COUNT-REGRESSION-BASED EMPIRICAL CAUSAL ANALYSIS FROM A POTENTIAL OUTCOMES PERSPECTIVE: ACCOUNTING FOR BOUNDEDNESS, DISCRETENESS, DISPERSION AND UNOBSERVABLE CONFOUNDING

Golnoush Kazeminezhad

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

June 2024

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, PhD, Chair

Christopher A. Harle, PhD

April 23, 2024

Wendy Morrison, PhD

Steven Russell, PhD

DEDICATION

For my loving husband, Maziar and my parents Morasa and Rahmat.

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude and acknowledge the following individuals who have played a significant role in the completion of my dissertation: First and foremost, I am deeply indebted to my advisor and the chair of my dissertation committee, Dr. Joseph Terza. His unwavering and invaluable support, guidance, patience, and mentorship have nurtured my personal and academic growth. None of my accomplishments on this journey would have been possible without him. I am forever grateful for his caring nature, profound intelligence, and the privilege of working with him.

I would also like to extend my appreciation to my other research committee members. To Dr. Wendy Morrison, the chair of the department of Economics at IUPUI whose compassion and affection have been instrumental in facilitating my academic endeavors. I am also grateful to Dr. Steven Russell and Dr. Chris Harle. Their support, insightful feedback and constructive criticism have greatly contributed to the development and refinement of my work.

A special mention goes to my dear friend, professor Mohammad Kaviani. I am humbled by his enormous sacrifice and can never thank him enough for his assistance and support. I would also like to express my gratitude to my professors in the economics department at IUPUI, Dr. Mark Wilhelm, Dr. Vidura Tennekoon, Dr. Subir Chakrabarti and Dr. Sumedha Gupta, for their contributions and valuable insights. Their expertise and guidance have been indispensable in shaping my research.

I am grateful to Terri Crews for her attentive assistance throughout my time in the department. Her kind smile and unwavering support have been sources of comfort and encouragement. To senior Ph.D. students in the Department of Economics, Daniel Asfaw, Haile Shone, Satabdi Adhikary, Taul Cheong, and my very best friend and cohort member Apurva Bendre, I extend my appreciation. Your camaraderie, collaboration, and shared experiences have played a vital role in my achievements today.

To my beloved husband, Maziar Aboualizadehbehbahani, I cannot thank you enough for your endless support, both academically and emotionally. Your love and encouragement have been my driving force, and I am forever grateful to have you by my side. I would also like to express my deepest gratitude to my parents, Morasa Jahangirrad and Rahmat Kazeminejad. Your belief in me and your sacrifices have made me who I am today. To my sister, Noushin, and my brother, Pasha, my brother-in-law, Ali, I am eternally thankful for your constant love and support. To my late father-in-law Rasoul Aboualizadehbehbahani, whose encouragement kept me going on this journey and in memory of my late mother-in-law Pari and my grandmother Forough. To my friends Niloufar, Zeinab, Maryam, Forough, Mehdi, Amirreza, Morgan, Rachel, Dr. Samad Hedayat, Dr. Ali Jafari, Maymanat Montaser, and Todd Sears, who listened to me and motivated me when I needed it the most. And to my precious nephew, Dara, my ray of sunshine.

And lastly, I would like to extend my sincere appreciation to Indiana University for providing me with this incredible opportunity. I am thankful for the welcoming community and the safe place I have found to call home during my academic pursuits. To all those mentioned and the countless others who have contributed to my growth and success, I offer my heartfelt thanks. Your support, encouragement, and belief in me have been instrumental, and I am truly grateful for your presence in my life.

Golnoush Kazeminezhad

COUNT-REGRESSION-BASED EMPIRICAL CAUSAL ANALYSIS FROM A POTENTIAL OUTCOMES PERSPECTIVE: ACCOUNTING FOR BOUNDEDNESS, DISCRETENESS, DISPERSION AND UNOBSERVABLE CONFOUNDING

Empirical economic research is primarily driven by the desire to offer scientific evidence that serves to inform the study of cause-and-effect. In this dissertation, I developed new models for count-regression-model-based (CRM-based) causal effect estimation in which the value for the outcome of interest is restricted to the non-negative integers. I implement first-order two-stage residual inclusion (FO-2SRI) methods, in the context of the general potential outcomes framework, that accommodate nonlinearities due to the intrinsic characteristics of count-valued outcomes such as boundedness (outcome nonnegative), discreteness (outcome has countable support) and dispersion (conditional variance and other higher order conditional moments of the outcome not necessarily equal to its conditional mean) of count data, and unobservable confounding. The focus here is on the case in which the causal variable is continuous.

The newly proposed causal effect estimators are compared with extant FO-2SRI estimators based on conventional control function methods and the linear instrumental variables (LIV) estimator. A series of simulation studies are performed to investigate the accuracy of the proposed estimators and compare the results with the extant estimators. In the simulation studies, the robustness of the fully nonlinear CRM-based FO-2SRI methods are investigated with attention to an important type of misspecification error. The models are also applied to a real-world data from Nigeria to investigate the effect of female education on their fertility decisions in a developing country.

The results of the simulation studies reveal that estimates obtained via the newly proposed estimators are very accurate and widely diverge from the results from the extant control function and LIV methods. Moreover, one of the new estimators, which allows dispersion flexibility, dominated all other estimators (aside from a few extreme dispersion cases) with regard to avoidance of misspecification bias. Finally, the results showed that same estimator to be quite accurate for a wide range of values of the dispersion parameter (which measures mean/variance divergence). Similar results were obtained via the real data analysis which indicates that increasing women's education decreases childbearing.

Joseph V. Terza, PhD, Chair

Christopher A. Harle, PhD

Wendy Morrison, PhD

Steven Russell, PhD

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

AIE	Average Incremental Effect
ASE	Asymptotic Standard Error
AVAR	Asymptotic Variance-Covariance Matrix
CE	Causal Effect
CF	Control Function
CI	Conditional Independence
CMP	Conway-Maxwell-Poisson
CPOM	Conditional Potential Outcomes Model
COI	Conditional Outcome Invariance
CRM	Count Regression Model
DGP	Data Generating Process
ECA	Empirical Causal Analysis
EDR	Equi-Dispersion Restriction
FIML-CMP	Full Information Maximum Likelihood Estimator with Exogenous Causal variable and Conway-Maxell-Poisson Specification for the outcome variable
FIML-Poisson	Full Information Maximum Likelihood Estimator with Exogenous Causal variable and Poisson Specification for the outcome variable
FO-2SRI	First-Order Two-Stage Residual Inclusion
GEC	Generalized Event Count
GEC_k	Generalized Event Count with an additional parameter (k)
GG	Generalized Gamma

GG-CMP	First-Order Two-Stage Residual Inclusion Estimator with Generalized Gamma specification in the First Stage and Conway-Maxwell-Poisson Specification in the Second Stage
GG-Poisson	First-Order Two-Stage Residual Inclusion Estimator with Generalized Gamma specification in the First Stage and Poisson Specification in the Second Stage
GMICS	Global Multiple Indicator Cluster Survey Program
GMM	Generalized Method of Moments
GP	Generalized Poisson
GPO	General Potential Outcomes Framework
HP	Hyper-Poisson
LIE	Law of Iterated Expectation
LIN-CMP	Two-Stage Control Function Estimator with Linear specification in the First Stage and Conway-Maxwell-Poisson Specification in the Second Stage
LIN-Poisson	Two-Stage Control Function Estimator with Linear specification in the First Stage and Poisson Specification in the Second Stage
LIV	Linear Instrumental Variable
MLE	Maximum Likelihood Estimator
MICSN	Multiple Indicator Cluster Survey of Nigeria
NB	Negative Binomial
OC	Unobservable Confounding
OLS	Ordinary Least Squares

PMF	Probability Mass Function
RGP	Restricted Generalized Poisson
SEP	Series-Expansion Poisson
UNICEF	United Nations International Children's Emergency Fund
2SLS	Two-Stage Least Squares
2SME	Two-Stage M-estimator
2SRI	Two-Stage Residual Inclusion

Chapter 1

Introduction, Background, Significance, and Summary

1.1 Overview

Empirical economic research is primarily driven by the desire to offer scientific evidence that helps assess policy relevant cause-and-effect. The approach most often applied in pursuit of this objective involves regression modeling and estimation. In this dissertation, we propose a novel approach to the use of *count regression models* (CRMs) for the specification, identification, and estimation of policy relevant causal effects (CEs) and related inference. We focus on the specification, estimation, and causal inference of a CE in the context of the CRM in which the value for the outcome of interest or the dependent variable is restricted to the non-negative integers and the presumed causal variable is continuous. We seek to develop and implement a regression model specification that accommodates the nonlinearity due to intrinsic characteristics of count-valued outcomes i.e., boundedness (being bounded from below at zero), discreteness (measured by counting) and dispersion (skewed and sparse) in the perspective of a causal framework as well as the nonlinearity that could be additionally imposed by the endogeneity of the presumed causal variable. We develop new models for CRM-based CE estimation that implement *two-stage residual inclusion* (2SRI) methods, as suggested by Terza et al. (2008). We assess the accuracy of our proposed new methods and compare them with extant 2SRI approaches.

1.2 Literature in Methodology and Application

Various CRM specifications have been developed to accommodate the inherent nonlinearity due to boundedness (nonnegative outcome), discreteness (outcome has countable support) and dispersion (conditional variance and other higher order conditional moments of the outcome not necessarily equal to its conditional mean) of count data. These models in which the outcome of interest is non-negative integer-valued, abound in the applied economics literature. First discussed by Jorgenson (1961), the most widely implemented CRM specification, is based on the Poisson probability mass function (pmf). Other early methodological studies of the Poisson CRM include Mahamunulu (1967), Weber (1971), Nelder & Wedderburn (1972), El-Sayyad (1973), Frome (1973), and Kumar and Shih (1978). Applications of the Poisson CRM are too numerous to list. Early examples in economics include Grogger (1990), Borsch-Supan (1990), Michener and Tighe (1992), Schwartz & Torous (1993), Khlat & Courbage (1997), and Yu et al. (2006). More recent studies that implement this model are Austin & Totaro (2011), Polasik et al. (2020), and García-Gómez & Parrado (2023). Because the Poisson CRM is saddled by the equi-dispersion restriction (conditional mean = conditional variance) [EDR], less dispersion-restrictive CRMs have been developed. The negative binomial (NB) regression model, introduced by Gourieroux et al. (1981, 1984), loosens the EDR somewhat by accommodating over-dispersion (conditional mean less than conditional variance). Hausman et al. (1984) detail the connection between the Poisson and NB specifications, highlighting how the presence of an additional parameter in the formulation of the latter allows for over-dispersion – a condition commonly encountered in economic data. As is the case for the Poisson, applications of the NB CRM are many. Early applications of this approach in economics can be found in Cameron et al. (1988), Ransom & Pope (1995) McCarthy (2003), and Weiss & Wittkopp (2005). The popularity of the NB CRM in empirical economics continues to the present [see Cohen et al. (2014), Kalist & Lee (2016), Schuettig & Sundmacher (2022), Dwomoh et al. (2023), and Arena et al. (2024)]. The attractiveness of these methods can be attributed to their relative simplicity and good behavior. [1]

Although many applied CRM contexts in economics are characterized by equi- or over-dispersion (applied contexts in economics characterized by overdispersion include health services utilization, health related behaviors and problems, unemployment, financial data, and international relations as described by King [1989a].), there are other important empirical questions whose answers require the analysis of under-dispersed count-valued data (applied contexts in economics characterized by under-dispersion include population studies, demography, and fertility es described by Winkelmann and Zimmermann [1994]). To resolve this issue, CRM specifications based on dispersion-flexible pmfs have been developed. Such pmf formulations include: the Conway-Maxwell-Poisson (CMP) [Conway & Maxwell, 1962]; the hyper-Poisson (HP) [Bardwell & Crow, 1964]; the generalized Poisson (GP) [Consul and Jain, 1973]; the restricted generalized Poisson (RGP) [Consul, 1989]; the generalized event count (GEC) [King, 1989a]; the GEC with an additional parameter (k) that allows for nonlinear mean-variance relationships (GEC_k) [Winkelmann and Zimmerman, 1991]; and series-expansion (or semi-non-parametric) Poisson (SEP) [Cameron & Johansson, 1997]. Corresponding CRMs have been developed

[1] For data with excessive number of zeros, zero-inflated and hurdle versions of these models are also used.

by: Sellers and Shmueli (2010) for the CMP[2]; Sáez-Castillo & Conde-Sánchez (2013) for the HP; Consul and Famoye (1992) for the GP; Famoye (1993) for the RGP; King (1989b) for the GEC; Winkelmann and Zimmerman, (1991) for the GEC_k; and Cameron & Johansson (1997) for the SEP. Applications of these CRMs in economics can be found in: Chuang & Oliva (2014) and Fraser (2020) for the CMP; English et al. (2002) for the HP; Colón-López & García (2022) and Issahaku et al. (2023) for the GP; Wang & Famoye (1997) and Agyeman (2021) for the RGP; Clarke et al. (1999), Chi (1998), Krain (1998) and Yan et al. (2019) for the GEC; Barbosa et al. (2004) and Bolancé et al. (2008) for the GEC_k; and Mainardi (2003) and Creel (2011) for the SEP.

The need to accommodate unobservable confounding in the implementation of the CRM for empirical causal analysis typically necessitates the introduction of an additional source of nonlinearity.[3, 4] For example, to investigate the effect of an education policy addressing population growth in a developing country, let $\mathbf{X}$ be the mother's educational attainment and $\mathbf{Y}$ represent her number of viable children. The causal relationship between the $\mathbf{X}$ and the $\mathbf{Y}$ is likely to be subject to unobservable confounding since characteristics such as ability, motivation and sociocultural factors that affect one's access to education could partially or entirely affect their fertility decisions.[5] Moreover, modeling approaches

---

[2] For more about the CMP distribution, see Shmueli et al. (2005); Huang (2017); Forthmann et al. (2020); and Sellers (2023).

[3] We use the term *Unobservable confounding* in reference to an empirical circumstance that others may refer to as involving *endogeneity*. We opt for this phrasing because it better comports with causal framing of confounder control. This will be made clear later in the discussion.

[4] Here we rely on the dictionary definition of the term *confounding*. Later in the discussion we offer formal definitions of this and related terms (e.g., *confounder*).

[5] To facilitate the discussion throughout the remainder of the dissertation, we will refer to this example (henceforth, the *education/fertility [E/F] example*).

that account for such unobservable confounding via a nonlinear regression specification of the **X** (e.g., education as a nonlinear, lower-bounded at zero, function of instrumental variables) introduce additional nonlinearity in the CRM context.

Empirical applications of CRMs that account for unobservable confounding and concomitant nonlinearity can be found in the health economics and health services research literatures. In many of these examples, the **X** is binary (Higuera and Prada, 2016; Lippi Bruni et al, 2016; Costa-Font and Vilaplana-Prieto, 2020; Landersø and Fallesen, 2021; Oyenubi and Kollamparambil, 2022; Serrano-Alarcón et al., 2022; Soltani et al, 2022). The methods used by these authors include those developed by Terza (1998); Romeu and Vera-Hernández (2005), and Bratti and Miranda, 2011. In other cases, the **X** is continuous or polychotomous (Coulson et al, 1995, Dusheiko and Gravelle, 2018; Costa-Font et al. 2018; Soltani et al, 2022). Methods covering such cases include: the generalized method of moments [GMM] (Windmeijer and Santos Silva, 1997; Mullahy, 1997; Blundell et al., 2002)[6,7]; quasi-limited information maximum likelihood methods (Wooldridge, 2014); control function methods (Wooldridge, 2015; Blundel and Powell, 2003); and the 2SRI method (Terza et al. 2008).[8]

[6] The empirical examples of the GMM in the health economics literature include Vera-Hernández (1999), Schellhorn (2001), Sulemana et al. (2018), Olayiwola and Kazeem (2019), Costa-Font and Vilaplana-Prieto (2020), Oyenubi and Kollamparambil (2022).
[7] Terza (2006) discusses the non-symmetry of the observable and unobservable regressors in typical applications of the GMM and notes the infeasibility of the GMM for estimating the CE under non-exponential specifications. Examples could be dispersion flexible models such as CMP that have extra terms in their conditional mean other than the exponential form needed to estimate the CE.
[8] Bayesian inference approaches have also been proposed for count data models in Winkelmann (2008); Xue-Dong (2009); Malyshkina et al. (2009); Neelon et al. (2010); Barua et al. (2014); Dimitrakopoulos (2019); Bansal et al. (2021).

1.3 Objectives

In this dissertation, we develop new models for CRM-based causal effect (CE) estimation that implement *two-stage residual inclusion* (2SRI) methods, as suggested by Terza et al. (2008). We assess the accuracy of our proposed new methods and compare them with extant 2SRI approaches. In the first level (stage) of a 2SRI protocol the causal variable is modeled, and the relevant related residual, aimed at capturing unobservable confounding influences, is specified. The outcome of interest is modeled in the 2SRI second level (stage) with the first level residual included to control for unobservable confounding. We focus on the class of 2SRI estimators in which the relevant residual is formulated in terms of the first-order conditional moment of the causal variable. We refer to this class of estimators as *first-order two-stage residual inclusion* (FO-2SRI) methods.

The CE estimators we consider differ with respect to how their corresponding underlying FO-2SRI specifications account for nonlinearity. The CRM features that are likely to induce nonlinearity at the FO-2SRI second level include discreteness of the outcome, boundedness from below at zero, and dispersion flexibility (i.e., loosening of the mean = variance restriction that plagues the most frequently applied CRM specification). Similar nonlinearity inducing aspects are typically present in the FO-2SRI first level. Such nonlinearities are usually related to inherent restrictions on the support of the causal variable.

A popular member of the FO-2SRI class of estimators is the conventional linear instrumental variables (LIV) estimator which offers no accommodation for possible inherent nonlinearity at either of the FO-2SRI levels. Other FO-2SRI estimators accommodate nonlinearity at the second level but impose linearity in the first. These are the so-called *control function* (CF) methods. The first objective of this dissertation is to compare CRM-based FO-2SRI estimators, that fully account for nonlinearity at both levels, to those that ignore nonlinearity (fully Linear Instrumental Variable [LIV] and partially linear CF methods). In particular, we seek to answer the question…Do fully nonlinear CE estimators produce results that substantively differ from results obtained using conventional linear methods (in particular, LIV)? We confront this query via simulated data and a real data example. Secondly, using simulated and real data we investigate whether allowing dispersion flexibility improves the accuracy of CE estimation in the CRM context. Finally, using the same simulated data, we investigate how robust the fully nonlinear CRM-based FO-2SRI methods are to an important type of misspecification error.

1.4 Main Findings and Significance

In answer to the query characterizing the first study objective, we found that causal effect (CE) estimates obtained from the fully nonlinear specifications are generally accurate and differ substantially from the partially (CF) and fully (LIV) linearized approaches – by as much as 107%. Moreover, in the context of a real data example, we found massive differences in the CE results from the nonlinear specifications vs. the linearized approaches (in the worst case, –.905 vs. –.035, respectively). Inferential statistics were similarly divergent (in the worst case, asymptotic t-stat = –18.91 vs. asymptotic t-stat = –1.45, respectively). To achieve the second study goal, in the context of our simulation study, we varied the level of dispersion flexibility in our data generator and found that the CE estimates that took account of all three types of dispersion generally dominated those obtained from the models that ignored the dispersion issue (dominance would have been uniform except for a few of the simulated cases featuring low over-dispersion levels). Finally, in support of our third study goal, we exploited the case in which our data generator was designed to produce samples that did not exactly coincide with draws from the would-be populations underlying any of the five FO-2SRI CRM-based CE estimation protocols. Therefore, in this case, in our simulated sampling regime, all five of our estimators are subject, by design, to misspecification error. Nevertheless, we found the FO-2SRI CRM-based CE estimator designed to account for dispersion flexibility and inherent nonlinearity at both levels, to be quite accurate over a wide range of values of the dispersion parameter.

1.5 Organization of the Dissertation

The remainder of the dissertation is organized as follows. In chapter 2, we detail a general potential outcomes causal framework for specification and estimation of the generic First-Order Two Stage Residual Inclusion (FO-2SRI) CRM-based causal effect (CE) of interest. Therein, we show how CE estimation relies on estimation of the parameters of the underlying CRM that follows from the relevant assumptions regarding dispersion and the parametric representation of unobservable confounding. In chapter 3, we detail three specific estimators with their asymptotic statistics for the case in which the intrinsic characteristics of count data such as boundedness, discreteness and dispersion are of primary interest and unobservable confounding is not present. In chapter 4, the five specific FO-2SRI CRM-based CE estimators highlighted in our study are detailed along with their attendant asymptotic inferential statistics for the case in which there is unobservable confounding. In chapter 5, we assess and compare the alternative estimators with respect to bias using simulated data. In chapter 6, the five estimators are applied in an analysis of the causal effect of education on fertility using real data. The final chapter summarizes and concludes.

Chapter 2

General Potential Outcomes Framework (GPOF) in a Count Data Setting

2.1 Overview

Almost all empirical research studies in health economics aim to estimate the "effect" of a presumably causal entity of interest (the **X**) on an outcome of interest (the **Y**). The term "effect" (and any synonym thereof commonly used in this context) connotes a difference in the **Y** that can be exclusively and causally attributed to some sort of difference in the **X**. To be truly representative of causation, the definition of the term "effect" should be cast in the counterfactual (what if?) realm and should answer a query of the following form… "If the **X** were, by mandate, hypothetically altered in a specified way, then what would be the consequent impact on the **Y**?" For the case in which the **Y** is count-valued, the following *general potential outcomes framework* (GPOF) offers a context in which such causal effects can be clearly and rigorously specified (defined), identified and estimated.[9]

2.2 The General Potential Outcomes Framework (GPOF): Introduction, Basic Concepts and Definitions

In this GPOF, a distinction is drawn between two versions of the **X**:

$X^*(\omega) \equiv$ a deterministic function (known by the policy maker and/or researcher) that, under a specified counterfactual scenario, maps individuals (ω) in the relevant

[9] The framework we offer here is a special case of that discussed by Terza (2020, 2024a-b-c).

population (Ω) to known real values [$X^*(\omega)$ is the generic representation of a counterfactually mandated version of the **X**].

X ≡ the random variable representing the observable (*factual*) version of the distribution of the **X** [X is the element of the data generating process (DGP) from which the sampled values of the **X** can be drawn].[10]

Correspondingly, there are two versions of the **Y**:

$Y_{X^*(\omega)}$ ≡ the random variable representing the count-valued *potential outcome*, representing the counterfactual distribution of values of the **Y** for individual ω at a particular counterfactually imposed value $X^*(\omega)$.[11]

and

[10] The DGP is defined as the joint distribution from which the observable data can be sampled.

[11] In its most fundamental form, $Y_{X^*(\omega)}$ is specific to individual ω and specific to the counterfactual value of **X** mandated for that individual. Technically speaking, it is the measurable function $Y_{X^*(\omega)}(\tau_{X^*(\omega)})\colon \mathcal{T}_{X^*(\omega)} \longrightarrow \mathcal{S} \subset \mathbb{R}$ that maps various states of the world specific to individual ω at value $X^*(\omega)$ ($\tau_{X^*(\omega)} \in \mathcal{T}_{X^*(\omega)}$), to count values in the support of $Y_{X^*(\omega)}$ ($\mathcal{S} = \{0, 1, 2, \ldots\}$). The primitive probability space for $Y_{X^*(\omega)}$ then is ($\mathcal{T}_{X^*(\omega)}$, $\mathcal{F}_{X^*(\omega)}$, $\mathcal{P}_{X^*(\omega)}$), where $\mathcal{F}_{X^*(\omega)}$ and $\mathcal{P}_{X^*(\omega)}$ are the relevant σ-algebra and probability measure. Likewise, the probability space that is induced by $Y_{X^*(\omega)}$ is ($\mathcal{S}$, $\mathcal{B}(\mathcal{S})$, P), where $\mathcal{B}(\mathcal{S})$ denotes the Borel sets on $\mathcal{S}$ and P is the probability measure induced by $Y_{X^*(\omega)}$ – for an event A in $\mathcal{S}$ [i.e., $A \in \mathcal{B}(\mathcal{S})$], $P(A) = \mathcal{P}_{X^*(\omega)}(Y_{X^*(\omega)}{}^{-1}(A))$. This is the technical sense in which $Y_{X^*(\omega)}$ is specific to individual ω and specific to the counterfactual value of **X** mandated for that individual.

Y ≡ the random variable representing the observable version of the **Y** (Y is the element of the DGP from which the sampled values of the **Y** can be drawn).

We also define

1) the counterfactually imposed pre-version of the **X**, $X^*(\omega) = X^{pre}(\omega)$, to which there corresponds a potential outcome version of the **Y**,

$Y_{X^*(\omega)} = Y_{X^{pre}(\omega)}$ [12]

and

2) a counterfactually imposed increment to $X^{pre}(\omega)$, $\Delta(\omega)$, that defines a counterfactually imposed post-version of the **X**,

$X^*(\omega) = X^{post}(\omega) = X^{pre}(\omega) + \Delta(\omega)$, with a corresponding potential outcome version of the **Y**, $Y_{X^*(\omega)} = Y_{X^{post}(\omega)} = Y_{X^{pre}(\omega) + \Delta(\omega)}$. Like $X^*(\omega)$, $\Delta(\omega)$ is a known deterministic function (known by the policy maker and/or researcher) that, under a specified counterfactual scenario, also maps individuals (ω) in the relevant population (Ω) to known counterfactually mandated real values. [13]

[12] Here we use the prefixes "pre-" and "post-" in reference to the change in the **X** to be hypothetically mandated as part of the relevant counterfactual query.

[13] Because $X^{pre}(\omega)$ and $X^{post}(\omega) = X^{pre}(\omega) + \Delta(\omega)$ are versions of the generic $X^*(\omega)$, the corresponding respective potential outcomes, $Y_{X^{pre}(\omega)}$ and $Y_{X^{pre}(\omega) + \Delta(\omega)}$, are versions of the generic $Y_{X^*(\omega)}$ and, as such are defined as specific to individual ω and specific to the values assigned to ω, viz. $X^{pre}(\omega)$ and $X^{post}(\omega)$, respectively.

