ESSAYS IN THE ESTIMATION OF SYSTEMS OF LIMITED DEPENDENT VARIABLES WITH APPLICATION TO DEMAND SYSTEMS

By

FAYSAL HABIB FAHS

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
School of Economic Sciences

August 2008

To the faculty of Washington State University:

The members of the committee appointed to examine the dissertation of FAYSAL HABIB FAHS find it satisfactory and recommend that it be accepted.

____

Dr. Ron C. Mittelhammer (Chair)

____

Dr. Jill McCluskey

____

Dr. Jonathan Yoder

ACKNOWLEDGMENTS

I would like to formally thank:

My wonderful parents, Habib and Ilham, who have raised me to be the person I am today. You have been with me every step of the way, through the good time and bad. Thank you for all the unconditional love, guidance, and support that you have always given me, helping me to succeed and instilling in me the confidence that I am capable of doing anything I put my mind to. Thank you for everything. I love you!

Dr. Ron C. Mittelhammer, my advisor, for your hard work and guidance throughout this entire dissertation process and for believing in my abilities. I have learned so much, and without you, this would not have been possible. Thank you so much for a great experience. You are truly an exceptional teacher and advisor!

Dr. Jill McCluskey and Dr. Jonathan Yoder for serving on my Dissertation committee, thank you for believing in me and for your insight and guidance throughout the past four years. Your excellent comments and generous input enriched the content of the research.

Dr. Scott Matulich and Dr. Kenneth Casavant, who both fueled my passion in teaching. Thank you for your sincere friendship and continuous support.

My brothers Rafic and Yasser, my sisters-in-law, Marica and Dareen, thanks for your love and support throughout the years.

The School of Economic Sciences and Washington State University, thank you for the financial and technical support.

My fellow graduate students: thank you for your friendship and support. The last years have been quite an experience and you have all made it a memorable time of my life. I will miss all of you. Good luck to each of you in your future endeavors.

My soul mate and confidant, Angela, you always have been there for me. Thank you for your continual love, support, and patience as I went through this journey. I could not have made it through without you by my side.

ESSAYS IN THE ESTIMATION OF SYSTEMS OF LIMITED DEPENDENT VARIABLES WITH APPLICATION TO DEMAND SYSTEMS

Abstract

by Faysal Habib Fahs, Ph.D.
Washington State University
School of Economic Sciences
August 2008

Chair: Ron C. Mittelhammer

This dissertation includes three essays in the estimation of Limited Dependent Variables and Demand System Models. In the first essay, we utilize the Generalized Method of Moments (GMM) approach to estimate censored equation systems. The GMM approach is based on a common set of marginal and bivariate moment relations that hold between the explanatory variables and the model noise. We review the computational problems involved in estimating Multivariate Tobit (MVT) Models of relatively high dimension, and then note how our GMM approach addresses the computational burden. The GMM estimator is consistent, asymptotically normally distributed, near-asymptotically efficient, and computationally easy and tractable as the dimensionality of the model increases. Finally Monte Carlo experiments were conducted to investigate and compare the performance of the GMM approach to the Simulated Maximum Likelihood (SML) estimator with different distributional assumptions. The GMM estimator demonstrates itself as an empirically tractable way of estimating systems of censored regressions involving large samples and high dimensional models.

The Second essay examines the impact of the *E.coli* outbreaks that occurred in 2006 on consumer demand for salad vegetables on the West Coast of the United States. The scanner-data set used in our analysis is obtained from a chain supermarket and is aggregated on a weekly basis for the consumption of salad vegetables. The data contain a significant portion of observations with zero consumption on one or more vegetable groups. Zero consumption may be reflecting consumer concern about the *E.coli* outbreaks, the effect of removal of vegetable groups from store shelves due to product recalls and/or the result of personal preferences with respect to consumption. We motivate the use of the Tobit model as a statistical representation of consumer behavior by specifying the Quadratic Almost Ideal Demand System (QUAIDS) with demographic effects under binding non-negativity constraints. To avoid violating the non-negativity constraints of the model and to overcome the computational burden of high dimensionality, the GMM approach, along with the virtual prices concept, are used for estimating the system of non-linear censored demand equations. The empirical results show that during the outbreak period lettuce and cabbage were substituted for spinach, indicating consumers' concern about the *E.coli* impact.

The third essay utilizes the Minimum Power Divergence (MPD) class of probability distributions to estimate censored regression models. Based on the minimization of the Cressie-Read (CR) power divergence function, we are able to implement an estimator that requires less priori model structure than conventional parametric models such as the Tobit estimator. Our estimator assumes that the distribution of the noise term is neither based on, nor restricted to, the conventional parametric families (normal, logistic) and suggests a range of CDFs that is based on the

MPD principle. The paper pursues two estimation approaches to estimate censored regression model using the MPD principle: 1) Generalized Method of Moments (GMM) and 2) Maximum Likelihood approach (ML). Monte Carlo sampling experiments suggest that the estimators within the CR class will be more robust than conventional methods often used in empirical practice while also producing estimation precision that rivals the tightly specified parametric approaches in the event that the data generating distributional assumptions underlying the parametric specifications are true.

TABLE OF CONTENTS

Page

LIST OF TABLES

ESSAY ONE:

ESSAY TWO:

ESSAY THREE:

LIST OF FIGURES

ESSAY ONE:

ESSAY TWO:

ESSAY THREE:

Dedication

This work is dedicated to my family

INTRODUCTION

Models with Limited Dependent Variables (LDV) have received considerable attention in the econometric literature. LDV models stem from the increasing use of household and firm-level microeconomic data in empirical analysis of consumer demand, labor supply, output supply, and input demand, among other empirical economic contexts.

In this dissertation we focus on a situation where the dependent variable is censored. With the assumption of normality, the Tobit model is one of the classical models that can address the situation of censoring using Maximum Likelihood (ML) estimation procedure. In a systems context, the Multivariate Tobit (MVT) Model typically has a correlated error structure and requires a joint estimation procedure based on a mixed continuous-discrete distribution consisting of the continuous probability density function and the discrete mass function for the zero observations. Using ML for estimating MVT models requires the evaluation of the partially integrated multivariate normal probability density function. This is known to be computationally inefficient and quickly becomes intractable as the dimensionality of the model increases.

The first essay in this dissertation suggests using the Generalized Method of Moments approach (GMM) as an alternative approach for estimating MVT models. The GMM approach is based on a common set of marginal and bivariate moment relations that hold between explanatory variables and model noise. Estimates obtained by the GMM procedure are consistent, asymptotically normally distributed, and near-asymptotically efficient. Other advantages of the GMM, in comparison to the ML approach, are that it is empirically tractable as the dimensionality of the model increases, computationally easy and relatively fast, and flexible for imposing restrictions.

The second essay provides a substantive illustration of the empirical application of the GMM approach, presenting a case study that analyzes and estimates the demand for vegetable salads after the 2006 *E. coli* outbreaks on the West Coast of the United States. A censored Quadratic Almost Ideal Demand System (QUAIDS) under binding non-negativity constraints is used as a functional form for the demand system. Because of the *E. coli* outbreaks, the expenditure data obtained from the retail store contains a high frequency of zero consumption observations for spinach. The zero consumption can be representative of consumer concern about the *E.coli* outbreaks, the effect of removal of vegetable groups from the shelves due to product recall, or the result of personal preferences. The GMM approach alone is not sufficient for the analysis, since there is a possibility of the violation of the non-negativity constrains and the adding up conditions. This additional econometric complication was addressed by introducing the virtual prices concept within the GMM estimation approach. The proposed approach guarantees that the regularity conditions and adding up constraints are satisfied in these models. The empirical results along with an elasticity analysis suggest the usefulness of this approach in the analysis of the *E. coli* outbreaks on consumer demand for salad vegetable.

The third essay utilizes the Minimum Power Divergence (MPD) class of probability distributions to estimate censored regression models. Based on the minimization of the Cressie-Read (CR) power divergence function, we are able to implement an estimator that requires less priori model structure than conventional parametric models such as the Tobit estimator. Our estimator assumes that the distribution of the noise term is neither based on, nor restricted to, the conventional parametric families (normal, logistic) and suggests a range of CDFs that is based on the MPD principle. The paper pursues two estimation approaches to estimate censored regression model using the MPD principle: 1) Generalized Method of Moments (GMM) and 2) Maximum Likelihood approach (ML). Monte Carlo sampling experiments suggest that the estimators within the CR class will be more robust than conventional methods often used in empirical practice while also producing estimation precision that rivals the tightly specified parametric approaches in the event that the data generating distributional assumptions underlying the parametric specifications are true.

GENERALIZED METHOD OF MOMENTS ESTIMATION OF CENSORED EQUATION SYSTEMS

Faysal Habib Fahs

School of Economic Sciences
Washington State University
Pullman, WA 99163
Email: faysal74@wsu.edu

ESSAY ONE:

GENERALIZED METHOD OF MOMENTS ESTIMATION OF CENSORED EQUATION SYSTEMS

Abstract

In this paper we utilize the Generalized Method of Moments (GMM) approach to estimate censored equation systems. The GMM approach is based on a common set of marginal and bivariate moment relations that hold between the explanatory variables and the model noise. We review the computational problems involved in estimating Multivariate Tobit (MVT) Models of relatively high dimension, and then note how our GMM approach addresses the computational burden. The GMM estimator is consistent, asymptotically normally distributed, near-asymptotically efficient, and computationally easy and tractable as the dimensionality of the model increases. Finally Monte Carlo experiments were conducted to investigate and compare the performance of the GMM approach to the Simulated Maximum Likelihood (SML) estimator with different distributional assumptions. The GMM estimator demonstrates itself as an empirically tractable way of estimating systems of censored regressions involving large samples and high dimensional models.

1. Introduction

In statistical applications, one often encounters a situation where the dependent variable is only observable under certain conditions. Censoring occurs when observations on the dependent variable are limited in range, whereby observations are restricted to one or both endpoints of an interval with positive probability mass.

A common example of the above situation is Tobin's (1958) study of household expenditures. The consumer maximizes his or her utility by purchasing durable goods under the constraint that total expenditures do not exceed income. The expenditure for durable goods must be at least equal to the cost of the least expensive item or else the outcome is censored at a zero value. Many other examples of censored outcomes can be found: hours worked by wives (Quester and Green, 1982), scientific publications (Stephan and Levin, 1992), extramarital affairs (Fair, 1978), foreign trade and investment (Eaton and Tamura, 1994), austerity protests in Third World countries (Walton and Ragin, 1990), damage caused by a hurricane (Fronstin and Holtmann, 1994) and in addition to a wide range of examples that researchers encounter in both economic and econometric analysis.

One of the econometric approaches for addressing censored dependent variables in a single equation setting is using the standard Tobit estimator while assuming that the error term is normally distributed. However, in system of equations with multiple censored variables the error terms are correlated and require a joint estimation procedure. Lee (1993) proposed a specification of a mixed continuous-discrete distribution which consists of a continuous probability density function for positive observation and a discrete probability mass function for the zero observations.

Maximum likelihood (ML) is a commonly used estimation approach in the context of censored regressions. It requires the evaluation of a partially integrated multivariate normal density function (in the case of a Tobit model), which is known to be a computationally inefficient, inaccurate and intractable approach as dimensionality of the model increases (Lee, 1993). Several studies attempted to apply alternative computationally more tractable techniques. For example, Pudney (1989) estimated a system of Tobit equations by applying the Tobit technique to each equation marginally. Amemiya (1974) proposed a model based only on jointly positive sample outcomes of all dependent variables. While both techniques are consistent and numerically tractable, the estimators are inefficient. The former ignores the inter-correlation between the equations and fails to impose any cross equation restrictions, such as symmetry conditions, while the latter does not consider cases where the number of observations with positive valued dependent variables is very small i.e., the probability of joint positivity is small. Maddala (1977) modified Amemiya's (1974) procedure so as to use all the sample observations pertaining to the model, but a major shortcoming of his procedure is the evaluation of partially integrated multivariate normal probability density functions.

Evaluating the probability of the censored observations involves high dimensional integration in the case of a system with a relatively large number of equations, and an appreciable incidence of censoring. In response, researchers often restrict their attention to special classes of multivariate Tobit models where the censored observations are ignored. Ignoring the specific data sampling characteristics of the zero observations while modeling and estimating the parameters of the model leads to biased estimates.

Excluding the censored observations will result in sample selection bias (Lee and Pitt, 1987).

Hajivassiliou and Ruud (1994) suggested the Simulated Maximum Likelihood (SML) approach based on the Geweke-Hajivassiliou-Keane (GHK) simulator for overcoming the problem of high dimensional numerical integration underlying the choice probabilities in systems of censored equations. Hasan and Mittelhammer (2001) developed an improved type of GHK Called the "ordered GHK" (ORDGHK) simulator which is based on efficient reordering of the integrations that occur in the definition of the conditional probabilities in the GHK simulators. The advantage of the reordering is the calculation of multivariate rectangular probabilities, which is less computationally difficult, more numerically accurate and faster in convergence compared to other simulators. Although a great deal of work has been done to overcome the problem of high dimensionality, a remaining concern was tractability, simplicity and convergence time.

This paper utilizes the GMM approach, first used by Fahs and Mittelhammer (2001), to estimate systems of MVT models. In order to increase efficiency, we introduce third moment bivariate conditions along with cross-moments in the GMM approach. Adding more moments conditions, where asymptotic efficiency comparisons suggest that more is better than less, will generally decrease the asymptotic covariance matrix of the GMM estimator. The GMM uses a common set of marginal and bivariate moment conditions that takes into account correlation among the random components of the model to calculate parameter estimates and covariance structure of the error terms. The GMM estimator is consistent, empirically tractable, asymptotically normally distributed, and near-asymptotically efficient while being more robust relative to alternative sampling distributions.

The reminder of the paper is organized as follows: Section 2 summarizes the basic structure of the MVT model. Section 3 describes the moment conditions used in the GMM approach. Section 4 introduces the GMM estimation of the MVT model. Section 5 illustrates the Monte Carlo experiments and summarizes the results. Lastly, Section 6 ends with our conclusions.

2. Multivariate Tobit Model

In a multivariate Tobit model we have

(2.1) $$\boldsymbol{Y}_i^* = \boldsymbol{X}_i\beta + \varepsilon_i \ , i = 1, \dots, n,$$

where $\varepsilon \sim N([0], \Sigma)$ and $\boldsymbol{Y}_i^* \sim N(\boldsymbol{X}_i\beta, \Sigma)$ and i denotes the observation number. In this model $\boldsymbol{Y}_i^*$ is a $J \times 1$ continuous latent variables that determine the choice probabilities of an optimizing agent,

$$\boldsymbol{X}_i\beta = \begin{bmatrix} \boldsymbol{X}_{1i}\beta_1 \\ \vdots \\ \boldsymbol{X}_{Ji}\beta_J \end{bmatrix} = \boldsymbol{U}_i = \begin{bmatrix} U_{1i} \\ \vdots \\ U_{Ji} \end{bmatrix}$$ and $U_{ji} = \boldsymbol{X}_{ji}\beta_j$, where $\boldsymbol{X}_{ji}$ is $1 \times K_j$ row vector of explanatory variables, β_j is $K_j \times 1$ column vector of parameters and Σ is $J \times J$ covariance matrix. Here, and elsewhere, the symbol U_{ji} is the same as $\boldsymbol{X}_{ji}\beta_j$. Assuming the censoring occurs at zero, the model is generally written as:

(2.2) $$Y_{ji} = \begin{cases} Y_{ji}^* & if \ \ Y_{ji}^* > 0 \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } i = 1, 2, \dots, n, \text{ and } \ j = 1, 2, \dots, J$$

where the subscript i denotes the observation number and the subscript j denotes choice alternative.

More generally, for a given individual decision maker, the latent vector $\boldsymbol{Y}_i^*$ can be written as:

$$(2.3) \qquad \boldsymbol{Y}_i^* = \begin{bmatrix} \boldsymbol{Y}_{id}^* \\ \\ \boldsymbol{Y}_{ic}^* \end{bmatrix}$$

where $\boldsymbol{Y}_{id}^*$ represents the discrete elements of $\boldsymbol{Y}_i^*$ associated with the outcomes $\boldsymbol{Y}_{id} = 0$, and $\boldsymbol{Y}_{ic}^*$ represents the non-censored observations corresponding to $\boldsymbol{Y}_{ic} = \boldsymbol{Y}_{ic}^* > 0$.

Let $\theta = [\beta, \Sigma]$ denote all the parameters of the model. Then the joint probability density of $(\boldsymbol{Y}_{ic}^*, \boldsymbol{Y}_{id}^*)$ can be represented in general by

$$(2.4) \qquad f(\boldsymbol{Y}_{id}^*, \boldsymbol{Y}_{ic}^*; \boldsymbol{X}, \theta) = f(\boldsymbol{Y}_{id}^* \mid \boldsymbol{Y}_{ic}^*; \boldsymbol{X}, \theta) \times f(\boldsymbol{Y}_{ic}^*; \boldsymbol{X}, \theta)$$

where $f(\boldsymbol{Y}_{id}^* \mid \boldsymbol{Y}_{ic}^*; \boldsymbol{X}, \theta)$ represents the discrete choice probability conditional on the continuous variable $\boldsymbol{Y}_{ic}^*$, and $f(\boldsymbol{Y}_{ic}^*; \boldsymbol{X}, \theta)$ is the probability density function for the continuous random variable. Thus the contribution of the i^{th} observation to the likelihood function will be

$$(2.5) \qquad L(\boldsymbol{Y}_{id}^*, \boldsymbol{Y}_{ic}^*; \boldsymbol{X}, \theta) = L(\boldsymbol{Y}_{ic}^*; \boldsymbol{X}, \theta) \times L(\boldsymbol{Y}_{id}^* \mid \boldsymbol{Y}_{ic}^*; \boldsymbol{X}, \theta).$$

The first component of the likelihood function, $L(\boldsymbol{Y}_{ic}^*; \boldsymbol{X}, \theta)$, is a J_c-dimensional multivariate normal probability density and its value is straightforward to calculate. The second part of the likelihood function, $L(\boldsymbol{Y}_{id}^* \mid \boldsymbol{Y}_{ic}^*; \boldsymbol{X}, \theta)$, is difficult to evaluate, where different simulators have been suggested to overcome this problem. These simulators become computationally intractable and slow in convergence, and accuracy decreases as the dimensionality of the model increases. We demonstrate these problems in section 5 as we compare the GMM approach with SML using an ordered GHK simulator (ORDGHK).

3. Moments of the MVT Model

Given the MVT model presented in section 2, this section presents the moment conditions that hold between the explanatory variables and the model noise. Two types of moment conditions are introduced. The first is the marginal moment conditions and the second type is the bivariate moments that help to identify and estimate the parameters involved in the covariance matrix across equation errors.

3.1 Marginal Moments Conditions

Considering the model given by equation (2.2) and conditioning on the positive observations, the first and second conditional moments of a truncated normal random variable first derived by Amemiya (1973) are given by

(3.1.1) $$E\left(Y_{j.} \mid Y_{j.}>0\right)=U_{j.}+\sigma_j \frac{\phi_j}{\Phi_j}$$

(3.1.2) $$E\left(Y_{j.}^2 \mid Y_{j.}>0\right)=U_{j.}E\left(Y_{j.} \mid Y_{j.}>0\right)+\sigma_j^2$$

where the subscript j denotes the j^{th} choice alternative[1]. Equation (3.1.1) represents the first order marginal moments condition on all the positive observations in $Y_{j.}$, ϕ_j is the

---

[1] $U_{j.}$ shorthand for $\begin{bmatrix} X_{1i}\beta_1 \\ \vdots \\ X_{Ji}\beta_J \end{bmatrix}$, $Y_{j.}$ shorthand for $\begin{bmatrix} Y_{1i} \\ \vdots \\ Y_{Ji} \end{bmatrix}$, *and* $X_{j.}$ shorthand for $\begin{bmatrix} X_{1i} \\ \vdots \\ X_{Ji} \end{bmatrix}$

probability density function (PDF)and Φ_j is the cumulative density function (CDF), and both are shorthand for $\phi\left(\frac{U_{j.}}{\sigma_j}\right)$ and $\Phi\left(\frac{U_{j.}}{\sigma_j}\right)$, respectively (these notations apply for the rest of the paper). Equation (3.1.2) represents the second order marginal moments condition on all the positive observations in $Y_{j.}$.

Using all the observations in $Y_{j.}$, instead of using the non-zero observations, we can represent two unconditional marginal moments (Heckman, 1976*b*) as:

(3.1.3) $$E\left(Y_{j.}\right) = U_{j.}\Phi_j + \sigma_j\phi_j$$

(3.1.4) $$E\left(Y_{j.}^2\right) = U_{j.}E\left(Y_{j.}\right) + \sigma_j^2\Phi_j$$

where ϕ_j and Φ_j represents the PDFs and the CDFs, respectively, equations (3.1.3) and (3.1.4) represent the first and the second order of the unconditional marginal moments respectively.

Based on asymptotic efficiency comparisons, the asymptotic covariance matrix of the GMM estimator generally becomes smaller as the number of nonredundant moments used increases. We can derive an additional moment condition by defining a binary variable, analogous to the Probit case:

$$Y_{Binary\,j.} = \begin{Bmatrix} 1 & if \quad Y_{j.} > 0 \\ 0 & \text{otherwise} \end{Bmatrix}$$

The binary marginal moment condition can be represented as

(3.1.5) $$\mathrm{E}(Y_{Binary\,j.}) = \Phi_j$$

where Φ_j represents the CDF and is a shorthand for $\Phi\left(\frac{U_{j.}}{\sigma_j}\right)$.

Gathering all of the marginal moment conditions (3.1.1)–(3.1.5) for all n observations, we can define the following relationships between $U_{j.}$, $Y_{j.}$ and disturbances $\xi_j^{(i)}$ that have zero expectations as:

$$(3.1.6)\qquad \left\{ \begin{array}{c} Y_{j.>0} = U_{j.>0} + \sigma_j \odot \dfrac{1}{\boldsymbol{\Phi}_{j>0}} \odot \boldsymbol{\phi}_{j>0} + \xi_j^{(1)} \\ Y_{j.>0}^2 = U_{j.>0} \odot E(Y_{j.>0}) + \sigma_j^{\,2} + \xi_j^{(2)} \\ Y_{j.} = U_{j.} \odot \boldsymbol{\Phi}_j + \sigma_j \odot \boldsymbol{\phi}_j + \xi_j^{(3)} \\ Y_{j.}^{\,2} = U_{j.} \odot E\left(Y_{j.}\right) + \sigma_j^{\,2} \odot \boldsymbol{\Phi}_j + \xi_j^{(4)} \\ Y_{Binary\ j.} = \boldsymbol{\Phi}_j + \xi_j^{(5)} \end{array} \right\}$$

where $Y_{j.>0}$ denotes the set of positive valued observations relating to the j^{th} choice alternative, $U_{j.>0}$ denotes the observations on the explanatory variables that correspond to the positive valued outcomes, $E\left(Y_{j.>0}\right)$ is shorthand notation for $EY_{j.>0} \equiv E\left(Y_{j.} \mid Y_{j.}>0\right) = \left(U\right)_{j.>0} + \sigma_j \odot \dfrac{1}{\boldsymbol{\Phi}_{j>0}} \odot \boldsymbol{\phi}_{j>0}$, $\odot$ denotes the Hadamard (elementwise) product. In this context, $\boldsymbol{\Phi}_j$ and $\boldsymbol{\phi}_j$ are vectors the cumulative distribution function and density function values of the standard normal distribution, evaluated at the vector $U_{j.} \odot \left(\dfrac{1}{\sigma_j}\right)$, $\boldsymbol{\Phi}_{j>0}$ and $\boldsymbol{\phi}_{j>0}$ are the subsets of those vectors corresponding to the positive valued observations $Y_{j.>0}$, and $\dfrac{1}{\boldsymbol{\Phi}_{j>0}}$ denotes a vector of reciprocals of the elements in $\boldsymbol{\Phi}_{j>0}$.

Because

$$\xi_j^{(1)} = Y_{j.>0} - U_{j.>0} \text{ - } \sigma_j \odot \frac{1}{\boldsymbol{\Phi}_{j>0}} \odot \boldsymbol{\phi}_{j>0},$$

$$\xi_j^{(2)} = Y_{j.>0}^2 - U_{j.>0} \odot E(Y_{j.>0}) \text{ - } \sigma_j^{\,2},$$

$$\xi_j^{(3)} = Y_{j.} - U_{j.} \odot \boldsymbol{\Phi}_j \text{ - } \sigma_j \odot \boldsymbol{\phi}_j,$$

$$\xi_j^{(4)} = Y_{j.}^{\,2} - U_{j.} \odot E\left(Y_{j.}\right) \text{ - } \sigma_j^{\,2} \odot \boldsymbol{\Phi}_j,$$

and

$$\xi_j^{(5)} = Y_{Binary\ j.} \text{ - } \Phi_j.$$

Orthogonality conditions can be defined as follows:

$$E\left[\boldsymbol{X}'_{j.>0}\left(Y_{j.>0} \text{ - } U_{j.>0} \text{ - } \sigma_j \odot \frac{1}{\boldsymbol{\Phi}_{j>0}} \odot \boldsymbol{\phi}_{j>0}\right)\right] = 0,$$

$$E\left[\boldsymbol{X}'_{j.>0}\left(Y_{j.>0}^2 \text{ - } U_{j.>0} \odot EY_{j>0} \text{ - } \sigma_j^{\,2}\right)\right] = 0,$$

$$E\left[\boldsymbol{X}'_{j.}\left(Y_{j.} \text{ - } U_{j.} \odot \boldsymbol{\Phi}_j \text{ - } \sigma_j \odot \boldsymbol{\phi}_j\right)\right] = 0,$$

$$E\left[\boldsymbol{X}'_{j.}\left(Y_{j.}^{\,2} \text{ - } U_{j.} \odot E\left(Y_{j.}\right) \text{ - } \sigma_j^{\,2} \odot \boldsymbol{\Phi}_j\right)\right] = 0,$$

and

$$E\left[\boldsymbol{X}'_{j.}\left(Y_{Binary\ j.} \text{ - } \boldsymbol{\Phi}_j\right)\right] = 0,$$

where $\boldsymbol{X}_{j.}$ denotes the observations on the explanatory variables for the j^{th} outcome and having dimension $(n \times K)$, $\boldsymbol{X}_{j.>0}$ denotes the observations on the explanatory variables that correspond to the j^{th} positive valued outcomes. We can define a $(5KJ \times 1)$ vector of moment conditions derived from the orthogonality conditions as

$$(3.1.7)\qquad E\left[\boldsymbol{h}_{\text{j Marginal}}(\boldsymbol{Y},\boldsymbol{X},\theta)\right]=E\begin{bmatrix} \boldsymbol{X}'_{j.>0}\left(Y_{j.>0}-U_{j.>0}-\sigma_j\odot\dfrac{1}{\boldsymbol{\Phi}_{j>0}}\odot\boldsymbol{\phi}_{j>0}\right) \\ \boldsymbol{X}'_{j.>0}\left(Y^2_{j.>0}-U_{j.>0}\odot E\left(Y_{j.>0}\right)-\sigma_j^{\;2}\right) \\ \boldsymbol{X}'_{j.}\left(Y_{j.}-U_{j.}\odot\boldsymbol{\Phi}_j-\sigma_j\odot\boldsymbol{\phi}_j\right) \\ \boldsymbol{X}'_{j.}\left(Y_{j.}^{\;2}-U_{j.}\odot E\left(Y_{j.}\right)-\sigma_j^{\;2}\odot\boldsymbol{\Phi}_j\right) \\ \boldsymbol{X}'_{j.}\left(Y_{Binary\,j.}-\boldsymbol{\Phi}_j\right) \end{bmatrix}=0.$$

Then the sample analog of the population moments displayed in (3.1.7) is

$$(3.1.8)\qquad \boldsymbol{h}_{\text{j Marginal}}\left(\boldsymbol{y},\boldsymbol{x},\theta\right)=\begin{bmatrix} \dfrac{\boldsymbol{x}'_{j.>0}}{n_j^{(1)}}\left(y_{j.>0}-u_{j.>0}-\sigma_j\odot\dfrac{1}{\boldsymbol{\Phi}_{j>0}}\odot\boldsymbol{\phi}_{j>0}\right) \\ \dfrac{\boldsymbol{x}'_{j.>0}}{n_j^{(2)}}\left(y^2_{j.>0}-u_{j.>0}\odot E\left(y_{j.>0}\right)-\sigma_j^{\;2}\right) \\ \dfrac{\boldsymbol{x}'_{j.}}{n_j^{(3)}}\left(y_{j.}-u_{j.}\odot\boldsymbol{\Phi}_j-\sigma_j\odot\boldsymbol{\phi}_j\right) \\ \dfrac{\boldsymbol{x}'_{j.}}{n_j^{(4)}}\left(y_{j.}^{\;2}-u_{j.}\odot E\left(y_{j.}\right)-\sigma_j^{\;2}\odot\boldsymbol{\Phi}_j\right) \\ \dfrac{\boldsymbol{x}'_{j.}}{n_j^{(5)}}\left(y_{Binary\,j.}-\boldsymbol{\Phi}_j\right) \end{bmatrix}=0,$$

where $n_j^{(i)}$ denotes the number of sample observations that correspond to the i^{th} set of moment conditions for the j^{th} choice alternative, $E\left(y_{j.>0}\right)$ is equal to $E\left(Y_{j.>0}\right)$ evaluated at sample outcomes for $y_{j.}$ and $\boldsymbol{x}_{j.}$ and at specified values for θ. Note we changed variables from capital letters to small letters to indicate that we are evaluation at the sample moments.

