated using the observable data (via the appropriately specified DGP). By the same token, these conditions ensure that the estimation results are CI. We note that said conditions

[2] "Cake debates" refers to a thread of econometrics literature debating on the relative strengths and weaknesses of two frequent approaches dealing with limited dependent variables: the Sample Selection Model by Heckman (1976, 1979) and the 2PM by Duan et al. (1983, 1984, 1985). The studies are not conclusive in the literature, however the simulation results slightly favor the 2PM (see Dow and Norton 2003, Leung and Yu 1996, Jones 1989, Manning, Duan and Rogers 1987, Hay, Leu and Rohrer 1987 and Hay and Olsen 1984). This paper does not attempt to further discuss the "cake debates" but the general framework provided here incorporates both models. A brief note on "cake debates" can be found in Appendix I.

are not generally satisfied but are seldom discussed. Thus, this aspect of our analysis constitutes an important contribution to applied research in health economics.

Finally we note that, although there has been much discussion in the literature regarding statistical testing and inference in the 2PM context, there has been virtually no attention paid the most important of all to the relevant null hypotheses: [3]

H_0 : No substantive EM/IM distinction – a two-part structure is not needed.

Within our general modeling and estimation framework, as an extension of the approach by Mullahy (1986), we devise a relatively simple test of the above null hypothesis.

The remainder of the dissertation is organized as follows.

In Chapter 2, we cast the 2PM in the PO context and thereby explicitly specify the targeted counterfactual CI policy effect parameter of interest. We propose a general FP2PM PO framework, which is very flexible, can be used for both continuous and count-valued outcomes, and allows for an easy-to-implement statistical test on 2PM specification in empirical context (H_o above). We also specify two commonly encountered policy effect parameters [average incremental effect (AIE) and average price elasticity of demand (AED)] in the FP2PM.

In Chapter 3, we detail the conditions under which the counterfactual CI parameters of interest (here, AIE and AED) can be estimated with observed (factual) data using the appropriately specified DGP. Therein we also show that the conventional

[3] There are some other tests proposed in the 2PM context. Examples include Taylor and Pollard (2009)'s test on equality of two distributions in the 2PM setting, Santos Silva and Winmeijer (2001)'s test on single spell hypothesis, and Leung and Yu (1996)'s test on model choice between Sample Selection Model and 2PM. But none of them explicitly considers the misspecification issue of the 2PM structure. The misspecification test proposed by Mullahy (1986) in the count-valued data is an exception, however, the Mullahy's test was not paid much attention by the literature.

versions of the FP2PM for the continuous outcomes and the count-valued can be cast as special cases of our general framework. We also offer details for the very general 2PM versions for both the continuous and count-valued cases (models that accommodate the likelihood flexibility in the continuous case and omni-dispersion in the count-valued case). Along with four conventional FP2PM, this chapter concludes with formulations for the key conditional mean functions corresponding with the general model and all of its aforementioned particular versions (the conventional and flexible models for the continuous and count-valued cases).

Chapter 4 details the affordable likelihood ratio test on the 2PM specification as H_o above. In order to develop the test, we propose four versions (2 cases for continuous outcomes and 2 cases for count-valued outcomes) of the general FP2PM PO framework assuming there is no structural difference between the EM and IM (NSD). The FP2PM with the NSD assumption allows for the 2PM specification, however, the relative robustness of these models remains unclear. We make use of several simulation studies to provide the evidence that imposition of the FP2PM with NSD does not materially affect the estimation of policy effect parameters. At the end of the chapter, we cast the 2PM specification test for count-valued outcomes by Mullahy (1986) into our FP2PM with NSD and extend the test to continuous case.

Chapter 5 illustrates the aforementioned models, methods, and testing protocols in the context of estimating price effects on the demand for alcohol using a real dataset.

Chapter 6 summarizes and concludes the dissertation.

Chapter 2.

The Fully Parametric Two-Part Model (FP2PM) and Causal Inference in the Potential Outcomes (PO) Framework

The main motivation for nearly all empirical economic research is to provide scientific evidence that can be used to assess past, current, and future policy.[4] Essential to such assessments is the rigorous specification and accurate estimation of parameters that characterize the causal relationship between a policy variable of interest, which to some degree is (or can be brought) under the control of a policy maker, and a specified outcome of policy interest. The PO framework which takes account of the counterfactual nature of such effect analyses provides a means of clearly and coherently defining the relevant parameters such that they are causally interpretable.

This chapter will present the general PO framework and give examples of two CI parameters that are relevant estimation objectives in a variety of empirical contexts. We then show how these counterfactual parameters can be rewritten so as to make them amenable to estimation via observable (factual) data. Next, the details of general PO framework as it pertains to 2PM are given for the continuous and count-valued outcome versions of the model. For each of these versions, the conventional specification is shown to be a special case of the general framework. We also detail flexible variants of the continuous and count-data models (a flexible likelihood specification for continuous outcomes and an omni-dispersed specification for count-valued models). The overarching goal of this chapter is the specification of the key conditional mean function to be implemented in rewriting the relevant CI parameter in a way that makes it estimable

[4] We use the term policy very broadly to mean any exogenous action taken by an economic agent aimed at achieving a specified effect.

using observable data (as mentioned earlier). The chapter concludes by offering a very general statistic, based on this conditional mean formulation, which is consistent for the CI parameter of interest. The implementation of this statistic, however, is predicated on the existence of a consistent estimator for the "deep" parameters of the underlying 2PM. Consistent estimation of such deep parameters in the context of 2PM is the subject of Chapter 3.

2.1 Specifying the Parameter of Interest in the PO Framework[5]

The focus here are the rigorous specification and the accurate estimation, in a 2PM context, of a parameter that characterizes the causal relationship between a policy variable of interest ($\mathbf{X_p}$), which to some degree is (or can be brought) under the control of a policy maker, and a specified outcome of policy interest ($\mathbf{Y}$).[6] In 2PMs, it is typical that the observed data on $\mathbf{Y}$ is characterized by a large proportion of zeros and it is reasonable to believe that the component of the structural model and data generating process (DGP) pertaining to such null values is distinguishable from the other aspects of the model and DGP. Later, to illustrate our proposed methods and tests, we consider the case in which $\mathbf{Y}$ is beer consumption and $\mathbf{X_p}$ is the market price of beer. Because there is a substantial proportion of individuals in the population who do not drink beer and because the decision and motivation to be a beer drinker (or non-drinker) may be systematically distinguishable (i.e., modeled differently from the decision as to how

[5] See Terza (2018) for a detailed and more general discussion of the potential outcomes framework.

[6] $\mathbf{X_p}$ and $\mathbf{Y}$ are to be taken as global replacements for the phrases "policy variable of interest" and "outcome of policy interest", respectively.

much beer to drink if one is a beer drinker), the alcohol demand analysis is best cast in a 2PM context.

We first draw the distinction between two versions of $\mathbf{X_p}$:[7]

$X_p \equiv$ the random variable representing the observable (*factual*) version of the distribution of $\mathbf{X_p}$ (sampled values of the policy variable are drawn from the distribution of X_p)

and,

$X_p^* \equiv$ the random variable representing the hypothetical (*counterfactual*) exogenously mandated version of the distribution of $\mathbf{X_p}$ that might result from a policy intervention (X_p^* is, by design, independent of all other variates germane to the present discussion).[8]

Likewise, we distinguish two versions of $\mathbf{Y}$:

$Y \equiv$ the random variable with support of $[0, \infty)$ representing the factual version of the distribution of $\mathbf{Y}$ (the sampled values of the outcome are drawn from the distribution of Y)

and,

[7] Henceforth, we will adhere to the following notational conventions: (1) uppercase letters for random variables (e.g., A, B, Z); (2) lowercase letters for particular values in the support of the random variable in a specific question (e.g., a, b, z); and (3) uppercase letters with a subscript "i" for the sampled version of the random variables in a specific question (e.g. A_i, B_i, Z_i with i = 1, …, n indicating the i^{th} observation from a sample of size n).

[8] In the present context, we use the term counterfactual in reference to random variables that are not able to be, to some extent, observed.

$Y_{X_p^*}$ ≡ the random variable with support of [0, ∞) representing the distribution of *potential outcome*, defined as the distribution of values of $\mathbf{Y}$ that would have manifested for a particular X_p^* (an exogenously mandated version of $\mathbf{X_p}$).

Throughout the remainder of the discussion, we will explicitly and implicitly reference a hypothetical (*counterfactual*) policy intervention in which $\mathbf{X_p}$ is exogenously changed from X_p^{pre} to X_p^{post} (from pre-policy to post-policy). Without loss of generality, we write $X_p^{post} = X_p^{pre} + \Delta$, where Δ is an observable random variable representing the policy induced (i.e., *exogenous*) change in the distribution of $\mathbf{X_p}$ for the relevant population. Note that X_p^{pre} and $X_p^{pre} + \Delta$ are both specific versions of X_p^*, therefore, they are independent of all other variates germane to the discussion. So is Δ.

In our illustrative empirical analysis of alcohol demand presented in Chapter 5, we will focus on the following two common causal parameters cast in the above PO framework:

average incremental effect of price on demand for alcohol (AIE)

$$AIE(\Delta) = E[Y_{X_p^{pre} + \Delta}] - E[Y_{X_p^{pre}}] \tag{1}$$

and

average price elasticity of demand for alcohol (AED)

$$AED(\Delta) = AIE(\Delta) \times \frac{E[X_p^{pre}]}{E[Y_{X_p^{pre}}]} . \tag{2}$$

Unfortunately, parameters defined like (1) and (2) cannot be estimated from data straightforwardly because both $Y_{X_p^{pre}}$ and $Y_{X_p^{pre}+\Delta}$ are counterfactual; in other words, they do not represent observable statistical populations from which samples are drawn. Let us suppose, however, that the continuous (count-valued) potential outcome $Y_{X_p^*}$ has the following conditional probability density function [pdf] (probability mass function [pmf]) given a vector of observable covariates X_o,

$$\text{pdf}(Y_{X_p^*} \mid X_o) = f_{(Y_{X_p^*}|X_o)}(Y_{X_p^*}, X_p^*, X_o; \pi) \tag{3}$$

where, $f_{(Y_{X_p^*}|X_o)}(\ )$ has a known form and π is a vector of unknown (or, “deep”) parameters. It follows from (3) that

$$E[Y_{X_p^*} \mid X_o] = m(X_p^*, X_o; \pi)\,. \tag{4}$$

Note that the conditional mean function has a known form. Using the law of iterated expectations as well as (4), we can rewrite (1) and (2) as

$$AIE(\Delta) = E[m(X_p^{pre} + \Delta, X_o; \pi)] - E[m(X_p^{pre}, X_o; \pi)] \tag{5}$$

and

$$AED(\Delta) = \left(E[m(X_p^{pre} + \Delta, X_o; \pi)] - E[m(X_p^{pre}, X_o; \pi)]\right) \times \frac{E[X_p^{pre}]}{E[m(X_p^{pre}, X_o; \pi)]}\,. \tag{6}$$

We now introduce a very general 2PM framework which encompasses all the 2PMs in the literature, and the estimation strategy for the counterfactual policy effect parameters (5) and (6) using observable data.

2.2 The General FP2PM PO Framework: The Relevant Conditional Probability Density/Mass Function (pdf/pmf) and Corresponding Conditional Mean Function

To fix ideas and to motivate the formulation of our proposed general 2PM specification, we begin the discussion by casting the "classical" continuous 2PM of Cragg (1971) in the PO framework. Here the EM and IM are specified, respectively, as:

$$Y_{X_p^*} = 0 \qquad \text{iff} \qquad X_p^*\beta_{1p} + X_o\beta_{1o} + \varepsilon^{EM} \leq 0 \tag{7}$$

$$\ln(Y_{X_p^*}) \mid Y_{X_p^*} > 0 = X_p^*\beta_{2p} + X_o\beta_{2o} + \varepsilon^{IM} \qquad \text{iff} \qquad X_p^*\beta_{1p} + X_o\beta_{1o} + \varepsilon^{EM} > 0 \tag{8}$$

where, $\left(\ln(Y_{X_p^*}) \mid Y_{X_p^*} > 0\right)$ denotes the log of the observed strictly positive outcome values, i.e., $\left(\ln(Y_{X_p^*}) \mid Y_{X_p^*} > 0\right)$ is defined if and only if $X_p^*\beta_{1p} + X_o\beta_{1o} + \varepsilon^{EM} > 0$; $\beta_1' = [\beta_{1p}\ \beta_{1o}']$ and $\beta_2' = [\beta_{2p}\ \beta_{2o}']$ are vectors of coefficient parameters; and, $(\varepsilon^{EM} \mid X_o)$ and $(\varepsilon^{IM} \mid X_o)$ have known distributions. Typically, $(\varepsilon^{EM} \mid X_o)$ is assumed to be standard normally distributed and $(\varepsilon^{IM} \mid X_o)$ is taken to be normally distributed with mean of 0 and variance of σ^2 . Under these assumptions, the relevant specification for the potential outcome $Y_{X_p^*}$ is,

$$f_{(Y_{X_p^*}|X_o)}(Y_{X_p^*}, X_p^*, X_o; \pi) = [1-\Phi(X_p^*\beta_{1p} + X_o\beta_{1o})]^{I(Y_{X_p^*}=0)}$$
$$\times\left(\Phi(X_p^*\beta_{1p} + X_o\beta_{1o})\ \varphi_{\ln}(Y_{X_p^*};\ X_p^*\beta_{2p} + X_o\beta_{2o}, \sigma^2)\right)^{1-I(Y_{X_p^*}=0)} \qquad (9)$$

where, $\Phi(\ )$ denotes the cumulative distribution function (cdf) of the standard normal distribution, $\varphi_{\ln}(a;b,c)$ denotes the log-normal pdf with the argument a and location and scale parameters of b and c, and $\pi' = [\beta_1' \quad \beta_2' \quad \sigma^2]$ is the vector of parameters. Henceforth, we will refer to this model as the *classical two-part model*.

As stated in the introduction, one of the objectives of this paper is to find generic and very parametrically flexible two-part model specifications for the continuous and count-valued outcome cases that afford both easy parametric estimation and relatively simple tests of the one-part null hypothesis (i.e., H_0 : A two-part structure is not needed). For example, in the continuous case we seek a very flexible generic specification for the pdf of the PO [made explicit in equation (3)] that nests a version of the model in which there is in some sense no substantive distinction between the EM and the IM (the so-called "one-part" null model). If we could construct such a "nestable" version of (3), then we would be able to apply a relatively simple likelihood ratio test of the one-part null hypothesis. We note here that it is not possible to construct such a "nestable" version of (9). This is probably why the question of testing the one-part null never arises in the context of applications of the classical two-part model.

With this main motivation, we propose a general and very flexible fully parametric 2PM (FP2PM) specification for (3) that, as we will later show, does nest a plausible one-part version and, therefore, affords a likelihood ratio test of the pertinent null hypothesis. Our proposed FP2PM generally applies to both continuous and count-valued outcome models. We will highlight specific differences between the continuous and count-valued cases as necessary.

2.2.1 Extensive Margin (EM)

$$Y_{X_p^*} = 0 \qquad \text{iff} \qquad U < \mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(\zeta_{EM}, X_p^*, X_o; \tau_{EM}) \tag{10}$$

where, U is uniformly distributed on the unit interval and $\mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(\zeta_{EM}, X_p^*, X_o; \tau_{EM})$ is the conditional cdf of $\zeta^*_{EM} | X_o$, written as a function of ζ^*_{EM}, X_p^*, X_o, and the vector relevant parameters for the EM (τ_{EM}), evaluated at ζ_{EM} (an unobserved parametric threshold point), and $\mathcal{G}_{(A|C)}(A, B, C; \psi)$ denotes the cdf of A conditional on C, written as a function of A, B, C with a parameter vector, ψ.

2.2.2 Intensive Margin (IM)

$$(Y_{X_p^*} \mid Y_{X_p^*} > \zeta_{IM}) \text{ has the cdf } \mathcal{G}^{IM*}(Y_{X_p^*}) \quad \text{iff} \quad U \geq \mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(\zeta_{EM}, X_p^*, X_o; \tau_{EM}) \tag{11}$$

where, $\mathcal{G}^{IM*}(Y_{X_p^*})$ is a shorthand notation for

$$\mathcal{G}^{IM*}_{(Y_{X_p^*} \mid Y_{X_p^*} > \zeta_{IM}, X_o)}(Y_{X_p^*}, X_p^*, X_o, \zeta_{IM}; \tau_{IM}) = \frac{\mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}(Y_{X_p^*}, X_p^*, X_o; \tau_{IM})}{1 - \mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}(\zeta_{IM}, X_p^*, X_o; \tau_{IM})}. \tag{12}$$

$\mathcal{G}^{\mathrm{IM}}_{(\zeta^*_{\mathrm{IM}}|\mathrm{X_o})}(\zeta_{\mathrm{IM}}, \mathrm{X}^*_{\mathrm{p}}, \mathrm{X_o}; \tau_{\mathrm{IM}})$ is the specified conditional cdf of ζ^*_{IM} given $\mathrm{X_o}$, written as a function of ζ_{IM}, $\mathrm{X_o}$, and the vector of relevant parameters for the IM (τ_{IM}); and ζ_{IM} is a scalar. When $\mathrm{Y}_{\mathrm{X}^*_{\mathrm{p}}}$ is continuous, ζ_{IM} is an unknown parametric scalar. When $\mathrm{Y}_{\mathrm{X}^*_{\mathrm{p}}}$ is count-valued, ζ_{IM} is often set equal to 0. Note that our specification of the IM [(11) and (12)] diverges slightly from that of the classical two-part model discussed above in that it is written in terms of a truncated continuous (count) pdf (pmf) whose truncation point is parametric and unknown. The reason for such truncation will be made clear later in the discussion.