Note that from the perspective of the researcher (or policy maker), neither $X^*(\omega)$ nor $\Delta(\omega)$ is a random variable. They are deterministic functions. Moreover, X and Y are components of the relevant DGP, but $X^*(\omega)$, $\Delta(\omega)$, $Y_{X^{pre}(\omega)}$ and $Y_{X^{pre}(\omega) + \Delta(\omega)}$ are not. In the present context Y and $Y_{X^*(\omega)}$ are count valued but X and $X^*(\omega)$ can be any type of variable. (In this dissertation we focus on the case in which X and $X^*(\omega)$ are continuous variables.)

2.3 Specifying the Causal Effect (CE) in the General Potential Outcomes Framework (GPOF)

To give quantitative specificity to the term "effect", we formalize the "if" component of the above counterfactual query using the following ordered pair representing *the relevant counterfactual* (or simply *the counterfactual*)

$$[X^{pre}(\omega), \Delta(\omega)]. \tag{1}$$

which is the deterministic function that maps an individual (ω) in the relevant population (Ω) to a real-valued pair representing the corresponding counterfactually mandated pre- $[X^{pre}(\omega)]$ and post- $[X^{pre}(\omega) + \Delta(\omega)]$ values of the **X**. The "then" part of the query (the impact on the **Y**) can be stated most generally in terms of the properties of $Y_{X^{pre}(\omega)}$ juxtaposed with those of $Y_{X^{pre}(\omega) + \Delta(\omega)}$ and formalized as a CE of the following form

$$CE(X^{pre}(\omega), \Delta(\omega)) = \mathcal{k}(Y_{X^{pre}(\omega)}, Y_{X^{pre}(\omega) + \Delta(\omega)}) \tag{2}$$

where $k(\ .\ ,\ .\ )$ is a function of $Y_{X^{pre}(\omega)}$ and $Y_{X^{pre}(\omega)+\Delta(\omega)}$. The CE in (2) is most often formulated in terms of the moments of $Y_{X^{pre}(\omega)}$ and $Y_{X^{pre}(\omega)+\Delta(\omega)}$. In the remainder of the discussion, we will focus on the following CE [the *average incremental effect (AIE)*]

$$\mathrm{AIE}(X^{pre}(\omega), \Delta(\omega)) = \mathrm{AVG}_{\omega \in \Omega}\{ E[Y_{X^{pre}(\omega)+\Delta(\omega)} - Y_{X^{pre}(\omega)}]\}$$

$$= \frac{\sum_{\omega \in \Omega} \{ E[Y_{X^{pre}(\omega)+\Delta(\omega)} - Y_{X^{pre}(\omega)}]\}}{n(\Omega)} \tag{3}$$

where n(Ω) denotes the size of the relevant population. We draw the distinction between the expected value and averaging operators (E[ ] vs. AVG{ }) in order to maintain the fundamental conceptual difference between the potential outcomes ($Y_{X^{pre}(\omega)}$ and $Y_{X^{pre}(\omega)+\Delta(\omega)}$) and the components of the corresponding counterfactual ($[X^{pre}(\omega), \Delta(\omega)]$). The former are random variables, and the latter are non-stochastic functions.[14] Henceforth, to simplify exposition, we use the shorthand notation $\mathcal{CE}$ for $\mathrm{CE}(X^{pre}(\omega), \Delta(\omega))$.

---

[14] Note that (3) is quite comprehensive in that it subsumes as special cases: the *average treatment effect* in which the relevant version of the counterfactual in (1) is [0, 1] for all ω ∈ Ω, and (3) becomes

$$\mathrm{AIE}(X^{pre}(\omega) = 0, \Delta(\omega) = 1) = \mathrm{AVG}_{\omega \in \Omega}\{ E[Y_{1(\omega)} - Y_{0(\omega)}]\}$$

$$= \frac{\sum_{\omega \in \Omega} \{ E[Y_{1(\omega)} - Y_{0(\omega)}]\}}{n(\Omega)}$$

where the notation 0(ω) and 1(ω) are used to maintain the ω-specificity of the potential outcomes $Y_{0(\omega)}$ and $Y_{1(\omega)}$, respectively; the *average partial effect* in which $E[Y_{X^{pre}(\omega)+\Delta(\omega)} - Y_{X^{pre}(\omega)}]$ is replaced by

$$\lim_{\Delta(\omega)\to 0} \frac{E[Y_{X^{pre}(\omega)+\Delta(\omega)} - Y_{X^{pre}(\omega)}]}{\Delta(\omega)}$$

so that (3) becomes

$\mathcal{CE}$ as specified in (2), is based on a policy relevant counterfactual and, therefore, is causally interpretable – i.e., the difference between the potential outcomes ($Y_{X^{pre}(\omega)}$ and $Y_{X^{pre}(\omega)\,+\,\Delta(\omega)}$), that $\mathcal{CE}$ represents, albeit counterfactually-based, is exclusively attributable to the imposed counterfactual change in the **X** from $X^{pre}(\omega)$ to $X^{pre}(\omega) + \Delta(\omega)$. Moreover, in the context of any regression-based empirical causal study, once the formulation of $\mathcal{k}(\ .\ ,\ .\ )$ in (2) is specified, the estimation objective is clearly and rigorously established. In the present dissertation, we focus on specification, identification, estimation of (and inference regarding) $\mathcal{CE}$ when the **Y** is non-negative integer valued and the **X** is continuous. First we consider the case in which the causal relationship between **Y** and the **X** is not obscured by the presence of requisite regression control variables that are unobservable – so-called *unobservable confounders* and then we extend the discussion to the case in which there is UC .[15] For example, in an illustration detailed later, we consider the case in which the **X** is the woman's educational attainment, and the **Y** is the number of viable children to whom she has given birth (a count-valued variable). Here, assessment of prospective education/fertility policy via regression-based estimation of the CE of education on fertility is likely to be thwarted by the presence of unobservable confounding (UC).

$$\text{AIE}(X^{pre}(\omega), \Delta(\omega) \rightarrow 0) = \sum_{\omega\,\in\,\Omega} \frac{1}{N(\Omega)} \left\{ \lim_{\Delta(\omega)\,\rightarrow\,0} \frac{E[Y_{X^{pre}(\omega)\,+\,\Delta(\omega)} - Y_{X^{pre}(\omega)}]}{\Delta(\omega)} \right\}.$$

[15] This circumstance is often referred to in the economics literature as involving *endogeneity*. We opt to cast the discussion in terms of unobservable confounding to avoid restrictions and anomalies imposed by the typical (textbook) definition of endogeneity. We provide a formal definition of the term confounding later in the discussion (cf. footnote 22)).

2.4 The Conditional Potential Outcome Model (CPOM): Identification, and Estimation of the Causal Effect (CE)

We turn now to the estimation of the AIE in (3). At issue here is the fundamental disconnect between the inherently counterfactual estimation objective (the CE) and the data available for achieving this objective (sampleable data from the DGP). Data drawn from the DGP are, in general, not sufficiently informative of the counterfactual potential outcomes that underly the CE. In other words, without additional conditions, the AIE in (3) is not identified. As a means of bridging the gap between the counterfactuality of the CE and the factuality of the DGP (thus establishing identification), we suggest a *conditional potential outcomes model* (CPOM), which specifies moments (up to a specified order) of the distribution of the relevant potential outcome (Y_{X^*}) conditional on a vector of regression control variables (V). In the present context, we posit a generic fully parametric CPOM in which the pmf of the conditional count-valued potential outcome ($Y_{X^*} \mid V$) is assumed to be of the following CRM-conformant generic form

$$f_{(Y_{X^*} \mid V)}(Y_{X^*}, V, X^*; \theta) \qquad (4)$$

where θ is the relevant vector of ("deep") regression parameters.[16] Henceforth, we refer to (4) and all its specific versions as the CRM-CPOM. As discussed by Terza (2020, 2024-a-b-c), two sufficient conditions are key to establishing the identification of the AIE in (3). First, the vector of confounder controls V should be sufficient to induce *conditional*

[16] Here and henceforth, we use "f" to represent the various pmf versions pertaining to the **Y**. In general, $f_{(A \mid B)}(A, B; \kappa)$ denotes the conditional pmf of A given B, written as a function of those variates and a parameter vector κ.

independence (CI) between Y_{X^*} and X. Formally, in terms of the CRM-CPOM (4), V induces CI between Y_{X^*} and X if

$$f_{(Y_{X^*} \mid X, V)}(Y_{X^*}, X, V, X^*; \theta) = f_{(Y_{X^*} \mid V)}(Y_{X^*}, V, X^*; \theta). \quad (5)$$

The second condition is *conditional outcome invariance* (COI), which is formally stated in general as

$$f_{(Y_{X^*(\omega)} \mid X, V)}(Y_{X^*(\omega)}, X = X^*(\omega), V, X^*(\omega); \theta)$$
$$= f_{(Y \mid V)}(Y_{X^*(\omega)}, X = X^*(\omega), V, X^*(\omega); \theta). \quad (6)$$

COI requires that, for a population member (ω) whose observable value of the **X** [X(ω)] is $X^*(\omega)$, the pmf of the potential outcomes version of the **Y** at $X^*(\omega)$, conditional on X = $X^*(\omega)$, V $[(Y_{X^*(\omega)} \mid X = X^*(\omega), V)]$, would be the same as the pmf of the observable version of the **Y** conditional on X = $X^*(\omega)$, V $[(Y \mid X = X^*(\omega), V)]$. In other words, regardless of whether the value of the **X** is mandated or produced by the DGP (e.g. is a product of individual choice), the pmf of the **Y** conditional on X = $X^*(\omega)$ and V will be the same. This condition is quite intuitive and generally satisfied. As will be shown in Chapters 3 (in which V comprises only observable regression confounder controls) and 4 (in which V comprises both observable and unobservable controls), if CI and COI hold [in the context of the GPOF and the CPOM in (4)], θ and the AIE in (3) are identified and can be estimated.

Chapter 3

Count Regression Model (CRM) Based Empirical Causal Analysis (ECA) When X is Exogenous

3.1 Overview

In this chapter, the proposed models are specified in the GPOF for the case in the relevant confounder control vector (which induces Conditional independence and conditional outcome invariance) is observable (absence of unobservable confounding), and **Y** is count-valued. In this case, we label the deep parameter vector (θ) as π to distinguish it from the case involving unobservable confounding. We consider three regression estimators for π and the AIE – the ordinary least squares (OLS) estimator and two other proposed estimators based on different CPOM based on the most widely used CRM, the Poisson specification, and the dispersion-flexible Conway-Maxwell-Poisson (CMP) specification (which accommodates all three types of dispersion).

3.2 The General Potential Outcomes Framework (GPOF) When X is Exogenous

All the basic concepts and definitions of the GPOF articulated in chapter 2 remain intact for the present discussion, including the representations of the generic causal effect (CE) and its AIE version in (2) and (3), respectively. As we saw in section 2.4, the CPOM (regression specification) is the means by which to control for the relevant confounders – variables that, if left uncontrolled, will preclude the conditional independence between the count-valued potential outcome and X; thus, impeding identification. When confounders are observable, we say that X is *exogenous*. As a baseline in our quest to account for endogeneity we posit the following version of the fully parametric CPOM in (4) as

$$f_{(Y_{X^*} \mid X_o)}(Y_{X^*}, X_o, X^*; \pi) \tag{7}$$

where X_o is a scalar comprising the observable confounder controls and π is the relevant parameter vector. The CPOM in (7) implies the functional form for the mean of the potential outcome conditional on the controls.

$$E[Y_{X^*} \mid X_o] = m(X_o, X^*; \pi). \tag{8}$$

where

$$m(X_o, X^*; \pi) = \sum_{Y_{X^*}=0}^{\infty} Y_{X^*}\, f_{(Y_{X^*} \mid X_o)}(Y_{X^*}, X_o, X^*; \pi).$$

Using (8) and the law of iterated expectation, the AIE in (3) can be written as

$$\begin{aligned} AIE_{Exog}(X^{pre}(\omega), \Delta(\omega)) &= AVG_{\omega \in \Omega}\{ E[Y_{X^{pre}(\omega)+\Delta(\omega)} - Y_{X^{pre}(\omega)}]\} \\ &= AVG_{\omega \in \Omega}\{ E[m(X_o, X^{pre}(\omega) + \Delta(\omega); \pi) \\ &\qquad - m(X_o, X^{pre}(\omega); \pi)]\}. \end{aligned} \tag{9}$$

Writing (9) in more compact notation we have

$$AIE_{Exog}(X^{pre}(\omega), \Delta(\omega)) = AVG_{\omega \in \Omega}\{E[aie_{Exog}(\pi, \omega)]\} \tag{10}$$

and

$$aie_{Exog}(\pi, \omega) = m(X_o, X^{pre}(\omega) + \Delta(\omega); \pi) - m(X_o, X^{pre}(\omega); \pi). \tag{11}$$

Given (9), it is clear that the AIE will be identified if π is identified.

3.3 Methods When X is Exogenous in General

To estimate the value of the AIE and its corresponding asymptotic inferential statistics in the case where X is exogenous (no unobservable confounding), first we need to estimate the deep parameters of the model in (7) and its asymptotic inferential statistics. The deep parameters π can be estimated using a linear specification of the conditional mean and applying the ordinary least squares (OLS) estimator or by applying the Maximum Likelihood estimator (MLE) using the log-likelihood function of observable version of the specified CPOM in (4). Once the vector of deep parameters has been estimated, using the appropriate version of (8), the AIE can also be estimated using the sample analog to (10).

3.3.1 Estimation of Deep Parameters and the Causal Effect (CE)

Turning to the estimation of π, we note that conditional independence and conditional outcome invariance (as implied by conditioning on X_o) yield[17]

$$\text{pmf}(Y \mid X_o, X) = f_{(Y_{X^*} \mid X_o)}(Y, X_o, X; \pi) \qquad (12)$$

where *pmf(A)* denotes the *probability mass function of the random variable A.* Note that the result in (12), the pmf of the DGP version of $(Y \mid X_o, X)$, is the same as the CPOM with Y and X substituted for Y_{X^*} and X^*, respectively. Thus, equation (12) represents the key identification result. It bridges the gap between the observability of DGP and the unobservability (counterfactuality) of the AIE. Moreover, it provides the means by which to estimate the deep parameter vector π and, therefore, the AIE via the sample analog to (10).

---

[17] See Terza (2020, 2024-a-b-c) for details.

It follows from (12) that π can be consistently estimated as the maximizer of the following log-likelihood function with respect to $\breve{\pi}$

$$\sum_{i=1}^{n} \ln \{ q(\breve{\pi}, Z_i) \} . \tag{13}$$

where $q(\breve{\pi}, Z_i) = f_{(Y_{X^*} \mid X_o)}(Y_i, X_{oi}, X_i; \breve{\pi})$ and $Z_i = [Y_i, X_{oi}, X_i]$. After obtaining the consistent estimate of π (say $\hat{\pi}$), the AIE (10) can be consistently estimated using the following sample analog statistic

$$\widehat{AIE}_{Exog}(X^{pre}, \Delta) = \frac{\sum_{i=1}^{n} aie_{Exog(i)}(\hat{\pi})}{n} \tag{14}$$

where

$$aie_{Exog(i)}(\hat{\pi}) = m(X_{oi}, X_i^{pre} + \Delta_i; \hat{\pi}) - m(X_{oi}, X_i^{pre}; \hat{\pi})$$

and, for the ith observation in a sample size of n (i =1, …, n), X_{oi} is the sampled value of X_o, and X_i^{pre} and Δ_i are the counterfactually mandated values of the X^{pre} and Δ, respectively.

3.3.2 Asymptotic Inference for the Deep Parameters and the Causal Effect (CE)

The asymptotic variance of the deep parameters can be written as

$$\widehat{AVAR}(\hat{\pi}) = E[\nabla_{\pi\pi}\, q(\pi, X)]^{-1} E[\nabla_{\pi}\, q(\pi, X)' \nabla_{\pi}\, q(\pi, X)] E[\nabla_{\pi\pi}\, q(\pi, X)]^{-1} \tag{15}$$

where q is the likelihood function defined as in (13) and $Z = [Y, X_o, X]$. Under general conditions $\hat{\pi}$ is asymptotically normal, i.e.

$$\widehat{\text{AVAR}}(\hat{\pi})^{-\frac{1}{2}}\sqrt{n}\,(\hat{\pi}-\pi)\ \overset{d}{\rightarrow}\ N(0, I). \tag{16}$$

where $\widehat{\text{AVAR}}(\hat{\pi})$ is the consistent estimator of the asymptotic variance-covariance matrix (AVAR) of $\hat{\pi}$.[18] Therefore, the "t-statistic"

$$\frac{\sqrt{n}(\hat{\pi}_k - \pi_k)}{\sqrt{\widehat{\text{AVAR}}(\hat{\pi})_k}} \tag{17}$$

for the kth element of π is asymptotically standard normally distributed [where $\widehat{\text{AVAR}}(\hat{\pi})_k$ denotes the kth diagonal element of $\widehat{\text{AVAR}}(\hat{\pi})$] and can be used to test the hypothesis that $\pi_k = \pi_k^0$ for π_k^0, a given null value of π_k. In practice, the following version of (17) is typically used

$$\frac{\hat{\pi}_k - \pi_k^0}{\sqrt{\widehat{\text{AVAR}}(\hat{\pi})_k^{\dagger}}} \tag{18}$$

where $\widehat{\text{AVAR}}(\hat{\pi})_k^{\dagger}$ denotes the kth diagonal element of $\widehat{\text{AVAR}}(\hat{\pi})^{\dagger}$ which is the same as $\widehat{\text{AVAR}}(\hat{\pi})$ divided by n. We will henceforth refer to the square roots of the diagonal elements of $\widehat{\text{AVAR}}(\hat{\pi})^{\dagger}$ as, the asymptotic standard errors (ASEs) of $\hat{\pi}$.

Turning now to inference regarding $\text{AIE}_{\text{Exog}}(X^{pre}, \Delta)$. The asymptotic variance of the CE is

[18] See Terza (2016-a-b, 2023-a-b, 2024c) for formular details of $\widehat{\text{AVAR}}(\hat{\pi})$ and its Stata 18© implementation.

$$\text{avar}(\widehat{\text{AIE}}_{\text{Exog}}) =$$

$$\text{E}[\nabla_\pi \text{aie}_{\text{Exog}}(\hat{\pi})\,]\text{AVAR}(\hat{\pi})\text{E}[\nabla_\pi \text{aie}_{\text{Exog}}\,(\hat{\pi})]'$$

$$+ \text{E}\left[(\text{aie}_{\text{Exog}}(\hat{\pi})] - \text{AIE}_{\text{Exog}}\,)^2\right] \tag{19}$$

Where AIE_{Exog} and $\widehat{\text{AIE}}_{\text{Exog}}$ are shorthand notation for (10) and (14), respectively; ∇_π denotes the gradient with respect to π; $\text{aie}_{\text{Exog}}(\hat{\pi})$ is defined in (14), and $\text{AVAR}(\hat{\pi})$ denotes the asymptotic covariance matrix of $\hat{\pi}$ as defined in (15). The consistent sample analog estimators of $\text{E}[\nabla_\pi \text{ aie}(\pi)]$ and $\text{E}\left[(\text{aie}_{\text{Exog}}(\pi)] - \text{AIE}_{\text{Exog}})^2\right]$ are, respectively

$$\hat{\text{E}}[\nabla_\pi \text{ aie}_{\text{Exog}}(\hat{\pi})] = \frac{\sum_{i=1}^{n} \nabla_\pi \text{ aie}_{\text{Exog}}(\hat{\pi})}{n} \tag{20}$$

And

$$\hat{\text{E}}\left[\text{aie}_{\text{Exog}}(\hat{\pi})] - \text{AIE}_{\text{Exog}})^2\right] = \frac{\sum_{i=1}^{n} (\nabla_\pi \text{ aie}_{\text{Exog}}(\hat{\pi})] - \widehat{\text{AIE}}_{\text{Exog}})^2}{n} \tag{21}$$

where $\nabla_\pi \text{ aie}_{\text{Exog}}(\hat{\pi})$ denotes $\nabla_\pi \text{ aie}_{\text{Exog}}(\pi)$ evaluated at $X_i, W_i, X_i^{pre}, \Delta_i$ and $\hat{\pi}$. The consistent sample analog estimator of (19) then is

$$\widehat{\text{avar}}(\widehat{\text{AIE}}_{\text{Exog}}) =$$

$$\hat{\text{E}}[\nabla_\pi \text{ aie}_{\text{Exog}}(\hat{\pi})]\ \widehat{\text{AVAR}}(\hat{\pi})\ \hat{\text{E}}[\nabla_\pi \text{ aie}_{\text{Exog}}(\pi)]'$$

$$+ \hat{\text{E}}\left[(\text{aie}_{\text{Exog}}(\hat{\pi})] - \text{AIE}_{\text{Exog}})^2\right] \tag{22}$$

where $\widehat{\text{AVAR}}(\hat{\pi})$ is the estimated AVAR of $\hat{\pi}$ as given in (15). Under general conditions, $\widehat{\text{AIE}}_{\text{Exog}}$ is consistent for AIE_{Exog} and is asymptotically normal, i.e.

$$\sqrt{\frac{n}{\widehat{\text{avar}}\left(\widehat{\text{AIE}}_{\text{Exog}}\right)}}\left(\widehat{\text{AIE}}_{\text{Exog}} - \text{AIE}_{\text{Exog}}\right) \overset{d}{\rightarrow} n(0, 1). \tag{23}$$

In other words, the "t-statistic"

$$\frac{\widehat{\text{AIE}}_{\text{Exog}} - \text{AIE}_{\text{Exog}}}{\sqrt{\widehat{\text{avar}}(\widehat{\text{AIE}}_{\text{Exog}})/n}}$$

is asymptotically standard normal distributed and can be used to test the hypothesis that $\text{AIE}_{\text{Exog}} = \text{AIE}_{\text{Exog}}{}^{0}$, where $\text{AIE}_{\text{Exog}}{}^{0}$ is a given null value of AIE_{Exog}.

3.4 Alternative Estimators When X is Exogenous: Full Information Maximum Likelihood (FIML) Poisson (FIML-Posson), CMP (FIML-CMP) and Ordinary Least Squares (OLS)

In the case of an exogenous X, the deep parameters of the model can be estimated using full information maximum likelihood (FIML) estimation if a functional form for the CPOM is assumed, or ordinary least squares (OLS) for the linearized version of the model. In this section we detail both approaches. For the FIML cases, two specifications of the fully parametric CPOM are discussed (Poisson and CMP distributions).

3.4.1 Full Information Maximum Likelihood Poisson (FIML-Poisson)

One version of CPOM is the Poisson pmf where its generic version can be written as

$$f_{(Y_{X^*} \mid X_o)}(Y_{X^*}, X_o, X^*; \pi) = POI(Y_{X^*}, \lambda^*) = \frac{(\lambda^*)^{Y_{X^*}} \exp(-\lambda^*)}{Y_{X^*}!} \quad (24)$$

where

$Y_{X^*} = 0, 1, 2, 3, ...$

$\lambda^* = \exp(X_o\beta_o + X^*\beta_x).$

$\pi' = [\beta_o' \quad \beta_x]$

If (24) holds, we get that the relevant version of (8) is

$$E[Y_{X^*} \mid X_o] = \exp(X_o\beta_o + X^*\beta_x). \quad (25)$$

Using (25), by applying the law of iterated expectations (LIE), we can rewrite the AIE in (10) for the Poisson case as

$$AIE_{(POI)}(X^{pre}(\omega), \Delta(\omega)) = AVG_{\omega \in \Omega}\{E[aie_{(POI)}(\pi, \omega)]\} \quad (26)$$

where

$$aie_{(POI)}(\pi, \omega)= E[\exp(X_o\beta_o + (X^{pre} + \Delta)\beta_x) - \exp(X_o\beta_o + X^{pre}\beta_x)]. \quad (27)$$

Given (26), it is clear that the AIE will be identified if π is identified. Now if CI and COI hold (as implied by conditioning on X_o), then

$$\text{pmf}(Y \mid X_o, X) = f_{(Y_{X^*} \mid X_o)}(Y, X_o, X; \pi) = \text{POI}(Y, \lambda) = \frac{\lambda^Y \exp(-\lambda)}{Y!} \tag{28}$$

so that π is indeed identified. Moreover, we get from (28) that π can be consistently estimated as $\hat{\pi}' = [\hat{\beta}_o{}' \quad \hat{\beta}_x]$, the maximizer of the following log-likelihood function with respect to $\breve{\pi}$

$$\sum_{i=1}^{n} \ln \{ q(\breve{\pi}, Z_i) \} . \tag{29}$$

where $q(\breve{\pi}, Z_i) = \text{POI}(Y_i, \lambda_i)$, $Z_i = [Y_i, X_{oi}, X_i]$, $\lambda_i = \exp(X_{oi}\breve{\beta}_o + X_i\breve{\beta}_x)$ and $\breve{\pi}' = [\breve{\beta}_o{}' \quad \breve{\beta}_x]$. In this case the sample analog estimator of (26) is the following version of (14)

$$\widehat{\text{AIE}}_{(\text{POI})}(X^{\text{pre}}, \Delta) = \frac{\sum_{i=1}^{n} \text{aie}_{(\text{POI})(i)}(\hat{\pi})}{n} \tag{30}$$

where

$$\text{aie}_{(\text{POI})(i)}(\hat{\pi}) = \exp(X_{oi}\hat{\beta}_o + [X_i^{\text{pre}} + \Delta_i]\hat{\pi}_x) - \exp(X_{oi}\hat{\beta}_o + X_i^{\text{pre}}\beta_x). \tag{31}$$

We also have that

$$\frac{\widehat{\mathrm{AIE}}_{(\mathrm{POI})} - \mathrm{AIE}_{(\mathrm{POI})}}{\sqrt{\widehat{\mathrm{avar}}(\widehat{\mathrm{AIE}}_{(\mathrm{POI})})/n}} \overset{d}{\to} n(0, 1). \tag{32}$$

where

$$\widehat{\mathrm{avar}}\left(\widehat{\mathrm{AIE}}_{(\mathrm{POI})}\right) = \hat{\Psi}_{(\mathrm{POI})}\ \widehat{\mathrm{AVAR}}_{(\mathrm{POI})}(\hat{\pi})\ \hat{\Psi}_{(\mathrm{POI})}' + \hat{\Lambda}_{(\mathrm{POI})} \tag{33}$$

$$\hat{\Psi}_{(\mathrm{POI})} = \frac{\sum_{i=1}^{n} \nabla_{\pi} \mathrm{aie}_{(\mathrm{POI})(i)}(\hat{\pi})}{n} \tag{34}$$

$$\hat{\Lambda}_{\mathrm{POI}} = \frac{\sum_{i=1}^{n}\left(\mathrm{aie}_{(\mathrm{POI})(i)}(\hat{\pi}) - \widehat{\mathrm{AIE}}\right)^2}{n} \tag{35}$$

and $\mathrm{aie}_{(\mathrm{POI})(i)}(\hat{\pi})$ is defined as in (31), and $\widehat{\mathrm{AVAR}}_{(\mathrm{POI})}(\hat{\pi})$ is the estimated AVAR of $\hat{\pi}$. As discussed earlier, inference regarding the elements of π can be conducted using $\widehat{\mathrm{AVAR}}_{(\mathrm{POI})}(\hat{\pi})$ based on the asymptotic normality of $\hat{\pi}$. With $\hat{\pi}$ and $\widehat{\mathrm{AVAR}}_{(\mathrm{POI})}(\hat{\pi})$ in hand, we can also conduct inference regarding $\mathrm{AIE}_{(\mathrm{POI})}(X^{pre}(\omega), \Delta(\omega))$ using (32).

