EXPLORING THE IMPORTANCE OF ACCOUNTING FOR NONLINEARITY IN CORRELATED COUNT REGRESSION SYSTEMS FROM THE PERSPECTIVE OF CAUSAL ESTIMATION AND INFERENCE

Yilei Zhang

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

July 2021

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

Doctoral Committee

Joseph V. Terza, Ph.D., Chair

Joshua R. Vest, Ph.D., MPH

June 1, 2021

Wendy Morrison, Ph.D.

Sumedha Gupta, Ph.D.

DEDICATION

For my husband James, my son Christopher, and my parents, Chen Liexiu and Zhang Lin.

ACKNOWLEDGEMENT

This dissertation has been done under supervision of my advisor and dissertation committee chair, Dr. Joseph Terza. I am grateful to learn from and work with Dr. Terza, who played critical role in guiding me and inspiring me to navigate in the research field that we are passionate in. He also helped me to build a solid foundation and help me to equip with the tools needed to start and thrive in my professional career. Being a student of Dr. Terza is the best thing happened in my entire doctoral study, as one cannot find a mentor who cares about his or her students more.

I am indebted to the rest of my dissertation committee: Dr. Joshua Vest, Dr. Wendy Morrison, and Dr. Sumedha Gupta for their helpful insights that improved the dissertation. I am especially grateful to Dr. Vest, who has been always supportive in my application of research datasets by sharing his resources with me.

I would like to also say thank you to other faculty and staff members in the Department of Economics. My special appreciation goes to Dr. Steven Russell, Professor and Chair, who is always there for me whenever I face any challenges or questions as a junior instructor, I am grateful that Dr. Russell he has been supportive and accommodating to me as a new mother, particularly during the pandemic year. My sincere thanks also go to Dr Ann Royalty, Dr. Kosali Simon, Dr. Thuy Nguyen, Dr. Dan Sacks, Dr. Henry Mak, Dr. Jaesoo Kim, Dr. Vidhura Tennekoon, Dr. Mark Ottoni-Wilhelm, Dr. Subir Chakrabarti, for their supports and helps in my coursework and research.

I am also thankful to Terri Crews and Dana Ward for their help with various administrative supportive. They are the experts, and I am always well informed of what to do in each step of my study.

I would like to thank my loving and supportive husband, James Fisher, who has reached beyond the role of a husband and spent countless hours helping me professionally. I would like to thank my son, Christopher who is the sweetest baby. He gives me motivation to work hard and excel. Last but not least, I would also like to thank my mother Chen Liexiu, and my father Zhang Lin who have played essential roles in my success by being a life mentor and offering helps taking care of my young son. They are the heroes.

Yilei Zhang

EXPLORING THE IMPORTANCE OF ACCOUNTING FOR NONLINEARITY IN CORRELATED COUNT REGRESSION SYSTEMS FROM THE PERSPECTIVE OF CAUSAL ESTIMATION AND INFERENCE

The main motivation for nearly all empirical economic research is to provide scientific evidence that can be used to assess causal relationships of interest. Essential to such assessments is the rigorous specification and accurate estimation of parameters that characterize the causal relationship between a presumed causal variable of interest, whose value is to be set and altered in the context of a relevant counterfactual and a designated outcome of interest. Relationships of this type are typically characterized by an effect parameter (EP) and estimation of the EP is the objective of the empirical analysis. The present research focuses on cases in which the regression outcome of interest is a vector that has count-valued elements (i.e., the model under consideration comprises a multi-equation system of equations). This research examines the importance of account for nonlinearity and cross-equation correlations in correlated count regression systems from the perspective of causal estimation and inference.

We evaluate the efficiency and accuracy gains of estimating bivariate count valued systems-of-equations models by comparing three pairs of models: (1) Zellner's Seemingly Unrelated Regression (SUR) versus Count-Outcome SUR - Conway Maxwell Poisson (CMP); (2) CMP SUR versus Single-Equation CMP Approach; (3) CMP SUR versus Poisson SUR.

We show via simulation studies that it is more efficient to estimate jointly than equation-by-equation, it is more efficient to account for nonlinearity. We also apply our

model and estimation method to real-world health care utilization data, where the dependent variables are correlated counts: count of physician office-visits, and count of non-physician health professional office-visits. The presumed causal variable is private health insurance status. Our model results in a reduction of at least 30% in standard errors for key policy EP (e.g., Average Incremental Effect). Our results are enabled by our development of a Stata program for approximating two-dimensional integrals via Gauss-Legendre Quadrature.

Joseph V. Terza, Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIE: Average Incremental Effects

ATE: Average Treatment Effects

MLE: Maximum Likelihood Estimation

PE: Policy Effects

EP: Effect Parameters

CDCR: Correlated Dispersion Flexible Count Regression

DGP: Data Generating Process

CPOM: Conditional Potential Outcomes Model

PO: Potential Outcome

FP: Fully Parametric

MP: Minimally Parametric

PMF: Probability Mass Function

PDF: Probability Density Function

GPOF: General Potential Outcome Framework

VAR: Variance

AVAR: Asymptotic variance

CMP: Conway-Maxwell-Poisson

SUR: Seemingly Unrelated Regression

CM: Conditional Mean

CV: Conditional Variance

GLQ: Gauss Legendre Quadrature

2D: Two Dimensional

AAPB: Average of the Absolute Percentage Bias

Chapter 1: Introduction

1.1: Overview

The main motivation for nearly all empirical economic research is to provide scientific evidence that can be used to assess causal relationships of interest. Essential to such assessments is the rigorous specification and accurate estimation of parameters that characterize the causal relationship between a presumed causal variable of interest, whose value is to be set and altered in the context of a relevant counterfactual and a designated outcome of interest. Relationships of this type are typically characterized by an effect parameter (EP) and estimation of the EP is the objective of the empirical analysis. The present research focuses on cases in which the regression outcome of interest is a vector that has count-valued elements (i.e., the model under consideration comprises a multi-equation system of equations). This dissertation examines the importance of account for nonlinearity and cross-equation correlations in correlated count regression systems from the perspective of causal estimation and inference.

1.2: Literature in Application

Correlated multivariate count outcomes are common in many areas, including:[1] biology – e.g., counts of RNA sequences (see, e.g., Zhang et. al, 2017), where they

---

[1] There are many other applications of the correlated count regression models in various fields. Examples from health economics include Atella et al. (2008), Cameron, Trivedi et al. (2004). Dong et al. (2011), Serhiyenko et al.(2018), Famoye (2010a), Famoye (2010b), Famoye (2015), Winkelmann (2000), Winkelmann (2012). Examples from other disciplines such as forest science include: Affleck et al. (2016), Fu et al. (2017), Ma et al.

examined regression models for multivariate counts with flexible mean–covariance and correlation structure using the counts data of DNA or RNA fragments within a genomic interval; civil engineering – e.g., crash related morbidity and mortality data (see, e.g., Ma et al., 2008), where they offered a multivariate Poisson-lognormal specification that simultaneously models crash counts by injury severity; and healthcare utilization metrics (see, e.g., Chib and Winkelmann, 2001); healthcare resources and costs: Mihaylova, Briggs, O'Hagan, & Thompson (2011) conducted a systematic review on models analyzing healthcare resources and costs. They suggested that future work should focus on using mixture models that account for correlation structures, specifically: "A major limitation of the implementation of more complicated models in the field of randomized trials is the need for the analytical framework to accommodate both costs and health effects and evaluate the summary cost-effectiveness measures. In doing so, the analysis should allow for the correlation structure of different outcomes. The future development of such approaches in different situations is recommended, perhaps especially for two-part models or mixture models (Mihaylova, Briggs, O'Hagan, & Thompson, 2011)."

(2015), Mehta (2014), Parresol et al. (2001), Vonderach et al. (2018), and also in transportation science: Afghari et al. (2018), Lord et al. (2008), Sfeir et al. (2020), Zamzuri (2016), in finance and risk analysis: Chua et al (2019), McElroy et al. (1988), Guikema et al. (2008), in pharmacological and biological research: Leung et al (1992), Gueorguieva (2001), and in sociology research: King (1989).

1.3: Literature in Methodology

The methodological work that has been done on correlated count regression model can be traced back to a few seminal works on Seemingly Unrelated Regression (SUR). Zellner (1962) is the first to propose a SUR framework where he introduces the linear seemingly unrelated regression model and showed efficiency gains of the SUR approach, Gallant (1973) extended Zellner's framework to non-linear case.

A strand of literature investigates multivariate count models (Gourieroux and Monfort (1984), Hausman et al. (1984), King (1989) and Winkelmann (2000)]. Gourieroux and Monfort (1984) specified a bivariate Poisson model using the nonlinear Least Square method to avoid the problems involved in integral computation related to maximum likelihood estimation. Gourieroux and Monfort (1984) specified a mixture model and analytically derived the likelihood function by assuming errors' to be gamma distributed, they eventually arrived at negative binomial model under certain regularity assumptions; note that their mixture model is based on a multiplicative structural errors that only allows for overdispersion via a parameter that also controls the structural errors correlation. King (1989) proposed a bivariate Poisson model with Poisson errors. Winkelmann (2000) gave a weighted least square estimator with Negative Binomial specification. His model improved upon the previous ones by allowing for separate estimation of correlation and dispersion, but it still only allows for over-dispersion. A common limitation of these models is that they are restricted in the covariances they can accommodate.

Researchers have attempted to come with up models with less restricted or un-restricted covariance. A few recent works include Famoye (2015) and Andreassen &

Jensen (2018). Famoye (2015) proposed a generalized restricted Poisson model, while Andreassen & Jensen (2018) specified the fully parametric model by deriving a three-dimensional probability function based on Negative Binomial distributed margins. This model relies on maximum likelihood estimation (MLE), and it is shown to be almost computationally infeasible when the dimension of the outcomes goes higher than four.

Historically, Aitchison and Ho (1989) offered the first general mixture model, where the conditional probability mass function's integrand component is specified with a marginal term and a cross-equation heterogeneity term. In this model, the cross-equation heterogeneity term is integrated out to get the marginal distribution of the outcomes given the covariates. This type of model is easily generalizable, while allowing for heteroskedasticity and a flexible cross-equation distribution, which enables greater generality on the outcomes' covariance structure than prior work. Since our model is a type of mixture model, it inherits these class properties.

As previously discussed, a common problem with mixture models is that the likelihood functions required for MLE involve high dimension integrals that are difficult to calculate. The related literature takes several approaches towards addressing the problem. At one end are studies, like Aitchison and Ho (1989), which leverage quadrature or simulation methods to directly compute the integrals and estimate the model. At the other end are studies that attempt to circumvent the problem. Such studies broadly fall into two camps: Bayesian methods and frequentist methods.

Bayesian approaches are frequently able to avoid the direct computation of the likelihood and instead employ simulation methods – e.g., Metropolis Hasting or Markov Chain Monte Carlo – to derive their posterior distributions and calculate the relevant expectations. Examples in this vein include Ma et al. (2008), Mehta (2014), and Chua & Tsiaplias (2019). Ma et al. (2008) specified multivariate Poisson-lognormal likelihood function and the estimation is achieved by Bayesian methods. Similarly, Mehta (2014) specified multivariate Conway-Maxwell-Poisson marginal with multivariate normal cross-equation heterogeneity, the estimation is also achieved by Bayesian methods. Chua & Tsiaplias (2019) proposed a lognormal mixture model with both quadrature method and Expectation Maximization (EM) method for integration approximation.

Frequentist have also sought to avoid the integration problem and typically do so by either (i) transforming count-outcomes into continuous or (ii) treating count-outcomes as though they were continuous, and thereby approximating them. Particularly interesting examples of (i) are Parresol (2001) and Vonderach (2018), who are interested in forest size and characteristics. Rather than using the number of trees in a forest as the outcome, they transform the problem instead use the total mass of all trees in the forest, which is a continuous response variable. An example of (ii) occurs whenever one estimates a linear regression for a count outcome.

We close by observing that the literature on Seemingly Unrelated Regressions (SUR) is the key inspiration for the core conjecture of this paper – that joint estimation is more efficient than equation-by-equation estimation. This additional efficiency was first observed by Zellner (1962) in systems of linear equations estimation for continuous outcome variables and it was extended to systems of non-linear equations estimation for continuous outcome variables by Gallant (1973). A key restriction of both of these seminal works is that they assume normal, additively separable errors. McElroy and Burmeister (1988) subsequently showed that this intuition also applies to non-normal errors by non-linear weighted least squares. This approach has been subsequently generalized and applied in a variety of studies – e.g., Delgado (1992), Parresol (2001), Marshall (2003), and Vonderach (2018).

1.4: Objectives

The statistics and econometric literature focused on causal inference in such outcomes does not typically examine the efficiency gains of linear SUR versus nonlinear SUR models, and joint estimation versus equation-by-equation estimation. Inspired by Zellner and Gallant's work where they demonstrate significant efficiency gains to joint estimation when the outcome variables are specified as correlated, continuous function of the parameters and covariates. We thus conjecture that similar efficiency gains also obtain when the outcomes are not-continuous, but rather correlated count-valued.
The primary objective of the study is to produce evidence that is informative with respect to answering the following question: "Is the analytic and programming effort needed to account for nonlinearity in such models 'worth it' relative to the (i) conventional linear specification and estimation approach, the (ii) single-isolated-equation estimation approach, and the (iii) specification that is absent of dispersion flexibility?"

To investigate this family of questions, we introduce a general (parametric) bivariate count model, wherein the joint distribution of the outcome variables is decomposed into two marginal distributions and a bivariate normal distribution, which links the two marginals together. We then employ simulation studies to compare the efficiency of: (1) Zellner's linear SUR model versus our Count-Outcome SUR model in terms of the policy effect parameters – either the average incremental effects (AIE, or average treatment effects (ATE) -- (2) joint estimation versus equation-by-equation estimation, (3) Count-Outcome SUR model that does not account for dispersion flexibility versus dispersion flexible Count-Outcome SUR; we are particularly interested in the accuracy and efficiency gains of treatment effect measures (AIE or ATE).

1.5: Main Findings and Significance

We show via simulation studies that it is typically more efficient and accurate to estimate these models using our correlated count regression system than it is to estimate them using linear seemingly unrelated regression models developed by Zellner (1962): the percent bias of effect parameter is reduced by at least 89%, and the averaged absolute percent bias (AAPB) of our effect parameter is reduced by at least 49% in the case of Conway-Maxwell-Poisson marginals.

We find that there are significant efficiency gains to joint estimation in the case of Poisson marginals. For instance, at certain parameter values, joint estimation results in standard errors for the Average Incremental Effect that are at least 40% smaller than those resulting from equation-by-equation estimation. In the case of Conway-Maxwell-Poisson (CMP) marginals, our findings are mixed -- there are certain parameter values with no efficiency gains attributable to joint estimation and others with small efficiency gains attributable to joint estimation. The general finding, however, is clear: there are typically efficiency gains to joint estimation.

We also find that accounting for dispersion flexibility matters. Our model with dispersion flexible marginals performs with more accuracy and efficiency in terms of effect parameters (AIEs) than the model with Poisson marginals (no flexibility in dispersion). For instance, when the cross-equation correlation rho = 0.75, the percent bias of effect parameter is reduced by 35%, and the averaged absolute percent bias (AAPB) of our effect parameter is reduced by 30% by introducing dispersion parameters to our model.

To explore the ramifications of this finding, we consider an application to health care, wherein a question of substantive interest is the extent to which the use of health services depends on insurance coverage (Chib and Winkelmann, 2001). Specifically, we employ data from the 1987 National Medical Expenditure Survey Data to estimate the policy effect of private health insurance status on two correlated measures of healthcare utilization: the number of doctor office visits; and the number of non-doctor office visits. Estimation of our joint model using Poisson marginals gives that the estimated ATE of having private insurance on the number of doctor office visits is 1.88 and the ATE of having private insurance on the number of non-doctor office visits is 4.00. Further, we find that joint estimation of our model results in at least a 30% reduction in the standard errors of the ATE as compared to equation-by-equation estimations.

Our model is estimated via MLE. Yet, our likelihood function contains several improper integrals, which do not admit closed forms. To surmount the associated technical challenges, we develop a Stata program for numerically approximating any two-dimensional (improper) integrals via Gauss-Legendre Quadrature methods. Our validation work shows the program to be fast and accurate, and we leverage the program as part of our estimation routine. The program will soon be available as a Stata package.

Our work contributes to a growing literature that explores systems of equations models with count-valued outcomes -- e.g., Aitchison and Ho (1989), Famoye (2015), Gourieroux and Monfort (1984), and King (1989), among others. These papers, however, make a key assumption: joint estimation is more efficient than equation-by-equation estimation. Our paper is the first, to our knowledge, to explicitly investigate and verify this assumption, with findings of large efficiency gains; we also relax certain key assumptions present in these earlier works and provide a new, more flexible numerical integration methods. Since commonly available estimation packages (e.g., in Stata) only offer equation-by-equation count-value models, a broader implication of our work is that many practitioners are leveraging larger standard errors than necessary.

1.6: Organization of the Dissertation

The balance of this dissertation is organized as follows. Section 2 introduces the model and describes its estimation, including the computation of the average incremental effect in the General Potential Outcomes Framework; Chapter 3 proposes the estimator for the deep parameters and the AIE in the seemingly unrelated count regressions context using Poisson marginals and Conway-Maxwell-Poisson marginals as examples. Chapter 4 covers our Gauss-Legendre Quadrature program and its validation. Section 5 presents the simulation study results exploring the importance accounting for nonlinearity, cross-equation correlation, and dispersion flexibility in count-outcome SUR Models. Section 6 discusses the application to healthcare utilization data. Section 7 concludes and discusses limitations and future work.

Chapter 2: Potential-Outcomes-Based Causal Effect Specification, Identification, Estimation, and Inference In the Correlated Count Regressions Context

In this chapter, the goal is to specify and estimate the causal effect of a presumed policy variable on a count-valued outcome using correlated dispersion-flexible count regression (CDCR) models to accommodate potential cross-equation correlations and dispersion flexibility. I begin with a general review of the PO framework as discussed in Terza (2019a). After detailing relevant concepts within the PO framework, I will use the concepts to specify the AIE in correlated count data context using the newly developed Correlated Dispersion Flexible Count Regression model, which I will discuss briefly in this chapter and detail (specifying their probability mass functions, conditional mean functions and the policy effects and the standard errors of the policy effects) in the next chapter. As already noted, and as it will be clearer later, casting the specification and estimation of the AIE is important to make conditions required for causality explicit.

2.1: Specification of the Treatment Effect of Interest in the Potential Outcomes Framework

The overarching objective of virtually every research study in empirical health economics, health services research and health policy analysis is to provide scientific evidence that can be used to assess the causal relationship between a X, whose value is to be exogenously set and altered in the context of the relevant counterfactual, and a designated outcome of interest Y.[2] For this reason, we take the general potential

[2]Henceforth, the symbol **Y** is used as the global replacement for the phrase "outcome of interest."

outcomes framework (GPOF) detailed by Terza (2019a) as the appropriate setting for our discussion of the specification, identification and estimation of a policy-relevant EP based on the correlated dispersion flexible count regression model. In the GPOF, the distinction is drawn between two versions 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 are drawn.[3]

and

> X^* ≡ the random variable representing a hypothetical (*counterfactual*) exogenously mandated version of the X; it is not an element of the DGP.

There are, likewise, two versions of the Y:

> Y ≡ the random variable representing the factual version of the distribution of the Y; Y is the element of the DGP from which the sampled values of the Y are drawn.

and

[3] The DGP is defined as the joint distribution from which data on the observable data can be drawn.

Y_{X^*} ≡ the random variable representing the distribution of *potential outcomes*, defined as the distribution of values of the Y that would have manifested for a particular X^*, which is not an element of the DGP.

In the present study, we focus on specification, identification, estimation and inference for a causal effect parameter when the Y is a vector of count variables that are potentially correlated. For example, in an illustration discussed later, we consider the case in which the Y is

$$\begin{pmatrix} \text{Number of phyisican offfice visits} \\ \text{Number of non-phyisican offfice visits} \end{pmatrix}$$

and the X is binary variable that represents the private health insurance status of a patient. Causal analysis and inference in the GPOF are based on a counterfactual in which:

I: a pre-counterfactual version of X^* is set ($X^* = X^{pre}$), to which there corresponds a potential outcome ($Y_{X^*} = Y_{X^{pre}}$);

and

II: as a "what if" thought experiment, X^{pre} is incremented by Δ so that the post-counterfactual version of X^* is $X^{post} = X^{pre} + \Delta$, to which there also corresponds a potential outcome ($Y_{X^{post}} = Y_{X^{pre}+\Delta}$).

The relevant EP is typically formulated as a function of the moments of $Y_{X^{pre}+\Delta}$ and $Y_{X^{pre}+\Delta}$. In the remainder of the discussion, we will focus on the following EP, the *average incremental effect*, which is

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

Where $\Delta = 1 \equiv$ an incremental change to the pre-counterfactual status of private insurance for each induvial in the population. In our empirical example of the policy effects of private health insurance on health care utilizations, $\Delta_i = 1$ means individual i counterfactually switch from having no private insurance to having private insurance. On the contrary, $\Delta_j = -1$ means individual j counterfactually switch from having private insurance to having no private insurance. Note that when the policy variable is binary and $\Delta_j = -1$, the AIE reverts to the ATE.

Using the GPOF in this way to specify the EP based on a policy relevant counterfactual (e.g., (1) or (2)] ensures that this parameter is indeed causally interpretable – the change in the potential outcome is exclusively attributable to the imposed counterfactual change in the X. Moreover, it serves to clearly and rigorously establish the EP as the estimation objective at the outset.

There is a fundamental disconnect between the inherently counterfactual targeted EP and the relevant DGP – the latter produces observable data, while the former is based on random variables from which data cannot be sampled (for the most part). This raises two questions:

How do we establish statistical identification of the EP, given that such identification must be based on the relevant DGP?

How do we consistently estimate the inherently counterfactual (unobservable) targeted EP using only DGP-produced factual (observable) sample data?

We explore answers to these questions in the context of the version of the GPOF in which the Y is fractional. Let us consider the case in which the X is exogenous.