2.2.3 Pdf (Pmf) of the Continuous (Count-Valued) PO

The relevant version of (3) then becomes

$$\mathrm{f}_{(\mathrm{Y}_{\mathrm{X}^*_{\mathrm{p}}}|\mathrm{X_o})}(\mathrm{Y}_{\mathrm{X}^*_{\mathrm{p}}}, \mathrm{X}^*_{\mathrm{p}}, \mathrm{X_o}; \pi) = \mathcal{G}^{\mathrm{EM}}_{(\zeta^*_{\mathrm{EM}}|\mathrm{X_o})}(\zeta_{\mathrm{EM}}, \mathrm{X}^*_{\mathrm{p}}, \mathrm{X_o}; \tau_{\mathrm{EM}})^{\mathrm{I}(\mathrm{Y}_{\mathrm{X}^*_{\mathrm{p}}}=0)}$$
$$\times \Big[1 - \mathcal{G}^{\mathrm{EM}}_{(\zeta^*_{\mathrm{EM}}|\mathrm{X_o})}(\zeta_{\mathrm{EM}}, \mathrm{X}^*_{\mathrm{p}}, \mathrm{X_o}; \tau_{\mathrm{EM}})$$
$$\times \mathcal{G}^{\mathrm{IM}*}_{(\zeta^*_{\mathrm{IM}}\,|\,\mathrm{Y}_{\mathrm{X}^*_{\mathrm{p}}} > \zeta_{\mathrm{IM}}, \mathrm{X_o})}{}'(\mathrm{Y}_{\mathrm{X}^*_{\mathrm{p}}}, \mathrm{X}^*_{\mathrm{p}}, \mathrm{X_o}, \zeta_{\mathrm{IM}}; \tau_{\mathrm{IM}}) \Big]^{1-\mathrm{I}(\mathrm{Y}_{\mathrm{X}^*_{\mathrm{p}}}=0)} \tag{13}$$

where, $\mathcal{G}^{\mathrm{IM}*}_{(\zeta^*_{\mathrm{IM}}\,|\,\mathrm{Y}_{\mathrm{X}^*_{\mathrm{p}}} > \zeta_{\mathrm{IM}}, \mathrm{X_o})}{}'(\mathrm{Y}_{\mathrm{X}^*_{\mathrm{p}}}, \mathrm{X}^*_{\mathrm{p}}, \mathrm{X_o}, \zeta_{\mathrm{IM}}; \tau_{\mathrm{IM}})$ denotes the continuous pdf (count pmf) corresponding to $\mathcal{G}^{\mathrm{IM}*}_{(\zeta^*_{\mathrm{IM}}\,|\,\mathrm{Y}_{\mathrm{X}^*_{\mathrm{p}}} > \zeta_{\mathrm{IM}}, \mathrm{X_o})}(\zeta^*_{\mathrm{IM}}, \mathrm{X}^*_{\mathrm{p}}, \mathrm{X_o}, \zeta_{\mathrm{IM}}; \tau_{\mathrm{IM}})$ evaluated at $\mathrm{Y}_{\mathrm{X}^*_{\mathrm{p}}}$ for $\mathrm{Y}_{\mathrm{X}^*_{\mathrm{p}}} > \zeta_{\mathrm{IM}}$; i.e.,

$$\mathcal{G}^{\text{IM}*}_{(\zeta^*_{\text{IM}} \mid Y_{X^*_p} > \zeta_{\text{IM}}, X_o)}{}'(Y_{X^*_p}, X^*_p, X_o, \zeta_{\text{IM}}; \tau_{\text{IM}}) = \frac{\mathcal{G}^{\text{IM}}_{(\zeta^*_{\text{IM}} | Y_{X^*_p} > \zeta_{\text{IM}}, X_o)}{}'(Y_{X^*_p}, X^*_p, X_o; \tau_{\text{IM}})}{1 - \mathcal{G}^{\text{IM}}_{(\zeta^*_{\text{IM}} | X_o)}(\zeta_{\text{IM}}, X^*_p, X_o; \tau_{\text{IM}})}. \tag{14}$$

$\mathcal{G}^{\text{IM}}_{(\zeta^*_{\text{IM}} | Y_{X^*_p} > \zeta_{\text{IM}}, X_o)}{}'(Y_{X^*_p}, X^*_p, X_o; \tau_{\text{IM}})$ is the continuous pdf (count pmf) corresponding to $\mathcal{G}^{\text{IM}}_{(\zeta^*_{\text{IM}} | X_o)}(\zeta^*_{\text{IM}}, X^*_p, X_o; \tau_{\text{IM}})$ evaluated at $Y_{X^*_p}$ for $Y_{X^*_p} > \zeta_{\text{IM}}$. Later we will give conditions under which (13) affords a version that comports with an one-part version of the model so that a straightforward likelihood ratio test of the "key" null hypothesis can be applied.

2.2.4 Conditional Mean Function for the PO

In the continuous case we have

$$E[Y_{X^*_p} \mid X_o] = m(X^*_p, X_o; \pi)$$

$$= \left(1 - \mathcal{G}^{\text{EM}}_{(\zeta^*_{\text{EM}} | X_o)}(\zeta_{\text{EM}}, X^*_p, X_o; \tau_{\text{EM}})\right) \times \frac{\int_{\zeta_{\text{IM}}}^{\infty} Y_{X^*_p} \mathcal{G}^{\text{IM}}_{(\zeta^*_{\text{IM}} | X_o)}{}'(Y_{X^*_p}, X^*_p, X_o; \tau_{\text{IM}})\, dY_{X^*_p}}{1 - \mathcal{G}^{\text{IM}}_{(\zeta^*_{\text{IM}} | X_o)}(\zeta_{\text{IM}}, X^*_p, X_o; \tau_{\text{IM}})} \tag{15}$$

whereas in the count-valued case

$$E[Y_{X^*_p} \mid X_o] = m(X^*_p, X_o; \pi)$$

$$= \left(1 - \mathcal{G}^{\text{EM}}_{(\zeta^*_{\text{EM}} | X_o)}(\zeta_{\text{EM}}, X^*_p, X_o; \tau_{\text{EM}})\right) \left(\frac{E[Y_{X^*_p}{}^{+} \mid X_o]}{1 - \mathcal{G}^{\text{IM}}_{(\zeta^*_{\text{IM}} | X_o)}(0, X^*_p, X_o; \tau_{\text{IM}})} \right) \tag{16}$$

noting that

$$E[Y_{X_p^*}{}^{+} \mid X_o] = \sum_{a=1}^{\infty} a \times \mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}{}'(a, X_p^*, X_o; \tau_{IM}).$$

2.2.5 Conventional Two-Part Model

We refer to the case, in which (10) through (16) hold and it assumes that

$$\mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(\zeta_{EM}, X_p^*, X_o; \tau_{EM}) = 1 - \Phi(X_p^*\beta_{p1} + X_o\beta_{o1}) \tag{17}$$

as the *conventional two-part model*, where, $\Phi(\ )$ denotes the standard normal cdf. Note that this comports with the EM assumption of the classical continuous two-part model. In the conventional two-part model, the relevant versions of (13) and (16) are

$$\begin{aligned} f_{(Y_{X_p^*}|X_o)}(Y_{X_p^*}, X_p^*, X_o; \pi) &= [1 - \Phi(X_p^*\beta_{p1} + X_o\beta_{o1})]^{I(Y_{X_p^*}=0)} \\ &\times \left(\Phi(X_p^*\beta_{p1} + X_o\beta_{o1})\mathcal{G}^{IM*}_{(\zeta^*_{IM} \mid Y_{X_p^*} > \zeta_{IM}, X_o)}{}'(Y_{X_p^*}, X_p^*, X_o, \zeta_{IM}; \tau_{IM})\right)^{1-I(Y_{X_p^*}=0)} \end{aligned} \tag{18}$$

and

$$\begin{aligned} E[Y_{X_p^*} \mid X_o] &= m(X_p^*, X_o; \pi) \\ &= \Phi(X_p^*\beta_{p1} + X_o\beta_{o1}) \times \frac{\int_{\zeta_{IM}}^{\infty} Y_{X_p^*}\mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}{}'(Y_{X_p^*}, X_p^*, X_o; \tau_{IM})\, dY_{X_p^*}}{1 - \mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}(\zeta_{IM}, X_p^*, X_o; \tau_{IM})} \end{aligned} \tag{19}$$

whereas for the count-valued case, from (16), we get

$$E[Y_{X_p^*} \mid X_o] = m(X_p^*, X_o; \pi)$$

$$= \Phi(X_p^*\beta_{p1} + X_o\beta_{o1})\left(\frac{E[Y_{X_p^*}{}^{+} \mid X_o]}{1-\mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}(0, X_p^*, X_o; \tau_{IM})}\right). \qquad (20)$$

2.2.6 Examples

Two examples of the conventional two-part model may help fix ideas.

Continuous Outcome Conventional Two-Part Model – Log-Normal IM:

Here we specify the relevant pdf as in (18) with

$$\mathcal{G}^{IM}_{(\zeta^*_{IM}|Y_{X_p^*}>\zeta_{IM}, X_o)}{}'(Y_{X_p^*}, X_p^*, X_o; \tau_{IM})$$

$$= \varphi_{\ln}(Y_{X_p^*}; X_p^*\beta_{p2} + X_o\beta_{o2}, \sigma_2^2)$$
$$= \frac{1}{Y_{X_p^*}\sqrt{2\pi\sigma_2^2}}\exp\left[-\frac{1}{2\sigma_2^2}\left\{\ln Y_{X_p^*} - (X_p^*\beta_{p2} + X_o\beta_{o2})\right\}^2\right] \qquad (21)$$

and

$$\mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}(\zeta_{IM}, X_p^*, X_o; \tau_{IM}) = \Phi\left(\frac{\ln(\zeta_{IM}) - (X_p^*\beta_{p2} + X_o\beta_{o2})}{\sigma_2}\right). \qquad (22)$$

The relevant version of the conditional mean in (19) is[9]

$$E[Y_{X_p^*} \mid X_o] = m(X_p^*, X_o; \pi)$$
$$= \Phi(X_p^*\beta_{p1} + X_o\beta_{o1}) \times \frac{\exp\left(X_p^*\beta_{p2} + X_o\beta_{o2} + \frac{\sigma_2^2}{2}\right)\left[1 - \Phi\left(\frac{\ln(\zeta_{IM})}{\sigma_2} - \sigma_2\right)\right]}{1 - \Phi\left(\frac{\ln(\zeta_{IM}) - (X_p^*\beta_{p2} + X_o\beta_{o2})}{\sigma_2}\right)}. \qquad (23)$$

[9] See Lemma 1 through Lemma 3 in the Appendix II for derivation of (23).

Count-Valued Outcome Conventional Two-Part Model – Poisson IM:

Here we specify the relevant pmf as in (18) with

$$\mathcal{G}^{\text{IM}}_{(\zeta^*_{\text{IM}}|Y_{X^*_p}>\zeta_{\text{IM}},X_o)}{}'(Y_{X^*_p}, X^*_p, X_o; \tau_{\text{IM}})$$

$$= \text{poi}(Y_{X^*_p}, X^*_p, X_o; \lambda_2)$$

$$= \frac{\lambda_2^{Y_{X^*_p}} \exp(-\lambda_2)}{Y_{X^*_p}!} \tag{24}$$

where, $\lambda_2 = \exp(X^*_p\beta_{p2} + X_o\beta_{o2})$ and

$$\mathcal{G}^{\text{IM}}_{(\zeta^*_{\text{IM}}|X_o)}(0, X^*_p, X_o; \tau_{\text{IM}}) = \text{poi}(0, X^*_p, X_o; \lambda_2) = \exp(-\lambda_2)\,. \tag{25}$$

The relevant version of the conditional mean in (20) is

$$E[Y_{X^*_p} \,|\, X_o] = m(X^*_p,\ X_o;\ \pi)$$

$$= \Phi(X^*_p\beta_{p1} + X_o\beta_{o1}) \times \left(\frac{\lambda_2}{1-\exp(-\lambda_2)}\right). \tag{26}$$

2.2.7 The Generic Two-Part PO Specification Revisited

The generic FP2PM specification detailed in section 2.2.1 and 2.2.2 above can be summarized in an instructive and useful way as follows.

Continuous Outcome Models:

Based on (10), (11), and (12), for continuous $Y_{X_p^*}$, we get

$$Y_{X_p^*} = I\left(U \geq \mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(\zeta_{EM}, X_p^*, X_o; \tau_{EM})\right) \times \mathcal{G}^{IM*}_{(U \geq \mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(\zeta_{EM}, X_p^*, X_o; \tau_{EM}), X_o)}{}^{-1}(U, X_p^*, X_o, \zeta_{IM}; \tau_{IM}) \quad (27)$$

where, $\mathcal{G}^{IM*}_{(U \geq \mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(\zeta_{EM}, X_p^*, X_o; \tau_{EM}), X_o)}{}^{-1}(U, X_p^*, X_o, \zeta_{IM}; \tau_{IM})$ denotes the inverse of the cdf $\mathcal{G}^{IM*}_{(Y_{X_p^*} \mid Y_{X_p^*} > \zeta_{IM}, X_o)}(Y_{X_p^*}, X_p^*, X_o, \zeta_{IM}; \tau_{IM})$ whose main argument U (unit uniform) is restricted to the interval $\left(\mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(\zeta_{EM}, X_p^*, X_o; \tau_{EM}), X_o), 1\right)$.

Count-Valued Outcome Models:

Based on (10), (11), and (12), for count-valued $Y_{X_p^*}$, ($Y_{X_p^*} = 0, 1, 2, \ldots, \infty$), we get

$$\begin{aligned} Y_{X_p^*} = j \times I\Bigg(& \mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(0, X_p^*, X_o; \tau_{EM}) \\ & + \left(1 - \mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(0, X_p^*, X_o; \tau_{EM})\right) \times \left(\frac{\mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}(j-1, X_p^*, X_o; \tau_{IM}) - \mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}(0, X_p^*, X_o; \tau_{IM})}{1 - \mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}(0, X_p^*, X_o; \tau_{IM})} \right) \\ & < U \leq \left(\mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(0, X_p^*, X_o; \tau_{EM}) \right) \\ & + \left(1 - \mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(0, X_p^*, X_o; \tau_{EM})\right) \times \left(\frac{\mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}(j, X_p^*, X_o; \tau_{IM}) - \mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}(0, X_p^*, X_o; \tau_{IM})}{1 - \mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}(0, X_p^*, X_o; \tau_{IM})} \right) \Bigg) \end{aligned} \quad (28)$$

for $j = 1, 2, \ldots, \infty$.

As soon as we propose the FP2PM in (10), (11), and (13), it is worth mentioning that the general framework does not restrict the error terms in the EM and IM [like in (7) and (8)] to be jointly distributed (e.g., bivariate normal distribution), however, the framework allows the two error terms to have any specific distribution(s). In addition, the general FP2PM allows for extent flexibilities of the EM and IM's distributions, which are most likely unknown in applied research. And thus, the general FP2PM framework eliminates the model selection concerns regarding "cake debate". Please refer to Appendix I for more details regarding the "cake debate".

Chapter 3.

Estimating Targeted Causal Parameters of Interest in the FP2PM PO Framework

Now we have the relevant functional forms for m( ) in both the continuous and count-data versions of the FP2PM PO framework. If we also have a consistent estimator for the vector of the "deep" parameters π (say, $\hat{\pi}$), we will be able to consistently estimate (5) and (6) using their following sample analogs[10]

$$\text{AIE}(\Delta) = \hat{\kappa}_2 - \hat{\kappa}_1 \tag{29}$$

and

$$\text{AED}(\Delta) = (\hat{\kappa}_2 - \hat{\kappa}_1)\frac{\overline{X}_p}{\hat{\kappa}_1} \tag{30}$$

where, $\hat{\kappa}_1 = \sum_{i=1}^{n} \frac{1}{n} m(X_{pi}, X_{oi}; \hat{\pi})$, $\hat{\kappa}_2 = \sum_{i=1}^{n} \frac{1}{n} m(X_{pi} + \Delta, X_{oi}; \hat{\pi})$, and $\overline{X}_p = \sum_{i=1}^{n} \frac{1}{n} X_{pi}$. It is to the consistent estimation of the vector of deep parameters, π, which we now turn. We begin our discussion by establishing conditions under which the PO pdf (pmf) in (13) implies a similar form for the DGP in the continuous (count-valued) case.

3.1. Reconciling the PO with the Data Generating Process (DGP) in the FP2P Modeling Framework

Our first inclination here is to simply replace $Y_{X_p^*}$ and X_p^* with Y and X_p respectively in (13) for the continuous (count-valued) case, and then assume that the

[10] The asymptotic standard errors of (7) and (8) can be obtained using the approach of Terza (2017).

resultant conditional pdf (pmf) represents the true DGP. In fact, in most applications of the 2PM, there is no discussion at all about the underlying PO framework and, therefore, no clear and rigorous specification given for the targeted causal parameter to be estimated. In such applications, model specification simply begins with an assumed form for the true DGP. This approach ignores the fact that a policy relevant causal parameter characterizes a counterfactual feature of interest and, therefore, can only be coherently specified in a counterfactual framework, like the PO framework. Moreover, it fails to acknowledge the fact that the relevant PO, e.g., characterized as (13) in the FP2PM context, may not coincide with the true DGP under the aforementioned substitutions. Such coincidence between the relevant PO and the DGP holds only under certain conditions that can only be specified in the PO framework. See Terza (2018) for a detailed discussion of said conditions. Placing Terza's (2018) discussion in the present context, we can show that these conditions are sufficient to ensure that given X_o, $Y_{X_p^*}$ and X_p are conditionally independent. This implies that the true DGP is

$$\text{pdf}(Y \mid X_p, X_o) = f_{(Y_{X_p^*} \mid X_o)}(Y, X_p, X_o; \pi) \tag{31}$$

where, $f_{(Y_{X_p^*} | X_o)}(\ )$ is defined as in (13). Two points to be made in review: first, equation (31) only holds under conditions that are specified in the relevant PO framework; and second, $f_{(Y_{X_p^*} \mid X_o)}(\ )$ is defined and specified in the relevant PO framework.