3.4.2 Full Information Maximum Likelihood Conway-Maxwell-Poisson (FIML-CMP)

As a more dispersion-flexible alternative to the Poisson, we consider the following version of the generic CRM-CPOM in (7)

$$f_{(Y_{X^*} \mid X_o)}(Y_{X^*}, X_o, X^*; \pi) = \mathrm{CMP}(Y_{X^*}; \lambda^*, \psi) = \frac{(\lambda^*)^{Y_{X^*}}}{Y_{X^*}!^{\exp(\psi)}\ Z(\lambda^*, \exp(\psi))} \tag{36}$$

where

$$Y_{X^*} = 0, 1, 2, 3, \ldots$$

$$\lambda^* = \exp(X_o\beta_o + X^*\beta_x)$$

$$\pi' = [\beta_o' \quad \beta_x \quad \psi]$$

$$Z(\lambda^*, \exp(\psi)) = \sum_{j=0}^{\infty} \frac{(\lambda^*)^j}{(j!)^{\exp(\psi)}}$$

$$-\infty < \psi = \ln(\upsilon) < \infty$$

and υ the dispersion parameter ($\upsilon > 0$). Equi-dispersion corresponds to the case when υ = 1 (i.e., $\psi = 0$), over-dispersion prevails when $\upsilon < 1$ (i.e., $\psi < 0$), and under-dispersion holds if $\upsilon > 1$ (i.e., $\psi > 0$). In this case If (36) holds, we have the relevant version of (8) as

$$E[Y_{X^*} \mid X_o] = m_{CMP}(\lambda^*, \psi) = \lambda^* \sum_{j=1}^{\infty} \frac{j(\lambda^*)^{j-1}}{(j!)^{\exp(\psi)} Z(\lambda^*, \exp(\psi))} \tag{37}$$

Using (37), by applying the LIE, we can rewrite the AIE in (10) for the CMP case as

$$AIE_{(CMP)}(X^{pre}(\omega), \Delta(\omega)) = AVG_{\omega \in \Omega}\{E[aie_{(CMP)}(\pi, \omega)]\} \tag{38}$$

where

$$aie_{(CMP)}(\pi, \omega) =$$

$$E[m_{CMP}(\exp(X_o\beta_o + (X^{pre} + \Delta)\beta_x), \psi)$$

$$- m_{CMP}(\exp(X_o\beta_o + X^{pre}\beta_x), \psi)]. \tag{39}$$

Given (26), it is clear that the AIE will be identified if π is identified. Now if CI and COI hold (as implied by conditioning on X_o), then

$$\text{pmf}(Y \mid X_o, X) = f_{(Y_{X^*} \mid X_o)}(Y, X_o, X; \pi) =$$

$$\text{CMP}(Y, \lambda, \psi) = \frac{(\lambda)^Y}{Y!^{\exp(\psi)}\, Z(\lambda, \exp(\psi))} \tag{40}$$

where

$$\lambda = \exp(X_o\beta_o + X\beta_x).$$

so that π is indeed identified. Moreover, we get from (40) that π can be consistently estimated as $\hat{\pi}' = [\hat{\beta}_o{}' \quad \hat{\beta}_x \quad \hat{\psi}]$, the maximizer of the following log-likelihood function with respect to $\breve{\pi}$

$$\sum_{i=1}^{n} \ln\{\, q(\breve{\pi}, Z_i)\} \,. \tag{41}$$

where $q(\breve{\pi}, Z_i) = \text{CMP}(Y_i, \lambda_i, \psi_i)$, $Z_i(\lambda_i, \exp(\psi_i)) = \sum_{j=0}^{\infty} \frac{(\lambda_i)^j}{(j!)^{\exp(\psi_i)}}$, $\lambda_i = \exp(X_{oi}\breve{\beta}_o + X_i\breve{\beta}_x)$ and $\breve{\pi}' = [\breve{\beta}_o{}' \quad \breve{\beta}_x \quad \breve{\psi}]$. In this case the sample analog estimator of (38) is the following version of (14)

$$\widehat{\text{AIE}}_{(\text{CMP})}(X^{pre}, \Delta) = \frac{\sum_{i=1}^{n} \text{aie}_{(\text{CMP})(i)}(\hat{\pi})}{n} \tag{42}$$

where

$$\text{aie}_{(CMP)(i)}(\hat{\pi}) = m_{CMP}(\exp(X_{oi}\hat{\beta}_o + [X_i^{pre} + \Delta_i]\hat{\beta}_x), \hat{\psi}) - m_{CMP}(\exp(X_{oi}\hat{\beta}_o + X_i^{pre}\hat{\beta}_x), \hat{\psi}). \tag{43}$$

We also have that

$$\frac{\widehat{AIE}_{(CMP)} - AIE_{(CMP)}}{\sqrt{\widehat{avar}(\widehat{AIE}_{(CMP)})/n}} \overset{d}{\rightarrow} n(0, 1). \tag{44}$$

where,

$$\widehat{avar}\left(\widehat{AIE}_{(CMP)}\right) = \hat{\Psi}_{(CMP)}\, \widehat{AVAR}_{(CMP)}(\hat{\pi})\, \hat{\Psi}_{(CMP)}' + \hat{\Lambda}_{(CMP)} \tag{45}$$

$$\hat{\Psi}_{(CMP)} = \frac{\sum_{i=1}^{n} \nabla_{\pi}\text{aie}_{(CMP)(i)}(\hat{\pi})}{n} \tag{46}$$

$$\hat{\Lambda}_{CMP} = \frac{\sum_{i=1}^{n}\left(\text{aie}_{(CMP)(i)}(\hat{\pi}) - \widehat{AIE}\right)^2}{n} \tag{47}$$

and $\text{aie}_{(CMP)(i)}(\hat{\pi})$ is defined as in (43), and $\widehat{AVAR}_{(CMP)}(\hat{\pi})$ is the estimated AVAR of $\hat{\pi}$. As discussed earlier, inference regarding the elements of π can be conducted using $\widehat{AVAR}_{(CMP)}(\hat{\pi})$ based on the asymptotic normality of $\hat{\pi}$. With $\hat{\pi}$ and $\widehat{AVAR}_{(CMP)}(\hat{\pi})$ in hand, we can also conduct inference conduct inference regarding $AIE_{(CMP)}(X^{pre}(\omega), \Delta(\omega))$ using (44).[19]

3.4.3 Linear Model (OLS)

As a baseline for comparison, we consider a less parametric CPOM that comports with the classical linear regression model – less parametric in that it specifies merely the mean rather that the full pmf of $(Y_{X^*} \mid X, X_o)$. The relevant CPOM here is

$$E[Y_{X^*} \mid X_o] = X_o\beta_o + X^*\beta_x \tag{48}$$

where, for this linear case, $\pi' = [\beta_o{}' \quad \beta_x]$, Using (48), by applying the LIE, we can rewrite the AIE in (10) as

$$\begin{aligned} &AIE_{(LIN)}(X^{pre}(\omega), \Delta(\omega)) \\ &\qquad = E[((X^{pre}(\omega) + \Delta(\omega))\beta_x + X_o\beta_o)] - E[(X^{pre}(\omega)\beta_x + X_o\beta_o)]] \\ &\qquad = E[\Delta(\omega)]\beta_x \end{aligned} \tag{49}$$

Given (49), it is clear that the AIE will be identified if β_x is identified. Now if CI and COI hold (as implied by conditioning on X_o), then

$$E[Y \mid X_o, X] = X_o\beta_o + X\beta_x \tag{50}$$

[19] There is currently no packaged command in Stata software to implement the CMP regression and it must be coded in Stata by the user (e.g., in Mata using the **`moptimize`** option).

so that β_x is indeed identified. Moreover, we get from (50) that β_x can be consistently estimated by applying OLS to the following classical linear regression model for a sample of size i = 1, . . ., n

$$Y_i = X_{oi}\beta_o + X_i\beta_x + e_i \tag{51}$$

where

$$E[Y_i \mid X_{oi}, X_i] = X_{oi}\beta_o + X_i\beta_x$$

and e_i denotes the classical regression error term ($E[e_i \mid X_{oi}, X_i] = 0$). With OLS estimate $\hat{\pi}' = [\hat{\beta}_o' \quad \hat{\beta}_x]$, we can estimate the value of the AIE using the following sample analog statistic

$$\widehat{AIE}_{(LIN)}(X^{pre}, \Delta) = \frac{1}{n}\sum_{i=1}^{n}\left\{X_i^{pre} + \Delta_i]\hat{\beta}_x + X_{oi}\hat{\beta}_o - (X_i^{pre}\hat{\beta}_x + X_{oi}\beta_o)\right\}$$
$$= \frac{1}{n}\sum_{i=1}^{n} aie_{(LIN)(i)}(\hat{\beta}_x) = \hat{\beta}_x \left(\frac{1}{n}\sum_{i=1}^{n}\Delta_i\right). \tag{52}$$

where $aie_{(LIN)(i)}(\hat{\beta}_x) = \hat{\beta}_x\Delta_i$ .We also have that

$$\frac{\widehat{AIE}_{(LIN)} - AIE_{(LIN)}}{\sqrt{\widehat{avar(\widehat{AIE}_{(LIN)})}/n}} \overset{d}{\to} n(0, 1). \tag{53}$$

It is easy to show, however, that

$$\frac{\widehat{AIE}_{(LIN)} - AIE_{(LIN)}}{\sqrt{\widehat{avar}(\widehat{AIE}_{(LIN)})/n}} = \frac{\hat{\beta}_x - \beta_x}{\sqrt{\widehat{avar}(\hat{\beta}_x)}} \tag{54}$$

where $\sqrt{\widehat{avar}(\hat{\beta}_x)}$ is the convention asymptotic OLS standard error for $\hat{\beta}_x$. Therefore (54) can be used to test the hypothesis that $AIE_{(LIN)} = AIE^0_{(LIN)}$ for $AIE^0_{(LIN)}$, a given null value of$AIE_{(LIN)}$.

Chapter 4

Count Regression Model Based (CRM-Based) Empirical Causal Analysis (ECA) When X Is Endogenous: Two-Stage Residual Inclusion (2SRI) Estimation

4.1 Overview

In this chapter, the proposed models are specified in the GPOF for the case in which the presumed causal entity **X** is endogenous, and **Y** is count-valued. Five regression estimators for the AIE and their asymptotic statistics discussed in this chapter are the Linear Instrumental Variable (LIV), two partially linear control function estimators and two other newly proposed estimators based on two distributions for the count data models. The count specific distributions include the most widely used Poisson specification and the dispersion-flexible CMP specification (to accommodate all three types of dispersion).

4.2 The General Potential Outcomes Framework (GPOF) Extended to Account for Unobservable Confounding (UC) [Endogeneity]

Considering the basic concepts and definitions of the GPOF articulated in sections 2.2 and 2.3, the CPOM (regression specification) is the means by which to control for the relevant confounders – variables that, if left uncontrolled, will preclude the conditional independence between the count-valued potential outcome and X; thus, impeding identification. When some of the confounders are not observable (and, therefore, cannot be directly included in the CPOM) we say that X is subject to unobservable confounding (i.e., X is *endogenous*). As a baseline in our quest to account for unobservable confounders we extend the fully parametric CPOM in (7) as

$$f_{(Y_{X^*} \mid X_o, X_u)}(Y_{X^*}, X_o, X_u, X^*; \gamma) \quad (55)$$

where X_u is a scalar comprising the unobservable confounders, X_o is a scalar comprising the observable confounders and γ is the relevant parameter vector. The additional conditioning on X_u serves to maintain conditional independence between Y_{X^*} and X, which implies that[20]

$$\text{pmf}(Y \mid X_o, X_u) = f_{(Y_{X^*} \mid X_o, X_u)}(Y, X, X_o, X_u; \gamma) \quad (56)$$

The main obstacle that must be confronted in the use of (56) for the estimation of the parameter vector π is, of course, the non-observability of X_u.

To deal with the non-observability of X_u, we follow the two-stage residual inclusion (2SRI) approach propounded by Terza et al. (2008) and replace it with a representative proxy that is observable up to a vector of estimable parameters. In the generic 2SRI protocol, we have the following *residual*

$$X_u = \mathcal{X}_u(X, W; \delta) \quad (57)$$

where δ is a vector of unknown parameters and $W = [X_o \quad W^+]$, with W^+ being a vector of observable identifying instrument variables (IVs).[21] Terza (2024b) discusses the conditions under which the following equality holds

$$f_{(Y_{X^*} \mid X, W)}(Y_{X^*}, X, W, X^*; \gamma) = f_{(Y_{X^*} \mid X_o, X_u)}(Y_{X^*}, X_o, \mathcal{X}_u(X, W; \delta), X^*; \gamma) \quad (58)$$

[20] See Terza (2020, 2024-a-b-c) for details.

[21] The term “residual” is a misnomer, it is the estimated version of (57) that is actually the “residual”.

which implies that we can write the corresponding conditional mean function as

$$E[Y_{X^*} \mid X, W] = m(X_o, \mathcal{X}_u(X, W; \delta), X^*; \gamma). \tag{59}$$

where

$$m(X_o, X_u, X^*; \gamma) = \sum_{Y_{X^*}=0}^{\infty} Y_{X^*}\, f_{(Y_{X^*} \mid X_o, X_u)}(Y_{X^*}, X_o, X_u, X^*; \gamma).$$

Then, using (59) and the LIE, we can write the relevant version of (3) [i.e., with unobservable confounding] as

$$\begin{aligned} AIE_{UC}(X^{pre}(\omega), \Delta(\omega)) &= AVG_{\omega \in \Omega}\{ E[Y_{X^{pre}(\omega)+\Delta(\omega)} - Y_{X^{pre}(\omega)}]\} \\ &= AVG_{\omega \in \Omega}\{ E[E[Y_{X^{pre}(\omega)+\Delta(\omega)} \mid X, W] - E[Y_{X^{pre}(\omega)} \mid X, W]] \} \\ &= AVG_{\omega \in \Omega}\{ E[m(X_o, \mathcal{X}_u(X, W; \delta), X^{pre}(\omega) + \Delta(\omega); \gamma) \\ &\qquad - m(X_o, \mathcal{X}_u(X, W; \delta), X^{pre}(\omega); \gamma)] \}. \end{aligned} \tag{60}$$

Writing (60) in more compact notation we have

$$AIE_{UC}(X^{pre}(\omega), \Delta(\omega)) = AVG_{\omega \in \Omega}\{E[aie_{UC}(\theta, \omega)]\} \tag{61}$$

where $\theta = [\delta' \quad \gamma']$ and

$$\begin{aligned} aie_{UC}(\theta, \omega) = \\ &m(X_o, \mathcal{X}_u(X, W; \delta), X^{pre}(\omega) + \Delta(\omega); \gamma) \\ &\qquad - m(X_o, \mathcal{X}_u(X, W; \delta), X^{pre}(\omega); \gamma). \end{aligned} \tag{62}$$

If we can establish the identification of θ and had an estimator of it (say $\hat{\theta}$) [to be discussed in the next section] we would, by implication, have the identification of $AIE_{UC}(X^{pre}(\omega), \Delta(\omega))$ in (61). Moreover, we could consistently estimate (61) via the following sample analog statistic

$$\widehat{AIE}_{UC}(X^{pre}, \Delta) = \frac{\sum_{i=1}^{n} aie_{UC(i)}(\hat{\theta})}{n} \quad (63)$$

where

$$aie_{UC(i)}(\hat{\theta}) = m(X_{oi}, \mathcal{X}_u(X_i, W_i; \hat{\delta}), X_i^{pre} + \Delta_i; \hat{\gamma}) - m(X_{oi}, \mathcal{X}_u(X_i, W_i; \hat{\delta}), X_i^{pre}; \hat{\gamma})$$

and, for the ith observation in a sample size of n (i =1, …, n), X_i^{pre} and Δ_i are the counterfactually mandated values of the X^{pre} and Δ, respectively; and $W_i = [X_{oi} \quad W_i^{+}]$ is the sampled value of W.

In this chapter, we will discuss the five alternative CRM-CPOM specifications in the presence of unobservable confounding that we consider in this dissertation and their implied versions of the conditional mean function m( ) in (63). As for the identification conditions, discussed by Terza (2020, 2024a-b-c), the vector of confounder controls V needs to be sufficient to induce CI between Y_{X^*} and X, formally defined as in (5), in terms of the CRM-CPOM (7). We formally define an element of V to be a *confounder* if dropping it from V would invalidate the equality in (5).[22] Finally, we say that V provides *comprehensive confounder control* if it includes all relevant confounders. In (55), we

[22] This provides the promised formal definition of the term *confounder/confounding* (cf. footnotes 3 and 18).

specify the vector of regression controls as $[X_o \quad X_u]$, which clearly provides comprehensive confounder control so that

$$f_{(Y_{X^*} \mid X_o, X_u, X)}(Y_{X^*}, X, X_o, X_u, X^*; \gamma) = f_{(Y_{X^*} \mid X_o, X_u)}(Y_{X^*}, X_o, X_u, X^*; \gamma) \qquad (64)$$

i.e., $[X_o \quad X_u]$ induces CI between Y_{X^*} and X. Confounders that are not included among the observables (X_o) will be captured by X_u.

The other salient condition for identification, the COI, is formally stated in the present UC context as

$$f_{(Y_{X^*(\omega)} \mid X_o, X_u, X)}(Y_{X^*(\omega)}, X = X^*(\omega), X_o, X_u, X^*(\omega))$$

$$= f_{(Y \mid X, X_o, X_u)}(Y_{X^*(\omega)}, X = X^*(\omega), X_o, X_u, X^*). \qquad (65)$$

COI requires that, for a population member (ω) whose observable value of the **X** [X(ω)] is $X^*(\omega)$, the pmf of the potential outcomes version of the **Y** at $X^*(\omega)$, conditional on $X = X^*(\omega)$, X_o and X_u $[(Y_{X^*(\omega)} \mid X = X^*(\omega), X_o, X_u)]$, would be the same as the pmf of the observable version of the **Y** conditional on $X = X^*(\omega)$, X_o and X_u $[(Y \mid X = X^*(\omega), X_o, X_u])]$. In other words, regardless of whether the value of the **X** is mandated or produced by the DGP (e.g. is a product of individual choice), the pmf of the **Y** conditional on $X = X^*(\omega)$, X_o and X_u will be the same. This condition is quite intuitive and generally satisfied. Based on the argument given by Terza (2024b), using (55), (64) and (65), we have that

$$f_{(Y \mid X, W)}(Y, X, W; \delta, \gamma) = f_{(Y_{X^*} \mid X_o, X_u)}(Y, X_o, \mathcal{X}_u(X, W; \delta), X; \gamma). \qquad (66)$$

Equation (66) establishes the identification of $\theta = [\delta' \quad \gamma']$ and, by implication, the identification of AIE_{UC} in (61). The subscript expression for "f" on the right-hand side of (66) implies that the functional form of the left-hand side is the same as that of (58). Note that (66) is obtained by substituting Y, X, and $\mathcal{X}_u(X, W; \delta)$ for Y_{X^*}, X^* and $X_u$in (56), respectively. Let us now turn to consistent estimation of θ and, by implication, consistent estimation of AIE_{UC}.

4.3 The Two Stage Residual Inclusion (2SRI) Approach in General: First-Order (FO) Error/Residuals

We focus here on the version of the 2SRI protocol in which the relevant form of the residual in (57) is

$$X_u = X - r(W; \delta) \tag{67}$$

where r(.) = E[X | W] is a known conditional mean regression function and δ is the vector of estimable regression parameters, so that X_u is the regression error term in

$$X = r(W; \delta) + X_u \tag{68}$$

with $E[X_u \mid W] = 0$. We refer to X_u defined in this way as a *first-order (FO) residual* because it is based solely on the conditional mean E[X | W] – the first-order conditional moment of X given W. Writing the (67) explicitly as function of X, W and δ we have

$$\mathcal{X}_u^{FO}(X, W; \delta) = X_u = X - r(W; \delta). \tag{69}$$

In this case, the relevant version of (61) is

$$\text{AIE}_{2SRI}(X^{pre}(\omega), \Delta(\omega)) = \text{AVG}_{\omega \in \Omega}\{E[\text{aie}_{2SRI}(\theta, \omega)]\} \tag{70}$$

where $\theta = [\delta' \quad \gamma']$ and

$$\text{aie}_{2SRI}(\theta, \omega) = m(X_o, \mathcal{x}_u^{FO}(X, W; \delta), X^{pre}(\omega) + \Delta(\omega); \gamma) \\ - m(X_o, \mathcal{x}_u^{FO}(X, W; \delta), X^{pre}(\omega); \gamma). \tag{71}$$

and the relevant form of m( ) is implied by the functional form specified for (66).

4.3.1 Estimation of Deep Parameters and the Causal Effect (CE) [The Average Incremental Effect (AIE)]

Given that θ is identified, it can be consistently estimated via the following 2SRI protocol:[23]

<u>First Stage</u>:

Consistently estimate δ using a regression method based on (68) or assumptions about the distribution of (X | W) that imply it, then calculate the relevant FO residuals using

$$\mathcal{x}_u^{FO}(X_i, W_i; \hat{\delta}) = X_i - r\left(W_i; \hat{\delta}\right) \tag{72}$$

where $\hat{\delta}$ is the first-stage consistent estimate of δ and the subscript i = 1, . . ., n denotes the ith member of a sample of size n.

[23] See Terza et al. (2008).

Second Stage:

Based on (66), consistently estimate γ as the maximizer of the following objective function with respect to $\breve{\gamma}$

$$\sum_{i=1}^{n} \ln \{ f_{(Y_{X^*} \mid X_o, X_u)}(Y_i, X_i, X_{oi}, \mathcal{X}_u(X_i, W_i; \hat{\delta}); \breve{\gamma})\} . \tag{73}$$

where $\hat{\delta}$ is assumed to be given and fixed.

After obtaining the 2SRI consistent estimators of δ and γ, the AIE (70) can be consistently estimated using the following version of the sample analog statistic in (70)

$$\widehat{AIE}_{2SRI}(X^{pre}, \Delta) = \frac{\sum_{i=1}^{n} aie_{2SRI(i)}(\hat{\theta})}{n} \tag{74}$$

where

$$aie_{2SRI(i)}(\hat{\theta}) = m(X_{oi}, \mathcal{X}_u(X_i, W_i; \hat{\delta}), X_i^{pre} + \Delta_i; \hat{\gamma}) - m(X_{oi}, \mathcal{X}_u(X_i, W_i; \hat{\delta}), X_i^{pre}; \hat{\gamma}).$$

4.3.2 Asymptotic Inference for Deep Parameters and the Causal Effect (CE) [The Average Incremental Effect (AIE)]

The vector of parameters to be estimated in the present 2SRI context ($\theta = [\delta' \quad \gamma']$), can be formally defined in the following way

$$\delta \;=\; \underset{\breve{\delta}}{\operatorname{argmax}}\; E[q_1(\breve{\delta}, X, W)] \tag{75}$$

and

$$\gamma \;=\; \underset{\breve{\gamma}}{\text{argmax}}\; E[q_2(\delta, \breve{\gamma}, Y, X, W)] \qquad (76)$$

where $\breve{\delta}$ and $\breve{\gamma}$ represent the generic arguments of (75) and (76), respectively

$$q_2(\delta, \breve{\gamma}, Y, W) \equiv \ln\{f_{(Y_{X^*} \mid X_o, X_u)}(Y, X, X_o, \mathcal{X}_u(X, W; \delta); \breve{\gamma})\} \qquad (77)$$

and

$q_1(\breve{\delta}, X, W) \equiv$ a similar function pertinent to the conditional mean regression in (68).

For example, we might specify

$$q_1(\breve{\delta}, X, W) = -\,(X - r(W; \delta))^2 \qquad (78)$$

or if the functional form of $f_{(X \mid W)}(X, W; \delta)$ were known, the relevant specification would be

$$q_1(\breve{\delta}, X, W) = \ln\{f_{(X \mid W)}(X, W; \delta)\} \qquad (79)$$

Based on (77) and (79) we can characterize the 2SRI estimator as the following two-stage M-estimator (2SME)

First Stage:

$$\hat{\delta} \;=\; \underset{\breve{\delta}}{\text{argmax}} \sum_{i=1}^{n} q_1(\breve{\delta}, X_i, W_i) \qquad (80)$$

and

Second Stage:

$$\hat{\gamma} = \underset{\breve{\gamma}}{\text{argmax}} \sum_{i=1}^{n} q_2(\hat{\delta}, \breve{\gamma}, Y_i, X_i, W_i) \tag{81}$$

where, for the definition of $\hat{\gamma}$ in (81), the estimate $\hat{\delta}$ is assumed to be given and fixed, and the "i" subscript refers to the ith member of a sample of size n (i = 1, ..., n), drawn from the data generating process [Y X W] – the vector of observable relevant random variables.