The dimension of $\boldsymbol{h}_{\text{Marginal}}(Y, X, \theta)$ is $\left(5KJ \times 1\right)$, which is greater than the number of unknown parameters in the model. Consequently, in general there will not exist a unique parameter vector θ that solves the sample moment conditions via the ordinary method of moment's approach, which attempts to find a θ that satisfies (3.1.8). We will deal with identification issues later in the paper.

3.2 Bivariate Moment Conditions

The bivariate moments help to identify and estimate the parameters involve in the covariance structure occurring across equation errors. In addition, the bivariate moments avoid the problem of evaluating the probability of the discontinuous part in higher dimensions because numerical integration is then only required in two dimensions, which is accurate and computationally fast.

Tallis (1961) derived the first two moments and the moment generating function of the truncated multivariate normal distribution. Following Tallis, Fahs and Mittelhammer (2001) were able to derive the first two moments of the truncated bivariate normal distribution based on a different parameterization of the model. In this paper we derive the first, second and third order bivariate moments along with all the cross moment conditions. Based on Tallis (1961), Fahs and Mittelhammer (2001), Appendix 2 shows all the calculation for any decision outcomes[2] $\left(Y_{j.}, Y_{k.}\right)$.

The MVT model is characterized by J alternative choices, so that there are $J\left(\frac{J-1}{2}\right)$ alternative pairs of decision outcomes that can be examined in a bivariate

[2] Note j and k denotes the choice alternatives and Capital K represents the number of columns in the explanatory variables, $\boldsymbol{X}_i{}'s$

manner. For example, in a five choice-model, there are ten bivariate pairs for each observation given by: $(y_{1i,}y_{2i}),(y_{1i,}y_{3i}),(y_{1i,}y_{4i}),(y_{1i,}y_{5i}),(y_{2i,}y_{3i}),(y_{2i,}y_{4i}),(y_{2i,}y_{5i}),(y_{3i,}y_{4i})$, $(y_{3i,}y_{5i})$ and $(y_{4i,}y_{5i})$. For each pair, one can derive nine moment conditions through the third order. For example for (Y_{1i},Y_{2i}) outcome we can represent the moments as:

The first order bivariate moments: $E(y_{1i} \mid y_{1i} \geq 0, y_{2i} \geq 0)$ and $E(y_{2i} \mid y_{1i} \geq 0, y_{2i} \geq 0)$

The second order bivariate moments: $E(y_{1i}^2 \mid y_{1i} \geq 0, y_{2i} \geq 0)$ and $E(y_{2i}^2 \mid y_{1i} \geq 0, y_{2i} \geq 0)$

The third order bivariate moments: $E(y_{1i}^3 \mid y_{1i} \geq 0, y_{2i} \geq 0)$ and $E(y_{2i}^3 \mid y_{1i} \geq 0, y_{2i} \geq 0)$

Lastly the cross bivariate moments: $E(y_{1i}y_{2i} \mid y_{1i} \geq 0, y_{2i} \geq 0)$, $E(y_{1i}y_{2i}^2 \mid y_{1i} \geq 0, y_{2i} \geq 0)$ and $E(y_{2i}^2 y_{1i} \mid y_{1i} \geq 0, y_{2i} \geq 0)$

Similar to the orthogonality conditions presented in section 3.1 relating to the marginal moments, and noting that for any pair of decision outcomes, $(y_{j.}, y_{k.})$, there are nine bivariate moments. We can define a $\left(\frac{9J(J-1)K}{2} \times 1\right)$ vector function of bivariate moment conditions based on the preceding results as:

$$(3.2.1) \qquad E\left[\boldsymbol{h}_{j\text{Bivariate}}(\boldsymbol{Y},\boldsymbol{X},\theta)\right] = \begin{bmatrix} \boldsymbol{X}_{j.}'\left(\boldsymbol{Y}_{j.} - E\left(\boldsymbol{Y}_{j.} \mid \boldsymbol{Y}_{j.} \geq 0, Y_{k.} \geq 0\right)\right) \\ \scriptstyle for\ j\ and\ k=1,2,\ldots J;\ j\neq k \\ \vdots \\ \boldsymbol{X}_{j.}'\left(\boldsymbol{Y}_{j.}^2 - E\left(\boldsymbol{Y}_{j.}^2 \mid \boldsymbol{Y}_{j.} \geq 0, \boldsymbol{Y}_{k.} \geq 0\right)\right) \\ \scriptstyle for\ j\ and\ k=1,2,\ldots J;\ j\neq k \\ \vdots \\ \boldsymbol{X}_{j.}'\left(\boldsymbol{Y}_{j.}^3 - E\left(\boldsymbol{Y}_{j.}^3 \mid \boldsymbol{Y}_{j.} \geq 0, \boldsymbol{Y}_{k.} \geq 0\right)\right) \\ \scriptstyle for\ j\ and\ k=1,2,\ldots J;\ j\neq k \\ \vdots \end{bmatrix} = [0]$$

$$
\begin{bmatrix}
\vdots \\
\boldsymbol{X}_{j.}^{'}\left(\boldsymbol{Y}_{j.}\boldsymbol{Y}_{k.} - E\left(\boldsymbol{Y}_{j.}\boldsymbol{Y}_{k.} \mid \boldsymbol{Y}_{j.} \geq 0, \boldsymbol{Y}_{k.} \geq 0\right)\right) \\
\textit{for } j \textit{ and } k = 1, 2, \ldots J; j > k \\
\vdots \\
\boldsymbol{X}_{j.}^{'}\left(\boldsymbol{Y}_{j.}^{2}\boldsymbol{Y}_{k.} - E\left(\boldsymbol{Y}_{j.}^{2}\boldsymbol{Y}_{k.} \mid \boldsymbol{Y}_{j.} \geq 0, \boldsymbol{Y}_{k.} \geq 0\right)\right) \\
\textit{for } j \textit{ and } k = 1, 2, \ldots J; j > k \\
\vdots \\
\boldsymbol{X}_{j.}^{'}\left(\boldsymbol{Y}_{j.}\boldsymbol{Y}_{k.}^{2} - E\left(\boldsymbol{Y}_{j.}\boldsymbol{Y}_{k.}^{2} \mid \boldsymbol{Y}_{j.} \geq 0, \boldsymbol{Y}_{k.} \geq 0\right)\right) \\
\textit{for } j \textit{ and } k = 1, 2, \ldots J; k > j
\end{bmatrix}
$$

The sample analog of the population moment condition (3.2.1)

(3.2.2)

$$
\boldsymbol{h}_{\text{j Bivariate}}\left(\boldsymbol{y}, \boldsymbol{x}, \theta\right) =
\begin{bmatrix}
\frac{\boldsymbol{x}_{j.}^{'}}{\boldsymbol{n}_{jk}^{(6)}}\left(\boldsymbol{y}_{j.} - E\left(\boldsymbol{y}_{j.} \mid \boldsymbol{y}_{j.} \geq 0, \boldsymbol{y}_{k.} \geq 0\right)\right) \\
\textit{for } j \textit{ and } k = 1,2,\ldots J; j \neq k \\
\vdots \\
\frac{\boldsymbol{x}_{j.}^{'}}{\boldsymbol{n}_{jk}^{(7)}}\left(\boldsymbol{y}_{j.}^{2} - E\left(\boldsymbol{y}_{j.}^{2} \mid \boldsymbol{y}_{j.} \geq 0, \boldsymbol{y}_{k.} \geq 0\right)\right) \\
\textit{for } j \textit{ and } k = 1,2,\ldots J; j \neq k \\
\vdots \\
\frac{\boldsymbol{x}_{j.}^{'}}{\boldsymbol{n}_{jk}^{(8)}}\left(\boldsymbol{y}_{j.}^{3} - E\left(\boldsymbol{y}_{j.}^{3} \mid \boldsymbol{y}_{j.} \geq 0, \boldsymbol{y}_{k.} \geq 0\right)\right) \\
\textit{for } j \textit{ and } k = 1,2,\ldots J; j \neq k \\
\vdots \\
\frac{\boldsymbol{x}_{j.}^{'}}{\boldsymbol{n}_{jk}^{(9)}}\left(\boldsymbol{y}_{j.}\boldsymbol{y}_{k.} - E\left(\boldsymbol{y}_{j.}\boldsymbol{y}_{k.} \mid \boldsymbol{y}_{j.} \geq 0, \boldsymbol{y}_{k.} \geq 0\right)\right) \\
\textit{for } j \textit{ and } k = 1, 2, \ldots J; j > k \\
\vdots
\end{bmatrix}
= \left[0\right]
$$

$$\begin{bmatrix} \vdots \\ \frac{\boldsymbol{x}_{j.}^{'}}{\overset{(10)}{n}_{jk}}\left(\boldsymbol{y}_{j.}^{2}\boldsymbol{y}_{k.}-E\left(\boldsymbol{y}_{j.}^{2}\boldsymbol{y}_{k.}\mid\boldsymbol{y}_{j.}\geq 0,\boldsymbol{y}_{k.}\geq 0\right)\right) \\ \text{for } j \text{ and } k=1,2,\ldots J;\, j>k \\ \vdots \\ \frac{\boldsymbol{x}_{j.}^{'}}{\overset{(11)}{n}_{jk}}\left(\boldsymbol{y}_{j.}\boldsymbol{y}_{k.}^{2}-E\left(\boldsymbol{y}_{j.}\boldsymbol{y}_{k.}^{2}\mid\boldsymbol{y}_{j.}\geq 0,\boldsymbol{y}_{k.}\geq 0\right)\right) \\ \text{for } j \text{ and } k=1,2,\ldots J;\, k>j \end{bmatrix}$$

where, $\overset{(i)}{n}_{jk}$ denotes the number of sample observations involved in the i^{th} set of moment conditions for the $(j,k)^{th}$ Choice pair. The number of equations in the vector $\boldsymbol{h}_{\text{Bivariate}}(Y, X, \theta)$, $\left(\frac{9J(J-1)K}{2}\times 1\right)$, is greater than the number of unknown parameters in the model $\left(\frac{J(J-1)}{2}+K\right)$. Consequently, there is no unique parameter vector (θ) that solves the sample moment conditions via the ordinary method of moments approach. We will deal with such issues in section 4 ahead.

4. GMM Estimation of the MVT System

In this section the GMM approach is applied to the MVT system by utilizing the general marginal and bivariate moment relations that were derived in Sections 3. In section 3.1 we defined a $(5K\times 1)$ vector of marginal moment conditions, and in section 3.2 we derived a $\left(\frac{9J(J-1)K}{2}\times 1\right)$ vector of the bivariate moment conditions. Then a cumulative set of moment conditions, including both marginal and bivariate moments, can be specified as:

(4.1) $$E\left(\boldsymbol{H}(\boldsymbol{Y},\boldsymbol{X},\theta)\right)=E\begin{bmatrix}\boldsymbol{h}_{\text{jMarginal}}(\boldsymbol{Y},\boldsymbol{X},\theta)\\ \boldsymbol{h}_{\text{jbivariate}}(\boldsymbol{Y},\boldsymbol{X},\theta)\end{bmatrix}=0$$

The sample estimating moment analog is $\boldsymbol{H}(\boldsymbol{y},\boldsymbol{x},\theta)=[0]$, which has dimension $\left(5KJ+\frac{9J(J-1)K}{2}\right)\times 1$ while the parameter vector $\theta=(\beta,\Sigma)$ contains $\left(\frac{J(J-1)}{2}+K\right)$ unique elements, where β *is* $(K\times 1)$ and $\Sigma=(J\times J)$.

The number of estimating functions in the preceding specification is greater than the number of unknown parameters θ, so the set of estimating equations is over-determined for estimating θ. Given $\boldsymbol{H}(\boldsymbol{y},\boldsymbol{x},\theta)=[0]$, under the GMM approach the parameter vector is chosen for which the sample moment conditions are as close to the zero vector as possible. To solve for the estimation problem we use the following weighted Euclidean distance as a measure of closeness:

(4.2) $$\min_{\theta}\left[Q(\boldsymbol{y},\boldsymbol{x},\theta)\right]=\min_{\theta}\left[\boldsymbol{H}(\boldsymbol{y},\boldsymbol{x},\theta)'\boldsymbol{W}\,\boldsymbol{H}(\boldsymbol{y},\boldsymbol{x},\theta)\right]$$

where $\boldsymbol{W}$ is a conformable positive definite symmetric weight matrix. Another way to view the GMM approach to the problem of solving over-determined sets of sample moment conditions is through the necessary condition for minimizing (4.2) by

(4.3) $$\frac{\partial Q(\boldsymbol{y},\boldsymbol{x},\theta)}{\partial\theta}=2\left[\frac{\partial\boldsymbol{H}(\boldsymbol{y},\boldsymbol{x},\theta)}{\partial\theta}\right]'\boldsymbol{W}\,\boldsymbol{H}(\boldsymbol{y},\boldsymbol{x},\theta)=0.$$

The condition (4.3) indicates that the problem of the equation system being over-determined is overcome by forming a $\left(\frac{J(J-1)}{2}+K\right)$ linear combination of the moment conditions based on the matrix $\left[\left(\frac{\partial\boldsymbol{H}}{\partial\theta}\right)'\boldsymbol{W}\right]$ to project the moment conditions to a

$\left(\frac{J(J-1)}{2}+K\right)$-dimensional space, in effect resulting in the same number of equations as unknowns.

One major difficulty in implementing the GMM estimator is the choice of the weighting matrix ***W***, which can affect the relative efficiency of the estimator. The choice of ***W*** that results in the asymptotically most efficient estimator within the class of GMM estimators is the inverse of the covariance matrix $\text{cov}[\boldsymbol{H}(\boldsymbol{y},\boldsymbol{x},\theta)]=(E(\boldsymbol{HH}'))^{-1}=\boldsymbol{W}^*$ (Hansen, 1982; Andrews, 1999). Optimality in the current context refers to choosing a *W* matrix in the definition of the GMM estimator

(4.4) $$\hat{\theta}_{GMM}(\boldsymbol{W})=\arg\min_{\theta}\,[\boldsymbol{H}(\boldsymbol{y},\boldsymbol{x},\theta)'\boldsymbol{W}\,\boldsymbol{H}(\boldsymbol{y},\boldsymbol{x},\theta)]$$

such that $\hat{\theta}_{GMM}(\boldsymbol{W})$ has the smallest asymptotic covariance matrix. Because the optimal weight matrix implied by $(E(\boldsymbol{HH}'))^{-1}=\boldsymbol{W}^*$ is generally unknown, and thus $\hat{\theta}_{GMM}(\boldsymbol{W})$ is not operational, a consistent estimator, $\hat{\boldsymbol{W}}_n$ of $\boldsymbol{W}^*$ is used. In practice, this is obtained by setting ***W*** = ***I*** and calculating $\hat{\boldsymbol{\theta}}(\boldsymbol{I})$ in (4.4) and in the process calculating the covariance matrix of ***H***. In the second step, the sample estimator of the optimal weighting matrix $\hat{\boldsymbol{W}}_n$ is substituted into (4.4) leading to the estimated optimal GMM defined by $\hat{\theta}_{GMM}(\hat{\boldsymbol{W}}_n)$. The estimated optimal GMM estimator will be consistent, asymptotically normal and asymptotically efficient estimator in the sense of making asymptotically efficient use of the given moment information used in estimation.

The above estimation procedure can be summarized as follows:

1. Use all conditional and unconditional (first and second) marginal moments along with the binary moments for each equation.

2. Use all possible combinations of the bivariate first, second and third order moments, including cross moments.
3. Define the sample estimating equations analogs to the population moments.
4. Use the GMM method with an estimated optimal weight matrix to estimate the model parameters.

An important question in seeking the minimum of the quadratic form in moment conditions that defines the GMM objective is how to provide starting values for the minimization algorithm used. We suggest using the univariate and bivariate Tobit estimator. In the first step, given that the MVT system has *J* equations, one can estimate *J*-univariate Tobit estimates of each equation marginally. Doing this we get consistent estimates of $\hat{\beta}_j's$ and $\hat{\sigma}_j's$. In the second step we apply the bivariate Tobit on every pair of equations to estimate $\hat{\beta}_j's$ and $\left(\frac{J(J-1)}{2}\right)$ covariances. The estimates of the first and second steps are then averaged and used as starting values for the GMM approach. We have found this approach to be very effective in providing starting values in the array of problems we have applied the procedure to.

5. Monte Carlo Experiments and Results

In this section we perform Monte Carlo experiments to compare the parameter estimates of the GMM to the SML using the ORDGHK simulator (Hasan and Mittelhammer, 2001). We start the first experiment with a Two-choice model and compare our GMM approach to a SML under the assumption of normality as well a number of skewed distributions that are described in greater detail below. Then we test moment equations validity under these different distributions. Finally, in the second experiment we extended the MVT system to a Five-choice model and compare the GMM to the SML under normality.

All calculations were obtained using the GAUSS 8.0 matrix programming language. All the simulation experiments were performed on the HP Pentium 4 machine (AMD Athlon ™ 64×2 Dual core processor 5600 +2.8 GHz) and 4 GHz of memory.

For the minimization of equation (4.2), we used the Nelder-Meade polytope direct-search method of optimization which only requires a continuous objective function, and evaluations of that function to achieve optimization. It is robust to non-differentiability and useful for functions whose derivatives cannot be calculated or approximated easily, or at all. A convergence criterion of 0.00001 was used for the difference between the maximum and the minimum objective function associated with the vertices of the Nelder-Meade simplex.

5.1 Sampling Experiments

We begin generating outcomes of the latent variable in equation (2.1) by sampling the X from a uniform distribution having support on the interval (-5, 5). The betas and the correlation matrix, $R,$ for the two-choice model and the five-choice model are represented in equation 5.1.1 and 5.1.2 respectively.

$$(5.1.1) \qquad \begin{bmatrix} Y_1^* \\ Y_2^* \end{bmatrix} = \begin{bmatrix} X_1\beta_1 \\ X_2\beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix},$$

where $\beta_1 = \begin{bmatrix} \beta_{11} \\ \beta_{12} \end{bmatrix} = \begin{bmatrix} .1 \\ .2 \end{bmatrix}, \beta_2 = \begin{bmatrix} \beta_{21} \\ \beta_{22} \end{bmatrix} = \begin{bmatrix} .3 \\ .4 \end{bmatrix},$ and $\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, R = \begin{bmatrix} 1 & .2 \\ .2 & 1 \end{bmatrix}\right), \begin{bmatrix} \sigma_{11} \\ \sigma_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$$(5.1.2) \qquad \begin{bmatrix} Y_1^* \\ Y_2^* \\ Y_3^* \\ Y_4^* \\ Y_5^* \end{bmatrix} = \begin{bmatrix} X_1\beta_1 \\ X_2\beta_2 \\ X_3\beta_3 \\ X_4\beta_4 \\ X_5\beta_5 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix},$$

where $\beta_1 = \begin{bmatrix} \beta_{11} \\ \beta_{12} \end{bmatrix} = \begin{bmatrix} .1 \\ .2 \end{bmatrix}, \beta_2 = \begin{bmatrix} \beta_{21} \\ \beta_{22} \end{bmatrix} = \begin{bmatrix} .3 \\ .4 \end{bmatrix}, \beta_3 = \begin{bmatrix} \beta_{31} \\ \beta_{32} \end{bmatrix} = \begin{bmatrix} .5 \\ .6 \end{bmatrix}, \beta_4 = \begin{bmatrix} \beta_{41} \\ \beta_{42} \end{bmatrix} = \begin{bmatrix} .7 \\ .8 \end{bmatrix}, \beta_5 = \begin{bmatrix} \beta_{51} \\ \beta_{52} \end{bmatrix} = \begin{bmatrix} .5 \\ .2 \end{bmatrix},$

and $\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, R = \begin{bmatrix} 1 & .1 & .2 & .3 & .4 \\ .1 & 1 & .6 & .4 & .25 \\ .2 & .6 & 1 & .11 & .12 \\ .3 & .4 & .11 & 1 & .15 \\ .4 & .25 & .12 & .15 & 1 \end{bmatrix} \right), \begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{44} \\ \sigma_{55} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 3 \\ 4 \end{bmatrix}.$

The correlation matrix was chosen so as to produce a fair degree of variability in the error term outcomes as well as to produce a variety of covariances between latent variables. Given all the information above we can generate the latent variables and then generate the data for the MVT model as shown in equation (2.2).

5.2 Two-Choice Tobit Model

Under the assumption of normality, Table 1 in Appendix 1 summarizes the comparison between the GMM and SML in the Two-choice model. With a sample size of 1,000 observations the SML approach, as expected, outperforms the GMM approach in term of the root means square errors[3] (RMSE). The RMSEs of the SML is slightly smaller in magnitude compared to the RMSEs of the GMM. These results are not surprising because of the asymptotic efficiency of the ML estimator, which achieves the Cramér-Rao lower bound when the assumption of normality is correct.

The Percentage probabilities of censored observations in each equation of the Two-choice model show a fair degree of censoring. In equation one (containing β_{11} and β_{12} parameters) the percentage probability of censored observations is 23%, while in equation two (containing β_{21} and β_{22} parameters) it is 20%. Reporting the percentage probability of censoring is very important as we will show later in the paper how high censoring can affect our results. To check if any of the estimation methods (GMM or SML) is an improvement over the other, two tests were performed to see if there were any statistical differences in the RMSEs and the sample means of the two approaches. The null hypothesis for the RMSEs is H_0: mean ($RMSE_{GMM}$) – mean ($RMSE_{SML}$) = 0 and the null hypothesis for the sample means is H_0: mean (β_{GMM}) – mean (β_{SML}) = 0. Simple T-tests are performed according to the following formulas:

$$T_{RMSE} = \frac{\text{Mean}(RMSE_{GMM})\text{-Mean}(RMSE_{SML})}{\sqrt{\frac{s^2_{GMM}}{n_1} + \frac{s^2_{SML}}{n_2}}} \quad and \quad T_{Sample\,Means} = \frac{\text{Mean}(\beta_{GMM})\text{-Mean}(\beta_{SML})}{\sqrt{\frac{s^2_{GMM}}{n_1} + \frac{s^2_{SML}}{n_2}}}.$$

---

[3] RMSE= $\sqrt{\frac{\sum_{i=1}^{n}\left(\theta_i - \hat{\theta}_i\right)^2}{n}}$, where θ_i is the vector of the true parameters and $\hat{\theta}_i$ is vector of estimated parameters

The results (Table 1) show that with the assumption of normality we fail to reject the null hypothesis[4] for the RMSEs and the sample means. Given that there are no differences between the RMSEs of the GMM and the SML estimators, the results suggest that the GMM estimator is nearly asymptotically efficient as the SML estimator in reasonably higher sample size. Beside this interesting result, the GMM has two advantages over the SML which are tractability and speed of convergence. The convergence time shows that GMM converges faster than the SML (GMM converges at 140 minuets while SML converges at 310 minutes).

Again with the assumption of normality, Table 2 in Appendix 2 shows that decreasing the sample size by one tenth of the observations (*n=100*) causes the RMSEs in both approach to increase (compared to larger sample size, Table 1), suggesting that sample size have direct effect on the RMSEs. The most interesting finding is that the magnitudes of the RMSEs in the GMM approach are smaller than the magnitudes of the RMSEs in the SML approach, suggesting that the GMM approach is an improvement over the SML approach. The T-tests for the RMSEs and the sample means confirm our conclusion, as the T-test of RMSEs indicates that we reject the null hypothesis indicating the means of the RMSEs are different, while the T-test for the sample means indicates that we fail to reject the null hypothesis[5]. Again the percentage probability of censoring for each equation has fair degree of censoring and The GMM approach has faster convergence.

[4] Large sample size (n=1,000), at 95% significance level the critical value is 1.96
[5] Small sample size (n=100), at 95% significance level the critical value is 1.98

As a conclusion, under the assumption of normality and under fair degree of censoring, the above results suggest that the GMM have three advantages over SML in the Two-Choice model:

1. GMM in large sample size is nearly asymptotically efficient as the SML.
2. The GMM is computationally tractable and has faster convergence in large and small sample sizes.
3. The GMM is more efficient in small sample size compared to the SML.

Further investigation was done, as we compared both approaches (GMM and SML) under the assumptions of different skewed distributions. We assumed three Types of Gamma distributions taking into consideration the level of skewness. Figure 1 in Appendix 1 shows the graphical representations of these distributions (where each was scaled to have a mean of zero and variance of one).

With a sample size of 1,000 observations, Tables 3,4, and 5 in Appendix 1 represent the comparison results between the GMM and the SML with the assumptions of Gamma (1,1), Gamma (3,2), and Gamma (4,3) distributions, respectively. The results tell us that the GMM is an improvement over the SML, as the RMSEs magnitudes of the GMM are smaller compared to those of SML. Also the T-test of the RMSEs indicates that we reject the null hypothesis, suggesting that there is a statistical difference between the means of the RMSEs of the two approaches. The sample mean test fail to reject the null hypothesis suggesting that there are no statistical differences in the sample means. This finding suggests that the ML estimator become inefficient as it deviates from the normality assumption, and suggests that the GMM estimator can be more robust towards these different distributions given that it does not use all of the information implied by the normal distribution assumption.