3.2 Full Information Maximum Likelihood Estimation of the Deep Parameters

We begin this section by detailing the generic log-likelihood function of the form

$$L(\pi \mid Y, X_p, X_o, \zeta) = \sum_{i=1}^{n} \ln \left(f_{(Y_{X_p^*} \mid X_o)}(Y_i, X_{pi}, X_{oi};\, \pi) \right) \tag{32}$$

where, Y_i, X_{pi} and X_{oi} are the values of Y, X_p and X_o observed for the i[th] individual in the sample (for i = 1, ..., n) for the continuous and count-data cases that would follow from the appropriately specified versions of (31). In each of these two relevant contexts, we also give the details of said log-likelihood functions for both the conventional version of the model and a more flexible specification. We then turn to a discussion of causal effect estimation in each of these four cases, i.e., two for continuous data and two for count-valued outcome data.

3.2.1 Log-Likelihood Function and Conditional Mean Function – Continuous Model

The generic log-likelihood function in the continuous case is based on the version of (31), in which

$$\begin{aligned} f_{(Y_{X_p^*} \mid X_o)}(Y_i, X_{pi}, X_{oi};\, \pi) = {} & \mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(\zeta_{EM}, X_{pi}, X_{oi}; \tau_{EM})^{I(Y_i=0)} \\ & \times \Bigg(1 - \mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(\zeta_{EM}, X_{pi}, X_{oi}; \tau_{EM}) \\ & \qquad \times \frac{\mathcal{G}^{IM}_{(\zeta^*_{IM}|Y_{X_p^*} > \zeta_{IM}, X_o)}{}'(Y_i, X_{pi}, X_{oi}, X_o; \tau_{IM})}{1 - \mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}(\zeta_{IM}, X_{pi}, X_{oi}; \tau_{IM})} \Bigg)^{1-I(Y_i=0)} . \end{aligned} \tag{33}$$

As we will later show, ζ_{EM} and ζ_{IM} can be consistently estimated using the minimum order statistic for the sample values of Y at the extensive margin.[11]

Case I: Continuous Conventional Two-Part Model with Probit EM and Log-Normal IM

Combining (18) with (21), under the requisite conditions (Terza, 2018), yields the following version of (33)

$$f_{(Y_{X_p^*}|X_o)}(Y_i, X_{pi}, X_{oi}; \pi) = [1-\Phi(X_{pi}\beta_{p1} + X_{oi}\beta_{o1})]^{I(Y_i=0)}$$
$$\times\left(\Phi(X_{pi}\beta_{p1} + X_{oi}\beta_{o1}) \frac{\varphi_{\ln[Y>\zeta_{IM}]}(Y_i;\, X_{pi}\beta_{p2} + X_{oi}\beta_{o2}, \sigma_2^2)}{1-\Phi\left(\dfrac{\ln(\zeta_{IM}) - (X_{pi}\beta_{p2} + X_{oi}\beta_{o2})}{\sigma_2}\right)}\right)^{1-I(Y_i=0)} \quad (34)$$

where, $\pi' = [\beta_1' \quad \beta_2' \quad \sigma_2^2]$ is the vector of parameters and

$$\varphi_{\ln[Y>\zeta_{IM}]}(Y_i; X_{pi}\beta_{p2} + X_{oi}\beta_{o2}, \sigma_2^2) = \frac{1}{Y_i\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{1}{2\sigma_2^2}\left\{\ln Y_i - (X_{pi}\beta_{p2} + X_{oi}\beta_{o2})\right\}^2\right] \quad (35)$$

is the log-normal pdf with domain restricted to $Y > \zeta_{IM}$. The relevant estimated conditional mean function for causal effect estimation is

$$m(X_{pi}, X_{oi}; \hat{\pi}) = \Phi(X_{pi}\hat{\beta}_{p1} + X_{oi}\hat{\beta}_{o1})\left(\frac{\exp\left(X_{pi}\hat{\beta}_{p2} + X_{oi}\hat{\beta}_{o2} + \dfrac{\hat{\sigma}_2^2}{2}\right)\left[1-\Phi\left(\dfrac{\ln(\hat{\zeta}_{IM})}{\hat{\sigma}_2} - \hat{\sigma}_2\right)\right]}{1-\Phi\left(\dfrac{\ln(\hat{\zeta}_{IM}) - (X_{pi}\hat{\beta}_{p2} + X_{oi}\hat{\beta}_{o2})}{\hat{\sigma}_2}\right)}\right) \quad (36)$$

[11] Minimum order statistic refers to the smallest value in the sample.

where, $\hat{\pi}' = [\hat{\beta}_1' \quad \hat{\beta}_2' \quad \hat{\sigma}_2^2]$ is the MLE estimator of π obtained via (34) and $\hat{\zeta}_{IM}$ is the minimum order statistic for the IM subsample on Y.

Case II: Continuous Conventional Two-Part Model with Probit EM and Generalized Gamma IM

Here, under the requisite conditions (Terza, 2018), we replace the log-normal specification for the IM in (34) with the three parameter generalized gamma distribution (GG) to obtain

$$f_{(Y_{X_p^*}|X_o)}(Y_i, X_{pi}, X_{oi}; \pi) = [1-\Phi(X_{pi}\beta_{p1} + X_{oi}\beta_{o1})]^{I(Y_i=0)} \times \left(\Phi(X_{pi}\beta_{p1} + X_{oi}\beta_{o1}) \frac{gg_{[Y>\zeta_{IM}]}(Y_i; X_{pi}\beta_{p2} + X_{oi}\beta_{o2}, \kappa_2, \sigma_2)}{1 - GG(\zeta_{IM}; X_{pi}\beta_{p2} + X_{oi}\beta_{o2}, \kappa_2, \sigma_2)} \right)^{1-I(Y_i=0)} \tag{37}$$

where, $\pi' = [\beta_1' \quad \beta_2' \quad \kappa_2 \quad \sigma_2^2]$ ($\beta_1' = [\beta_{1p} \quad \beta_{1o}']$ and $\beta_2' = [\beta_{2p} \quad \beta_{2o}']$) is the vector of parameters; $gg_{[A>a]}(A; b, c, d)$ denotes the pdf of a generalized gamma variate A with parameters b, c and d, and domain restricted to $A > a$; and GG(A; b, c, d) is the generalized gamma cdf with parameters b, c and d, evaluated at A. Specifically,

$$gg(Y_i; X_{pi}\beta_{p2} + X_{oi}\beta_{o2}, \kappa_2, \sigma_2) = \frac{\nu^\nu}{\sigma_2 Y_i \sqrt{\nu}\Gamma(\nu)} \exp[z_i\sqrt{\nu} - u_i] \tag{38}$$

and

$$GG(\zeta_{IM}; X_{pi}\beta_{p2} + X_{oi}\beta_{o2}, \kappa_2, \sigma_2) = \begin{cases} SG(\nu, (\zeta_{IM} / \alpha_i)^p) & \text{if } p > 0 \\ \\ 1 - SG(\nu, (\zeta_{IM} / \alpha_i)^p) & \text{if } p < 0 \end{cases} \tag{39}$$

where, $\nu = |\kappa|^{-2}$, $z_i = \text{sign}(\kappa)[\log(Y_i) - (X_{pi}\beta_{p2} + X_{oi}\beta_{o2})] / \sigma_2$, $u_i = \nu \times \exp(|\kappa| z_i)$, $\alpha_i = \dfrac{\exp(X_{pi}\beta_{p2} + X_{oi}\beta_{o2})}{\nu^{\frac{1}{p}}}$, $p = \dfrac{\kappa_2}{\sigma_2}$, and SG(h, j) denotes the cdf of the standard gamma distribution evaluated at h with shape parameter j, specifically

$$SG(h,\ j) = \frac{\int_0^j t^{h-1} e^{-t}\, dt}{\Gamma(h)}$$

[see (Yang 2016): Section 2.4.1]. The relevant estimated conditional mean function for causal effect estimation is

$$m(X_{pi},\ X_{oi};\ \hat{\pi}) = \Phi(X_{pi}\hat{\beta}_{p1} + X_{oi}\hat{\beta}_{o1}) \left(\frac{\int_{\hat{\zeta}_{IM}}^{\infty} Y_i\ gg_{[Y_{X_p^*} > \zeta_{IM}]}(Y_i; X_{pi}\hat{\beta}_{p2} + X_{oi}\hat{\beta}_{o2}, \hat{\kappa}_2, \hat{\sigma}_2)\, dY_i}{1 - GG(\hat{\zeta}_{IM}; X_{pi}\hat{\beta}_{p2} + X_{oi}\hat{\beta}_{o2}, \hat{\kappa}_2, \hat{\sigma}_2)} \right) \tag{40}$$

where, $\hat{\pi}' = [\hat{\beta}_1' \quad \hat{\beta}_2' \quad \hat{\kappa}_2 \quad \hat{\sigma}_2^2]$ is the MLE estimator of π obtained via (37) and $\hat{\zeta}_{IM}$ is the minimum order statistic for the IM subsample on Y.

The three-parameter GG has been discussed and utilized in applied econometrics due to its high degree of model flexibility (Manning et al. 2005 and Liu et al. 2010).

Corresponding to different parameter configurations, the GG nests several common distributions, such as gamma, Weibull, exponential, and log-normal.

3.2.2 Log Likelihood Function and Conditional Mean Function – Count-Data Model

The generic log-likelihood function in the count-data case is based on the version of (31), in which

$$
\begin{aligned}
f_{(Y_{X_p^*} \mid X_o)}(Y_i, X_{pi}, X_{oi};\, \pi) = {} & \mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(\zeta_{EM}, X_{pi}, X_{oi}; \tau_{EM})^{I(Y_i=0)} \\
& \times \Bigg[1 - \mathcal{G}^{EM}_{(\zeta^*_{EM}|X_o)}(\zeta_{EM}, X_{pi}, X_{oi}; \tau_{EM}) \\
& \qquad \times \frac{\mathcal{G}^{IM}_{(\zeta^*_{IM}|Y_{X_p^*}>0, X_o)}{}'(Y_i, X_{pi}, X_{oi}, X_o; \tau_{IM})}{1 - \mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}(0, X_{pi}, X_{oi}; \tau_{IM})} \Bigg]^{1-I(Y_i=0)}
\end{aligned} \tag{41}
$$

recalling in this case that $\mathcal{G}^{IM}_{(\zeta^*_{IM}|X_o)}{}'(Y_i, X_{pi}, X_{oi}, X_o; \tau_{IM})$ is a count-data pmf.

Case III: Count-Valued Conventional Two-Part Model with Probit EM and Poisson IM

Combining (18) with (24), under the requisite conditions (Terza, 2018), yields the following version of (41)

$$
\begin{aligned}
f_{(Y_{X_p^*}|X_o)}(Y_i, X_{pi}, X_{oi}; \pi) = {} & [1 - \Phi(X_{pi}\beta_{p1} + X_{oi}\beta_{o1})]^{I(Y_i=0)} \\
& \times \left(\Phi(X_{pi}\beta_{p1} + X_{oi}\beta_{o1}) \, \frac{poi_{[Y>0]}(Y_i, X_{pi}, X_{oi}; \lambda_{2i})}{1 - \exp(-\lambda_{2i})} \right)^{1-I(Y_i=0)}
\end{aligned} \tag{42}
$$

where, $\lambda_{2i} = \exp(X_{pi}\beta_{p2} + X_{oi}\beta_{o2})$, $\pi' = [\beta_1' \quad \beta_2' \quad \sigma_2^2]$ ($\beta_1' = [\beta_{1p} \quad \beta_{1o}']$ and $\beta_2' = [\beta_{2p} \quad \beta_{2o}']$) is the vector of parameters and

$$\text{poi}_{[Y>0]}(Y_i, X_{pi}, X_{oi}; \lambda_{2i}) = \frac{\lambda_{2i}{}^{Y_i} \exp(-\lambda_{2i})}{Y_i\,!} \tag{43}$$

is the Poisson pmf with domain restricted to $Y > 0$. The relevant estimated conditional mean function for causal effect estimation is

$$m(X_{pi},\, X_{oi};\, \hat{\pi}) = \Phi(X_{pi}\hat{\beta}_{p1} + X_{oi}\hat{\beta}_{o1})\left(\frac{\hat{\lambda}_{2i}}{1-\exp(-\hat{\lambda}_{2i})}\right) \tag{44}$$

where, $\hat{\pi}' = [\hat{\beta}_1' \quad \hat{\beta}_2' \quad \hat{\sigma}_2^2]$ is the MLE estimator of π obtained via (42), and $\hat{\lambda}_{2i} = \exp(X_{pi}\hat{\beta}_{p2} + X_{oi}\hat{\beta}_{o2})$.

Case IV: Count-Valued Conventional Two-Part Model with Probit EM and Conway-Maxwell Poisson IM

Here, under the requisite conditions in (Terza, 2018), we replace the Poisson specification for the IM in (42) with the Conway-Maxwell Poisson (CMP) to obtain

$$\begin{aligned} f_{(Y_{X_p^*}|X_o)}(Y_i, X_{pi}, X_{oi}; \pi) &= [1-\Phi(X_{pi}\beta_{p1} + X_{oi}\beta_{o1})]^{I(Y_i=0)} \\ &\times\left(\Phi(X_{pi}\beta_{p1} + X_{oi}\beta_{o1})\ \frac{\text{cmp}_{[Y>0]}(Y_i; X_{pi}\beta_{p2} + X_{oi}\beta_{o2}, \sigma_2)}{1 - \text{cmp}(0; X_{pi}\beta_{p2} + X_{oi}\beta_{o2}, \sigma_2)}\right)^{1-I(Y_i=0)} \end{aligned} \tag{45}$$

where, $\pi' = [\beta_1' \quad \beta_2' \quad \sigma_2^2]$ ($\beta_1' = [\beta_{1p} \quad \beta_{1o}']$ and $\beta_2' = [\beta_{2p} \quad \beta_{2o}']$) is the vector of parameters; and $\text{cmp}_{[A>a]}(A; b, c)$ denotes the pmf of a Conway-Maxwell Poisson variate A with parameters b and c, and domain restricted to $A > a$. Specifically,

$$\text{cmp}(Y_i, X_{pi}, X_{oi}; \lambda_{2i}, \sigma_2) = \frac{\lambda_{2i}{}^{Y_i}}{(Y_i\,!)^{\sigma_2} \times Z(\lambda_{2i}, \sigma_2)} \tag{46}$$

where, $\lambda_{2i} = \exp(X_{pi}\beta_{p2} + X_{oi}\beta_{o2})$ and $Z(\lambda_{2i}, \sigma_2) = \sum_{j=0}^{\infty} \frac{\lambda_{2i}{}^{j}}{(j!)^{\sigma_2}}$. Note also that for this random variable,

$$E[Y \mid X_p, X_o] = \lambda_2 \left(\frac{\sum_{j=1}^{\infty} \frac{j\lambda_2^{j-1}}{(j!)^{\sigma_2}}}{\sum_{j=0}^{\infty} \frac{\lambda_2^{j}}{(j!)^{\sigma_2}}} \right). \tag{47}$$

According to (Sellers et al. 2010), the mean function in (47) can be also approximated by

$$E[Y \mid X_p, X_o] \approx \lambda_2^{\frac{1}{\sigma_2}} - \frac{\sigma_2 - 1}{2\sigma_2}. \tag{48}$$

The relevant estimated conditional mean function for causal effect estimation is

$$m(X_{pi},\, X_{oi};\, \hat{\pi}) = \Phi(X_{pi}\hat{\beta}_{p1} + X_{oi}\hat{\beta}_{o1}) \left(\frac{\hat{\lambda}_{2i}^{\frac{1}{\hat{\sigma}_2}} - \frac{\hat{\sigma}_2 - 1}{2\hat{\sigma}_2}}{1 - \frac{1}{Z(\hat{\lambda}_{2i}, \hat{\sigma}_2)}} \right). \tag{49}$$

where, $\hat{\pi}' = [\hat{\beta}_1' \quad \hat{\beta}_2' \quad \hat{\sigma}_2^2]$ is the MLE estimator of π obtained via (45) and $\hat{\lambda}_{2i} = \exp(X_{pi}\hat{\beta}_{p2} + X_{oi}\hat{\beta}_{o2})$.

The CMP nests the standard Poisson distribution when $\sigma_2 = 1$. The data is over-dispersed if $\sigma_2 < 1$, and under-dispersed if $\sigma_2 > 1$. In addition, the limiting case of a

CMP includes two other common count-valued specifications: the Geometric distribution when $\sigma_2 = 0$ and $\lambda_2 < 1$, and the Bernoulli distribution when $\sigma_2 \to \infty$ with probability of $\frac{\lambda_2}{1+\lambda_2}$. The fact that the Poisson is nested in the CMP allows for a simple statistical test of whether or not the specification varies significantly from the standard Poisson. Another advantage of using the CMP distribution in the IM is that the CMP is theoretically unlimited in the range of dispersion, and is even to model binary outcomes, which gives it an unmatched flexibility among fully parametric models.

Chapter 4.

Testing the Non-Two-Part Model Null

There is a vast literature on model specification testing in the 2PM setting. Nearly all of it, however, focuses on testing the 2PM specification against the classical sample selection model of Heckman (1976).[12] By the same token, this literature virtually ignores what should be the key null hypothesis related to model selection in the 2PM context, viz. that a 2PM is not warranted in a particular empirical context (H_0 as given in the Introduction above).[13] Formally,

H_0 : The 2PM is not needed – special modeling consideration need not be given to the zero-valued outcomes.

H_A : The zeros warrant special consideration in modeling – as in the 2PM.

In this section, within the context of our general 2PM framework, we propose a statistical test of this key null. In developing this test, we re-cast the approach of Mullahy (1986) for count-data models and extend it to accommodate models with continuous outcomes. In all of the 2PM specifications considered above, we have assumed that the EM and IM have essentially distinct stochastic structures. As background for the development of our proposed testing protocol, we first examine whether such a structural distinction is necessary. We begin by defining the term *no structural difference* (NSD) in reference to the EM and IM in a 2PM.

---

[12] See the footnote 2 for a list of literature of tests in the 2PM context.

[13] (Mullahy 1986) is an early paper that takes account of the misspecification of the 2PM specification but it has been virtually ignored. As far as we can find, (Arulampalam and Booth 1977) is the only paper which applies Mullahy's misspecification test on the 2PM structure.