2.2: Fully Parametric (FP) Specification of Models in the Potential Outcomes Framework

Effect parameters like AIEs and ATEs are not directly estimable from data because X^{pre} and $Y_{X^{pre}}$ are counterfactual – they do not represent observable statistical populations from which samples can be drawn. As discussed by Terza (2019a and 2019b), this gap between the estimation objective (the inherently counterfactual EP) and the observable data (the data generating process) can be bridged via parametric specification of the conditional probability distributions of the relevant potential outcome (Y_{X^*}) given a vector of control covariates (X_o). To this end, following the approach proposed by Terza (2019a), I posit the following fully parametric structural specification for the distribution of ($Y_{X^*}|X_o$):

$$\mathrm{pmf}(Y_{jX^*} \mid X_o)=f_{(Y_j \mid X_{jo})}(Y_j, X, X_{jo}; \pi) \qquad (2)$$

where $\text{pmf}(Y_{jX^*} \mid X_o)$ is the conditional probability mass function of Y_{X^*} given X_o, for j = 1, 2 and where the "deep" parameters of the model are π.

From the pmf expressions, it follows that the conditional mean of Y_{jX^*} is

$$E[Y_{jX^*} \mid X_o] = m_j(X^*, X_{jo};\ \pi), \tag{3}$$

where

$$m_j(X^*, X_{jo};\ \pi) = E[Y_{jX^*} \mid X_o] = \int_{-\infty}^{\infty} Y_{jX^*}\, f_{(Y_{jX^*} \mid X_{jo})}(Y_{jX^*}, X^*, X_{jo}; \pi)\, dY_{jX^*}$$

For j = 1, 2

Without loss of generality, we for now omit the subscript j (a.k.a., the bivariate outcome context) for now and look at the general expression of AIE.

Using the conditional mean function expression and the law of iterated expectation (LIE), the AIE can be written as

$$\text{AIE}(\Delta) = E[m(X^{pre} + \Delta, X_o;\ \pi)] - E[m(X^{pre}, X_o;\ \pi)] \tag{4}$$

It is relatively easy to show the under general conditions, given a consistent estimate of π (say $\hat{\pi}$), AIE can be consistently estimated using the following sample analog to the AIE expression above.

$$\widehat{\text{AIE}(\Delta)} = \sum_{i=1}^{n} \frac{1}{n} \{m(X_i^{pre} + \Delta_i, X_{oi}; \hat{\pi})] - E[m(X_i^{pre}, X_{oi}; \hat{\pi})\}$$

(5)

where X_i^{pre} and Δ_i are the exogenously determined values of X_i^{pre} and Δ for the i-th observation in a sample of size n (i = 1 , … , n); and X_{oi} is the sampled value of X_o.

2.3: Estimating the Treatment Effect of Interest

In this section, I will first introduce the concept of Conditional Potential Outcomes Model (CPOM), first proposed by Terza (2019a), in the generic context, and also in the context of our particular model: the correlated count regression, then review conditions under which DGP version of CPOM is legitimate, and lastly detail the asymptotic properties, especially the asymptotic standard error of the AIE estimator.

The conditional potential outcomes model (CPOM) specifies all moments of the distribution of $(Y_{X^*} \mid X)$ up to a given order. The class of CPOM that we call minimally parametric (MP) (fully parametric (FP)]) comprises those for which it is assumed that,

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

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

where m(.) is a known function, π is the vector of parameters. The FP CPOM of course implies a known form of the MP CPOM.

In our correlated count regression context, the CPOM is fully parametric, since we specify the joint probability mass function of the two correlated outcomes, which is detailed in the equations at the beginning of the next chapter. Generically, our FP CPOM has the form: $f(Y_1, Y_2 \mid X, X_o; \pi) = f_{([Y_{1X^*}, Y_{2X^*}] \mid X_o)}(Y_{1X^*}, Y_{2X^*}, X_o; \pi)]$. This of course implies that we have the knowledge of the conditional mean function (first order) and higher order conditional moment. Note that it might or not might have closed forms.

The identification of the relevant EP is tantamount to the identification of the relevant version of the CPOM. As is made clear by the expression of estimated AIE, consistent estimation of the EP in (4) hinges on the existence of a consistent estimate of π. With this in mind, and under conditions articulated by Terza (2019a), the following is legitimate:

$$\text{pmf/pdf}(Y \mid X_o) = f_{(Y_{X^*} \mid X_o)}(Y, X, X_o; \pi). \tag{6}$$

In other words, under certain condition[4], the relevant DGP can be obtained from our fully parametric specification by replacing the counterfactual random variables Y_{X^*} and X^* with the observable random variables Y and X, respectively.

[4] Here, we formally establish conditions under which the above replacement holds. In the context of the FP CPOM, in which it is implicit that (a) X_o induces conditional independence between Y_{X^*} and X, if (b) conditional outcome invariance holds and (c) overlap holds, then $\text{pmf/pdf}(Y \mid X_o) = f_{(Y_{X^*} \mid X_o)}(Y, X, X_o; \pi)$. For context, Terza 2019 (a), defines conditional independence under the GPOF as follows: let A, B and C be vector or scalar variates, then B induces conditional independence between A and C if pmf/pdf(A|B,C) = pmf/pdf(A|B). Additionally, the definition of overlap under the GPOF

From above replacement, it follows that π can be consistently estimated as the MLE obtained as

$$\hat{\pi} = \text{argmax} \sum_{i=1}^{n} q(\breve{\pi}, Z_i) \tag{7}$$

Where

$$q(\breve{\pi}, Z_i) = ln[f_{(Y_{X^*} \mid X_o)}(Y_i, X_i, X_{oi}; \breve{\pi})] \tag{8}$$

and

$$Z_i = [Y_i, X_i, X_{oi}]$$

In addition to consistently estimate the deep parameters π , we need the the asymptotic standard errors of the effect parameters (EP) – the square root of estimated asymptotic variance. Terza (2016a and b) shows that the asymptotically correct standard error of AIE is:

$$\text{Asy VAR(AIE)} = \left(\frac{\sum_{i=1}^{n} \nabla_\pi \widehat{aie}_i}{n}\right)\left(\widehat{AVAR}(\hat{\pi})\right)\left(\frac{\sum_{i=1}^{n} \nabla_\pi aie_i}{n}\right)' + \frac{\sum_{i=1}^{n} (\widehat{aie}_i - AIE)^2}{n}$$

is the following: overlap holds if $0 < p(X| X_o)(x | x_o) < 1$, where $p(X| X_o)(x | x_o)$ denotes the conditional probability mass/density function of X given $X_o = x_o$ evaluated at X = x.

where aie_i is shorthand notation for $\text{aie}(X_{pi}^{pre}, \Delta_i, X_{oi}; \hat{\pi})$, $A\widehat{VAR}(\hat{\pi})$ is the estimated asymptotic covariance matrix of $\hat{\pi}$, $\nabla_\pi\text{aie}$ denotes the gradient of aie with respect to π, and $\nabla_\pi\widehat{\text{aie}}_i$ represents $\nabla_\pi\text{aie}$ with X_{pi}^{pre}, X_{oi} and $\hat{\pi}$ substituted for X_p^{pre}, X_o and π respectively.

Now we have detailed the general fully parametric specification of our correlated count regression model in the Potential Outcomes Framework. In Chapter 3, we will introduce a particular fully parametric model: Correlated Dispersion-Flexible Count Regression (Seemingly Unrelated Count Regressions) model and cast our causal inference discussion into this model.

Chapter 3: A Proposed Estimator for the Deep Parameters and the AIE in the Seemingly Unrelated Count Regressions Context

The main objective of this chapter is to introduce our model -- a novel fully parametric specification that is particularly suitable for corelated count-valued data. Understanding the specification of this model is essential to compare and contrast with existing system-of-equations models that, for instance, estimate linear system of equations, or impose restriction on the correlation structures across equations. By doing so, we introduce the conditional probability mass functions, and then introduce the underlying likelihood functions for MLE of all deep parameters. Then we provide the formulation and estimation of our effect parameter (AIE) based on the conditional mean functions.

There are two major components under Chapter 3. In Section 3.1, introduce the generic version of our newly proposed model, which is an extension of Aitchison & Ho (1989) and discuss both its estimation of deep parameters via maximum likelihood methods and the estimation of the effect parameter (AIE); to lay the mathematical backgrounds necessary for the analyses in Chapters 5 and 6, to concretely illustrate abstract concepts, we focus on the cases of Poisson as our baseline model, and extend our baseline example to Conwy-Maxwell-Poisson marginal distributions, which is a distribution that accounts for dispersion flexibility that we introduce in Section 3.2.

3.1: The Generic Correlated Dispersion-Flexible Count Regression Model and the Conventional Poisson Specification

Motivated by the potential short-comings of conventional linear system of equation estimation models (e.g, Zeller 1962), we focus on count-valued outcomes. In this section, we introduce the mathematical framework for the analysis.

Specifically, in Section 3.1.1, I first introduce the generic correlated dispersion-flexible count Potential Outcomes specification (i.e., the probability mass functions (PMF)), then I present an example, which utilizes the classic Poisson marginals. In this example, I provide the expressions of the PMF, the conditional mean functions, and the likelihood functions. As discussed before, the focus in this dissertation is on policy effect estimation for the case in which the outcomes (the Ys) are count valued, I thus also provide the expressions of the effect parameter and its standard errors.

3.1.1: Probability Mass Function

A single observation of our data is a quadruple (Y_1, Y_2, X, X_0), where (Y_1, Y_2) is the bivariate, count-valued outcome vector, X_0 is a vector of covariates, and X is the policy variable of interest. Taking (X, X_0) as given, the conditional distribution of (Y_1, Y_2) is

$$f(Y_1, Y_2 \mid X, X_o; \pi) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [f_{(Y_1 \mid X, X_0, \eta_1)}(Y_1, X, X_o, \eta_1; \pi) \times f_{(Y_1 \mid X, X_0, \eta_2)}(Y_2, X, X_0, \eta_2; \pi)\, g(\eta_1, \eta_2; \pi)]\, d\eta_1\, d\eta_2 \tag{9}$$

where: (i) π is the vector of parameters, (ii) $f_{(Y_j \mid X_p, X_0, \eta_j)}(Y_j, X, X_0, \eta_j; \pi)$ is the conditional marginal distribution of Y_j given (X, X_0, η_j) for j = 1, 2; (iii) (η_1, η_2) are the "structural cross-equation heterogeneity terms" that serve to link the marginal distributions together and are independent of the covariates; and (iv) $g(\eta_1, \eta_2; \pi)$ is the probability density function of (η_1, η_2). Without loss of generality, we take $g(\eta_1, \eta_2; \pi)$ to be a bivariate normal distribution with mean vector 0, marginal standard deviations of 1, and correlation parameter of ρ_{12}. Formally stated,

$$(\eta_1, \eta_2) \sim N(0, \Sigma) \tag{10}$$

where

$$\Sigma = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix} \tag{11}$$

We thus write $g(\eta_1, \eta_2; \rho_{12})$ instead of $g(\eta_1, \eta_2; \pi)$ for the balance of the paper.

This type of "mixture" probability mass function is well-studied in the count-outcomes literature -- e.g., Aitchison and Ho (1989). One of the key advantages of this type of model is that it relaxes strong homoscedastic features of earlier bivariate count models and allows for flexible heteroscedasticity.

Example (Case I): Poisson Model PMF

For the Poisson case and j = 1, 2, the marginal distribution of the j-th outcome is,

$$f_{(Y_j \mid X, X_0, \eta_1)}(Y_j, X, X_0, \eta_j; \pi) = \frac{(\lambda_j)^{Y_j} \exp(-\lambda_j)}{Y_j!}$$

(12)

where

$$\lambda_j = \exp(X\beta_j + X_0\beta_{0j} + \sigma_j\eta_j)$$

(13)

The vector of parameters is thus $\pi = \left[\beta_{1,}\ \beta_{01,}\ \beta_{2,}\ \beta_{02,}\ \sigma_1, \sigma_2, \rho_{12}\right]$.[5]

3.1.2: Conditional Mean Functions

Given X_0, the "Average Incremental Effect (AIE)" is the expected change in the outcome vector that results from an increase in the policy variable from X^{pre} to $X^{pre} + \Delta$ where $\Delta > 0$ is the size of the increment. Formally stated,

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

(14)

where m(.) is the expected value of the outcome vector given (X, X_0), i.e.,

---

σ_1 and σ_2 play roles analogous to the marginal standard deviations of η_1 and η_2 respectively. Thus, if (η_1, η_2) were not standardized to have variances of one, then the Poisson case would fail to be identified. We model the marginal variation in (η_1, η_2) through (σ_1, σ_2) to simplify the integration come estimation.

$$m(X, X_0; \pi) = E[Y \mid X,\ X_0; \pi]$$

$$= \begin{bmatrix} \sum_{y_1=1}^{\infty} [y_1 \cdot\ f(Y_1 = y_1 \mid X, X_0; \pi)] \\ \sum_{y_2=1}^{\infty} [y_2 \cdot\ f(Y_2 = y_1 \mid X, X_0; \pi)] \end{bmatrix}$$

(15)

where

$$f(Y_j = y \mid X^{pre}, X_0; \pi) = \sum_{y_j=1}^{\infty} f(Y_1, Y_2 \mid X, X_0;\ \pi)$$

(16)

is the marginal distribution of Y_j for j = 1, 2. We refer to $m(X,\ X_0; \pi)$ as the "conditional mean function."

The AIE is a key object for the analyses in Chapters 5 and 6 and general policy analysis since it estimates the shift in outcomes resulting from a change in policy. To foreshadow, in Chapter 6, we use the AIE to estimate the impact of expansions of health insurance on health care utilization and evaluate hypotheses concerning this estimate.

Example (Case I): Poisson Model Conditional Mean Function

In the case of the Poisson specification, we have that

$$m(X, X_0; \pi) = \begin{bmatrix} E(Y_1 | X, X_0; \pi) \\ E(Y_2 | X, X_0; \pi) \end{bmatrix}$$

$$= \begin{bmatrix} \int_{-\infty}^{\infty} \exp\left(X\beta_1 + X_0\beta_{01} + \sigma_1\eta_1\right) \int_{-\infty}^{\infty} g\left(\eta_1, \eta_2; \widehat{\rho_{12}}\right) d\eta_2\, d\eta_1 \\ \int_{-\infty}^{\infty} \exp\left(X\beta_2 + X_0\beta_{02} + \sigma_2\eta_2\right) \int_{-\infty}^{\infty} g\left(\eta_1, \eta_2; \widehat{\rho_{12}}\right) d\eta_1\, d\eta_2 \end{bmatrix}$$

(17)

at the parameter vector $\pi = \left[\beta_{1,}\ \beta_{01,}\ \beta_{2,}\ \beta_{02,}\ \sigma_1, \sigma_2, \rho_{12}\right]$. To see this, write for j = 1, 2,

$$E\left(Y_j \mid X, X_0; \pi\right) = \int_{-\infty}^{\infty} E\left(Y_j \,\middle|\, X, X_0, \eta_j; \pi\right) f_{\eta_j \mid X, X_0}\left(\eta_j \,\middle|\, X, X_0; \pi\right) d\eta_j$$

$$= \int_{-\infty}^{\infty} E\left(Y_j \,\middle|\, X, X_0, \eta_j; \pi\right) \int_{-\infty}^{\infty} g(\eta_1, \eta_2; \widehat{\rho_{12}}) d\eta_{-j}\, d\eta_j$$

$$= \int_{-\infty}^{\infty} \exp\left(X\beta_j + X_0\beta_{0j} + \sigma_j\eta_j\right) \int_{-\infty}^{\infty} g\left(\eta_1, \eta_2; \widehat{\rho_{12}}\right) d\eta_{-j}\, \partial\eta_j$$

(18)

where $E\left(Y_j \,\middle|\, X, X_0, \eta_j; \pi\right)$ is the conditional expectation of Y_j given (X, X_0, η_j), where $f_{\eta_j|X,X_0}\left(\eta_j \,\middle|\, X, X_0; \pi\right)$ is the conditional distribution of η_j given (X, X_0), and where $\int_{-\infty}^{\infty} g\left(\eta_1, \eta_2; \widehat{\rho_{12}}\right) d\eta_{-j}$ is the marginal distribution of η_j. The first equality follows from the definitions of marginal and conditional probabilities; see below. The second equality follows from the fact that the structural cross equation errors are independent of the covariates under the overall specification, so the conditional distribution equals the marginal distribution, i.e., $f_{\eta_j|X,X_0}\left(\eta_j \,\middle|\, X, X_0; \pi\right) = \int_{-\infty}^{\infty} g(\eta_1, \eta_2; \widehat{\rho_{12}}) d\eta_{-j}$. The last equality follows from the Poisson assumption.

To elaborate on the first equality in the display equation, let A, B, and C be random variables with joint distribution $f_{ABC}(A, B, C) > 0$, where A is discrete, and B and C are continuous with support equal to the real line and where f(.) is continuous. Then,

$$E(A \mid B = b) = \int_{-\infty}^{\infty} E(A \mid B = b\ , C = c) f_{C|B}(c|b) db\,.$$

(19)

The argument is straightforward,

$$\int_{-\infty}^{\infty} E(A|B{=}b,\ C{=}c) f_{C|B}(c|b)dc = \int_{-\infty}^{\infty} \sum_{a=1}^{\infty} a\ f_{A|BC}(a|b,c) f_{C|B}(c|b)dc$$

$$= \sum_{a=1}^{\infty} a \int_{-\infty}^{\infty} f_{A|BC}(a|b,c) f_{C|B}(c|b)dc$$

$$= \sum_{a=1}^{\infty} a \int_{-\infty}^{\infty} f_{ABC}(a,b,c)/f_B(b)dc$$

$$= \sum_{a=1}^{\infty} a\ f_{AB}(a,b)/f_B(b) = E(A|B{=}b)$$

The first equality is definitional. The second equality follows by Fubini's Theorem since continuous functions are measurable. The third equality follows from the fact that $f_{A|BC}(a|b,c)f_{C|B}(c|b) = (f_{ABC}(a,b,c)/f_{BC}(b,c))(f_{BC}(b,c)/f_B(b)) = f_{ABC}(a,b,c)/f_B(b)$ by the definition of marginal and conditional probabilities. The fourth equality follows from the definition of marginal probability. The last equality is definitional.

By setting $A = Y_j$, $B = (X, X_0)$, and $C = \eta_j$, it is easily seen that equation (19) implies the first equality in equation (18) since the joint distribution of the random variables is continuous and strictly positive.

3.1.3: Maximum Likelihood Estimation (MLE)

Maximum Likelihood Estimation (MLE) is used to estimate the parameters of the model from the data. Specifically, my data $\{(Y_{1i}, Y_{2i}, X_i, X_{0i})\}_{i=1}^{N}$ consist of N independent observations of the outcomes and covariates, each indexed i = 1 … N. The log-likelihood of the data is thus

$$L(\pi) = \sum_{i=1}^{N} \{\ln f_{(Y_1, Y_2 \mid X, X_o)}(Y_{1i}, Y_{2i}, X_i, X_{oi}; \pi)\}$$

$$= \sum_{i=1}^{N} \{\ln \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [f_{(Y_1 \mid X, X_o, \eta_1)}(Y_{1i}, X_i, X_{oi}, \eta_1; \pi) \times$$

$$f_{(Y_2 \mid X, X_o, \eta_2)}(Y_{2i}, X_i, X_{oi}, \eta_2; \pi)\, g(\eta_1, \eta_2; \rho_{12})]\, d\eta_1\, d\eta_2\} \tag{20}$$

We estimate the unknown, true parameters π from the data using maximum likelihood methods. The integrals in equation (20) generally lack closed form; this poses a central challenge in the estimation routine. We surmount this problem by developing a new Stata program that accurately and quickly approximates the integrals using the Gauss-Legendre Quadrature; see Chapter 4 for details. This program allows us to maximize equation (20) using Stata's "moptimize" command, which employs hill-climbing algorithms (e.g., variants of Newton-Raphson and Berndt–Hall–Hall–Hausman), and obtain the MLE estimates $\hat{\pi}$ of π.

Example (Case I): Poisson Model Maximum Likelihood Function

The relevant version of the log likelihood function in Poisson case is:

$$L(\pi) =$$

$$\sum_{i=1}^{N} \ln\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\frac{(\lambda_{1i})^{Y_{1i}}\exp(-\lambda_{1i})}{Y_{1i}!}\times\frac{(\lambda_{2i})^{Y_{2i}}\exp(-\lambda_{2i})}{Y_{2i}!}\cdot g(\eta_1,\eta_2;\rho_{12})\right] d\eta_1\, d\eta_2 \qquad (21)$$

Where $\lambda_{ji} = \exp(X_i\beta_j + X_{oi}\beta_{oj} + \sigma_j\eta_j)$ for j =1, 2; and for i = 1…N.

3.1.4: Estimation of the Average Incremental Effect and Its Standard Error

Given X_o, the "Average Incremental Effect (AIE)" is the expected change in the outcome vector that results from an increase in the policy variable from X^{pre} to $X^{pre} + \Delta$, where $\Delta > 0$ is the size of the increment. In symbols,

$$AIE(\Delta) = E[m(X^{pre} + \Delta, X_o;\ \pi)] - E[m(X^{pre}, X_o;\ \pi)] \qquad (22)$$

where m(.) is the expected value of the outcome vector given $(X,\ X_0)$, i.e.,

$$m(X, X_0; \pi) = E[Y \mid X,\ X_0;\ \pi] = \begin{bmatrix}\sum_{y_1=1}^{\infty}[y_1 \cdot f(Y_1 = y_1 \mid X,\ X_0;\ \pi)]\\ \sum_{y_2=1}^{\infty}[y_2 \cdot f(Y_2 = y_1 \mid X,\ X_0;\ \pi)]\end{bmatrix} \qquad (23)$$

where

$$f(Y_j = y \mid X^{pre}, X_0; \pi) = \sum_{y_j=1}^{\infty} f(Y_1, Y_2 \mid X, X_0; \pi)$$

(24)

is the marginal distribution of Y_j for j = 1, 2. We refer to $m(X, X_0; \pi)$ as the "conditional mean function."