Regarding the number of moment equations to use in the GMM approach, asymptotic efficiency comparisons suggest more is better than less. Unfortunately, it is not necessarily the case that more moments are better than less in terms of finite sampling properties. There is a tendency for finite sample bias to increase as more estimating equations are used for estimation. To validate the moment equations in the GMM approach, we construct a Chi-square test that is based on the asymptotic normal distribution of the estimating equations. In particular, if $\Psi^{-\frac{1}{2}}h(Y,X,\theta)\xrightarrow{d}N(0,1)$, and $\Psi = \text{cov}[h(Y,X,\theta)] = E[h(Y,X,\theta)h(Y,X,\theta)^{'}]$ then:

$$Q_{GMM(TEST)} = h(Y,X,\hat{\theta})'\left[E\left[h\left(Y,X,\theta\right)h\left(Y,X,\theta\right)'\right]\right]^{-1} h(Y,X,\theta)\xrightarrow{d}\chi^2(m) \text{ under } H_0$$

where m is the number estimating equations. If $Q_{GMM(TEST)} \geq \chi^2(m)$, we will reject the validity of the estimating moments. Table 6, in Appendix 1, shows the GMM moments validity test for the Two- Choice model under the assumptions of normality and Gamma distributions. The degree of freedom for the Two- Choice Tobit model is 38 with a critical value of 53.10 at 95% significant level. The results suggest that we fail to reject the null hypothesis. Failing to reject the null hypothesis can signal that there is no violation of the moments conditions added.

5.3 Five-Choice Tobit Model

Assuming that the error terms follow a normal distribution (standard normal) and assuming that sample size equal to 1,000 observations, the two-choice MVT model is extended to a Five-choice model as in equation (5.1.2). Monte Carlo experiment with 1,000 repetitions is designed to estimate and compare the Five-choice model using both GMM and SML approaches. Table 7 in Appendix 1 shows the comparison between the two estimation approaches. The results show that the RMSEs of the GMM have smaller magnitude than those of the SML. The T-test of the RMSEs shows that we reject the null hypothesis, while the T-test of the sample means suggest the opposite (fail to reject the null hypothesis), meaning that GMM is an improvement over SML in higher dimensions. These results are surprising since the GMM uses subset of moments conditions while the SML is considered asymptotically efficient when the assumption of normality is correct. We speculate that the reason for these results lies in the percentage censoring probabilities for each equation in MVT systems. In the Two-choice model the percentage censoring probabilities vary from 20% to 23%, while in Five-choice model they varies from 42% to 53%. In the Two-choice model with low censoring, the RMSEs of the SML were smaller, meaning that the SML outperforms the GMM approach. In the Five-choice model with high censoring, the RMSEs of the GMM are smaller, meaning that the GMM outperforms the SML approach. This finding could be related to the likelihood function in equation (2.4). As the likelihood function is composed of two parts the continuous part (L_c), and the discontinuous ($L_{d|c}$) part. Simulating the discontinuous likelihood part and imposing high censoring could be a reasonable explanation for these results. Beside efficiency, another advantage of the GMM over the SML is tractability and speed of convergence, the GMM approach converges in 10 hours while the SML approach converges in 48 hours.

6. Conclusions

In this paper we utilize the GMM approach to estimate systems of censored equations. In the GMM approach we use the marginal and all possible bivariate moment conditions (up to the third order) to help increase the efficiency of the GMM estimator. The GMM estimator is compared with an efficient full information ML estimator obtained by simulating the likelihood function by ORDGHK algorithm. Monte Carlo experiments reveal that the GMM estimator under the assumption of normality and with small sample size is more efficient than that of SML. Furthermore, the GMM estimator can be more robust toward different skewed distributions given that it does not use all of the information implied by the normal distribution assumption. In the presence of high censoring, high dimensional systems, large sample size, and with the assumption of normality the GMM estimator is more efficient than SML.

The GMM parameter estimates from a variety of Monte Carlo experiments appear quit accurate, and illustrate the potential of the proposed GMM estimation method. Estimates obtained by this procedure are consistent, asymptotically normally distributed and appear overall near-asymptotically efficient, making asymptotic efficient use of the rather high order of moment information used. Besides providing a tractable way of estimating systems of MVT models, the GMM demonstrates faster convergence in simulations.

Another advantage of the GMM over the SML, in practice there is often insufficient information to specify the parametric form of the likelihood function underlying the data sampling process. Given this situation, the GMM method in this paper can be applied to non-normally distributed data sampling processes by defining the analogs of the moment conditions used here. Furthermore, even using moment conditions based on the normal distribution, the fact that not all of the information of the normal distribution is utilized in the specification can afford the GMM approach a robustness advantage when the distribution of the error term is not normal. The principle contribution of this paper is to introduce a method of estimating a system of censored demand equations that is computationally tractable, consistent, near overall asymptotically efficient, and asymptotically efficient relative to the moment conditions used.

References

Amemiya, T., (1973) "Regression analysis when the dependent variable is truncated normal" Econometrica, 41, 997–1016

Amemiya, T., (1974) "Multivariate regression and simultaneous equation models when the dependent variable is truncated normal" Econometrica, 42, 999–1013

Amemiya, T., (1979a) "The estimation of simultaneous equation Tobit model," International Economic Review, 20, 169–181

Amemiya, T., (1979b) "Tobit models: a survey," Journal of Econometrics, 24, 3–61

Andrews, D.W.K., (1999) "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," Econometrica, 67, 543–564

Blundell, R.W. and Meghir C., (1987) "Bivariate Alternatives to the Univariate Tobit Model," Journal of Econometrics, 34, 179–200

Chiang, J and Lee L., (1992) Discrete /continuous modes of consumer demand with binding non-negativity constrains," Journal of econometrics, 23, 59-80

Chung, C.E, and Goldberger, A.S., (1984) "Proportional Projections in Limited Dependent Variable Models," Econometrica , 52, 531–34

Chavas, J.P., and Perali, F., (2000) "Estimation of Censored Demand Equations from Large Cross-Sectional Data," Amer. J. Agr. Econ. 82, 1022–1037

Deaton, A., (1986) "Demand Analysis, Handbook of Econometrics," vol. 3, chap. 30, pp. 1769–1839

Fahs, R., and Mittelhammer, R., (2001) "Semiparametric Estimation and Inference in a System of Censored Demand Equation," Working paper, Washington State University

Greene, W.H., (1981) "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model," Econometrica, 49, 505–13

Hasan, M., and Mittelhammer, R., (2001) "Simulated maximum Likelihood estimation for multivariate limited dependent variable models using improved smooth simulators with application to estimating demand systems," Working paper, Washington State University

Hajivassiliou, V. and Ruud, P., (1994) "Classical estimation methods for LDV models using simulation," In Handbook of Econometrics, Vol. 4, Chap. 40, North Holland, Amsterdam

Hajivassiliou, V., (1993b) "Simulation estimation methods for limited dependent variables," Handbook of statistics, vol.2, 519-543

Hajivassiliou, V.A., D. McFadden, and Ruud, P., (1996) "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results," J. Econometrics, 72, 85–134

Hansen, L. and Singleton, (1982) "Generalized Instrumental Variable Estimators of Nonlinear Rational Expectations Models," Econometrica, 50, 1269–1286

Heckman, J. J., (1976*b*) "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and A Simple Estimator for Such Models," Annals Econ. and Soc. Meas., 5, 475–92

Lee, L., (1993) "Multivariate Tobit models in econometrics, Handbook of statistics," vol. 2, 145–173

Lee, L. and Pitt, M., (1986) "Microeconometric demand systems with binding non-negativity constrains: the dual approach," Econometrica, 54, 1237–1242

Lee, L. and Pitt, M., (1987) "Microeconometric models of rationing, imperfect market and non-negativity constrains," Journal of Econometrics, 36, 89–110

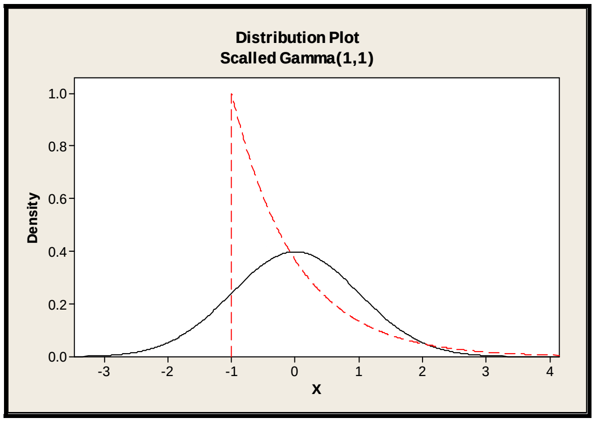
Maddala, G.S., (1977) "Identification and estimation problems in limited Dependent Variable Models," In A.S.Blinder and P.Friedman (eds.), Natural Resources, Uncertainty and General Equilibirum Systems: Essays in Memory of Rafael Lusky, pp.219-39. New York: Academic.

Pudney, S., (1989) "Modeling Individual Choice," Oxford: UK Basil Blackwell

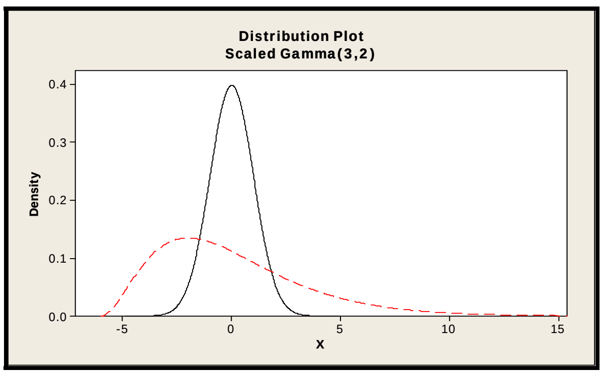
Tallis, G. M., (1961) "The Moment Generating Function of the Truncated Multi-Normal Distribution," Journal of the Royal Statistical Society (Series B), 23, 223–229

Appendix 1

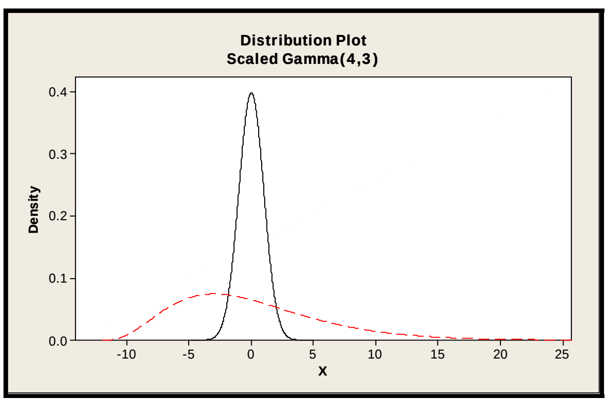
Figure 1: Graphical representations of the scaled Gamma distributions



Dashed line: Scaled Gamma (1,1) distribution
Solid line : N(0,1) distribution



Dashed line: Scaled Gamma (3,2) distribution
Solid line : N(0,1) distribution



Dashed line: Scaled Gamma (4,3) distribution
Solid line : N(0,1) distribution

Table 1: Comparison between GMM and SML, Two-choice model, large sample size and under the assumption of normality

Sample Size=1,000, Repetitions=1,000								
Parameters	**True Values**	**GMM (MEAN)**	**GMM (RMSE)**	**SML (MEAN)**	**SML (RMSE)**	**T-Test Values (RMSE)**	**T-Test Values (MEAN)**	**% Probability of Censored Observation in each Equation**
β_{11}	0.1000	0.1054	0.0475	0.1032	0.0462	0.7888	0.1080	23%
β_{12}	0.2000	0.1998	0.0310	0.1991	0.0299	0.7280	0.0409	
β_{21}	0.3000	0.2940	0.0742	0.2957	0.0731	0.9775	0.0682	20%
β_{22}	0.4000	0.3996	0.0482	0.4008	0.0442	1.1747	0.0605	
ρ	0.2000	0.2088	0.0088	0.2013	0.0083	0.8375	0.0523	
		GMM		**SML**				
Time of Convergence (Minutes)		140		310				

Table 2: Comparison between GMM and SML, Two-choice model, small sample size and under the assumption of normality

Sample Size=100, Repetitions=1,000

Parameters	True Values	GMM (MEAN)	GMM (RMSE)	SML (MEAN)	SML (RMSE)	T-Test Values (RMSE)	T-Test Values (MEAN)	% Probability of Censored Observation in each Equation
β_{11}	0.1000	0.0922	0.1016	0.0952	0.1670	1.9930	0.0779	23%
β_{12}	0.2000	0.2012	0.0629	0.1974	0.1119	2.0669	0.1293	
β_{21}	0.3000	0.2948	0.1767	0.3042	0.2387	1.9942	0.4245	20%
β_{22}	0.4000	0.4125	0.0999	0.4139	0.1499	2.0206	0.0389	
ρ	0.2000	0.2094	0.0130	0.1870	0.0294	1.9984	0.1252	

	GMM	SML
Time of Convergence (Minutes)	97	210

Table 3: Comparison between GMM and SML, Two-choice model, large sample size and under the assumption of Gamma (1,1)

Sample Size=1000, Repetitions=1000								
Parameters	**True Values**	**GMM (MEAN)**	**GMM (RMSE)**	**SML (MEAN)**	**SML (RMSE)**	**T-Test Values (RMSE)**	**T-Test Values (MEAN)**	**% Probability of Censored Observation in each Equation**
β_{11}	0.1000	0.0867	0.0156	0.1238	0.0447	2.0772	0.9876	56.6%
β_{12}	0.2000	0.2117	0.0182	0.2206	0.0825	3.1256	0.4657	
β_{21}	0.3000	0.2768	0.0268	0.2555	0.3697	2.8754	0.5862	49.5%
β_{22}	0.4000	0.447	0.0109	0.5011	0.0161	1.9824	0.6024	
ρ	0.2000	0.1983	0.0102	0.2102	0.0417	2.2332	0.6213	
		GMM		**SML**				
Time of Convergence (Minutes)		196		412				

Table 4: Comparison between GMM and SML, Two-choice model, large sample size and under the assumption of Gamma (3, 2)

Sample Size=1000, Repetitions=1000								
Parameters	**True Values**	**GMM (MEAN)**	**GMM (RMSE)**	**SML (MEAN)**	**SML (RMSE)**	**T-Test Values (RMSE)**	**T-Test Values (MEAN)**	**% Probability of Censored Observation in each Equation**
β_{11}	0.1000	0.0899	0.0190	0.0845	0.0237	2.3241	0.2321	50.1%
β_{12}	0.2000	0.2070	0.0202	0.2167	0.0225	1.9846	0.3452	
β_{21}	0.3000	0.2204	0.0179	0.2349	0.0398	2.0231	1.2321	46.6%
β_{22}	0.4000	0.3928	0.0240	0.4166	0.0292	1.9975	0.7652	
ρ	0.2000	0.1932	0.0068	0.1677	0.0323	2.4345	0.6534	
		GMM		**SML**				
Time of Convergence (Minutes)		175		389				

Table 5: Comparison between GMM and SML, Two-choice model, large sample size and under the assumption of Gamma (4, 3)

Sample Size=1000, Repetitions=1000								
Parameters	**True Values**	**GMM (MEAN)**	**GMM (RMSE)**	**SML (MEAN)**	**SML (RMSE)**	**T-Test Values (RMSE)**	**T-Test Values (MEAN)**	**% Probability of Censored Observation in each Equation**
β_{11}	0.1000	0.0394	0.0149	0.0404	0.0169	2.8940	0.0740	49.4%
β_{12}	0.2000	0.2043	0.0155	0.2129	0.0162	2.7452	0.7766	
β_{21}	0.3000	0.3199	0.0237	0.2747	0.0369	2.3180	1.2000	46.3%
β_{22}	0.4000	0.3887	0.0105	0.4124	0.0396	2.8918	1.6028	
ρ	0.2000	0.2061	0.0083	0.1917	0.0123	1.9986	0.6243	
		GMM		**SML**				
Time of Convergence (Minutes)		191		394				

Table 6: Chi-Square tests for the GMM Two-choice Tobit model with the assumption of different distributions on the error terms; the degree of freedom is 38 and the critical value is 53.10 at 95% significance level

GMM		
Distributions	**Sample size**	**Moment validity Chi-square test**
Normal (0,1)	100	15.2
Normal (0,1)	1000	13.3
Gamma(1,1)	1000	26.5
Gamma(3,2)	1000	20.4
Gamma(4,3)	1000	22.6

Table 7: Comparison between GMM and SML, Five-choice model, large sample size and under the assumption of normality

Sample Size=1000, Repetitions=1000								
Parameter	**True Value**	**GMM (Mean)**	**GMM (RMSE)**	**SML (RMSE)**	**SML (RMSE)**	**T-Test Values (RMSE)**	**T-Test Values (Mean)**	**% Probability of Censored Observations in each equation**
β_{11}	0.1000	0.1022	0.0224	0.1086	0.0784	2.1354	0.6324	48%
β_{12}	0.2000	0.2003	0.0130	0.2043	0.0253	2.0323	0.5324	
β_{21}	0.3000	0.3032	0.0326	0.3134	0.0472	1.9981	0.5623	53%
β_{22}	0.4000	0.4010	0.0142	0.4080	0.0568	2.3213	0.4562	
β_{31}	0.5000	0.5021	0.0412	0.5141	0.0819	2.3421	0.5214	42%
β_{32}	0.6000	0.6011	0.0241	0.6094	0.0964	2.4213	0.5642	
β_{41}	0.7000	0.7021	0.0324	0.6982	0.0985	2.2243	0.4896	51.5%
β_{42}	0.8000	0.8031	0.0143	0.8130	0.0432	2.5342	0.3213	
β_{51}	0.9000	0.9100	0.0241	0.9181	0.0452	2.4325	0.6354	47%
β_{52}	0.1100	0.1101	0.0128	0.1193	0.0845	2.3351	0.4532	
ρ_{12}	0.1000	0.1014	0.0141	0.1207	0.0549	2.1352	0.2232	
ρ_{13}	0.2000	0.2013	0.0136	0.2131	0.0562	2.1213	0.2865	
ρ_{14}	0.3000	0.3009	0.0245	0.3090	0.0568	1.9875	0.3248	
ρ_{15}	0.4000	0.4015	0.0314	0.3970	0.0454	2.0562	0.4356	
ρ_{23}	0.6000	0.4022	0.0233	0.4103	0.0426	1.9895	0.3521	
ρ_{24}	0.7000	0.7013	0.0138	0.7087	0.0542	2.1542	0.3641	
ρ_{25}	0.2500	0.2500	0.0122	0.2530	0.0565	2.1438	0.2013	
ρ_{34}	0.1100	0.1100	0.0130	0.1121	0.0423	2.0532	0.2831	
ρ_{35}	0.1200	0.1190	0.0521	0.1300	0.0849	2.0462	0.3261	
ρ_{45}	0.1500	0.1501	0.0233	0.1498	0.0989	2.1024	0.3352	
		GMM		**SML**				
Time of Convergence (Hours)		10		48				

Appendix 2

Suppose $(y_1, y_2) \sim N\left(\begin{pmatrix} X_1\beta_1 \\ X_2\beta_2 \end{pmatrix}, \Sigma\right)$, and then consider that the random variables are truncated as $y_1 \geq 0,\ y_2 \geq 0$. We can standardize the variables so that we are dealing with correlated standard normal random variables.

$$y_1 \geq 0 \iff z_1 = \frac{y_1 - x_1\beta_1}{\sigma_1} \geq \frac{-x_1\beta_1}{\sigma_1} = a_1$$
$$y_2 \geq 0 \iff z_2 = \frac{y_2 - x_2\beta_2}{\sigma_2} \geq \frac{-x_2\beta_2}{\sigma_2} = a_2$$

All the derivations below can be obtained from the author upon request.

$$E(z_1 \mid z_1 \geq a_1, z_2 \geq a_2) = \int_{a_1}^{\infty} z_1 \int_{a_2}^{\infty} \frac{1}{(2\pi)(1-\rho^2)^{1/2}} e^{-1/2 z' R^{-1} z} dz_2 dz_1$$
$$where\, R = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad |R| = 1-\rho^2,$$
$$and\ R^{-1} = \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} . \left(\frac{1}{1-\rho^2}\right) = \begin{pmatrix} (1-\rho^2)^{-1} & \frac{-\rho}{1-\rho^2} \\ \frac{-\rho}{1-\rho^2} & (1-\rho^2)^{-1} \end{pmatrix}$$

$$E(z_1 \mid z_1 \geq a_1, z_2 \geq a_2) = \int_{a_1}^{\infty} z_1 \int_{a_2}^{\infty} \frac{1}{(2\pi)(1-\rho^2)^{1/2}} \exp\left(\frac{-1}{2(1-\rho^2)}\left[z_1^2 + z_2^2 - 2\rho z_1 z_2\right]\right) dz_2 dz_1$$
$$= \frac{1}{F^*(a_1, a_2)}\left[\phi(a_1)\Phi\left(\frac{a_2 - \rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho\phi(a_2)\Phi\left(\frac{a_1 - \rho a_2}{(1-\rho^2)^{1/2}}\right)\right]$$

where $F^*(a_1, a_2) = \int_{a_1}^{\infty}\int_{a_2}^{\infty} BISN(z;\rho)dz$ represents integral of a bivariate standard normal random variable with correlation ρ, $\phi\, and\, \Phi$ represents the probability density function (PDF) and the cumulative density function (CDF) respectively.

$$E(z_1^2 \mid z_1 \geq a_1, z_2 \geq a_2) = 1 + \frac{1}{F^*(a_1,a_2)} \begin{bmatrix} a_1\phi(a_1)\Phi\left(\dfrac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho^2 a_2 \phi(a_2)\Phi\left(\dfrac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \\ +\rho \dfrac{(1-\rho^2)}{\sqrt{(1-\rho^2)}} \phi(a_2)\Phi\left(\dfrac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \end{bmatrix}$$

$$E(z_1^3 \mid z_1 \geq a_1, z_2 \geq a_2) = \frac{1}{F^*(a_1,a_2)} \begin{bmatrix} 3 \times \left[\phi(a_1)\Phi\left(\dfrac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho\phi(a_2)\Phi\left(\dfrac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +\rho^3 a_2^2 \phi(a_2)\Phi\left(\dfrac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) + 2\rho^2 a_2 \phi(a_2)\Phi\left(\dfrac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\left(\dfrac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \\ +\rho\phi(a_2)\Phi\left(\dfrac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\left(1-\rho^2\right) + a_1^2\phi(a_1)\Phi\left(\dfrac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) \end{bmatrix}$$

$$E(z_1 z_2 \mid z_1 \geq a_1, z_2 \geq a_2) = \rho + \frac{1}{F^*(a_1,a_2)} \begin{bmatrix} \rho a_1\phi(a_1)\Phi\left(\dfrac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho a_2 \phi(a_2)\Phi\left(\dfrac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \\ +\left(\dfrac{1-\rho^2}{\sqrt{1-\rho^2}}\right)\phi(a_1)\Phi\left(\dfrac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) \end{bmatrix}$$

$$E(z_1 z_2^2 \mid z_1 \geq a_1, z_2 \geq a_2) = \frac{1}{F^*(a_1,a_2)} \begin{bmatrix} \left[\phi(a_1)\Phi\left(\dfrac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho\phi(a_2)\Phi\left(\dfrac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +2\rho\left[\rho\phi(a_1)\Phi\left(\dfrac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \phi(a_2)\Phi\left(\dfrac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +\rho^2 a_1^2 \phi(a_1)\Phi\left(\dfrac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) + \rho a_1 \phi(a_1)\Phi\left(\dfrac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)\left(\dfrac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \\ +\rho a_2^2 \phi(a_2)\Phi\left(\dfrac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) + a_2 \phi(a_2)\Phi\left(\dfrac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\left(\dfrac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \end{bmatrix}$$

$$E(z_1^2 z_2 \mid z_1 \ge a_1, z_2 \ge a_2) = \frac{1}{F^*(a_1,a_2)} \begin{bmatrix} 2\rho\left[\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+\rho\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +\left[\rho\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +\rho^2 a_2^2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)+\rho a_2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \\ +\rho a_1^2\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+a_1\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \end{bmatrix}$$

Then

$$E(z_2 \mid z_1 \ge a_1, z_2 \ge a_2) = \int_{a_2}^{\infty} z_2 \int_{a_1}^{\infty} \frac{1}{(2\pi)(1-\rho^2)^{1/2}} \exp\left(\frac{-1}{2(1-\rho^2)}\left[z_1^2+z_2^2-2\rho z_1 z_2\right]\right) dz_1 dz_2$$

$$= \frac{1}{F^*(a_1,a_2)}\left[\rho\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right]$$

$$E(z_2^2 \mid z_1 \ge a_1, z_2 \ge a_2) = 1 + \frac{1}{F^*(a_1,a_2)} \begin{bmatrix} \rho^2 a_1\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+a_2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \\ +\rho\frac{(1-\rho^2)}{\sqrt{(1-\rho^2)}}\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right) \end{bmatrix}$$

$$E(z_2^3 \mid z_1 \ge a_1, z_2 \ge a_2) = \frac{1}{F^*(a_1,a_2)} \begin{bmatrix} 3\times\left[\rho\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +\rho^3 a_1^2\phi(a_1)\Phi\left(\frac{a_2-\rho a_2}{(1-\rho^2)^{1/2}}\right)+2\rho^2 a_1\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \\ +\rho\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)\left(1-\rho^2\right)+a_2^2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \end{bmatrix}$$

$$E(z_2 z_1^2 \mid z_1 \geq a_1, z_2 \geq a_2) = \frac{1}{F^*(a_1,a_2)} \begin{bmatrix} 2\rho\left[\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+\rho\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +\left[\rho\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +\rho^2 a_2^2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)+\rho a_2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \\ +\rho a_1^2\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+a_1\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \end{bmatrix}$$

$$E(z_2^2 z_1 \mid z_1 \geq a_1, z_2 \geq a_2) = \frac{1}{F^*(a_1,a_2)} \begin{bmatrix} \left[\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+\rho\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +2\rho\left[\rho\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\right] \\ +\rho^2 a_1^2\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+\rho a_1\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \\ +\rho a_2^2\phi(a_2)\Phi\left(\frac{a_1-\rho a_1}{(1-\rho^2)^{1/2}}\right)+a_2\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right)\left(\frac{1-\rho^2}{\sqrt{1-\rho^2}}\right) \end{bmatrix}$$

Then the first order moments are

$$E(y_1 \mid y_1 \geq 0, y_2 \geq 0) = \frac{\sigma_1}{F^*(a_1,a_2)} \begin{bmatrix} \phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+ \\ \rho\phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \end{bmatrix} + x_1\beta_1$$

$$E(y_2 \mid y_1 \geq 0, y_2 \geq 0) = \frac{\sigma_2}{F^*(a_1,a_2)} \begin{bmatrix} \rho\phi(a_1)\Phi\left(\frac{a_2-\rho a_1}{(1-\rho^2)^{1/2}}\right)+ \\ \phi(a_2)\Phi\left(\frac{a_1-\rho a_2}{(1-\rho^2)^{1/2}}\right) \end{bmatrix} + x_2\beta_2$$

The second moments are

$$E(y_1^2 \mid y_1 \geq 0, y_2 \geq 0) = E((\sigma_1 z_1 + x_1\beta_1)^2 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= E((\sigma_1^2 z_1^2 + 2(x_1\beta_1)\sigma_1 z_1 + (x_1\beta_1)^2) \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= \sigma_1^2 E(z_1^2 \mid z_1 \geq a_1, z_2 \geq a_2) + 2\sigma_1(x_1\beta_1)E(z_1 \mid z_1 \geq a_1, z_2 \geq a_2) + (x_1\beta_1)^2$$

and

$$E(y_2^2 \mid y_1 \geq 0, y_2 \geq 0) = E((\sigma_2 z_2 + x_2\beta_2)^2 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= E((\sigma_2^2 z_2^2 + 2\sigma_2(x_2\beta_2)z_2 + (x_2\beta_2)^2 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= \sigma_2^2 E(z_2^2 \mid z_1 \geq a_1, z_2 \geq a_2) + 2\sigma_2(x_2\beta_2)E(z_2 \mid z_1 \geq a_1, z_2 \geq a_2) + (x_2\beta_2)^2$$