4.1 No Structural Difference between the EM and the IM (NSD)

We begin this section by recalling the generic form of the relevant log-likelihood function is given in (32), specific versions of which for continuous and count-valued outcomes would follow from the appropriately specified versions of (31). In (33) [(41)], the generic version of the relevant pdf [pmf] underlying (32) for the continuous [count-valued] case is given. Cases I through Case IV detail various EM specifications for the conventional version of the 2PM (Probit IM). In all of these cases, we see that the specifications for the EM and IM may differ both *structurally* and *parametrically*. *Structural difference* amounts the possibility that $\mathcal{G}^{\text{EM}}_{(\zeta^*_{\text{EM}}|\text{X}_\text{o})}(\zeta_{\text{EM}}, \text{X}^*_\text{p}, \text{X}_\text{o}; \tau_{\text{EM}})$ in (13) may have a functional form that is different from that of $\mathcal{G}^{\text{IM}}_{(\zeta^*_{\text{IM}}|\text{X}_\text{o})}(\zeta_{\text{IM}}, \text{X}^*_\text{p}, \text{X}_\text{o}; \tau_{\text{IM}})$. Similarly, *parametric difference* amounts to the possibility that in parametrization of (13), τ_{EM} may differ from τ_{IM}.

In the remainder of this chapter, we explore the plausibility (robustness) of assuming that there is *no structural difference* (NSD) (but possibly parametric difference) between the EM and IM in both the continuous and count-data cases. We begin by developing the NSD versions of all the specific continuous and count-data cases detailed above (continuous – Cases I and Case II; count-data – Case III and Case IV).

4.1.1 Log-Likelihood Function and Conditional Mean Function – Continuous Outcome Model with NSD

The log-likelihood function for the generic NSD continuous case follows from the version of (33) in which

$$\mathcal{G}^{\text{EM}}_{(a|\text{X}_\text{o})}(a, \text{X}^*_\text{p}, \text{X}_\text{o}; b) = \mathcal{G}^{\text{IM}}_{(a|\text{X}_\text{o})}(a, \text{X}^*_\text{p}, \text{X}_\text{o}; b) \tag{50}$$

where, $\mathcal{a}$ denotes the argument of the function and $\mathcal{b}$ denotes the parameter vector. We now detail the versions of (50) that correspond with Cases I and II discussed above.

Case V: Continuous Two-Part Model with Log-Normal EM and IM (NSD)

Here, $\mathcal{G}^{\text{IM}}_{(a|X_o)}(a, X^*_p, X_o; b)$ in (50) is the cdf of a log-normal variate with location and scale parameters c and d in the parameter vector b [under the NSD condition (50), $\mathcal{G}^{\text{EM}}_{(a|X_o)}(a, X^*_p, X_o; b)$ has the same functional form]. Therefore, the distinction between the EM and IM in this case is purely parametric. In the EM, the location parameter is $X_{pi}\beta^{o}_{p1} + X_{oi}\beta^{o}_{o1}$ and the scale parameter is σ^2_1. Whereas in the IM, the location parameter is $X_{pi}\beta_{p2} + X_{oi}\beta_{o2}$ and the scale parameter is σ^2_2. Here, (33) becomes

$$f_{(Y_{X^*_p}|X_o)}(Y_i, X_{pi}, X_{oi}; \pi) = \left(1 - \Phi(X_{pi}\beta^{\dagger}_{p1} + X_{oi}\beta^{\dagger}_{o1})\right)^{I(Y_i=0)}$$
$$\times \left(\Phi(X_{pi}\beta^{\dagger}_{p1} + X_{oi}\beta^{\dagger}_{o1}) \left(\frac{\varphi_{\ln\,[Y>\zeta_{IM}]}(Y_i;\, X_{pi}\beta_{p2} + X_{oi}\beta_{o2}, \sigma^2_2)}{1 - \Phi\left(\dfrac{\ln(\zeta_{IM}) - (X_{pi}\beta_{p2} + X_{oi}\beta_{o2})}{\sigma_2}\right)} \right) \right)^{1-I(Y=0)} \tag{51}$$

where, $\pi' = [\beta^{\dagger\prime}_1 \quad \beta'_2 \quad \sigma^2_2]$ ($\beta^{\dagger}_1 = [\beta^{\dagger}_{p1} \quad \beta^{\dagger}_{po}{}']$, $\beta_2 = [\beta_{p2} \quad \beta'_{o2}]$, $\beta^{\dagger}_{p1} = \beta^{o}_{p1} / \sigma_1$ and $\beta^{\dagger}_{po}$ is the same as $\beta^{o}_{po} / \sigma_1$ with the intercept term shifted by $-\ln(\zeta) / \sigma_1$).[14] The relevant estimated conditional mean function for causal effect estimation is

[14] See Lemma 1 through Lemma 3 in the Appendix II for derivation of (51).

$$
\begin{aligned}
&\mathrm{m}(X_{pi},\, X_{oi};\, \hat{\pi}) \\
&= \Phi(X_{pi}\hat{\beta}^{\dagger}_{p1} + X_{oi}\hat{\beta}^{\dagger}_{o1})\left(\frac{\exp\left(X_{pi}\hat{\beta}_{p2} + X_{oi}\hat{\beta}_{o2} + \dfrac{\hat{\sigma}_2^2}{2}\right)\left[1 - \Phi\left(\dfrac{\ln(\hat{\zeta}_{IM})}{\hat{\sigma}_2} - \hat{\sigma}_2\right)\right]}{1 - \Phi\left(\dfrac{\ln(\hat{\zeta}_{IM}) - (X_{pi}\hat{\beta}_{p2} + X_{oi}\hat{\beta}_{o2})}{\hat{\sigma}_2}\right)} \right)
\end{aligned}
\tag{52}
$$

where, $\hat{\pi}' = [\hat{\beta}_1^{\dagger\prime} \quad \hat{\beta}_2' \quad \hat{\sigma}_2^2]$ is the MLE estimator of π obtained via (51) and $\hat{\zeta}_{IM}$ is the minimum order statistic for the IM subsample on Y.

Case VI: Continuous Two-Part Model with Generalized Gamma EM and IM (NSD)

Here, $\mathcal{G}^{IM}_{(a|X_o)}(a, X_p^*, X_o; b)$ in (50) is the cdf of a generalized gamma variate with two shape parameters of c and d, and one scale parameters of e in the parameter vector b [under the NSD condition (50), $\mathcal{G}^{EM}_{(a|X_o)}(a, X_p^*, X_o; b)$ has the same functional form]. Therefore, as in Case V, the distinction between the EM and IM in this case is purely parametric. For reasons that will become clear subsequently, we begin the discussion in this section at the level of the PO. In this case, the pdf of the PO, given in general form as in (13), is

$$
\begin{aligned}
f_{(Y_{X_p^*}|X_o)}(Y_{X_p^*}, X_p^*, X_o; \pi) &= GG(\zeta_{EM}; X_p^*\beta_{p1} + X_o\beta_{o1}, \kappa_1, \sigma_1)^{I(Y_{X_p^*}=0)} \\
&\quad \times \left[1 - GG(\zeta_{EM}; X_p^*\beta_{p1} + X_o\beta_{o1}, \kappa_1, \sigma_1)\right. \\
&\quad\quad \left. \times \frac{gg_{[Y_{X_p^*} > \zeta_{IM}]}(Y_{X_p^*}; X_p^*\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)}{1 - GG(\zeta_{IM}; X_p^*\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)}\right]^{1 - I(Y_{X_p^*}=0)}
\end{aligned}
\tag{53}
$$

where $\pi' = [\beta_1' \quad \beta_2' \quad \kappa_1 \quad \sigma_1^2 \quad \kappa_2 \quad \sigma_2^2]$ ($\beta_1' = [\beta_{1p} \quad \beta_{1o}']$ and $\beta_2' = [\beta_{2p} \quad \beta_{2o}']$) is the vector of parameters; the EM shape and scale parameters are $X_p^*\beta_{p1} + X_o\beta_{o1}$, κ_1 and σ_1, respectively; and the IM shape and scale parameters are $X_{pi}\beta_{p2} + X_o\beta_{o2}$, κ_2 and σ_2, respectively. The relevant conditional mean function for causal effect specification is

$$\begin{aligned} E[Y_{X_p^*} \mid X_o] &= m(X_p^*, X_o; \pi) \\ &= \left(1 - GG(\zeta_{EM}; X_p^*\beta_{p1} + X_o\beta_{o1}, \kappa_1, \sigma_1)\right) \\ &\quad \times \left(\frac{\int_{\hat{\zeta}_{IM}}^{\infty} Y_{X_p^*}\, gg_{[Y_{X_p^*} > \zeta_{IM}]}(Y_{X_p^*}; X_p^*\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)\, dY_{X_p^* i}}{1 - GG(\zeta_{IM}; X_p^*\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)} \right). \quad (54) \end{aligned}$$

We postpone the further discussion of this case until section 4.2, in which we discuss the robustness of the NSD condition (50) in the various cases.

4.1.2 Log-Likelihood Function and Conditional Mean Function – Count-Valued Outcome Model with NSD

Here, as in the previous section, the log-likelihood function for the generic NSD continuous case follows from the version of (33) in which the condition in (50) holds.

Case VII: Count-Data Two-Part Model with Poisson EM and IM (NSD)

Here, $\mathcal{G}^{IM}_{(a|X_o)}(a, X_p^*, X_o; b)$ in (50) is the cdf of a Poisson variate a with location parameter b. [under the NSD condition (50), $\mathcal{G}^{EM}_{(a|X_o)}(a, X_p^*, X_o; b)$ has the same functional form]. Therefore, the distinction between the EM and IM in this case is purely parametric. In the EM, the location parameter is $\lambda_1 = \exp(X_{pi}\beta_{p1} + X_{oi}\beta_{o1})$; and in the IM, the location parameter is $\lambda_2 = \exp(X_{pi}\beta_{p2} + X_{oi}\beta_{o2})$. Here, (33) becomes

$$f_{(Y_{X_p^*}|X_o)}(Y_i, X_{pi}, X_{oi}; \pi) = \exp(-\lambda_{1i})^{I(Y_i=0)} \times \left[\left(1 - \exp(-\lambda_{1i})\right) \frac{poi_{[Y>0]}(Y_i, X_{pi}, X_{oi}; \lambda_{2i})}{1 - \exp(-\lambda_{2i})} \right]^{1-I(Y_i=0)} \quad (55)$$

where, $\pi' = [\beta_1' \quad \beta_2']$ ($\beta_1' = [\beta_{1p} \quad \beta_{1o}']$ and $\beta_2' = [\beta_{2p} \quad \beta_{2o}']$) is the vector of parameters. The relevant estimated conditional mean function for causal effect estimation is

$$m(X_{pi},\ X_{oi};\ \hat{\pi}) = \left(1 - \exp(-\hat{\lambda}_{1i})\right)\left(\frac{\hat{\lambda}_{2i}}{1 - \exp(-\hat{\lambda}_{2i})}\right) \quad (56)$$

where, $\hat{\pi}' = [\hat{\beta}_1' \quad \hat{\beta}_2']$ is the MLE of π obtained via (55), $\hat{\lambda}_{1i} = \exp(X_{pi}\hat{\beta}_{p1} + X_{oi}\hat{\beta}_{o1})$ and $\hat{\lambda}_{2i} = \exp(X_{pi}\hat{\beta}_{p2} + X_{oi}\hat{\beta}_{o2})$.[15]

Case VIII: Count-Data Two-Part Model with Conway-Maxwell Poisson EM and IM (NSD)

Here, $\mathcal{G}^{IM}_{(a|X_o)}(a, X_p^*, X_o; b, c)$ in (50) is the cdf of a Conway-Maxwell Poisson variate a with location parameter b and dispersion parameter c [under the NSD condition (50), $\mathcal{G}^{EM}_{(a|X_o)}(a, X_p^*, X_o; b, c)$ has the same functional form]. Therefore, the distinction between the EM and IM in this case is purely parametric. In the EM, the location and dispersion parameters are $\lambda_{1i} = \exp(X_{pi}\beta_{p1} + X_{oi}\beta_{o1})$ and σ_1, respectively;

[15] This is the model considered by Mullahy (1986).

and in the IM, the location and dispersion parameters are $\lambda_{2i} = \exp(X_{pi}\beta_{p2} + X_{oi}\beta_{o2})$ and σ_2, respectively. Here, (33) becomes

$$
\begin{aligned}
f_{(Y_{X_p^*}|X_o)}(Y_i, X_{pi}, X_{oi}; \pi) &= \text{cmp}(0; X_{pi}, X_{oi}, \lambda_{1i}, \sigma_1)^{I(Y_i=0)} \\
&\times \left[\left(1 - \text{cmp}(0; X_{pi}, X_{oi}, \lambda_{1i}, \sigma_1)\right) \frac{\text{cmp}_{[Y>0]}(Y_i; X_{pi}, X_{oi}, \lambda_{2i}, \sigma_2)}{1 - \text{cmp}(0; X_{pi}, X_{oi}, \lambda_{2i}, \sigma_2)} \right]^{1-I(Y_i=0)} \\
&= \left(\frac{1}{Z(\lambda_1, \sigma_1)} \right)^{I(Y_i=0)} \\
&\times \left[\left(1 - \frac{1}{Z(\lambda_1, \sigma_1)}\right) \left(\frac{\text{cmp}_{[Y>0]}(Y; X_{pi}, X_{oi}, \lambda_{2i}, \sigma_2)}{\left(1 - \frac{1}{Z(\lambda_2, \sigma_2)}\right)} \right) \right]^{1-I(Y_i=0)}
\end{aligned}
\qquad (57)
$$

where, $\pi' = [\beta_1' \quad \beta_2' \quad \sigma_1 \quad \sigma_2]$ ($\beta_1' = [\beta_{1p} \quad \beta_{1o}']$ and $\beta_2' = [\beta_{2p} \quad \beta_{2o}']$) is the vector of parameters. The relevant estimated conditional mean function for causal effect estimation is

$$
m(X_{pi}, X_{oi}; \hat{\pi}) = \left(1 - \frac{1}{Z(\hat{\lambda}_{1i}, \sigma_1)}\right) \times \left(\frac{\hat{\lambda}_{2i}^{\frac{1}{\hat{\sigma}_2}} - \frac{\hat{\sigma}_2 - 1}{2\hat{\sigma}_2}}{\left(1 - \frac{1}{Z(\hat{\lambda}_{2i}, \hat{\sigma}_2)}\right)} \right) \qquad (58)
$$

where, $\hat{\pi}' = [\hat{\beta}_1' \quad \hat{\beta}_2' \quad \hat{\sigma}_1 \quad \hat{\sigma}_2]$ is the MLE estimator of π obtained via (57), $\hat{\lambda}_{1i} = \exp(X_{pi}\hat{\beta}_{p1} + X_{oi}\hat{\beta}_{o1})$ and $\hat{\lambda}_{2i} = \exp(X_{pi}\hat{\beta}_{p2} + X_{oi}\hat{\beta}_{o2})$.

4.2 Exploring the Robustness of Maintaining NSD

As the likelihood expressions in (51), (53), (55), and (57) clearly demonstrate, under the NSD assumption, the "no 2PM needed" null (the one-part model null) is nested in the general framework as the case in which the parameters of the EM and IM are set equal to each other. We will later discuss the likelihood ratio test that follows naturally from this fact. With a view to implementing this very simple test of the very important "no 2PM needed" null hypothesis, we here conduct a preliminary examination of the restrictiveness of the NSD assumption based on theoretical considerations and simulated data. As a by-product of this analysis, we will develop and validate data generation software: Case VI in the continuous outcome context and Case IV for count-valued outcome models. Moreover, we develop and test MLE software for Cases VI, IV, and VIII. We begin by examining continuous two-part models.

4.2.1 Robustness of NSD in a Continuous Two-Part Model: Case II vs. Case VI

We focus here on a comparison of Cases II (Probit EM – GG IM) and VI (GG EM and IM). We will show there is no need for data simulation in making this comparison by showing that Case VI effectively nests Case II. Thereby we show that the Case VI model (which imposes NSD) cannot possibly be restrictive. We will, however: 1) give full analytic details of an "approximate" MLE for Case VI; 2) develop Stata/Mata software for its implementation; 3) develop relevant (Case VI) data simulation software; 4) assess the statistical consistency of our *approximate MLE* estimator/software (with regard to AIE estimation) by applying it to samples of increasing size produced by our simulation software.

Analytic Argument for Robustness of Case VI vis-a-vis Case II

Recall that the GG distribution nests the log-normal as a special case. Therefore, if the NSD is imposed as in Case VI, we know that the corresponding GG EM specification is exactly amenable to (capable of representing) the log-normal EM specification as in the EM for Case V. The EM specification in Case V is, however, identical in all relevant respects to the Probit EM specification in Case II. Therefore, Case VI (NSD) is robust vis-a-vis Case II.