Given the MLE estimate $\hat{\pi}$, an estimate of the Average Incremental Effect for an increment of $\Delta > 0$ is

$$\widehat{AIE}(\Delta) = \hat{\kappa}_1(\Delta) - \hat{\kappa}_2(\Delta)$$

(25)

where

$$\hat{\kappa}_1(\Delta) = \sum_{i=1}^{N} \frac{1}{N} m(X_i + \Delta, X_{0i}; \hat{\pi})$$

(26)

and

$$\hat{\kappa}_2(\Delta) = \sum_{i=1}^{N} \frac{1}{N} m(X_i, X_{0i}; \hat{\pi})$$

(27)

Thus, we also have that

$$\widehat{AIE} = \frac{\sum_{i=1}^{n} \widehat{aie}_i}{n}$$

(28)

Also, to make statistical inferences regarding the effect parameter, we would like to seek the estimated asymptotically correct variance of AIE, the square root of which is the correct asymptotic standard error.

Terza (2016a and b) has provided a generic formula for deriving the asymptotic variance of AIE as a function of estimated deep parameters, observation by observation AIE and the gradient of individual AIEs. This formulation is suitable for any casual estimation and inferences of effect parameters that involve maximum likelihood methods or pseudo-maximum likelihood estimation methods, and it is readily applicable in our model after obtaining the prerequisite elements.

Terza (2016a and b) shows that the formulation of the estimated asymptotically correct variance of AIE is:

$$\text{Asy VAR(AIE)} = \left(\frac{\sum_{i=1}^{n} \nabla_{\pi}\widehat{aie}_i}{n}\right)\left(\widehat{AVAR}(\hat{\pi})\right)\left(\frac{\sum_{i=1}^{n} \nabla_{\pi}aie_i}{n}\right)' + \frac{\sum_{i=1}^{n} (\widehat{aie}_i - AIE)^2}{n}$$

(29)

where $\hat{\pi}$ represents the vector of all deep parameter estimates, aie_i is shorthand notation for $aie(X_{pi}^{pre}, \Delta_i, X_{oi}; \hat{\pi})$, $\widehat{AVAR}(\hat{\pi})$ is the estimated asymptotic covariance matrix of $\hat{\pi}$ $\nabla_{\pi}aie$ denotes the gradient of aie with respect to π, and $\nabla_{\pi}\widehat{aie}_i$ represents $\nabla_{\pi}aie$ with X_{pi}^{pre}, X_{oi} and $\hat{\pi}$ substituted for X_p^{pre}, X_o and π respectively.

In terms of the procedure of obtaining each component of Asy VAR(AIE) in Stata/Mata, AVAR($\hat{\pi}$) can be obtained directly from the Stata output from the relevant State regression command. The term $\frac{\sum_{i=0}^{n}(\nabla_\pi \text{aie}_i\text{-ME})^2}{n}$ is easily calculated using Mata, given that $\widehat{\text{AIE}} = \frac{\sum_{i=1}^{n}\widehat{\text{aie}}_i}{n}$ has already been calculated. Direct calculation of the remaining component, $\frac{\sum_{i=1}^{n}\nabla_\pi\widehat{\text{aie}}_i}{n}$, requires analytic derivation of $\nabla_\pi aie$ and Mata coding of $\nabla_\pi\widehat{\text{aie}}_i$. alternatively, we can use the Mata DERIV (which calculates the vectorized analytical derivation value) command to calculate the estimated asymptotically correct variance and corresponding t-stat without having the exact formulation of $\nabla_\pi\widehat{\text{aie}}_i$. Use of the DERIV command allows us to avoid having to derive the explicit form of $\frac{\sum_{i=1}^{n}\nabla_\pi\widehat{\text{aie}}_i}{n}$ because it affords a way to numerically approximate the components of this gradient vector.

At last, we can obtain the standard error of the average incremental effect by taking the square root of Asy VAR(AIE), i.e., s.e. (AIE) = $\sqrt{\text{Asy VAR(AIE)}}$.

Example (Case I): Poisson Model Average Incremental Effects

The above equations imply that the estimated AIE for the bivariate Poisson model is:

$$\widehat{\text{AIE}}(\Delta) = \sum_{i=1}^{N}\frac{1}{N}\begin{bmatrix} \left(\int_{-\infty}^{\infty}\left(\exp\left((X_i+\Delta)\widehat{\beta_1} + X_{0i}\widehat{\beta_{01}} + \widehat{\sigma_1}\eta_1\right) - \exp\left((X_i)\widehat{\beta_1} + X_{0i}\widehat{\beta_{01}} + \widehat{\sigma_1}\eta_1\right)\right) \right. \\ \left. \times \int_{-\infty}^{\infty} g(\eta_1,\eta_2;\widehat{\rho_{12}})d\eta_2\,\partial\eta_1\right) \\ \left(\int_{-\infty}^{\infty}\left(\exp\left((X_i+\Delta)\widehat{\beta_2} + X_{0i}\widehat{\beta_{02}} + \widehat{\sigma_2}\eta_2\right) - \exp\left((X_i)\widehat{\beta_2} + X_{0i}\widehat{\beta_{02}} + \widehat{\sigma_2}\eta_2\right)\right) \right. \\ \left. \int_{-\infty}^{\infty} g(\eta_1,\eta_2;\widehat{\rho_{12}})d\eta_1\,\partial\eta_2\right) \end{bmatrix} \tag{30}$$

The analytical expression for the standard error of the $\widehat{\text{AIE}}$ in the Poisson case follows as:

$$\frac{\sum_{i=1}^{n} \nabla_{\pi}\widehat{\text{aie}_i}}{n}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(\exp((X_{pi}+1)\widehat{\beta_p} + X_{oi}\widehat{\beta_{o1}} + \widehat{\sigma_1}\eta_1)\left[[X_{pi}+1]\ \ X_o\ \ \eta_1\right]$$

$$- \exp(X_{pi}\widehat{\beta_1} + X_{0i}\widehat{\beta_{01}} + \widehat{\sigma_1}\eta_1)[X_{pi}\ \ X_o\ \eta_1]) \tag{31}$$

3.2: Generalization: Dispersion-Flexible Marginal Distributions – Conway-Maxwell-Poisson (CMP)

By using Poisson marginals, we have implicitly restricted the conditional mean (CM) of each outcome variable to equal its conditional variance (CV) given the covariates and structural errors; this is called the equi-dispersion restriction. To mitigate this restriction, we consider incorporate dispersion-flexible, count-valued marginal distributions that allow for both under-dispersion (i.e., where the CV < CM) and over-dispersion (i.e., CV > CM). As a first step in this direction, we extend our simulation work to cover Conway-Maxwell-Poisson (CMP) marginals.

3.2.1: Probability Mass Function of the CMP Case

For j = 1, 2, a straight-forward implementation of the standard Conway-Maxwell-Poisson (CMP) marginals yields,

$$f_{(Y_j \mid X, X_0, \eta_1)}(Y_j, X, X_0, \eta_j; \pi) = \frac{(\lambda_j^o)^{Y_j}}{(Y_j!)^{\exp(\omega_j)} Z(\lambda_j^o, \omega_j)},$$

(32)

where

$$\lambda_j^o = \exp(X\beta_j + X_0\beta_{0j} + \sigma_j \eta_j)^{\frac{1}{\sigma_j}}$$

(33)

and

$$Z(\lambda_j^o, \omega_j) = \sum_{r=0}^{\infty} \frac{(\lambda_j^o)^r}{(r!)^{\exp(\omega_j)}}$$

(34)

The vector of parameters is thus $\pi = \left[\beta_{1,}\ \beta_{01,}\ \beta_{2,}\ \beta_{02,}\ \sigma_1, \sigma_2, \omega_1, \omega_2,\ \rho_{12}\right]$. In this specification, the parameter ω_j controls dispersion – i.e., over-dispersion occurs when $\omega_j < 0$, equi-dispersion when $\omega_j = 0$, and under-dispersion when $\omega_j > 0$ – and so contributes to the variance of Y_j. The CMP nests the case of Poisson marginals when $\omega_j = 0$ – and so allows for a great range of flexibility in applied modeling. In addition, the CMP limits towards two other common count-valued specifications: the Geometric distribution when $\omega_j \rightarrow -\infty$ and $\lambda_j^o < 1$ and the Bernoulli distribution when $\omega_j \rightarrow \infty$ with probability of success $\lambda_j^o / 1 + \lambda_j^o$. The nested Poisson allows for a simple statistical test of whether or not the specification varies significantly from the standard Poisson.

Unfortunately, this specification cannot be estimated as written dure to an identification problem – e.g., both σ_j and ω_j control the variance of Y_j and so are individually indeterminant. We thus reparametrize equation (34) by dividing the terms inside the exponential by σ_j so they become $\beta_j^{\oplus} = \beta_j/\sigma_j$ and $\beta_{0j}^{\oplus} = \beta_{0j}/\sigma_j$ respectively and we get

$$\lambda_j^o = \exp(X\beta_j^{\oplus} + X_0\beta_{0j}^{\oplus} + \eta_j) \tag{35}$$

This version of the model does not suffer from the identification concern and is rendered "neutral" with respect to the parameter σ_j. Modeling in this way: 1) allows for possible correlation between equations; 2) accommodates possible non-equi-dispersion in the data; and 3) avoids identification problems surrounding the multiplicity of parameters.

3.2.2: Conditional Mean Functions of the CMP Case

The conditional means function of the potential outcomes Y_{jX^*} under CM-Poisson marginals is as below.

$$E[Y_{jX^*} | X_0] = m\left(X^*, X_0; \pi\right) = \lambda_j \cdot \sum_{m=0}^{\infty} \frac{m \cdot (\lambda_j)^{m-1}}{(m!)^{\exp(\omega_j)} \cdot Z(\lambda_j, \exp(\omega_j))} \tag{36}$$

where

$$Z(\lambda_j, \exp(\omega_j)) = \sum_{m=0}^{\infty} \frac{(\lambda_j)^m}{(m!)^{\exp(\omega_j)}}$$

(37)

for j = 1, 2, and $\lambda_j = \exp(X^*\beta_{pj} + X_0\beta_{0j} + \eta_j^*)$, ω_{1j} are the dispersion parameters for each outcome, and $\pi = [\beta_{1,}\ \beta_{01,}\ \beta_{2,}\ \beta_{02,}\ \omega_1,\ \omega_2]$,

3.2.3: Maximum Likelihood Estimation (MLE) of the CMP case

For correlated bivariate count outcomes (Y_1, Y_2), the log-likelihood function in the Conway-Maxwell-Poisson mixture single equation model Y_j is:

$$f_{(Y_{1X_p^*}|\ X_p,\ X_0)}(Y_{1i}, X_{pi}, X_{0i}; \pi) = \int_{-\infty}^{\infty} cmp(Y_{1i}, X_{pi}, X_{0i}; \lambda_1, \omega_1)\ g^*(\eta_1^*; \sigma_1)\ d\eta_1^*$$

$$f_{(Y_{2X_p^*}|\ X_p,\ X_0)}(Y_{2i}, X_{pi}, X_{0i}; \pi) = \int_{-\infty}^{\infty} cmp(Y_{2i}, X_{pi}, X_{0i}; \lambda_2, \omega_2)\ g^*(\eta_2^*, \sigma_2)\ d\eta_2^*$$

(38)

where $cmp(Y_{ji}, X_{pi}, X_{0i}; \lambda_j, \nu_j)$ are the Conway-Maxwell-Poisson density and $\lambda_j = \exp(X^*\beta_j + X_0\beta_{0j} + \eta_j^*)$ for j = 1, 2, and $g^*(\eta_1^*)$ and $g^*(\eta_2^*)$ are the univariate standard normal pdfs with mean 0, 0 and variance 1 and 1, and $\pi = (\beta_1, \beta_2, \omega_1, \omega_2)$.

3.2.4: Estimation of the Average Incremental Effect and Its Standard Error for the CMP case

Similarly, we would be able to consistently estimate the effect parameter (AIE) of the bivariate CMP model from the generic conditional mean function using their following sample analogs:

$$\mathrm{AIE}(\Delta) = \sum_{i=1}^{n} \frac{1}{n} \{ m\left(X_{pi}^{*}+\Delta_i, X_{0i}; \hat{\pi}\right) - m\left(X_{pi}^{*}, X_{0i}; \hat{\pi}\right) \}$$

$$= \sum_{i=1}^{n} \frac{1}{n} \{ \lambda_{ji} \cdot \sum_{m=0}^{\infty} \frac{m \cdot (\lambda_{ji}(X_{pi}^{pre}+\Delta))^{m-1}}{(m!)^{\exp(\omega_j)} \cdot Z(\lambda_{ji}(X_{pi}^{pre}+\Delta), \exp(\omega_j))}$$

$$- \lambda_{ji} \cdot \sum_{m=0}^{\infty} \frac{m \cdot (\lambda_{ji}(X_{pi}^{pre}))^{m-1}}{(m!)^{\exp(\omega_j)} \cdot Z(\lambda_{ji}(X_{pi}^{pre}), \exp(\omega_j))} \}$$

(39)

where

$$\lambda_{ji}\left(X_{pi}^{pre} + \Delta\right) = \exp((X_{pi}^{pre} + \Delta)\widehat{\beta_{pj}} + X_{0i}\widehat{\beta_{0j}} + \eta_j^{*})$$

(40)

And

$$\lambda_{ji}\left(X_{pi}^{pre}\right) = \exp(X_{pi}^{pre}\widehat{\beta_{pj}} + X_{0i}\widehat{\beta_{0j}} + \eta_j^{*})$$

(41)

The standard error of the AIE for the CMP case again follows Terza 2016 (a) and (b)'s formulation of a generic, simplified version of AIE's asymptotic variance.

$$\text{s.e.}(\text{AIE})^{CMP} = \sqrt{\text{Asy VAR}(\text{AIE})^{CMP}}$$

(42)

where:

$$\text{Asy VAR}(\text{AIE})^{CMP} =$$

$$\left(\frac{\sum_{i=1}^{n} \nabla_{\pi}\widehat{\text{aie}_i cmp}}{n}\right)\left(\widehat{\text{AVAR}}(\hat{\pi})\right)\left(\frac{\sum_{i=1}^{n} \nabla_{\pi}\text{aie}_i cmp}{n}\right)' + \frac{\sum_{i=1}^{n}(\widehat{\text{aie}_i cmp} - \text{AIE})^2}{n}$$

(43)

Where we replace the first component the CMP version observation wise AIE's derivative with respect to the full set of parameters, π.

$$\frac{\sum_{i=1}^{n} \nabla_{\pi}\widehat{\text{aie}_i \text{cmp}}}{n} =$$

$$\sum_{i=1}^{n}\frac{1}{n}\nabla_{\pi}\{\ \lambda_{ji}\cdot\sum_{m=0}^{\infty}\frac{m\cdot(\ \lambda_{ji}(X_{pi}^{pre}+\Delta))^{m\text{-}1}}{(m!)^{\exp(\omega_j)}\cdot Z(\ \lambda_{ji}(X_{pi}^{pre}+\Delta),\exp(\omega_j))}$$

$$\lambda_{ji}\cdot\sum_{m=0}^{\infty}\frac{m\cdot(\ \lambda_{ji}(X_{pi}^{pre}))^{m\text{-}1}}{(m!)^{\exp(\omega_j)}\cdot Z(\ \lambda_{ji}(X_{pi}^{pre}),\exp(\omega_j))}\}$$

(44)

Since the formulation of above is analytically daunting, we actually do not need to explicit derive the expression above, instead, we use the Mata's DERIV command to approximate its value evaluated at the value of our deep parameter estimates.

And the last component ($\widehat{aie_i cmp}$) with CMP version observation wise AIE with estimated full set of parameters $\hat{\pi}$, which we also have obtained when calculation the AIE; formally, we have that

$$\widehat{aie_i cmp}=$$

$$\sum_{i=1}^{n}\frac{1}{n}\{\,\lambda_{ji}\cdot\sum_{m=0}^{\infty}\frac{m\cdot(\,\lambda_{ji}(X_{pi}^{pre}+\Delta))^{m-1}}{(m!)^{\exp(\omega_j)}\cdot Z(\,\lambda_{ji}(X_{pi}^{pre}+\Delta),\exp(\omega_j))}$$

$$\lambda_{ji}\cdot\sum_{m=0}^{\infty}\frac{m\cdot(\,\lambda_{ji}(X_{pi}^{pre}))^{m-1}}{(m!)^{\exp(\omega_j)}\cdot Z(\,\lambda_{ji}(X_{pi}^{pre}),\exp(\omega_j))}\}$$

(45)

Now we have introduced most of the analytical components of our model and prerequisites for causal estimation and inference. However, there is one additional technical challenge we need to solve before we could arrive at desired estimation results. As we have mentioned when we introduce our non-closed form likelihood functions in Section 3.1.3, the values of our target likelihood function are virtually impossible to calculate given existing built-in packages or commands within Stata/Mata or most other statistical software, as it involves high dimensional integration. In Chapter 4, we introduce our newly developed numerical integral approximation algorithm implemented in Stata/Mata that take care of the above issues. After understanding the theories and applications of this approximation program, the readers will be able to implement our model and, without technical challenges, produce or replicate our estimation results in Chapters 5 and 6.

Chapter 4: Bivariate Gauss-Legendre Quadrature

In this section, we describe the details of our Gauss-Legendre quadrature (GLQ) methods for accurately approximating two-dimensional integrals by first overviewing the available methods, then reviewing the math behind GLQ and our implementation of it in Mata, before closing with two validation studies.

4.1: Literature on Numerical Approximation of Non-Closed Form Integrals

Many applications in empirical econometrics require the evaluation of two-dimensional (2D) integrals – e.g., equations (20) and (21), and more generally in the computation of likelihood functions for MLE, expectations for M-estimators, and so on. Yet, it is generically impossible to derive an analytic closed form for such integrals outside of special cases. This poses a significant challenge to estimation, inference, and prediction.

Researchers have attempted to solve this "2D integration problem" in several ways. One approach is implements Gaussian quadrature. In Gaussian quadrature, one approximates an integral by evaluating its integrand at select points (i.e., "abscissas") and then taking their weighted sum. There are several ways in which to compute the abscissas, each with their advantages and disadvantages. For instance, Aitchison and Ho (1989), Chua and Tsiaplias (2019), and Kim et al. (2015) employ Gauss-Hermitian quadrature, whereas we employ GLQ. Gauss-Hermitian Quadrature is a type of quadrature well-suited to approximating 2D wherein the method for selecting the abscissas and weights requires the integrand be written as $f(x, y)\, e^{-x^2-y^2}$, where the variables of integration are (x, y), and allows for improper integrals. While many distributions can be transformed to fit into this form, others cannot. In contrast, the

GLQ's method for selecting the abscissas and weights only requires that the integrand be of the form $f(x, y)$ and so is more flexible, but requires that the integral be taken over a closed and bounded domain $[-M, M]^2$, with bound $M < \infty$. Thus, its use for improper integrals always entails a small approximation error, which vanishes as $M \to \infty$. From a practical standpoint, we find that the flexibility of GLQ outweighs the approximation error, especially when the bound is large.

Another approach to solving the 2D integration problem is Monte-Carlo Simulation, wherein one randomly selects a large number of points in the domain of the integral, evaluates the integrand at each point, and then takes the average. By the law of large numbers, the average converges to the value of the integral as the number of points grows large. Such methods are common in applied work, including Chua and Tsiaplias (2019), Heiss and Winschel (2008), Train (2000), and Skrainka and Judd (2011). Skrainka and Judd conduct a detailed comparison of quadrature and Monte-Carlo methods, wherein they conclude that the accuracy and computation of the former often dominate the latter.

In Bayesian estimation, it is common to circumvent the 2D integration problem by using methods that update the integrand without computing the integral and then employing (complex) simulation exercises to evaluate the updated integral. Typical works in this literature include Chib and Winkelmann (2001), Hirano and Porter (2003), Mehta (2014), and Stegmueller (2013). The focus of our paper, however, lies on classical econometric methods and hypothesis testing for which the 2D integration problem poses challenges.[6]

4.2: Approximation Algorithms and Mata Software Development

In this section, we outline the methods of GLQ, with a focus on our Mata implementation of these procedures. (Mata is Stata's low-level matrix programming language.)

GLQ is based around the approximation of definite integrals. To fix ideas, consider the following target of approximation:

$$I = \int_a^b \int_c^d f(x, y)\, dy\, dx \tag{46}$$

where the domain of (x, y) is $[a, b] \times [c, d]$, and the integrand is a generic continuous function $f(x, y)$.

As the first step of implementing the Gauss-Legendre quadrature approximation algorithm, transformation of the domain is required to the unit square. Here, we transform the arguments domains from $[c, d] \times [a, b]$ to $[-1, 1] \times [-1, 1]$ by the following steps:

[6] Regarding the broader, philosophical debate between Bayesians classical econometricians and statisticians, we do not take a side beyond observing that classical econometric methods and hypothesis testing are still central to many fields of economics, policy work, and other domains.

$$I = \int_a^b \int_c^d f(x,y)\,dy\,dx = \int_a^b \int_{-1}^1 f(x, \frac{(d-c)u + (d+c)}{2})(\frac{d-c}{2})\,du\,dx$$

$$= \int_{-1}^1 \int_{-1}^1 f(\frac{(b-a)v + (b+a)}{2}, \frac{(d-c)u + (d+c)}{2})(\frac{(d-c)(b-a)}{4})\,du\,dv$$

(47)

As the next step, we select the number of abscises to use, *n*. The GLQ methods allows us to approximate *I* by as follows,

$$I \approx \frac{(d-c)(b-a)}{4} \sum_{j=1}^{n} \sum_{i=1}^{n} w_{i,1} \cdot w_{j,2} \cdot f(\frac{(b-a)abs_{i,1} + (b+a)}{2}, \frac{(d-c)abs_{j,2} + (d+c)}{2})$$

(48)

where the weights $(w_{i,1}, w_{j,2})$ and the abscises $(abs_{i,1}, abs_{j,2})$ for each i, j pair are determined by the GLQ rules. These rules are based on the roots of the Legendre polynomials and the order of the polynomials used in the approximation *n*; for details see Mander and Bowden (2012). Observe that the higher number of quadrature points yields more accurate approximation.

We implement equation (48) in Mata, along with appropriate code to compute the abscises and weights. The results are presently available upon request and will be released as a Stata program for wide-spread distribution, which will accept user-generated 2D integrands, the limits of integration, and the number of abscises and then compute the integrals.