The third moments are

$$E(y_1^3 \mid y_1 \geq 0, y_2 \geq 0) = E((\sigma_1 z_1 + x_1\beta_1)^3 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= E((\sigma_1^3 z_1^3 + 3(\sigma_1^2 z_1^2)(x_1\beta_1) + 3(\sigma_1 z_1)(x_1\beta_1)^2 + (x_1\beta_1)^3) \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= \sigma_1^3 E(z_1^3 \mid z_1 \geq a_1, z_2 \geq a_2) + 3\sigma_1^2(x_1\beta_1)E(z_1^2 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$+ 3\sigma_1(x_1\beta_1)^2 E(z_1 \mid z_1 \geq a_1, z_2 \geq a_2) + (x_1\beta_1)^3$$

$$E(y_2^3 \mid y_1 \geq 0, y_2 \geq 0) = E((\sigma_2 z_2 + x_2\beta_2)^3 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= E((\sigma_2^3 z_2^3 + 3(\sigma_2^2 z_2^2)(x_2\beta_2) + 3(\sigma_2 z_2)(x_2\beta_2)^2 + (x_2\beta_2)^3) \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= \sigma_2^3 E(z_2^3 \mid z_1 \geq a_1, z_2 \geq a_2) + 3\sigma_2^2(x_2\beta_2)E(z_2^2 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$+ 3\sigma_2(x_2\beta_2)^2 E(z_2 \mid z_1 \geq a_1, z_2 \geq a_2) + (x_2\beta_2)^3$$

The cross section moments are

$$E(y_1 y_2 \mid y_1 \geq 0, y_2 \geq 0) = E((\sigma_1 z_1 + (x_1\beta_1))(\sigma_2 z_2 + (x_2\beta_2)) \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= E((\sigma_1\sigma_2 z_1 z_2 + (x_2\beta_2)\sigma_1 z_1 + (x_1\beta_1)\sigma_2 z_2 + (x_1\beta_1)(x_2\beta_2)) \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= \sigma_1\sigma_2 E(z_1 z_2 \mid z_1 \geq a_1, z_2 \geq a_2) + (x_2\beta_2)\sigma_1 E(z_1 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$+ (x_1\beta_1)\sigma_2 E(z_2 \mid z_1 \geq a_1, z_2 \geq a_2) + (x_1\beta_1)(x_2\beta_2)$$

$$E(y_2y_1 \mid y_1 \geq 0, y_2 \geq 0) = E((\sigma_2 z_2 + (x_2\beta_2)(\sigma_1 z_1 + (x_1\beta_1))) \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= E((\sigma_1\sigma_2 z_2 z_1 + (x_2\beta_2)\sigma_1 z_1 + (x_1\beta_1)\sigma_2 z_2 + (x_1\beta_1)(x_2\beta_2)) \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= \sigma_1\sigma_2 E(z_2 z_1 \mid z_1 \geq a_1, z_2 \geq a_2) + (x_2\beta_2)\sigma_1 E(z_1 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$+ (x_1\beta_1)\sigma_2 E(z_2 \mid z_1 \geq a_1, z_2 \geq a_2) + (x_1\beta_1)(x_2\beta_2)$$

$$E(y_1^2 y_2 \mid y_1 \geq 0, y_2 \geq 0) = E((\sigma_1 z_1 + (x_1\beta_1))^2(\sigma_2 z_2 + (x_2\beta_2)) \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= E(\sigma_1^2 z_1^2 \sigma_2 z_2 + 2\sigma_2 z_1 z_2 \sigma_1 (x_1\beta_1) + \sigma_2 z_2 (x_1\beta_1)^2 + \sigma_1^2 z_1^2 (x_2\beta_2)$$
$$+ 2\sigma_1 z_1 (x_1\beta_1)(x_2\beta_2) + (x_1\beta_1)^2 (x_2\beta_2)) \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= \sigma_1^2\sigma_2 E(z_1^2 z_2 \mid z_1 \geq a_1, z_2 \geq a_2) + 2\sigma_1\sigma_2 (x_1\beta_1) E(z_1 z_2 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$+ \sigma_2 (x_1\beta_1)^2 E(z_2 \mid z_1 \geq a_1, z_2 \geq a_2) + \sigma_1^2 (x_2\beta_2) E(z_1^2 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$+ 2\sigma_1 (x_1\beta_1)(x_2\beta_2) E(z_1 \mid z_1 \geq a_1, z_2 \geq a_2) + (x_1\beta_1)^2 (x_2\beta_2)$$

$$E(y_2 y_1^2 \mid y_1 \geq 0, y_2 \geq 0) = E((\sigma_2 z_2 + (x_2\beta_2)(\sigma_1 z_1 + (x_1\beta_1))^2) \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= E(\sigma_1^2 \sigma_2 z_2 z_1^2 + 2\sigma_2 z_2 \sigma_1 z_1 (x_1\beta_1) + \sigma_2 z_2 (x_1\beta_1)^2 + \sigma_1^2 z_1^2 (x_2\beta_2)$$
$$+ 2\sigma_1 z_1 (x_1\beta_1)(x_2\beta_2) + (x_1\beta_1)^2 (x_2\beta_2)) \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= \sigma_1^2\sigma_2 E(z_2 z_1^2 \mid z_1 \geq a_1, z_2 \geq a_2) + 2\sigma_1\sigma_2 (x_1\beta_1) E(z_2 z_1 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$+ \sigma_2 (x_1\beta_1)^2 E(z_2 \mid z_1 \geq a_1, z_2 \geq a_2) + \sigma_1^2 (x_2\beta_2) E(z_1^2 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$+ 2\sigma_1 (x_1\beta_1)(x_2\beta_2) E(z_1 \mid z_1 \geq a_1, z_2 \geq a_2) + (x_1\beta_1)^2 (x_2\beta_2)$$

$$E(y_1 y_2^2 \mid y_1 \geq 0, y_2 \geq 0) = E((\sigma_1 z_1 + (x_1\beta_1))(\sigma_2 z_2 + (x_2\beta_2))^2 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= E(\sigma_2^2 z_1 z_2^2 \sigma_1 + 2\sigma_1 z_1 \sigma_2 z_2 (x_2\beta_2) + \sigma_1 z_1 (x_2\beta_2)^2 + \sigma_2^2 z_2^2 (x_1\beta_1)$$
$$+ 2\sigma_2 z_2 (x_2\beta_2)(x_1\beta_1) + (x_2\beta_2)^2 (x_1\beta_1)) \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$= \sigma_2^2\sigma_1 E(z_1 z_2^2 \mid z_1 \geq a_1, z_2 \geq a_2) + 2\sigma_1\sigma_2 (x_2\beta_2) E(z_1 z_2 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$+ \sigma_1 (x_2\beta_2)^2 E(z_1 \mid z_1 \geq a_1, z_2 \geq a_2) + \sigma_2^2 (x_1\beta_1) E(z_2^2 \mid z_1 \geq a_1, z_2 \geq a_2)$$
$$+ 2\sigma_2 (x_1\beta_1)(x_2\beta_2) E(z_2 \mid z_1 \geq a_1, z_2 \geq a_2) + (x_2\beta_2)^2 (x_1\beta_1)$$

$$
\begin{aligned}
E(y_2^2 y_1 \mid y_1 \geq 0, y_2 \geq 0) &= E((\sigma_2 z_2 + (x_2\beta_2))^2(\sigma_1 z_1 + (x_1\beta_1)) \mid z_1 \geq a_1, z_2 \geq a_2) \\
&= E(\sigma_2^2 z_2^2 \sigma_1 z_1 + 2\sigma_1 z_2 z_1 \sigma_2 (x_2\beta_2) + \sigma_1 z_1 (x_2\beta_2)^2 + \sigma_2^2 z_2^2 (x_1\beta_1) \\
&\quad + 2\sigma_2 z_2 (x_2\beta_2)(x_1\beta_1) + (x_2\beta_2)^2 (x_1\beta_1)) \mid z_1 \geq a_1, z_2 \geq a_2) \\
&= \sigma_2^2 \sigma_1 E(z_2^2 z_1 \mid z_1 \geq a_1, z_2 \geq a_2) + 2\sigma_1\sigma_2 (x_2\beta_2) E(z_2 z_1 \mid z_1 \geq a_1, z_2 \geq a_2) \\
&\quad + \sigma_1 (x_2\beta_2)^2 E(z_1 \mid z_1 \geq a_1, z_2 \geq a_2) + \sigma_2^2 (x_1\beta_1) E(z_2^2 \mid z_1 \geq a_1, z_2 \geq a_2) \\
&\quad + 2\sigma_2 (x_1\beta_1)(x_2\beta_2) E(z_2 \mid z_1 \geq a_1, z_2 \geq a_2) + (x_2\beta_2)^2 (x_1\beta_1)
\end{aligned}
$$

GMM AND VIRTUAL PRICES APPROACH FOR ESTIMATING SYSTEM OF CENSORED DEMAND EQUATIONS: A CASE STUDY FOR ANALYZING THE IMPACT OF THE *E.COLI* OUTBREAKS ON CONSUMER DEMAND FOR SALAD VEGETABLES

Faysal Habib Fahs

School of Economic Sciences
Washington State University
Pullman, WA 99163-4741
(509)-339-3376
Email: faysal74@wsu.edu

ESSAY TWO:

GMM AND VIRTUAL PRICES APPROACH FOR ESTIMATING SYSTEM OF CENSORED DEMAND EQUATIONS: A CASE STUDY FOR ANALYZING THE IMPACT OF THE *E.COLI* OUTBREAKS ON CONSUMER DEMAND FOR SALAD VEGETABLES

Abstract

This paper examines the impact of the *E.coli* outbreaks that occurred in 2006 on consumer demand for salad vegetables on the West Coast of the United States. The scanner-data set used in our analysis is obtained from a chain supermarket and is aggregated on a weekly basis for the consumption of salad vegetables. The data contain a significant portion of observations with zero consumption on one or more vegetable groups. Zero consumption may be reflecting consumer concern about the *E.coli* outbreaks, the effect of removal of vegetable groups from store shelves due to product recalls and/or the result of personal preferences with respect to consumption. We motivate the use of the Tobit model as a statistical representation of consumer behavior by specifying the Quadratic Almost Ideal Demand System (QUAIDS) with demographic effects under binding non-negativity constraints. To avoid violating the non-negativity constraints of the model and to overcome the computational burden of high dimensionality, the GMM approach, along with the virtual prices concept, are used for estimating the system of non-linear censored demand equations. The empirical results show that during the outbreak period lettuce and cabbage were substituted for spinach, indicating consumers' concern about the *E.coli* impact.

1. Introduction

Consumption of salad vegetables has increased significantly in recent years because of health interests and the dietary trends of consumers. The popularity of salad bars, ethnic dishes and vegetable appetizers in restaurants contributed to the increase in the consumption of vegetables. Freshly prepared salads are offered in more than 70% of fast food establishments and family restaurant chains (Hedberg *et al*., 1994). In addition, fresh cut vegetables, including salad vegetables, are sold in supermarkets. According to the U.S. Department of Agriculture (USDA), in 2000, Americans consumed 20% more vegetables than in the 1970s. Figure 1 in the Appendix shows how the annual per capita vegetable consumption has continued to increase over time[1].

Fresh salad vegetables are generally considered safe to eat by consumers. However, a survey conducted by the Center for Disease Control and Prevention (CDC) revealed that vegetables, as a category, contributed to 5% of the foodborne disease outbreaks in the World, 2% of which occurred in the USA, during 1973 and 1987 (Bean and Griffin, 1990). Outbreaks of foodborne diseases caused by *E. coli* (*Escherichia coli*)[2] bacteria have become a serious problem in the United States. The CDC estimates 73,000 cases of infection with *E. coli* 0157:H7 and 61 deaths on average occur in the USA every year.[3]

In 2006 there were two major outbreaks of the *E. coli* 0157:H7 in North America: the first outbreak occurred in September and the second outbreak happened in November. Both outbreaks were tracked to organic bagged fresh spinach that was grown in San

---

[1] Source: USDA, ERS Briefing rooms, Vegetables and Melons

[2] Escherichia coli: is a bacterium that is commonly found in the lower intestine of warm-blooded animals. Most *E. coli* strains are harmless, but some such serotype O157:H7, can cause serious food poisoning in humans, and are occasionally responsible for costly product recalls

[3] Center for Disease Control and Prevention, Strategies to Reduce Person-to-Person Transmission during Widespread Escherichia coli O157:H7 Outbreak, Edmund Y.W. Seto, Jeffrey A. Soller, and John M. Colford Jr

Benito County, California. The CDC investigators initially speculated that the dangerous strains of the *E. coli* 0157:H7 bacteria originated from the irrigation water contaminated with cattle feces or from grazing deer[4]. Federal and state public health officials issued a nationwide health alert warning. Although the Food and Drug Administration (FDA) had no mandatory recall authority, Natural Selection Food, other processors, and retailers quickly initiated a voluntary recall of all products containing fresh spinach. As a result of the voluntary removal of spinach from retail shelves, new illness outbreaks were curtailed by early October and December for both outbreaks, respectively.

The long term impact of the *E. coli* O157:H7 outbreak is still uncertain; however, the broad recall could have lasting effects on consumers. The economic impacts of the recall had the greatest effect on growers and retailers, when sales of salad mix experienced a roughly 50% reduction during the two outbreaks, according to Pacific International Marketing.

In this paper we estimate the effect of *E.coli* on consumer demand for salad vegetables (*Tomato*, *Onion*, *Cabbage*, *Lettuce* and *Spinach*). The data used in this analysis are a scanner-data set, provided by a well known retail chain[5] that has many supermarkets throughout the United States. For this analysis we collected data from stores located on the West Coast of the United States (California, Oregon and Washington).

The Quadratic Almost Ideal Demand System (QUAIDS) is specified for estimating the demand for salad vegetables. For the outbreak period, the data contain a significant portion of observations in which the expenditure on one or more vegetable groups is zero.

---

[4] Source: FDA Warning on Serious Foodborne *E.coli* O157:H7 Outbreak. FDA (September 14, 2006)
[5] The name of the retail store is suppressed for confidentiality reasons

The presence of zero consumption observations reflects positive probabilities that choices occur at the kink or boundary points of the feasible choice set. This implies that the dependent variable is censored at zero. In a systems approach the censored regression equations typically have correlated error terms and require a joint estimation procedure, which generally leads to the specification of a mixed distribution consisting of continuous probability density functions for the positive observations and discrete probability mass functions for the zero observations (Lee, 1993).

In the literature, the Multivariate Tobit Model (MVT) has been frequently used to represent the data generating process underlying consumer demand models that exhibit binding non-negativity constraints (Wales and Woodland, 1983; Ransom, 1987; Hausman, 1985; Lee and Pitt, 1986; Heien and Wessells, 1990, and Lee, 1993). Estimating the MVT model using classical maximum likelihood methods requires the evaluation of a partially integrated multivariate normal probability density function, which is known to be computationally inefficient, inaccurate and intractable as the dimensionality of the integration problem increases much beyond three.

Fahs and Mittelhammer (2007) attempted to circumvent the estimation tractability problem of systems of censored demand equations by utilizing a Generalized Method of Moments (GMM) approach in place of Maximum Likelihood. The authors calculate parameter estimates and the covariance structure of the error terms based on a set of marginal univariate and bivariate moment conditions that account for correlation among random components of the model.

In our analysis the response variables in the QUAIDS model are bounded between zero and one, and the sum of the response variables must equal one. It is possible in these types of models that predicted shares violate the non-negativity constraints and adding up conditions implied by Neoclassical economic theory. We address this problem by applying the concept of virtual prices that was first suggested by Lee and Pitt (1986). Using the virtual prices concept along with the GMM approach, we ensure that the predicted shares are between zero and one, eliminating the possibility of a violation of the non-negativity constraints and adding up conditions.

The reminder of the paper is organized as follows: Section 2 describes the characteristics of the data set, variables, and aggregation procedure. Section 3 introduces an overview of the QUAIDS model and its functional specifications with binding non-negativity constraints. Section 4 focuses on the discussion of the virtual prices concept and the use of the GMM estimation approach to calculate these prices. Section 5 summarizes the GMM approach (Fahs & Mittelhammer, 2007) for estimating system of censored demand equations. Section 6 discusses the GMM estimation procedure for the QUAIDS model incorporating estimated virtual prices. Section 7 summarizes, interprets, and analyzes the estimation results. Lastly, Section 8 ends with our conclusions.

2. Data, Variables and Related Issues

The scanner-data set used originally contained observations referring to three states that are located on the West Coast of the United States (California, Oregon and Washington). Ten stores from each state were selected taking into consideration geographical spread, and representation of rural and urban areas. Figure 2 in the Appendix presents the geographical location of these retail stores[6]. The daily data observations represent consumer purchases of salad vegetables for one calendar year

[6] Note that these retail stores belong to one supermarket chain

(January 1, 2006 to December 31, 2006), as well as household expenditures, quantities purchased and socio-demographic variables that are expected to influence consumer behavior.

Five groups of vegetables suitable for salad consumption were identified: *Tomato*, *Onion*, *Cabbage*, *Lettuce* and *Spinach*. Each group is divided into two subgroups: *bulk* and *bagged*; those, in turn, vary by different vegetable types as shown in Table 1 of the Appendix.

Socio-demographic variables are important factors in influencing the variety as well as the quantity of vegetable consumption by households. In this study these variables are defined as:

- *Age*: is a continuous variable that represents the age of the head of the household.
- *Sex*: is an indicator variable that represents the gender of the head of the household, where male is set as the base indicator value of 0 (Female =1 and male= 0).
- *Children*: is a variable that represents the number of children in the household.
- *Status*: is an indicator variable that represents the marital status of the head of the household, where single is the base indicator value of 0 (single=0 and Married=1).
- $Income\text{-}D_1$: is a dummy variable that indicates consumers' income in the range between $25,000 and $50,000 ($Income\text{-}D_1$=1 and 0 otherwise).
- $Income\text{-}D_2$: is a dummy variable that indicates consumers' income greater than $50,000 ($Income\text{-}D_2$ =1 and 0 otherwise).

- *Location*: is an indicator variable, where the location of the household is identified, the north states (Oregon and California) being the base indicator value of 0 (Oregon and California =0 and California =1).
- *D-September*: is a dummy variable that represents the month of the first outbreak (September), during which the consumers purchased salad vegetables (*D-September*=1 and 0 otherwise).
- *D-October*: is a dummy variable that represents the month after the first outbreak (October), during which the consumers purchased salad vegetables (*D-October* =1 and 0 otherwise).
- *D-November*: is a dummy variable that represents the month of the second outbreak (November), during which the consumers purchased salad vegetables (*D-November*=1 and 0 otherwise).
- *D-December*: is a dummy variable that represents the month after the second outbreak (December), during which the consumers purchased salad vegetables (*D-December* =1 and 0 otherwise).
- *Time* and $Time^2$: are the variables that account for time trends, where *Time* represents the months from January to December (*t*=1, 2…, 12).

The original data consisted of five different groups of vegetables with a number of types and varieties as indicated in Table 1. The data were aggregated in two steps: first, into five major groups (*Tomato*, *Onion*, *Cabbage*, *Lettuce* and *Spinach*); second, on a weekly observation basis. Note that all the quantities sold are measured in pounds of group product (lb unit) and the expenditures are measured in dollars. Associated with the aggregated quantities and expenditures, the weighted (by quantities) average prices for each of the vegetable groups consumed were calculated. For those households who did not purchase a particular vegetable group, it was assumed that they paid the average price associated with the vegetable groups for the store they made other purchases from.

After aggregating, the data contained five different groups of prices along with the demographic variables. The demographic variables were obtained from a third party contractor and it contained missing data. *Age*, *Sex*, *Status*, and *Income* variables had the same percentage of missing data i.e. 38% (the data on those demographics are either completely missing or non-missing) and the variable *Children* had 45% missing data. Three different types of statistical analysis were conducted to test if the missing values were missing completely at random (MCAR). MCAR exists when missing values are randomly distributed across all observations. If data are MCAR, then the researcher may choose to delete the missing observations[7].

The prices of vegetables were divided into two groups (missing and non-missing) matching the corresponding observations of missing and non-missing in the demographic variables. Three types of analysis were conducted for the two groups of prices created (missing and non-missing). The first analysis was to check if the means of those two groups were the same. The null hypothesis is: sample mean of prices with missing observations equal to the sample mean of prices with non-missing observations. A simple T-test was performed and the values were reported in Table 2 in the Appendix. The results show that at 95% significance level (critical value=1.96) we fail to reject the null hypothesis (t-values are less than the critical value). This finding indicates that the means

[7] Missing Data: A Gentle Introduction, by Patrick E. McKnight, Souraya Sidani, Aurelio Jose Figueredo

between the two groups (missing and non-missing) for all types of prices are statistically the same.

The second analysis was to check if the two groups of prices (missing and non-missing) have similar distributions. We generated the empirical cumulative density functions for the two groups of prices as shown in Figures 3 –12. The patterns of curves are noted to be nearly similar, which suggest that the distributions of the prices with missing and non-missing observations are almost identical. For example, Figure 3 shows that 80% of the onion prices with missing observations and 80% of onion prices with non-missing observations have prices less than 1.13 dollars (horizontal axis represents prices in dollars and vertical axis represents percentage of data).

The third analysis was conducted to check whether the two distributions of prices (missing and non-missing) were statistically identical. The Mann-Whitney Test, which is a non-parametric test to determine whether two populations have the same population median, was used for this analysis[8]. The Mann-Whitney test does not require the data to come from normally distributed populations, but it does make the following assumptions: 1) the populations of interest have the same shape; 2) the populations are independent. The null hypothesis is: the two population medians are equal or have identical distributions. Table 3 in the Appendix presents the p-values of the Mann-Whitney test. It indicates that the p-values are not significant at alpha =0.05, which suggests that that we fail to reject the null hypothesis.

The above three analysis suggested that there were no statistical differences between the two groups of prices (missing and non-missing). According to those tests, we

[8] Statistics for Biologists by Richard Colin Campbell

assumed that the data were missing completely at random. Deleting the missing cases form the data, with the assumption of MCAR, the sample selectivity would not be an issue for furtherer analysis.

Table 4 in the appendix represents the summary statistics for all variables along with the market shares for each vegetable group. The sample of observations on the consumption of the five vegetable groups contained 377,149 observations.

3. A QUAIDS Consumer Demand System under Binding Non-Negativity Constraints

It is desirable to specify the indirect utility function in a flexible way, retaining theoretical consistency but also producing a general and flexible demand system. There are several flexible forms introduced in literature. In this paper, we used the non-linear form of QUAIDS to represent our model because it has better approximation of the non-linear Engel curves in empirical analysis, aggregate perfectly over consumers and capable of imposing restrictions of homogeneity and Slutsky symmetry.

3.1 Overview of the QUAIDS Model

The Almost Ideal Demand System (AIDS) of Deaton and Muellbauer (1980) has been the most widely used system approach for modeling consumption behavior for grouped commodities after it was claimed to possess the best properties of both the Translog and Rotterdam models, including approximating any demand system arbitrarily to first-order, aggregating perfectly over consumers, satisfying the axioms of choice and capable of imposing restrictions of homogeneity and Slutsky symmetry.

However, the AIDS model has difficulty capturing the effects of non-linear Engel curves. In order to maintain the attractive properties of the AIDS model, while maintaining consistency with both Engel curve and relative price effects within the utility maximization framework, a quadratic term in log income or log total expenditure is added to the AIDS model and leads to Quadratic Almost Ideal Demand System (QUAIDS) model specification (Lewbel, 1997). Increased flexibility of the demand system representation is thus achieved in a parsimonious way through the addition of the quadratic term.

The generalized linear form of rank two (Gorman, 1981) is necessary and sufficient for aggregation of demands[9]. Rank two demand models include Linear AIDS, Translog, Linear Expenditure, Price-Independent Generalized linear (PIGL) and Price-Independent Generalized Log (PIGLOG) systems. These locally flexible functional forms possess a relatively small regular region according to Cooper and McLaren (1996), and they can only provide a local approximation within a small neighborhood of the true data generation function. The Translog has been criticized for mistakenly classifying goods as complements when they are substitutes, and it loses flexibility when semi-definiteness (curvature) is imposed (Diewert and Wales, 1987), while the Linear Expenditure Systems has been criticized for its additive preference structure.

Because of these problems, researchers focused their attention on developing globally flexible functional forms that have higher ranks. The QUAIDS has rank three, and better approximates non-linear Engel curves in empirical analysis. Since the QUAIDS model produces a considerably larger regular region than the locally flexible

---

[9] Rank is the maximum dimension of the function space spanned by Engel curves of the demand system (Lewbel, 1991).

forms, it can be classified as "effectively globally regular", which is a label defined by Cooper and McLaren (1996) for locally flexible functional forms with larger regularity regions, whose corresponding direct and indirect utility functions and cost functions satisfy their theoretical properties for all non-negative demand, price and all utility levels as appropriate.

3.2 Functional Specification of the QUAIDS Model

Assume a rational consumer has an income y to spend on the purchase of M different goods $Q = (Q_1, \ldots, Q_M)$. The consumer considers both income and prices of goods $p = (p_1, \ldots, p_M)$ to be exogenous, and furthermore, she has the possibility of consuming the desired quantities and does not face transaction costs. Given the prices of the goods and income, the individual chooses a particular consumption vector, $q = (q_1, \ldots, q_M)$. This vector belongs to the consumption set defined as the non-negative space $\mathbf{R}_{\geq}^{M}$ and is the non-negative quantities of the M consumption goods that maximize utility, where these quantities do not imply expenditure higher than available income. In this paper we focus on the demand for a specific group of commodities, and use total expenditure on those commodities, y, in place of income.

Assuming the existence of a utility function $U(q)$ that represents the preferences of the consumer, one can formulate the consumer equilibrium as the constrained optimization problem

(3.2.1) $$\max U(q) \ s.t. \ y = pq$$

whose first-order conditions can be solved to derive the Marshallian demand system $q_i = q_i\left(p, y\right)$, for $\left(i = 1,\ldots,M\right)$.

Now consider the QUAIDS model as a generalization of PIGLOG preferences (Muellbauer, 1975 and 1976), which can be derived from the following indirect utility function for *M* commodities as:

(3.2.2) $$\ln V = \left\{ \left[\frac{\ln y - \ln a\left(p\right)}{b\left(p\right)} \right]^{-1} + \lambda\left(p\right) \right\}^{-1}$$

where

(3.2.3) $$\ln a\left(p\right) = \alpha_0 + \sum_{i=1}^{M} \alpha_i \ln p_i + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} \gamma_{ij} \ln p_i \ln p_j$$

(3.2.4) $$b\left(p\right) = \prod_{i=1}^{M} p_i^{\beta_i}$$ is the simple Cobb-Douglas price aggregator

(3.2.5) $$\lambda\left(p\right) = \sum_{i=1}^{M} \lambda_i \ln p_i$$

and $\sum_{i=1}^{M} \lambda_i = 0$, All $a\left(p\right), b\left(p\right)$ and $\lambda\left(p\right)$ are defined to be homogenous functions of degree zero in prices.

Applying Roy's identity to equation (3.2.2) obtains:

(3.2.6) $$w_i = \frac{\partial \ln a(p)}{\partial \ln p_i} + \frac{\partial \ln b(p)}{\partial \ln p_i}\left(\ln x\right) + \frac{\partial \lambda}{\partial \ln p_i}\frac{1}{b(p)}\left(\ln x\right)^2$$

where $\ln x = \ln(y) - \ln a(p)$, and w_i denotes the expenditure share of the i^{th} commodity. Inserting the appropriate derivative expressions in 3.2.6 obtains Marshallian demand equation for the QUAIDS model in budget share form as:

(3.2.7) $$w_i = \alpha_i + \sum_{j=1}^{M} \gamma_{ij} \ln p_j + \beta_i \ln\left(\frac{y}{a(p)}\right) + \frac{\lambda_i}{b(p)}\left\{\ln\left(\frac{y}{a(p)}\right)\right\}^2 + \varepsilon_i$$

and note that *i* and *j* denote commodities, which are *M* in numbers.