Approximate MLE for Case VI

It is easy to see from (53) that κ_1 and σ_1 are unidentified parameters. Moreover, an admissible reduction of the original parameter vector that effectively eliminates all "non-coefficient" parameters in the specification, does not appear to exist.[16] We can, however, re-parameterize the model in a useful way. Our focus here is on the EM for Case VI. In particular, we seek to re-parameterize

$$\Pr(Y_{X_p^*} = 0) = \Pr(\zeta_{EM}^* < \zeta_{EM}) = GG(\zeta_{EM}; X_p^*\beta_{p1} + X_o\beta_{o1}, \kappa_1, \sigma_1)$$

where, $(\zeta_{EM}^* \mid X_o)$ is GG distributed with parameters $X_p^*\beta_{p1} + X_o\beta_{o1}$, κ_1 and σ_1, and cdf $GG(\zeta_{EM}^*; X_p^*\beta_{p1} + X_o\beta_{o1}, \kappa_1, \sigma_1)$. Let

$$\alpha = \frac{\exp(X_p^*\beta_{p1} + X_o\beta_{o1})}{\left(\frac{1}{|\kappa_1|^2}\right)^{\frac{\sigma_1}{\kappa_1}}}, \ \nu = \frac{1}{|\kappa_1|^2} \text{ and } p = \frac{\kappa_1}{\sigma_1},$$

[16] See Terza (1985) for a discussion of *admissible reductions*.

and note that $\left(\frac{\zeta_{EM}^{*}}{\alpha}\right)^{p}$ is standard gamma distributed with shape parameter ν. We have that

$$\zeta_{EM}^{*} \leq \zeta_{EM}$$

$\Leftrightarrow$

$$\left(\frac{\zeta_{EM}^{*}}{\alpha}\right)^{p} \leq \left(\frac{\zeta_{EM}}{\alpha}\right)^{p}.$$

Now

$$\left(\frac{\zeta_{EM}}{\alpha}\right)^{p} = \exp\left(\ln\left[\left(\frac{\zeta_{EM}}{\alpha}\right)^{p}\right]\right)$$

$$= \exp\left(p \ln\left(\frac{\zeta_{EM}}{\alpha}\right)\right)$$

$$= \exp\left(p \ln(\zeta_{EM}) - p\left[\ln\left(\exp(X_{p}^{*}\beta_{p1} + X_{o}\beta_{o1})\right) - \ln(\text{const})\right]\right)$$

$$= \exp\left(p \ln(\zeta_{EM}) + p \ln(\text{const}) - p(X_{p}^{*}\beta_{p1} + X_{o}\beta_{o1})\right)$$

$$= \exp\left(p[\ln(\zeta_{EM}) + \ln(\text{const}) - (X_{p}^{*}\beta_{p1} + X_{o}\beta_{o1})]\right)$$

$$= \exp(X_{p}^{*}\beta_{p1}^{o} + X_{oi}\beta_{o1}^{o})$$

where, $\text{const} = \left(\frac{1}{|\kappa_1^2|}\right)^{\frac{\sigma_1}{\kappa_1}}$, $\beta_{p1}^o = p\beta_{p1}$ and β_{o1}^o is the same as $p\beta_{o1}$ with its constant term shifted by $+p[\ln(\zeta_{EM}) + \ln(\text{const})]$. Therefore, we have

$$\Pr(\zeta_{EM}^* < \zeta_{EM}) = GG(\zeta_{EM}; X_p^*\beta_{p1} + X_o\beta_{o1}, \kappa_1, \sigma_1) = SG(\exp(X_p^*\beta_{p1}^o + X_o\beta_{o1}^o); \nu) \tag{59}$$

where, SG( ) denotes the cdf of the standard gamma random variable with shape parameter ν. Given (59), we propose the following approximation to (53)

$$\begin{aligned} f_{(Y_{X_p^*}|X_o)}^{\sim}(Y_{X_p^*}, X_p^*, X_o; \pi) = {} & SG(\exp(X_p^*\beta_{p1}^o + X_o\beta_{o1}^o); 1)^{I(Y_{X_p^*}=0)} \\ & \times \Bigg(\Big(1 - SG(\exp(X_p^*\beta_{p1}^o + X_o\beta_{o1}^o); 1)\Big) \\ & \times \frac{gg_{[Y_{X_p^*} > \zeta_{IM}]}(Y_{X_p^*}; X_p^*\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)}{1 - GG(\zeta_{IM}; X_p^*\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)}\Bigg)^{1-I(Y_{X_p^*}=0)} \end{aligned} \tag{60}$$

where, $\pi' = [\beta_1^{o\prime} \quad \beta_2' \quad \kappa_2 \quad \sigma_2^2]$ ($\beta_1^{o\prime} = [\beta_{1p}^o \quad \beta_{1o}^{o}{}']$ and $\beta_2 = [\beta_{2p} \quad \beta_{2o}']$) is the vector of parameters; $X_p^*\beta_{1p}^o + X_o\beta_{1o}^o$ is the EM index and the IM parameters are $X_p^*\beta_{p2} + X_o\beta_{o2}$, κ_2 and σ_2, respectively. We then base our *approximate MLE* on the following

$$\begin{aligned} f_{(Y_{X_p^*}|X_o)}^{\sim}(Y_i, X_{pi}, X_{oi}; \pi) = {} & SG(\exp(X_{pi}\beta_{p1}^o + X_{oi}\beta_{o1}^o); 1)^{I(Y_i=0)} \\ & \times \quad 1 - SG(\exp(X_{pi}\beta_{p1}^o + X_{oi}\beta_{o1}^o); 1) \\ & \times \frac{gg_{[Y>\zeta_{IM}]}(Y_i; X_{pi}\beta_{p2} + X_{oi}\beta_{o2}, \kappa_2, \sigma_2)}{1 - GG(\zeta_{IM}; X_{pi}\beta_{p2} + X_{oi}\beta_{o2}, \kappa_2, \sigma_2)}\Bigg\}^{1-I(Y_i=0)} . \end{aligned} \tag{61}$$

The relevant estimated conditional mean function for causal effect estimation is

$$m(X_{pi}, X_{oi}; \hat{\pi}) = \left(1 - SG(\exp(X_{pi}\beta^{o}_{p1} + X_{oi}\beta^{o}_{o1}); 1)\right) \times \left(\frac{\int_{\hat{\zeta}_{IM}}^{\infty} Y_i \, gg_{[Y_{X^*_p} > \zeta_{IM}]}(Y_i; X_{pi}\beta_{p2} + X_{oi}\beta_{o2}, \kappa_2, \sigma_2)\, dY_i}{1 - GG(\hat{\zeta}_{IM}; X_{pi}\beta_{p2} + X_{oi}\beta_{o2}, \kappa_2, \sigma_2)} \right) \qquad (62)$$

where, $\hat{\pi}' = [\hat{\beta}^{o\prime}_1 \;\; \hat{\beta}'_2 \;\; \hat{\kappa}_2 \;\; \hat{\sigma}^2_2]$ is the approximate MLE estimator of π obtained via (61) and $\hat{\zeta}_{IM}$ is the minimum order statistic for the IM subsample on Y.

Simulating Case VI Data

An issue here is the statistical consistency of the AIE estimator based on (29), (60), and (62). Although we conjecture that this estimator is indeed consistent, we do not (as yet) have formal proofs. For this reason, we developed Stata/Mata code to simulate data for the true Case VI model. The protocol for the simulator is as follows:

1) Choose values for the elements of the parameter vector

$$\pi' = [\beta'_1 \;\; \beta'_2 \;\; \kappa_1 \;\; \sigma^2_1 \;\; \kappa_2 \;\; \sigma^2_2] \; (\beta'_1 = [\beta_{1p} \;\; \beta'_{1o}] \text{ and } \beta'_2 = [\beta_{2p} \;\; \beta'_{2o}])$$

and

$$\zeta_{IM} .$$

2) Generate a sample of simulated data on X_p and X_o; each assumed to be uniformly distributed with means and variances chosen as part of the sampling design.

3) Generate a sample of outcomes at the extensive margin (Y = 0 or not) using

$$EM = I\left(U > GG(\zeta_{EM}; X_p\beta_{p1} + X_o\beta_{o1}, \kappa_1, \sigma_1)\right)$$

4) Complete the construction of the simulated sample by generating a subsample of Y values at the IM (i.e., only for those whose EM = 1).

This last step warrants some discussion because these IM values of Y must be drawn from an appropriately specified truncated GG distribution. Recall equation (39) from which it follows that (for the case in which $\kappa_2 / \sigma_2 > 0$)

$$GG^{-1}(\mathcal{P};\ X_p\beta_{p2} + X_o\beta_{o2},\ \kappa_2, \sigma_2) = \alpha \times \left(SG^{-1}(\mathcal{P}; \nu, 1)^{\frac{1}{p}} \right) \qquad (63)$$

where, $\mathcal{P}$ is a value in the unit interval

$$\alpha = \frac{\exp(X_p\beta_{p2} + X_o\beta_{o2})}{\left(\frac{1}{|\kappa_2|^2} \right)^{\frac{\sigma_2}{\kappa_2}}},\ \nu = \frac{1}{|\kappa_2|^2},\ p = \frac{\kappa_2}{\sigma_2}.$$

$GG^{-1}(\mathcal{P}; c, d, e)$ represents the inverse cdf of the GG cdf with parameters c, d, and e evaluated at $\mathcal{P}$. Now recall that (in a shorthand version of our notation above)

$$GG^{*}(Y;\ X_p\beta_{p2} + X_o\beta_{o2},\ \kappa_2, \sigma_2) = \frac{GG_{[\zeta_{IM} \le Y]}(Y;\ X_p\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)}{1 - GG(\zeta_{IM};\ X_p\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)} \qquad (64)$$

where, $GG^{*}(Y; c, d, e, \zeta)$ denotes the cdf of the GG with parameters c, d, and e truncated at ζ; and $GG(Y; c, d, e)$ represents the cdf of the GG with parameters c, d, and e. Now (64) implies that we can generate a truncated GG random variable y based on

$$[1-\mathrm{GG}(\zeta_{\mathrm{IM}};\ X_p\beta_{p2}+X_o\beta_{o2},\ \kappa_2,\sigma_2)]\times U[0,\,1]=\mathrm{GG}_{[\zeta_{\mathrm{IM}}\le Y]}(\mathcal{Y};\ X_p\beta_{p2}+X_o\beta_{o2},\ \kappa_2,\sigma_2)\,. \tag{65}$$

Ultimately, we want to be able to use $\mathrm{GG}^{-1}(\mathcal{P};\ c,\ d,\ e)$ to generate the desired random variate, but not $\mathrm{GG}^{-1}_{[\zeta\le Y]}(\mathcal{P};\ c,\ d,\ e)$. With this in mind, note that adding $\mathrm{GG}(\zeta_{\mathrm{IM}};\ X_p\beta_{p2}+X_o\beta_{o2},\ \kappa_2,\sigma_2)$ to both sides of (65), we get

$$\begin{aligned}&[1-\mathrm{GG}(\zeta_{\mathrm{IM}};\ X_p\beta_{p2}+X_o\beta_{o2},\ \kappa_2,\sigma_2)]\times U[0,\,1] + \mathrm{GG}(\zeta_{\mathrm{IM}};\ X_p\beta_{p2}+X_o\beta_{o2},\ \kappa_2,\sigma_2)\\ &\quad=\mathrm{GG}_{[\zeta\le Y]}(\mathcal{Y};\ X_p\beta_{p2}+X_o\beta_{o2},\ \kappa_2,\sigma_2) + \mathrm{GG}(\zeta_{\mathrm{IM}};\ X_p\beta_{p2}+X_o\beta_{o2},\ \kappa_2,\sigma_2)\end{aligned} \tag{66}$$

but

$$\begin{aligned}&\mathrm{GG}_{[\zeta\le Y]}(\mathcal{Y};\ X_p\beta_{p2}+X_o\beta_{o2},\ \kappa_2,\sigma_2) + \mathrm{GG}(\zeta_{\mathrm{IM}};\ X_p\beta_{p2}+X_o\beta_{o2},\ \kappa_2,\sigma_2)\,.\\ &\quad= \mathrm{GG}(\mathcal{Y};\ X_p\beta_{p2}+X_o\beta_{o2},\ \kappa_2,\sigma_2)\,.\end{aligned} \tag{67}$$

Therefore, based on (66) and (67), we have

$$\begin{aligned}&[1-\mathrm{GG}(\zeta_{\mathrm{IM}};\ X_p\beta_{p2}+X_o\beta_{o2},\ \kappa_2,\sigma_2))]\times U[0,\,1] + \mathrm{GG}(\zeta_{\mathrm{IM}};\ X_p\beta_{p2}+X_o\beta_{o2},\ \kappa_2,\sigma_2)\\ &\quad=\mathrm{GG}(\mathcal{Y};\ X_p\beta_{p2}+X_o\beta_{o2},\ \kappa_2,\sigma_2)\,,\end{aligned}$$

from which, it follows that the desired pseudo-random variate $\mathcal{Y}$ can be generated as

$$\mathcal{Y} = \mathrm{GG}^{-1}\left(\mathrm{A}; X_p\beta_{p2}+X_o\beta_{o2},\ \sigma_2,\ \kappa_2\right). \tag{68}$$

where,

$$A = [1 - GG(\zeta_{IM};\, X_p\beta_{p2} + X_o\beta_{o2},\, \kappa_2, \sigma_2)] \times U[0,\,1] + GG(\zeta_{IM};\, X_p\beta_{p2} + X_o\beta_{o2},\, \kappa_2, \sigma_2)$$

and $GG^{-1}(\bullet;\, \bullet,\, \bullet,\, \bullet)$ is defined as in (63). Therefore, to complete Step 4 above, we need to

4-a) Generate a unit uniform (U[0,1]) for the full sample

4-b) Calculate A using the design parameters, X_p, X_o and the unit uniform in 4-a

4-c) Generate y for the full sample using (68)

To complete the sample simulation, generate the full sample of Y values as

$$Y = EM \times y.$$

Validating the Consistency of the Approximate MLE of the AIE Using Simulated Data

To validate the statistical consistency of the AIE estimator based on (29), (61) and (62), we simulated samples of increasing size using the data generator detailed in the previous section and applied the approximate MLE (AMLE) estimator of the AIE to each of them.

As for the sampling design, κ_1 and σ_1 for the GG distribution determining the EM are selected to be 4 and 1, and ζ_{EM} is chosen to be 0.5. It is not necessary to change the ζ_{EM}, κ_1 and σ_1 values for the EM in different sampling designs because, as is shown in the above discussion (*Approximate MLE for Case VI*) they are not individually identified. Both X_p and X_o are drawn from the unit uniform distribution with mean of 1 and variance of 1. β_1' and β_2' are both set to be $[0.5 \ -0.5 \ 0.25]$. However, the sampling design varies along three margins:

1) 4 sets of { κ_2, σ_2 } pairs are used to determine the shape and scale of the truncated GG distribution and its other variants in the IM:

{ κ_2, σ_2 } = {1.5, 2} (truncated Generalized Gamma),

{ κ_2, σ_2 } = {2, 2} (truncated Gamma),

{ κ_2, σ_2 } = {1, 2} (truncated Weibull),

{ κ_2, σ_2 } = {0.1, 2} (truncated Log-normal);

2) 3 chosen ζ_{IM} values in the IM determining the minimum order statistic of the outcome in the IM:

ζ_{IM} = 0.1,

ζ_{IM} = 1,

ζ_{IM} = 10;

3) 9 sample sizes testing the estimation consistency of NSD models:

n = 1,000,

n = 2,500,

n = 5,000,

n = 10,000,

n = 25,000,

n = 50,000,

n = 100,000,

n = 250,000,

n = 500,000.

Based on the above sampling design, we are going to conduct 4×3×9 = 108 experiments. In each experiment, we

1) Calculate the true AIE, which should be the same regardless of different sample sizes;

2) Estimate the deep parameters using 2PMs with NSD assumption;

3) Calculate the estimated AIEs, which are compared to the true AIE.

The comparison criteria is Absolute Percentage Bias (APB) calculated by the following formula,

$$\text{APB } \text{AIE}(\Delta) = \left| \frac{\text{AIE}(\Delta) - \text{AIE}(\Delta)}{\text{AIE}(\Delta)} \right| \qquad (70)$$

where, $\text{AIE}(\Delta)$ denotes the estimated AIE with the increment being equal to Δ and $\text{AIE}(\Delta)$ denotes the true AIE value. Without loss of generality, we choose $\Delta = 1$. Table 1-(1) through Table 1-(4) display the simulation results, which empirically validate the theoretical consistency of *approximate MLE based AIE estimator* [(29), (61) and (62)]. In addition, the results in Table 1-(1) through Table 1-(4) have other important implications. First, the zeta values in the IM do not play an important role in order to obtain the unbiased estimated AIEs. In other words, the minimum order statistic in the IM should not be a concern while estimating the policy parameters. This supports the practicality of the *approximate MLE based AIE estimator* (AMLE-AIE) for real data analyses, wherein the minimum order statistic may vary widely across empirical estimation contexts. Second, the AMLE-AIE works well for continuous outcomes having different distributions.

4.2.2 Robustness of NSD in a Count-Valued Two-Part Model: Case IV vs. Case VIII

To shed light on the robustness of the NSD assumption in two-part models with count-valued outcomes, we compare Case IV (Probit EM – CMP IM; the conventional two-part model with CMP IM) with case VIII (CMP EM and IM). Unlike the continuous case discussed in section 4.3.1, there is no clear way to conclusively make this comparison analytically. Instead, we base the present comparison on simulated data. We do this by: 1) generating data based on a sampling design that comports with Case IV (the conventional two-part model with Probit EM and CMP IM); 2) calculating the true AIE based on the chosen sampling design, the true AIE formulation in (5), and the specification of the relevant conditional mean given in (49); 3) applying the MLEs based on (45) and (57) to this Case IV simulated data; 4) estimating the AIE as in (29) using the deep parameter estimates from step (3), based first on (49) and then on (58). If step (4) yields similar (near identical) results we will conclude that the NSD assumption is not restrictive (is robust) in this count-valued context.

Simulating Case IV Data

In order to conduct the simulation study, we developed Stata/Mata code to simulate data for the true Case IV model. The protocol for the simulator is as follows:

1) Choose values for the elements of the parameter vector

$$\pi' = [\beta_1' \quad \beta_2' \quad \sigma_2] \; (\beta_1' = [\beta_{1p} \quad \beta_{1o}'] \text{ and } \beta_2' = [\beta_{2p} \quad \beta_{2o}']).$$

2) Generate a sample of simulated data on X_p and X_o; each assumed to be uniformly distributed with means and variances chosen as part of the sampling design.

3) Generate a sample of outcomes at the extensive margin ($Y = 0$ or not) using

$$EM = I\left(X_p\beta_{p1} + X_o\beta_{o1} + \varepsilon^{EM} > 0\right)$$

4) Complete the construction of the simulated sample by generating a subsample of Y values at the IM (i.e., only for those whose $EM = 1$).