Equation (46) concerns a definite integral. Yet, most of the integrals in the present paper are improper and of the form

$$I = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dy\, dx \tag{49}$$

The definition of an improper integral gives that there is a real number $M < \infty$ such that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dy\, dx \approx \int_{-M}^{M} \int_{-M}^{M} f(x, y)\, dy\, dx. \tag{50}$$

In addition, it is easily seen that if $f(x, y) \geq 0$, then the above approximation becomes monotonically more accurate as $M \to \infty$. Since the improper integral we seek to approximate in our count-valued models have strictly positive integrands, we approximate it by first choosing a suitably large M and then applying the methods discussed above.

In our application, where the structural errors in our main specification are normally distributed, we find that setting M larger than five to eight standard deviations away from the mean does not lead to noticeable increases in the values returned from quadrature. The implication is that five to eight standard deviations are sufficient for accurate approximation.

4.3: Validation of our Bivariate Gauss-Legendre Quadrature Software

To validate the accuracy of our approximation software, we conducted two validation exercises. First, we applied our 2D Gauss-Legendre Quadrature program to approximate the cumulative distribution function of a bivariate Poisson model and compared the results to the empirical cumulative distribution function computed from simulated data, which was drawn in the same manner as described in Chapter 3. Table 1 provides the results and shows that the software accurately reproduces the true marginal cumulative distributions of both variables. The support of simulated bivariate Poisson distributed Y1 is [0, 18], and the support of simulated Y2 is [0, 19]. As we can observe, our approximation software can at least approximate the cumulative density for two decimal points. For instance, according to our 2D Gauss-Legendre Quadrature, Prob (Y1 = 0) = 0.823859309, while the actual relative frequency of Y1 = 0 in the simulated dataset with sample size of 500,000 is equal to 0.824746.

We not only looked at the CDFs, but also the joint PMFs during our validation. we used our program to approximate the joint distribution of the same model and compared it to the joint empirical distribution from the same simulated data. Figures 1 present the results and show that our program returns an accurate probability mass function.

Secondly, we applied our 2D Gauss-Legendre Quadrature program to approximate the cumulative distribution function of a bivariate Conway Maxwell Poisson (CMP) model and compared the results to the empirical cumulative distribution function computed from simulated data. For simplicity, we set the linear indexes to be constant valued vector with each element equal to 0.5. The dispersion parameters (omegas) are set at 0.75 for both count variables (Thus: we simulated under dispersed count variables), and the correlation coefficient rho is set at 0.5. Table 2 provides the bivariate CMP data generator validation result. The support of simulated Y1 is [0, 14], and the support of simulated Y2 is [0, 13]. Again, we can observe that our approximation software can approximate the cumulative density for at least two decimal points. For instance, according to our 2D Gauss-Legendre Quadrature, Prob (Y1 = 0) = 0.3278748, while the actual relative frequency of Y1 = 0 in the simulated Bivariate CMP dataset with sample size of 500,000 is equal to 0.32888.

4.4: Advantages

There are three main advantages of our Bivariate Gauss-Legendre Quadrature software. (1) It can approximate double integrals without any restrictions on the form of the integrand. (Hermitian Quadrature has such restrictions since it requires necessary transformations of the underlying integrand before it can be approximated using Hermitian integration); (2) It can accurately approximate any double integrals with quadrature points around 30 or higher. We will show validation of the accuracy of our software in the next section; (3) It significantly reduces the computational time compared to Monte Carlo based integral approximation, and Bayesian methods. The author has attempted one of the alternative methods: Monte Carlo integration for approximating double integrals during preparing the dissertation, the time needed for simulating 50,000 observations of bivariate Poisson data with 5000 Halton draws is roughly 5 hours, for CMP distribution, the time needed is even much longer. While the 2D Gauss-Legendre Quadrature program takes less than 5 seconds for simulating 500,000 observations, thanks to the efficiency of Stata/Mata's matrices operations.

Chapter 5: Simulation Study of the Importance Accounting for Nonlinearity, Cross-Equation Correlation and Dispersion Flexibility in Count-Outcome SUR Models

In this section, we evaluate, via simulation studies, the efficiency and accuracy gains of estimating a bivariate count valued systems-of-equations models via three comparisons:

(1) Zellner SUR and Count-Outcome SUR (CMP);

(2) CMP SUR Joint Estimation and Single Equation CMP Estimation;

(3) CMP SUR Joint Estimation and Poisson SUR Joint Estimation.

The first comparison allows us to assess the importance of "accounting for nonlinearity," i.e., of using count-outcome estimators on count data, as opposed to using linear SUR estimators. The second comparison allows us to assess the importance of joint versus equation-by-equation estimation the presence of cross-equation correlation. The third comparison allows us to assess the role of the CPM versus Poisson marginals.

In the subsequent subsection -- 5.1, 5.2 and 5.3 -- we detail the models to be compared (i.e., probability mass functions, conditional mean functions, likelihood functions and effect parameters and the standard errors of the effect parameters), the simulation study designs, as well as the simulation results of the three pairs of model respectively.

5.1: Comparison Between Linear SUR and Count-Outcome SUR (CMP) Via a Simulation Study (Study I)

In this section, we conduct the first pair of analytical comparison via simulation to answer the question: "does modeling nonlinearity matter, i.e., does the use of count-outcome estimators on count data offer improvement over the use of linear estimators?"

To answer this question, we compare the accuracy and efficiency gains of the average incremental effect (AIE) estimation between Zellner (1962)'s Linear Seeming Unrelated Regression model and our Count-outcome Seemingly Unrelated Regression model.

It is known that linear method is widely used in empirical economics and health science research, i.e., difference-in-differences, fixed effects estimations, as well as linear instrumental variable approach. The main motivation for linear models is the simplicity in its analytical forms, the known properties of its estimator, and ease in execution by existing built-in computational software. However, a priori there are potential shortcomings of conventional linear models when used with count-valued outcome(s). first, OLS does not account for the fact that count data are truncated at zero; thus, it could predict negative values for count-valued outcomes which are inherently restricted to positive (Woodridge, 2010; Gardner et al., 1995). Second, since count data are skewed to the right, they are unlikely to satisfy the normality assumption of OLS, making statistical tests based on this assumption invalid (Cameron and Trivedi, 2005; Gardner et al., 1995). Third, the validity of hypothesis tests in the OLS also depends on assumptions about homoscedasticity, which are unlikely to be met in count data (Winkelmann, 2018; Gardner et al., 1995, Cameron and Trivedi, 2005). Lastly, there are potential efficiency gains in terms of averaged treatment effects or averaged incremental effects when accounting for the nonlinearity (against using simple linear mode) to fit correlated count data. To date, not much work has been done to empirically or theoretically explore these shortcomings of applying linear system-of-equation models in count-data settings.

Our model takes an alternative approach to account for the nonlinear nature of the correlated count data, and in turn introduces complexity in model specification and potential difficulties in estimation. The main motivation for our modeling approach is the possible accuracy and efficiency gains in deep parameters as well as effect parameter estimations in the context of Seemingly Unrelated Regression. We thus conjecture that there are accuracy and efficiency gains in accounting for the nonlinearity of our count data. We will conduct a series of simulation studies to compare the AIE estimates by using Zeller (1962)'s linear SUR model with our model to investigate these hypotheses.

5.1.1: Zeller (1962)'s Linear SUR Model Specification

The data generating process of Zellner's (1962) original specification is:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_2 \end{pmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & X_J \end{bmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_J \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_J \end{pmatrix} \tag{51}$$

Where the error terms u has variance covariance structure of:

$$V(u) = \Sigma = \begin{bmatrix} \sigma_{11}I & \sigma_{12}I & \cdots & \sigma_{1J}I \\ \sigma_{21}I & \sigma_{22}I & \cdots & \sigma_{2J}I \\ \vdots & \vdots & & \vdots \\ \sigma_{J1}I & \sigma_{J2}I & \cdots & \sigma_{JJ}I \end{bmatrix} \tag{52}$$

where I is a unit matrix of order $T \times T$, σ_{ij}=E($u_{it}u_{jt}$) for $t = 1,2 \ldots., T$, and i, j=1, 2 , …, M.

The data generating process of the Zellner's model is as below, the first moment has the form:

$$E(Y|X) = \left[X_1\beta_1,\ldots,X_J\beta_J\right]'_{J\times 1}$$

(53)

And the second moment of Zellner's model has the form:

$$Var(Y|X) = \begin{bmatrix} \sigma_{11}I & \sigma_{12}I & \cdots & \sigma_{1J}I \\ \sigma_{21}I & \sigma_{22}I & \cdots & \sigma_{2J}I \\ \vdots & \vdots & & \vdots \\ \sigma_{J1}I & \sigma_{J2}I & \cdots & \sigma_{JJ}I \end{bmatrix}$$

(54)

where X is the matrix of explanatory variables and Y is the J×1 vector of outcomes, β_i represent equation $\boldsymbol{i}$'s deep parameters ($i = 1, \ldots J$), σ_{ij} is the covariance of equation i and $j's$ error terms, σ_{ii} is the variance of equation $i's$ error term, and I is a unit matrix of order N × N. This model is conditionally first-and second-order homoscedastic. It is assumed in Zellner's model, in each equation, the variances of the errors (or dependence variable) are the same for all observations, the means are the same for all observations.

The linear SUR estimation procedure is as below. The linear seemingly unrelated regression estimator by Zellner is a two-step estimator. First, I estimate each equation using ordinary least squares and save the predicted residuals. Second, I use the predicted residuals to form a weight matrix and estimate the parameters using weighted least square.

The estimator has the form:

$$\hat{\beta}_{sur} = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

(55)

Where X is the matrix of predictors of all equations stacked together, and V the Kronecker product of S and I, and it has the form below:

$$V = S \otimes I_N \tag{56}$$

Where *S* is the variance covariance matrix of the OLS residuals, and *I* is an identity matrix of size equal to the numbers of observations. The CMP SUR model is detailed in Chapter 3 under the generalization section.

5.1.2: Conditional Mean Function and AIE Estimation of Zeller (1962)'s Linear SUR Model

For the case where j = 2, given X_0, the "Average Incremental Effect (AIE)" under Zellner's linear SUR model is the expected change in the outcome vector that results from an increase in the policy variable from X^{pre} to $X^{pre} + \Delta$, where $\Delta > 0$ is the size of the increment. In symbols,

$$AIE(\Delta) = E[m(X^{pre} + \Delta, X_o;\ \pi)] - E[m(X^{pre}, X_o;\ \pi)] \tag{57}$$

where m(.) is the expected value of the outcome vector given $(X,\ X_0)$, i.e.,

$$m(X, X_0; \pi) = E[Y \mid X,\ X_0; \pi] = \begin{bmatrix} X\beta_1 + X_0\beta_{01} \\ X\beta_2 + X_0\beta_{02} \end{bmatrix}$$

(58)

The estimated AIE is:

$$\widehat{AIE}(\Delta) = \sum_{i=1}^{N} \frac{1}{N} \begin{bmatrix} (X_i + \Delta)\widehat{\beta_1} + X_{oi}\widehat{\beta_{o1}} - ((X_i)\widehat{\beta_1} + X_{oi}\widehat{\beta_{o1}}) \\ (X_i + \Delta)\widehat{\beta_2} + X_{oi}\widehat{\beta_{o2}} - ((X_i)\widehat{\beta_2} + X_{oi}\widehat{\beta_{o2}}) \end{bmatrix}$$

(59)

and the standard error of the average incremental effect is formulated as:

The standard error of the AIE for the Zellner SUR estimator is:

$$\text{s.e.}(AIE)^{Zellner} = \sqrt{\text{Asy VAR}(AIE)^{Zellner}}$$

where

$$\text{Asy VAR}(AIE)^{Zellner} =$$

$$= \left(\frac{\sum_{i=1}^{n} \nabla_\pi \widehat{aie}_i}{n}\right)\left(\widehat{AVAR}(\hat{\pi})\right)\left(\frac{\sum_{i=1}^{n} \nabla_\pi aie_i}{n}\right)' + \frac{\sum_{i=1}^{n} (\widehat{aie}_i - AIE)^2}{n}$$

$$= \left(\frac{\sum_{i=1}^{n} \Delta}{n}\right)\left(\widehat{AVAR}(\hat{\pi})\right)\left(\frac{\sum_{i=1}^{n} \Delta}{n}\right)' + \frac{\sum_{i=1}^{n} \Delta\widehat{\beta}_1 - \widehat{\beta}_1)^2}{n}$$

$$= 1\left(\widehat{AVAR}(\hat{\pi})\right)1 + 0$$

$$= \widehat{AVAR}(\hat{\pi}).$$

(60)

5.1.3: Design and Results

This section shows the deep parameter estimation results of Zellner's SUR model versus Count-Outcome SUR model (CMP case). Before showing the results, we introduce our simulation design.

5.1.3.1: Simulation Design

We focused on simulating a series of correlated and over-dispersed data. We would like to use our bivariate CMP distributed simulated data as a benchmark, so we could fit and test the performances of different candidate models.

The main motivation of simulating such a dataset is that we often encounter health care utilization counts that have similar properties – correlated, highly skewed, and (sometimes or often) over-dispersed, i.e., where the mean of the outcome variable is small (smaller than the variance). This focus is motivated as follows: when the mean is large and the variance is small, the conditional distribution of count data resembles the classical bell shape required by least squares and thus Zellner's SUR. However, when the means are small and the variance (comparatively large), the conditional distribution of count data exhibits a skew shape that is not well-captured by the classical bell shape. Thus, least squares and derivative methods are expected to have performance challenges on such data. Such data, consequently, provides fertile ground to investigate the similarity/differences. of the estimators.

We developed Stata/Mata code to simulate data for bivariate CMP data. The protocol for the simulator is as follows:

1) Choose values for the elements of the parameter vector

$$\pi' = [\beta'_1, \beta'_2, \omega_1, \omega_2]\ (\beta'_1 = [\beta_{1p}, \beta'_{1o}] \text{ and } \beta'_2 = [\beta_{2p}, \beta'_{2o}])$$

where ω_1, ω_2 are the dispersion parameters of each outcome,

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

3) Generate a sample of simulated data of η_1^*, η_2^*; assume to be bivariate standard normal with mean zeros and variance parameters both 1, and correlation coefficient ρ_{12} chosen as part of the sampling design,

4) Generate a sample of outcomes vector by approximating the value of the joint probability mass function:

$$f(Y_{1X^*}, Y_{2X^*} \mid X, X_0) =$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \text{cmp}(Y_{1X^*}, X^*, X_0, \eta_1^*, \omega_1)\ \text{cmp}(Y_{2X^*}, X^*, X_0, \eta_2^*, \omega_2)\ \ g^*(\eta_1^*, \eta_2^*)\ d\eta_1^*\ d\eta_2^*$$

To compare the efficiency of the estimated averaged AIE estimator based on corresponding equations in Chapter 3, we simulated samples with fixed sample size with large number of replications, using the data generator detailed in the previous section and applied the bivariate CMP MLE of the AIE to each of them, whose likelihood functions were approximated by Gauss-Legendre Quadrature rules.

Regarding our parameter choices, for all our sub-studies, we take X to be a uniformly distributed random variable with support [0.13, 1.87] and $X_0 = (X_{01}, X_{02})$ to be a random vector where X_{01} is uniformly distributed on [0.13, 1.87] and $X_{02} = 1$ is a constant. In the sub-study on under-dispersion, for simplicity, β_1,β_2 are set to be the same: we take $\beta_1 = \beta_2 = 1,\ \beta_{01} = \beta_{02} = (-1, 0)',\ \sigma_1 = \sigma_2 = 1,\ \omega_1 = \omega_2 = -0.1.$

However, the sampling design varies along one margin: four chosen ρ_{12} values determining the correlation coefficient of the bivariate normal distributed cross-equation heterogeneity terms η_1^*, η_2^*. These values are the following:

$\rho_{12} = 0.75$ (strong positively correlated)

$\rho_{12} = 0.5$

$\rho_{12} = 0.25$

$\rho_{12} = 0$ (uncorrelated)

(Note that $\rho_{12} = 0.9$ was not included in the sampling design under CMP model because, in this case, the MLE typically did not converge and it usually took much longer for the optimization program to finish.) This design is summarized in Table 20.

The summary statistics of one simulated dataset is displayed in Table 21. As we can see, the outcome variables have means of 2.3, and standard deviations of 3.9. it is apparently over-dispersed (CM < CV). To illustrate the skewness of this simulated data, I created Figure 3, which illustrate the distributions of all simulated variables. As we can see, the count data, conditionally on all covariates, is not normally distributed – it is clustered around its (very small-valued) mean. Moreover, the logarithm of the outcomes is also not close to normally distributed. This nature of the data necessitates the introduction of some type of nonlinear specifications that align with the true data generating process.

Note that we chose the sampling design especially to exploit the skewness of the distribution, where the correlated count model likely will show advantages. As the conditional mean of the count outcome gets large while the variance is very small, the data likely can be well fitted with a linear specification.

The simulation study also included a design of replication for the purpose of exploring the asymptotic statistical properties of the deep parameters and effect parameters.

For each of these designs, 100 sample of size 10,000 were generated. In each replication, we

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

2) Estimate the effect parameters using CMP SUR model,

3) Estimate the effect parameters using Zellner Linear SUR model (not accounting for nonlinearity of the outcome data),

4) Repeat 2) and 3) with 100 replications,

5) Compared the 100-replication averaged AIE to the true AIE using criteria: the grand average of the absolute percentage bias (AAPB).

The first comparison criteria the Grand Average of the Absolute Percentage Bias is calculated by the following formula,

$$\text{AAPB } \widehat{\text{AIE}(\Delta)} = \frac{1}{R} \times \sum_{r=1}^{R} \left| \frac{\widehat{\text{AIE}(\Delta)}_r - \text{AIE}(\Delta)}{\text{AIE}(\Delta)} \right| \tag{61}$$

where, r denotes the r^{th} replication, $\widehat{\text{AIE}(\Delta)}$ denotes the estimated AIE with the increment being equal to Δ and $\text{AIE}(\Delta)$denotes the true AIE value. Without loss of generality, we choose $\Delta = 1$.

We conducted similar data generator validation exercise for our Bivariate CMP data as in Chapter 4.3. We display the result for one sampling design, where $\rho_{12} = 0.5$, and $\beta_1 = \beta_2 = 1,\ \beta_{01} = \beta_{02} = (-1, 0)',\ \sigma_1 = \sigma_2 = 1,\ \omega_1 = \omega_2 = -0.1$. The other sampling designs can be validated using the same method and code.

5.1.3.2: Results

Table 13 displays the estimation of deep parameters of Zellner's Linear Seemingly Unrelated Regression (SUR) Model with simulated over-dispersed data (N = 50,000), the true parameter values are also shown for comparison purposes. As we can see, the linear SUR model tends to overestimate all coefficient parameters.

Table 16 displays the effect parameter (Average Incremental Effect - AIE) estimation results of the first outcome variable Y_1 under both Linear SUR Model and Count-Outcome SUR model (CMP case). We report the running average of the estimated AIE over 100 replications of simulated data and 100 replications of the corresponding AIE based on the MLE results, we also report the averaged absolute percent bias of the AIE estimates.

As we can see, for instance, the average AIE under Linear SUR model is 2.185 (with averaged absolute percent bias 53%), while our Count-Outcome SUR model estimated average AIE is 4.502 (with averaged absolute percent bias 10%). This implies that our bivariate CMP model estimation is more accurate (|4.5 – 4.765|/4.765 < (|2.185 – 4.765|/4.765 in terms of effect parameter estimation), and also more efficient (with much less variation in 100 replications: 10.48% < 53.06%). The Linear SUR model tend to under-estimate the AIE when data are over-dispersed.

5.2: Comparison between CMP SUR and Single Equation CMP Approach via a Simulation Study (Study II)

In this section, we present the results of simulation studies that compare the accuracies and efficiencies of the parameter and AIE estimates when the model is estimated jointly versus on an equation-by-equation basis.

For context, in Chapter 2, described the joint estimation of the model by maximizing the joint log-likelihood of (Y_1,Y_2), to obtain estimates of the full parameter vector. In contrast, estimation-by-equation estimation of the model involves two steps. First, we focus on Y_1 and then maximize the log-likelihood of its *marginal* distribution (in our data) to estimate only the parameters that are relevant to Y_1. Second, we focus on Y_2 and then maximize the log-likelihood of its *marginal* distribution to estimate only the parameters that are relevant to Y_2. Key to performing equation-by-equation estimation are the marginal distributions of Y_1 and Y_2, which we refer to as the "single equation model."

In our simulation exercises, we generate a sample of size N according to the chosen distribution at pre-specified true parameters. We will then estimate the parameters both in joint and equation-by-equation fashions and compute the associated AIEs under each approach. We then write these to a database and replicate the sampling-estimation exercise 100 times. We finally leverage these replications to explore the efficiency and accuracy of the estimation methods.

To illustrate our simulation result, we chose to compare between Count Outcome SUR model and Single Equation Count model. The CMP SUR model is again detailed in Chapter 3 under the generalization section. Here, we will focus on detailing the generic single equation count model specification and estimation.