The adding-up restriction, $\sum_{i=1}^{M} w_i = 1$ implies that

(3.2.8) $$\sum_{i=1}^{M} \alpha_i = 1$$

(3.2.9) $$\sum_{i=1}^{M} \lambda_i = 0$$

(3.2.10) $$\sum_{i=1}^{M} \beta_i = 0$$

(3.2.11) $$\sum_{i=1}^{M} \gamma_{ij} = 0 \ \ \forall j$$

Since the Marshallian demands are homogenous of degree zero in (p, y)

(3.2.12) $$\sum_{i=1}^{M} \gamma_{ij} = 0 \ \ \forall j$$

Slutsky symmetry implies that:

(3.2.13) $$\gamma_{ij} = \gamma_{ji} \ \ \forall i \neq j$$

As can be seen, the expenditure shares, which are quadratic in the logarithm of expenditure, have been derived from PIGLOG preferences. Therefore, they maintain the relevant properties of its linear counterpart, the AIDS, allowing for exact aggregation over households. It is evident that the QUAIDS budget shares reduce to those of AIDS if $\lambda_i = 0$, for all *i*, in which case the rank three Engel curves for the QUAIDS reduces to the rank two Working-Leser Engel curves.

Demographic effects on household consumption are introduced in this model, where constant term and total expenditure coefficients are specified to depend on the vector of household characteristics Z^h (Moschini and Rizzi, 1997). Including these demographic effects, the share model is expressed as:

$$(3.2.14) \qquad w_i = \alpha_i + \sum_s \alpha_{is} z_s^h + \sum_{j=1}^{n} \gamma_{ij} \ln p_j + \beta_i \ln\left(\frac{y}{a\left(p,z^h\right)}\right) + \frac{\lambda_i}{b\left(p\right)}\left\{\ln\left(\frac{y}{a\left(p,z^h\right)}\right)\right\}^2 + \varepsilon_i$$

Note that s denotes the demographic variables, which are seven in number. And

$$(3.2.15) \qquad \ln a\left(p,z\ \right) = \alpha_0 + \sum_s \alpha_{0s} z_s^h + \sum_i \left(\alpha_i + \sum_s \alpha_{is} z_s^h\right) \ln p_i + \frac{1}{2}\sum_i \sum_j \gamma_{ij} \ln p_i \ln p_j$$

and the following parameter restriction applies

$$(3.2.16) \qquad \sum_i \alpha_{is} = 0 \quad \forall s$$

Following Banks *et al*. (1997), expenditure elasticities are calculated by differentiating the log of equation (3.2.14) with respect to log of total expenditure ($\ln y$), yielding:

$$(3.2.17) \qquad e_i = 1 + \frac{1}{w_i}\left\{\beta_i + \frac{2\lambda_i}{b\left(p\right)} \ln\left(\frac{y}{a\left(p,z^h\right)}\right)\right\}$$

The uncompensated price elasticities are calculated by differentiating the log of equation (3.2.14) with respect to the log of price ($\ln p_j$):

$$(3.2.18) \quad e_{ij}^{U} = -\delta_{ij} + \frac{1}{w_i}\left\{\gamma_{ij} - \left(\beta_i + \frac{2\lambda_i}{b(p)}\ln\left(\frac{y}{a(p,z^h)}\right)\right)\left(\alpha_j + \sum_s \alpha_{js} z_s^h + \sum_r \gamma_{jr}\ln(p_r)\right) - \frac{\lambda_i \beta_j}{b(p)}\left(\ln\left(\frac{y}{a(p,z^h)}\right)\right)^2\right\}$$

where δ_{ij} is the Kronecker's delta, which equals one if $i = j$ (i.e., for own price elasticities) and zero otherwise.

The compensated price elasticities are deduced from the Slutsky's formula as follow:

$$(3.2.19) \quad e_{ij}^{C} = e_{ij}^{U} + w_j e_i$$

Finally, the QUAIDS model can be represented in regression form as:

$$(3.2.20) \quad w_i = \alpha_i + \sum_s \alpha_{is} z_s^h + \sum_{j=1}^{M} \gamma_{ij} \ln p_j + \beta_i \ln\left(\frac{y}{a(p,z^h)}\right) + \frac{\lambda_i}{b(p)}\left\{\ln\left(\frac{y}{a(p,z^h)}\right)\right\}^2 + \varepsilon_i$$

If it is assumed in (3.2.14) that $\varepsilon \sim N([0],\Sigma)$ and $\Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1M} \\ \vdots & \ddots & \vdots \\ \sigma_{M1} & \cdots & \sigma_{MM} \end{bmatrix}$, so that

$w_i \sim N(u,\Sigma)$ where

$u_i = \alpha_i + \sum_s \alpha_{is} z_s^h + \sum_{j=1}^{M} \gamma_{ij} \ln p_j + \beta_i \ln\left(\frac{y}{a(p,z^h)}\right) + \frac{\lambda_i}{b(p)}\left\{\ln\left(\frac{y}{a(p,z^h)}\right)\right\}^2$, then the

demand model in equation (3.2.14) is a censored demand equation system and can be represented in the MVT model form as:

$$(3.2.21) \quad w_i = \begin{cases} w_i \ \textit{if} \ w_i \geq 0 \\ 0 \ \textit{otherwise} \end{cases}$$

where w_i represents the observed level of consumption for the vegetable groups (*Tomato*, *Cabbage*, *Lettuce* and *Spinach*). Note that we drop one of the vegetable groups (Onion) from the MVT-QUAIDS model to deal with the inherent covariance matrix singularity in the model.

4. Virtual Price and Internal Consistency of the Model

Given the QUAIDS model presented in equation (3.2.14), two econometric problems generally arise when estimating these type of models; internal theoretical Neoclassical consistency of the model and evaluation of high dimensional multivariate integrals corresponding to the corner solution decisions.

The first problem arises because each share variable in the QUAIDS models is bounded between zero and one, and the sum of the shares must equal one. It is possible that the predicted shares violate the non-negativity constraint and the adding up conditions. The second problem arises in evaluating the probability distribution of the mixed continuous-discrete density function. This latter issue will be discussed in section 5.

Lee and Pitt (1986a) introduced the concept of virtual prices when dealing with corner solutions. In this section, to elucidate the difficulties arising from the improper use of market prices when corner solutions are present, we initially investigate a simplified model of consumer choice. We illustrate that the correct price to be used in the case of corner point solutions is the virtual price, which is the price that results in the quantity demanded equalling zero. This approach can be used to insure that budget shares are bounded between zero and one, and the sum of the shares is equal to one.

Assume a choice set involving two goods, and let the utility function *U (q_1, q_2)* be smooth, continuously differentiable, strictly increasing and strictly quasi-concave utility of q_1 and q_2.

The consumer seeks to maximize utility subject to:

(4.1) $$P_{q_1} q_1 + P_{q_2} q_2 \le Y$$

where P_{q_1} and P_{q_2}, are the respective prices of the goods, respectively, Y is income, and $q_1 \ge 0, q_2 \ge 0$ are the non-negativity constraints.

The demand functions can be solved by maximizing the following Lagrangian:

(4.2) $$L = U\left(q_1, q_2\right) - \lambda\left(Y - P_{q_1} q_1 - P_{q_2} q_2\right) + \delta q_1 + \eta q_2$$

where λ, δ and η are LaGrange multipliers, respectively.

Given that a strictly increasing utility function implies budget exhaustion, the Kuhn-Tucker conditions can be written as:

(4.3) $$\begin{aligned} \frac{\partial L}{\partial q_1} &= \frac{\partial U\left(q_1, q_2\right)}{\partial q_1} - \lambda P_{q_1} + \delta \\ \frac{\partial L}{\partial q_2} &= \frac{\partial U\left(q_1, q_2\right)}{\partial q_2} - \lambda P_{q_2} + \eta \end{aligned}$$

(4.4) $$\begin{aligned} &\Rightarrow \frac{\partial U\left(q_1, q_2\right) / \partial q_1}{\lambda} = P_{q_1} - \delta / \lambda \\ &\Rightarrow \frac{\partial U\left(q_1, q_2\right) / \partial q_2}{\lambda} = P_{q_2} - \eta / \lambda \end{aligned}$$

and

(4.5) $$\delta q_1 = 0, \quad \eta q_2 = 0$$

From equations (4.1) to (4.5), the consumer decides the optimal consumption bundle $q_1^*\left(P_{q_1}, P_{q_2}, Y\right), q_2^*\left(P_{q_1}, P_{q_2}, Y\right)$ where the optimal Lagrange multipliers are such that $\lambda^*(\bullet) > 0$,

$\delta^*(\bullet) \geq 0 \, and \, \eta^*(\bullet) \geq 0$. Inserting these values into (4.3) and (4.4) allows defining the Marshallian (virtual) prices for q_1 and q_2 as:

(4.6)
$$P_{q_1}^v\left(P_{q_1}, P_{q_2}, Y\right) = \frac{\partial U\left(q_1^*, q_2^*\right)/\partial q_1}{\lambda^*} = P_{q_1} - \delta^*/\lambda^*$$
$$P_{q_2}^v\left(P_{q_1}, P_{q_2}, Y\right) = \frac{\partial U\left(q_1^*, q_2^*\right)/\partial q_2}{\lambda^*} = P_{q_2} - \eta^*/\lambda^*$$

Equation (4.6) suggests that $P_{q_1}^v$ and $P_{q_2}^v$ equal their corresponding market prices only if $\delta^* = 0$ and $\eta^* = 0$ or equivalently if $q_1^* > 0$ or $q_2^* > 0$, given the complementary slackness conditions represented by equation (4.5).

If we assume that the consumer decides to consume bundle q_1 only, then $P_{q_1}^v = P_{q_1}$ and $q_2 = 0 \Rightarrow P_{q_2}^v \neq P_{q_2}$ and $P_{q_2}^v = P_{q_2} - \eta^*/\lambda^*$. In other words, at equilibrium the prices that determine the quantity demanded of q_1 and q_2 are P_{q_1} for positively consumed commodities and the virtual prices for q_2, $P_{q_2}^v = P_{q_2} - \eta^*/\lambda^*$, the prices that set the demand for q_2 to zero. Thus, at equilibrium the quantity demanded of q_1 and q_2 in this situation is:

(4.7)
$$q_1^* = q_1\left(P_{q_1}, P_{q_2}^v\left(P_{q_1}, P_{q_2}, Y\right), Y\right)$$
$$q_2^* = q_2\left(P_{q_1}, P_{q_2}^v\left(P_{q_1}, P_{q_2}, Y\right), Y\right)$$

An intuitive explanation of the use of virtual prices in a two-choice model is provided in Figure 3 of the Appendix. The utility-maximizing observed consumption bundle in this case is a corner solution, where $q_2 = 0$ at market prices (P_{q_1}, P_{q_2}). If the utility function was maximized without regard to non-negativity constraints, the solution would be the notional demands $(q_1, -q_2)$, where the good q_2 is consumed at a negative level. The virtual price $P_{q_2}^V$ for the q_2 good is a reservation price at which consumption of the good is induced to be exactly zero. By using the price ratio $\frac{P_{q_1}}{P_{q_2}^V}$ in place of $\frac{P_{q_1}}{P_{q2}}$, one can characterize the tangency condition for the observed consumption bundle. Also note that in the case of a corner solution the market price is greater than the virtual price of a non-consumed good. Comparison of the virtual price to the market price can, therefore, be used to identify which goods are non-consumed.

Now consider how virtual prices can be calculated empirically. Given the aggregated data and market prices as described in section 2, suppose the MVT-QUAIDS model is estimated (with all restrictions that apply: homogeneity, adding up and symmetry). The parameter estimates resulting from the estimation procedure are economically meaningless if there is a violation of the non-negativity constraints, which would be evident if some of the predicted budget shares are negative valued. Consistent with the concept of virtual prices, set the negative values of the predicted shares to zeros and solve for the respective virtual prices within the vector p that satisfy the following system:

(4.8)

$$0 = \hat{\alpha}_i + \sum_s \hat{\alpha}_{is} z_s^h + \sum_{j=1}^{M} \hat{\gamma}_{ij} \ln p_j + \hat{\beta}_i \ln\left(\frac{y}{a\left(p, z^h\right)}\right) + \frac{\hat{\lambda}_i}{b(p)} \left\{ \ln\left(\frac{y}{a\left(p, z^h\right)}\right) \right\}^2, \; i \in \{i : \hat{w}_i < 0\}$$

Then using the price vector (p^*) consisting of market prices for goods consumed at positive values and virtual prices for those not consumed (zero-valued) values of predicted shares can be recalculated, reflecting the appropriate corner point solution for the non-consumed good.

5. GMM Estimation of the MVT System

In estimating the QUAIDS model, a difficulty arises when evaluating the probability distribution of the mixed continuous-discrete density function. This probability density function exhibits both discrete (probability of observing zero outcomes) and continuous (probability density of $w > 0$) components (Lee, 1993). Researchers recognized that if the number of decision outcomes, and thus the number of equations in the system, is relatively large, the propensity for a notable subset of the decision outcomes in the system to occur at kink or boundary points is generally increased. In this case ML estimation of unknown parameters of the system will involve integration problems at each iteration of a likelihood maximization algorithm that are equal in number to the number of sample observations having binding non-negativity constraints. Moreover, the dimension of each of the integrals equals the number of elements contained in the discontinuous component of the observation. While numerical integration methods can evaluate multiple integrals in two or even in three dimensions with reasonable accuracy and speed, integration becomes progressively slower, less accurate, and /or intractable as the dimension of integration increases.

Circumventing the problem of tractability, we use the GMM approach (Fahs and Mittelhammer, 2007) to estimate the QUAIDS model. The GMM approach can be applied to the QUAIDS model by utilizing the general marginal and bivariate moment relations that hold between explanatory variables and model noise. The GMM can use both univariate and bivariate moments sufficient to mitigate information loss while avoiding the use of higher order moments and higher dimensional probability integrals.

The general form of the MVT model is represented as:

(5.1)
$$w_{ji} = \begin{cases} w_{ji} = u_{ji} & if \quad w_{ji} > 0 \\ 0 & otherwise \end{cases}$$

where now the subscript j denotes the commodity[10] and subscript i denotes the sample observation, and u_{ji} is an observation on the appropriate function of the explanatory variables.

5.1 Marginal Moments

Fahs and Mittelhammer (2007) derive all the marginal moments that are based on equation (5.1) as:

(5.1.1)
$$E(W_{j.} \mid W_{j.} > 0) = U_{j.} + \sigma_j \frac{\phi_j}{\Phi_j}$$

(5.1.2)
$$E(W_{j.}^{2} \mid W_{j.} > 0) = U_{j.} E(W_{j.} \mid W_{j.} > 0) + \sigma_j^2$$

(5.1.3)
$$E(W_{j.}) = U_{j.} \Phi_j + \sigma_j \phi_j$$

(5.1.4)
$$E(W_{j.}^{2}) = U_{j.} E(W_{j.}) + \sigma_j^2 \Phi_j$$

(5.1.5)
$$E(W_{Binary\ j.}) = \Phi_j$$

[10] For simplicity of presenting the moment equation we used the notation J to represent the commodities instead of M (only for section 5)

where the subscript j denotes the j^{th} commodity[11], ϕ_j and Φ_j are shorthand for $\phi\left(\frac{U_{j.}}{\sigma_j}\right)$ and $\Phi\left(\frac{U_{j.}}{\sigma_j}\right)$, respectively, $U_{j.}$ denotes the observations on the explanatory variables relating to the j^{th} commodity and $W_{Binary\,j.}$ is defined as:

$$W_{Binary\,j.} = \begin{Bmatrix} 1 & if \quad W_{j.} > 0 \\ 0 & \text{otherwise} \end{Bmatrix}.$$

Equations (5.1.1) and (5.1.2) represent the conditional first and second order moment conditions respectively, and equations (5.1.3) and (5.1.4) represent the unconditional first and second order moment conditions respectively. Finally, equation (5.1.5) represents the binary moment conditions.

Gathering all of the marginal moment conditions (5.1.1)–(5.1.5) for all n observations, we can define the following relationships between $U_{j.}, W_{j.}$ and disturbances $\xi_j^{(i)}$ that have zero expectations as:

$$(5.1.6) \qquad \begin{Bmatrix} W_{j.>0} = U_{j.>0} + \sigma_j \odot \dfrac{1}{\boldsymbol{\Phi}_{j>0}} \odot \boldsymbol{\phi}_{j>0} + \xi_j^{(1)} \\ W_{j.>0}^2 = U_{j.>0} \odot E(W_{j.>0}) + \sigma_j^{\,2} + \xi_j^{(2)} \\ W_{j.} = U_{j.} \odot \boldsymbol{\Phi}_j + \sigma_j \odot \boldsymbol{\phi}_j + \xi_j^{(3)} \\ W_{j.}^{\,2} = U_{j.} \odot E\left(W_{j.}\right) + \sigma_j^{\,2} \odot \boldsymbol{\Phi}_j + \xi_j^{(4)} \\ W_{Binary\,j.} = \boldsymbol{\Phi}_j + \xi_j^{(5)} \end{Bmatrix}$$

where $\xi_j^{(i)}$ denotes the disturbances that correspond to the i^{th} set of moment conditions for the j^{th} commodity, $W_{j.>0}$ denotes the set of positive valued observations relating to the j^{th}

---

[11] $U_{j.}$ shorthand for $\begin{bmatrix} U_{1i} \\ \vdots \\ U_{Ji} \end{bmatrix}$, $W_{j.}$ shorthand for $\begin{bmatrix} W_{1i} \\ \vdots \\ W_{Ji} \end{bmatrix}$

commodity, $(U)_{j.>0}$ denotes the observations on the explanatory variables that correspond to the positive valued outcomes, $E(W_{j.>0})$ is shorthand notation for $EW_{j>0} \equiv E(W_{j.} \mid W_{j.} > 0) = (U)_{j.>0} + \sigma_j \odot \frac{1}{\boldsymbol{\Phi}_{j>0}} \odot \boldsymbol{\phi}_{j>0}$, $\odot$ denotes the Hadamard (elementwise) product, In this context, $\boldsymbol{\Phi}_j$ and $\boldsymbol{\phi}_j$ are vectors of the CDFs and PDFs of the standard normal distribution, evaluated at the vector $\left(\frac{U_{j.}}{\sigma_j}\right)$, respectively, $\boldsymbol{\Phi}_{j>0}$ and $\boldsymbol{\phi}_{j>0}$ are the subsets of those vectors corresponding to the positive valued observations $W_{j.>0}$, and $\frac{1}{\boldsymbol{\Phi}_{j>0}}$ denotes a vector of reciprocals of the elements in $\boldsymbol{\Phi}_{j>0}$.

Because

$$\xi_j^{(1)} = W_{j.>0} - U_{j.>0} - \sigma_j \odot \frac{1}{\boldsymbol{\Phi}_{j>0}} \odot \boldsymbol{\phi}_{j>0},$$

$$\xi_j^{(2)} = W_{j.>0}^2 - U_{j.>0} \odot E(W_{j.>0}) - \sigma_j^{\,2},$$

$$\xi_j^{(3)} = W_{j.} - U_{j.} \odot \boldsymbol{\Phi}_j - \sigma_j \odot \boldsymbol{\phi}_j,$$

$$\xi_j^{(4)} = W_{j.}^{\,2} - U_{j.} \odot E(W_{j.}) - \sigma_j^{\,2} \odot \boldsymbol{\Phi}_j,$$

and

$$\xi_j^{(5)} = W_{Binary\, j.} - \boldsymbol{\Phi}_j.$$

Orthogonality moment conditions can be defined as follows:

$$E\left[\boldsymbol{Z}_{j.>0}'\left(W_{j.>0} - U_{j.>0} - \sigma_j \odot \frac{1}{\boldsymbol{\Phi}_{j>0}} \odot \boldsymbol{\phi}_{j>0}\right)\right] = 0,$$

$$E\left[\boldsymbol{Z}_{j.>0}'\left(W_{j.>0}^2 - U_{j.>0} \odot EW_{j>0} - \sigma_j^{\,2}\right)\right] = 0,$$

$$E\left[\boldsymbol{Z}_{j.}'\left(W_{j.} - U_{j.} \odot \boldsymbol{\Phi}_j - \sigma_j \odot \boldsymbol{\phi}_j\right)\right] = 0,$$

$$E\left[\boldsymbol{Z}_{j.}'\left(W_{j.}^{2} - U_{j.} \odot E\left(W_{j.}\right) - \sigma_j^{2} \odot \boldsymbol{\Phi}_j\right)\right] = 0,$$

and

$$E\left[\boldsymbol{Z}_{j.}'\left(W_{Binary\ j.} - \boldsymbol{\Phi}_j\right)\right] = 0,$$

where $\boldsymbol{Z}_{j.}$ denotes the observations on the exogenous explanatory variables (prices, demographics and expenditure), $\boldsymbol{Z}_{j.>0}$ denotes the observations on the exogenous explanatory variables that correspond to the positive valued outcomes and having dimension $(n \times K)$.

We can define a $(5KJ\times 1)$ vector of moment conditions derived from the orthogonality condition as:

(5.1.7) $$E\left[\boldsymbol{h}_{j\ \text{Marginal}}(W,\boldsymbol{Z},\theta)\right] = E\begin{bmatrix} \boldsymbol{Z}_{j.>0}'\left(W_{j>0} - U_{j.>0} - \sigma_j \odot \dfrac{1}{\boldsymbol{\Phi}_{j>0}} \odot \boldsymbol{\phi}_{j>0}\right) \\ \boldsymbol{Z}_{j.>0}'\left(W_{j>0}^{2} - U_{j.>0} \odot E\left(W_{j>0}\right) - \sigma_j^{2}\right) \\ \boldsymbol{Z}_{j.}'\left(W_{j.} - U_{j.} \odot \boldsymbol{\Phi}_j - \sigma_j \odot \boldsymbol{\phi}_j\right) \\ \boldsymbol{Z}_{j.}'\left(W_{j.}^{2} - U_{j.} \odot E\left(W_{j.}\right) - \sigma_j^{2} \odot \boldsymbol{\Phi}_j\right) \\ \boldsymbol{Z}_{j.}'\left(W_{Binary\ j.} - \boldsymbol{\Phi}_j\right) \end{bmatrix} = 0.$$

The sample analog of the population moments displayed in (5.1.7) is

$$
(5.1.8)\qquad \boldsymbol{h}_{\text{j Marginal}}(w,\boldsymbol{z},\theta)=\begin{bmatrix}
\dfrac{\mathbf{z}_{j.>0}'}{n_j^{(1)}}\left(w_{j>0}-u_{j.>0}-\sigma_j\odot\dfrac{1}{\boldsymbol{\Phi}_{j>0}}\odot\boldsymbol{\phi}_{j>0}\right)\\
\dfrac{\mathbf{z}_{j.>0}'}{n_j^{(2)}}\left(w_{j>0}^2-u_{j>0}\odot E\left(w_{j>0}\right)-\sigma_j^{\,2}\right)\\
\dfrac{\mathbf{z}_{j.}'}{n_j^{(3)}}\left(w_{j.}-u_{j.}\odot\boldsymbol{\Phi}_j-\sigma_j\odot\boldsymbol{\phi}_j\right)\\
\dfrac{\mathbf{z}_{j.}'}{n_j^{(4)}}\left(w_{j.}^{\,2}-u_{j.}\odot E\left(w_{j.}\right)-\sigma_j^{\,2}\odot\boldsymbol{\Phi}_j\right)\\
\dfrac{\mathbf{z}_{j.}'}{n_j^{(5)}}\left(w_{Binary\,j.}-\boldsymbol{\Phi}_j\right)
\end{bmatrix}=0,
$$

where $n_j^{(i)}$ denotes the number of sample observations that correspond to the i^{th} set of moment conditions for the j^{th} commodity; the notations $\mathbf{z}_{j.}$, $w_{j.}$ and $u_{j.}$ are the same as $\boldsymbol{Z}_{j.}$, $W_{j.}$ and $U_{j.}$ respectively (sample moments notations). $E\left(w_{j.>0}\right)$ is equal to $E\left(W_{j.>0}\right)$ evaluated at sample outcomes for $w_{j.}$ and $\boldsymbol{z}_{j.}$ and at specified values for θ ,where θ represents the parameters to be estimated.

5.2 Bivariate Moments

Fahs and Mittelhammer (2007) were able to derive the first, second and third order moment conditions for the truncated bivariate normal distribution. Bivariate moment conditions help to identify and estimate the parameters involved in the covariance or the correlation structure occurring across equation errors. In addition, the bivariate moments avoid the problem of evaluating the probability of the discontinuous part in higher dimensions as numerical integration is then only required in two dimensions, which is quite accurate and fast computationally.

In the MVT model, there are J different dependent variables. There are $\left(\frac{J(J-1)}{2}\right)$ alternative pairs of decision outcomes relating to these dependent variables that can be examined in a bivariate manner. For example, in the empirical demand model there are four groups of vegetables (one being suppressed to address the issue of covariance matrix singularity), leading to six pairs of bivariate outcomes that can be represented as $(w_{1i}, w_{2i}), (w_{1i}, w_{3i}), (w_{1i}, w_{4i}), (w_{2i}, w_{3i}), (w_{2i}, w_{4i})$ and (w_{3i}, w_{4i}). For each pair, one can derive first, second, third and cross moment conditions.

Using the bivariate moment conditions (Fahs and Mittelhammer, 2007), and similar to the orthogonality conditions of the marginal moments above, and noting that for any pair of decision outcomes, $(y_{j.}, y_{k.})$, there are nine bivariate moments, we can define a vector of bivariate-type as:

$$(5.2.1) \qquad E\left[\boldsymbol{h}_{\text{Bivariate}}(\boldsymbol{W}, \boldsymbol{Z}, \theta)\right] = \begin{bmatrix} \boldsymbol{Z}_{j.}'\left(\boldsymbol{W}_{j.} - E\left(\boldsymbol{W}_{j.} \mid \boldsymbol{W}_{j.} \geq 0, \boldsymbol{W}_{k.} \geq 0\right)\right) \\ \scriptstyle for\ j\ and\ k=1,2,\ldots,J;\ j\neq k \\ \vdots \\ \boldsymbol{Z}_{j.}'\left(\boldsymbol{W}_{j.}^{2} - E\left(\boldsymbol{W}_{j.}^{2} \mid \boldsymbol{W}_{j.} \geq 0, \boldsymbol{W}_{k.} \geq 0\right)\right) \\ \scriptstyle for\ j\ and\ k=1,2,\ldots,J;\ j\neq k \\ \vdots \\ \boldsymbol{Z}_{j.}'\left(\boldsymbol{W}_{j.}^{3} - E\left(\boldsymbol{W}_{j.}^{3} \mid \boldsymbol{W}_{j.} \geq 0, \boldsymbol{W}_{k.} \geq 0\right)\right) \\ \scriptstyle for\ j\ and\ k=1,2,\ldots,J;\ j\neq k \\ \vdots \\ \boldsymbol{Z}_{j.}'\left(\boldsymbol{W}_{j.}\boldsymbol{W}_{k.} - E\left(\boldsymbol{W}_{j.}\boldsymbol{W}_{k.} \mid \boldsymbol{W}_{j.} \geq 0, \boldsymbol{W}_{k.} \geq 0\right)\right) \\ \scriptstyle for\ j\ and\ k=1,2,\ldots,J;\ j>k \\ \vdots \\ \boldsymbol{Z}_{j.}'\left(\boldsymbol{W}_{j.}^{2}\boldsymbol{W}_{k.} - E\left(\boldsymbol{W}_{j.}^{2}\boldsymbol{W}_{k.} \mid \boldsymbol{W}_{j.} \geq 0, \boldsymbol{W}_{k.} \geq 0\right)\right) \\ \scriptstyle for\ j\ and\ k=1,2,\ldots,J;\ j>k \\ \vdots \\ \boldsymbol{Z}_{j.}'\left(\boldsymbol{W}_{j.}\boldsymbol{W}_{k.}^{2} - E\left(\boldsymbol{W}_{j.}\boldsymbol{W}_{k.}^{2} \mid \boldsymbol{W}_{j.} \geq 0, \boldsymbol{W}_{k.} \geq 0\right)\right) \\ \scriptstyle for\ j\ and\ k=1,2,\ldots,J;k>j \end{bmatrix} = [0],$$

where $\boldsymbol{Z}_{j.}$ denotes the observations on the exogenous explanatory variables (prices, demographics and expenditure) and it have a dimension of *(n×K)*, all the expectation terms are derived in Fahs and Mittelhammer (2007) paper, the dimension of $E\left[h_{\text{Bivariate}}\left(W,Z,\theta\right)\right]$ is $\left(\frac{9J(J-1)K}{2}\times 1\right)$, where *J* is the number of commodities and *K* is the number of columns in the explanatory variable $\boldsymbol{Z}_{j.}$.