This last step warrants some discussion because these IM values of Y must be drawn from an appropriately specified truncated CMP distribution. Here, expression (28) serves as the template. From (28) it follows that simulated outcome draws for the IM can be obtained from the following

$$\mathcal{Y} = \begin{cases} 1 \;\text{ iff }\; \mathcal{G}_0^{EM} \le U < \mathcal{G}_0^{EM} + (1-\mathcal{G}_0^{EM})\dfrac{cmp(1)}{1-cmp(0)} \\ 2 \;\text{ iff }\; \mathcal{G}_0^{EM} + (1-\mathcal{G}_0^{EM})\dfrac{cmp(1)}{1-cmp(0)} \le U < \mathcal{G}_0^{EM} + (1-\mathcal{G}_0^{EM})\left[\dfrac{cmp(1)+cmp(2)}{1-cmp(0)}\right] \\ \ldots \\ j \;\text{ iff }\; \mathcal{G}_0^{EM} + (1-\mathcal{G}_0^{EM})\left[\dfrac{\sum_{m=1}^{j-1} cmp(m)}{1-cmp(0)}\right] \le U < \mathcal{G}_0^{EM} + (1-\mathcal{G}_0^{EM})\left[\dfrac{\sum_{m=1}^{j} cmp(m)}{1-cmp(0)}\right] \\ \ldots \end{cases} \tag{71}$$

where, $\mathcal{G}_0^{EM}$ is shorthand for $\mathcal{G}_{(0|X_o)}^{EM}(0, X_p, X_o; \tau_{EM})$ which, in the Case IV model, is replaced by $1-\Phi(X_p\beta_{p1} + X_o\beta_{o1})$, and $cmp(\mathcal{Y})$ is shorthand for the CMP pmf, $cmp(\mathcal{Y}; X_p\beta_{p2} + X_o\beta_{o2}, \sigma_2)$ for $\mathcal{Y} = 1,\ldots,\infty$. Therefore, to complete Step 4 above, we need to

4-i) Generate a unit uniform $(\mathrm{U}[0,1])$ for the full sample

4-ii) Calculate

$$\mathcal{G}_0^{\mathrm{EM}} + (1-\mathcal{G}_0^{\mathrm{EM}})\left[\frac{\sum_{m=1}^{j-1}\mathrm{cmp(m)}}{1-\mathrm{cmp(0)}}\right]$$

for all relevant values of j using the design parameters, $\mathrm{X_p}$ and $\mathrm{X_o}$

4-iii) Use $\mathrm{U}[0,1]$ from 4-i) and the values obtained from 4-ii) to generate y for the full sample via (71)

To complete the sample simulation, generate the full sample of Y values as

Y = EM × y.

Investigating the Robustness of the NSD (Case IV vs. Case VIII) Using Simulated Data

To investigate the robustness of the AIE estimator by the count-valued 2PM with NSD assumption, we simulated different samples using the data generator detailed in the previous section and applied the MLE-based estimator of the AIE to each of them.

As for the sampling design, both $\mathrm{X_p}$ and $\mathrm{X_o}$ are drawn from unit uniform distributions with mean and variance equal to 1. The sampling design varied along two margins:

1) 4 EM specifications, i.e., $\beta_1' = [\beta_{1p} \quad \beta_{1o}']$ corresponding to different probabilities for non-zero outcome values:

$\beta_1' = [0.5 \;\; -2 \;\;\; -1]$ – 85% of non-zero outcome values,

$\beta_1' = [0.5 \;\; -2 \;\; 0.25]$ – 68% of non-zero outcome values,

$\beta_1' = [0.5\ \ -0.5\ \ 0.25]$ – 43% of non-zero outcome values,

$\beta_1' = [1\ \ 0.5\ \ 0.25]$ – 13% of non-zero outcome values;

2) 5 values of the dispersion parameter corresponding to the count-data dispersion scenarios:

$\sigma_2 = -1$,

$\sigma_2 = -0.5$,

$\sigma_2 = 0$,

$\sigma_2 = 0.5$,

$\sigma_2 = 1$.

For simplicity, β_2 is set equal to β_1.[17] The 20 resulting sampling designs are detailed in Table 2-(1). For each of these designs, 100 samples of size 10,000 were generated. Two performance criteria were used: the grand average of the estimated AIEs

$$\text{Average AIE}(\Delta) = \frac{1}{100} \times \sum_{j=1}^{100} \left(\text{AIE}(\Delta)_j \right) \tag{72}$$

where, j denotes the j^th replication, and the grand average of the absolute percentage bias

$$\text{AAPB AIE}(\Delta) = \frac{1}{100} \times \sum_{j=1}^{100} \left| \frac{\text{AIE}(\Delta)_j - \text{AIE}(\Delta)}{\text{AIE}(\Delta)} \right|. \tag{73}$$

[17] In this simulation study, the true specifications of EM and IM have different functional forms: more specifically, Probit and truncated GG. Thus, β_2 in the IM can be any values because they are of no special interest.

The policy increment Δ is set equal to 1. True AIE, Average $\text{AIE}(\Delta)$ and AAPB $\text{AIE}(\Delta)$ for each design are shown in Table 2-(2). The main take-away from the results in Table 2-(2) is that the AIE estimates obtained from the version of the model with the incorrect NSD assumption imposed are virtually the same as those obtained using the unrestricted model that comports with the design used to simulate the data. We take this as preliminary evidence that imposition of the NSD condition may not materially affect the results.

4.3 A Test of No Parametric Difference between EM and IM with NSD Maintained

If, as the preliminary simulation results of the previous sections support, estimation of the targeted causal effect is robust to the assumption of NSD between the EM and IM, then it is clear from the likelihood specifications in (53) and (61), for the continuous case, and (57), for the count-data case, that a conventional likelihood ratio statistic can be used to test the null hypothesis that a 2PM structure is not needed (the one-part model null). In the GG-based context for the continuous case, this null is tantamount to imposing the restriction that $[\beta_1' \quad \kappa_1 \quad \sigma_1] = [\beta_2' \quad \kappa_2 \quad \sigma_2]$ and $\zeta_{\text{EM}} = \zeta_{\text{IM}}$ on (53). Similarly, in the CMP-based context for the count-valued case, the one-part null corresponds to imposing the restriction that $[\beta_1' \quad \sigma_1] = [\beta_2' \quad \sigma_2]$ on (57). In this section, we detail the one-part null versions of the relevant continuous and count-data models, along with the corresponding likelihood ratio tests.

4.3.1 Generalized Gamma Model with NSD Under the One-Part Null

Recall that the likelihood function of the GG model with NSD (Case VI) follows from the specification in (53). Under the one-part null hypothesis

$(H_o: [\beta_1' \quad \kappa_1 \quad \sigma_1] = [\beta_2' \quad \kappa_2 \quad \sigma_2] = [\beta' \quad \kappa \quad \sigma]$ and $\zeta_{EM} = \zeta_{IM} = \zeta)$, (53) becomes

$$\begin{aligned} f_{(Y_{X_p^*}|X_o)}(Y_{X_p^*}, X_p^*, X_o; \pi) &= GG(\zeta; X_p^*\beta_p + X_o\beta_o, \kappa, \sigma)^{I(Y_{X_p^*}=0)} \\ &\quad \times \left(1 - GG(\zeta; X_p^*\beta_p + X_o\beta_o, \kappa, \sigma) \right. \\ &\quad \left. \times \frac{gg_{[Y_{X_p^*}>\zeta]}(Y_{X_p^*}; X_p^*\beta_p + X_o\beta_o, \kappa, \sigma)}{1 - GG(\zeta; X_p^*\beta_p + X_o\beta_o, \kappa, \sigma)} \right)^{1-I(Y_{X_p^*}=0)} \\ &= GG(\zeta; X_p^*\beta_p + X_o\beta_o, \kappa, \sigma)^{I(Y_{X_p^*}=0)} \\ &\quad \times gg_{[Y_{X_p^*}>\zeta]}(Y_{X_p^*}; X_p^*\beta_p + X_o\beta_o, \kappa, \sigma)^{1-I(Y_{X_p^*}=0)} \end{aligned} \tag{74}$$

where, $\pi' = [\beta' \quad \kappa \quad \sigma]$ $(\beta' = [\beta_p \quad \beta_o'])$. Therefore, in the continuous (GG-based) case, the relevant log-likelihood function under the one-part null follows from

$$\begin{aligned} f_{(Y_{X_p^*}|X_o)}(Y_i, X_{pi}, X_{oi}; \pi) &= GG(\zeta; X_{pi}\beta_p + X_{oi}\beta_o, \kappa, \sigma)^{I(Y_i=0)} \\ &\quad \times gg_{[Y_i>\zeta]}(Y_i; X_{pi}\beta_p + X_{oi}\beta_o, \kappa, \sigma)^{1-I(Y_i=0)}. \end{aligned} \tag{75}$$

Given (75) and (61), a simple likelihood ratio (LR) statistic can be used to test the NSD (unconstrained) specification against the one-part (constrained) specification, viz.,

$$LR = -2 \times (\hat{L}_{\text{one-part}} - \hat{L}_{NSD}) \tag{76}$$

where, $\hat{L}_{\text{one-part}}$ denotes the maximized version of the log-likelihood function with (75) with respect to π and $\hat{\pi}$ being its maximizer; $\hat{L}_{NSD}$ denotes the maximized version of the log-likelihood function with (61) with respect to π and $\hat{\pi}$ being estimated parameters by the NSD model. In this case, the LR follows χ_q^2 distributions with degrees of freedom q being the difference in dimensionality of deep parameters between two model specifications in (75) and (61).

4.3.2 Conway-Maxwell Poisson Model with NSD Under the One-Part Null

Recall that the likelihood function of the Conway-Maxwell Poisson model with NSD (Case VIII) follows from the specification in (57). Under the one-part null hypothesis ($H_o : [\beta_1' \quad \sigma_1] = [\beta_2' \quad \sigma_2] = [\beta' \quad \sigma]$ which implies $\lambda_1 = \lambda_2 = \lambda$), (57) becomes

$$f_{(Y_{X_p^*}|X_o)}(Y_i, X_{pi}, X_{oi}; \pi) = \left(\frac{1}{Z(\lambda, \sigma)}\right)^{I(Y_i=0)}$$

$$\times \left[\left(1 - \frac{1}{Z(\lambda, \sigma)}\right)\left(\frac{cmp_{[Y>0]}(Y; X_{pi}, X_{oi}, \lambda_i, \sigma)}{\left(1 - \frac{1}{Z(\lambda, \sigma)}\right)}\right)\right]^{1-I(Y_i=0)}$$

$$= \left(\frac{1}{Z(\lambda, \sigma)}\right)^{I(Y_i=0)} \times cmp_{[Y>0]}(Y; X_{pi}, X_{oi}, \lambda_i, \sigma)^{1-I(Y_i=0)}$$

$$= cmp(Y; X_{pi}, X_{oi}, \lambda_i, \sigma) \qquad (77)$$

where, $\pi' = [\beta' \quad \kappa \quad \sigma]$ ($\beta' = [\beta_p \quad \beta_o']$) and $\lambda_i = \exp(X_{pi}\beta_p + X_{oi}\beta_o)$. Therefore, in the count-valued (CMP-based) case, the relevant log-likelihood function under the one-part null is that of the simple CMP model. Given (77) and (57), a simple likelihood ratio (LR) statistic akin to (76) for the continuous (GG-based) case can be used to test the NSD (unconstrained) specification against the one-part (constrained) specification.

Table 1-(1). Simulation Results of Continuous 2PM with NSD

$\{\kappa_2, \sigma_2\} = \{1.5, 2\}$

	Zeta=0.1		Zeta=1		Zeta=10	
True AIE	1.0711		1.3839		2.4714	
n	Est AIE	Abs Pct Bias	Est AIE	Abs Pct Bias	Est AIE	Abs Pct Bias
1,000	1.1670	8.95%	1.4117	2.00%	2.2734	8.01%
2,500	0.8663	19.12%	1.1508	16.84%	2.1379	13.49%
5,000	1.0276	4.07%	1.3575	1.91%	2.5723	4.08%
10,000	1.0431	2.61%	1.3633	1.49%	2.5893	4.77%
25,000	1.0279	4.04%	1.3323	3.73%	2.3713	4.05%
50,000	1.0940	2.14%	1.4197	2.58%	2.5676	3.89%
100,000	1.0664	0.44%	1.3773	0.48%	2.4402	1.26%
250,000	1.0892	1.68%	1.4052	1.54%	2.5182	1.90%
500,000	1.0903	1.79%	1.4056	1.56%	2.4959	0.99%

Table 1-(2). Simulation Results of Continuous 2PM with NSD

$\{\kappa_2, \sigma_2\} = \{2, 2\}$

	Zeta=0.1		Zeta=1		Zeta=10	
True AIE	0.8669		1.1544		2.0749	
n	Est AIE	Abs Pct Bias	Est AIE	Abs Pct Bias	Est AIE	Abs Pct Bias
1,000	0.9434	8.82%	1.1557	0.12%	1.8422	11.21%
2,500	0.7188	17.08%	0.9935	13.94%	1.7963	13.42%
5,000	0.8408	3.02%	1.1477	0.58%	2.1763	4.89%
10,000	0.8570	1.14%	1.1523	0.18%	2.2051	6.28%
25,000	0.8316	4.07%	1.1150	3.41%	1.9680	5.15%
50,000	0.8873	2.34%	1.1869	2.82%	2.1438	3.32%
100,000	0.8665	0.04%	1.1501	0.37%	2.0364	1.85%
250,000	0.8801	1.52%	1.1704	1.39%	2.1023	1.32%
500,000	0.8801	1.52%	1.1698	1.34%	2.0749	0.00%

Table 1-(3). Simulation Results of Continuous 2PM with NSD

$\{\kappa_2, \sigma_2\} = \{1, 2\}$

	Zeta=0.1		Zeta=1		Zeta=10	
True AIE	1.4275		1.7800		3.1621	
n	Est AIE	Abs Pct Bias	Est AIE	Abs Pct Bias	Est AIE	Abs Pct Bias
1,000	1.5386	7.78%	1.8578	4.37%	3.0000	5.13%
2,500	1.1157	21.85%	1.4064	20.99%	2.6570	15.98%
5,000	1.3547	5.10%	1.7142	3.70%	3.2129	1.61%
10,000	1.3634	4.49%	1.7196	3.39%	3.2202	1.84%
25,000	1.3709	3.96%	1.7052	4.20%	3.0353	4.01%
50,000	1.4514	1.68%	1.8185	2.16%	3.2787	3.68%
100,000	1.4113	1.13%	1.7663	0.77%	3.1261	1.14%
250,000	1.4542	1.87%	1.8118	1.79%	3.2219	1.89%
500,000	1.4569	2.06%	1.8134	1.87%	3.2075	1.43%

Table 1-(4). Simulation Results of Continuous 2PM with NSD

$\{\kappa_2, \sigma_2\} = \{0.5, 2\}$

	Zeta=0.1		Zeta=1		Zeta=10	
True AIE	2.1254		2.5414		4.4353	
n	Est AIE	Abs Pct Bias	Est AIE	Abs Pct Bias	Est AIE	Abs Pct Bias
1,000	2.2328	5.05%	2.7645	8.78%	4.4559	0.46%
2,500	1.6022	24.61%	1.8870	25.75%	3.5214	20.60%
5,000	2.0109	5.39%	2.3948	5.77%	4.3440	2.06%
10,000	1.9995	5.92%	2.4086	5.23%	4.3963	0.88%
25,000	2.0351	4.25%	2.4090	5.21%	4.2170	4.92%
50,000	2.1416	0.76%	2.5777	1.43%	4.5778	3.21%
100,000	2.0806	2.11%	2.5103	1.23%	4.4041	0.70%
250,000	2.1691	2.06%	2.5984	2.24%	4.5311	2.16%
500,000	2.1721	2.20%	2.5967	2.18%	4.5195	1.90%

Table 2-(1). Simulation Parameter Design of Count-Valued 2PM with NSD

Design	P^0	β_1' (EM)	β_2' (IM)	ν
[1]	84.82%	[0.5, -2, -1]	[0.5, -2 -1]	1; 0.5; 0; -0.5; -1
[2]	68.33%	[0.5, -2, 0.25]	[0.5, -2, 0.25]	
[3]	42.83%	[0.5, -0.5, 0.25]	[0.5, -0.5, 0.25]	
[4]	12.79%	[1, 0.5, 0.25]	[1, 0.5, 0.25]	

Table 2-(2). Simulation Results of Count-Valued 2PM with NSD

			CMP – Unrestricted		CMP – With NSD	
Design	ν	True AIE	Average AIE(Δ)	AAPB AIE(Δ)	Average AIE(Δ)	AAPB AIE(Δ)
[1]	-1	1.865	1.655	10.25%	1.680	8.95%
	-0.5	0.749	0.663	10.46%	0.671	9.44%
	0	0.277	0.250	9.29%	0.249	9.43%
	0.5	0.142	0.135	5.79%	0.134	6.21%
	1	0.089	0.091	4.80%	0.090	4.46%
[2]	-1	7.524	7.268	3.69%	7.249	3.80%
	-0.5	3.397	3.265	3.31%	3.264	3.32%
	0	0.968	0.928	3.37%	0.927	3.53%
	0.5	0.322	0.314	3.00%	0.312	3.25%
	1	0.157	0.157	2.98%	0.156	2.93%
[3]	-1	8.601	8.108	4.74%	8.268	2.92%
	-0.5	3.382	3.221	3.81%	3.280	2.35%
	0	1.063	1.040	2.59%	1.055	2.45%
	0.5	0.455	0.472	4.87%	0.477	5.92%
	1	0.241	0.274	14.43%	0.275	15.20%
[4]	-1	140.085	141.526	3.94%	141.561	3.95%
	-0.5	64.458	64.790	2.90%	64.797	2.91%
	0	17.670	17.742	1.83%	17.742	1.84%
	0.5	3.033	3.038	1.33%	3.038	1.31%
	1	0.933	0.923	1.33%	0.922	1.31%

Chapter 5.