Henceforth, we maintain the following notational conventions:

$\nabla_\delta q_1$ is shorthand for $q_1(\breve{\delta}, X, W)\big|_{\breve{\delta}=\delta}$

$\nabla_\delta q_2$ is shorthand for $\nabla_{\breve{\delta}} q_2(\delta, \breve{\gamma}, Y, X, W)\big|_{\breve{\delta}=\delta,\, \breve{\gamma}=\gamma}$

$\nabla_\gamma q_2$ is shorthand for $\nabla_{\breve{\gamma}} q_2(\delta, \breve{\gamma}, Y, X, W)\big|_{\breve{\delta}=\delta,\, \breve{\gamma}=\gamma}$

$\nabla_{\gamma\delta} q_2$ denotes the matrix whose typical (jk^{th}) element is

$$\left.\frac{\partial^2 q_2(\breve{\delta}, \breve{\gamma}, Y, X, W)}{\partial\breve{\gamma}_j\, \partial\breve{\delta}_k}\right|_{\breve{\delta}=\delta,\, \breve{\gamma}=\gamma} \tag{82}$$

where the gradients $\nabla_\delta q_1$, $\nabla_\delta q_2$ and $\nabla_\gamma q_2$ are row vectors; the j and k subscripts refer to specific elements of γ and δ, respectively; and the row dimension of $\nabla_{\gamma\delta} q_2$ (j = 1, …, J) corresponds to that of its first subscript and its column dimension (k = 1, …, K) to that of its second subscript.

Terza (2016a, 2023-a-b) shows that the asymptotic variance (AVAR) of $\hat{\theta} = [\hat{\delta}'\ \hat{\gamma}']$ can be written

$$\text{AVAR}(\hat{\theta}) = \begin{bmatrix} D_{11} & D_{12} \\ D'_{12} & D_{22} \end{bmatrix} \tag{83}$$

where

$$D_{11} = \text{AVAR}(\hat{\delta}) = E\left[\nabla_{\delta\delta} q_1\right]^{-1} E\left[\nabla_{\delta} q_1{}' \nabla_{\delta} q_1\right] E\left[\nabla_{\delta\delta} q_1\right]^{-1} \tag{84}$$

$$D_{12} = -\,\text{AVAR}(\hat{\delta})\; E[\nabla_{\gamma\delta} q_2]' E[\nabla_{\gamma\gamma} q_2]^{-1} \tag{85}$$

$$D_{22} = E[\nabla_{\gamma\gamma} q_2]^{-1} \left[E[\nabla_{\gamma\delta} q_2]\; \text{AVAR}(\hat{\delta})\; E[\nabla_{\gamma\delta} q_2]' + E[\nabla_{\gamma} q_2'\; \nabla_{\gamma} q_2]\right] E[\nabla_{\gamma\gamma} q_2]^{-1} \tag{86}$$

and $\text{AVAR}(\hat{\delta})$ denotes the AVAR of $\hat{\delta}$. To conduct inference regarding $\theta = [\delta' \quad \gamma']$ we will need a consistent estimator of (83). The requisite consistent estimator of $\text{AVAR}(\hat{\theta})$ is obtained by the following substitutions in (84), (85), and (86):

$$\widehat{\text{AVAR}}(\hat{\delta}) \qquad \text{for AVAR}(\hat{\delta})\text{, where} \tag{87}$$

$\widehat{\text{AVAR}}(\hat{\delta})$ = estimated AVAR for the 2SRI first-stage estimator of δ

$$\hat{E}[\nabla_{\gamma} q_2'\; \nabla_{\gamma} q_2\,] = \textstyle\sum_{i=1}^{n} \nabla_{\gamma} \hat{q}_{2i}'\; \nabla_{\gamma} \hat{q}_{2i} \qquad \text{for} \qquad E[\nabla_{\gamma} q_2'\; \nabla_{\gamma} q_2\,]\text{, where} \tag{88}$$

$$\nabla_{\gamma} \hat{q}_{2i} = \nabla_{\breve{\gamma}} q_2(\breve{\delta}, \breve{\gamma}, Y_i, X_i, W_i)\Big|_{\breve{\delta}=\hat{\delta},\, \breve{\gamma}=\hat{\gamma}}$$

$$\hat{E}[\nabla_{\gamma\gamma} q_2]^{-1} = [\textstyle\sum_{i=1}^{n} \nabla_{\gamma\gamma} \hat{q}_{2i}]^{-1} \qquad \text{for} \qquad E[\nabla_{\gamma\gamma} q_2]^{-1} \tag{89}$$

$$\hat{E}[\nabla_{\gamma\delta} q_2] = \textstyle\sum_{i=1}^{n} \nabla_{\gamma\delta} \hat{q}_{2i} \qquad \text{for} \qquad E[\nabla_{\gamma\delta} q_2]\text{, where} \tag{90}$$

$$\nabla_{ab} \hat{q}_{2i} = \nabla_{ab} q_2(\breve{\delta}, \breve{\gamma}, Y_i, X_i, W_i)\Big|_{\breve{\delta}=a,\, \breve{\gamma}=b} .$$

These substitutions yield

$$\widehat{\text{AVAR}}(\hat{\theta}) = \begin{bmatrix} \hat{D}_{11} & \hat{D}_{12} \\ \hat{D}'_{12} & \hat{D}_{22} \end{bmatrix} \tag{91}$$

where $\hat{D}_{11}$, $\hat{D}_{12}$ and $\hat{D}_{22}$ are the sample analogs to their population counterparts. Under general conditions $\hat{\theta}$ is asymptotically normal, i.e.

$$\widehat{\text{AVAR}}(\hat{\theta})^{-\frac{1}{2}} \sqrt{n}\,(\hat{\theta} - \theta) \overset{d}{\rightarrow} N(0, I). \tag{92}$$

where $\widehat{\text{AVAR}}(\hat{\theta})$ is the consistent estimator of the asymptotic variance-covariance matrix (AVAR) of $\hat{\theta} = [\hat{\delta}' \quad \hat{\gamma}']$.[24] Therefore, the "t-statistic"

$$\frac{\sqrt{n}(\hat{\theta}_k - \theta_k)}{\sqrt{\widehat{\text{AVAR}}(\hat{\theta})_k}} \tag{93}$$

for the kth element of θ is asymptotically standard normally distributed [where $\widehat{\text{AVAR}}(\hat{\theta})_k$ denotes the kth diagonal element of $\widehat{\text{AVAR}}(\hat{\theta})$] and can be used to test the hypothesis that $\theta_k = \theta_k^0$ for θ_k^0, a given null value of θ_k. In practice, the following version of (93) is typically used

$$\frac{\hat{\theta}_k - \theta_k^0}{\sqrt{\widehat{\text{AVAR}}(\hat{\theta})_k^{\dagger}}} \tag{94}$$

[24] See Terza (2016a, 2024-a-b) for formular details of $\widehat{\text{AVAR}}(\hat{\theta})$ and its Stata 18© implementation.

where $\widehat{AVAR}(\hat{\theta})_k^{\dagger}$ denotes the kth diagonal element of $\widehat{AVAR}(\hat{\theta})^{\dagger}$ which is the same as $\widehat{AVAR}(\hat{\theta})$ divided by n. We will henceforth refer to the square roots of the diagonal elements of $\widehat{AVAR}(\hat{\theta})^{\dagger}$ as, the ASEs of $\hat{\theta}$.

Turning now to inference regarding $AIE_{2SRI}(X^{pre}(\omega), \Delta(\omega))$ as defined in (70), we note that as a special case of the formulation offered by Terza (2016a, 2016b, 2017, 2023a, 2023b), the asymptotic variance of the sample analog estimator (74) is

$$\text{avar}(\widehat{AIE}_{2SRI}) = E[\nabla_{\theta} aie_{2SRI}(\theta)] AVAR(\hat{\theta}) E[\nabla_{\theta} aie_{2SRI}(\theta)]' + E\left[(aie_{2SRI}(\theta)] - AIE_{2SRI})^2\right] \tag{95}$$

where AIE_{2SRI} and $\widehat{AIE}_{2SRI}$ are shorthand notation for (70) and (74), respectively; ∇_{θ} denotes the gradient with respect to θ; $aie_{2SRI)}(\theta)$ is defined in (71), and $AVAR(\hat{\theta})$ denotes the asymptotic covariance matrix of $\hat{\theta}$ as defined in (83). The consistent sample analog estimators of $E[\nabla_{\theta} aie_{2SRI}(\theta)]$ and $E\left[(aie_{2SRI}(\theta)] - AIE_{2SRI})^2\right]$ are, respectively.

$$\hat{E}[\nabla_{\theta} aie_{2SRI}(\theta)] = \frac{\sum_{i=1}^{n} \nabla_{\theta} aie_{2SRI(i)}(\hat{\theta})}{n} \tag{96}$$

and

$$\hat{E}\left[(aie_{2SRI}(\theta)] - AIE_{2SRI})^2\right] = \frac{\sum_{i=1}^{n} (aie_{2SRI(i)}(\hat{\theta})] - \widehat{AIE}_{2SRI})^2}{n} \tag{97}$$

where $\nabla_\theta \text{aie}_{2SRI(i)}(\hat{\theta})$ denotes $\nabla_\theta \text{aie}_{2SRI}(\theta)$ evaluated at X_i, W_i, X_i^{pre}, Δ_i and $\hat{\theta}$.[25] The consistent sample analog estimator of (95) then is

$$\widehat{\text{avar}}(\widehat{\text{AIE}}_{2SRI}) = \hat{E}[\nabla_\theta \text{aie}_{2SRI}(\theta)]\ \widehat{\text{AVAR}}(\hat{\theta})\ \hat{E}[\nabla_\theta \text{aie}_{2SRI}(\theta)]' + \hat{E}\left[(\text{aie}_{2SRI}(\theta)] - \text{AIE}_{2SRI})^2\right] \quad (98)$$

where $\widehat{\text{AVAR}}(\hat{\theta})$ is the estimated AVAR of $\hat{\theta}$ as given in (91). Under general conditions, $\widehat{\text{AIE}}_{2SRI}$ is consistent for AIE_{2SRI} and is asymptotically normal, i.e.

$$\sqrt{\frac{n}{\widehat{\text{avar}}\left(\widehat{\text{AIE}}_{2SRI}\right)}}\left(\widehat{\text{AIE}}_{2SRI} - \text{AIE}_{2SRI}\right) \overset{d}{\to} n(0, 1).$$

In other words, the "t-statistic"

$$\frac{\widehat{\text{AIE}}_{2SRI} - \text{AIE}_{2SRI}}{\sqrt{\widehat{\text{avar}}(\widehat{\text{AIE}}_{2SRI})/n}}$$

is asymptotically standard normal distributed and can be used to test the hypothesis that $\text{AIE}_{2SRI} = \text{AIE}_{2SRI}^0$, where AIE_{2SRI}^0 is a given null value of AIE_{2SRI}.

[25] Note that, although (74) could be calculated using the **`margins`** command, it will be more convenient to obtain this estimate as a by-product of evaluating $\hat{E}[\nabla_\theta \text{aie}_{2SRI}(\theta)]$ via the Mata **`deriv`** function.

4.4 Two Stage Residual Inclusion (2SRI) in the Count Regression Model (CRM) Context with Continuous X

We have detailed the generic aspects of the count regression model potential outcomes framework with unobservable confounding. Now let us consider FO-2SRI estimation for four combinations of residual specifications (GG and linear) and alternative CRM-CPOM specifications (Poisson, CMP). As a baseline for comparison, we include the conventional linear instrumental variables (LIV) estimator in which both the residual (implicitly) and the CPOM have first-order linear forms so that all nonlinearity inducing aspects of the CRM are ignored. For the remainder of the discussion, we assume that X is continuous and conditional on W, either follows a generalized gamma (GG) probability law or has a mean that is linear in W.

4.4.1 Generalized Gamma Distributed Unobservables: First-Order (FO) Residuals

Following Manning et al. (2005), we write the GG pdf of (X | W) as

$$gg_{(X \mid W)}(X, W, \delta) = \frac{\nu^{\nu}}{\sigma X \sqrt{\nu}\,\Gamma(\nu)} \exp(z\sqrt{\nu} - u) \tag{99}$$

where $\delta' = [\alpha' \quad \sigma \quad \kappa]$, $\nu = |\kappa|^{-2}$, $z = \frac{\text{sign}(\kappa)[\log(X) - \mu]}{\sigma}$, $u = \nu \times \exp(|\kappa| z)$, and $\Gamma(\ )$ is the gamma function. This flexible three parameter (μ [location], σ [scale] and κ [shape]) distribution subsumes several popular distributions that are commonly used for non-negative random variables, such as the Weibull, exponential and log-normal among others. In this case, in the context of (67) the relevant version of $r(W; \delta)$ is

$$r_{GG}(W; \delta) = E[X \mid W] =$$

$$\exp\{W\alpha + (\sigma/\kappa) \times \ln(\kappa^2) + \ln[\Gamma(1/\kappa^2) + (\sigma/\kappa)] - \ln[\Gamma(1/\kappa^2)]\} \tag{100}$$

where $\delta' = [\alpha' \quad \sigma \quad \kappa]$, so that the relevant version of (69) is

$$\mathcal{X}^{FO}_{uGG}(X, W; \delta) =$$

$$X - \exp\{W\alpha + (\sigma/\kappa) \times \ln(\kappa^2) + \ln[\Gamma((1/\kappa^2) + (\sigma/\kappa))] - \ln[\Gamma(1/\kappa^2)]\} \tag{101}$$

4.4.2 Generalized Gamma First-Order (FO) Residual and Poisson CPOM (GG-Poisson)

The Poisson pmf version of the CPOM in (55) is

$$f_{(Y_{X^*} \mid X_o, X_u)}(Y_{X^*}, X_o, X_u, X^*; \gamma) = POI(Y_{X^*}, \lambda^*) = \frac{(\lambda^*)^{Y_{X^*}} \exp(-\lambda^*)}{Y_{X^*}!} \tag{102}$$

where $Y_{X^*} = 0, 1, 2, 3, ...$; $\gamma' = [\gamma_o' \quad \gamma_u \quad \gamma_x]$ and $\lambda^* = \exp(X_o\beta_o + X_u\beta_u + X^*\beta_x)$. If (100) holds, we get that the relevant version of (59) is

$$E[Y_{X^*} \mid X, W] = \exp(X_o\gamma_o + \mathcal{X}^{FO}_{uGG}(X, W; \delta)\gamma_u + X^*\gamma_x) \tag{103}$$

where $\mathcal{X}^{FO}_{uGG}(X, W; \delta)$ is defined as in (101). It follows that the relevant version of (61) is

$$AIE_{UC(POI)}(X^{pre}(\omega), \Delta(\omega)) = AVG_{\omega \in \Omega}\{E[aie_{UC(POI)}(\theta, \omega)]\} \tag{104}$$

where

$$\text{aie}_{UC(POI)}(\theta, \omega) =$$

$$E[\exp(X_o\gamma_o + \mathcal{X}^{FO}_{uGG}(X, W; \delta)\gamma_u + (X^{pre} + \Delta)\gamma_x)$$

$$- \exp(X_o\gamma_o + \mathcal{X}^{FO}_{uGG}(X, W; \delta)\gamma_u + X^{pre}\gamma_x)]. \tag{105}$$

Given (105), it is clear that the AIE (104) will be identified if θ is identified. If (102), CI (64) and COI (65) hold, we get the following version of (66)

$$f_{(Y \mid X, W)}(Y, X, W; \delta, \gamma) = POI(Y, \lambda) = \frac{\lambda^Y \exp(-\lambda)}{Y!} \tag{106}$$

where

$$\lambda = \exp(X_o\gamma_o + \mathcal{X}^{FO}_{uGG}(X, W; \delta)\gamma_u + X\gamma_x).$$

To estimate the deep parameters of the model $\theta = [\delta' \quad \gamma'])$, we use the following version of the general 2SRI protocol discussed above.

First Stage:

Consistently estimate δ using full information maximum likelihood estimation based on the GG distribution (99). Then calculate the first-order residuals as

$$\mathcal{X}^{FO}_{uGG}(X_I, W_I; \hat{\delta}) =$$

$$X_i - \exp\left\{W_i\hat{\alpha} + (\hat{\sigma}/\hat{\kappa}) \times \ln(\hat{\kappa}^2) + \ln[\Gamma(1/\hat{\kappa}^2) + (\hat{\sigma}/\hat{\kappa})] - \ln[\Gamma(1/\hat{\kappa}^2)]\right\} \tag{107}$$

where $\hat{\delta}' = [\hat{\alpha}' \quad \hat{\sigma} \quad \hat{\kappa}]$ is the first-stage consistent estimate of δ and the subscript $i = 1, \ldots, n$ denotes the ith sample member. This can be accomplished using the Stata `streg` command and postestimation `predict` command with the `mean time` option.

<u>Second Stage</u>:

Consistently estimate γ as the maximizer of

$$\ln[\mathrm{POI}(Y_i, \lambda_i)] \tag{108}$$

with respect to $\breve{\gamma}$, where

$$\mathrm{POI}(Y_i, \breve{\lambda}_i) = \frac{(\breve{\lambda}_i)^{Y_i}\exp(-\breve{\lambda}_i)}{Y_i!} \tag{109}$$

and

$$\breve{\lambda}_i = \exp(X_{oi}\breve{\gamma}_o + \mathcal{X}_{uGG}^{FO}(X_i, W_i; \hat{\delta})\breve{\gamma}_u + X_i\breve{\gamma}_x).$$

Next, consistently estimate the AVAR of $\hat{\theta} = [\hat{\delta}' \quad \hat{\gamma}']$ using $\widehat{\mathrm{AVAR}}_{UC(POI)}(\hat{\theta})$ as in (91). As discussed earlier, inference regarding the elements of θ can be conducted using $\widehat{\mathrm{AVAR}}_{UC(POI)}(\hat{\theta})$ based on the asymptotic normality of $\hat{\theta}$. With $\hat{\theta}$ in hand, we estimate $\mathrm{AIE}_{UC(POI)}(X^{pre}(\omega), \Delta(\omega))$ in (104) using its sample analog. In this case the sample analog estimator of (104) is the following version of (63)

$$\widehat{\mathrm{AIE}}_{UC(POI)}(X^{pre}, \Delta) = \frac{\sum_{i=1}^{n} \mathrm{aie}_{UC(POI)(i)}(\hat{\theta})}{n} \tag{110}$$

where

$$\text{aie}_{\text{UC(POI)(i)}}(\hat{\theta}) = \exp(\,X_{oi}\hat{\gamma}_o + \mathcal{X}^{\text{FO}}_{\text{uGG}}(X_i, W_i; \hat{\delta})\hat{\gamma}_u + [X^{\text{pre}}_i + \Delta_i]\hat{\gamma}_x) - \exp(\,X_{oi}\hat{\gamma}_o + \mathcal{X}^{\text{FO}}_{\text{uGG}}(X_i, W_i; \hat{\delta})\hat{\gamma}_u + X^{\text{pre}}_i\hat{\gamma}_x). \tag{111}$$

We also have that

$$\frac{\widehat{\text{AIE}}_{\text{UC(POI)}} - \text{AIE}_{\text{UC(POI)}}}{\sqrt{\widehat{\text{avar}}(\widehat{\text{AIE}}_{\text{UC(POI)}})/n}} \xrightarrow{d} n(0, 1). \tag{112}$$

where

$$\widehat{\text{avar}}\left(\widehat{\text{AIE}}_{\text{UC(POI)}}\right) = \hat{\Psi}_{\text{UC(POI)}}\, \widehat{\text{AVAR}}_{\text{UC(POI)}}(\hat{\theta})\, \hat{\Psi}_{\text{UC(POI)}}' + \hat{\Lambda}_{\text{UC(POI)}} \tag{113}$$

$$\hat{\Psi}_{\text{UC(POI)}} = \frac{\sum_{i=1}^{n} \nabla_{\gamma}\text{aie}_{\text{UC(POI)(i)}}(\hat{\theta})}{n} \tag{114}$$

$$\hat{\Lambda}_{\text{POI}} = \frac{\sum_{i=1}^{n}\left(\text{aie}_{\text{UC(POI)(i)}}(\hat{\theta}) - \widehat{\text{AIE}}\right)^2}{n} \tag{115}$$

$\text{aie}_{\text{UC(POI)(i)}}(\hat{\theta})$ is defined as in (111), and $\widehat{\text{AVAR}}_{\text{UC(POI)}}(\hat{\theta})$ is the estimated AVAR of $\hat{\theta}$. We then consistently estimate the asymptotic variance of $\widehat{\text{AIE}}_{\text{UC(POI)}}(X^{\text{pre}}, \Delta)$ as in (113) and, therewith [based on the asymptotic normality of $\widehat{\text{AIE}}_{\text{UC(POI)}}(X^{\text{pre}}, \Delta)$], conduct inference regarding $\text{AIE}_{\text{UC(POI)}}(X^{\text{pre}}(\omega), \Delta(\omega))$.

4.4.3 Generalized Gamma First-Order (FO) Residual and Conway-Maxwell-Poisson CPOM (GG-CMP)

As a more dispersion-flexible alternative to the Poisson, we consider the following version of the generic CRM-CPOM in (55)

$$f_{(Y_{X^*} \mid X_{o}, X_{u})}(Y_{X^*}, X_o, X_u, X^*; \gamma) = \text{CMP}(Y_{X^*}; \lambda^*, \psi) = \frac{(\lambda^*)^{Y_{X^*}}}{Y_{X^*}!^{\exp(\psi)}\, Z(\lambda^*, \exp(\psi))} \tag{116}$$

where $Y_{X^*} = 0, 1, 2, 3, \ldots; \gamma' = [\beta_o' \quad \beta_u \quad \beta_x \quad \psi]$, $\lambda^* = \exp(X_o\beta_o + X_u\beta_u + X^*\beta_x)$,

$$Z(\lambda^*, \exp(\psi)) = \sum_{j=0}^{\infty} \frac{(\lambda^*)^j}{(j!)^{\exp(\psi)}},$$

$-\infty < \psi = \ln(\upsilon) < \infty$, and $\upsilon > 0$ is the dispersion parameter. Equi-dispersion corresponds to the case when $\upsilon = 1$ (i.e., $\psi = 0$), over-dispersion prevails when $\upsilon < 1$ (i.e., $\psi < 0$), and under-dispersion holds if $\upsilon > 1$ (i.e., $\psi > 0$). Given (116), we get that the relevant version of (59) is

$$E[Y_{X^*} \mid X, W] = m_{\text{CMP}}(\lambda^*, \psi) = \lambda^* \sum_{j=1}^{\infty} \frac{j(\lambda^*)^{j-1}}{(j!)^{\exp(\psi)} Z(\lambda^*, \exp(\psi))} \tag{117}$$

wherein, to be explicit, we rewrite λ^* as

$$\lambda^* = \exp(X_o\beta_o + x_{uGG}^{FO}(X, W; \delta)\beta_u + X^*\beta_x).$$

It follows that the relevant version of (61) is

$$\mathrm{AIE}_{\mathrm{UC(CMP)}}(\mathrm{X}^{\mathrm{pre}}(\omega), \Delta(\omega)) = \mathrm{AVG}_{\omega \in \Omega}\{\mathrm{E}[\mathrm{aie}_{\mathrm{UC(CMP)}}(\theta, \omega)]\} \tag{118}$$

where

$$\mathrm{Aie}_{\mathrm{UC(CMP)}}(\theta, \omega) =$$

$$\mathrm{E}[m_{\mathrm{CMP}}(\exp(\mathrm{X_o}\beta_o + x^{\mathrm{FO}}_{\mathrm{uGG}}(\mathrm{X}, \mathrm{W}; \delta)\beta_u + (\mathrm{X}^{\mathrm{pre}} + \Delta)\beta_x), \psi)$$

$$- m_{\mathrm{CMP}}(\exp(\mathrm{X_o}\beta_o + x^{\mathrm{FO}}_{\mathrm{uGG}}(\mathrm{X}, \mathrm{W}; \delta)\beta_u + \mathrm{X}^{\mathrm{pre}}\beta_x), \psi)\ ]. \tag{119}$$

If (116), CI (64) and COI (66) hold, we get

$$f_{(\mathrm{Y}\,|\,\mathrm{X},\,\mathrm{W})}(\mathrm{Y}, \mathrm{X}, \mathrm{W}; \delta, \gamma) = \mathrm{CMP}(\mathrm{Y}, \lambda, \psi) = \frac{(\lambda)^{\mathrm{Y}}}{\mathrm{Y}!^{\exp(\psi)}\, \mathrm{Z}(\lambda, \exp(\psi))} \tag{120}$$

where

$$\lambda = \exp(\mathrm{X_o}\beta_o + x^{\mathrm{FO}}_{\mathrm{uGG}}(\mathrm{X}, \mathrm{W}; \delta)\beta_u + \mathrm{X}\beta_x).$$

To estimate the deep parameters of the model $\theta = [\delta' \quad \gamma']$), we use the following version of the general 2SRI protocol discussed above.

First Stage:

This stage is the same as for the Poisson version of the CPOM.