5.2.1: Generic Single Equation Model

In the single equation model, for j = 1, 2, the distribution of Y_j given covariates X and X_0 is given by the marginal distribution of Y_j under the model, i.e.,

$$f(Y_j \mid X, X_0; \pi) = \sum_{y_{-j}=0}^{\infty} f(Y_j, Y_{-j} = y_{-j} \mid X, X_0; \pi)$$

$$= \int_{-\infty}^{\infty} f_{(Y_j \mid X, X_0, \eta_1)}(Y_j \mid X, X_0, \eta_j; \pi)\, g(\eta_j)\, d\eta_j$$

(62)

where $g(\eta)$ is a univariate standard normal probability density function and $f_{(Y_j \mid X, X_0, \eta_1)}(\cdot)$ is the same as our baseline model. The second equality follows from a bit of algebra:

$$\sum_{y_{-j}=0}^{\infty} f(Y_j, Y_{-j} = y_{-j} \mid X, X_0; \pi) = \sum_{y_{-j}=0}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [f_{(Y_j \mid X, X_0, \eta_1)}(Y_j \mid X, X_0, \eta_j; \pi)$$

$$\times f_{(Y_{-j} \mid X, X_0, \eta_{-j})}(y_{-j} \mid X, X_0, \eta_2; \pi)\, g(\eta_j, \eta_{-j}; \widehat{\rho_{12}})] d\eta_j\, d\eta_{-j}$$

$$= \int_{-\infty}^{\infty} [f_{(Y_j \mid X, X_0, \eta_1)}(Y_j \mid X, X_0, \eta_j; \pi)$$

$$\times \int_{-\infty}^{\infty} g(\eta_j, \eta_{-j}; \widehat{\rho_{12}}) \sum_{y_{-j}=0}^{\infty} f_{(Y_{-j} \mid X, X_0, \eta_{-j})}(y_{-j} \mid X, X_0, \eta_2; \pi)]\, d\eta_{-j} d\eta_j$$

$$= \int_{-\infty}^{\infty} f_{(Y_j \mid X, X_0, \eta_1)}(Y_j \mid X, X_0, \eta_j; \pi)\, g(\eta_j)\, d\eta_j$$

(63)

The first equality is definitional – see Chapter 3. The second equality is due to a change in the order of integration. And the third equality is due to the fact that the sum of any marginal distribution over its support is equal to one, i.e., $\sum_{y_{-j}=0}^{\infty} f_{(Y_{-j} \mid X, X_0, \eta_{-j})}(y_{-j}, X, X_0, \eta_2; \pi) = 1$, and that $g(\eta_j) = \int_{-\infty}^{\infty} g(\eta_j, \eta_{-j}; \widehat{\rho_{12}})\, d\eta_{-j}$.

Observe that the single-equation model above does not involve ρ_{12}. Thus, the maximum likelihood estimation of the single equation model for either j = 1 or j = 2 does not provide an estimate of the correlation of the structural errors.

The single equation model is estimated via MLE using univariate Gauss Legendre Quadrature. For j = 1, 2, the associated log-likelihood is

$$L_j(\pi) = \sum_{i=1}^{N} \{\ln f(Y_{ji}| X_i, X_{oi}; \pi)\}$$

$$= \sum_{i=1}^{N} \ln \int_{-\infty}^{\infty} f_{(Y_j \mid X, X_0, \eta_1)}(Y_j, X, X_0, \eta_j; \pi)\, g(\eta_j)\, d\eta_j$$

(64)

This equation is maximized using methods analogous to those employed in Section 2.

The above equation implies that, for j = 1, 2, the conditional mean function of the single equation model is,

$$m_j(X, X_0; \pi) = E(Y_j|X, X_0; \pi) = \sum_{y=1}^{\infty} y\, f(Y_j = y \mid X, X_0).$$

(65)

Example Case I: Poisson Single Equation Model

For j = 1, 2, the single observation probability mass function in the Poisson model is

$$f_{(Y_j \mid X, X_0)}(Y_j|X, X_{0i}; \pi) = \int_{-\infty}^{\infty} \frac{(\lambda_j)^{Y_j} \exp(-\lambda_j)}{Y_j!} g(\eta_j)\, d\eta_j$$

(66)

where $\lambda_j = \exp(X\beta_j + X_{0i}\beta_{0j} + \sigma_j\eta_j)$. And the conditional mean function of the potential outcomes Y_j for j = 1, 2 is

$$m_j(X, X_0; \pi) = E(Y_j|X,\ X_0; \pi) = \int_{-\infty}^{\infty} \exp\left(X\beta_j + X_0\beta_{0j} + \sigma_j\eta_j\right) g(\eta_j)d\eta_j\ . \tag{67}$$

We first describe the details of the simulation and then we investigate the efficiency properties of joint versus equation-by-equation estimation in terms of efficiency and accuracy as the sample size increases. Throughout, we focus on the case of Poisson marginals.

5.2.2: Example Case I: Poisson Single Equation Model

5.2.2.1: Simulation Results

In this study, we use a Bivariate Poisson generator as the benchmark for joint versus single equation model comparison, the simulation design mirrors the design of the bivariate CMP generator in Study I, although there are slight changes on the parameters. See Table 19 for the parameter settings for this simulation. In terms of the procedure for replications, we again follow Study I but replace step (2) and (3) with Poisson SUR and Poisson Single-Equation models respectively.

Tables 2 presents an example of the deep parameter estimation results under the joint estimation using for a trial run with 50,000 observations. The table shows that all deep parameters are close to their true values and thereby validates the core accuracy of our estimation approach, including quadrature methods. Further, estimation of these results took less than 30 minutes on a modern laptop, which shows that our methods are computationally efficient.

Table 4 reports the true AIE and the estimated AIEs for each method, as well as the AAPB. The table shows, for instance, that at a $\rho_{12} = 0.9$, the AIE of the Poisson joint and equation-by-equation estimates are close to the true AIE; but that the joint estimate is about twice as efficient as the equation-by-equation estimate – e.g., AAPB of 3.51% versus 6.76%. Since this is a common result for each row of the table, we conclude that joint estimation is generally more efficient than equation-by-equation estimation in the case of Poisson marginals.

5.2.3: Example Case II: Conway-Maxwell-Poisson Single Equation Model

The single equation CMP model is straight forward: the probability mass function under the CMP distribution is given by the marginal distribution of Y_j for j = 1, 2. Specifically,

$$f_{(Y_j \mid X, X_0)}(Y_j, X, X_{0i}; \pi) = \int_{-\infty}^{\infty} \frac{(\lambda_j^o)^{Y_j}}{(Y_j!)^{\exp(\omega_j)} Z(\lambda_j^o, \omega_j)} g(\eta_j)\, d\eta_j \tag{68}$$

where Z $(\cdot)$ and λ_j^o are given in Section 3.2.1.

5.2.3.1: Simulation Results

In this simulation study, we follow the same simulation design and use the same bivariate CMP outcome generator as Study I. In terms of the procedure for replications, we again follow Study I but replace step (3) with CMP Single-Equation model.

Table 14 presents an example of the deep parameter estimation results under the joint estimation using for a trial run with 50,000 observations of the over-disperse sub-simulation. The table shows that all parameters estimated using our CMP model are close to the true parameter values, thereby validating the effectiveness of the core estimation routine.

Table 17 reports the true AIE and the estimated AIEs for each method, as well as the AAPB, for the over-disperse sub-simulation. The table shows, for instance, that at a $\rho_{12} = 0.5$, the AIE of the joint and equation-by-equation estimates are approximately equal to the true AIE and that both methods have an AAPB of approximately 25%, with joint estimation being slightly more efficient. Other rows of the table – e.g., $\rho_{12} = 0.75$, show that both estimation methods have approximately the same efficiency, with equation-by-equation estimation being slightly more efficient. The overall finding for CMP marginals with over-dispersed data is thus that joint estimation does not perform worse than equation-by-equation estimations, which is consistent with the conclusion that joint estimation is slightly more efficient than equation-by-equation estimation.

5.3: Comparison Between Count-Outcome Conway-Maxwell-Poisson SUR and Count-Outcome Poisson SUR via a Simulation Study (Study III)

In Study III, we conduct a simulation comparison between our bivariate CMP model and bivariate Poisson model. We again follow the same simulation design and use the same bivariate count outcome generator as Study I. In terms of the procedure for replications, we follow Study I but replace step (3) with Poisson SUR model.

Table 15 Columns 3-4 and 5-6 display all MLE deep parameter estimates and the standard errors under: (1) Count-Outcome SUR Model (Poisson case) and, (2) dispersion Flexible Count-Outcome SUR Model (Conway-Maxwell-Poisson case) on 50,000 simulated over-dispersed data with correlation parameter sets to 0.75 and dispersion parameter for both outcomes set to -0.1. The outcome variable Y_1 has mean of 2.3671, variance of 15.45, minimum of 0, and maximum of 25. The second outcome variable Y_2 has mean of 2.337, variance of 15.13, minimum of 0 and maximum of 25. This data is over-dispersed since the conditional mean is smaller than the conditional variance. Traditional Poisson distribution unfortunately is not able to model any data generating process that has non-equi dispersion, since Poisson mean parameter is equal to its variance parameter. This suggests that we might need to introduce Conway-Maxwell-Poisson marginals to better analyze data of this type.

5.3.1: Simulation Results

Table 18 shows the simulation results of our two effect parameters [AIEs] over 100 replications with 10,000 observations for each replication) as well as the averaged absolute percent bias in both models. We chose four correlation designs: $\rho_{12} \in \{0, 0.25, 0.5, 0.75\}$.

As we can see, our dispersion flexible SUR model performs with more accuracy and efficiency in terms of effect parameters (AIEs). For instance, when rho = 0.75, the Conway-Maxwell-Poisson SUR model estimated averaged AIE is 4.502 (with 10.48% average absolute percent bias), while the Poisson SUR model estimated averaged AIE is 4.034 (with 34.19% average absolute percent bias). In this model, the true AIE is 4.765. The CMP SUR model estimates the AIE with better precision: (|4.765 - 4.502|/4.765 < (|4.765 - 4.034|/4.765) as well as smaller variation: 10.48% < 34.19%. The takeaway is that the Poisson SUR model is insufficiently flexible to handle certain types of over-disperse data, necessitating the use of the CMP SUR model.

Chapter 6: Dispersion-Flexible Count-Outcome SUR Estimation in a Real-Data Context

We have observed from the previous simulation studies that there might be significant differences using Linear SUR model versus our CDCR model in terms of the accuracy and efficiency of policy effect parameter estimations. To illustrate the utility of the Count Outcome SUR methods built in previous chapters of this dissertation, we consider an application of the Poisson and Conway-Maxwell Poisson marginal model to medical utilization data. We conduct a series of estimations and comparisons to illuminate the tradeoffs between the models, specifically: (1) Comparison between Linear SUR and Count-Outcome SUR (CMP); (2) Comparison between CMP SUR and Single Equation CMP Approach; (3) Comparison between Count-Outcome CMP SUR and Count-Outcome Poisson SUR. Note that (i) all comparisons will be based on accuracy and efficiency in estimation of the effect parameters - EP (in this case the average incremental effect AIE) and (ii) that, per our simulation results in Chapter 5, we regard the Count Outcome CMP SUR model as the benchmark. Thus, one may think of the present exercise as illustrating the magnitudes of possible biases and efficiency loss in effect parameter estimations if we do not account for nonlinearity, cross equation correlation, or dispersion flexibility.

6.1: Data Overview

A central question of substantive policy interest is the extent to which the use of health services depends on insurance coverage (e.g., Chib and Winkelmann, 2001). To address this question, we apply our model to medical care utilization data from the 1987 National Australian Medical Expenditure Survey. This data is used by many previous works such as Deb and Trivedi (1997), Chib and Winkelmann (2001), and Famoye (2015).

The 1987 National Australian Medical Expenditure Survey Data reports several types of correlated count, medical utilization data – e.g., number of physician office visits, number of non-physician office visits, number of emergency room visits, etc. – and additional covariates – e.g., private insurance, geography, and so on – for elderly Australians over the age of 65. Table 5 provides detailed summary statistics.

We use a subset of the variables as part of the present exercise. Our outcomes are (i) the number of physician office visits, denoted Y_o, and (ii) the number of non-physician office visits, denoted Y_1. Our covariates consist of a constant and all patient characteristics listed in Table 5, they are denoted X_o. And our policy variable is an indicator variable of a whether a patient has private health insurance, which is denoted X. For simplicity, we assume that the policy variable is exogenous to the other covariates and the outcomes; This assumption allows us to focus on the comparison of the methods without the complication of developing a control strategy for endogeneity (e.g., instrumental variables). One of the goals of future work is to fully incorporate econometric approaches for handling with endogenous policy variables.

We choose $\Delta = 1$, so that the AIE gives (i) the difference between physician office visits in a counterfactual where everyone has private insurance and a counterfactual where no one has private insurance and (ii) the difference between non-physician office visits in a counterfactual where everyone has private insurance and a counterfactual where no one has private health insurance. In other words, it answers the questions: what happens to average physician and non-physician office visits as everyone in the population is switched from not having private insurance to having private insurance?

6.2: Comparison Between Linear SUR and Count-Outcome SUR (CMP) in Real Data

In this section, we compare the empirical results of the effect parameters (AIEs) under linear system of equation models versus count-outcome system of equation model. The main objective for such comparison is to answer the question: *how much does accounting for nonlinearity matter in policy effect estimation?*

Table 6 shows the deep parameter estimations of the two correlated count outcomes under Zellner's Linear SUR model. (I also included OLS model estimation results and put it next to Zellner's Linear SUR model, so we can easily observe the potential efficiency gains in some deep parameters.)

Table 9 shows the Policy Effect Parameter (AIE) Estimation Results using National Medical Expenditure Survey Data between Zellner's Linear SUR model and our count-outcome SUR model – Conway Maxwell Poisson case.

As we can see, the AIE(Y_1) of our first outcome under Count-Outcome SUR Model - Conway-Maxwell-Poisson case is 2.5637, with standard error of AIE(Y_1) equal to 0.763. The AIE (Y_2) of our second outcome under Count-Outcome SUR Model - Conway-Maxwell-Poisson case is 22.358, with standard error of AIE(Y_2) equal to 7.611.

Linear SUR model estimated AIE(Y_1) is equal to 1.6302, with standard error 0.2784, and the Linear SUR Model estimated AIE(Y_2) is equal to 0.5958, with standard error 0.2288. The effect parameter estimates between Linear SUR model and our Count-Outcome SUR (CMP case) model differs significantly. The Linear SUR model tends to underestimate the average incremental effects when the outcome variable shows nonlinearity. To visually illustrate the nonlinearity, see Figure 2 -- Histograms of Highly Skewed Distributed Count Outcomes (Dependent Variables) of 1987 National Australian Medical Expenditure Survey Data. We also observe that the data exerts nonlinearity even after taking logarithmic transformation.

The standard errors of AIE under Linear SUR model are much smaller than those from and our Count-Outcome CMP model. This is a classical bias-efficiency tradeoff.

6.3: Comparison Between CMP SUR and Single Equation CMP Approach in Real Data

In this section, we compare the empirical results of the effect parameters (AIEs) under Single Equation Count-Outcome model versus Count-Outcome SUR model (Conway-Maxwell-Poisson case). The main objective for such comparison is to answer the question: *how much does accounting for potential cross-equation correlation (estimation under a system-of-equation or SUR framework) matter in policy effect estimation?*

6.3.1: Poison Case Empirical Results

Table 7 and Table 10 give the estimation results for both the joint and equation-by-equation Poisson model. Specifically, Table 7 reports the deep parameters and Table 10 reports the AIEs. In terms of deep parameter estimation, Table 7 shows that the joint estimation method typically achieves strictly larger t-statistics and thus smaller standard errors than under the equation-by-equation estimation method.

In terms of EP estimation, Table 10 shows, under joint estimation, that the policy effect of having private insurance on the number of doctor office visits is 1.88 and that the policy effect of having private insurance on the number of non-doctor office visits is 4.00. Further, under equation-by-equation estimation that ignores the cross-equation correlation, the table shows that the policy effect of having private insurance on the number of doctor office visits is 2.53 and that the policy effect of having private insurance on the number of non-doctor office visits is 2.98. The single equation model overestimates the policy effect on the number of doctor office visits and underestimates the policy effect on the number of non-doctor office visits.

In terms of the standard errors of our target EPs, Table 10 shows that the standard errors of the AIEs from the joint estimation are 0.504 and 0.478 for physician and non-physician visits respectively, while the standard errors of the policy effects under equation-by-equation estimation much larger at 0.734 and 1.590 for physician and non-physician visits respectively. Thus, we again see that joint estimation is more efficient than equation-by-equation estimation.

6.3.2: CMP Case Empirical Results

As we can see from the data descriptive, the outcome variables are over-dispersed, i.e., the conditional mean is smaller than the conditional variance, for instance, the mean of our second outcome Y_2 - number of non-doctor office visits is 1.62, while the variance of Y_2 is 28.27, the conditional mean of our first outcome Y_1 - number of doctor office visits is 5.77, while the conditional variance of Y_1 is 45.56. This also argues against the appropriateness of the Linear SUR case. This indicates that we need a model that is able to account for flexible dispersions, which traditional Poisson model cannot account for (due to its single parameter restriction that forces mean = variance).

In terms of deep parameter estimation, Table 8 shows the MLE deep parameter estimations of the two correlated count outcomes under our count-outcome SUR model – Conway Maxwell Poisson case. The comparison of single equation Conway Maxwell Poisson model with bivariate Conway Maxwell Poisson model are put side by side, so we can easily observe the potential efficiency gains in some deep parameters. We can also refer to Table 7 to compare our CMP model's deep parameter estimation results with Poisson model's deep parameter estimation results. Note that the additional parameters under CMP case are the dispersion parameters omega1 and omega2.

In terms of EP estimations, Table 11 shows the effect parameter estimation results (AIE) under our count-outcome SUR model – Conway Maxwell Poisson case. As we can see, the AIE(Y_1) of our first outcome under Count-Outcome SUR Model - Conway-Maxwell-Poisson case is 2.564, with standard error of AIE(Y_1) equal to 0.763. The AIE(Y_2) of our second outcome under Count-Outcome SUR Model - Conway-Maxwell-Poisson case is 22.358, with standard error of AIE(Y_2) equal to 7.611. The single equation Conway-Maxwell-Poisson model estimated AIE(Y_1) is equal to 2.644, with standard error 0.8597, and the single equation Conway-Maxwell-Poisson model estimated AIE(Y_2) is equal to 21.652, with standard error 7.251. The effect parameter point estimates are similar between single equation CMP and bivariate CMP model, and the standard errors results are mixed, there seems to be an improvement in efficiency for the AIE of first outcome (0.8597 > 0.7630), while in the case of our second outcome, the standard error of the bivariate CMP model is slightly larger: 7.2507 < 7.6115.

6.4: Comparison between Count-Outcome CMP SUR and Count-Outcome Poisson SUR

In this section, we compare the empirical results of the effect parameters (AIEs) between Dispersion Flexible Count-Outcome SUR (Conway-Maxwell-Poisson) and Count-Outcome Poisson SUR model. The main objective for such comparison is to answer the question: *how much does accounting for dispersion flexibility matter in policy effect estimation?*

Table 8 Columns 2 and 4, and Table 7 Columns 2 and 4 shows the MLE deep parameter estimations of the two correlated count outcomes under our count-outcome SUR model – Conway Maxwell Poisson case and our count-outcome SUR model – Poisson case.

In terms of EP estimations, Table 12 shows the effect parameter estimation results (averaged incremental effects -- AIEs) comparison between the count-outcome SUR model – Conway Maxwell Poisson case and the count-outcome SUR model – Poisson case. As we can see, the AIE (Y_1) under Count-Outcome SUR Model - Conway-Maxwell-Poisson case is 2.564, with standard error of AIE(Y_1) equal to 0.763. The AIE(Y_2) under Count-Outcome SUR Model - Conway-Maxwell-Poisson case is 22.358, with standard error of AIE(Y_2) equal to 7.611. The count-outcome SUR model – Poisson case estimated AIE(Y_2) is equal to 1.883, with standard error 0.504, and the count-outcome SUR model – Poisson case estimated AIE(Y_2) is equal to 4.009, with standard error 0.478.

We observe that, for this over-dispersed health care utilization data, the Poisson SUR model tend to underestimate the EPs, since it is not able to model the dispersion level of the data generating process. The standard errors of the AIE under Poisson SUR model is smaller for both outcomes (0.763 > 0.503; and 7.612 > 0.478).

Chapter 7: Summary, Discussion and Conclusions

In this section, we: i) discuss the significance of this dissertation; ii) list the limitation of the study; iii) conclude by pointing out a few avenues for future work.

7.1: Summary

This dissertation focuses on bivariate, count-value outcomes. Using an Aitchison and Ho (1989) type mixture approach in the context of the policy evaluation (and thus potential outcomes – e.g., Terza, 2019a), we examine whether joint estimation or equation-by-equation estimation provides greater efficiency for the deep parameters and estimated treatment effects. Based on simulation studies and applications to healthcare utilization data, we find strong evidence that joint estimation is typically more efficient.

The contribution of this dissertation is two folds. First, to the best of our knowledge, we are the first test and verify the classic assumptions in the count-outcomes literature that equation-by-equation estimation is less efficient than joint estimation for both deep parameters. Second, the model allows for omni-dispersion and thus relaxes classical dispersion-related restrictions present in prior works – e.g., Aitchison and Ho (1989) who use a Poisson-based approach. Third, most papers in the count-outcomes literature do not address treatment effects like Average Incremental Effects – e.g., Aitchison and Ho (1989), Chua and Tsiaplias (2019), and Kim et al. (2015) only focus on deep parameter estimation. This dissertation shows how to incorporate treatment effects into a flexible count-outcome framework and then demonstrates that joint estimation of the parameters improves the efficiency of treatment effects estimation; thereby providing an important new tool for policy analysis.

7.2: Discussion

Here are a few limitations of this study which might deserve further investigation: (1) The efficiency gains we observed might come from simulator's aspects; (2) It would be a plus to also provide theoretical proofs of such efficiency gain that we observed in our simulation studies; (3) We did not account for possible endogeneity issues, which will be a natural extension of this dissertation.

7.3: Conclusion

This dissertation represents the first step in an important line of research and much remains to be done. Presently, I am in the process of exploring additional marginal distributions and building out our two-dimensional quadrature program for general use. I also plan to extend the present framework to account for endogenous variables (via analogues of two-stage residual inclusion).