The sample analog of the population moments condition are:

$$(5.2.2) \qquad \boldsymbol{h}_{\text{Bivariate}}\left(\mathbf{y},\mathbf{z},\theta\right)=\begin{bmatrix} \frac{z'_{j.}}{n^{(6)}_{jk}}\left(\boldsymbol{w}_{j.}-E\left(\boldsymbol{w}_{j.}\mid \boldsymbol{w}_{j.}\geq 0,\boldsymbol{w}_{k.}\geq 0\right)\right) \\ \text{for } j \text{ and } k=1,2,\ldots,J;\ j\neq k \\ \vdots \\ \frac{\mathbf{z}'_{j.}}{n^{(7)}_{jk}}\left(w^2_{j.}-E\left(\boldsymbol{w}^2_{j.}\mid \boldsymbol{w}_{j.}\geq 0,\boldsymbol{w}_{k.}\geq 0\right)\right) \\ \text{for } j \text{ and } k=1,2,\ldots,J;\ j\neq k \\ \vdots \\ \frac{\mathbf{z}'_{j.}}{n^{(8)}_{jk}}\left(\boldsymbol{w}^3_{j.}-E\left(\boldsymbol{w}^3_{j.}\mid \boldsymbol{w}_{j.}\geq 0,\boldsymbol{w}_{k.}\geq 0\right)\right) \\ \text{for } j \text{ and } k=1,2,\ldots,J;\ j\neq k \\ \vdots \\ \frac{\mathbf{z}'_{j.}}{n^{(9)}_{jk}}\left(\boldsymbol{w}_{j.}\boldsymbol{w}_{k.}-E\left(\boldsymbol{w}_{j.}\boldsymbol{w}_{k.}\mid \boldsymbol{w}_{j.}\geq 0,\boldsymbol{w}_{k.}\geq 0\right)\right) \\ \text{for } j \text{ and } k=1,2,\ldots,J;\ j>k \\ \vdots \\ \frac{\mathbf{z}'_{j.}}{n^{(10)}_{jk}}\left(\boldsymbol{w}^2_{j.}\boldsymbol{w}_{k.}-E\left(\boldsymbol{w}^2_{j.}\boldsymbol{w}_{k.}\mid \boldsymbol{w}_{j.}\geq 0,\boldsymbol{w}_{k.}\geq 0\right)\right) \\ \text{for } j \text{ and } k=1,2,\ldots,J;\ j>k \\ \vdots \\ \frac{\mathbf{z}'_{j.}}{n^{(11)}_{jk}}\left(\boldsymbol{w}_{j.}\boldsymbol{w}^2_{k.}-E\left(\boldsymbol{w}_{j.}\boldsymbol{w}^2_{k.}\mid \boldsymbol{w}_{j.}\geq 0,\boldsymbol{w}_{k.}\geq 0\right)\right) \\ \text{for } j \text{ and } k=1,2,\ldots,J;\ k>j \end{bmatrix}=\left[0\right],$$

where as before, $n_{jk}^{(i)}$ denotes the number of sample observations involved in the i^{th} set of moment conditions for the $(j,k)^{th}$ Choice pair. The cumulative set of moment conditions, including both marginal and bivariate moments, can be specified as:

$$(5.2.3) \qquad E\boldsymbol{H}(W\ ,z\ ,\theta)=E\begin{bmatrix}\boldsymbol{h}_{\text{Marginal}}(W\ ,z,\theta)\\ \boldsymbol{h}_{\text{Bivariate}}(W\ ,z\ ,\theta)\end{bmatrix}=0$$

The sample counterpart of the population moments is $\boldsymbol{H}(w,z,\theta)=\left[0\right]$. It has a dimension of $\left(5KJ+\frac{9J(J-1)K}{2}\right)\times 1$, in our empirical analysis J=4 (number of share equations after dropping share of Onion), K=19 (prices, expenditure and demographics) the dimension of the $\boldsymbol{H}$ vector defined above will be (1406×1). Given this moment information, we need to estimate a parameter vector that has $\frac{J(J-1)}{2}+K$ elements, which are unique. The number of moment conditions is clearly greater than the number of unknown parameters we need to estimate. Hence, the set of moment equations is over determined. In the GMM approach, the parameter vector is chosen for which the sample moment conditions are as close to the zero vector in weighted Euclidian distance as possible. We use the following measure closeness, and to generate GMM estimates of the parameters of the model:

$$(5.2.4) \qquad min_{\theta}\left[Q(w_{j.},z_{j.},\theta)\right]=min_{\theta}\left[\ \boldsymbol{H}(w\ ,z\ ,\theta)'\,\hat{\boldsymbol{W}}_{opt}\ \boldsymbol{H}(w,z,\theta)\right]$$

where $\hat{\boldsymbol{W}}_{opt}$ denotes an estimate of the optimal weight matrix.

5.3 Starting Values for GMM Estimation Method

The convergence of the non-linear optimization problem in equation (5.2.4) is affected by the quality of the starting values for the parameters. To find starting values we followed a two step approach. In the first step, with the starting values obtained from the procedure described in Fahs and Mittelhammer (2007), we solved the optimization problem for all of the marginal moment conditions with symmetry, homogeneity and adding up restrictions imposed. The identity matrix was used as the weight matrix in the GMM criterion function. Then the parameter estimates obtained from this first step were used as the starting values for the second step. The optimal GMM weight matrix in the second step was defined by finding the inverse of the covariance matrix for the set of marginal and bivariate moment conditions, and then they were inserted into the weight matrix. Using this estimated optimal weight matrix and starting values derived from the first step, we solved the optimization problem for all the moment conditions, i.e., marginal and bivariate conditions, with symmetry, homogeneity and adding up restriction imposed.

6. Model Estimation Using Virtual Prices

Using the market prices (prices from the original data) to estimate the share equations in model (3.2.20) by the GMM approach will generate predicted shares with negative values. This will result in violating the non-negativity constraints. Setting those negative values (predicted shares) to zeros and solving for the virtual prices as in equation (4.8), we can form a new set of prices that contain both market price and virtual prices(p^*). Using this new set of prices we can recalculate the new predicted shares.

The parameter estimates from the GMM approach, the new set of prices containing both market and virtual prices and the new calculated predicted shares can be used in elasticity analysis.

The MVT-QUAIDS model estimation using the GMM approach along with the virtual prices concept is carried out by the following steps:

1. Obtain starting values for the parameter estimates for the MVT-QUAIDS model using the market price data (section 5.3).
2. Estimate the MVT-QUAIDS model with market price data using the GMM approach (section 5).
3. Set the negative predicted shares obtained in step 2 to zeros (section 4).
4. Calculate the virtual prices by solving the subset of the demand system for the prices equate the respective budget shares to zero.
5. Use the virtual prices calculated in step 4 together with market prices for the goods consumed at positive levels to recalculate predictions of the budget shares.
6. Use the parameter estimates together with the market and virtual prices to perform an elasticity analysis of the demand system.

7. Estimation Results Using the GMM–Virtual Prices Approach

This section is divided into two parts: the first part presents the parameter estimate of the QUAIDS using the GMM approach and illustrates the usefulness of the GMM-virtual price approach in correcting the corner point and improving predictions of the observed share values. The second part presents the effect of the *E.coli* on the consumption of salad vegetables during the non-outbreak period (January to August) versus the outbreak periods (i.e. September and November).

Note that in this paper, we used the SAS statistical software package for data recoding, cleaning and transformation. Estimation is conducted using GAUSS 8.0 programming software.

7.1 Results of the Demand for Salad Vegetables

The QUAIDS model is estimated using the GMM approach, with theoretical restrictions of adding-up, homogeneity, and symmetry imposed during estimation. Four share equations are estimated (*Tomato*, *Cabbage*, *Lettuce* and *Spinach*), where the share equation of *Onion* is dropped to deal with the inherent covariance matrix singularity in the model.

In order to estimate the QUAIDS model presented above one should minimize equation (5.2.4), which is non-linear, we used the Nelder-Meade polytope direct-search method of optimization, which is a direct search method that only requires objective function evaluations for optimization. As a result, it is robust to non-differentiability (but requires continuity) and useful for functions whose derivatives are difficult to calculate, or that cannot be calculated or approximated at all. A convergence criterion of 0.00001 was used for the difference between the maximum and the minimum objective function associated with the vertices of the Nelder-Meade simplex.

Table 5 in the Appendix presents the estimated coefficients for the QUAIDS model. At the 95% significance level, the parameters associated with the expenditure terms (β and λ) are highly significant; this demonstrates the importance of the quadratic terms in real expenditure for all five types of salad vegetables under the GMM estimation criterion. The coefficients associated with the prices (γ) are also highly significant; the interpretation of these parameters can be better explained using elasticity analysis (procedures of elasticity analysis during the non-outbreak and outbreak periods are described in section 7.2). Besides the price and expenditure parameters it is important to examine the effects of socio-demographic variables on the consumption of vegetables. Households with different socio-demographic profiles may have different consumption preferences and exhibit different consumption patterns. During the outbreak period, the coefficient estimates of the socio-demographic variables suggest that:

- *Age*: is not statistically significant for all groups of salad vegetables.
- *Sex*: indicates that female households prefer more of Tomato, Lettuce and Spinach, and less of Cabbage and Onion.
- *Children:* indicates that households with children prefer more of Tomato, Cabbage, Lettuce and Onion, and less of Spinach. This finding suggests that consumers are concern about their children safety during the outbreak period.
- *Status*: is not statistically significant for all groups of salad vegetables.
- *Income-D_1*: indicates that households with income between \$25,000 and \$50,000 prefer more of Tomato, Cabbage, Lettuce and Onion, and less of Spinach.
- *Income-D_2*: indicates that households with income greater than \$50,000 prefer more of Tomato, Lettuce, Cabbage and Onion, and less of Spinach.

Note that households with higher income consume less of Spinach and more of other vegetables than those with lower income.

- *Location*: indicates that households located in California prefer more of Tomato, Cabbage and Lettuce, and less of Spinach and Onion. This could be explained as the closer the consumer to the epidemic area, the more he/she is concern about the *E.coli* effect.
- *D-September*: indicates that households during the first outbreak prefer more of Tomato, Cabbage, Lettuce and Onion, and less of Spinach. This indicates that consumers are either concern about their safety or there is a lack of Spinach due to the product recalls during the first outbreak.
- *D-October*: indicates that households after the first outbreak prefer more of Tomato, Cabbage and Lettuce, and less of Spinach and Onion. However, the households' preferences for Tomato, Cabbage and Lettuce are lower than their preferences for the same vegetables in September (for example, Lettuce was 0.414 in September and dropped down to 0.322 in October). Households' preferences for Spinach start increasing significantly in October (from -0.8761in September to -0.0012 in October). These changes indicate that the consumers start adjusting their preferences for vegetables after the first outbreak.
- *D-November*: indicates that households during the second outbreak prefer more of Tomato, Cabbage, Lettuce and Onion and less of Spinach. Note that during the second outbreak (November) households prefer less of Spinach and more of other vegetables compared to the first outbreak

(September). The explanation of this finding could be that households during the second outbreak are more alert and more educated about the *E.coli* warnings, as the result of the first outbreak.

- *D-December*: indicates that households after the second outbreak prefer more of Tomato, Cabbage and Lettuce, and less of Spinach and Onion. Again, in this month households start adjusting their preferences for vegetables after the second outbreak.

The time trend suggests that households consume more of salad vegetables in the summer months as shown in Figure 14 in the Appendix. The consumption of Tomato, Cabbage, Lettuce and Spinach reaches the peak in the summer, while the consumption of Onion decreases. This finding coincides with the fact that households in the United States tend to consume more of salad vegetables during summer months.

In order to calculate the marginal effects of socio-demographics on consumption and elasticities of demand, we need to use the virtual price concept to correct for the predicted shares. The predicted shares obtained from the GMM estimator had negative values; the presence of these negative values would be evident that we would be violating the non-negativity constraints. Setting the predicted negative shares to zeros and solving for the virtual prices, we created a new set of prices that contained both market (goods consumed at positive values) and virtual prices (goods that are not consumed). The new predicted shares, which were calculated from those new set of prices, would reflect the appropriate corner point solution for the non-consumed goods. Table 6 in the Appendix summarizes the actual, new predicted shares, and the root mean square errors [12](RMSE). The RMSE of the actual and new predicted shares are reasonably small for all the groups of vegetables. This finding supports the usefulness of the GMM–virtual prices approach in correcting the corner point and improving predictions of the observed share values.

7.2 The Effect of the *E.coli* Outbreak on the Demand for Salad Vegetables

In order to explain the effect of *E.coli* on the consumer demand for salad vegetables, we computed and compared the elasticities of the model before the outbreak (i.e. January to August) to the elasticities of the first outbreak (September) and the second outbreak (November). Then we calculated the percentage change of marginal effects of the socio-demographic variables on the consumption of salad vegetables before and during the outbreaks.

The elasticities before the outbreak period (January to August) are calculated according to the following procedure:

- Set the dummies for September, October, November and December to zero (switch off); this will eliminate the effect of these months.
- Set the right-hand-side (RHS) demographic variables in equations (3.2.20) to the mean levels of their values in January to August.
- Predict the shares for salad vegetables using GMM approach.
- Set the negative predicted shares obtained to zeros and solve for the Virtual prices according to equation (4.8).

[12] $RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(w_i - \hat{w}_i\right)^2}{n}}$, where w_i : actuals shares, and $\hat{w}_i$: new predicted shares

- Recalculate the predicted shares using both market and virtual prices.
- Use the parameter estimates and the recalculated predicted shares to calculate the expenditure and price elasticities.

To compute the elasticities of the September outbreak period:

- Set the dummy of September to one and the dummies for October, November and September to zero (switch off).
- Set time = 9 indicating September month.
- Set the right-hand-side (RHS) demographic variables in equations (3.2.20) to the mean levels of their values in September.
- Predict the shares for salad vegetables using GMM approach.
- Set the negative predicted shares to zeros and solve for the Virtual prices.
- Recalculate the predicted shares using both market and virtual prices.
- Use the parameter estimates and the recalculated predicted shares to calculate the expenditure and price elasticities.

To compute the elasticities of the November outbreak period we follow the same procedure as for the September outbreak, the only difference is to account for November instead of September.

Tables 7 and 8 in the Appendix present the predicted shares of the non-outbreak period (January to August) and the outbreak periods (September and November). The results show that the consumption of Tomato, Cabbage, Lettuce and Onion increases by 8.6%, 32.6%, 31.8% and 0.7%, respectively, by September. While the consumption of Spinach decreases by 50.6%. Similarly, the consumption of Tomato, Cabbage, Lettuce and Onion increases by 13.5%, 34.4%, 33.4% and 1.3%, respectively, by November. While the consumption of Spinach decreases by 54.2%. The results depict that Spinach is being substituted with other salad vegetables, which suggests that consumers either concern about the *E.coli* effect or they are not able to purchase Spinach due to the product recalls. The results also show that the second outbreak has greater impact on the consumption of salad vegetables as consumers seem to be more alert and worried about the *E.coli* effect.

Table 9 in the Appendix presents the expenditure elasticities during the non-outbreak and the outbreak periods. The results indicate that during the non-outbreak period, Tomato (1.22), Lettuce (1.13) and Spinach (1.73) are expenditure elastic, while Cabbage (0.86) and Onion (0.72) are expenditure inelastic; the latter indicates that Cabbage and Onion are not responsive to consumer change in expenditure. During the first outbreak (September) the expenditure elasticities of Tomato (1.35), Cabbage (1.05) and Lettuce (1.31) are elastic, while Onion (0.83) and Spinach (0.81) are inelastic. The results for the second outbreak (November) indicate that the expenditure elasticities of Tomato (1.37), Cabbage (1.06) and Lettuce (1.51) are more elastic compared to the first outbreak, while Spinach (0.85) becomes less inelastic. Three main comparative observations can be made:

1. The expenditure elasticity of Cabbage changes from inelastic (during the non-outbreak period) to elastic (during both outbreaks), which indicates that consumers are willing to purchase more of Cabbage if their expenditures increase.

2. The expenditure elasticity of Spinach changes from elastic (during the non-outbreak period) to inelastic (during both outbreaks), which indicates that consumers are not responsive to expenditure changes. This could be because households are concern about their safety.
3. Tomato, Cabbage and Lettuce are more elastic and Spinach is more inelastic during the second outbreak compared to the first outbreak. This indicates that the November outbreak has greater impact on the expenditure elasticities than the September outbreak.

Tables 10, 11 and 12 represent the compensated price elasticities for the non-outbreak, the first outbreak and the second outbreak periods, respectively. All compensated own price elasticities are negative and have reasonable magnitudes consistent with the economic theory.

Table 10 shows that Tomato (-1.33), Lettuce (-1.421) and Spinach (-1.52) are price elastic, while Cabbage (-0.81) and Onion (-0.812) are price inelastic. The cross price elasticities contain positive and negative values indicating that the set of vegetables contains both substitute and complement goods. For example, Spinach and Cabbage are substitutes, while Spinach and Lettuce are complement goods.

Table 11 indicates that during the first outbreak the own price elasticities of Tomato (-1.43), Cabbage (-1.12) and Lettuce (-1.51) are more price elastic, while Spinach (-0.80) and Onion (-0.84) are price inelastic. The results also show that Spinach is substituted with Tomato, Cabbage and Lettuce.

Table 12 indicates that during the second outbreak the own price elasticities of Tomato (-1.51), Cabbage (-1.32) and Lettuce (-1.63) are more price elastic, while Spinach (-0.68) is more price inelastic compared to the first outbreak. The cross price elasticities also show that Spinach is substituted with Tomato, Cabbage and Lettuce.

Five main observations can be made from the elasticity analysis results:

1. The own price elasticity of Cabbage changes from price inelastic (during the non-outbreak period) to price elastic (during both outbreaks), which indicates that consumers are willing to purchase more of Cabbage if the price of Cabbage is increased.
2. The own price elasticity of Spinach changes from price elastic (during the non-outbreak period) to price inelastic (during both outbreaks), which means that consumers are not responsive to price changes during the outbreaks.
3. In the second outbreak the own price elasticities of Tomato, Cabbage, and Lettuce are more price elastic, while Spinach is more inelastic compared to the first outbreak.
4. Spinach is substituted with Tomato, Cabbage, and Lettuce during both outbreak periods.
5. The substitution of Spinach with Tomato, Cabbage, and Lettuce is greater during the second outbreak.

The summary of the results above indicates that the consumption of Spinach decreases, while the consumption of other vegetables increases significantly during the outbreak periods. The cross price elasticities results suggest that Spinach is substituted with other vegetables, such as Cabbage and Lettuce. This finding has two possible interpretations: consumers either are concerned about their health or there is a shortage in the supply of Spinach due to product recalls. The results also indicate that the second outbreak has greater impact on the consumption of salad vegetables as consumers seem to be more alert and worried about the *E.coli* effect.

Because of the non-linearity in the demand system and the fact we are dealing with a censored data generating, the values of the parameter estimates do not provide a simple measure of marginal effects of socio-demographic variable on consumption.

In this study we compared the percentage change of marginal effects of the socio-demographic variable on consumption. The comparison is computed for the non-outbreak period versus the outbreak periods.

To calculate the marginal effect of socio-demographic variables on consumption during the non-outbreak period versus the outbreak periods, the procedure is based on the following steps:

For non-outbreak period

1. Create three types of income Ranges

   i) Set $Income\text{-}D_1$=0 and $Income\text{-}D_2$=0 (Income < $25,000)

   ii) Set $Income\text{-}D_1$=1 and $Income\text{-}D_2$=0 ($25,000 <Income < $50,000)

   iii) Set $Income\text{-}D_1$=0 and $Income\text{-}D_2$=1 (Income >$50,000)

2. Set the dummies for September, October, November and December to zero (switch off); this will eliminate the effect of these months.

3. Next, set the RHS demographic variables in equations (3.2.20) to the mean levels of their values in January to August.
4. Set the appropriate demographic variable of interest to zero or one (if it is an indicator variable), and set the variable to its mean (if it is not an indicator variable).
5. Then predict the share using GMM approach for each type of income range in step 1, taking into account the variable of interest in step 4.
6. Set the negative predicted shares to zeros and solve for the Virtual prices.
7. Recalculate the predicted shares of the demographic variable of interest using both market and virtual prices.

For the September outbreak

1. Create three types of income Ranges
   i) Set $Income\text{-}D_1$=0 and $Income\text{-}D_2$=0 (Income < $25,000)
   ii) Set $Income\text{-}D_1$=1 and $Income\text{-}D_2$=0 ($25,000 <Income < $50,000)
   iii) Set $Income\text{-}D_1$=0 and $Income\text{-}D_2$=1 (Income >$50,000)
2. Set the dummy of September to one and the dummies for October, November and September to zero (switch off).
3. Set time = 9 indicating September month.
4. Then set the RHS demographic variables in equations (3.2.20) to the mean levels of their values in September.
5. Set the appropriate demographic variable of interest to zero or one (if it is an indicator variable), and set the variable to its mean (if it is not an indicator variable).

6. Then predict the share using GMM approach.
7. Set the negative predicted shares to zeros and solve for the Virtual prices.
8. Recalculate the predicted shares of the demographic variable of interest using both market and virtual prices.

<u>For the November outbreak</u>

- For November outbreak we follow the same procedure as the September outbreak, the only difference is to account for November instead of September.

<u>Comparison between the non-outbreak period versus the outbreak periods</u>

- Calculate the percentage change of the recalculated shares during the non-outbreak versus the outbreak periods.

Table 13 in the Appendix represents the percentage changes of marginal effects of income on the consumption of salad vegetables during the non-outbreak versus the outbreak periods. Three major observations can be detected:

1. Higher income households consume more of Tomato, Cabbage, and Lettuce and less of Spinach and Onion compared to households with lower income during both outbreak periods (September and November).
2. Regardless of the household income, the consumption of Tomato, Cabbage, Lettuce and Onion increases, while consumption of Spinach decreases during both outbreak periods.

3. November outbreak has greater impact on the consumption of salad vegetables, where households tend to consume less of Spinach and more of other vegetables compared to September outbreak.

As a conclusion income has a significant effect on the consumption of vegetable during the outbreak periods, where it is evident that households with higher income tend to consume less of Spinach and more of other vegetables (Tomato, Cabbage and Lettuce) than households with lower income with the exception of Onion.

The percentage changes of marginal effects of *Sex* on the consumption of salad vegetables during the non-outbreak versus the outbreak periods are presented in Table 14. The results depict four major observations:

1. Females tend to consume more of Tomato, Spinach and Lettuce, and consume less of Onion and Cabbage compared to males, during both outbreak periods. A reasonable explanation of this finding is that females consider salad as a main constituent of their diet.
1. Consumers with higher income consume more of Tomato, Cabbage and Lettuce and less of Onion and Spinach.
2. Regardless of the household gender and income, the consumption of Tomato, Cabbage, Lettuce and Onion increases, while consumption of Spinach decreases during both outbreak periods.
3. November outbreak has greater impact on the consumption of salad vegetables, where households tend to consume less of Spinach and more of other vegetables compared to the first outbreak.

Table 15 in the Appendix represents the percentage changes of marginal effects of *Location* on the consumption of salad vegetables during the non-outbreak versus the outbreak period. The results depict three major observations:

1. During both outbreak periods, Consumers located in California tend to consume more of Tomato, Cabbage and Lettuce and less of Onion and Spinach than those consumers located in Oregon and Washington. The decrease in the consumption of spinach in California indicates that consumers closer to the epidemic area are more concerns about the *E.coli* effect.
2. Consumers with higher income consume more of Tomato, Cabbage and Lettuce and less of Onion and Spinach.
3. November outbreak has greater impact on the consumption of salad vegetables, where households tend to consume less of Spinach and Onion and more of other vegetables compared September outbreak.

The percentage changes of marginal effects of *Status* on the consumption of salad vegetables during the non-outbreak versus the outbreak period are presented in Table 16. Note that all the parameters for *Status* were found to be insignificant.

Table 17 in the Appendix represents the percentage changes of marginal effects of *Children* on the consumption of salad vegetables during the non-outbreak versus the outbreak period. Four categories are identified for the number of children: 1) consumers with no children; 2) consumers with one child; 3) consumers with 2 children; 4) consumers with three or more children. Four major observations can be detected:

1. In general households with children consume more of Tomato, Cabbage, Onion and Lettuce and consume less of Spinach compared to those households without children, during both outbreak periods.
2. Consumers with higher income tend to consume less of Spinach and Onion, and more of other vegetables than those with lower income, during both outbreaks.
3. November outbreak has greater impact on the consumption of salad vegetables, where households tend to consume less of Spinach and more of other vegetables compared to September outbreak.

Finally Table 18 in the Appendix represents the elasticity of Age on the consumption of salad vegetables, where the elasticity of age is calculated by differentiating the share model with respect to *Age* in equation (3.2.20). Note that all parameter estimates for age are not significant.

As a summary, this study shows that during the outbreak periods the consumption of Spinach decreases significantly, while the consumption of Tomato, Onion, Cabbage and Lettuce increases. The decrease in Spinach consumption can be explained as: 1) The consumer is either concern about his/her health or 2) the supply of Spinach is decreased due to the recalls of spinach. The results also show that after the outbreaks (October and December) consumers adjusted their behavior and the consumption of Spinach increases significantly. The elasticity analysis shows that households substituted spinach with other leafy vegetables such as Cabbage and Lettuce; this finding justifies the increase in the consumption of Tomato, Onion, Cabbage and Lettuce.

The socio-demographic effects indicate that households with higher income consume more of Tomato, Cabbage and Lettuce and consume less of Spinach and Onion. It is also shown that consumers with children tend to consume less of Spinach and substitute Spinach with other vegetables, as consumers are concern about their children safety. The study also reveals that households close to the epidemic area (California) consume less of spinach and more of alternative vegetables than those households in Oregon and Washington. Finally the study shows that consumers are more affected by the *E.coli* in the second outbreak compared to the first outbreak; as consumers seem to be more alert and more knowledgeable about the *E.coli* effect.