Second Stage:

Consistently estimate γ as the maximizer of

$$\ln[\mathrm{CMP}(\mathrm{Y}_i, \lambda_i)] \tag{121}$$

with respect to $\breve{\gamma}$, where

$$\mathrm{CMP}(Y_i, \breve{\lambda}_i) = \frac{(\breve{\lambda}_i)^{Y_i}}{Y_i!^{\exp(\psi)}\, Z(\breve{\lambda}_i, \exp(\psi))} \tag{122}$$

And

$$\breve{\lambda}_i = \exp(X_{oi}\breve{\gamma}_o + \mathcal{X}^{FO}_{uGG}(X_i, W_i; \hat{\delta})\breve{\gamma}_u + X_i\breve{\gamma}_x).$$

Next, consistently estimate the AVAR of $\hat{\theta} = [\hat{\delta}' \quad \hat{\gamma}']$ using $\widehat{AVAR}_{UC(CMP)}(\hat{\theta})$ as in (91). As discussed earlier, inference regarding the elements of θ can be conducted using $\widehat{AVAR}_{UC(CMP)}(\hat{\theta})$ based on the asymptotic normality of $\hat{\theta}$. With $\hat{\theta}$ in hand, we estimate $AIE_{UC(CMP)}(X^{pre}(\omega), \Delta(\omega))$ in (118) using its sample analog. In this case the sample analog estimator of (118) is the following version of (63)

$$\widehat{AIE}_{UC(CMP)}(X^{pre}, \Delta) = \frac{\sum_{i=1}^{n} aie_{UC(CMP)(i)}(\hat{\theta})}{n} \tag{123}$$

where

$$aie_{UC(CMP)(i)}(\hat{\theta}) =$$

$$m_{CMP}(\exp(X_{oi}\hat{\beta}_o + \mathcal{X}^{FO}_{uGG}(X_i, W_i; \hat{\delta})\hat{\beta}_u + [X_i^{pre} + \Delta_i]\hat{\beta}_x), \hat{\psi})$$

$$- m_{CMP}(\exp(X_{oi}\hat{\beta}_o + \mathcal{X}^{FO}_{uGG}(X_i, W_i; \hat{\delta})\hat{\beta}_u + X_i^{pre}\hat{\beta}_x), \hat{\psi}). \tag{124}$$

We also have that

$$\frac{\widehat{AIE}_{UC(CMP)} - AIE_{UC(CMP)}}{\sqrt{\widehat{avar}(\widehat{AIE}_{UC(CMP)})/n}} \overset{d}{\rightarrow} n(0, 1). \tag{125}$$

where

$$\widehat{avar}\left(\widehat{AIE}_{UC(CMP)}\right) = \widehat{\Psi}_{UC(CMP)}\ \widehat{AVAR}_{UC(CMP)}(\hat{\theta})\ \widehat{\Psi}_{UC(CMP)}{}' + \widehat{\Lambda}_{UC(CMP)} \tag{126}$$

$$\widehat{\Psi}_{UC(CMP)} = \frac{\sum_{i=1}^{n} \nabla_{\gamma} aie_{UC(CMP)(i)}(\hat{\theta})}{n} \tag{127}$$

$$\widehat{\Lambda}_{CMP} = \frac{\sum_{i=1}^{n} \left(aie_{UC(CMP)(i)}(\hat{\theta}) - \widehat{AIE}\right)^2}{n} \tag{128}$$

$aie_{UC(CMP)(i)}(\hat{\theta})$ is defined as in (124), and $\widehat{AVAR}_{UC(CMP)}(\hat{\theta})$ is the estimated AVAR of $\hat{\theta}$. We then consistently estimate the asymptotic variance of $\widehat{AIE}_{UC(CMP)}(X^{pre}, \Delta)$ as in (126) and therewith [based on the asymptotic normality of $\widehat{AIE}_{UC(CMP)}(X^{pre}, \Delta)$] conduct inference regarding $AIE_{UC(CMP)}(X^{pre}(\omega), \Delta(\omega))$.

4.4.4 Linear in Mean Unobservables: First-Order (FO) Residuals

Here we consider the case in which X is continuous and has a mean that is linear in W and assume that, in the context of (67), we have

$$r_{LIN}(W; \delta) = W\delta \tag{129}$$

so that the relevant version of the first order residual in (69) is

$$x_{uLIN}^{FO}(X, W; \delta) = X - W\delta. \tag{130}$$

4.4.5 Linear First-Order (FO) Residual and Poisson CPOM (LIN-Poisson)

Here the Poisson pmf version of the CRM-CPOM is as specified in (102), and if (55) holds, we get that the relevant version of (59) is

$$E[Y_{X^*} \mid X, W] = \exp(X_o\gamma_o + \mathcal{X}^{FO}_{uLIN}(X, W; \delta)\gamma_u + X^*\gamma_x). \qquad (131)$$

where $\mathcal{X}^{FO}_{uLIN}(X, W; \delta)$ is defined as in (130). It follows that the relevant version of (61) is the same as (104) with $\mathcal{X}^{FO}_{uGG}(X, W; \delta)$ replaced by $\mathcal{X}^{FO}_{uLIN}(X, W; \delta)$. The details of the asymptotic properties of the AIE estimator are analogous to those given in the discussion surrounding equations (112) through (115). If (55), CI (64) and COI (65) hold, we get the following version of (66)

$$f_{(Y \mid X, W)}(Y, X, W; \delta, \gamma) = POI(Y, \lambda) = \frac{\lambda^Y \exp(-\lambda)}{Y!} \qquad (132)$$

where

$$\lambda = \exp(X_o\gamma_o + \mathcal{X}^{FO}_{uLIN}(X, W; \delta)\gamma_u + X\gamma_x).$$

To estimate the deep parameters of the model $\theta = [\delta' \quad \gamma'])$, we use the following version of the general 2SRI protocol discussed above.

<u>First Stage</u>:

Consistently estimate δ using the ordinary least squares (OLS) estimator based on (129). Then calculate the first-order residuals as

$$\mathcal{X}^{FO}_{uLIN}(X_i, W_i; \hat{\delta}) = X_i - W_i\hat{\delta} \qquad (133)$$

where $\hat{\delta}$ is the OLS estimate of δ and the subscript i = 1, . . ., n denotes the ith sample member. This can be accomplished using the Stata `regress` command and postestimation `predict` command with the `r` option.

Second Stage:

Consistently estimate γ as the maximizer of

$$\ln[\mathrm{POI}(Y_i, \lambda_i)] \tag{134}$$

with respect to $\breve{\gamma}$, where

$$\mathrm{POI}(Y_i, \lambda_i) = \frac{(\lambda_i)^{Y_i}\exp(-\lambda_i)}{Y_i!} \tag{135}$$

and

$$\lambda_i = \exp(X_{oi}\breve{\gamma}_o + x^{FO}_{uLIN}(X_i, W_i; \hat{\delta})\breve{\gamma}_u + X_i\breve{\gamma}_x).$$

Next, consistently estimate the AVAR of $\hat{\theta} = [\hat{\delta}' \quad \hat{\gamma}']$ and use the result and the asymptotic normality of $\hat{\theta}$ to draw inferences regarding the elements of θ. With $\hat{\theta}$ in hand, estimate the AIE and its asymptotic variance and, therewith [based on the asymptotic normality of the AIE estimator], conduct causal inference.

4.4.6 Linear First-Order (FO) Residual and Conway-Maxwell-Poisson CPOM (LIN-CMP)

Here the CMP pmf version of the CRM-CPOM is as specified in (116), and if (55) holds, we get that the relevant version of (59) is

$$E[Y_{X^*} \mid X, W] = m_{CMP}(\lambda^*, \psi) = \lambda^* \sum_{j=1}^{\infty} \frac{j(\lambda^*)^{j-1}}{(j!)^{\exp(\psi)} Z(\lambda^*, \exp(\psi))} \tag{136}$$

where

$$\lambda^* = \exp(X_o\beta_o + x^{FO}_{uLIN}(X, W; \delta)\beta_u + X^*\beta_x). \tag{137}$$

and $x^{FO}_{uLIN}(X, W; \delta)$ is defined as in (130). It follows that the relevant version of (61) is the same as (118) with $x^{FO}_{uGG}(X, W; \delta)$ replaced by $x^{FO}_{uLIN}(X, W; \delta)$. The details of the asymptotic properties of the AIE estimator are analogous to those given in the discussion surrounding equations (125) through (128). If (55), CI (64) and COI (65) hold, we get the following version of (116)

$$f_{(Y \mid X, W)}(Y, X, W; \delta, \gamma) = CMP(Y, \lambda, \psi) = \frac{(\lambda)^Y}{Y!^{\exp(\psi)}\, Z(\lambda, \exp(\psi))} \tag{138}$$

where

$$\lambda = \exp(X_o\gamma_o + x^{FO}_{uLIN}(X, W; \delta)\gamma_u + X\gamma_x).$$

To estimate the deep parameters of the model $\theta = [\delta' \quad \gamma'])$, we use the following version of the general 2SRI protocol discussed above.

<u>First Stage</u>:

Consistently estimate δ using the ordinary least squares (OLS) estimator based on (129). Then calculate the first-order residuals as

$$\mathcal{X}^{FO}_{uLIN}(X_i, W_i; \hat{\delta}) = X_i - W_i\hat{\delta} \qquad (139)$$

where $\hat{\delta}$ is the OLS estimate of δ and the subscript i = 1, . . ., n denotes the ith sample member. This can be accomplished using the Stata **`regress`** command and postestimation **`predict`** command with the **`r`** option.

<u>Second Stage</u>:

Consistently estimate γ as the maximizer of

$$\ln[CMP(Y_i, \lambda_i, \psi)] \qquad (140)$$

with respect to $\breve{\gamma}' = [\breve{\beta}_o' \quad \breve{\beta}_u \quad \breve{\beta}_x \quad \breve{\psi}]$, where

$$CMP(Y_i, \lambda_i) = \frac{(\lambda_i)^{Y_i}}{Y_i!^{\exp(\psi)}\, Z(\lambda_i, \exp(\psi))} \qquad (141)$$

and

$$\lambda_i = \exp(X_{oi}\breve{\beta}_o + \mathcal{X}^{FO}_{uLIN}(X_i, W_i; \hat{\delta})\breve{\beta}_u + X_i\breve{\beta}_x).$$

Next, consistently estimate the AVAR of $\hat{\theta} = [\hat{\delta}' \quad \hat{\gamma}']$ and use the result and the asymptotic normality of $\hat{\theta}$ to draw inferences regarding the elements of θ. With $\hat{\theta}$ in hand, estimate the AIE and its asymptotic variance and, therewith [based on the asymptotic normality of the AIE estimator], conduct causal inference.

4.4.7 Linear First-Order (FO) Residual and Linear CPOM (LIV)

In this case, the relevant CPOM is first-order (i.e., based solely on the relevant conditional mean) rather than fully parametric (i.e., full-order) as in (55). Here we have

$$E[Y_{X^*} \mid X, W] = X_o\gamma_o + X_u\gamma_u + X^*\gamma_x \tag{142}$$

where Y_{X^*} can be any value on the real line and $\gamma' = [\gamma_o' \quad \gamma_u \quad \gamma_x]$. It follows that the relevant version of (3) is

$$AIE_{UC(LIN)}(X^{pre}(\omega), \Delta(\omega)) = AVG_{\omega \in \Omega}\{E[aie_{UC(LIN)}(\theta, \omega)]\} \tag{143}$$

where

$$aie_{UC(LIN)}(\theta, \omega) = AVG_{\omega \in \Omega}\{(X_o\gamma_o + X_{uGG}\gamma_u + (X^{pre}(\omega) + \Delta(\omega))\gamma_x) - (X_o\gamma_o + X_{uGG}\gamma_u + X^{pre}(\omega)\gamma_x)\} = \Delta\hat{\gamma}_x. \tag{144}$$

In this case the sample analog estimator of (143) is

$$\widehat{AIE}_{UC(LIN)}(X^{pre}, \Delta) = \frac{\sum_{i=1}^{n} aie_{UC(LIN)(i)}(\hat{\theta})}{n} \tag{145}$$

where

$$aie_{UC(LIN)(i)}(\hat{\theta}) = (X_{oi}\hat{\gamma}_o + X_{uGG}\hat{\gamma}_u + [X_i^{pre} + \Delta_i]\hat{\gamma}_x) - (X_{oi}\hat{\gamma}_o + X_{uGG}\hat{\gamma}_u + X_i^{pre}\hat{\gamma}_x) = \Delta_i\hat{\gamma}_x. \tag{146}$$

Therefore

$$\widehat{AIE}_{UC(LIN)}(X^{pre}, \Delta) = \left(\frac{\sum_{i=1}^{n} \Delta_i}{n}\right)\hat{\gamma}_x \tag{147}$$

To estimate the deep parameters of the model $\theta = [\delta' \quad \gamma'])$, we can use the conventional linear instrumental variables [LIV] estimator.

Chapter 5

Assessing and Comparing the Estimators: Simulation Study

5.1 Overview

In this chapter, we investigate the properties of the estimators presented in chapters 3 and 4 using a series of simulation studies aimed at validating and comparing their accuracy in estimating the relevant AIE. We start with the case with no unobservable confounding (X is exogenous) and examine how the FIML-CMP, FIML-Poisson and OLS methods perform for estimating the value of the AIE compared to the true value of the AIE which is calculated based on the sample design. Next, we investigate the case where unobservable confounding is present (X is endogenous) and report the performance of the FO-2SRI estimators (GG-CMP, GG-Poisson, LIN-CMP, LIN-Poisson, and LIV) comparing the estimated values of AIE with the true value of the AIE calculated based on the sample design. In the sampling design for the endogenous X case, we consider two different approaches for generating the residuals. The performance of the estimators are investigated at different levels of dispersion in each case.

5.2 Exogenous X

Following the discussion in chapter 3, here we consider the case in which X is exogenous (not subject to UC). We run a simulation study to validate and compare the consistency properties of the AIE estimators.

5.2.1 Data Generation Protocol – No Unobservable Confounding (UC)

We assess and compare the bias properties of the three methods discussed in chapter 3 (FIML-Poisson, FIML-CMP, and OLS) using simulated data. We generated the values of $\mathcal{X}_{oj}$, $\mathcal{X}_{j}$ and $\mathcal{Y}_{j}$, the simulated values of X_j, X_{oj}, Y_j, respectively, for each of 200 samples of size 150,000 (j = 1, …, 150, 000) using specified values for the elements of the CMP parameter vector (say, $\bar{\gamma}' = [\bar{\beta}_o' \quad \bar{\beta}_x \quad \bar{\psi}]$), the generated sample values for $\mathcal{X}_{oj}$ and $\mathcal{X}_{j}$, and a generated purely random component. We simulated the sample of values of $\mathcal{Y}_{j}$ under the assumption that it is CMP distributed as in (40).[26]

As the estimation objective for the simulations, we chose the version of the AIE in (10) implied by the following counterfactual [a version of (1)]

$$[X^{pre}(\omega) = X(\omega), \Delta(\omega) = -X(\omega)] \tag{148}$$

where X(ω) denotes the deterministic function underlying the observable random variable X.[27] In this case, the two relevant counterfactual scenarios are:

pre- counterfactual scenario: each individual (ω) in the relevant population Ω is mandated a value of the $\mathbf{X}$ that is equal to their observable value $[X^{pre}(\omega) = X(\omega)]$

post- counterfactual scenario: each individual is mandated a value of 0 for the $\mathbf{X}$ $[\Delta(\omega) = -X(\omega)$ so that $X^{post}(\omega) = X^{pre}(\omega) + \Delta(\omega) = X(\omega) - X(\omega) = 0]$

---

[26] An appendix detailing the numerical aspects of the sampling designs and the corresponding Stata/Mata implementation code will be supplied upon request.

[27] $X(\omega): \Omega \longrightarrow S \subset \mathbb{R}$, where Ω is the primitive sample space for the random variable X, $\omega \in \Omega$, and S is the support of the random variable X. X(ω) is a deterministic function in the sense that it is (can be) defined without regard to the probability measures or σ-algebras defined on Ω and S.

and the relevant counterfactual query is

"Starting with a counterfactually imposed scenario in which the value of the **X** for each member of the relevant population is equal to their observable value at the time of the data survey, suppose we counterfactually imposed a change in which the value of the **X** for each member of the relevant population is set equal to 0, what would be the consequent causal effect on the **Y**?"

This type of counterfactual is useful in a variety of empirical causal analysis (ECA) contexts. As an example, consider an ECA motivated by a prospective intervention program aimed at reducing smoking during pregnancy to zero. It is also noteworthy that $\Delta(\omega)$ in (148) is not constant across members of the relevant population.

Under the above data simulation protocol, the counterfactual in (148) we define and calculate the "TRUE" values of the AIE in (38) as

$$AIE_{TRUE}(X^{pre}(\omega) = X(\omega), \Delta(\omega) = -X(\omega)) = AVG_{\omega \in \Omega}\{E[aie_{TRUE}(\bar{\theta}, \omega)]\} \tag{149}$$

where

$$aie_{TRUE}(\theta, \omega) = E[\mathcal{m}_{CMP}(\exp(\mathcal{X}_o\bar{\beta}_o), \bar{\psi}) - \mathcal{m}_{CMP}(\exp(\mathcal{X}_o\bar{\beta}_o + X(\omega)\bar{\beta}_x), \bar{\psi})]$$

$\mathcal{X}_o$ denotes the pseudo-random variables from which $\mathcal{X}_{oj}$ is drawn

and $\mathcal{m}_{CMP}(.\ ,.\ )$ is defined as in (37). For a particular sampling design specification of the CMP deep parameter vector $\bar{\gamma}' = [\bar{\beta}_o' \quad \bar{\beta}_x \quad \bar{\psi}]$ we approximated the "TRUE" value of the AIE in (149) as

$$\text{AIE}_{\text{TRUE}} = \frac{\sum_{j*=1}^{n(\infty)} \text{aie}_{\text{TRUE}(j*)}}{n(\infty)} \tag{150}$$

where j* denotes the j*th simulated observation in a "super sample" of size n(∞) = 1.5M,

$$\text{aie}_{\text{TRUE}(j*)} = m_{\text{CMP}}(\exp(\mathcal{X}_{oj*}\bar{\beta}_o), \bar{\psi}) - m_{\text{CMP}}(\exp(\mathcal{X}_{oj*}\bar{\beta}_o + \mathcal{X}_{j*}\bar{\beta}_x), \bar{\psi})\ ]$$

and j* = 1, …, n(∞).

To the kth simulated sample (k = 1,…, 200) we applied each of the three methods (EST = FIML-CMP, FIML-Poisson, and OLS) from chapter 3 and obtained a corresponding estimate of the deep parameter vector (say, $\hat{\theta}_{\text{EST(k)}}$). Using $\hat{\theta}_{\text{EST(k)}}$, we then estimated the AIE using the following version of (14)

$$\widehat{\text{AIE}}_{\text{EST(k)}} = \frac{\sum_{i=1}^{n} \text{aie}_{(i)}(\hat{\theta}_{\text{EST(k)}})}{n} \tag{151}$$

where

$$\text{aie}_{(i)}(\hat{\theta}_{\text{EST(k)}}) = m(X_{oi}, 0; \hat{\gamma}_{\text{EST(k)}}) - m(X_{oi}, X_i; \hat{\gamma}_{\text{EST(k)}})$$

and m(   ) is defined in (14).

To assess the performance of each of the three methods, we used the following cross-sample average of (151)

$$(\text{AVG } \widehat{\text{AIE}}_{\text{EST}}) = \sum_{k=1}^{200} \frac{1}{200} \widehat{\text{AIE}}_{\text{EST}}(k) \tag{152}$$

where EST = Poisson, CMP, or OLS] for the kth sample (k = 1, …, 200). We also calculated the following average absolute percentage bias statistic

$$(\text{AVG } |\% \text{ Bias}|)_{EST} = \sum_{k=1}^{200} \frac{1}{200} \left| \frac{\widehat{AIE}_{EST}(k) - AIE_{TRUE}}{AIE_{TRUE}} \right| \times 100\% . \tag{153}$$

We note that FIML-Poisson and OLS estimators are easy to apply using conventional econometric software. For this reason, it is useful to get a sense of the size of the potential bias cost that might be incurred by applied researchers seeking simplicity and convenience through the use of these estimators compared to the more complex dispersion-flexible FIML-CMP estimator. The statistics in (152) and (153) should be informative in this regard.

5.2.2 Data Generated without Unobservable Confounding (UC) – Sampling Design

Considering the simulation protocol, to compare the performance of the alternative estimators (i.e. FIML-CMP, FIML-Poisson, and OLS), the following sampling design is considered.

1) The simulated data on X and X_o are generated using uniform distribution with means and variances as in $E[X_o] = 0$, $E[X] = 0$, $Var[X_o] = 1$ and $Var[X] = 1$.

2) To generate the pseudo values for Y, the following parameter values are set

$$[\bar{\beta}_o' \quad \bar{\beta}_x \quad \bar{\psi}] = [0.5 \quad 0 \quad -0.5]$$

where $\beta_o' = [\beta_{X_o} \quad \beta_{cons}]$ are the coefficients for X_o and the intercept, respectively and the following values for the log-transformed version of the dispersion parameter are considered.

ψ = -2 , -1.75, -1.5, -1.25, -1, -0.75, -0.5, -0.25, 0, 0.25, 0.75, 1, 1.25, 1.5, 1.75, 2

5.2.3 Simulation Results – No Unobservable Confounding (UC): FIML-CMP, FIML-Poisson, and OLS

The complete results of the simulation study in the exogenous X case are displayed in Tables 1 and 2; for the average AIE statistic (152) and the average absolute percentage bias (153), respectively. We compare the amount of bias among the estimators based on what differentiates them. As detailed in chapter 3, the three proposed estimators include FIML-CMP, FIML-Poisson, and linear OLS estimators, and they differ with regard to how they account for the various sources of nonlinearity, including aspects of the CRM i.e. discreteness, boundedness, and dispersion flexibility. Therefore, we compare these estimators based on each of these margins (sources) of nonlinearity and compare the average of absolute percentage bias for the estimators.

The OLS estimator does not take account of the boundedness and discreteness of the count outcome data (FIML-CMP, FIML-Poisson, both account for these intrinsic characteristics). The OLS and FIML-Poisson estimators do not allow for dispersion-flexibility – a possibly important aspect of count data (FIML-CMP is dispersion-flexible). We compare the estimators with respect to these characteristics.

Table 1: Simulation Results for Estimated Average AIE When X is Exogenous

Dispersion ψ	**True AIE**	**FIML-CMP**	**FIML-Poisson**	**OLS**
-2	-0.561	-0.566	-1.966	-0.001
-1.75	-0.794	-0.800	-1.940	-0.001
-1.5	-1.101	-1.105	-1.849	-0.001
-1.25	-1.390	-1.390	-1.651	-0.001
-1	-1.363	-1.361	-1.309	0.000
-0.75	-0.943	-0.941	-0.856	0.000
-0.5	-0.503	-0.502	-0.460	0.000
-0.25	-0.263	-0.262	-0.249	0.000
0	-0.147	-0.147	-0.147	0.000
0.25	-0.086	-0.086	-0.092	0.000
0.5	-0.052	-0.052	-0.061	0.000
0.75	-0.032	-0.032	-0.043	0.000
1	-0.020	-0.020	-0.031	0.000
1.25	-0.012	-0.012	-0.023	0.000
1.5	-0.006	-0.006	-0.018	0.000
1.75	-0.003	-0.003	-0.015	0.000
2	-0.001	-0.001	-0.014	0.000

200 replications with 150,000 sample size.

With regard to accounting for boundedness and discreteness, the results in Table 2 indicate that at different levels of dispersion (under-dispersion, equi-dispersion, and over-dispersion), the linear OLS estimator underestimates the value of AIE by almost 100% which makes this estimator the most biased of all. Moreover, the amount of bias does not increase or decrease much with the data being more over or under-dispersed. For example, at ψ = –1.25 the true value of the AIE is –1.390 and the OLS estimated average AIE is -0.001, as compared to –1.651 by the Posson estimator; while the CMP estimate the value of the AIE is identical to the true value of the AIE. Similar results are nearly uniformly true at all other dispersion levels.

With regard to dispersion-flexibility, the FIML-CMP is the only estimator that accommodates this aspect of count data . We observe in Tables 1 and 2 that FIML-CMP produces smaller bias across all specified dispersion levels and, as the data becomes more extremely under or over dispersed, the FIML-CMP clearly dominates the other estimators. Only at ψ = 0 are the FIML-CMP and FIML-Poisson estimates are comparable, having identical average percentage bias. Another noteworthy result is that at more intensified underdispersion (ψ = –2 and ψ =–1.75) and over dispersion levels (ψ = 1.75 and ψ = 2), the FIML-Poisson estimator yields the most amount of bias, even more than the linear OLS estimator.

Table 2: Simulation Results for Average Absolute % Bias of Estimated AIE When X is Exogenous

Dispersion ψ	**True AIE**	**FIML-CMP**	**FIML-Poisson**	**OLS**
-2	-0.561	2.6%	250.5%	99.9%
-1.75	-0.794	1.8%	144.3%	99.9%
-1.5	-1.101	1.2%	67.9%	99.9%
-1.25	-1.390	0.9%	18.8%	100.0%
-1	-1.363	0.7%	3.9%	100.0%
-0.75	-0.943	0.8%	9.3%	100.0%
-0.5	-0.503	0.9%	8.7%	100.0%
-0.25	-0.263	1.0%	5.4%	100.0%
0	-0.147	1.2%	1.2%	99.9%
0.25	-0.086	1.4%	7.3%	99.9%
0.5	-0.052	1.6%	17.3%	99.9%
0.75	-0.032	1.9%	31.6%	99.9%
1	-0.020	2.3%	54.2%	99.9%
1.25	-0.012	3.1%	96.1%	99.8%
1.5	-0.006	4.7%	191.3%	99.7%
1.75	-0.003	9.4%	461.1%	99.4%
2	-0.001	27.3%	1462.5%	98.3%

200 replications with 150,000 sample size.

5.3 Endogenous X

Following the discussion in chapter 4, here we consider the case in which X is subject to UC (endogenous). We run a simulation study to validate the AIE estimators and compare their performances using two separate sample designs for creating the residuals.

In each case we first estimate the AIE ignoring the presence of UC and then estimate the AIE considering the presence of endogeneity or UC.

5.3.1 Data Generation Protocol – Implementing First-Order (FO) Residuals to Account for Unobservable Confounding (UC)

We assess and compare the bias properties of the estimators detailed in chapter 3 and 4 using simulated data under the assumption that the true data generating process is subject to UC. We generated the values of $\mathcal{W}_j^+$, $\mathcal{X}_{oj}$, $\mathcal{X}_j$ and $\mathcal{Y}_j$, the simulated values of W_j^+, X_j, X_{oj}, Y_j, respectively, for each of 200 samples of size 150,000 (j = 1, …, 150, 000) in the following two steps;

1. Using specified values for the elements of the GG parameter vector (say, $\bar{\delta}' = [\bar{\alpha}' \quad \bar{\sigma} \quad \bar{\kappa}]$), the generated sample of values for $\mathcal{W}_j = [\mathcal{W}_j^+ \quad \mathcal{X}_{oj}]$, and a generated purely random component, U_j, we simulated the sample of values of $\mathcal{X}_j$ using the inverse transform method (see Ross, 1997) under the assumption that it is GG distributed as in (99).
2. Using specified values for the elements of the CMP parameter vector (say, $\bar{\gamma}' = [\bar{\beta}_o' \quad \bar{\beta}_u \quad \bar{\beta}_x \quad \bar{\psi}]$), the generated sample values for $\mathcal{W}_j$ and $\mathcal{X}_j$, and a generated purely random component, we simulated the sample of values of $\mathcal{Y}_j$ under the assumption that it is CMP distributed as in (126).[28]
3. Instead of plugging U_j into the CMP data generator in the first step of the sample design, we replace it with the residual as in (101)

[28] An appendix detailing the numerical aspects of the sampling designs and the corresponding Stata/Mata implementation code will be supplied upon request.