Table 1

Validation of Gauss-Legendre Quadrature Software's Accuracy: Comparison of GLQ-Based Marginal CDFs Versus Empirical Marginal CDFs From Simulated Bivariate Poisson Data

Support	Y1		Y2	
	GLQ-Based Marginal CDF	Empirical Marginal CDF	GLQ-Based Marginal CDF	Empirical Marginal CDF
0	0.823859309	0.824746	0.823859244	0.82493
1	0.965752483	0.966086	0.965752361	0.965966
2	0.99158493	0.991692	0.991584771	0.991814
3	0.997423585	0.997406	0.997423404	0.99741
4	0.999057349	0.99898	0.999057154	0.999038
5	0.999603662	0.999544	0.999603459	0.999578
6	0.999814232	0.999752	0.999814023	0.999808
7	0.999905068	0.999872	0.999904855	0.999892
8	0.999947951	0.999934	0.999947737	0.999942
9	0.999969734	0.99996	0.999969518	0.999972
10	0.99998148	0.999976	0.999981263	0.999984
11	0.999988136	0.99998	0.999987918	0.999994
12	0.999992073	0.99998	0.999991854	0.999998
13	0.999994492	0.999984	0.999994273	0.999998
14	0.999996029	0.99999	0.99999581	0.999998
15	0.999997031	0.999992	0.999996812	0.999998
16	0.999997698	0.999994	0.999997478	0.999998
17	0.999998149	0.999998	0.999997929	0.999998
18	0.999998461	1	0.999998241	0.999998
19			0.999998461	1

N=500,000

Figure 1

Gauss-Legendre Quadrature Approximated Joint PMF Versus Joint PMF of Simulated Bivariate Poisson Data

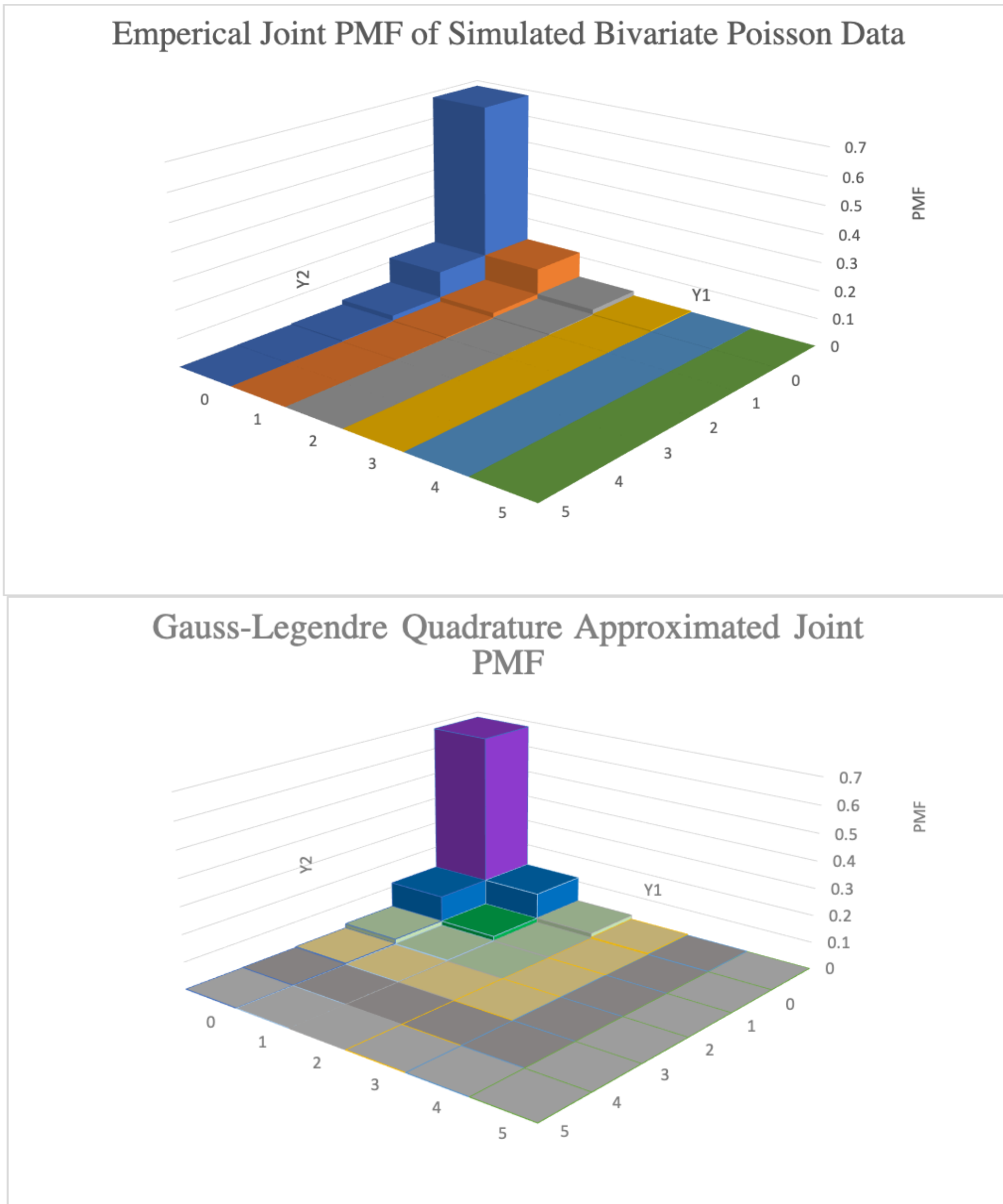


Table 2

Validation of Gauss-Legendre Quadrature Software's Accuracy: Comparison of GLQ-Based Marginal CDFs Versus Empirical Marginal CDFs From Simulated Bivariate Conway-Maxwell-Poisson Data

Support	Y1		Y2	
	GLQ-Based Marginal CDF	**Empirical Marginal CDF**	**GLQ-Based Marginal CDF**	**Empirical Marginal CDF**
0	0.3278748	0.328492	0.3278748	0.32888
1	0.7092633	0.709816	0.7092635	0.710044
2	0.8984288	0.89853	0.8984293	0.898684
3	0.9673462	0.967728	0.9673469	0.967364
4	0.9896771	0.989954	0.9896781	0.989898
5	0.9966822	0.996734	0.9966834	0.996738
6	0.9988979	0.998882	0.9988991	0.998926
7	0.9996185	0.999556	0.9996198	0.999622
8	0.9998617	0.999824	0.999863	0.999864
9	0.9999472	0.99994	0.9999485	0.99995
10	0.9999784	0.999974	0.9999798	0.999984
11	0.9999904	0.99998	0.9999917	0.999992
12	0.9999951	0.999992	0.9999964	0.999998
13	0.999997	0.999996	0.9999984	1
14	0.9999978	0.999998		
15	0.9999982	0.999998		
16	0.9999984	1		

omega1= omega2 = 0.75, rho = 0.5, N=500,000

Table 3

MLE of Deep Parameters of Bivariate Poisson Model and True Parameters

Two Outcomes	**Y1**	**Y2**	**Y1**	**Y2**
	Bivariate Poisson (Joint Estimation)		**TRUE**	
beta1	-.977 (0.0217)	-.9845 (0.022)	-1	-1
beta2	-1.020 (.0218)	-1.025 (.0221)	-1	-1
constant	.0045 (.0295)	-.0172 (.0298)	0	0
sigma squared	1.0102 (.0134)	1.0179 (.0135)	1	1
rho12	.8697 (.0137)		0.9	

N = 50,000, standard errors in parentheses

Table 4

Average Incremental Effect (AIE) Estimates using Bivariate Poisson Model Versus Single Equation Poisson Model: AIE, AAPB

	Based on 100 replication of size n = 10,000				
	True Model	**Single Equation Model**		**Bivariate Poisson Model**	
		Avg AIE	**AAPB**	**Avg AIE**	**AAPB**
rho12 = 0.9	-0.1801	-0.1869	6.76%	-0.1778	3.51%
rho12 = 0.75	-0.1801	-0.1930	9.61%	0.1749	2.81%
rho12 = 0.5	-0.1801	-0.1822	5.83%	-0.1703	3.48%
rho12 = 0.25	-0.1801	-0.1910	8.67%	-0.1750	3.10%

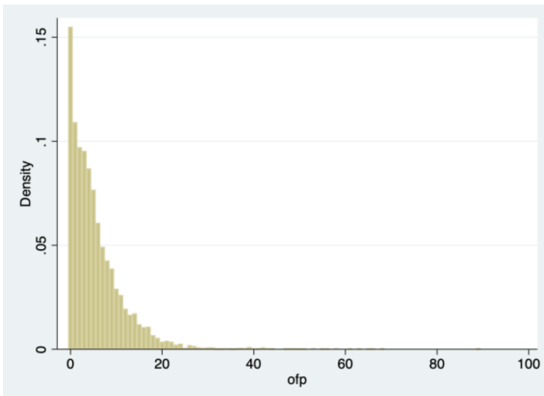
Table 5

Summary Statistics of 1987 National Australian Medical Expenditure Survey (NMES) Data

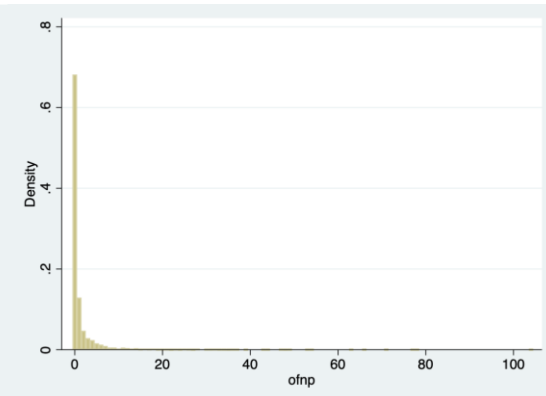
Variable Types	Variable Names	Mean	Std. Dev.	Min	Max
Policy Variable	**Private Insurance**	0.77644	0.416676 9	0	1
Correlated Count Outcome Variables	**Count of Physician Office Visits**	5.7744	6.7592	0	89
	Count of Non-Physician Office Visits	1.61802	5.3170	0	104
	Count of Physician Outpatient Visits	0.75079	3.6527	0	141
	Count of Non-Physician Outpatient Visits	0.53609	3.8795	0	155
	Count of Emergency Room Visits	0.2635	0.70365	0	12
	Count of Hospitalizations	0.29596	0.74639	0	8
Patient Characteristics	**Excellent Health**	0.07785	0.26796	0	1
	Poor Health	0.12574	0.33159	0	1
	Number of Chronic Diseases	1.54199	1.3496	0	8
	Daily Activity Difficulty Index	0.20404	0.40304	0	1
	North East	0.18997	0.39232	0	1
	Midwest	0.2626	0.44009	0	1
	West	0.18112	0.38515	0	1
	Age	7.40241	0.63340	6.6	10.9
	Black	0.11711	0.32159	0	1
	Male	0.40354	0.49066	0	1
	Married	0.54607	0.49792	0	1
	Schooling Level	10.2903	3.73873	0	18
	Family Income (Normalized)	2.52713	2.92464	-1.012	54.8 3
	Employed	0.10327	0.30434	0	1
	Australian Medicaid	0.09124	0.287981 7	0	1
Obs: 4406					

Figure 2

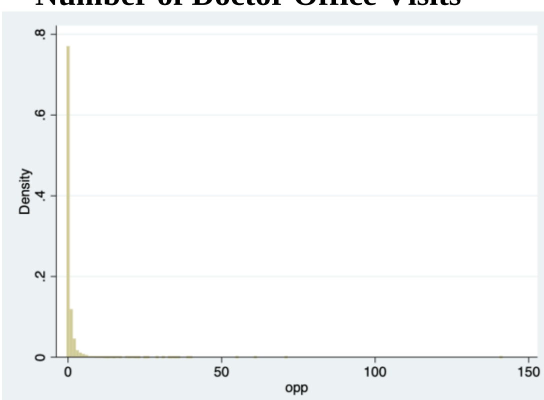
Histograms of Highly Skewed Distributed Count Outcomes (Dependent Variables) of 1987 National Australian Medical Expenditure Survey Data



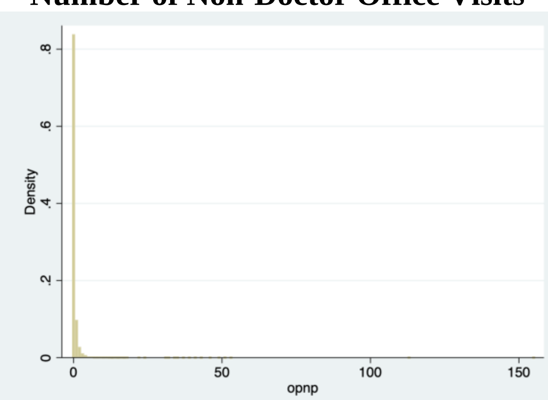
Number of Doctor Office Visits



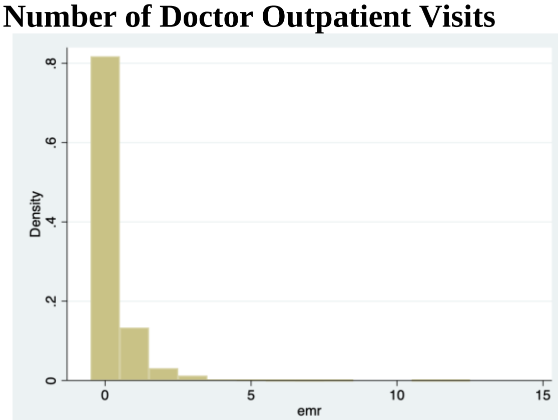
Number of Non-Doctor Office Visits



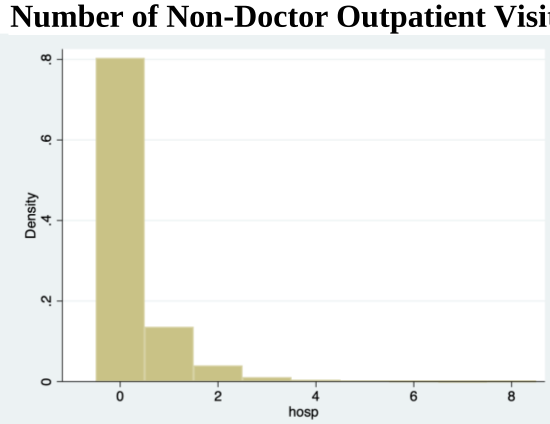
Number of Doctor Outpatient Visits



Number of Non-Doctor Outpatient Visits



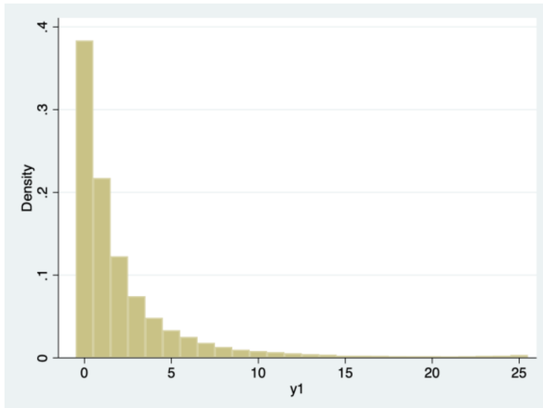
Number of ER Visits



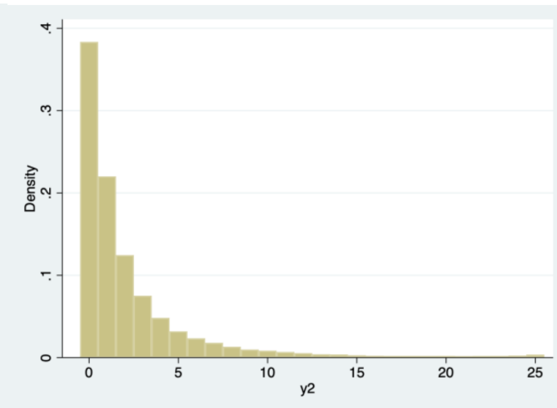
Number of Hospitalization

Figure 3

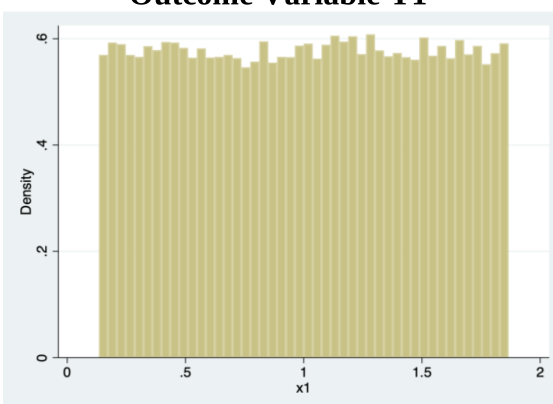
Histograms of Simulated Bivariate Over dispersed CMP Data (Rho=0.25)



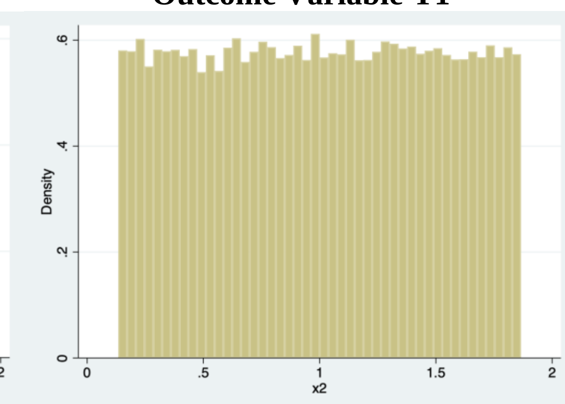
Outcome Variable Y1



Outcome Variable Y1



Policy Variable X



Dependent Variable X_{01}

Rho=0.25

Table 6

MLE of Deep Parameters of Zellner Linear SUR Model Using National Medical Expenditure Survey (NMES) Data

Variables	Outcome Variables: # of Physician Office Visits		Outcome Variables: # of Non-Physician Office Visits	
	OLS (single equation)	Zellner Linear SUR	OLS (single equation)	Zellner Linear SUR
Private Insurance	1.6302 (0.27)	1.6302 (0.27)	0.5959 (0.229)	0.5958 (0.229)
Excellent Health	-1.514 (-4.10)	-1.514 (-4.10)	-0.0458 (-0.15)	-0.0458 (-0.1)
Poor Health	2.152 (6.62)	2.152 (6.63)	-0.2747 (-1.03)	-0.2747 (-1.03)
Number of Chronic Diseases	1.073 (13.79)	1.0737 (13.81)	0.1780 (2.78)	0.1780 (2.79)
Daily Activity Difficulty Index	.6036 (2.20)	0.6036 (2.20)	0.4189 (1.85)	0.4189 (1.86)
Midwest	-0.0194 (-0.08)	-0.0194 (-0.08)	0.8277 (4.01)	0.8277 (4.01)
West	0.7415 (2.60)	0.7415 (2.60)	1.140 (4.86)	1.140 (4.87)
Age	-0.3514 (-2.10)	-0.3514 (-2.11)	-0.4113 (-3.00)	-0.4113 (-3.00)
Black	-0.337 (-1.03)	-0.337 (-1.03)	-0.3107 (-1.16)	-0.3107 (-1.16)
Male	-0.375 (-1.74)	-0.375 (-1.74)	-0.249 (-1.40)	-0.249 (-1.40)
Married	-0.2459 (-1.08)	-0.2459 (-1.0)	0.0142 (0.08)	0.0142 (0.08)
Schooling Level	0.1410 (4.82)	0.1410 (4.83)	0.0873 (3.63)	0.0873 (3.64)
North East	0.6368 (2.31)	0.6368 (2.31)	0.5608 (2.47)	0.5608 (2.48)
Family Income	-0.0197 (-0.55)	-0.0197 (-0.56)	-0.0276 (-0.94)	-0.0276 (-0.95)
Employed	0.27421 (0.83)	0.27421 (0.8)	-0.345 (-1.27)	-0.345 (-1.27)
Australian Medicaid	1.5039 (3.79)	1.5039 (3.80)	0.289 (0.89)	0.289 (0.89)
Constant	3.6866 (2.77)	3.6866 (2.77)	2.657 (2.43)	2.657 (2.43)

T-Statistics in Parenthesis, same in Table 7 and Table 8

Table 7

MLE of Deep Parameters of Bivariate Poisson Model Using National Medical Expenditure Survey (NMES) Data

Variables	Outcome Variables: # of Physician Office Visits		Outcome Variables: # of Non-Phys. Office Visits	
	Single Equation	Bivariate	Single Equation	Bivariate
Private Insurance	0.411 (9.233)	0.424 (9.529)	0.908 (8.636)	0.888 (8.981)
Excellent Health	-0.355 (6.661)	-0.393 (6.672)	-0.297 (10.334)	-0.399 (12.991)
Poor Health	0.307 (8.757)	0.286 (8.981)	-0.384 (10.135)	-0.372 (12.818)
Number of Chronic Diseases	0.242 (34.024)	0.226 (35.28)	0.130 (37.832)	0.173 (41.789)
Daily Activity Difficulty Index	0.027 (9.703)	0.041 (10.482)	0.157 (14.269)	0.270 (17.442)
Midwest	0.098 (9.789)	0.071 (9.672)	0.572 (8.484)	0.253 (13.928)
West	-0.013 (10.168)	-0.001 (10.723)	0.737 (8.386)	0.520 (13.964)
Age	0.143 (9.347)	0.136 (9.575)	1.175 (7.820)	0.916 (13.381)
Black	-0.006 (14.877)	0.007 (15.598)	-0.052 (18.418)	-0.060 (20.951)
Male	-0.147 (7.517)	-0.119 (8.189)	-0.239 (7.864)	-0.205 (8.839)
Married	-0.142 (12.339)	-0.151 (11.914)	-0.678 (15.325)	-0.611 (17.710)
Schooling Level	0.004 (11.338)	0.010 (11.584)	0.224 (15.3180	0.129 (17.911)
North East	0.028 (91.933)	0.022 (87.991)	0.042 (114.784)	0.037 (129.874)
Family Income	-0.003 (68.406)	0.003 (71.126)	0.011 (113.060)	0.020 (118.028)
Employed	0.034 (8.497)	0.016 (9.116)	0.229 (9.698)	0.182 (12.320)
Australian Medicaid	0.387 (6.692)	0.344 (6.870)	-0.059 (7.375)	-0.341 (7.352)
Constant	0.339 (1.893)	0.292 (1.941)	-3.096 (2.325)	-2.678 (2.538)
sigma	0.956 (55.151)	0.900 (33.93)	1.453 (56.08)	1.874 (39.74)
rho	N/A	0.312 (14.436)	N/A	0.312 (14.436)

Table 8

MLE of Deep Parameters of Bivariate Conway-Maxwell-Poisson Model Using National Medical Expenditure Survey (NMES) Data

Variables	Outcome Variables: # of Physician Office Visits		# of Non-Physician Office Visits	
	Single Equation	Bivariate	Single Equation	Bivariate
Private Insurance	0.4535 (8.31)	0.471 (8.85)	0.472 (10.11)	0.424 (9.08)
Excellent Health	-0.368 (-5.15)	-0.378 (-5.31)	-0.180 (-3.24)	-0.152 (-2.00)
Poor Health	0.443 (8.21)	0.503 (8.99)	0.230 (3.17)	0.235 (4.63)
Number of Chronic Diseases	0.225 (15.20)	0.222 (15.00)	0.060 (5.14)	0.054 (4.98)
Daily Activity Difficulty Index	0.0758 (1.52)	0.085 (1.70)	0.169 (4.42)	0.142 (3.93)
Midwest	0.048 (1.05)	0.057 (1.26)	0.446 (11.42)	0.405 (10.57)
West	0.128 (2.51)	0.079 (1.58)	0.338 (8.27)	0.3032 (7.62)
Age	-0.028 (-0.92)	-0.051 (-1.66)	-.176 (-5.76)	-0.161 (-5.53)
Black	-0.125 (-2.05)	-.13473 (-2.27)	-.3069 (-5.92)	-0.286 (-5.41)
Male	-0.0787 (-2.03)	-0.04916 (-1.25)	-0.075 (-2.05)	-0.071 (-2.08)
Married	-0.142 (0.0187)	0.0038 (0.09)	.0194 (0.60)	0.034 (0.98)
Schooling Level	0.036 (7.02)	0.038 (7.48)	.06130 (12.66)	0.059 (10.91)
North East	0.129 (2.57)	.0504 (2.91)	0.468 (10.79)	0.417 (9.48)
Family Income	-0.008 (-1.36)	-0.009 (-1.65)	-0.018 (-3.95)	-0.019 (-3.88)
Employed	0.1127 (1.7)	0.11269 (1.70)	-0.0386 (-0.63)	-0.025 (-0.51)
Australian Medicaid	0.380 (4.76)	0.441 (5.79)	0.373 (5.09)	0.3551 (4.70)
Constant	0.355 (1.43)	0.4756 (1.94)	-1.3460 (-5.45)	-1.362 (-5.55)
omega	0.0148 (0.91)	0.0109 (0.66)	-1.408 (-33.06)	-1.5001 (-32.53)
rho	N/A	0.396163 (16.50)	N/A	0.396163 (16.50)

Table 9

Comparison 1 (Zellner's SUR versus CMP SUR): Average Incremental Effect (AIE) Estimates using National Medical Expenditure Survey (NMES) Data

Average Incremental Effects: Private Insurance Status on Two Correlated Health Care Utilization Counts								
	Zellner's Linear Seemingly Unrelated Regression (SUR) Model				**Count-Outcome SUR Model (Conway-Maxwell-Poisson case)**			
	AIE	**S.E.**	**T-Stat**	**P-Value**	**AIE**	**S.E.**	**T-Stat**	**P-Value**
Count of Physician Office Visits (Y_1)	1.6302	0.2784	5.8536	0.0000	2.5637	0.7630	1.0604	0.2889
Count of Non-Physician Office Visits (Y_2)	0.5958	0.2288	2.6034	0.0092	22.358	7.6115	2.9374	0.0033

Standard errors are asymptotic standard errors as given in Terza (2017a); T-Statistics and P-Values are derived from these asymptotic standard errors.