An Application: Estimating Price Effects on Alcohol Demand

This chapter illustrates the econometric issues discussed in the previous chapters by presenting an empirical analysis of price effects on alcohol demand. Accurate estimation of price effects is especially important to the assessment of policies aimed at reducing alcohol abuse and alcoholism (Coate and Grossman 1988, Manning et al. 1995 and Chaloupka et al. 2002). As Manning et al. (1995) point out, alcohol demand is well-suited to two-part modeling because many individuals choose not to drink and that pick-up decision may differ systematically from quantity of consumption decisions made by those who have chosen to drink. We will focus on the two potential outcomes based causally interpretable price effect parameters defined in (1) and (2) – the *average incremental effect of price on demand for alcohol* (AIE) and the *average price elasticity of demand for alcohol* (AED). We will estimate the aforementioned effect parameters under two different model specifications: 1) the conventional 2PM with Probit EM and GG IM using MLE based on (37) (Case II); 2) the GG 2PM with NSD assumption using the *approximate MLE* based on (61) (Case VI); 3) the GG one-part model using MLE based on (75). In the Case II context, for the AIE we have

$$\mathrm{AIE}(\Delta)^{\mathrm{II}}$$

$$= \mathrm{E}\left[\Phi((X_p^* + \Delta)\beta_{p1} + X_o\beta_{o1}) \frac{\int_{\zeta_{IM}}^{\infty} Y_{X_p^*} gg_{[Y_{X_p^*} > \zeta_{IM}]}(Y_{X_p^*}; (X_p^* + \Delta)\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)\, dY_{X_p^*}}{1 - GG(\zeta_{IM}; (X_p^* + \Delta)\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)}\right]$$

$$
- \mathrm{E}\left[\Phi(X_p^* \beta_{p1} + X_o \beta_{o1}) \frac{\int_{\zeta}^{\infty} Y_{X_p^*} gg_{[Y_{X_p^*} > \zeta_{IM}]}(Y_{X_p^*}; X_p^* \beta_{p2} + X_o \beta_{o2}, \kappa_2, \sigma_2)\, dY_{X_p^*}}{1 - GG(\zeta_{IM}; X_p^* \beta_{p2} + X_o \beta_{o2}, \kappa_2, \sigma_2)} \right] \tag{78}
$$

and its corresponding estimator can be specified as

$$
\mathrm{AIE}(\Delta)^{\mathrm{II}}
$$

$$
= \sum_{i=1}^{n} \frac{1}{n} \left[\Phi((X_{pi} + \Delta)\hat{\beta}_{p1} + X_o \hat{\beta}_{o1}) \frac{\int_{\hat{\zeta}_{IM}}^{\infty} Y_i gg_{[Y_{X_p^*} > \zeta_{IM}]}(Y_i; (X_{pi} + \Delta)\hat{\beta}_{p2} + X_{oi} \hat{\beta}_{o2}, \hat{\kappa}_2, \hat{\sigma}_2)\, dY_i}{1 - GG(\hat{\zeta}_{IM}; (X_{pi} + \Delta)\hat{\beta}_{p2} + X_{oi} \hat{\beta}_{o2}, \hat{\kappa}_2, \hat{\sigma}_2)} \right]
$$

$$
- \sum_{i=1}^{n} \frac{1}{n} \left[\Phi(X_{pi} \hat{\beta}_{p1} + X_o \hat{\beta}_{o1}) \frac{\int_{\hat{\zeta}_{IM}}^{\infty} Y_i gg_{[Y_{X_p^*} > \zeta_{IM}]}(Y_i; X_{pi} \hat{\beta}_{p2} + X_{oi} \hat{\beta}_{o2}, \hat{\kappa}_2, \hat{\sigma}_2)\, dY_i}{1 - GG(\hat{\zeta}_{IM}; X_{pi} \hat{\beta}_{p2} + X_{oi} \hat{\beta}_{o2}, \hat{\kappa}_2, \hat{\sigma}_2)} \right]. \tag{79}
$$

For the AIE in the Case VI context, we have

$$
\mathrm{AIE}(\Delta)^{\mathrm{VI}}
$$

$$
= \mathrm{E}\Big[\ 1 - SG\big[\exp\ (X_p^* + \Delta)\beta_{p1}^{o} + X_o \beta_{o1}^{o}\ ,1\big]
$$

$$
\times \frac{\int_{\zeta_{IM}}^{\infty} Y_{X_p^*} gg_{[Y_{X_p^*} > \zeta_{IM}]}(Y_{X_p^*}; (X_p^* + \Delta)\beta_{p2} + X_o \beta_{o2}, \kappa_2, \sigma_2)\, dY_{X_p^*}}{1 - GG(\zeta_{IM}; (X_p^* + \Delta)\beta_{p2} + X_o \beta_{o2}, \kappa_2, \sigma_2)} \Bigg]
$$

$$
- \mathrm{E}\Big[\ 1 - SG\big[\exp\ X_p^* \beta_{p1}^{o} + X_o \beta_{o1}^{o}\ ,1\big]
$$

$$\times \frac{\int_{\zeta}^{\infty} Y_{X_p^*} gg(Y_{X_p^*}; X_p^*\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)\, dY_{X_p^*}}{1 - GG(\zeta; X_p^*\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)}\Bigg] \tag{80}$$

and its corresponding estimator can be specified as

$$AIE(\Delta)^{VI}$$

$$= \sum_{i=1}^{n} \frac{1}{n}\Bigg[1 - SG\Big[\exp\ (X_{pi} + \Delta)\hat{\beta}_{p1}^{o} + X_{oi}\hat{\beta}_{o1}^{o}\ \ ,1\Big]$$
$$\times \frac{\int_{\hat{\zeta}_{IM}}^{\infty} Y_i gg_{[Y_{X_p^*} > \zeta_{IM}]}(Y_i; (X_{pi} + \Delta)\hat{\beta}_{p2} + X_o\hat{\beta}_{o2}, \hat{\kappa}_2, \hat{\sigma}_2)\, dY_i}{1 - GG(\hat{\zeta}_{IM}; (X_{pi} + \Delta)\hat{\beta}_{p2} + X_o\hat{\beta}_{o2}, \hat{\kappa}_2, \hat{\sigma}_2)}\Bigg]$$
$$- \sum_{i=1}^{n} \frac{1}{n}\Bigg[1 - SG\Big[\exp\ X_{pi}\hat{\beta}_{p1}^{o} + X_{oi}\hat{\beta}_{o1}^{o}\ \ ,1\Big]$$
$$\times \frac{\int_{\hat{\zeta}_{IM}}^{\infty} Y_i gg_{[Y_{X_p^*} > \zeta_{IM}]}(Y_i; X_{pi}\hat{\beta}_{p2} + X_o\hat{\beta}_{o2}, \hat{\kappa}_2, \hat{\sigma}_2)\, dY_i}{1 - GG(\hat{\zeta}_{IM}; X_{pi}\hat{\beta}_{p2} + X_o\hat{\beta}_{o2}, \hat{\kappa}_2, \hat{\sigma}_2)}\Bigg]. \tag{81}$$

In the Case II context for the AED we have

$$AED(\Delta)^{II}$$

$$= \frac{AIE(\Delta)^{II}}{\Delta} \times \frac{E[X_p^*]}{\Phi(X_p^*\beta_{p1} + X_o\beta_{o1}) \frac{\int_{\zeta}^{\infty} Y_{X_p^*} gg_{[Y_{X_p^*} > \zeta_{IM}]}(Y_{X_p^*}; X_p^*\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)\, dY_{X_p^*}}{1 - GG(\zeta_{IM}; X_p^*\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)}} \tag{82}$$

and its corresponding estimator

$$
\text{AED}(\Delta)^{\text{II}}
$$

$$
= \frac{\text{AIE}(\Delta)^{\text{II}}}{\Delta} \times \frac{\sum_{i=1}^{n} \frac{1}{n} \ X_{pi}}{\sum_{i=1}^{n} \frac{1}{n} \left[\Phi(X_{pi}\hat{\beta}_{p1} + X_{o}\hat{\beta}_{o1}) \frac{\int_{\hat{\zeta}_{IM}}^{\infty} Y_i gg_{[Y_{X_p^*} > \zeta_{IM}]}(Y_i; X_{pi}\hat{\beta}_{p2} + X_{oi}\hat{\beta}_{o2}, \hat{\kappa}_2, \hat{\sigma}_2) \ dY_i}{1 - GG(\hat{\zeta}_{IM}; X_{pi}\hat{\beta}_{p2} + X_{oi}\hat{\beta}_{o2}, \hat{\kappa}_2, \hat{\sigma}_2)} \right]} . \tag{83}
$$

For the AED in the Case VI context we have

$$
\text{AED}(\Delta)^{\text{VI}}
$$

$$
= \frac{\text{AIE}(\Delta)^{\text{VI}}}{\Delta}
$$

$$
\times \frac{E[X_p^*]}{E\left[1 - SG\left[\exp \ \ X_p^*\beta_{p1}^o + X_o\beta_{o1}^o \ \ ,1\right] \times \frac{\int_{\zeta}^{\infty} Y_{X_p^*} gg(Y_{X_p^*}; X_p^*\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2) \ dY_{X_p^*}}{1 - GG(\zeta; X_p^*\beta_{p2} + X_o\beta_{o2}, \kappa_2, \sigma_2)} \right]} \tag{84}
$$

and its corresponding estimator can be specified as

$$
\text{AED}(\Delta)^{\text{VI}}
$$

$$
= \frac{\text{AIE}(\Delta)^{\text{VI}}}{\Delta}
$$

$$
\times \frac{\sum_{i=1}^{n} \frac{1}{n} \ X_{pi}}{-\sum_{i=1}^{n} \frac{1}{n} \left[1 - SG\left[\exp \ \ X_{pi}\hat{\beta}_{p1}^o + X_{oi}\hat{\beta}_{o1}^o \ \ ,1\right] \times \frac{\int_{\hat{\zeta}_{IM}}^{\infty} Y_i gg_{[Y_{X_p^*} > \zeta_{IM}]}(Y_i; X_{pi}\hat{\beta}_{p2} + X_o\hat{\beta}_{o2}, \hat{\kappa}_2, \hat{\sigma}_2) \ dY_i}{1 - GG(\hat{\zeta}_{IM}; X_{pi}\hat{\beta}_{p2} + X_o\hat{\beta}_{o2}, \hat{\kappa}_2, \hat{\sigma}_2)} \right]} . \tag{85}
$$

5.1 Data

The data used in this chapter combines the Uniform Product Code (UPC) and the National Epidemiological Survey of Alcohol and Related Conditions (NESARC). This combined dataset has been recognized as a reliable dataset for estimating the price elasticity of alcohol demand and has been widely used in the economics research (Ruhm et al., 2012; Terza et al., 2016). The UPC dataset collected by AC Nielsen in grocery stores from 51 U.S. markets provides accurate information on beer prices by type of beverage and packaging size and it is also a reliable dataset to estimate price elasticity of beer consumption. As Ruhm et al. (2012) mention, beer price as provided by the UPC is a volume-weighted average price in the available grocery stores. The NESARC dataset sponsored by the National Institute on Alcohol Abuse and Alcoholism has two waves (NESARC1 2001-2002 and NESARC2 2004-2005) and provides nationally representative information regarding alcohol consumption, alcohol use disorders, treatment services as well as demographic characteristics of respondents. Due to inherent limitations on the UPC data, only NESARC2 can be merged with UPC beer price information. The sample size for this study is 23,743.

The outcome variable is average daily ethanol consumption in ounces from beer during the past year. The policy variable of main interest is the nominal price of beer in U.S. dollars per ounce of ethanol. The other covariates include respondent's gender, marital status, age, race, family size, education, census region, occupation, and household income. The descriptive statistics for analysis variables can be found in Table 3. It is noticed that 36% of respondents report any beer consumption in the past year and there are slight differences in the sample means of the covariates for beer drinkers and non-drinkers.

5.2 Empirical Results

Table 4 presents fully parametric maximum likelihood estimation results based on the conventional 2PM (Case II), 2PM with NSD (Case VI), and one-part model characterized in (37), (59) and (75) respectively. Column (1) and Column (2) present coefficient estimations for whether the respondent is a beer drinker or not, i.e., the EM of the 2PM, for the Probit EM and GG EM, respectively. There is not much sense to compare the signs and magnitudes of estimators in these two columns due to their different functional forms. Column (3) presents coefficient estimation results for the level of beer consumption conditional on any positive consumption, i.e., the GG IM. The price of beer per ounce of ethanol has significant negative effects on both being a beer drinker and the volume of beer consumed if one is a drinker. Column (4) presents coefficient estimation results for the GG one-part model, assuming that there is no systematic difference between drinkers and non-drinkers. Only in the GG one-part model, beer price does not show a statistically significant effect on alcohol consumption.

Table 5 presents the AED and AIE estimates, the parameter estimates of particular policy interest, which are calculated using the regression results and the formulations in (78) through (85) given above. The estimated AEDs for Cases II and VI are -0.6838 and -0.6762, respectively. Estimated AIEs are -0.0950 and -0.0926, which are also quantitatively and qualitatively similar between the two models; in other words, a hypothetical increase in beer price by 1 dollar will cause the overall mean of daily beer consumption to decrease by similar amount of ounces under both models. As the empirical results show, the estimated parameters for the Case II and Case VI are very close to each other. However, under the GG one-part model, estimated AED and AIE are very different from 2PMs in magnitudes. Even though we do not know the true DGP in this case, Chapter 4.4 provides a statistical test on whether the two-part structure is needed.

5.3 Likelihood Ratio Test on One-Part Null

As shown in Chapter 4.4, we apply the likelihood ratio test in this application to test whether the distinction between the EM and IM is necessary. $\hat{L}_{\text{one-part}} = -11770.42$, $\hat{L}_{\text{NSD}} = -8996.97$, $LR = -2\times(\hat{L}_{\text{one-part}} - \hat{L}_{\text{NSD}}) = 5546.9$, and $\Pr(\chi^2_{21} > 5546.9) = 0$. Based on the LR statistic, the null hypothesis that no two-part structure is needed gets rejected. The AED and AIE estimations under the 2PMs should be considered in this empirical study.

Table 3. Descriptive Statistics

Variable Name	Full Sample		Beer Drinker		Non-Beer Drinker	
	Mean	SD	Mean	SD	Mean	SD
Outcomes						
If Any Beer	0.360	0.480				
Beer Ethanol Consumption	0.153	0.694	0.424	1.106		
Policy Variable						
Nominal Beer Price	1.253	0.104				
Covariates						
ln (Age)	3.818	0.365	3.742	0.341	3.861	0.371
ln (Income)	10.580	0.923	10.782	0.895	10.467	0.920
ln (Family Size)	0.819	0.578	0.847	0.566	0.804	0.585
Proportion of Sample That Is:						
Female	0.581	0.493	0.403	0.490	0.681	0.466
Married	0.505	0.500	0.520	0.500	0.496	0.500
Black	0.207	0.405	0.152	0.359	0.237	0.426
Hispanic	0.220	0.414	0.218	0.413	0.221	0.415
Other Race	0.044	0.206	0.039	0.194	0.047	0.212
No High School	0.156	0.363	0.114	0.318	0.180	0.384
Some College	0.317	0.465	0.328	0.470	0.311	0.463
College	0.275	0.446	0.328	0.470	0.245	0.430
Midwest Region	0.215	0.411	0.229	0.420	0.207	0.406
South Region	0.390	0.488	0.358	0.479	0.408	0.492
West Region	0.249	0.433	0.271	0.444	0.237	0.425
Blue Collar	0.152	0.359	0.186	0.389	0.133	0.340
White Collar	0.538	0.499	0.591	0.492	0.509	0.500
Service Occupation	0.151	0.358	0.140	0.347	0.157	0.364
N	23,743		8,542		15,201	

Table 4. Regression Results

	Probit EM			GG EM			GG IM			GG One-Part Model		
	Column (1)			Column (2)			Column (3)			Column (4)		
	beta	s.e.		beta	s.e.		beta	s.e.		beta	s.e.	
beerprice	-0.272	0.119	**	0.257	0.120	**	-0.631	0.280	**	-0.504	0.370	
female	-0.667	0.019	***	0.695	0.020	***	-1.290	0.046	***	-1.761	0.069	***
married	-0.103	0.030	***	0.121	0.022	***	-0.285	0.053	***	-0.384	0.069	***
lnage	-0.474	0.029	***	0.501	0.031	***	-1.105	0.075	***	-1.360	0.112	***
black	-0.294	0.025	***	0.283	0.024	***	0.003	0.063		-0.259	0.082	***
hispanic	-0.113	0.024	***	0.120	0.025	***	-0.352	0.058	***	-0.510	0.076	***
other	-0.305	0.044	***	0.305	0.043	***	-0.241	0.112	**	-0.490	0.140	***
lnincome	0.129	0.012	***	-0.125	0.012	***	-0.062	0.028	**	-0.029	0.038	
lnfamsize	-0.030	0.019		0.036	0.020	*	-0.146	0.047	***	-0.038	0.059	
nohs	-0.063	0.030	**	0.056	0.029	*	0.116	0.073		0.321	0.095	***
somecllg	0.032	0.024		-0.029	0.024		-0.249	0.057	***	-0.040	0.075	
college	0.118	0.026	***	-0.117	0.027	***	-0.440	0.065	***	-0.406	0.084	***
midwest	0.015	0.037		-0.018	0.037		0.168	0.089	*	0.203	0.116	*
south	-0.054	0.029	*	0.053	0.029	*	0.298	0.071	***	0.305	0.091	***
west	0.075	0.029	***	-0.074	0.030	**	0.183	0.072	**	0.195	0.093	**
bluecllr	0.283	0.035	***	-0.240	0.034	***	0.428	0.087	***	0.620	0.110	***
whitcllr	0.265	0.031	***	-0.227	0.028	***	0.070	0.080		0.270	0.094	***
servwrkr	0.253	0.035	***	-0.222	0.032	***	0.191	0.088	**	0.491	0.106	***
_cons	0.687	0.227	***	-1.243	0.231	***	3.887	0.547	***	4.628	0.736	***
Ancillary Parameters												
sigma							1.926	0.019	***	4.237	0.065	***
kappa							0.028	0.042	***	4.412	0.086	***
N	23,743			23,743			8,542			23,743		

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 5. Estimated Price Effects on Alcohol Consumption

	Conventional 2PM Probit EM and GG IM			2PM with NSD GG EM and IM			One-Part Model GG	
	beta	s.e.		beta	s.e.		beta	s.e.
AED(1 dollar)	-0.6838	0.1631	***	-0.6762	0.1672	***	-0.4655	0.3006
AIE(1 dollar)	-0.0950	0.0225	***	-0.0926	0.0230	***	-0.0586	0.0379

Note: *** p<0.01, ** p<0.05, * p<0.1

Chapter 6.