$\mathcal{X}^{FO}_{uGG}(X, W; \delta) =$

$$X - \exp\{W\alpha + (\sigma/\kappa) \times \ln(\kappa^2) + \ln[\Gamma((1/\kappa^2) + (\sigma/\kappa))] - \ln[\Gamma(1/\kappa^2)]\}.$$

As the estimation objective for the simulations, we chose the version of the AIE in (61) implied by the counterfactual defined as in (148). Definitions for X(ω) and X remain the same and the two relevant counterfactual scenarios (pre-counterfactual scenario and post-counterfactual scenarios) remain as in the exogenous case. We seek to investigate the same query.

"Starting with a counterfactually imposed scenario in which the value of the **X** for each member of the relevant population is equal to their observable value at the time of the data survey, suppose we counterfactually imposed a change in which the value of the **X** for each member of the relevant population is set equal to 0, what would be the consequent causal effect on the **Y**?"

As in the exogenous case, the Δ(ω) in (149) is not constant across members of the relevant population. Under the above two-step data simulation protocol, the counterfactual in (148) implies the following version of version of the AIE in (61)

$$AIE_{TRUE}(X^{pre}(\omega) = X(\omega), \Delta(\omega) = -X(\omega)) = AVG_{\omega \in \Omega}\{E[aie_{TRUE}(\bar{\theta}, \omega)]\} \tag{154}$$

where

$aie_{TRUE}(\theta, \omega) =$

$$E[m_{CMP}(\exp(\mathcal{X}_o\bar{\beta}_o + U\bar{\beta}_u), \bar{\psi})$$
$$- m_{CMP}(\exp(\mathcal{X}_o\bar{\beta}_o + U\bar{\beta}_u + X(\omega)\bar{\beta}_x), \bar{\psi})\,]$$

$\mathcal{X}_o$ and U denote the pseudo-random variables from which $\mathcal{X}_{oj}$ and , U_j are drawn, respectively

$\bar{\theta} = [\bar{\delta}' \quad \bar{\gamma}']$

and $m_{CMP}(.\ ,.)$ is defined as in (117). For a particular sampling design specification of the CMP deep parameter vector $\bar{\gamma}' = [\bar{\beta}_o' \quad \bar{\beta}_u \quad \bar{\beta}_x \quad \bar{\psi}]$ we approximated the "true" value of the AIE in (154) as

$$AIE_{TRUE} = \frac{\sum_{j*=1}^{n(\infty)} aie_{TRUE(j*)}}{n(\infty)} \tag{155}$$

where j* denotes the j*th simulated observation in a "super sample" of size $n(\infty) = 1.5M$,

$$aie_{TRUE(j*)} = m_{CMP}(\exp(\mathcal{X}_{oj*}\bar{\beta}_o + U_{j*}\bar{\beta}_u), \bar{\psi}) - m_{CMP}(\exp(\mathcal{X}_{oj*}\bar{\beta}_o + U_{j*}\bar{\beta}_u + \mathcal{X}_{j*}\bar{\beta}_x), \bar{\psi})]$$

and $j* = 1, \ldots, n(\infty)$.

To the kth simulated sample (k = 1,…, 200) we applied each of the five methods (EST = GG-Poisson, GG-CMP, LIN-Poisson, LIN-CMP, and LIV) and obtained a corresponding estimate of the deep parameter vector (say, $\hat{\theta}_{EST(k)} = [\hat{\delta}'_{EST(k)} \quad \hat{\gamma}'_{EST(k)}]$). Using $\hat{\theta}_{EST(k)}$, we then estimated the AIE using the following version of (63)

$$\widehat{AIE}_{EST(k)} = \frac{\sum_{i=1}^{n} aie_{(i)}(\hat{\theta}_{EST(k)})}{n} \tag{156}$$

where

$$\text{aie}_{(i)}(\hat{\theta}_{EST(k)}) = m(X_{oi}, \mathcal{X}_u(X_i, W_i; \hat{\delta}_{EST(k)}), 0; \hat{\gamma}_{EST(k)})$$

$$- m(X_{oi}, \mathcal{X}_u(X_i, W_i; \hat{\delta}_{EST(k)}), X_i; \hat{\gamma}_{EST(k)})$$

and m(   ) is defined in (59).

To assess the performance of each of the five methods, we used the following cross-sample average of (156)

$$(\text{AVG } \widehat{\text{AIE}}_{EST}) = \sum_{k=1}^{200} \frac{1}{200} \widehat{\text{AIE}}_{EST}(k) \tag{157}$$

where $\widehat{\text{AIE}}_{EST}(k)$ denotes the AIE estimator in (156) [EST = GG-Poisson, GG-CMP, LIN-Poisson, LIN-CMP, or LIV] for the kth sample (k = 1, …, 200). We also calculated the following average absolute percentage bias statistic

$$(\text{AVG } |\% \text{ Bias}|)_{EST} = \sum_{k=1}^{200} \frac{1}{200} \left| \frac{\widehat{\text{AIE}}_{EST}(k) - \text{AIE}_{TRUE}}{\text{AIE}_{TRUE}} \right| \times 100\% . \tag{158}$$

We note that none of the five FO-2SRI estimators under consideration is consistent for the "true" (under our sampling design) AIE in (154). The reason for this is the appearance in (154) of U instead of $\mathcal{X}_u(X, W; \delta)$ as the argument that is multiplicative with $\bar{\beta}_u$ in $m_{CMP}(\,.\,,.\,)$. Although the use of U in the specification of the DGP seems natural and reasonable, the design and implementation of an estimator for the parameters of such a process is not simple and straightforward. On the other hand, FO-2SRI estimators are intuitive and easy to apply using conventional econometric software. For this reason, it is useful to get a sense of the size of the potential bias cost that might be incurred by applied researchers seeking simplicity and convenience through the use of FO-2SRI methods. The statistics in (156) and (158) should be informative in this regard.

5.3.2 Data Generated with Unobservable Confounding (UC) Using First-Order (FO) Residuals –Sampling Design

Considering the simulation protocol, to compare the performance of the alternative estimators (i.e. FIML-CMP, FIML-Poisson, and OLS), the following sampling design is considered.

1) To generate pseudo values for X, we set

$$\bar{\alpha}'_{W} = [\bar{\alpha}_{X_o} \quad \bar{\alpha}_{con} \quad \bar{\alpha}_{W^+}] = [-0.75 \quad 1.25 \quad 0.75]$$

for the parameter vector. The means and variances of X_o and W^+ are set to be $E[X_o] = 0$, $E[W^+] = 0$, $Var[X_o] = 1$ and $Var[W^+] = 1$. The values for the shape parameters are set as $\log\text{-}\bar{\sigma} = -2$ and $\bar{\kappa} = 2$.

3) Similarly, to generate the pseudo values for Y, the following parameter values are set

$$[\bar{\beta}_o' \quad \bar{\beta}_u \quad \bar{\beta}_x \quad \bar{\psi}] = [0.5 \quad 0 \quad 0.5 \quad -0.5 \quad \bar{\psi}]$$

where $\bar{\beta}_o' = [\bar{\beta}_{X_o} \quad \bar{\beta}_{cons}]$ are the coefficients for X_o and the intercept, respectively and the following values for the log-transformed version of the dispersion parameter are considered to account for different levels of dispersion.

$\bar{\psi}$ = -2 , -1.75, -1.5, -1.25, -1, -0.75, -0.5, -0.25, 0, 0.25, 0.75, 1, 1.25, 1.5, 1.75, 2

5.3.3 Estimation Results – Data Generated Using First-Order (FO) Residuals – Unobservable Confounding (UC) Ignored in Estimation: FIML-CMP, FIML-Poisson, OLS

When estimating the value of the AIE, we ignore the presence of UC. Results in Table 3 demonstrate the estimated AIE corresponding to the equation (156) when applying the FIML-CMP, FIML-Poisson and linear OLS estimators as defined in chapter 3. Only the OLS estimator does not take account of the boundedness and discreteness of count data (CMP and Poisson account for these characteristics) and only the CMP takes account of the dispersion characteristic of count data. None of these estimators account for UC (the endogeneity of X) and the additional nonlinearity that could therefore be induced.

Table 3: Estimated Average AIE (157), Simulated FO Residual and Ignoring UC

Dispersion ψ	**True AIE**	**FIML-CMP**	**FIML-Poisson**	**OLS**
-2	8.143	4.984	5.175	-0.020
-1.75	7.334	4.326	4.065	-0.008
-1.5	6.251	3.496	2.926	0.014
-1.25	4.902	2.545	2.004	0.037
-1	3.539	1.732	1.424	0.051
-0.75	2.499	1.224	1.074	0.057
-0.5	1.800	0.917	0.848	0.058
-0.25	1.325	0.714	0.687	0.058
0	0.993	0.576	0.573	0.056
0.25	0.761	0.474	0.485	0.053
0.5	0.600	0.399	0.419	0.051
0.75	0.489	0.342	0.368	0.049
1	0.412	0.300	0.329	0.047
1.25	0.358	0.269	0.302	0.045
1.5	0.322	0.248	0.283	0.044
1.75	0.300	0.236	0.272	0.044
2	0.289	0.230	0.267	0.043

200 replications with 150,000 sample size.

Results in Table 4 demonstrate the average absolute percentage bias of the AIE estimator corresponding to equation (158).

Table 4: Average Absolute % Bias of Estimated AIE, Simulated FO Residual and Ignoring UC

Dispersion ψ	**True AIE**	**FIML-CMP**	**FIML-Poisson**	**OLS**
-2	8.143	38.8%	36.4%	100.2%
-1.75	7.334	41.0%	44.6%	100.1%
-1.5	6.251	44.1%	53.2%	99.8%
-1.25	4.902	48.1%	59.1%	99.2%
-1	3.539	51.1%	59.8%	98.6%
-0.75	2.499	51.0%	57.0%	97.7%
-0.5	1.800	49.0%	52.9%	96.8%
-0.25	1.325	46.1%	48.1%	95.7%
0	0.993	42.0%	42.3%	94.4%
0.25	0.761	37.7%	36.2%	93.0%
0.5	0.600	30.2%	33.5%	91.5%
0.75	0.489	24.8%	30.1%	90.0%
1	0.412	27.2%	20.0%	88.6%
1.25	0.358	24.9%	15.8%	87.3%
1.5	0.322	22.9%	12.1%	86.2%
1.75	0.300	21.4%	9.3%	85.4%
2	0.289	20.4%	7.7%	85.0%

200 replications with 150,000 sample size.

We observe that on the margin of boundedness and discreteness, the results in Table 4 indicate that, at different levels of dispersion, the OLS estimator is most biased, but the bias slightly decreases for the OLS estimator as the data becomes more underdispersed. For example, at $\psi = -1.75$ the averages percentage bias for OLS is 100.1% while at $\psi =$

1.75 this value is 85.4%. The results for the FIML-CMP and Poisson estimators are mixed at different dispersion levels. The FIML-CMP estimator does not produce smaller bias compared to other estimators across all specified dispersion levels but at some dispersion levels such as $-1.75 \leq \psi \leq$ and $0.5 \leq \psi \leq 0.75$, the FIML-CMP is less biased. The Poisson estimator performs better at some dispersion levels, especially as data the becomes more underdispersed; in which case the average bias gets smaller for this estimator and remains smaller than the bias for the dispersion flexible FIML-CMP estimator.

5.3.4 Estimation Results – Data Generated Using First-Order (FO) Residuals – Unobservable Confounding (UC) Accounted for in Estimation: GG-CMP, GG-Poisson, LIN-CMP, LIN-Poisson, and LIV

Similar to section 5.3.2, we implement the FO residuals in the CMP data generator but this time we consider the presence of the UC when estimating the value AIE. This substitution results in simulated data that exactly comports with the GG-CMP FO-2SRI estimator since the data generator was designed to produce samples that exactly coincide with draws from the would-be populations underlying GG-CMP FO-2SRI CRM-based CE estimation protocol. Therefore, we can use it to validate that our GG-CMP FO-2SRI estimator design is correct since the data simulation design is not subject to misspecification error. We report the results of applying the FO-2SRI estimators detailed in chapter 4 (GG-CMP, GG-Poisson, LIN-CMP, LIN-Poisson, and LIV).

The LIV estimator is the only estimator that does not account for any of the nonlinear features of count data or the UC. With regard to dispersion-flexibility the LIV, LIN-Poisson and GG-Poisson are the estimators which do not take account of this feature of count data. Only the GG-CMP and GG-Poisson estimators take account of nonlinearity induced by UC. GG-CMP is the only estimator which considers all the aforementioned possible sources of endogeneity. We compare the performance of the estimators with respect to each of these aspects of the data.

Table 5: Estimated Average AIE, Simulated FO Residual and Considering UC

Dispersion ψ	**True AIE**	**GG-CMP**	**GG-Poisson**	**LIN-CMP**	**LIN-Poisson**	**LIV**
-2	8.143	8.138	574.153	2.515	2.064	2.564
-1.75	7.334	7.331	218.388	2.225	1.534	2.115
-1.5	6.251	6.253	51.591	1.845	1.176	1.610
-1.25	4.902	4.906	12.329	1.415	0.944	1.148
-1	3.539	3.545	4.877	1.061	0.794	0.828
-0.75	2.499	2.505	2.775	0.822	0.684	0.627
-0.5	1.800	1.806	1.878	0.659	0.594	0.495
-0.25	1.325	1.326	1.384	0.541	0.519	0.400
0	0.993	0.998	1.083	0.454	0.458	0.333
0.25	0.761	0.765	0.876	0.387	0.406	0.281
0.5	0.600	0.603	0.729	0.335	0.363	0.242
0.75	0.489	0.491	0.623	0.294	0.328	0.211
1	0.412	0.414	0.547	0.263	0.301	0.188
1.25	0.358	0.359	0.493	0.240	0.281	0.172
1.5	0.322	0.323	0.458	0.225	0.267	0.160
1.75	0.300	0.301	0.437	0.216	0.259	0.154
2	0.289	0.290	0.428	0.211	0.255	0.150

200 replications with 150,000 sample size.

On the margin of boundedness and discreteness, the results in Table 6 indicate that at different levels of dispersion (under-dispersion, equi-dispersion, and over-dispersion), the LIV estimator underestimates the value of AIE by more than 48% (corresponding to equation (158)) and this percentage bias decreases for more under-dispersed data. The amount of bias for the LIV estimator is comparable in some cases with the LIN-CMP and LIN-Poisson estimators that are linear in their first stage while taking account of the boundedness and discreteness. (especially in the overdispersion case at $-2 \leq \psi \leq -0.25$)

On the margin of dispersion-flexibility, for the GG-CMP and LIN-CMP which are the only dispersion flexible models, we observe that GG-CMP produces significantly smaller percentage bias compared to LIN-CMP across all specified dispersion levels specially in the overdispersion case. Moreover, on the margin of nonlinearity due to the UC, among the estimators that account for this aspect (GG-CMP and GG-Poisson), estimated AIE values from GG-CMP estimator are dominantly closer to the true value of the AIE compared to GG-Poisson across all specified dispersion levels. The amount of bias being around 1% at all dispersion levels for this estimator. The only other estimator for which estimated the value of AIE is close the true value at a few dispersion levels (close to equi-dispersion) is the GG-POI. At $\psi = -0.25$ and $\psi = -0.25$ the percentage bias is less than 6% which is still large in comparison to the GG-CMP results.

Table 6: Average Absolute % Bias of Estimated AIE, Simulated FO Residual and Considering UC

Dispersion ψ	True AIE	GG-CMP	GG-Poisson	LIN-CMP	LIN-Poisson	LIV
-2	8.143	0.55%	6951.26%	69.11%	74.66%	68.51%
-1.75	7.334	0.64%	2877.93%	69.66%	79.08%	71.16%
-1.5	6.251	0.79%	725.29%	70.49%	81.19%	725.29%
-1.25	4.902	1.02%	151.49%	71.14%	80.75%	76.59%
-1	3.539	1.27%	37.79%	70.01%	77.56%	76.61%
-0.75	2.499	1.45%	11.20%	67.10%	72.65%	74.89%
-0.5	1.800	1.60%	5.10%	63.37%	66.98%	72.51%
-0.25	1.325	1.60%	4.95%	59.17%	60.80%	69.78%
0	0.993	1.69%	9.06%	54.25%	53.89%	66.49%
0.25	0.761	1.62%	15.07%	49.14%	46.65%	63.06%
0.5	0.600	1.44%	21.46%	44.22%	39.50%	59.74%
0.75	0.489	1.28%	27.38%	39.91%	32.93%	56.82%
1	0.412	1.20%	32.74%	36.18%	26.96%	54.28%
1.25	0.358	1.10%	37.60%	32.96%	21.63%	52.10%
1.5	0.322	1.02%	42.06%	30.26%	17.10%	50.26%
1.75	0.300	0.95%	45.56%	28.25%	13.74%	48.90%
2	0.289	0.89%	47.77%	27.06%	11.79%	48.08%

200 replications with 150,000 sample size.

5.3.5 Data Generation Protocol – Implementing Uniform (“Ideal”) Residuals to Account for Unobservable Confounding (UC)

The data generation protocol is the same as the case detailed in section 5.3.1 with only one exception and that is the protocol used to generate the residuals. Here, instead of generating the residual as in (101), we generate the data using a unit uniform that is used to generate the X variable and we call it the *ideal* residual. That is, we first generate X_o and W_j^+ using the specified values for the means and variances of X_o and W_j^+ in the sampling design and generate random unit uniform U_X. We then use X_o and W_j^+ and U_X to generate a vector of GG distributed X values and generate random unit uniform U_Y. Finally, we use X_o, $X_u = U_X$, X and U_Y to generate a vector of CMP distributed Y values.

5.3.6 Sampling Designs – Data Generated with Unobservable Confounding (UC) Using Uniform ("Ideal") Residuals

The sampling design follows the same designated values for the parameters as in section 5.3.2.

5.3.7 Simulation Results – Data Generated Using Uniform ("Ideal") Residuals – Unobservable Confounding (UC) Ignored in Estimation: OLS, FIML-Poisson, FIML-CMP

In this section we consider the exact sampling design detailed in section 5.3.1 but ignore the presence of UC when estimating the value of the AIE. Table 7 displays the results from estimating the value of the AIE using (156) based on the FIML-CMP, FIML-Poisson and linear OLS estimators as defined in chapter 3. Only the OLS estimator does not take account of the boundedness and discreteness of count data (CMP and Poisson account for these characteristics) and only the CMP takes account of the dispersion characteristic of count data. None of the estimators account for UC (the endogeneity of X) and the nonlinearity that could be induced thereby.

Table 7: Estimated Average AIE, Simulated IDEAL Residual and Ignoring UC

Dispersion ψ	**True AIE**	**FIML-CMP**	**FIML-Poisson**	**OLS**
-2	7.973	6.780	8.135	0.067
-1.75	7.269	6.135	6.934	0.042
-1.5	6.349	5.316	5.465	0.031
-1.25	5.156	4.299	3.893	0.042
-1	3.743	3.119	2.587	0.064
-0.75	2.469	2.059	1.766	0.078
-0.5	1.666	1.418	1.295	0.082
-0.25	1.203	1.044	1.000	0.081
0	0.913	0.803	0.807	0.077
0.25	0.719	0.642	0.661	0.073
0.5	0.582	0.526	0.558	0.069
0.75	0.484	0.442	0.481	0.065
1	0.413	0.380	0.425	0.062
1.25	0.362	0.335	0.384	0.060
1.5	0.328	0.305	0.357	0.058
1.75	0.307	0.287	0.342	0.057
2	0.297	0.278	0.334	0.056

200 replications with 150,000 sample size.

The results in Table 8 display the average absolute percentage biases of the estimators.

Table 8: Average Absolute % Bias of Estimated AIE, Simulated IDEAL Residual and Ignoring UC

Dispersion ψ	**True AIE**	**FIML-CMP**	**FIML-Poisson**	**OLS**
-2	7.973	15.0%	2.0%	99.2%
-1.75	7.269	15.6%	4.6%	99.4%
-1.5	6.349	16.3%	13.9%	99.5%
-1.25	5.156	16.6%	24.5%	99.2%
-1	3.743	16.7%	30.9%	98.3%
-0.75	2.469	16.6%	28.5%	96.8%
-0.5	1.666	14.9%	22.3%	95.1%
-0.25	1.203	13.2%	16.9%	93.3%
0	0.913	12.0%	11.6%	91.5%
0.25	0.719	10.6%	8.0%	89.8%
0.5	0.582	9.6%	4.2%	88.1%
0.75	0.484	8.8%	1.0%	86.5%
1	0.413	8.0%	2.9%	85.0%
1.25	0.362	7.4%	6.2%	83.5%
1.5	0.328	6.8%	9.1%	82.3%
1.75	0.307	6.5%	11.3%	81.5%
2	0.297	6.3%	12.5%	81.0%

200 replications with 150,000 sample size.

We observe that on the margin of boundedness and discreteness, the results in table 8 indicate that at different levels of dispersion, the OLS estimator is the most biased among the estimators, but the bias slightly decreases for the OLS estimator as the data becomes more underdispersed. For example, at $\psi = -2$ the average percentage bias for the OLS is

92% while at $\psi = 2$ this value is 81%. The results for the FIML-CMP and FIML-Poisson estimators are mixed at different dispersion levels. The FIML-CMP estimator does not produce smaller bias compared to other estimators across all specified dispersion levels but at some dispersion levels such as $-1.25 \leq \psi \leq -0.25$ and $1.5 \leq \psi \leq 2$, the FIML-CMP performs better relative to the other estimators. The FIML-Poisson estimator on the other hand performs relatively well at some dispersion levels. For example, at $\psi = -2$ the amount of bias is 2% but this amount increases as the data gets closer to the equi-dispersion case. The bias remains fairly low at lower levels of under dispersion $0.25 \leq \psi \leq 1.25$ but increases as the data becomes more extremely under dispersed.

5.3.8 Simulation Results – Data Generated Using Uniform ("Ideal") Residuals – Unobservable Confounding (UC) Accounted for in Estimation: GG-CMP, GG-Poisson, LIN-CMP, LIN-Poisson, and LIV

Like the previous section, here we consider the exact sampling design detailed in section 5.3.1; but this time consider the presence of UC when estimating the value of the AIE. It is worth noting that this generator does not align with any of the five models under consideration in that it replaces the FO-2SRI assumption with U_j. The complete results of the simulation study applying the FO-2SRI estimators accounting for UC (the endogeneity of X) are displayed in Tables 9 and 10; for the average AIE statistic (156) and the average absolute percentage bias (158), respectively.

We compare the amount of bias among the estimators based on what differentiates them. As detailed in chapter 4, the five proposed FO-2SRI estimators include GG-CMP, GG-Poisson, LIN-CMP, LIN-Poisson, and LIV estimators are different with regard to how they account for the various sources of nonlinearity, including aspects of the CRM i.e. discreteness and boundedness, dispersion flexibility, and UC. Therefore, we compare these estimators based on each of these margins (sources) of nonlinearity by comparing the average of absolute percentage bias for the estimators.

On the margin of boundedness and discreteness of the count data, only the LIV estimator does not take account of these characteristics of count data (GG-CMP, GG-Poisson, LIN-CMP, and LIN-Poisson, all account for these intrinsic characteristics). On the margin of dispersion-flexibility, the linear LIV and Poisson based estimators (LIN-Poisson and GG-Poisson) are the estimators which do not take account of the dispersion characteristic of count data. And on the margin of nonlinearity induced by the UC, among the proposed estimators only GG-CMP and GG-Poisson take account of this aspect (GG-CMP is the only estimator which considers all afore mentioned possible sources of nonlinearity). We focus on each of these margins and compare the performance of the estimators that take account of these sources of nonlinearity with the estimators that (at least to some degree) do not.

Table 9: Estimated Average AIE, Simulated IDEAL Residual and Considering UC

Dispersion ψ	True AIE	GG-CMP	GG-Poisson	LIN-CMP	LIN-Poisson	LIV
-2	7.973	9.196	19.713	4.193	5.556	-0.725
-1.75	7.269	8.573	19.526	3.755	4.047	-0.642
-1.5	6.349	7.693	15.685	3.268	2.719	-0.534
-1.25	5.156	6.381	8.327	2.698	1.831	-0.403
-1	3.743	4.555	3.947	2.042	1.372	-0.278
-0.75	2.469	2.797	2.304	1.481	1.130	-0.190
-0.5	1.666	1.775	1.577	1.114	0.957	-0.137
-0.25	1.203	1.236	1.173	0.870	0.816	-0.105
0	0.913	0.924	0.920	0.699	0.701	-0.083
0.25	0.719	0.717	0.744	0.574	0.608	-0.068
0.5	0.582	0.578	0.620	0.481	0.533	-0.057
0.75	0.484	0.479	0.530	0.411	0.473	-0.049
1	0.413	0.408	0.464	0.358	0.427	-0.043
1.25	0.362	0.357	0.418	0.320	0.394	-0.038
1.5	0.328	0.323	0.387	0.293	0.371	-0.035
1.75	0.307	0.303	0.369	0.278	0.358	-0.033
2	0.297	0.293	0.361	0.270	0.351	-0.033

200 replications with 150,000 sample size.

Results in Table 10 demonstrate the average absolute percentage bias of the AIE estimator corresponding to the equation (158).