Table 10

Comparison 2-A (Single-Equation Poisson versus Poisson SUR): Average Incremental Effect (AIE) Estimates using National Medical Expenditure Survey (NMES) Data

Average Incremental Effects: Private Insurance Status on Two Correlated Health Care Utilization Counts								
	Single-Equation Estimation (Poisson case)				**Count-Outcome SUR Model (Poisson case)**			
	AIE	**S.E.**	**T-Stat**	**P-Value**	**AIE**	**S.E.**	**T-Stat**	**P-Value**
Count of Physician Office Visits (Y_1)	2.5290	0.7340	3.4400	0.0006	1.8830	0.5036	3.7400	0.0002
Count of Non-Physician Office Visits (Y_2)	2.9808	1.5900	1.8740	0.0600	4.0088	0.4783	3.7470	0.0002

Standard errors are asymptotic standard errors as given in Terza (2017a); T-Statistics and P-Values are derived from these asymptotic standard errors.

Table 11

Comparison 2-B (Single-Equation CMP versus CMP SUR): Average Incremental Effect (AIE) Estimates using National Medical Expenditure Survey (NMES) Data

Average Incremental Effects: Private Insurance Status on Two Correlated Health Care Utilization Counts								
	Single-Equation Estimation (Conway-Maxwell-Poisson case)				Count-Outcome SUR Model (Conway-Maxwell-Poisson case)			
	AIE	S.E.	T-Stat	P-Value	AIE	S.E.	T-Stat	P-Value
Count of Physician Office Visits (Y_1)	2.6444	0.8597	3.0759	0.0020	2.5637	0.7630	1.0604	0.288
Count of Non-Physician Office Visits (Y_2)	21.652	7.2507	2.9862	0.0028	22.358	7.6115	2.9374	0.003

Standard errors are asymptotic standard errors as given in Terza (2017a); T-Statistics and P-Values are derived from these asymptotic standard errors.

Table 12

Comparison 3 (CMP SUR versus Poisson SUR): Average Incremental Effect (AIE) Estimates using National Medical Expenditure Survey (NMES) Data

Average Incremental Effects: Private Insurance Status on Two Correlated Health Care Utilization Counts								
	Count-Outcome SUR Model (Conway-Maxwell-Poisson case)				Count-Outcome SUR Model (Poisson case)			
	AIE	S.E.	T-Stat	P-Value	AIE	S.E.	T-Stat	P-Value
Count of Physician Office Visits (Y_1)	2.5637	0.7630	1.0604	0.2889	1.8830	0.5036	3.7400	0.0002
Count of Non-Physician Office Visits (Y_2)	22.3582	7.6115	2.9374	0.0033	4.0088	0.4783	3.7470	0.0002

Standard errors are asymptotic standard errors as given in Terza (2017a); T-Statistics and P-Values are derived from these asymptotic standard errors.

Table 13

MLE of Deep Parameters of Zellner's Linear Seemingly Unrelated Regression (SUR) Model with Simulated Over-Dispersed Data

Two Outcomes	**Y1**	**Y2**	**Y1**	**Y2**
	Zellner's Linear Seemingly Unrelated Regression (SUR) Model		**TRUE**	
beta1	2.192 (0.0700)	2.224 (0.0707)	1	1
beta2	-2.239 (0.0706)	-2.268 (0.0713)	-1	-1
constant	2.658 (0.6953)	2.649 (0.7027)	0	0
rho12	N.A.		0.75	
omega	N.A.	N.A.	-0.1	-0.1

N = 50,000, standard errors in parentheses

Table 14

MLE of Deep Parameters of Bivariate Conway-Maxwell Poisson Model with Simulated Over-Dispersed Data

Two Outcomes	**Y1**	**Y2**	**Y1**	**Y2**
	Count-Outcome SUR Model (Conway-Maxwell-Poisson case)		***Single-Equation* Estimation (Conway-Maxwell-Poisson case)**	
beta1	0.9963 (0.0125)	1.00225 (0.0127)	0.9963 (0.0126)	1.0015 (0.0129)
beta2	-1.00173 (0.0126)	-0.9929 (0.0127)	-1.0019 (0.0126)	-0.9986 (0.0129)
constant	.01298 (0.0193)	.00372 (0.0192)	0.0111 (0.01941)	0.0088 (0.0194)
rho12	0.743392 (.00492)		N.A.	
omega	-0.08856 (0.00674)	-0.0872 (0.0066)	-0.0935 (0.0068)	-0.0896 (0.0068)

N = 50,000, standard errors in parentheses

Table 15

MLE of Deep Parameters of Zellner's SUR Versus Poisson SUR Versus CMP SUR using Simulated Over-Dispersed Data

Two Outcomes	Y1	Y2	Y1	Y2	Y1	Y2
	Zellner's Linear SUR Model		Count-Outcome SUR (Poisson case)		Count-Outcome SUR (Conway-Maxwell-Poisson case)	
beta1	2.192 (0.0700)	2.224 (0.0707)	1.051 (0.0288)	1.0115 (.02985)	0.9963 (0.0125)	1.00225 (0.0127)
beta2	-2.239 (0.0706)	-2.268 (0.0713)	-1.0375 (.02907)	-1.0565 (.02988)	-1.00173 (0.01264)	-0.9929 (0.0127)
constant	2.658 (0.6953)	2.649 (0.7027)	.04875 (.0430)	.1077 (.0437)	.01298 (0.0193)	.00372 (0.01917)
rho12	N.A.		0.477		0.743392 (.00492)	
sigma squared	N.A.	N.A.	1.039 (0.0138)	1.055 (0.0147)	N.A.	N.A.
sigma12			0.5002 (0.0146)		N.A.	
omega	N.A.	N.A.	N.A.	N.A.	-0.08856 (0.00674)	-0.0872 (0.0066)

N = 50,000, standard errors in parentheses

Table 16

Comparison 1 (Zellner's SUR versus CMP SUR): Average Incremental Effect (AIE) Estimates Using Simulated Over-Dispersed Data

			Count-Outcome SUR (Conway-Maxwell-Poisson case)		**Zellner's Linear SUR**	
Design	**rho12**	**True AIE**	**Average AIE()**	**AAPB AIE()**	**Average AIE()**	**AAPB AIE()**
1 (Over-Dispersed Correlated Counts) Omega = -0.1	0.75	4.765	4.502	10.48%	2.185	53.06%
	0.5	4.767	4.865	25.48%	2.191	52.89%
	0.25	4.767	4.502	19.00%	2.213	52.57%
	0	4.767	4.7079	26.87%	2.209	52.64%

100 replications with 10,000 observations for each replication

Table 17

Comparison 2 (CMP SUR versus Single-Equation CMP): Average Incremental Effect (AIE) Estimates Using Simulated Over-Dispersed Data

			Count-Outcome SUR (Conway-Maxwell-Poisson case)		***Single-Equation* Model (Conway-Maxwell-Poisson case)**	
Design	**rho12**	**True AIE**	**Average AIE()**	**AAPB AIE()**	**Average AIE()**	**AAPB AIE()**
1 (Over-Dispersed Correlated Counts) Omega = -0.1	0.75	4.765	4.502	10.48%	4.583	11.31%
	0.5	4.767	4.865	25.48%	4.845	24.61%
	0.25	4.767	4.502	19.00%	4.419	21.40%
	0	4.767	4.708	26.87%	4.710	26.46%

100 replications with 10,000 observations for each replication

Table 18

Comparison 3 (CMP SUR versus Poisson SUR): Average Incremental Effect (AIE) Estimates Using Simulated Over-Dispersed Data

			Count-Outcome SUR (Conway-Maxwell-Poisson case)		Bivariate Poisson	
Design	**rho12**	**True AIE**	**Average AIE()**	**AAPB AIE()**	**Average AIE()**	**AAPB AIE()**
1 (Over-Dispersed Correlated Counts) Omega = -0.1	0.75	4.765	4.502	10.48%	4.035	34.19%
	0.5	4.767	4.865	25.48%	4.265	36.93%
	0.25	4.767	4.502	19.00%	4.147	35.53%
	0	4.767	4.7079	26.87%	4.216	36.24%

100 replications with 10,000 observations for each replication

Table 19

Simulation Parameter Designs for Bivariate Poisson Models

Simulation Variables/Parameters	Value
X and X_{01}	Uniform [0.13, 1.87]
X_{02}	Equal to 1
$\beta_1 = \beta_2$	Equal to 1
$\beta_{01} = \beta_{02}$	Equal to $(-1,0)'$
$\sigma_1 = \sigma_2$	Equal to 1
ρ_{12}	Element of $\in \{0.25, 0.5, 0.75, 0.9\}$.

Table 20

Simulation Parameter Designs for Bivariate Over-Dispersed CMP Models

	Over-Disperse Data
Simulation Variables/Parameters	**Values**
X and X_{01}	Uniform [0.13, 1.87]
X_{02}	Equal to 1
$\beta_1 = \beta_2$	Equal to 1
$\beta_{01} = \beta_{02}$	Equal to $(-1,0)'$
$\sigma_1 = \sigma_2$	Equal to 1
$\omega_1 = \omega_2$	Equal to -1
ρ_{12}	Element of $\{0, 0.25, 0.5, 0.75\}$.

Table 21

Summary Statistics of Simulated Bivariate Over-Dispersed CMP Data

Variable	Mean	Std. Dev.	Min	Max
Y1	2.36714	3.931251	0	25
Y2	2.32628	3.869724	0	25
X	1.00061	0.5002346	0.134089	1.866013
$\mathbf{X_{01}}$	1.00059	0.4996162	0.1339784	1.866003

Based on 50,000 observations (the summary statistics are similar for other sample size simulations such as 10,000, 500,000 observations); Rho=0.25.

References

Affleck, D. L. R., & Diéguez-Aranda, U. (2016). Additive Nonlinear Biomass Equations: A Likelihood-Based Approach. *Forest Science*, *62*(2), 129–140. https://doi.org/10.5849/forsci.15-126

Afghari, A. P., Washington, S., Haque, Md. M., & Li, Z. (2018). A Comprehensive Joint Econometric Model Of Motor Vehicle Crashes Arising From Multiple Sources Of Risk. *Analytic Methods in Accident Research*, *18*, 1–14. https://doi.org/10.1016/j.amar.2018.03.002

Aitchison, J., & Ho, C. H. (1989). The Multivariate Poisson-Log Normal Distribution. *Biometrika*, *76*(4), 643–653. https://doi.org/10.1093/biomet/76.4.643

Alfò, M., Maruotti, A., & Trovato, G. (2011). A Finite Mixture Model For Multivariate Counts Under Endogenous Selectivity. *Statistics and Computing*, *21*(2), 185–202. https://doi.org/10.1007/s11222-009-9159-2

Amemiya, T. (1974). Bivariate Probit Analysis: Minimum Chi-Square Methods. *Journal of the American Statistical Association*, *69*(348), 940–944. https://doi.org/10.1080/01621459.1974.10480232

Amemiya, T. (1977). The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model. *Econometrica*, *45*(4), 955. https://doi.org/10.2307/1912684

Amemiya, T. (1983). Chapter 6 Non-Linear Regression Models. In *Handbook of Econometrics* (Vol. 1, pp. 333–389). Elsevier. https://doi.org/10.1016/S1573-4412(83)01010-7

Amiri, R., & Behnia, F. (2017). An Efficient Weighted Least Squares Estimator for Elliptic Localization in Distributed MIMO Radars. *IEEE Signal Processing Letters*, *24*(6), 902–906. https://doi.org/10.1109/LSP.2017.2697500

Andreassen, C. M., & Jensen, J. L. (2018). On The Three-Dimensional Negative Binomial Distribution. *Communications in Statistics - Theory and Methods*, *47*(10), 2314–2326. https://doi.org/10.1080/03610926.2013.864766

Ashford, J. R., & Sowden, R. R. (n.d.). *Multi-Variate Probit Analysis*. 13.

Atella, V., & Deb, P. (2008). Are Primary Care Physicians, Public And Private Sector Specialists Substitutes Or Complements? Evidence From A Simultaneous Equations Model For Count Data. *Journal of Health Economics*, *27*(3), 770–785. https://doi.org/10.1016/j.jhealeco.2007.10.006

Austin, P. C., Naylor, C. D., & Tu, J. V. (2001). A Comparison Of A Bayesian Vs. A Frequentist Method For Profiling Hospital Performance. *Journal of Evaluation in Clinical Practice*, *7*(1), 35–45. https://doi.org/10.1046/j.1365-2753.2001.00261.x

Babuska, I., Henshaw, W. D., Oliger, J. E., Flaherty, J. E., Hopcroft, J. E., & Tezduyar, T. (Eds.). (1995). *Modeling, Mesh Generation, and Adaptive Numerical Methods for Partial Differential Equations*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4612-4248-2

Browne. (2006). A Comparison Of Bayesian And Likelihood-Based Methods For Fitting Multilevel Models. *Bayesian Analysis (2006)*, *1, Number 3, pp. 473–514*.

Burda, M., Harding, M., & Hausman, J. (2012). A Poisson Mixture Model Of Discrete Choice. *Journal of Econometrics*, *166*(2), 184–203. https://doi.org/10.1016/j.jeconom.2011.09.001

Cameron, A. C., Li, T., Trivedi, P. K., & Zimmer, D. M. (2004). Modelling The Differences In Counted Outcomes Using Bivariate Copula Models With Application To Mismeasured Counts. *The Econometrics Journal*, *7*(2), 566–584. https://doi.org/10.1111/j.1368-423X.2004.00144.x

Cappellari, L., & Jenkins, S. P. (2003). Multivariate Probit Regression using Simulated Maximum Likelihood. *The Stata Journal: Promoting Communications on Statistics and Stata*, *3*(3), 278–294. https://doi.org/10.1177/1536867X0300300305

Chib, S., & Winkelmann, R. (2001). Markov Chain Monte Carlo Analysis of Correlated Count Data. *Journal of Business & Economic Statistics*, *19*(4), 428–435. https://doi.org/10.1198/07350010152596673

Chua, C. L., & Tsiaplias, S. (2019). Information Flows And Stock Market Volatility. *Journal of Applied Econometrics*, *34*(1), 129–148. https://doi.org/10.1002/jae.2649

Cragg, J. G. (1971). Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica*, *39*(5), 829. https://doi.org/10.2307/1909582

Deb, P., & Trivedi, P. (1997). Demand For Medical Care By The Elderly: A Finite Mixture Approach. *Journal of Applied Econometrics*, *Vol. 12*(Issue 3), p313-336. https://doi.org/10.1002/(SICI)1099-1255(199705)12:3<313::AID-JAE440>3.0.CO;2-G

Delgado, M. A. (1992). Semiparametric Generalized Least Squares in the Multivariate Nonlinear Regression Model. *Econometric Theory*, *8*(2), 203–222. https://doi.org/10.1017/S0266466600012767

Dong, X., Chintagunta, P. K., & Manchanda, P. (2011). A New Multivariate Count Data Model To Study Multi-Category Physician Prescription Behavior. *Quantitative Marketing and Economics*, *9*(3), 301–337. https://doi.org/10.1007/s11129-011-9102-7

Dufour, J.-M., & Khalaf, L. (2002). Exact Tests For Contemporaneous Correlation Of Disturbances In Seemingly Unrelated Regressions. *Journal of Econometrics*, *106*(1), 143–170. https://doi.org/10.1016/S0304-4076(01)00093-8

Durham, G. B., & Gallant, A. R. (2002). Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes. *Journal of Business & Economic Statistics*, *20*(3), 297–338. https://doi.org/10.1198/073500102288618397

Egan, K., & Herriges, J. (2006). Multivariate Count Data Regression Models With Individual Panel Data From An On-Site Sample. *Journal of Environmental Economics and Management*, *52*(2), 567–581. https://doi.org/10.1016/j.jeem.2006.03.002

Eriksson, J., & Viberg, M. (2004). Asymptotic Properties of Nonlinear Weighted Least Squares in Radar Array Processing. *IEEE Transactions on Signal Processing*, *52*(11), 3083–3095. https://doi.org/10.1109/TSP.2004.836459

Fahrmeir, L., & Osuna, L. (2003). *Structured count data regression*. https://doi.org/10.5282/UBM/EPUB.1712

Famoye, F. (2010a). A New Bivariate Generalized Poisson Distribution. *Statistica Neerlandica*, *64*(1), 112–124. https://doi.org/10.1111/j.1467-9574.2009.00446.x

Famoye, F. (2010b). On The Bivariate Negative Binomial Regression Model. *Journal of Applied Statistics*, *37*(6), 969–981. https://doi.org/10.1080/02664760902984618

Famoye, F. (2015). A Multivariate Generalized Poisson Regression Model. *Communications in Statistics - Theory and Methods*, *44*(3), 497–511. https://doi.org/10.1080/03610926.2012.743565

Farbmacher, H. (2011). Estimation of Hurdle Models for Overdispersed Count Data. *The Stata Journal: Promoting Communications on Statistics and Stata*, *11*(1), 82–94. https://doi.org/10.1177/1536867X1101100105

Faurot. (1986). A Computer Program For Seemingly Unrelated Nonlinear Regressions. *Econometrica (Pre-1986)*.

Fortunati, S., Farina, A., Gini, F., Graziano, A., Greco, M. S., & Giompapa, S. (2011). Least Squares Estimation and Cramér–Rao Type Lower Bounds for Relative Sensor Registration Process. *IEEE Transactions on Signal Processing*, *59*(3), 1075–1087. https://doi.org/10.1109/TSP.2010.2097258

Foschi, P., Belsley, D. A., & Kontoghiorghes, E. J. (2003). A Comparative Study Of Algorithms For Solving Seemingly Unrelated Regressions Models. *Computational Statistics & Data Analysis*, *44*(1–2), 3–35. https://doi.org/10.1016/S0167-9473(03)00028-8

Fu, L., Sharma, R. P., Wang, G., & Tang, S. (2017). Modelling A System Of Nonlinear Additive Crown Width Models Applying Seemingly Unrelated Regression For Prince Rupprecht Larch In Northern China. *Forest Ecology and Management*, *386*, 71–80. https://doi.org/10.1016/j.foreco.2016.11.038

Gallant, A. R. (1975). Seemingly Unrelated Nonlinear Regressions. *Journal of Econometrics*, *3*(1), 35–50. https://doi.org/10.1016/0304-4076(75)90064-0

Gamerman, Victoria Alexandria, "Statistical Methods for Time-Conditional Survival Probability and Equally Spaced Count Data" (2016). *Publicly Accessible Penn Dissertations*. 1729. https://repository.upenn.edu/edissertations/1729

Kim SH., Bao Y., Horan E., Kim M., Cohen A.S. (2015) Gauss–Hermite Quadrature in Marginal Maximum Likelihood Estimation of Item Parameters. In: van der Ark L., Bolt D., Wang WC., Douglas J., Chow SM. (eds) Quantitative Psychology Research. Springer Proceedings in Mathematics & Statistics, vol 140. Springer, Cham. https://doi.org/10.1007/978-3-319-19977-1_4

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression Analyses Of Counts And Rates: Poisson, Overdispersed Poisson, And Negative Binomial Models. *Psychological Bulletin*, *118*(3), 392–404. https://doi.org/10.1037/0033-2909.118.3.392

Geweke, J., & Tanizaki, H. (2003). Note on the Sampling Distribution for the Metropolis-Hastings Algorithm. *Communications in Statistics - Theory and Methods*, *32*(4), 775–789. https://doi.org/10.1081/STA-120018828

Geweke, J., Gowrisankaran, G., & Town, R. J. (2003). Bayesian Inference for Hospital Quality in a Selection Model. Econometrica, 71(4), 1215–1238. https://doi.org/10.1111/1468-0262.00444

Gourieroux, C., Monfort, A., & Trognon, A. (1984a). Pseudo Maximum Likelihood Methods: Applications to Poisson Models. *Econometrica*, *52*(3), 701. https://doi.org/10.2307/1913472

Gourieroux, C., Monfort, A., & Trognon, A. (1984b). Pseudo Maximum Likelihood Methods: Theory. *Econometrica*, *52*(3), 681. https://doi.org/10.2307/1913471

Greene, W. H. (2003). *Econometric analysis* (5th ed). Upper Saddle River, N.J: Prentice Hall.