8. Conclusions

The 2006 *E.coli* outbreaks affecting salad vegetables, and in particular *Spinach,* raise concerns about the impact of these outbreaks on consumers' consumption of salad vegetables. A system of demand equations for four groups of vegetables was estimated to estimate the quantitative impacts of the outbreak on consumption, and the system was based on the QUAIDS model under binding non-negativity constraints with demographic effects and allowing for non-linear Engel curves. Two prominent econometric issues were addressed in the analysis. The first resulted from the presence of censoring because budget shares (w) should not violate the non-negativity constraints and the adding up conditions of the QUAIDS model. The virtual price approach of Lee and Pitt (1986) was used to address the problem. Their empirical implementation of the virtual price approach is troubled by the computational complexity of maximizing the likelihood function. In this paper, we utilize a new GMM approach for estimating systems of non-linear censored demand equations that addresses both the computational burden of high dimensional integrals and also prevents the violation of the non-negativity constraints (Fahs and Mittelhammer, 2007). The method utilizes general marginal and bivariate moment relations that hold between explanatory variables and model noise. The estimates obtained by this approach are consistent, asymptotically normal, and near-asymptotically efficient, and are computationally relatively straightforward and tractable as the dimensionality of the model is increased. Another advantage of the GMM approach is its ability to impose side constraints on the parameters that add information to the data with the potential of further increasing the precision of the estimates.

Regarding the empirical results, the proposed model was useful for analyzing the effect of the *E.coli* outbreaks on consumers' consumption of salad vegetables. The results suggest that during the outbreak period, *Spinach* was affected the most; this was expected because of the warnings issued by the CDC and the recalls affecting this product. Households with Children were especially concerned about the outbreaks as indicated in our analysis, with the consumption of Cabbage, Tomato, Onion and Lettuce increasing significantly while the consumption of *Spinach* decreased during the outbreak periods. Further the elasticity analysis showed that consumers substitute away from spinach and towards alternative vegetables.

References

Abdul-Raouf, U.M., Beuchat, L.R. and Ammar, M.S. (1993), "Survival and growth of Escherichia cofi 0157:H7 on salad vegetables" Appl. Environ. Microbial. 59, 1999-2006

Amemiya, T. (1974), "Multivariate regression and simultaneous equation models when the dependent variable is truncated normal" Econometrica, 42, 999–1013

Amemiya, T. (1979a), "The estimation of simultaneous equation Tobit model" International Economic Review, 20, 169–181

Amemiya, T. (1979b), "Tobit models: a survey" Journal of Econometrics, 24, 3–61

Andrews, D.W.K. (1991), "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation" Econometrica, 67, 543–564

Banks, J., Blundell, R., Lewbel, A. (1997), "Quadratic Engle Curves and Consumer Demand" The review of economics and statistics, 527-539

Bean, H., Griffin, P. (1990), "Foodborne disease outbreaks in the United States, 1973-1987: pathogens, vehicles, and trends" Journal of food protection vol. 53, No.9, 804-817 (46 ref.)

Betti, G. (2000), "Quadratic Engle Curve and Household Equivalence scales: the Case of Italy 1985-1994" Working paper, London School of Economics

Blundell, R.W. and Meghir, C. (1987), "Bivariate Alternatives to the Univariate Tobit Model" Journal of Econometrics, 34, 179–200

Blundell, R., Pashardes, P. and Weber, A. (1993), "what do we learn about consumer demand patterns from micro data?" The American Economic Review, 83,570-597

Chavas, J.P., and Perali, F. (2000), "Estimation of Censored Demand Equations from Large Cross-Sectional Data" Amer. J. Agr. Econ. 82, 1022–1037

Chiang, J. and L. Lee (1992), "Discrete /continuous modes of consumer demand with binding non-negativity constrains" Journal of econometrics, 23, 59-80

Chia-Min Lin, Samuel Y. Fernando and Cheng-i Wei (1996), **"**Occurrence of *Listeria monocytogenes, Salmonella* spp., *Escherichia coli* and *E. coli* O157:H7 in vegetable salads" Food Control, Vol. 7, No. 3, pp. 135-140

Chern, W., Wang, G. (1994), "The Engle Function and Complete Food Demand System for Chinese Urban Households". China Economic Review 4: 35-47

Chung, C.E, and Goldberger, A.S. (1984), "Proportional Projections in Limited Dependent Variable Models" Econometrica , 52, 531–34

Christensen, L., Jorgensen, D. and L.J. Lau (1975), "Transcendental Logarithmic Utility Functions" American Economic Review vol. 53

Cooper, RJ. and McLaren, K.R. (1996), "A System of demand equations satisfying effectively global regularity conditions" The Review of Econometric and Statistics vol.78 No.2, 359-364

Deaton, A. (1986) "Demand Analysis, Handbook of Econometrics" vol. 3, chap. 30, pp. 1769–1839

Deaton, A. (1987), "Estimation of Owen- and cross-Price Elasticities from Households Survey Data" Journal of Econometrics 36: 7-30

Deaton, A. (1997) "The Analysis of Household Surveys: A Microeconomic Approach to Development Policy" Baltimore: Published for the World Bank, John Hopkins University Press

Deaton, A., Muellbauer, J. (1980), "An Almost Ideal Demand System" The American Economic Review 70(3): 312-326

Deaton, A., Irish, M. (1984), "Statistical model for zero expenditures in household budgets" Journal of Public Economics 23: 59-80

Decoster, A., Vermeulen F. (1998), "Evaluation of the Empirical Performance of Two-Stage Budgeting AIDS, QUAIDS and Rotterdam Models Based on Weak Separability" Department of Economics, Katholieke Universiteit Leuven

Diewert, W.E. and Wales, T.J. (1987), "Flexible Functional Forms and Global Curvature Conditions" Econometrica vol. 55, no. 1

Fahs, F. and Mittelhammer, R. (2007), "Conditional Moment Estimation of Censored Equation System" Working paper, Washington State University

Gallant, A.R. and Golub, E.H. (1984), "Imposing Curvature Restrictions on Flexible Functional Forms" Journal of Econometrics vol. 26

Greene, W.H. (1981), "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model". Econometrica, 49, 505–13

Hasan, M. and Mittelhammer, R. (2003), "Using Improved GHK Simulator in Simulated maximum Likelihood estimation for a Consumer Demand System under Non-

Negativity Constraints: An Application to Chinese Household Food Expenditure Data" Working paper, Washington State University

Hansen, L. and Singleton (1982), "Generalized Instrumental Variable Estimators of Nonlinear Rational Expectations Models" Econometrica, 50, 1269–1286

Hausman, J. (1985), "The Econometrics of Nonlinear Budget Sets" Econometrica Vol.53

Heckman, J.J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and A Simple Estimator for Such Models" Annals Econ. and Soc. Meas, 5, 475–92

Hedberg, C.W., MacDonald (1994) "Changing epidemiology of food-born disease; a Minnesota perspective" Clin. Infect. Dis. 18, 671-682

Heien, D., Wessells, C. (1990), "Demand systems estimation with micro data: A Censored regression approach" Journal of Business & Economic Statistics 8: 365-371

Lau, L.J. (1978), "Testing and Imposing Monotonicity, Convexity, and Quasi-Convexity Constraints" in Production Economics: A Dual Approach to Theory and Applications.vol. 1. Ed.by M. Fuss and D. McFadden. Amsterdam: North-Holland

Lee, L. (1993), "Multivariate Tobit models in econometrics, Handbook of statistics" vol. 2, 145–173

Lee, L. and Pitt, M. (1986), "Microeconometric demand systems with binding non-negativity constrains: the dual approach" Econometrica, 54, 1237–1242

Lee, L. and Pitt, M. (1987), "Microeconometric models of rationing, imperfect market and non-negativity constrains" Journal of Econometrics, 36, 89–110

Lee, Jong-Ying (1987), "The Demand for Varied Diet with Econometric Models for Count Data" American Journal of Agricultural Economics, 687-692

Lewbel, A. (1991), "The Rank of Demand Systems: Theory and Nonparametric Estimation" Vol. 59, No. 3, 711-730

MaCurdy, T., Green, D. and Paarsch, H. (1990), "Assessing Empirical Approaches for Analyzing Taxes and Labor Supply" Journal of Human Resources vol. 25

Maddala, G.S. (1977), "Identification and estimation problems in limited Dependent Variable Models" In A.S.Blinder and P.Friedman (eds.), Natural Resources, Uncertainty and General Equilibirum Systems: Essays in Memory of Rafael Lusky, pp.219-39. New York: Academic

McCulloch, R. and Rossi, P.E. (1994), "An Exact Likelihood Analysis of the Multinomial Probit Model" Journal of Econometrics vol. 64

Mittelhammer, R. (1996), "Mathematical statistics for economics and business" Springer-Verlag

Mittelhammer, R. and Judge, G. (1996), "Econometric Foundations" Cambridge University Press

Moschini, G., Rizzi, P.L. (1997) "La struttura dei consume delle famigilie in Italia" Published in Rivista Internationale di Scienze Sociali, Vol. 4, 449-471

Muellbauer, J. (1975) "Aggregation, income distribution and consumer demand" Review of Economic Studies, 62, 525-543

Muellbauer, J. (1976) "Community preferences and the representative consumer" Econometrica, 44, 979-999

Pashardes, P. (1993) "Bias in Estimating the Almost Ideal Demand System with Stone Index Approximation" The Economic Journal 103: 908-915

Pudney, S. (1989), "Modelling Individual Choice" Oxford: UK Basil Blackwell

Ransom, M. (1987), "A comment on consumer demand systems with binding non-negativity constrains" Econometrica 34: 355-359

Tallis, G.M. (1961), "The Moment Generating Function of the Truncated Multi-Normal Distribution" Journal of the Royal Statistical Society (Series B), 23, 223-229

Terrell, D. (1996), "Incorporating Monotonicity and Concavity Conditions in Flexible Functional Forms" Journal of Applied Econometrics vol. 11

Van Soest, A. and Kooreman, P. (1990), "Coherency of the Indirect Translog Demand System with Binding Nonnegativity Constraints" Journal of Econometrics vol. 44

Wales, T., Woodland, A. (1983), "Consumer demand systems with binding non-negativity constrains" Journal of Econometrics 21: 263-285

Appendix

Figure 1: Annual per capita vegetables consumption (1970–2002)

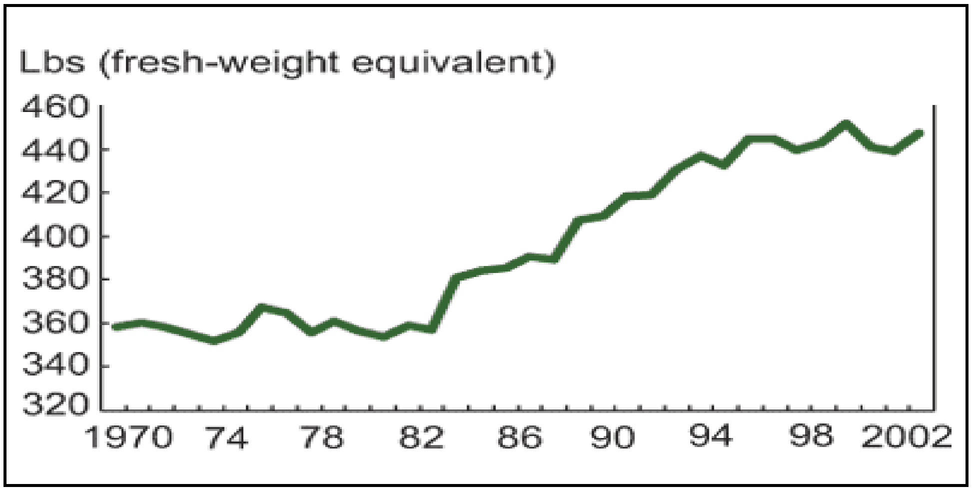


Figure 2: Geographical spread of the retail stores in California, Oregon and Washington

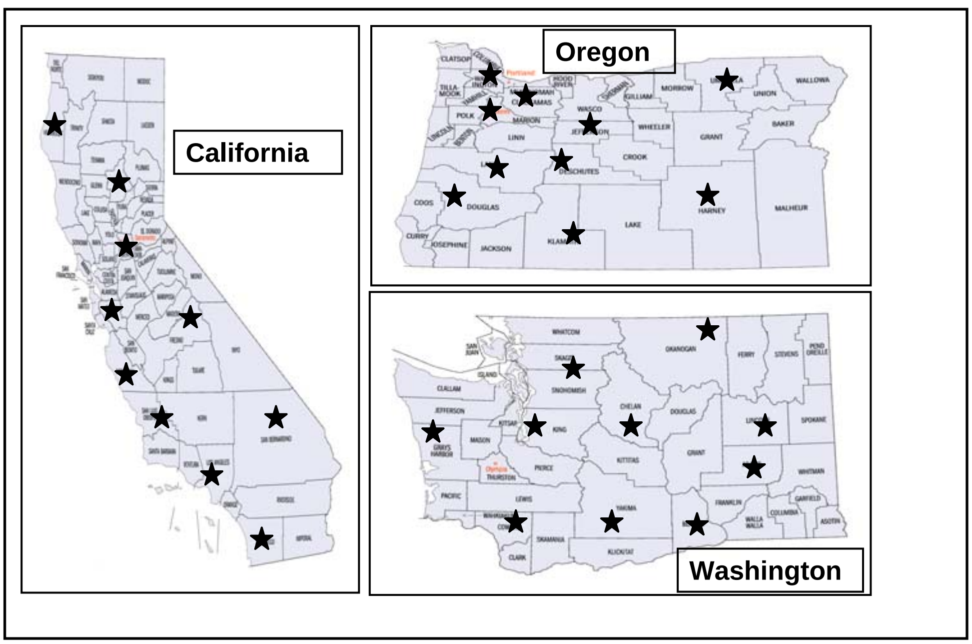


Figure 3: Empirical Cumulative Distribution Functions, Price of Onion, indicator variables: Age, Sex, Status, and Income Range

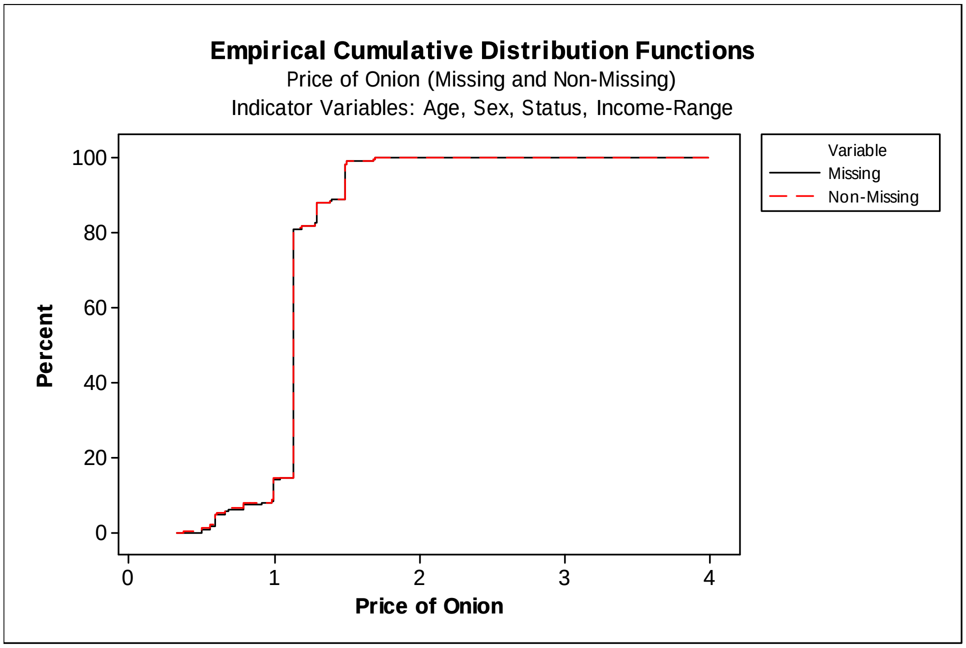


Figure 4: Empirical Cumulative Distribution Functions, Price of Onion, indicator variables: Number of Children

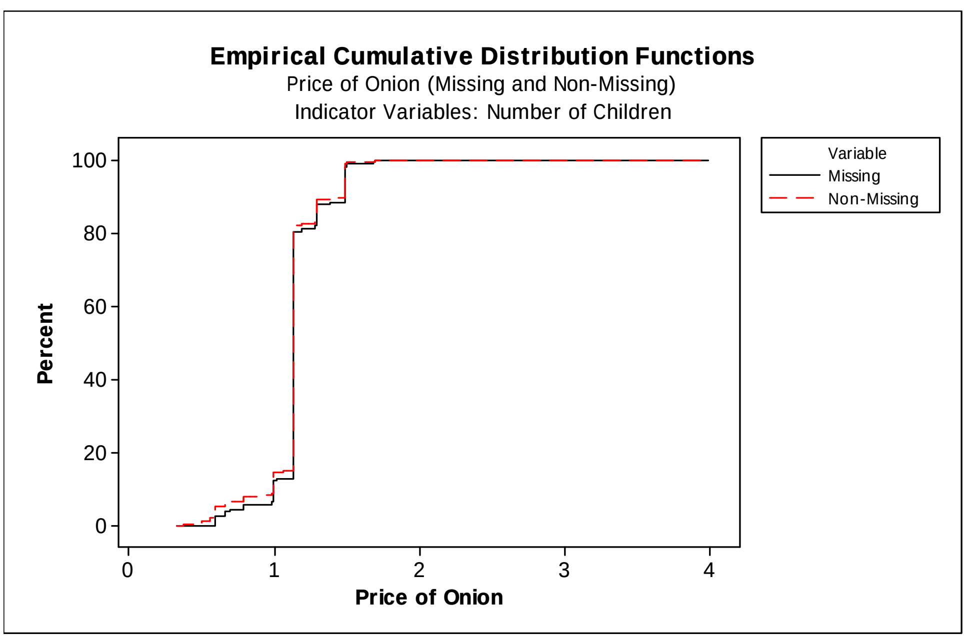


Figure 5: Empirical Cumulative Distribution Functions, Price of Cabbage, Demographic variables: Age, Sex, Status, and Income Range

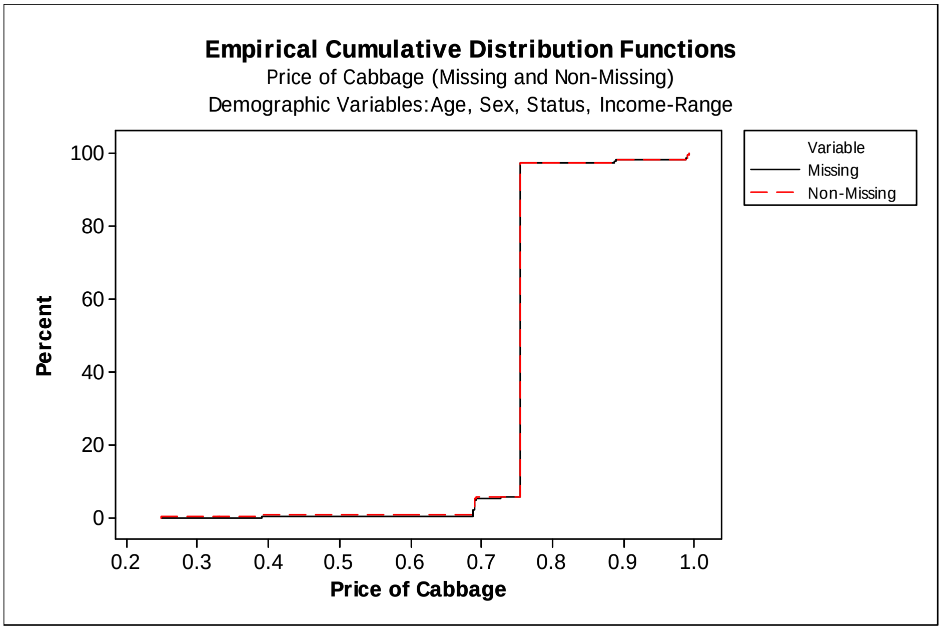


Figure 6: Empirical Cumulative Distribution Functions, Price of Cabbage, Demographic variables: Number of Children

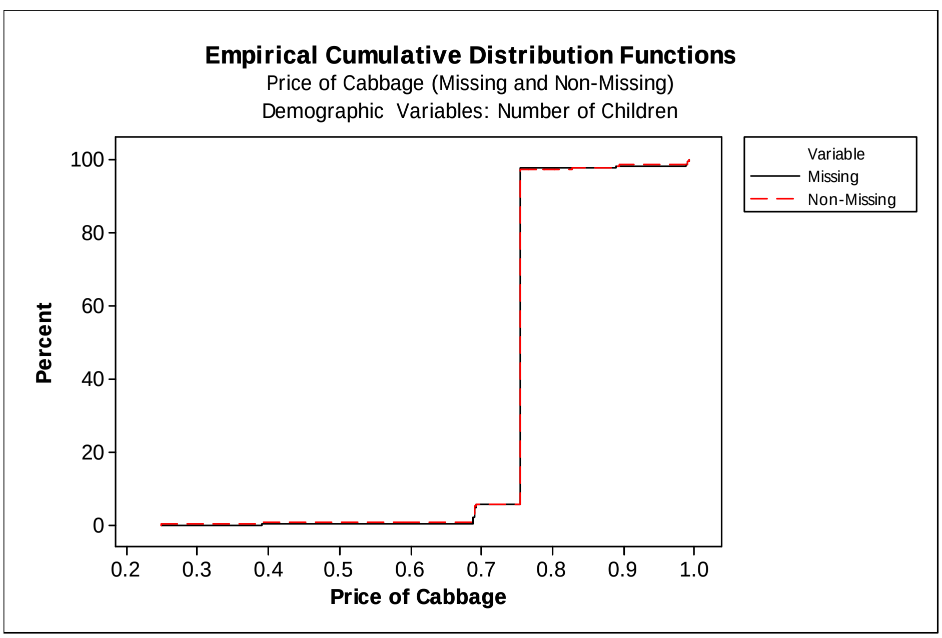


Figure 7: Empirical Cumulative Distribution Functions, Price of Lettuce, Demographic variables: Age, Sex, Status, and Income Range

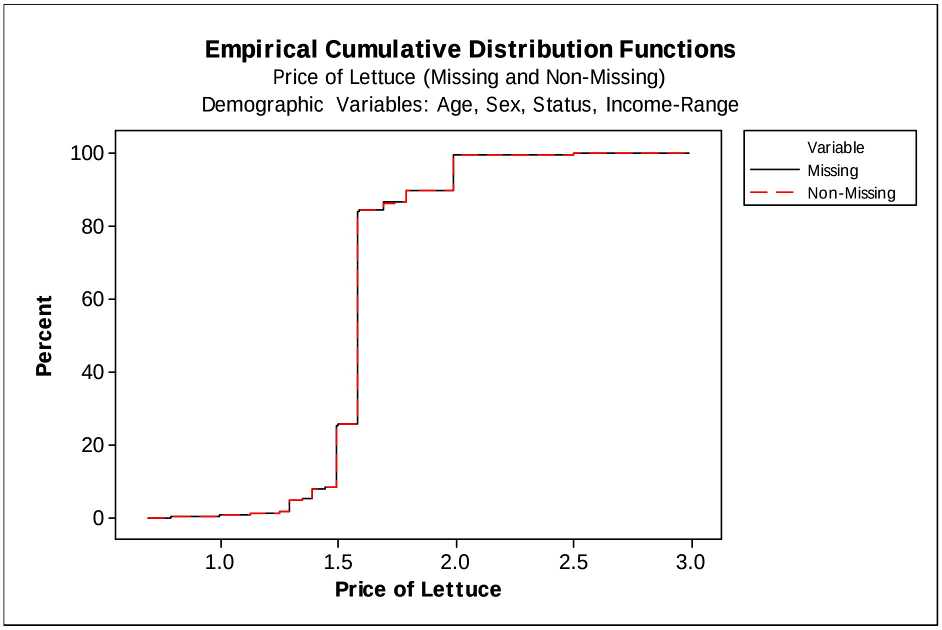


Figure 8: Empirical Cumulative Distribution Functions, Price of Lettuce, Demographic variables: Number of Children

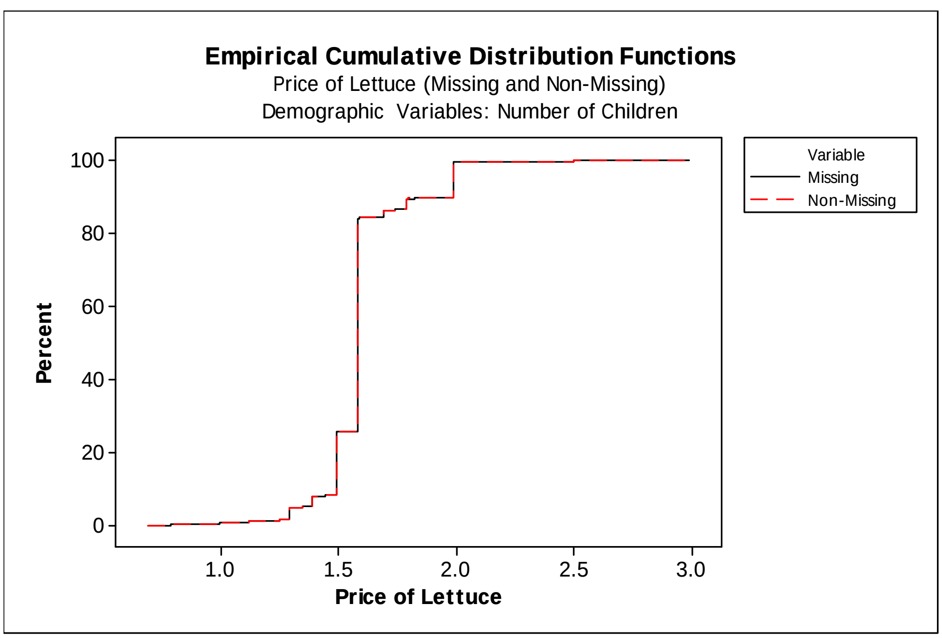


Figure 9: Empirical Cumulative Distribution Functions, Price of Spinach, Demographic variables: Age, Sex, Status, and Income Range

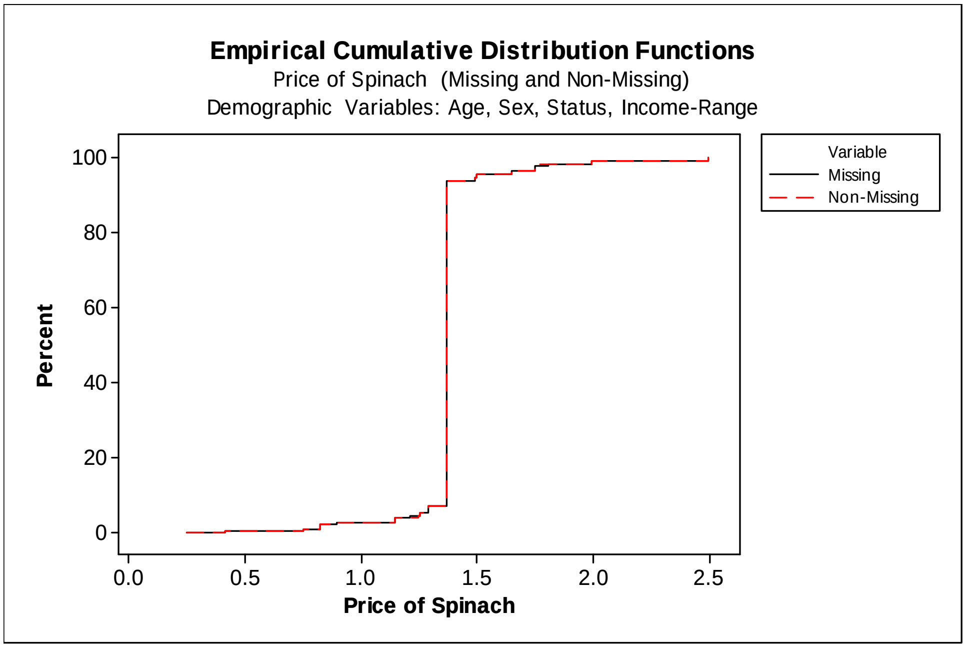


Figure 10: Empirical Cumulative Distribution Functions, Price of Spinach, Demographic variable: Number of Children