Table 10: Average Absolute % Bias of Estimated AIE, Simulated IDEAL Residual and Considering UC

Dispersion ψ	**True AIE**	**GG-CMP**	**GG-Poisson**	**LIN-CMP**	**LIN-Poisson**	**LIV**
-2	7.973	15.34%	147.25%	47.40%	30.32%	109.09%
-1.75	7.269	17.94%	168.63%	48.34%	44.32%	108.83%
-1.5	6.349	21.18%	147.06%	48.52%	57.17%	108.41%
-1.25	5.156	23.77%	61.51%	47.66%	64.48%	107.82%
-1	3.743	21.69%	5.44%	45.45%	63.36%	107.42%
-0.75	2.469	13.25%	6.71%	40.01%	54.23%	107.68%
-0.5	1.666	6.52%	5.37%	33.14%	42.57%	108.25%
-0.25	1.203	2.78%	2.58%	27.68%	32.16%	108.72%
0	0.913	1.44%	1.22%	23.44%	23.19%	109.11%
0.25	0.719	1.02%	3.52%	20.06%	15.35%	109.45%
0.5	0.582	1.16%	6.51%	17.32%	8.41%	109.76%
0.75	0.484	1.34%	9.48%	15.10%	2.22%	110.04%
1	0.413	1.46%	12.53%	13.23%	3.54%	110.30%
1.25	0.362	1.54%	15.50%	11.67%	8.77%	110.54%
1.5	0.328	1.56%	18.24%	10.41%	13.26%	110.74%
1.75	0.307	1.56%	20.30%	9.54%	16.52%	110.88%
2	0.297	1.53%	21.50%	9.03%	18.36%	110.96%

200 replications with 150,000 sample size.

On the margin of boundedness and discreteness, results in table 10 indicate that at different levels of dispersion (under-dispersion, equi-dispersion, and over-dispersion), the LIV estimator underestimates the value of AIE by more than 100% which makes this estimator the most biased among all. For example, at ψ = 1.25 the true value of the AIE is

.362 and the estimated average AIE is –.381 by the LIV which is opposite in sign relative to the true value of the AIE. The amount of bias does not increase or decrease much with the data being more over or under-dispersed.

On the margin of dispersion-flexibility, the GG-CMP and LIN-CMP are the only estimators that are dispersion flexible. We observe in table 10 that GG-CMP produces smaller bias compared to LIN-CMP across all specified dispersion levels (high level of precision with percentage bias being around 1% at $\Upsilon \geq 0$). The amount of bias for both estimators decrease as data becomes more underdispersed but GG-CMP consistently remains the dominant estimator in terms of smaller bias.

And finally on the margin of nonlinearity due to the UC, the only estimators which take account of this matter are the GG-CMP and GG-Poisson estimators. Comparing the estimated values from these estimators with the true value of the AIE indicates that estimated values from GG-CMP are much closer to the true value of the AIE compared to GG-Poisson across all specified dispersion levels. The only exception is for dispersion values $0 \leq \psi \leq -1$ where the GG-Poisson is slightly less biased.

5.4 Discussion

In this chapter a simulation analysis is conducted to demonstrate the implementation of the exogenous estimators (FIML-CMP, FIML-Poisson, and OLS) as well as the FO-2SRI estimators (GG-CMP, GG-Poisson, LIN-CMP, LIN-Poisson, and LIV) and investigate the amount of bias yielded from applying these estimators compared to the true value of the AIE. A comparison in the absence of UC shows that, if the outcome of interest is count valued, there is an advantage in utilizing the nonlinear methods in estimating the CE, especially in more severely under or over dispersion data . In such cases, it is more essential to use a dispersion flexible specification such as the CMP. In the presence of UC, we show that the proposed GG-CMP estimator is the least biased estimator at all dispersion levels even when the simulated sampling regime is subject to the misspecification error while the conventional LIV is the most biased estimator. This emphasizes the importance of using nonlinear models when analyzing count data; especially in the presence of UC and the additional nonlinearity that it might be induced. With the results obtained from the Poisson based estimators, the practicality of these estimators tends to depend on the dispersion level of the count data. While the GG-Poisson estimator is useful at the more under-dispersion and equi-dispersion levels, the LIN-Poisson estimator shows that at some underdispersion levels it might be preferred to the GG-Poisson. And we also notice that if we ignore the endogeneity of the presumably causal entity of interest, the proposed GG-CMP is still the least biased estimator among all estimators. We thus conclude that accounting for nonlinearity and dispersion characteristic of count data in the presence and absence of endogeneity improves the estimation of the CE and that this is more pronounced in the case of endogeneity and more severe underdispersion and overdispersion.

Chapter 6

Real Data Application: Effect of Education on Fertility Decisions

6.1 Overview

To illustrate and compare the models and estimators detailed in the previous chapters, we applied them in the context of the E/F example. Fertility decisions have consequences for economic development (Ashraf et al., 2013; Fox et al., 2019), child health and educational attainment (Blaabæk et al., 2020; Lee, 2008) and the environment (Stephenson et al., 2010). Research has shown that increasing females' educational attainment decreases early fertility (Osili & Long, 2008), desire to have children (Kebede et al., 2021) and the number of children per woman (Ali & Gurmu, 2018; Güneş, 2016; Amin & Behrman, 2014; Currie & Moretti, 2003).

In this illustration, let us suppose that we seek to estimate the CE [on fertility] of a hypothetical counterfactual change in years of education that would bring all women to *at least* the secondary diploma equivalent (12 years).[29] We begin by placing the discussion in the context of the GPOF detailed in chapter 2. In the GPOF we have

$\mathbf{X}$ ≡ education

$\mathbf{Y}$ ≡ fertility

$X^*(\omega)$ ≡ a deterministic function that maps women (ω) in the population (Ω) to counterfactually mandated levels (years) of education.

[29] In Nigeria, the education system includes six years of primary school followed by 3 years of junior secondary school and 3 years of senior secondary. Only primary and junior secondary schooling is compulsory (9 years). The equivalent of a U.S. high school diploma in Nigeria is completion of the full 12 years of primary and secondary education.

$X \equiv$ the random variable representing the distribution of observable levels of education for women in the population (the component of the DGP from which education levels can be sampled)

$Y_{X^*(\omega)} \equiv$ the random variable representing the distribution of counterfactual fertility *potential outcome* levels (number of viable children) for a woman (ω) that would have held true if $X^*(\omega)$ years of education were counterfactually mandated for her.

$Y \equiv$ the random variable representing the distribution of observable fertility levels for women in the population (the component of the DGP from which fertility levels can be sampled).

$X^{pre}(\omega) \equiv$ the counterfactually imposed pre- value of the **X** for individual ω;

$Y_{X^{pre}(\omega)} \equiv$ fertility *potential outcome* for the pre- counterfactual scenario corresponding to $X^{pre}(\omega)$

$\Delta(\omega) = \max(0, 12 - X^{pre}(\omega)) \equiv$ counterfactually imposed increment to $X^{pre}(\omega)$ that defines the counterfactually imposed post- version of the **X** for individual ω,

$$X^{post}(\omega) = X^{pre}(\omega) + \Delta(\omega) = X^{pre}(\omega) + \max(0, 12 - X^{pre}(\omega))$$

and

$Y_{X^{post}(\omega)} = Y_{X^{pre}(\omega) + \Delta(\omega)} = Y_{X^{pre}(\omega) + \max(0, 12 - X^{pre}(\omega))} \equiv$ fertility *potential outcome* for the post- counterfactual scenario corresponding to

$= X^{pre}(\omega) + \max(0, 12 – X^{pre}(\omega)).$

for this illustration, we will assume that $X^{pre}(\omega) = X^{cm}(\omega)$, where $X^{cm}(\omega)$ denotes the observable value of the **X** (the level of education, represented by the random variable X) pertinent to individual ω. In other words, we choose the relevant pre- value of $X^*(\omega)$ to be the same as the value of X(ω) *as if it were counterfactually mandated* (hence the "cm" superscript) rather than a possible outcome in sampling from the DGP. Under this assumption we have:

$$Y_{X^{pre}(\omega)} = Y_{X^{cm}(\omega)}$$

$$\Delta(\omega) = \max(0, 12 – X^{cm}(\omega))$$

$$X^{post}(\omega) = X^{cm}(\omega) + \max(0, 12 – X^{cm}(\omega))$$

$$Y_{X^{post}(\omega)} = Y_{X^{cm}(\omega) + \max(0, 12 – X^{cm}(\omega))}$$

The relevant counterfactual comprises the change from pre-scenario mandated education levels to post- scenario mandated education levels. In other words, the relevant version of the formal representation of the counterfactual in (1) for this example is

$$[X^{cm}(\omega), \max(0, 12 – X^{cm}(\omega)]. \qquad (159)$$

In defining the relevant CE we pose the following counterfactual query

> "If the relevant counterfactual [as given in (159)] were imposed, how would the fertility potential outcome change?"

6.2 Exogenous X

First, we consider the case in which the mother's education (X) is not subject to UC and apply the estimators detailed in chapter 3 to estimate the CE [on fertility] of a hypothetical counterfactual change in years of education that would bring all women to *at least* the secondary diploma equivalent (12 years) while controlling for observed confounders such as age, marital status, etc. The data used for this illustrative example were taken from the sixth round of the Multiple Indicator Cluster Survey of Nigeria (MICSN) conducted in 2021 (United Nations Children's Fund [UNICEF], 2021) by the National Bureau of Statistics of Nigeria as part of the Global Multiple Indicator Cluster Survey Program (GMICS).[30] The GMICS, developed by UNICEF in the 1990s, is an international household survey program that enables countries to collect comparable data on the health and education of women and children. The individuals sampled in the MICSN are women of reproductive age (15 – 49), defined as such by the World Health Organization. Variable names and definitions in the sample are reported in Table 11.

To quantify the CE (the answer to the counterfactual query in section 6.1) we use the three versions of the generic average incremental effect (AIE) in (14) that correspond to the FIML-Poisson, FIML-CMP, and OLS models, respectively. To estimate the deep parameters on which these models are based, we used the corresponding protocols detailed in chapter 3.

[30] The MICSN 2021 is publicly available from UNICEF upon registration.

Table 11: Variable Names and Definitions

Variable Name	Definition
Count-Valued Outcome Variable of Interest (Y)	
CSURV	number of viable children born to a woman
Presumed Causal Variable (X)	
years_education	woman's number of years of schooling
Identifying Instrumental Variable (W^+)	
first_half	woman was born in the first half of the year
Regression Control Variables (X_o)	
mother_age	woman's age
married	woman is currently married
married_before	woman has been married before but is not married now
never_married	woman has never been married
Hausa	head of the household belongs to the Hausa ethnic group
Igbo	head of the household belongs to the Igbo ethnic group
Yoruba	head of the household belongs to the Yoruba ethnic group
Fulani	head of the household belongs to the Fulani ethnic group
Kanuri	head of the household belongs to the Kanuri ethnic group
Ijaw	head of the household belongs to the Ijaw ethnic group
Tiv	head of the household belongs to the Tiv ethnic group
Ibibio	head of the household belongs to the Ibibio ethnic group
Edo	head of the household belongs to the Edo ethnic group
urban	woman lives in an urban area
North_Central	woman lives in the north-central region
North_East	woman lives in the northeast region
North_West	woman lives in the northwest region
South_East	woman lives in the southeast region
South_South	woman lives in the south-south region
Television	woman watches television at least once per week
Radio	woman listens to radio at least once per week
wscore	composite measure of a household's living standard and economic status, calculated based on a household's ownership of selected assets, housing characteristics, and types of water access and sanitation facilities

Descriptive statistics for the analytic sample are displayed in Table 12.

Table 12: Descriptive Statistics

	Mean	SD	min	max
Count-Valued Outcome Variable of Interest (Y)				
CSURV	1.96	2.26	0	14
Presumed Causal Variable (X)				
years_education	10.48	3.35	1	18
Identifying Instrumental Variable (W^+)				
first_half	0.56	0.50	0	1
Regression Control Variables (X_o)				
mother_age	27.96	9.79	15	49
married	0.52	0.50	0	1
married_before	0.06	0.24	0	1
never_married	0.42	0.49	0	1
Hausa	0.16	0.36	0	1
Igbo	0.19	0.39	0	1
Yoruba	0.16	0.37	0	1
Fulani	0.03	0.18	0	1
Kanuri	0.02	0.12	0	1
Ijaw	0.03	0.17	0	1
Tiv	0.02	0.16	0	1
Ibibio	0.03	0.17	0	1
Edo	0.02	0.15	0	1
urban	0.40	0.49	0	1
North_Central	0.23	0.42	0	1
North_East	0.16	0.36	0	1
North_West	0.13	0.33	0	1
South_East	0.16	0.37	0	1
South_South	0.17	0.37	0	1
Television	0.31	0.46	0	1
Radio	0.36	0.48	0	1
wscore	0.28	0.99	-1.91	3.1
N				26,807

Among the regression controls (X_o), we included the respondent's age, religion of the household head, the region of residence, ethnic group, wealth score index and access to media as observable confounders. The average age of the women in the study is almost

28 years old and they live in different geographical regions of Nigeria. On average, each woman has two children and has about ten years of schooling. About 52% of the sample are married and 40% of the sample live in an urban residence. On average, 36% percent of women have access to radio and 31% have access to television.

6.3 Endogenous X

In this section we look at identifying the effect of mother's education on their fertility decisions with a new lens, the nature of the mother's education (X) and the probable endogeneity thereof. The mother's education can be endogenous due to unobserved factors such as mother's motivation and preferences, household resources and cultural norms that can affect both the mother's education and the number of children that she will end up having in life. Ignoring the endogeneity of mother's education can lead to biased estimates of the AIE and jeopardize the causal interpretability of the estimated values of that effect. Therefore, the identification strategy relies on using an appropriate IV which is a binary variable indicating whether the individual was born in the first half of her birth year.[31]

[31] Other studies of the effect of education on fertility decisions have implemented this IV (Zanin et al, 2014; Sobotka et al., 2013). Sulemana et al. (2018) use a binary indicator for being born before September. Angrist and Krueger (1991) were the first to use this type of IV (dummy variables for quarter of birthday). Justification for the use of this IV in the present context rests on the fact that schooling is compulsory in Nigeria for youths between the ages of 6 and 15 (generally from grade one to nine) [Statista, 2023]. Although children who are born earlier in the year have the same integer-valued age as those born later in the year, they are fractionally older when they start school and, therefore, tend to get to the legal dropout age having less years of education (see Angrist and Krueger, 1991). The validity of this IV is based on the argument that it is unlikely to: a) be correlated with the unobservable factors that confound the causal effect of education on fertility (i.e., preferences, abilities of the mother, etc.); and b) have a direct effect the on the number of children the mother has.

To quantify the causal effect (the answer to the counterfactual query in section 6.1) we use the five versions of the generic average incremental effect (AIE) in (63) that correspond to the GG-Poisson, GG-CMP, and LIV models, respectively. To estimate the deep parameters on which the GG-Poisson, GG-CMP and AIE estimators are based, we used the corresponding 2SRI protocol detailed in sections 4.4. For the linear AIE estimator we used the LIV method.

6.4 Results: Comparison of the Estimates and Discussion

For the Exogenous X case or the case in which the UC is ignored, the results for the AIE estimates are reported for each of the three model specifications (FIML-CMP, FIML-Poisson, and OLS) in Table 13. The results indicate that the estimated AIE is almost identical for FIML-CMP and FIML-Poisson estimators. The linear OLS estimator yielded a bigger estimated CE compared to the nonlinear estimators. Overall, the results indicate that women would have 0.12 to 0.19 less children as a result of increasing the level of education for women to at least the high school diploma level. The estimated deep parameters for each model are reported in Table A1 in appendix A.

Table 13: Estimated AIE on Fertility of Counterfactually Mandating Minimum 12 Years of Education, Ignoring UC

Method	**FIML-CMP**	**FIML-Poisson**	**OLS**
AIE	-0.121 (-17.15)	-0.120 (-15.62)	-0.192 (-32.10)
Log-Dispersion Parameter (ψ)	0.217 (16.75)	-	-

Asymptotic t-statistics are reported in parentheses.

For the case in which we account for UC (endogenous X), the results for the AIE estimates for each of the five model specifications are reported in Table 14. The most striking difference that emerges from the table is between those methods that account for nonlinearity in both FO-2SRI modeling levels (GG-CMP and GG-Poisson) and those for which either or both of the modeling levels is assumed to be linear (LIN-CMP [2nd level], LIN-Poisson 2nd level] and LIV [both levels]). The former produce estimates that are substantially larger (in absolute value) than the latter (nearly a 1 child decrease in fertility vs. an estimated CE about half that size) for the latter. Moreover, AIE estimates from GG-CMP and the GG-Poisson models are highly statistically significant while the remaining models (whose specifications impose at least some linearity) yielded insignificant results. This indicates that modeling linearity does matter. It is also noteworthy that among the models whose 2nd level is nonlinear (GG-CMP, GG-Poisson, LIN-CMP and LIN-Poisson), accounting for 1st level nonlinearity makes a difference.

Table 14: Estimated AIE on Fertility of Counterfactually Mandating Minimum 12 Years of Education, Considering UC

Method	GG-CMP	GG-Poisson	LIN-CMP	LIN- Poisson	LIV
AIE	-0.905 (-18.91)	-0.892 (-18.34)	-0.035 (-0.147)	-0.035 (-0.145)	-0.077 (-0.43)
$\hat{\mathbf{X}}_{\mathbf{u}}^{\mathbf{2SRI}}$	0.485 (5.24)	0.401 (5.25)	-0.020 (-0.37)	-0.016 (-0.37)	-
Log-Dispersion Parameter (ψ)	0.224 (15.61)	-	0.217 (15.20)	-	-

Asymptotic t-statistics are reported in parentheses.

We next consider differences AIE estimates that may be attributable to whether or not dispersion was accounted for. It is often the case that fertility data are underdispersed (Winkelmann & Zimmermann, 1994; Winkelmann, 1995; Wang & Famoye, 1997; Santos Silva & Covas, 2000). That this is indeed a feature of our sample, is supported by the GG-CMP and LIN-CMP results. For both of these models, the dispersion parameter (ψ) estimate is positive and significant. That said, the AIE estimate for GG-CMP (which accounts for dispersion) and the estimate from GG-Poisson (which does not account for dispersion) are quite similar in value. This comports, somewhat, with the simulation results. As can be seen in Table 1, at low levels of underdispersion ($\psi < .25$) AIE estimates do not diverge much. It is at higher levels of dispersion ($\psi < -.5$ and $\psi > .5$) that ignoring dispersion leads to pronounced bias (see Table 10). The results of the second stage of the 2SRI estimation are reported in Table B1 in appendix B.

Chapter 7

Summary, Discussion and Conclusions

We have developed two new estimators for CEs based on FO-2SRI CRM-based model for cases in which the causal variable is continuous [$\widehat{\text{AIE}}_{\text{UC(POI)}}(X^{\text{pre}}, \Delta)$ in (106) and $\widehat{\text{AIE}}_{\text{UC(CMP)}}(X^{\text{pre}}, \Delta)$ in (120)]. For simplicity of exposition let us refer to these estimators as GG-CMP and GG-Poisson, AS we did in chapter 4. We compare these new CE estimators with extant FO-2SRI estimators (those based on control function methods [called LIN-Poisson, LIN-CMP in chapter 4] and the popular fully linear specification (the conventional two-stage least squares or linear instrumental variables estimator [called LIV in chapter 4]. In answer to our first study objective, we find that, based on studies of simulated and real data, estimates obtained via our newly proposed estimators (GG-CMP and GG-Poisson) widely diverge from the results from the extant linearized methods [LIN-CMP, LIN-Poisson and LIV]. The difference is particularly pronounced for the LIV estimator. This may serve as a cautionary tale for those seeking to use LIV causal analysis in the CRM context. To summarize our findings regarding the second study goal, we found that our proposed GG-CMP CE estimator, which allows dispersion flexibility, dominated all other estimators (aside from a few extreme dispersion cases). Finally, in support of our third study goal, we exploited the fact that our data generator was designed to produce samples that did not exactly coincide with draws from the would-be populations underlying any of the five FO-2SRI CRM-based CE estimation protocols. Therefore, in our simulated sampling regime, all five of our estimators were subject to misspecification error. Nevertheless, we found the GG-CMP CE estimator, which is designed to fully account for dispersion flexibility and inherent nonlinearity, to be quite accurate for a wide range of values of the dispersion parameter.

Appendices

Appendix A

Table A1: Deep Parameter Regression Estimates obtained from FIML-CMP, FIML-Poisson, and OLS models When Education is Exogenous

VARIABLE	FIML-CMP	FIML-Poisson	OLS
years_education	-0.027 (-17.10)	-0.023 (-15.82)	-0.095 (-33.27)
mother_age	0.054 (62.85)	0.045 (79.71)	0.117 (100.68)
married	0.363 (2.45)	0.305 (2.24)	0.843 (3.13)
married_before	0.028 (2.45)	0.025 (0.18)	0.064 (0.23)
never_married	-3.160 (0.19)	-3.013 (-21.31)	-0.821 (-3.04)
Hausa	0.136 (-20.64)	0.113 (6.28)	0.194 (5.84)
Igbo	-0.063 (6.85)	-0.053 (-1.94)	-0.101 (-2.07)
Yoruba	-0.071 (-2.12)	-0.060 (-2.86)	-0.044 (-1.19)
Fulani	0.149 (-3.11)	0.124 (4.63)	0.220 (4.43)
Kanuri	0.195 (5.05)	0.163 (4.05)	0.251 (3.64)
Ijaw	0.193 (4.43)	0.160 (6.19)	0.415 (7.71)
Tiv	0.028 (6.78)	0.024 (0.82)	0.009 (0.16)
Ibibio	-0.018 (0.89)	-0.015 (-0.51)	-0.128 (-2.41)
Edo	0.036 (-0.58)	0.030 (0.92)	0.054 (0.9)
urban	0.066 (1.00)	0.055 (4.65)	0.067 (3.13)
North_Central	0.044 (5.05)	0.037 (1.77)	0.092 (2.49)
North_East	0.180	0.150	0.300

	(1.95)	(6.32)	(6.97)
North_West	0.184 (6.91)	0.153 (5.78)	0.347 (7.16)
South_East	0.137 (6.34)	0.114 (3.8)	0.213 (3.95)
South_South	0.091 (4.15)	0.075 (3.27)	0.186 (4.39)
Television	0.003 (3.58)	0.003 (0.29)	0.003 (0.16)
Radio	-0.012 (0.30)	-0.010 (-1.08)	-0.019 (-1.06)
wscore	-0.095 (-1.17)	-0.079 (-12.55)	-0.108 (-9.65)
Constant	-0.534 (-3.51)	-0.529 (-3.77)	-0.611 (-2.22)
Log-Dispersion Parameter (ψ)	0.217 (16.75)	-	-

Asymptotic t-statistics are reported in parentheses.

Appendix B

Table B1: Deep Parameter Regression Estimates obtained from the Second Stage of the FO-2SRI GG-CMP, GG-Poisson, LIN-CMP, LIN-Poisson and OLS models When Education is Endogenous

VARIABLE	GG-CMP	GG-Poisson	LIN-CMP	LIN-Poisson	LIV
years_education	-0.512 (-5.53)	-0.423 (-5.54)	-0.008 (-0.14)	-0.006 (-0.14)	-0.038 (-0.43)
mother_age	0.069 (21.30)	0.06 (22.59)	0.054 (53.46)	0.045 (75.51)	0.117 (89.81)
married	-0.020 (-0.07)	-0.01 (-0.06)	0.357 (2.43)	0.300 (2.40)	0.826 (3.03)
married_before	-0.399 (-1.45)	-0.33 (-1.43)	0.025 (0.17)	0.023 (0.19)	0.057 (0.21)
never_married	-3.151 (-11.74)	-3.00 (-13.33)	-3.186 (-18.50)	-3.034 (-20.46)	-0.897 (-3.02)
Hausa	-0.219 (-2.94)	-0.18 (-2.92)	0.160 (2.30)	0.134 (2.30)	0.265 (2.27)
Igbo	-0.093 (-1.79)	-0.08 (-1.79)	-0.065 (-2.16)	-0.054 (-2.15)	-0.105 (-2.12)
Yoruba	-0.061 (-1.60)	-0.05 (-1.62)	-0.075 (-3.15)	-0.063 (-3.16)	-0.054 (-1.33)
Fulani	-0.276 (-2.93)	-0.23 (-2.91)	0.180 (2.01)	0.150 (2.01)	0.310 (2.07)
Kanuri	0.062 (0.81)	0.052 (0.82)	0.215 (2.98)	0.179 (2.98)	0.308 (2.71)
Ijaw	0.180 (3.45)	0.149 (3.44)	0.195 (6.37)	0.161 (6.36)	0.419 (7.67)
Tiv	0.059 (1.16)	0.050 (1.17)	0.030 (0.99)	0.026 (1.00)	0.014 (0.26)
Ibibio	-0.105 (-1.89)	-0.086 (-1.88)	-0.014 (-0.41)	-0.011 (-0.39)	-0.115 (-2.03)
Edo	-0.046 (-0.72)	-0.037 (-0.72)	0.039 (1.07)	0.033 (1.08)	0.062 (1.00)
urban	0.112 (4.73)	0.093 (4.76)	0.062 (3.77)	0.052 (3.80)	0.056 (2.03)
North_Central	0.242 (4.61)	0.200 (4.61)	0.034 (0.98)	0.028 (0.98)	0.063 (1.07)

North_East	0.500 (6.77)	0.413 (6.81)	0.160 (2.69)	0.134 (2.69)	0.243 (2.43)
North_West	0.240 (4.94)	0.198 (4.95)	0.176 (4.78)	0.146 (4.77)	0.322 (5.22)
South_East	0.379 (5.15)	0.314 (5.17)	0.124 (2.61)	0.104 (2.61)	0.177 (2.25)
South_South	0.242 (4.65)	0.200 (4.66)	0.080 (2.18)	0.067 (2.18)	0.157 (2.51)
Television	0.055 (2.66)	0.045 (2.66)	0.001 (0.06)	0.001 (0.07)	-0.004 (-0.19)
Radio	0.095 (3.51)	0.078 (3.50)	-0.020 (-0.84)	-0.017 (-0.85)	-0.042 (-1.04)
wscore	0.447 (4.33)	0.369 (4.33)	-0.126 (-1.49)	-0.105 (-1.49)	-0.198 (-1.40)
$\hat{X}_u^{2SRI}$ **education**	0.485 (5.24)	0.401 (5.25)	-0.020 (-0.37)	-0.016 (-0.37)	-
Constant	4.012 (4.45)	3.231 (4.33)	-0.709 (-1.44)	-0.674 (-1.63)	-1.120 (-1.33)
Log-Dispersion Parameter (ψ)	0.224 (15.61)	-	0.217 (15.20)	-	-

Asymptotic t-statistics are reported in parentheses.