Summary, Discussion and Conclusions

The Fully Parametric Two-Part Model (FP2PM) is one of the most widely applied estimation frameworks in empirical economics research. It is applied to cases in which the outcome of interest is nonnegative with a large fraction of zeros. By the model design, the FP2PM allows for two systematically different data generating processes: the Extensive Margin (EM) governs zero or not and the Intensive Margin (IM) determines positive levels given non-zero. The conventional 2PM in the literature genuinely assumes a Probit EM and another distribution with positive support for the IM. This dissertation provides a general FP2PM potential outcomes (PO) framework, in which causally interpretable policy effect parameters can be identified and estimated in the 2PM context. Except for taking account of causal policy effect parameters, the framework presented in this dissertation also permits an alternative 2PM with non-structural difference between the EM and IM (NSD), in which the EM and the IM have the same underlying distribution. Several simulation studies compare the policy effect parameter estimations fit by the 2PM with NSD to those fit by the conventional 2PM, and simulation results show that the 2PM with NSD is a reliable model and not inferior to the conventional 2PM. Maintaining the same structural assumption, the 2PM with NSD allows an easy-to-implement likelihood ratio test on the null hypothesis that the two-part structure is not needed, which is a fundamental model specification question researchers should ask. An empirical study on alcohol demand is used to illustrate the aforementioned econometric issues. In virtually all 2PM instances, we suggest that the affordable likelihood ratio test on the null hypothesis that the two-part structure is not needed should be implemented before any further analysis.

Appendix I.

A Brief Note on "Cake Debate"

It should be noticed that in Cragg's model, ε^{EM} and ε^{IM} are assumed to have two separate distributions, as in all types of 2PMs, rather than a bivariate normal distribution with a significant correlation, as in Sample Selection Models. Both 2PMs and Sample Selection Model are well known to deal with limited dependent variables, e.g., the outcome variable of interest having a significant number of zeros. 2PMs have been favored due to more structural flexibilities and more robust statistical properties, however, Sample Selection Model and its two-step estimation procedure have also been employed in numerous applied microeconomic studies since Heckman (1976, 1979). Applied researchers are often confused about model selections and have almost never discussed any convincing reason why they choose one model over another. In the econometrics literature, there have been a series of intense debates regarding the model selection between 2PMs and Sample Selection Models since 1980s.

Duan et al. (1983, 1984, 1985) and Manning, Duan and Rogers (1987) are the early works showing that 2PMs dominate in the scenarios of limited dependent variables by providing theoretical and practical evidence, supported by simulation results. Duan and his associates have the following main arguments to advocate 2PMs. First, the structures of two parts are not restricted to any joint distribution in 2PMs, and it is hard to prove there to be a specific correlation between the two parts in most empirical works, and thus the model structures of 2PMs are more flexible. Second, the inverse Mill's ratio that is used to correct the selection bias is likely to be highly correlated with other regressors in the second step, which leads to the weaker statistical power of coefficient estimates in Sample Selection Models. Third, the comprehensive Monte Carlo simulation experiments show that despite the situation where Sample Selection Model is the true specification, 2PMs have better performance in prediction powers if there is not a strong exclusion restriction, i.e., the same set of regressors are used in participation and level decisions. In addition, there is never a certain reason to believe the true specification is a Sample Selection Model in practice, which makes using data-analytic 2PMs more convincing. Last but not least, as clarified by Duan et al. (1983) and Dow and Norton (2003), predictions on actual outcome are more interesting in many studies particularly on health related outcomes, for example, health care expenditure, where zero means actual outcome rather than missing values.

It is worth bearing in mind that there is another thread of literature defensing for Sample Selection Models. Hay and Olsen (1984) starts the “cake debate” and attacks the theoretical framework of 2PMs by arguing that 2PMs are nested in generalized Sample Selection Models. However, Hay, Leu and Rohrer (1987) use Monte Carlo simulation results based on an individual-level health care expenditure data to support the claim that estimators of 2PMs are more robust and thus 2PMs outperform the Sample Selection Model. By redesigning Monte Carlo simulation experiment based on (Manning, Duan and Rogers, 1987), Leung and Yu (1996) show that Sample Selection Models perform better than 2PMs only on the conditions that Sample Selection Models are true models and there is no collinearity problem. But they conclude that the results do not support the superiority of either model to another since each model has its advantage under different settings.

Appendix II.

Three Lemmas

Lemma 1: If z is log-normal with pdf f(z) and parameters a and b then

$$\int_d^\infty f(z)\, dz = \Phi\left(\frac{a - \ln(d)}{\sqrt{b}}\right)$$

and

$$\int_0^d f(z)\, dz = 1 - \Phi\left(\frac{a - \ln(d)}{\sqrt{b}}\right)$$

where, a, b and d are scalar constants.

Proof:

Because z is log-normal we know $\frac{\ln(z) - a}{\sqrt{b}}$ is standard normal. So

$$\int_d^\infty f(z)\, dz = \Pr(z > d) = \Pr\left(\frac{\ln(z) - a}{\sqrt{b}} > \frac{\ln(d) - a}{\sqrt{b}}\right)$$

$$= 1 - \Phi\left(\frac{\ln(d) - a}{\sqrt{b}}\right) = \Phi\left(\frac{a - \ln(d)}{\sqrt{b}}\right)$$

It can similarly be shown that

$$\int_0^d f(z)\, dz = 1 - \Phi\left(\frac{a - \ln(d)}{\sqrt{b}}\right)$$

∎

Lemma 2: If z is normal with mean a and variance b then

$$\int_{d}^{\infty} \exp(cz)\, f(z)\, dz = \exp\left(ca + \frac{1}{2}c^2 b\right)\left[1 - \Phi \;\; d - c\sqrt{b} \;\; \right]$$

and

$$\int_{-\infty}^{d} \exp(cz)\, f(z)\, dz = \exp\left(ca + \frac{1}{2}c^2 b\right)\Phi \;\; d - c\sqrt{b}$$

where, c and d are scalar constants.

Proof:

$$\int_{d}^{\infty} \exp(cz)\, f(z)\, dz = \int_{d}^{\infty} \exp(cz)\, \frac{1}{\sqrt{2\pi b}} \exp\left(-\frac{1}{2b}(z - a)^2\right) dz\,.$$

Now make the substitution $q = \dfrac{z - a}{\sqrt{b}}$ and get

$$\int_{d}^{\infty} \exp(cz)\, f(z)\, dz = \frac{1}{\sqrt{2\pi}} \int_{d}^{\infty} \exp \;\; c(\sqrt{b}q + a) \;\; \exp\left(-\frac{1}{2}q^2\right) dq\,.$$

$$= \exp(ca)\, \frac{1}{\sqrt{2\pi}} \int_{d}^{\infty} \exp \;\;\; c\sqrt{b}q \;\; \exp\left(-\frac{1}{2}q^2\right) dq$$

$$= \exp(ca) \frac{1}{\sqrt{2\pi}} \int_{d}^{\infty} \exp\left(c\sqrt{b}q - \frac{1}{2}q^2\right) dq$$

$$= \exp(ca) \frac{1}{\sqrt{2\pi}} \int_{d}^{\infty} \exp\left(-\frac{1}{2} \;\; q - c\sqrt{b} \;\;^2 + \frac{1}{2}c^2 b\right) dq$$

$$= \exp\left(ca + \frac{1}{2}c^2 b\right) \frac{1}{\sqrt{2\pi}} \int_{d}^{\infty} \exp\left(-\frac{1}{2} \;\; q - c\sqrt{b} \;\;^2 \;\; \right) dq\,.$$

But $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(q - c\sqrt{b}\right)^2\right)$ is the density of a normal random variable with mean $c\sqrt{b}$ and variance 1, so

$$\frac{1}{\sqrt{2\pi}} \int_d^{\infty} \exp\left(-\frac{1}{2}\left(q - c\sqrt{b}\right)^2\right) dq = 1 - \Phi\left(d - c\sqrt{b}\right)$$

and

$$\int_d^{\infty} \exp(cz)\, f(z)\, dz = \exp\left(ca + \frac{1}{2}c^2 b\right)\left[1 - \Phi\left(d - c\sqrt{b}\right)\right].$$

It can similarly be shown that

$$\int_{-\infty}^{d} \exp(cz)\, f(z)\, dz = \exp\left(ca + \frac{1}{2}c^2 b\right)\Phi\left(d - c\sqrt{b}\right).$$

■

Lemma 3: If z is log-normal with pdf f(z) and parameters a and b then

$$\int_d^{\infty} z\, f(z)\, dz = \exp\left(a + \frac{b}{2}\right)\left[1 - \Phi\left(\frac{\ln(d)}{\sqrt{b}} - \sqrt{b}\right)\right]$$

Proof:

We have

$$\int_d^{\infty} z\, f(z)\, dz = \int_d^{\infty} \frac{1}{\sqrt{2\pi b}} \exp\left[\frac{1}{2b}(\ln(z) - a)^2\right] dz$$

Using the change-of-variable

$$y = \frac{\ln(z) - a}{\sqrt{b}}$$

we get

$$z = \exp(\sqrt{b}\, y + a)$$

with Jacobian

$$\frac{dz}{dy} = \exp(\sqrt{b}\, y + a)\sqrt{b}$$

and

$$\int_{d}^{\infty} z\ f(z)\ dz = \int_{\frac{\ln(d)-a}{\sqrt{b}}}^{\infty} \frac{\exp(\sqrt{b}\, y + a)}{\sqrt{2\pi}} \exp\left[\frac{1}{2} y^2\right] dy$$

$$= \exp(a) \int_{\frac{\ln(d)-a}{\sqrt{b}}}^{\infty} \frac{\exp(\sqrt{b}\, y)}{\sqrt{2\pi}} \exp\left[\frac{1}{2} y^2\right] dy.$$

Now using Lemma 2 we get

$$\int_{d}^{\infty} z\ f(z)\ dz = \exp(a) \exp\left(\frac{b}{2}\right)\left[1 - \Phi\left(\frac{\ln(d)}{\sqrt{b}} - \sqrt{b}\right)\right]$$

$$= \exp\left(a + \frac{b}{2}\right)\left[1 - \Phi\left(\frac{\ln(d)}{\sqrt{b}} - \sqrt{b}\right)\right].$$

References

Arulampalam, Wiji and Alison L. Booth (1997): "Who Gets over the Training Hurdle? A Study of the Training Experiences of Young Men and Women in Britain," *Journal of Population Economics*, 10(2), 197-217.

Bradford, W.D., A.N. Kleit, M.A. Krousel-Wood and R.N. Re (2002): "A Two-Part Model of Treatment for Benign Prostatic Hyperplasia and The Impact of Innovation," *Applied Economics*, 34, 1291-1299.

Brown, Sarah, Pulak Ghosh, Li Su and Karl Taylor (2015): "Modelling Household Finances: a Bayesian Approach to a Multivariate Two-Part Model," *Journal of Empirical Finance*, 33, 190-207.

Buntin, Melinda Beeuwkes and Alan M. Zaslavsky (2004): "Too Much Ado about Two-Part Models and Transformation? Comparing Methods of Modeling Medicare Expenditures," *Journal of Health Economics*, 23: 525-542.

Burney, Nadeem A., Mohammad Alenezi, Nadia Al-Musallam and Ahmed Al-Khayat (2016): "The Demand for Medical Care Services: Evidence from Kuwait Based on Households' Out-of-Pocket Expenses," *Applied Economics*, 48(28), 2636-2650.

Chaloupka, Frank J., Michael Grossman and Henry Saffer (2002): "The Effects of Price on Alcohol Consumption and Alcohol-Related Problems," *Alcohol Research and Health*, 26(1), 22-34.

Chang, Hung-Hao and Chad D. Meyerhoefer (2016): "The Causal Effect of Education on Farm-Related Disability: Evidence from a Compulsory Schooling Reform in Taiwan," *American Journal of Agricultural Economics*, 98(5), 1545-1557.

Coate, Douglas, and Michael Grossman (1988): "Effects of Alcoholic Beverage Prices and Legal Drinking Ages on Youth Alcohol Use," *Journal of Law and Economics*, 31(1), 145-171.

Cragg, John G. (1971): "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," *Econometrica*, 39(5), 824-844.

Duan, Naihua, Willard G. Manning, Carl N. Morris and Joseph P. Newhouse (1983): "A Comparison of Alternative Models for the Demand for Medical Care," *Journal of Business & Economic Statistics*, 1(2), 115-126.

Fang, Kuangnan, Xiaoyan Wang, Ben-Chang Shia and Shuangge Ma (2017): "Identification of Proportionality Structure with Two-Part Models Using Penalization," *Computational Statistics and Data Analysis*, 99, 12-24.

Gurmu, Shiferaw and Pravin K. Trivedi (1996): "Excess Zeros in Count Models for Recreational Trips," *Journal of Business & Economic Statistics*, 14(4), 469-477.

Han, Cong and Richard Kronmal (2006): "Two-Part Models for Analysis of Agatston Scores with Possible Proportionality Constraints," *Communications in Statistics–Theory and Methods*, 35, 99-111.

Hertz, Tom (2010): "Heteroskedasticity-Robust Elasticities in Logarithmic and Two-Part Models," *Applied Economics Letters*, 17, 225-228.

Heckman, James J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5(4), 120-137.

Hyun, Kyung-Rae, Sungwook Kang and Sunmi Lee (2016): "Population Aging and Healthcare Expenditure in Korea," *Health Economics*, 25, 1239-1251.

J.S. Ramalho, Joaquim and Jacinto Vidigal Da Silva (2009): "A Two-Part Fractional Regression Model for the Financial Leverage Decisions of Micro, Small, Medium and Large Firms," *Quantitative Finance*, 9(5), 621-636.

Kim, Youngkoung and Bengt O. Muthén (2009): "Two-Part Factor Mixture Modeling: Application to an Aggressive Behavior Measurement Instrument," *Structural Equation Modeling*, 16(4), 602-624.

Leung, Siu Fai and Shihti Yu (1996): "On the Choice between Sample Selection and Two-Part Models," *Journal of Econometrics*, 72(1-2), 197-229.

Li, Bernadette, John Cairns, James Fotheringham and Rommel Ravanan (2016): "Predicting Hospital Costs for Patients Receiving Renal Replacement Therapy to Inform an Economic Evaluation," *European Journal of Health Economics*, 17, 659-668.

Liu, Lei, Robert L. Strawderman, Mark E. Cowen and Ya-Chen T. Shih (2010): "A Flexible Two-Part Random Effects Model for Correlated Medical Costs," *Journal of Health Economics*, 29, 110-123.

Madden, David (2008): "Sample Selection Versus Two-Part Models Revisited: The Case of Female Smoking and Drinking," *Journal of Health Economics*, 27, 300-307.

Manning, Willard G., Anirban Basu and John Mullahy (2005): "Generalized Modeling Approaches to Risk Adjustment of Skewed Outcomes Data," *Journal of Health Economics*, 24, 465-488.

Manning, Willard G., Linda Blumberg and Lawrence H. Moulton (1995): "The Demand for Alcohol: The Differential Response to Price," *Journal of Health Economics*, 14, 123-148.

Mullahy, John (1986): "Specification and Testing of Some Modified Count Data Models," *Journal of Econometrics*, 33, 341-365.

Ross, Hana and Frank J. Chaloupka (2003): "The Effect of Cigarette Prices on Youth Smoking," *Health Economics*, 12, 217-230.

Ruhm, Christopher J., Alison Snow Jones, Kerry Anne McGeary, William C. Kerr, Joseph V. Terza, Thomas K. Greenfield and Ravi S Pandian (2012): "What U.S. Data Should Be Used to Measure the Price Elasticity of Demand for Alcohol?" *Journal of Health Economics*, 31, 851-862.

Santos Silva, Joao M.C. and Frank Windmeijer (2001): "Two-Part Multiple Spell Models for Health Care Demand," *Journal of Econometrics*, 10, 67-89.

Seller, Kimberly F. and Galit Shmueli (2010): "A Flexible Regression Model for Count Data," *The Annals of Applied Statistics*, 4(2), 943-961.

Taylor, Sandra and Katherine Pollard (2009): "Hypothesis Tests for Point-Mass Mixture Data with Application to Omics Data with Many Zero Values," *Statistical Applications in Genetics and Molecular Biology*, 8(1), Article 8.

Terza, J. (1985): "Reduced-Form Trinomial Probit: A Quantal Response Model without A Priori Restrictions," *The Journal of Business and Economic Statistics*, 3, 54-59.

Terza, Joseph V. (2017): "Econometric Methods for Policy Analysis in Health Economics and Health Services Research, Chapter 4," manuscript in progress.

Terza, Joseph V. (2018): "Regression-Based Causal Analysis from the Potential Outcomes Perspective," Working Paper.

Terza, Joseph V., Alison Snow Jones, Srikant Devaraj, Thomas K. Greenfield, William C. Kerr, Kerry Anne McGeary and Christopher J. Ruhm (2016): "Specification and Estimation of Own-Price Elasticity in Two-Part Models of Alcohol Demand," Working paper.

Yang, Yan (2017): The Effect of Sugar-Sweetened Beverage Consumption on Childhood Obesity - Causal Evidence (Doctoral dissertation). Retrieved from http://hdl.handle.net/1805/11647.

Curriculum Vitae

Zhuang Hao

Education

Ph.D., Economics, Indiana University at IUPUI, Indianapolis, IN, 2018
M.S., International Economics, Suffolk University, Boston, MA, 2013
B.A., Economics, Hohai University, Nanjing, Jiangsu, China, 2011

Fields of Interest

Applied Econometrics; Health Economics

Work Experience

Analyst, Congressional Budget Office, Washington, D.C., 2018 - Present
Instructor, IUPUI, Indianapolis, IN, 2015 - 2018
Summer Associate, Discover Financial Services, Phoenix, AZ, 2016

Teaching Experience (with full teaching responsibilities)

Undergraduate Level:
- International Economics (2018 Spring, 2017 Fall/Spring, 2016 Fall)
- Introduction to Economic and Business Statistics (2015 Fall)

Ph.D. Level:
- Math Camp (2015 Summer)

Grants and Scholarships

Graduate-Professional Educational Grants, IUPUI, 2018/2016
Loretta Lunsford Scholarship, IUPUI, 2017
Robert B. Harris Graduate Teaching Recognition Scholarship, IUPUI, 2017
Innovation-to-Enterprise Central Scholarship, IUPUI, 2015