Gueorguieva, R. (2001). A Multivariate Generalized Linear Mixed Model For Joint Modelling Of Clustered Outcomes In The Exponential Family. *Statistical Modelling*, *1*(3), 177–193. https://doi.org/10.1191/147108201128159

Guikema, S. D., & Goffelt, J. P. (2008). A Flexible Count Data Regression Model for Risk Analysis. *Risk Analysis*, *28*(1), 213–223. https://doi.org/10.1111/j.1539-6924.2008.01014.x

Gurmu, S., & Elder, J. (2000). Generalized Bivariate Count Data Regression Models. *Economics Letters*, *68*(1), 31–36. https://doi.org/10.1016/S0165-1765(00)00225-1

Hao, Z. (2018a). *Casual Analysis Using Two-Part Models: A General Framework For Specification, Estimation and Inference*. IUPUI ScholarWorks. https://doi.org/10.7912/C2F35D

Hausman, J., Hall, B. H., & Griliches, Z. (1984). Econometric Models for Count Data with an Application to the Patents-R&D Relationship. *Econometrica*, *52*(4), 909. https://doi.org/10.2307/1911191

Heckman, J. J. (1978). Dummy Endogenous Variables in a Simultaneous Equation System. *Econometrica*, *46*(4), 931. https://doi.org/10.2307/1909757

Heckman, J. J., & Vytlacil, E. (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation1. *Econometrica*, *73*(3), 669–738. https://doi.org/10.1111/j.1468-0262.2005.00594.x

Heiss, F., & Winschel, V. (2008). Likelihood Approximation By Numerical Integration On Sparse Grids. *Journal of Econometrics*, *144*(1), 62–80. https://doi.org/10.1016/j.jeconom.2007.12.004

Herriges, J. A., Phaneuf, D. J., & Tobias, J. L. (2008). Estimating Demand Systems When Outcomes Are Correlated Counts. *Journal of Econometrics*, *147*(2), 282–298. https://doi.org/10.1016/j.jeconom.2008.09.026

Hirano, K., & Porter, J. R. (2003). Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support. *Econometrica*, *71*(5), 1307–1338. https://doi.org/10.1111/1468-0262.00451

Hofer, V., & Leitner, J. (2012). A Bivariate Sarmanov Regression Model For Count Data With Generalised Poisson Marginals. *Journal of Applied Statistics*, *39*(12), 2599–2617. https://doi.org/10.1080/02664763.2012.724661

Holoborodko, P. (2012, May). *Abscissas and Weights of Classical Gaussian Quadrature Rules*.

Holtmann, J., Koch, T., Lochner, K., & Eid, M. (2016). A Comparison of ML, WLSMV, and Bayesian Methods for Multilevel Structural Equation Models in Small Samples: A Simulation Study. *Multivariate Behavioral Research*, *51*(5), 661–680. https://doi.org/10.1080/00273171.2016.1208074

Chuck Huber, 2017. "Introduction to Bayesian Analysis Using Stata," Canadian Stata Users' Group Meetings 2017 14, Stata Users Group.

Jensen, M. J. (1995). A Monte Carlo Study On Two Methods Of Calculating The MLE's Covariance Matrix In A Seemingly Unrelated Nonlinear Regression. *Econometric Reviews*, *14*(3), 315–330. https://doi.org/10.1080/07474939508800323

Jones, O., Maillardet, R., & Robinson, A. (2014). *Introduction To Scientific Programming And Simulation Using R* (Second edition). Boca Raton, FL: CRC Press, Taylor & Francis Group.

Jorgenson, Laffont. (1974). Efficient Estimation of Nonlinear Simultaneous Equations with Additive Disturbances. *Annals of Economic and Social Measurement, Volume 3, number 4*.

Keshavarzzadeh, V., Kirby, R. M., & Narayan, A. (2018). Numerical Integration in Multiple Dimensions with Designed Quadrature. *SIAM Journal on Scientific Computing*, *40*(4), A2033–A2061. https://doi.org/10.1137/17M1137875

Kim, S.-H., Bao, Y., Horan, E., Kim, M., & Cohen, A. S. (2015). Gauss–Hermite Quadrature in Marginal Maximum Likelihood Estimation of Item Parameters. In L. A. van der Ark, D. M. Bolt, W.-C. Wang, J. A. Douglas, & S.-M. Chow (Eds.), *Quantitative Psychology Research* (pp. 43–58). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-19977-1_4

King, G. (1989). A Seemingly Unrelated Poisson Regression Model. *Sociological Methods & Research*, *17*(3), 235–255. https://doi.org/10.1177/0049124189017003001

Leung, E., Michelson, S., Villarubia, C., Perkins, L. A., & Eglen, R. M. (1992). Analysis Of Concentration-Response Relationships By Seemingly Unrelated Nonlinear Regression (SUNR) Technique. *Journal of Pharmacological and Toxicological Methods*, *28*(4), 209–216. https://doi.org/10.1016/1056-8719(92)90006-M

Leung, S. F., & Yu, S. (1996). On The Choice Between Sample Selection And Two-Part Models. *Journal of Econometrics*, *72*(1–2), 197–229. https://doi.org/10.1016/0304-4076(94)01720-4

Lord, D., Guikema, S. D., & Geedipally, S. R. (2008). Application Of The Conway–Maxwell–Poisson Generalized Linear Model For Analyzing Motor Vehicle Crashes. *Accident Analysis & Prevention*, *40*(3), 1123–1134. https://doi.org/10.1016/j.aap.2007.12.003

Lukacsy, K. (n.d.). Generating Random Samples From User-Defined Distributions. *The Stata Journal (2011)*, *(2011) 11, Number 2*, 299–304.

Ma, J., Kockelman, K. M., & Damien, P. (2008). A Multivariate Poisson-Lognormal Regression Model For Prediction Of Crash Counts By Severity, Using Bayesian Methods. *Accident Analysis & Prevention*, *40*(3), 964–975. https://doi.org/10.1016/j.aap.2007.11.002

Ma, W., & Lei, X. (2015). Nonlinear Simultaneous Equations for Individual-Tree Diameter Growth and Mortality Model of Natural Mongolian Oak Forests in Northeast China. *Forests*, *6*(12), 2261–2280. https://doi.org/10.3390/f6062261

Manalew, W. S. (2020, February). *A Switching Regressions Framework For Models With Count-Valued Omni-Dispersed Outcomes: Specification, Estimation And Causal Inference*. IUPUI ScholarWorks

Mander, A., & Bowden, J. (2012). Numerical Integration With An Application In Sample Size Reestimation. *United Kingdom Stata Users' Group Meetings 2012*. Retrieved from https://ideas.repec.org/p/boc/usug12/03.html

Marshall, S. L. (2003). Generalized Least-Squares Parameter Estimation From Multiequation Implicit Models. *AIChE Journal*, *49*(10), 2577–2594. https://doi.org/10.1002/aic.690491012

Maus, M., Cotlet, M., Hofkens, J., Gensch, T., De Schryver, F. C., Schaffer, J., & Seidel, C. A. M. (2001). An Experimental Comparison of the Maximum Likelihood Estimation and Nonlinear Least-Squares Fluorescence Lifetime Analysis of Single Molecules. *Analytical Chemistry*, *73*(9), 2078–2086. https://doi.org/10.1021/ac000877g

McElroy, M. B., & Burmeister, E. (1988). Arbitrage Pricing Theory as a Restricted Nonlinear Multivariate Regression Model Iterated Nonlinear Seemingly Unrelated Regression Estimates. *Journal of Business & Economic Statistics*, *6*(1), 29–42. https://doi.org/10.1080/07350015.1988.10509634

Mehta, G. S. (2014). *Analyzing Crash Frequency And Severity Data Using Novel Techniques*. University of Alabama, Institutional Repository

Mihaylova, B., Briggs, A., O'Hagan, A., & Thompson, S. G. (2011). Review of Statistical Methods For Analyzing Healthcare Resources And Costs. Health Economics, 20(8), 897–916. https://doi.org/10.1002/hec.1653

Mullahy, J. (1997). Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior. *Review of Economics and Statistics*, *79*(4), 586–593. https://doi.org/10.1162/003465397557169

Nguyen, H. B. (2010). Estimating A Fractional Response Model With A Count Endogenous Regressor And An Application To Female Labor Supply. In W. Greene & R. Carter Hill (Eds.), *Maximum Simulated Likelihood Methods and Applications* (pp. 253–298). Emerald Group Publishing Limited. https://doi.org/10.1108/S0731-9053(2010)0000026012

Oosterlee, K. (2009). *Monte Carlo Integration*. Hitotsubashi University, Summer School in Computational Finance

Pagan, A. (1984). Econometric Issues in the Analysis of Regressions with Generated Regressors. *International Economic Review*, *25*(1), 221. https://doi.org/10.2307/2648877

Parresol, B. R. (2001). Additivity Of Nonlinear Biomass Equations. *Canadian Journal of Forest Research*, *31*(5), 865–878. https://doi.org/10.1139/x00-202

Rickert, J. (2016). *Simulating from the Bivariate Normal Distribution in R*. Revolutionary Analysis.

Roodman, D. (2011). Fitting Fully Observed Recursive Mixed-process Models with CMP. *The Stata Journal: Promoting Communications on Statistics and Stata*, *11*(2), 159–206. https://doi.org/10.1177/1536867X1101100202

Rosen, O., Jiang, W., King, G., & Tanner, M. A. (2001). Bayesian and Frequentist Inference for Ecological Inference: The RxC Case. *Statistica Neerlandica*, *55*(2), 134–156. https://doi.org/10.1111/1467-9574.00162

Sag, T. W., & Szekeres, G. (1964). Numerical evaluation of high-dimensional integrals. *Mathematics of Computation*, *18*(86), 245–245. https://doi.org/10.1090/S0025-5718-1964-0165689-X

Sarada, J., & Nagaraja, K. V. (2011). Generalized Gaussian Quadrature Rules Over Two-Dimensional Regions With Linear Sides. *Applied Mathematics and Computation*, *217*(12), 5612–5621. https://doi.org/10.1016/j.amc.2010.12.039

Sellers, K. F., & Morris, D. S. (2017). Underdispersion Models: Models That Are “Under the Radar.” *Communications in Statistics - Theory and Methods*, *46*(24), 12075–12086. https://doi.org/10.1080/03610926.2017.1291976

Sellers, K. F., Morris, D. S., & Balakrishnan, N. (2016). Bivariate Conway–Maxwell–Poisson Distribution: Formulation, Properties, And Inference. *Journal of Multivariate Analysis*, *150*, 152–168. https://doi.org/10.1016/j.jmva.2016.04.007

Sellers, K. F., & Shmueli, G. (2010). A Flexible Regression Model For Count Data. *The Annals of Applied Statistics*, *4*(2), 943–961. https://doi.org/10.1214/09-AOAS306

Serhiyenko, V., Ravishanker, N., & Venkatesan, R. (2018). Multi-Stage Multivariate Modeling Of Temporal Patterns In Prescription Counts For Competing Drugs In A Therapeutic Category. *Applied Stochastic Models in Business and Industry*, *34*(1), 61–78. https://doi.org/10.1002/asmb.2232

Sfeir, G., Abou-Zeid, M., & Kaysi, I. (2020). Multivariate Count Data Models For Adoption Of New Transport Modes In An Organization-Based Context. *Transport Policy*, S0967070X19305128. https://doi.org/10.1016/j.tranpol.2020.03.014

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. https://doi.org/10.1111/j.1467-9876.2005.00474.x

Skrainka, B. S., & Judd, K. L. (2011). High Performance Quadrature Rules: How Numerical Integration Affects a Popular Model of Product Differentiation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1870703

StataCorp LLC. (2019). *Mata Reference Manual Release 16*. StataCorp LLC.

Stegmueller, D. (2013). How Many Countries for Multilevel Modeling? A Comparison of Frequentist and Bayesian Approaches. *American Journal of Political Science*, *57*(3), 748–761. https://doi.org/10.1111/ajps.12001

Terza, J. V. (1985). Reduced-Form Trinomial Probit: A Quantal Response Model Without A Priori Restrictions. *Journal of Business & Economic Statistics*, *3*(1), 54–59. https://doi.org/10.1080/07350015.1985.10509426

Terza, J. V. (1998). Estimating Count Data Models With Endogenous Switching: Sample Selection and Endogenous Treatment Effects. *Journal of Econometrics*, *84*(1), 129–154. https://doi.org/10.1016/S0304-4076(97)00082-1

Terza, J. V. (2009). Parametric Nonlinear Regression with Endogenous Switching. *Econometric Reviews*, *28*(6), 555–580. https://doi.org/10.1080/07474930802473751

Terza, J. V. (2016a). Inference Using Sample Means of Parametric Nonlinear Data Transformations. *Health Services Research*, *51*(3), 1109–1113. https://doi.org/10.1111/1475-6773.12494

Terza, J. V. (2016b). Simpler Standard Errors for Two-stage Optimization Estimators. *The Stata Journal: Promoting Communications on Statistics and Stata*, *16*(2), 368–385. https://doi.org/10.1177/1536867X1601600206

Terza, J. V. (2017). Causal Effect Estimation and Inference Using Stata. *The Stata Journal: Promoting Communications on Statistics and Stata*, *17*(4), 939–961. https://doi.org/10.1177/1536867X1801700410

Terza, J. V. (2018). Two-Stage Residual Inclusion Estimation in Health Services Research and Health Economics. *Health Services Research*, *53*(3), 1890–1899. https://doi.org/10.1111/1475-6773.12714

Terza, J. V. (2019a). *Mata Implementation Of Gauss-Legendre Quadrature In The M-Estimation Context: Correcting For Sample-Selection Bias In A Generic Nonlinear Setting*. 2019 Stata Conference 31, Stata Users Group.

Terza, J. V. (2020). Regression-Based Causal Analysis from the Potential Outcomes Perspective. *Journal of Econometric Methods*, *0*(0). https://doi.org/10.1515/jem-2018-0030

Terza, J. V. (2019). *Mata Implementation Of Gauss-Legendre Quadrature In The M-Estimation Context: Correcting For Sample Selection Bias In A Generic Nonlinear Setting*.

Terza, J. V. (Working paper). *Multivariate Probit Analysis via Seemingly Unrelated Nonlinear Regressions*.

Terza, J. V., Basu, A., & Rathouz, P. J. (2008). Two-Stage Residual Inclusion Estimation: Addressing Endogeneity In Health Econometric Modeling. *Journal of Health Economics*, *27*(3), 531–543. https://doi.org/10.1016/j.jhealeco.2007.09.009

Terza, J. V., & Welland, U. (1991). A Comparison Of Bivariate Normal Algorithms. *Journal of Statistical Computation and Simulation*, *39*(1–2), 115–127. https://doi.org/10.1080/00949659108811343

Terza, J. V., & Zhang, A. (2020, July 30). Two-Dimensional Gauss-Legendre Quadrature: Seemingly Unrelated Dispersion-Flexible Count Regressions. 2020 Stata Conference. Retrieved from https://www.stata.com/meeting/us20/slides/us20_Terza.pdf

Thomopoulos, N. T. (2017). Bivariate Lognormal. In N. T. Thomopoulos, *Statistical Distributions* (pp. 165–169). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-65112-5_20

Train, K. (2000). Halton Sequences for Mixed Logit. *UC Berkeley: Department of Economics*. Retrieved from https://escholarship.org/uc/item/6zs694tp

Train, K. (2003). Discrete Choice Methods with Simulation by Kenneth E. Train.

Vasilyev, D., Busygin, V., & Busygin, S. (2017). Testing And Interpreting Uncovered Interest Parity In Russia. *Russian Journal of Economics*, *3*(2), 158–173. https://doi.org/10.1016/j.ruje.2017.06.003

Vest, J. R., Unruh, M. A., Casalino, L. P., & Shapiro, J. S. (2020). The Complementary Nature Of Query-Based And Directed Health Information Exchange In Primary Care Practice. *Journal of the American Medical Informatics Association*, *27*(1), 73–80. https://doi.org/10.1093/jamia/ocz134

Vest, J. R., Unruh, M. A., Shapiro, J. S., & Casalino, L. P. (2019). The Associations Between Query-Based And Directed Health Information Exchange With Potentially Avoidable Use Of Health Care Services. *Health Services Research*, *54*(5), 981–993. https://doi.org/10.1111/1475-6773.13169

Vonderach, C., Kändler, G., & Dormann, C. F. (2018). Consistent Set Of Additive Biomass Functions For Eight Tree Species In Germany Fit By Nonlinear Seemingly Unrelated Regression. *Annals of Forest Science*, *75*(2), 49. https://doi.org/10.1007/s13595-018-0728-4

Wagner, M., Grabarczyk, P., & Hong, S. H. (2020). Fully Modified OLS Estimation And Inference For Seemingly Unrelated Cointegrating Polynomial Regressions And The Environmental Kuznets Curve For Carbon Dioxide Emissions. *Journal of Econometrics*, *214*(1), 216–255. https://doi.org/10.1016/j.jeconom.2019.05.012

Walker, A. J. (1977). An Efficient Method for Generating Discrete Random Variables with General Distributions. *ACM Transactions on Mathematical Software*, *3*(3), 253–256. https://doi.org/10.1145/355744.355749

Wang, W., & Famoye, F. (1997). Modeling Household Fertility Decisions With Generalized Poisson Regression. *Journal of Population Economics*, *10*(3), 273–283. https://doi.org/10.1007/s001480050043

Winkelmann, R. (2000). Seemingly Unrelated Negative Binomial Regression. *Oxford Bulletin of Economics and Statistics*, *62*(4), 553–560. https://doi.org/10.1111/1468-0084.00188

Winkelmann, R. (2012). Copula Bivariate Probit Models: With An Application To Medical Expenditures: Copula Bivariate Probit Models. *Health Economics*, *21*(12), 1444–1455. https://doi.org/10.1002/hec.1801

Wooldridge, J. M. (1999). Distribution-Free Estimation Of Some Nonlinear Panel Data Models. *Journal of Econometrics*, *90*(1), 77–97. https://doi.org/10.1016/S0304-4076(98)00033-5

Yildirim, I. (2012). *Bayesian Inference: Metropolis-Hastings Sampling*. Unpublished Manuscript

Zamzuri, Z. H. (2016). *Selected models for correlated traffic accident count data*. 060024. Johor Bahru, Malaysia. https://doi.org/10.1063/1.4954629

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association*, *57*(298), 348–368. https://doi.org/10.1080/01621459.1962.10480664

Zhang, H., Xu, J., Jiang, N., Hu, X., & Luo, Z. (2015). PLNseq: A Multivariate Poisson Lognormal Distribution For High-Throughput Matched RNA-Sequencing Read Count Data: Plnseq: A Multivariate Poisson Lognormal Distribution For High-Throughput Matched RNA-Sequencing Read Count Data. *Statistics in Medicine*, *34*(9), 1577–1589. https://doi.org/10.1002/sim.6449

Zhang, Y., Zhou, H., Zhou, J., & Sun, W. (2017). Regression Models for Multivariate Count Data. *Journal of Computational and Graphical Statistics*, *26*(1), 1–13. https://doi.org/10.1080/10618600.2016.1154063

Curriculum Vitae

Yilei Zhang

Education

Ph.D. (Economics) Indiana University School of Liberal Arts at Indiana University-Purdue University Indianapolis, July 2021

M.A. (Economics) University of Michigan, Ann Arbor, May 2014

B.A. (Economics) Northeast Normal University, June 2010

Fields of Interest

Applied Econometrics; Health Economics

Awards and Honors

Award for Best Instructor Among PhD Students, Indiana University School of Liberal Arts at IUPUI, 2019-2020

Full Teaching Assistantship, University of Arizona, 2014-2015

JASSO Honors Scholarship, Okayama University, 2009-2010

Merit-Based Presidential Scholarship, Northeast Normal University, 2009-2010

Work Experiences

Assistant Professor, Henan University, 2021 – present

Research Analyst, Illuminate Health Inc., 2017 summer

Instructor, IUPUI, Indianapolis, IN, 2018 – 2021

Teaching Assistant, Eller School of Business, University of Arizona, 2014-2015

Research Assistant, Ross School of Business, University of Michigan, 2013-2014

Teaching Experiences

Undergraduate Level: Survey of International Economics (2018 Fall, 2019 Fall, 2020 Fall/Spring, 2021 Spring), Introductory Microeconomics (2018 Spring, 2020 Summer), Intermediate Microeconomics Theory (2020 Spring), IUPUI, Managerial Economics and Business Strategy (2019 Spring), Kelly School of Business, Indiana University, Bloomington

Masters Level: MBA Managerial Economics (2020 Summer), H-E-B School of Business, University of Incarnate Word

Ph.D. Level: Microeconomic Theory I (TA, 2017 Fall), IUPUI

Working Papers

Terza, J. and Zhang, A., "Exploring the Importance of Accounting for Nonlinearity in Correlated Count Regression Systems from the Perspective of Causal Estimation and Inference"

Terza, J. and Zhang, A., “Hospitals Integration and Quality Decisions after the Implementation of Hospital Value-Based Purchasing Program” (CMS DUA: RSCH-2020-5483)