References

Agyeman, A. (2021). Multivariate Analysis of Fertility: An Application of the Generalized Poisson Regression Model. *Statistika: Statistics and Economy Journal, 101*(2), 218-236.

Ali, F. R. M., & Gurmu, S. (2018): "The impact of female education on fertility: a natural experiment from Egypt," *Review of Economics of the Household, 16*(3), 681-712.

Amin, V., & Behrman, J. R. (2014): "Do more-schooled women have fewer children and delay childbearing? Evidence from a sample of US twins," *J Popul Econ, 27*(1), 1-31. doi:10.1007/s00148-013-0470-z

Angrist, J. D., & Krueger, A. B. (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," *The Quarterly Journal of Economics*, 106(4), 979–1014.

Arena, G., Cumming, C., Lizama, N., Mace, H., & Preen, D. B. (2024). Hospital length of stay and readmission after elective surgery: a comparison of current and former smokers with non-smokers. *BMC health services research, 24*(1), 85. doi:10.1186/s12913-024-10566-3

Ashraf, Q. H., Weil, D. N., & Wilde, J. (2013): "The Effect of Fertility Reduction on Economic Growth," *Population and development review, 39*(1), 97-130.

Austin, W. A., & Totaro, M. W. (2011). Gender differences in the effects of Internet usage on high school absenteeism. *The Journal of Socio-Economics, 40*(2), 192-198. doi:https://doi.org/10.1016/j.socec.2010.12.017

Barua, S., El-Basyouny, K., & Islam, M. T. (2014): "A Full Bayesian multivariate count data model of collision severity with spatial correlation," *Analytic Methods in Accident Research, 3-4*, 28-43.

Bansal, P., Krueger, R., & Graham, D. J. (2021): "Fast Bayesian estimation of spatial count data models," *Computational Statistics & Data Analysis, 157*, 107152.

Barbosa, N., Guimarães, P., & Woodward, D. (2004). Foreign firm entry in an open economy: the case of Portugal. *Applied Economics, 36*(5), 465-472. doi:10.1080/00036840410001682160

Bardwell, G. E., & Crow, E. L. (1964): "A Two-Parameter Family of Hyper-Poisson Distributions," *Journal of the American Statistical Association, 59*(305), 133-141.

Blaabæk, E. H., Jæger, M. M., & Molitoris, J. (2020): "Family Size and Educational Attainment: Cousins, Contexts, and Compensation," *European Journal of Population, 36*(3), 575-600.

Blundell, R., Griffith, R., & Windmeijer, F. (2002): "Individual effects and dynamics in count data models," *Journal of Econometrics, 108*(1), 113-131.

Blundell, R., Powell, J.L., 2003: "Endogeneity in nonparametric and semiparametric regression models," In: Dewatripont, M., Hansen, L., Turnovsky, S. (Eds.), Advances in Economics and Econometrics: Theory and Applications, vol. II. Cambridge University Press.

Bolancé, C., Guillén, M., & Pinquet, J. (2008). On the link between credibility and frequency premium. *Insurance: Mathematics and Economics, 43*(2), 209-213. doi:https://doi.org/10.1016/j.insmatheco.2008.05.015

Borsch-Supan, A. (1990). Education and Its Double-Edged Impact on Mobility. *Economics of Education Review, 9*(1), 39-53.

Bratti, M., & Miranda, A. (2011): "Endogenous treatment effects for count data models with endogenous participation or sample selection," *Health Econ, 20*(9), 1090-1109.

Cameron, A. C., and Johansson, P. (1997): "Count Data Regression Using Series Expansions: with Applications," *Journal of Applied Econometrics, 12*(3), 203-223.

Cameron, A. C., Trivedi, P. K., Milne, F., & Piggott, J. (1988). A Microeconometric Model of the Demand for Health Care and Health Insurance in Australia. *The Review of Economic Studies, 55*(1), 85-106. doi:10.2307/2297531

Chi, C. H. (1998): “An event count model for studying health services utilization,” *Medical Care, 36*(12), 1639-1659. doi:10.1097/00005650-199812000-00003

Chuang, H. H.-C., & Oliva, R. (2014). Estimating retail demand with Poisson mixtures and out-of-sample likelihood. *Applied Stochastic Models in Business and Industry, 30*(4), 455-463. doi:https://doi.org/10.1002/asmb.1986

Clarke, H. D., Feigert, F. B., Seldon, B. J., & Stewart, M. C. (1999). More time with my money: Leaving the house and going home in 1992 and 1994. *Political Research Quarterly, 52*(1), 67-85. doi:https://doi.org/10.1177/106591299905200103

Cohen, D., Manuel, D. G., Tugwell, P., Sanmartin, C., & Ramsay, T. (2014). Direct healthcare costs of acute myocardial infarction in Canada’s elderly across the continuum of care. *The Journal of the Economics of Ageing, 3*, 44-49. doi:https://doi.org/10.1016/j.jeoa.2014.05.002

Colón-López, A., & García, C. (2022). 20th Century Puerto Rico and Later-Life Health: The Association Between Multigenerational Education and Chronic Conditions in

Island-Dwelling Older Adults. *Journal of Aging and Health, 35*(1-2), 3-22. doi:10.1177/08982643221097532

Consul, P. C. (1989). Generalized Poisson Distributions : Properties and Applications. *Marcel Dekker Inc*., New York.

Consul, P. C. and Famoye, F. (1992). Generalized Poisson Regression Model. *Comm. Statist. - Theor. & Meth. 2101,* 89-109.

Consul, P. C., & Jain, G. C. (1973). A Generalization of the Poisson Distribution. *Technometrics, 15*, 791-799.

Conway, R. W., & Maxwell, W. L. (1962): "A Queuing Model with State Dependent Service Rates," *Journal of Industrial Engineering, 12*: 132–136.

Costa-Font, J., Jimenez-Martin, S., & Vilaplana, C. (2018): "Does Long-Term Care Subsidization Reduce Hospital Admissions and Utilization?," *Journal of Health Economics, 58*, 43-66.

Costa-Font, J., & Vilaplana-Prieto, C. (2020): "'More than one red herring'? Heterogeneous effects of ageing on health care utilization," *Health Econ, 29*(S1), 8-29.

Coulson, E., Neslusan, C., Stuart, B., and Terza J. (1995): "Estimating the Moral Hazard Effect of Supplemental Medical Insurance in the Demand for Prescription Drugs by the Elderly," *American Economic Review - Papers and Proceedings,* 85, 122-126.

Creel, M., & Farell, M. (2011): "Modelling usage of medical care services: the medical expenditure panel survey data, 1996–2000," *Applied Economics, 43*(18), 2287-2302.

Currie, J., & E. Moretti. (2003) "Mother's education and the intergenerational transmission of human capital: Evidence from college openings," *Quarterly Journal of Economics, 118*(4): 1495–1532.

Dimitrakopoulos, S. (2019): "Bayesian estimation of panel count data models: Dynamics, latent heterogeneity, serial error correlation, and nonparametric structures," In *Panel Data Econometrics: Theory* (pp. 147-173).

Dusheiko, M., & Gravelle, H. (2018): "Choosing and booking—and attending? Impact of an electronic booking system on outpatient referrals and non-attendances," *Health Econ, 27*(2), 357-371.

Dwomoh, D., Agyabeng, K., Tuffour, H. O., Tetteh, A., Godi, A., & Aryeetey, R. (2023). Modeling inequality in access to agricultural productive resources and socioeconomic determinants of household food security in Ghana: a cross-sectional study. *Agricultural and Food Economics, 11*(1), 24. doi:10.1186/s40100-023-00267-6

El-Sayyad, G.M., (1973). "Bayesian and Classical Analysis of Poisson Regression," *Journal of the RQy~LStat;stical Society (Series B), 35*, 445-451.

English, D. R., Hien, T. V. V., & Knuiman, M. W. (2002): “The impact of smoking on use of hospital services: the Busselton study,” *Aust N Z J Public Health, 26*(3), 225-230.

Famoye, F. (1993): “Restricted Generalized Poisson Regression Model,” *Communications in Statistics-Theory and Methods, 22*(5), 1335-1354.

Forthmann, B., Gühne, D., & Doebler, P. (2020): “Revisiting dispersion in count data item response theory models: The Conway–Maxwell–Poisson counts model,” *British Journal of Mathematical and Statistical Psychology, 73*(S1), 32-50.

Fox, J., Klüsener, S., & Myrskylä, M. (2019): “Is a Positive Relationship Between Fertility and Economic Development Emerging at the Sub-National Regional Level? Theoretical Considerations and Evidence from Europe,” *European Journal of Population, 35*(3), 487-518.

Fraser, T. (2020): “Japan’s resilient, renewable cities: how socioeconomics and local policy drive Japan’s renewable energy transition,” *Environmental Politics*, *29*(3), 500-523. doi:10.1080/09644016.2019.1589037

Frome, E.L., Kutner, M.H., and Beauchamp, J.J., (1973). "Regression Analysis of Poisson distributed Oata," *Journal of the American Statistical Association*, *68*, 935-940.

García-Gómez, J., & Parrado, E. (2023). Early Childbearing of Immigrant Women and Their Descendants in Spain. *Population Research and Policy Review, 42*(4), 54. doi:10.1007/s11113-023-09802-1

GOURIEROUX, C., A. MONFORT, AND A. TROGNON: "Pseudo Maximum Likelihood Methods: Application to Poisson Models," *INSEE mimeo*, Paris, 1981.

Gourieroux, C., Monfort, A., & Trognon, A. (1984): “Pseudo Maximum Likelihood Methods: Applications to Poisson Models,” *Econometrica*, *52*(3), 701-720.

Grogger, J. (1990). The Deterrent Effect of Capital Punishment: An Analysis of Daily Homicide Counts. *Journal of the American Statistical Association, 85*(410), 295-303. doi:10.2307/2289764

Güneş, P. M. (2016): “The Impact of Female Education on Teenage Fertility: Evidence from Turkey,” *The B.E. Journal of Economic Analysis & Policy, 16*(1), 259-288.

Hausman, J., Hall, B.H., & Griliches, Z. (1984): "Econometric models for count data with an application to the patents-R&D relationship," *Econometrica 52*, 909-938.

Higuera, L., & Prada, S. I. (2016). Barrier to Access or Cost Share? Coinsurance and Dental-Care Utilization in Colombia. *Applied Health Economics and Health Policy*, *14*(5), 569-578. doi:http://dx.doi.org/10.1007/s40258-016-0251-4

Huang, A. (2017): "Mean-parametrized Conway–Maxwell–Poisson regression models for dispersed counts," *Statistical Modelling, 17*(6), 359-380.

Issahaku, H., Muhammed, M. A., & Abu, B. M. (2023). A count model of financial inclusion in Ghana: evidence from living standards surveys. *Journal of Economics, Finance and Administrative Science, 28*(56), 303-318. doi:10.1108/JEFAS-10-2021-0204

Jorgenson, D.W. (1961): "Multiple regression analysis of a Poisson process," *Journal of the American Statistical Association 56*, 235-245.

Kalist, D. E., & Lee, D. Y. (2016). The National Football League:Does Crime Increase on Game Day? *Journal of Sports Economics, 17*(8), 863-882. doi:10.1177/1527002514554953

Kebede, E., Striessnig, E., & Goujon, A. (2021): “The relative importance of women’s education on fertility desires in sub-Saharan Africa: A multilevel analysis,” *Population Studies*, 1-20.

Khlat, M., Deeb, M., & Courbage, Y. (1997). Fertility Levels and Differentials in Beirut during Wartime: An Indirect Estimation Based on Maternity Registers. *Population Studies, 51*(1), 85-92.

King, G. (1989a): “Event Count Models for International Relations: Generalizations and Applications,” *International Studies Quarterly, 33*(2), 123-147. doi:10.2307/2600534

King, G. (1989b): “Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator,” *American Journal of Political Science*, *33*(3), 762-784.

Krain, M. (1998): “Contemporary Democracies Revisited: Democracy, Political Violence, and Event Count Models,” *Comparative Political Studies, 31*(2), 139-164. doi:10.1177/0010414098031002001

Kumar, T.K., and Shih, W.P., (1978). "An Application of a Multiple Regression Model of a Poisson Process to the Murder Supply Equation," *American Statistical Association Proceedings of the Business and Economics Statistics Section 715-720.*

Landersø, R., & Fallesen, P. (2021): "Psychiatric Hospital Admission and Later Crime, Mental Health, and Labor Market Outcomes," *Health Econ*, 30(1), 165-179.

Lee, J. (2008): "Sibling Size and Investment in Children's Education: An Asian Instrument," *Journal of Population Economics, 21*(4), 855-875.

Lippi Bruni, M., Mammi, I., & Ugolini, C. (2016): "Does the extension of primary care practice opening hours reduce the use of emergency services?," *Journal of Health Economics, 50,* 144-155.

Mahamunulu, D.M., (1967). "A Note on Regression in the Multivariate Poisson Distribution," *Journal of the American Statistical Association, 58,* 251-258.

Mainardi, S. (2003): "Testing convergence in life expectancies: count regression models on panel data," *Prague Economic Papers, 12*(4), 350-370.

Malyshkina, N. V., Mannering, F. L., & Tarko, A. P. (2009): "Markov switching negative binomial models: An application to vehicle accident frequencies," *Accident Analysis & Prevention, 41*(2), 217-226.

Manning, W.G., Basu, A.,and Mullahy, J. (2005): "Generalized Modeling Approaches to Risk Adjustment of Skewed Outcomes Data," *Journal of Health Economics*,20,465–488.

McCarthy, P. (2003). Alcohol-Related Crashes and Alcohol Availability in Grass-Roots Communities. *Applied Economics, 35*(11), 1331-1338.

Michener, R., & Tighe, C. (1992). A Poisson Regression Model of Highway Fatalities. *American Economic Review, 82*(2), 452-456.

Mullahy, J. (1997): "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior," *Review of Economics and Statistics, 79*, 586-593.

Neelon, B. H., O'Malley, A. J., & Normand, S.-L. T. (2010): "A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use," *Statistical Modelling, 10*(4), 421-439.

Nelder JA, & Wedderburn RWM,. (1972): "Generalized Linear Models," Journal of the Royal Statistical Society A, 135, 370–384.

Olayiwola, S. O., & Kazeem, B. L. O. (2019). Count data modelling of health insurance and health care utilisation in Nigeria. *Journal of Economics and Management, 35*(1), 106-123. doi:doi:10.22367/jem.2019.35.06

Osili, U. O., & B. T. Long. (2008): "Does female schooling reduce fertility? Evidence from Nigeria," *Journal of Development Economics 87*(1): 57–75.

Oyenubi, A., & Kollamparambil, U. (2022): "Does the child support grant incentivise childbirth in South Africa?," *Economic Analysis and Policy, 73*, 812-825. doi:10.1016/j.eap.2022.01.005

Polasik, M., Huterska, A., Iftikhar, R., & Mikula, Š. (2020). The impact of Payment Services Directive 2 on the PayTech sector development in Europe. *Journal of Economic Behavior & Organization, 178*, 385-401. doi:https://doi.org/10.1016/j.jebo.2020.07.010

Ransom, M. R., & Pope, C. A., III. (1995). External Health Costs of a Steel Mill. *Contemporary Economic Policy, 13*(2), 86-97.

Romeu, A., & Vera-Hernández, M. (2005), "Counts with an endogenous binary regressor: A series expansion approach," *The Econometrics Journal, 8*(1), 1-22. http://www.jstor.org/stable/23114964

Ross, S. M. (1997): “Simulation,” San Diego: Academic Press.

Sáez-Castillo, A. J., & Conde-Sánchez, A. (2013): “A hyper-Poisson regression model for overdispersed and underdispersed count data,” *Computational Statistics & Data Analysis*, *61*, 148-157.

Santos Silva, J. M. C., & Covas, F. (2000). A modified hurdle model for completed fertility. *J Popul Econ, 13*(2), 173-188. doi:10.1007/s001480050132

Schellhorn, M. (2001). The effect of variable health insurance deductibles on the demand for physician visits. *Health Econ, 10*(5), 441-456. doi:https://doi.org/10.1002/hec.630

Schuettig, W., & Sundmacher, L. (2022). The Impact of Ambulatory Care Spending, Continuity and Processes of Care on Ambulatory Care Sensitive Hospitalizations. *European Journal of Health Economics, 23*(8), 1329-1340. doi:https://doi.org/10.1007/s10198-022-01428-y

Schwartz, E. S., & Torous, W. N. (1993). Mortgage Prepayment and Default Decisions: A Poisson Regression Approach. *American Real Estate and Urban Economics Association Journal, 21*(4), 431-449.

Sellers, K. F. (2023): "The Conway–Maxwell–Poisson Distribution," Cambridge: Cambridge University Press.

Sellers, K. F., & Shmueli, G. (2010): "A flexible regression model for count data," *The Annals of Applied Statistics, 4*(2), 943-961, 919.

Serrano-Alarcón, M., Hernández-Pizarro, H., López-Casasnovas, G., & Nicodemo, C. (2022): "Effects of long-term care benefits on healthcare utilization in Catalonia," *Journal of Health Economics, 84*, 102645.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., & Boatwright, P. (2005): "A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution," *Journal of the Royal Statistical Society: Series C (Applied Statistics), 54*(1), 127-142.

Sobotka, F., Radice, R., Marra, G., & Kneib, T. (2013): "Estimating the relationship between women's education and fertility in Botswana by using an instrumental variable approach to semiparametric expectile regression," *Journal of the Royal Statistical Society: Series C (Applied Statistics), 62*(1), 25-45. doi:https://doi.org/10.1111/j.1467-9876.2012.01050.x

Soltani, M., Batt, R. J., Bavafa, H., & Patterson, B. W. (2022): "Does What Happens in the ED Stay in the ED? The Effects of Emergency Department Physician Workload on

Post-ED Care Use," *Manuf Serv Oper Manag, 24*(6), 3079-3098. doi:10.1287/msom.2022.1110

Stephenson, J., Newman, K., & Mayhew, S. (2010): "Population dynamics and climate change: what are the links?," *Journal of Public Health, 32*(2), 150-156.

Sulemana Abdul-Salam, Shei Sayibu Baba & Haruna Jabir. (2018): "The Impact of Mothers Education on Fertility in Ghana," *International Journal of Probability and Statistics, 7*(2), 31-43. doi:10.5923/j.ijps.20180702.01

Terza, J.V. (1998): "Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects," *Journal of Econometrics, 84*(1), 129-154.
Doi: https://doi.org/10.1016/S0304-4076(97)00082-1

Terza, J. V. (2006): "Estimation of policy effects using parametric nonlinear models: a contextual critique of the generalized method of moments," *Health Services and Outcomes Research Methodology, 6*(3), 177-198. doi:10.1007/s10742-006-0013-0

Terza, J. V. (2016a): "Simpler standard errors for two-stage optimization estimators," *Stata Journal* 16: 368–385.

Terza, J. V. (2016b): "Inference Using Sample Means of Parametric Nonlinear Data

Transformations," *Health Services Research, 51*, 1109-1113.

Terza, J. V. (2017). Causal Effect Estimation and Inference Using Stata. *The Stata Journal, 17*(4), 939-961. doi:10.1177/1536867x1801700410

Terza, J. V. (2020): "Regression-Based Causal Analysis from the Potential Outcomes Perspective," *Journal of Econometric Methods,* 9, published online, DOI: https://doi.org/10.1515/jem-2018-0030.

Terza, J.V. (2023a): "Simpler Standard Errors for Two-Stage Optimization Estimators Revisited," *Stata Journal*, 23, 1057-1061.

Terza, J.V. (2023b): "Supplementary Appendix to 'Simpler Standard Errors for Two-Stage Optimization Estimators Revisited'," *Stata Journal*, 23, forthcoming online.

Terza, J.V. (2024a): "Introducing Ph.D. Students to Empirical Causal Analysis in the Context of the Simple Linear Regression Model," Unpublished manuscript, Department of Economics, Indiana University School of Liberal Arts at IUPUI.

Terza, J. V. (2024b): "Empirical Causal Analysis Based on First-Order Residual Inclusion Regression: Specification, Identification and Estimation from a Potential Outcomes Perspective," Unpublished manuscript, Department of Economics, Indiana University School of Liberal Arts at IUPUI.

Terza, J. V. (2024c): “Standard Errors for Regression-Based Causal Effect Estimates in Economics Using Numerical Derivatives,” *Computational Economics*, published online.

Terza, J. V., Basu, A., & Rathouz, P. J. (2008): “Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling,” *J Health Econ, 27*(3), 531-543. doi:10.1016/j.jhealeco.2007.09.009

Vera-Hernández, Á. M. (1999). Duplicate coverage and demand for health care. The case of Catalonia. *Health Econ, 8*(7), 579-598. doi:https://doi.org/10.1002/(SICI)1099-1050(199911)8:7<579::AID-HEC478>3.0.CO;2-P

Weber, D.C., (1971). "Accident Rate Potential: An Application of Multiple Regression Analysis of a Poisson Process," *Journal of the American Statistical Association, 66*, 431-441.

Wang, W., & Famoye, F. (1997). Modeling Household Fertility Decisions with Generalized Poisson Regression. *J Popul Econ, 10*(3), 273-283.

Weiss, C. R., & Wittkopp, A. (2005). Retailer Concentration and Product Innovation in Food Manufacturing. *European Review of Agricultural Economics, 32*(2), 219-244.

Windmeijer, F. A. G., & Santos Silva, J. M. C. (1997): "Endogeneity in count data models: An application to demand for health care," *Journal of Applied Econometrics, 12*(3), 281-294.

Winkelmann, R. (1995). Duration Dependence and Dispersion in Count-Data Models. *Journal of Business & Economic Statistics, 13*(4), 467-474. doi:10.2307/1392392

Winkelmann, R. (2008): Bayesian Analysis of Count Data. In Econometric Analysis of Count Data (pp. 241-250). Berlin, Heidelberg: Springer Berlin Heidelberg.

Winkelmann, R., & Zimmermann, K. F. (1991): "A New Approach for Modeling Economic Count Data," *Economics Letters, 37*, 139-143.

Winkelmann, R., & Zimmermann, K. F. (1994): "Count data models for demographic data," *Mathematical Population Studies, 4*(3), 205-221.

Wooldridge, J. M. (2014): "Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables." *Journal of Econometrics, 182*(1), 226-234. doi:https://doi.org/10.1016/j.jeconom.2014.04.020

Wooldridge, J. M. (2015). Control Function Methods in Applied Econometrics. *The Journal of Human Resources, 50*(2), 420-445.

Xue-Dong, C. (2009): “Bayesian analysis of semiparametric mixed-effects models for zero-inflated count data,” *Communications in Statistics - Theory and Methods, 38*(11), 1815-1833.

Yan, X. C., Wang, T., Chen, J., Ye, X. F., Yang, Z., & Bai, H. (2019): “Analysis of the Characteristics and Number of Bicycle-Passenger Conflicts at Bus Stops for Improving Safety,” *SUSTAINABILITY, 11*(19). doi:10.3390/su11195263

Yu, P. L. H., Chan, J. S. K., & Fung, W. K. (2006). Statistical Exploration from SARS. *The American Statistician, 60*(1), 81-91. Retrieved from http://www.jstor.org.proxy.ulib.uits.iu.edu/stable/27643734

Zanin, L., Radice, R., & Marra, G. (2014). A comparison of approaches for estimating the effect of women's education on the probability of using modern contraceptive methods in Malawi. *The Social Science Journal*, *51*(3), 361-367. doi:10.1016/j.soscij.2013.12.008

CURRICULUM VITAE

Golnoush Kazeminezhad

Education

- PhD Economics, Indiana University, earned at Indiana University-Purdue University Indianapolis (IUPUI), 2024
- B.Sc. Economics, University of Tehran, 2012

Honors, Awards, Fellowships Research and Training Experience

- Graduate Student Travel Fellowship Award, IUPUI, 2024
- Featured in IUPUI Student Success Webpage, 2024
- Carlin Award for Outstanding Graduate Student Paper in Empirical Economics, IUPUI, 2023
- Robert B. Harris Graduate Teaching Recognition Scholarship in Economics, IUPUI, 2022
- Full Tuition Scholarship and Graduate Assistantship, IUPUI, 2017-2022

Professional Experience

- Instructor, IUPUI, 2020-2023
- Research Assistant, Kittle Property Group Inc, 2017-2020
- Intern, CourseNetworking LLC/IUPUI Cyberlab, 2016-2017
- Economic Consultant, Secure Business Founders Ltd, 2012-2014

Teaching Experience

- Instructor, Undergraduate Statistics, IUPUI, Fall 2020, Spring 2022, Spring 2023 Summer 2023
- Instructor, Undergraduate Microeconomics, IUPUI, Fall 2022, Spring 2023

- Instructor, Statistics Boot Camp for Economics Ph.D. Students, IUPUI, Summer 2021
- Teaching Assistant, Graduate Econometrics, IUPUI, 2020-2021
- Tutor, Undergraduate Statistics, IUPUI, Fall 2022

Conferences Attended

- Allied Social Science Associations/American Economic Association (ASSA/AEA), 2024
- American Society of Health Economists (ASHEcon), 2023
- American Public Health Association (APHA), 2022
- 21st Annual International Conference on Health Economics, Management and Policy, Athens, Greece, 2022
- Stata Conference, 2021
- Health Economics Seminar, Department of Economics at IUPUI, 2019 and 2020

Working Papers

- Kazeminezhad, G. and Terza, J.V.: ”Count-Regression-Based First-Order Two-Stage Residual Inclusion Estimation of a Continuous Causal Effect: The Impact of Schooling on Fertility”
- Terza, J.V. and Kazeminezhad, G.: ”Two-Part Models and Causal Inference: Simpler Standard Errors for the Two-Stage Residual Inclusion Estimator Using Numerical Derivatives”