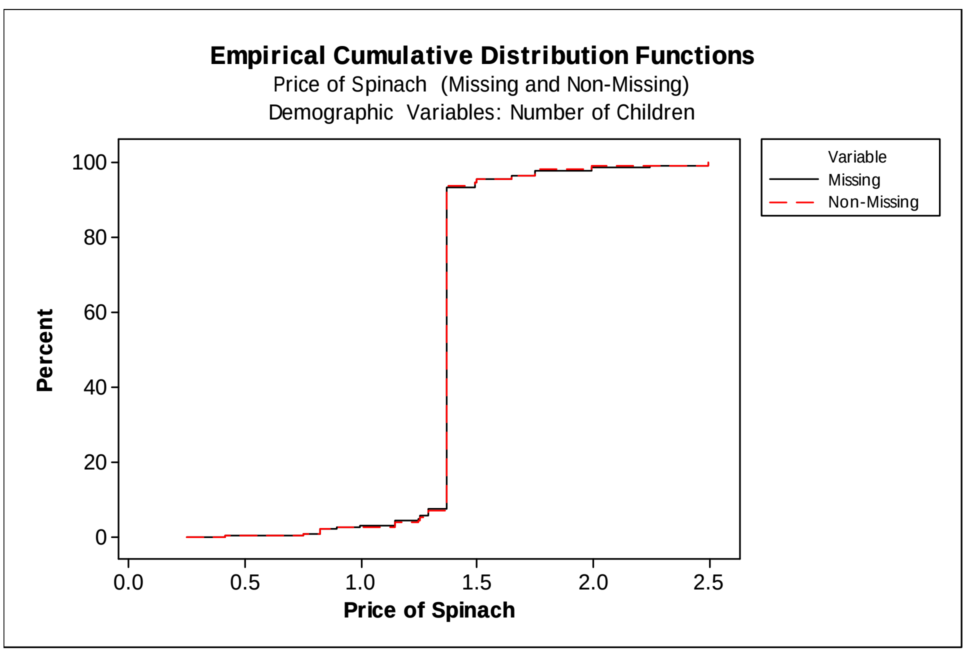


Figure 11: Empirical Cumulative Distribution Functions, Price of Tomato, Demographic variables: Age, Sex, Status, and Income Range

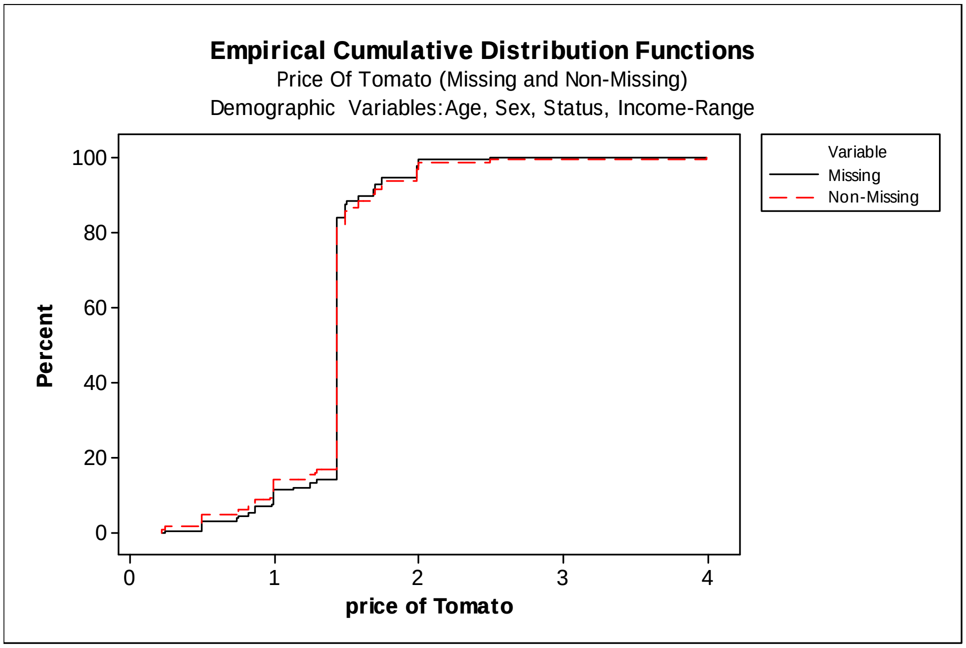


Figure 12: Empirical Cumulative Distribution Functions, Price of Tomato, Demographic variables: Number of Children

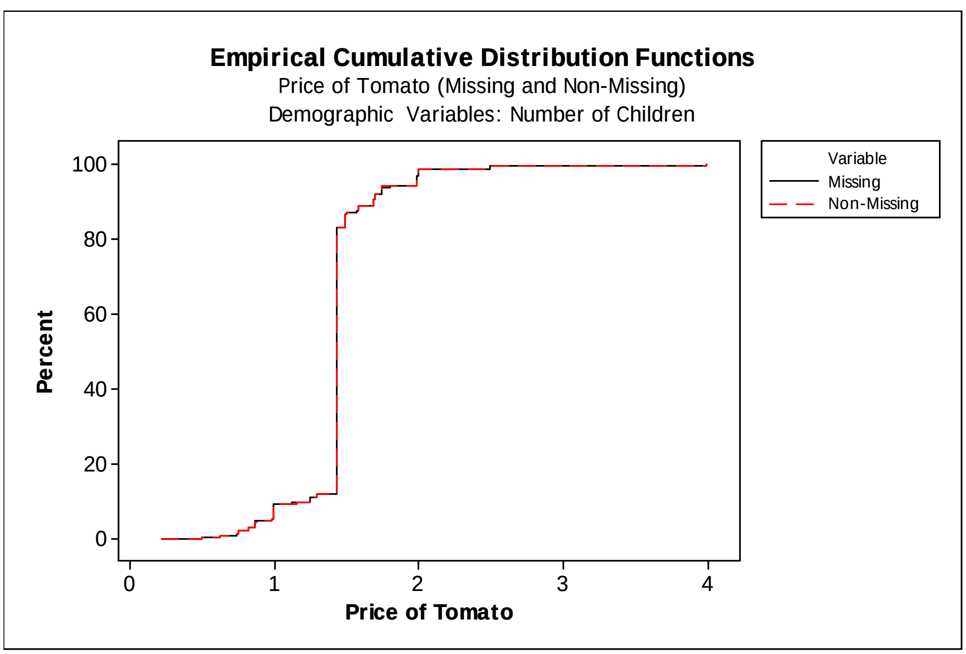


Figure 13: Example of corner solution and virtual price

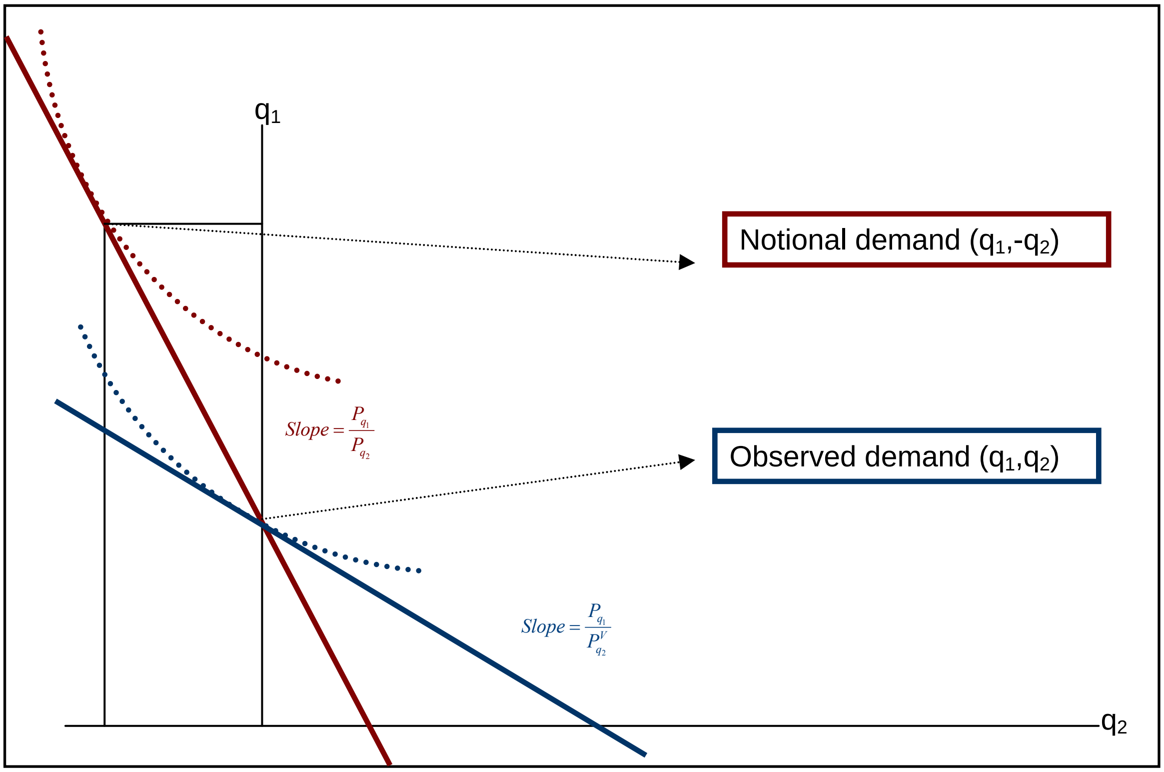
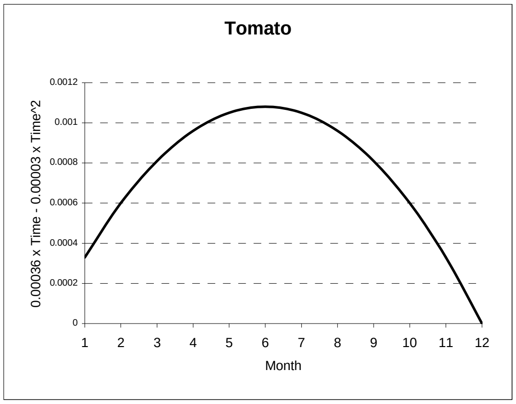
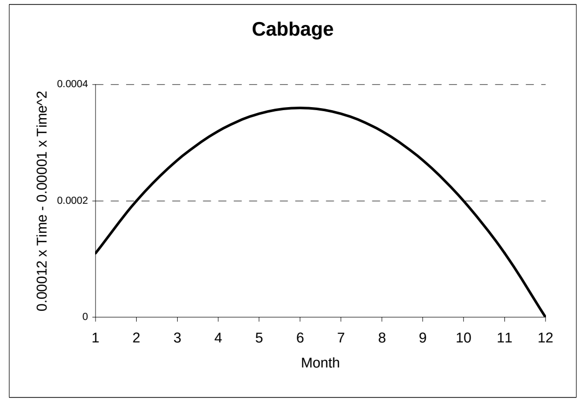
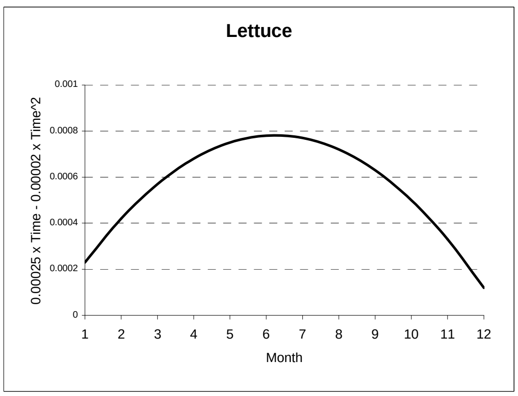
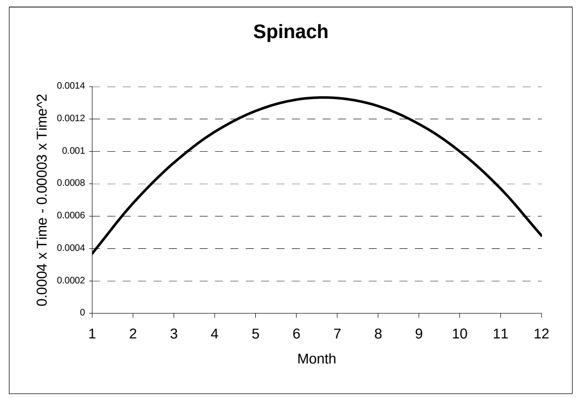


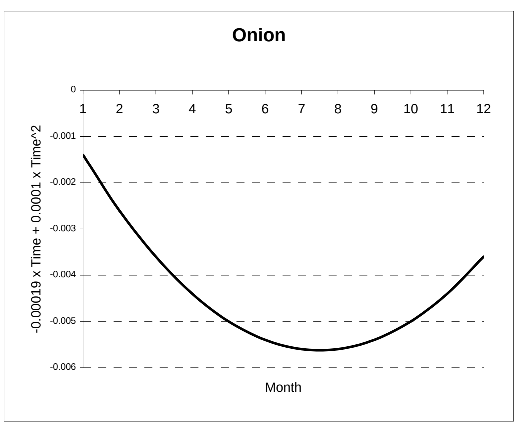
Figure 14: Time trend Graphs











Note:

- Horizontal line represents the number of months, t =1, 2… 12.
- Vertical line represents time trend function

Table 1: Vegetable groups, subgroups and types

Groups	Bulk	Packed
	Subgroups	
Tomato	148 Types	91 Types
Onion	95 Types	126 Types
Cabbage	16 Types	–
Lettuce	106 Types	43 Types
Spinach	31 Types	53 Types

Table 2: T-tests statistics for comparing the means of prices with missing observations with the means of prices with non-missing observations

	Null Hypothesis H_0: mean of price with missing observations equal to the mean of price with non-missing observations				
Price	**Age**	**Sex**	**Status**	**Income**	**Children**
			T-test values		
Tomato	0.40	0.40	0.40	0.40	0.72
Onion	0.52	0.52	0.52	0.52	0.78
Cabbage	0.48	0.48	0.48	0.48	0.65
Lettuce	0.63	0.63	0.63	0.63	0.82
Spinach	0.98	0.98	0.98	0.98	1.20

At 95% significant level, t-critical=1.96

Table 3: The Mann-Whitney Test

Null Hypothesis H_0: The two populations have identical distribution functions					
Price	**Age**	**Sex**	**Status**	**Income**	**Children**
			p-values		
Tomato	0.79	0.79	0.79	0.79	0.89
Onion	0.62	0.62	0.62	0.62	0.71
Cabbage	0.83	0.83	0.83	0.83	0.95
Lettuce	0.86	0.86	0.86	0.86	0.92
Spinach	0.91	0.91	0.91	0.91	0.98

At 95% significant level, p-critical=0.05

Table 4: Summary statistics for the variables after aggregation

Variables	Mean	Standard Deviation	Minimum	Maximum
Share-Tomato	0.2327	0.3849	0	1
Share-Onion	0.2842	0.4134	0	1
Share-Cabbage	0.0627	0.2223	0	1
Share-Lettuce	0.3380	0.4325	0	1
Share-Spinach	0.0823	0.255	0	1
Price Tomato	1.4479	0.3735	0.219	3.99
Price Onion	1.1358	0.2134	0.2983	3.99
Price Cabbage	0.7564	0.0516	0.2494	1.4948
Price Lettuce	1.6015	0.1782	0.265	2.99
Price Spinach	1.3651	0.1502	0.25	2.495
Total Expenditure	3.0488	2.5166	1.01	418.5
Age	56.1891	13.0264	18	98
Sex	0.184	0.3875	0	1
Children	0.7316	0.9561	0	7
Status	0.7842	0.4113	0	1
Income-D_1	0.3314	0.2824	0	1
Income-D_2	0.4823	0.3862	0	1
D-September	0.00833	0.2763	0	1
D-October	0.00824	0.2756	0	1
D-November	0.00866	0.2788	0	1
D-December	0.00889	0.2767	0	1
Location	0.2747	0.4464	0	1
Total number of observations			377149	

Table 5: GMM parameter estimates for vegetable salads and the p-values

GMM QUAIDS Model					
Parameter	**Variables**	**Tomato**	**Cabbage**	**Lettuce**	**Spinach**
α_i	Intercept	0.5214 (0.6120)	0.42654 (0.1521)	0.4253 (0.7521)	0.55.22 (0.4685)
	Age	-0.0498 (0.1030)	0.0354 (0.1101)	0.0135 (0.2121)	0.0244 (0.1250)
	Sex	0.0535 (0.0008)	-0.0623 (0.0000)	0.0623 (0.0221)	0.0055 (0.0020)
	Children	0.0525 (0.0001)	0.0635 (0.0010)	0.0345 (0.0000)	-0.2412 (0.0000)
	Status	-0.4210 (0.2621)	0.0312 (0.1151)	0.0312 (0.4202)	0.0313 (0.3010)
	Income-D_1 ($25,000-$50,000)	0.0345 (0.0010)	0.1232 (0.0120)	0.2352 (0.0005)	-0.4022 (0.0011)
	Income-D_2 (>$50,000)	0.1362 (0.0201)	0.2520 (0.0510)	0.2433 (0.0015)	-0.5082 (0.0150)
α_{is}	Location	0.0055 (0.410)	0.1244 (0.0000)	0.2534 (0.0021)	-0.1334 (0.0051)
	D-September	0.1343 (0.1240)	0.2232 (0.0010)	0.4146 (0.0011)	-0.8761 (0.0042)
	D-October	0.1212 (0.0214)	0.1811 (0.0225)	0.3221 (0.0113)	-0.0012 (0.0000)
	D-November	0.1553 (0.0902)	0.3012 (0.0000)	0.5021 (0.0030)	-0.9642 (0.0001)
	D-December	0.1201 (0.0011)	0.2432 (0.0010)	0.3757 (0.0250)	-0.0022 (0.0011)
	Time	0.0003 (0.0111)	0.0001 (0.0000)	0.0002 (0.0000)	0.0004 (0.0010)
	$Time^2$	-0.00003 (0.0512)	-0.00001 (0.0010)	-0.00002 (0.0000)	-0.00003 (0.0000)
	Price-Tomato	-0.1762 (0.0000)	-0.0552 (0.0001)	0.1456 (0.0000)	-0.0561 (0.0000)
γ_i	Price-Cabbage		-0.2534 (0.0000)	-0.0562 (0.0001)	-0.3356 (0.0000)
	Price-Lettuce			-0.5950 (0.0000)	-0.1422 (0.0001)
	Price-Spinach				-0.2971 (0.0000)
β_i	Expenditure	0.1321 (0.0000)	0.16254 (0.0005)	0.1032 (0.0000)	-0.1055 (0.0000)
λ_i	Expenditure-Squared	0.0856 (0.0002)	0.0765 (0.0000)	0.0720 (0.0210)	-0.2432 (0.0000)

The values in parenthesis are the p-values at 95% confidence level

Table 6: Actual, new predicted shares and RMSE

Model actual Shares, predicted shares and RMSE			
Vegetables	**Mean of Actual Share**	**Mean of new Predicted shares**	**RMSE =** $\sqrt{\frac{\sum_{i=1}^{n}(w_i - \hat{w}_i)^2}{n}}$
Tomato	0.2225	0.2095	0.0041
Onion	0.2831	0.2211	0.0033
Cabbage	0.0627	0.0632	0.0039
Lettuce	0.3390	0.3378	0.0035
Spinach	0.0723	0.0668	0.0062

Table 7: Predicted shares of salad vegetables during the non-outbreak period and the first outbreak

Predicted Shares of Vegetables during the Non-outbreak and September					
	Tomato	**Cabbage**	**Lettuce**	**Spinach**	**Onion**
Mean Predicted shares during September Outbreak	0.248	0.081	0.385	0.048	0.289
Mean Predicted shares Non-Outbreak (January to August)	0.228	0.061	0.290	0.097	0.287
% Change between Non-outbreak and September	8.661	32.67	31.83	-50.61	0.731

Table 8: Predicted shares of salad vegetables during the non-outbreak period and the Second outbreak

Predicted Shares of Vegetables during the Non-outbreak and November					
	Tomato	**Cabbage**	**Lettuce**	**Spinach**	**Onion**
Mean Predicted shares during November Outbreak	0.259	0.082	0.3896	0.0444	0.291
Mean Predicted shares Non-Outbreak (January to August)	0.228	0.061	0.290	0.097	0.287
% Change between Non-outbreak and November	13.59	34.42	33.42	-54.21	1.39

Table 9: Expenditure elasticities during the Non-outbreak and the Outbreak periods

Expenditure Elasticities during the Non-outbreak and the Outbreak periods						
		Tomato	**Cabbage**	**Lettuce**	**Spinach**	**Onion**
Non-Outbreak	Expenditure Elasticities	1.2201 (0.0000)	0.8691 (0.0232)	1.1342 (0.0010)	1.7312 (0.0021)	0.7224 (0.0320)
September Outbreak	Expenditure Elasticities	1.3542 (0.0000)	1.0561 (0.0212)	1.3132 (0.0001)	0.8125 (0.0032)	0.8312 (0.0432)
November Outbreak	Expenditure Elasticities	1.3712 (0.0020)	1.0631 (0.0352)	1.5122 (0.0001)	0.7321 (0.0001)	0.8511 (0.0032)

The values in parenthesis are the p-values at 95% confidence level

Table 10: Compensated price elasticities during the non-outbreak period

	Compensated price elasticities during the Non-outbreak period				
	Tomato	**Cabbage**	**Lettuce**	**Spinach**	**Onion**
P-Tomato	-1.333 (0.000)	0.161 (0.003)	-0.244 (0.020)	-0.432 (0.021)	0.565 (0.020)
P-Cabbage	0.151 (0.022)	-0.813 (0.010)	0.489 (0.048)	0.324 (0.030)	0.123 (0.021)
P-Lettuce	-0.321 (0.010)	0.502 (0.044)	-1.421 (0.000)	-0.342 (0.000)	-0.443 (0.032)
P-Spinach	-0.420 (0.010)	0.422 (0.020)	-0.344 (0.010)	-1.521 (0.000)	-0.484 (0.042)
P-Onion	0.522 (0.021)	0.245 (0.021)	-0.421 (0.022)	-0.424 (0.020)	-0.812 (0.000)

The values in parenthesis are the p-values at 95% confidence level

Table 11: Compensated price elasticities during the first outbreak

	Compensated price elasticities during September				
	Tomato	**Cabbage**	**Lettuce**	**Spinach**	**Onion**
P-Tomato	-1.433 (0.010)	-0.173 (0.002)	-0.341 (0.000)	0.678 (0.000)	0.568 (0.015)
P-Cabbage	-0.162 (0.002)	-1.124 (0.000)	0.805 (0.000)	0.882 (0.035)	0.264 (0.000)
P-Lettuce	-0.331 (0.021)	0.785 (0.031)	-1.512 (0.000)	0.585 (0.021)	-0.540 (0.031)
P-Spinach	0.625 (0.022)	0.922 (0.000)	0.562 (0.000)	-0.801 (0.000)	-0.501 (0.045)
P-Onion	0.554 (0.010)	0.232 (0.001)	-0.450 (0.001)	-0.432 (0.034)	-0.844 (0.000)

The values in parenthesis are the p-values at 95% confidence level

Table 12: Compensated price elasticities during the second outbreak

	Tomato	Compensated elasticities during November			
		Cabbage	**Lettuce**	**Spinach**	**Onion**
P-Tomato	-1.514 (0.000)	-0.281 (0.003)	-0.398 (0.005)	0.788 (0.015)	0.368 (0.015)
P-Cabbage	-0.252 (0.000)	-1.321 (0.000)	0.701 (0.030)	0.881 (0.035)	0.232 (0.000)
P-Lettuce	-0.441 (0.021)	0.656 (0.025)	-1.632 (0.000)	0.885 (0.042)	-0.540 (0.001)
P-Spinach	0.825 (0.011)	0.962 (0.030)	0.632 (0.010)	-0.689 (0.000)	-0.522 (0.001)
P-Onion	0.324 (0.010)	0.258 (0.001)	-0.530 (0.000)	-0.536 (0.000)	-0.954 (0.000)

The values in parenthesis are the p-values at 95% confidence level

Table 13: Percentage changes of marginal effects of income on the consumption of salad vegetables during the non-outbreak versus the outbreak periods

Percentage change of Marginal Effect of Income on the consumption of salad vegetables during the non-outbreak versus the outbreak periods		
	Tomato	
	Non-outbreak versus September	Non-outbreak versus November
Income<$25,000	5.50%	7.80%
$25,000<Income<$50,000	6.60%	8.23%
Income>$50,000	8.65%	10.20%
	Cabbage	
	Non-outbreak versus September	Non-outbreak versus November
Income<$25,000	25.00%	27.50%
$25,000<Income<$50,000	25.68%	28.62%
Income>$50,000	26.20%	29.80%
	Lettuce	
	Non-outbreak versus September	Non-outbreak versus November
Income<$25,000	30.20%	31.50%
$25,000<Income<$50,000	34.30%	34.82%
Income>$50,000	37.60%	37.62%
	Spinach	
	Non-outbreak versus September	Non-outbreak versus November
Income<$25,000	-37.33%	-39.50%
$25,000<Income<$50,000	-43.31%	-44.45%
Income>$50,000	-49.25%	-50.42%
	Onion	
	Non-outbreak versus September	Non-outbreak versus November
Income<$25,000	6.43%	7.85%
$25,000<Income<$50,000	5.32%	7.29%
Income>$50,000	4.88%	6.06%

Table 14: Percentage changes of marginal effects of sex on the consumption of salad vegetables during the non-outbreak versus the outbreak periods

Percentage change of Marginal Effect of Sex on the consumption of salad vegetables during the non-outbreak versus the outbreak periods

	Income<$25,000	$25,000<Income<$50,000	Income>$50,000
		Tomato	
		Non-outbreak versus September	
Male	7.80%	8.25%	14.20%
Female	25.25%	28.36%	32.36%
		Non-outbreak versus November	
Male	9.00%	9.45%	15.40%
Female	26.45%	29.56%	33.56%
		Cabbage	
		Non-outbreak versus September	
Male	18.54%	19.70%	13.20%
Female	8.27%	9.80%	8.56%
		Non-outbreak versus November	
Male	19.74%	20.90%	14.40%
Female	9.47%	11.00%	9.76%
		Lettuce	
		Non-outbreak versus September	
Male	17.80%	18.25%	24.20%
Female	35.25%	38.36%	42.36%
		Non-outbreak versus November	
Male	17.92%	18.37%	24.32%
Female	35.37%	38.48%	42.48%
		Spinach	
		Non-outbreak versus September	
Male	-29.56%	-33.56%	-40.53%
Female	-14.56%	-18.56%	-25.53%
		Non-outbreak versus November	
Male	-29.46%	-33.46%	-40.43%
Female	-14.46%	-18.46%	-25.43%
		Onion	
		Non-outbreak versus September	
Male	28.54%	29.70%	23.20%
Female	18.27%	19.80%	18.56%
		Non-outbreak versus November	
Male	28.84%	30.00%	23.50%
Female	18.57%	20.10%	18.86%

Table 15: Percentage changes of marginal effects of Location on the consumption of salad vegetables during the non-outbreak versus the outbreak periods

Percentage change of Marginal Effect of Location on the consumption of salad vegetables during the non-outbreak versus the outbreak periods			
	Income<$25,000	$25,000<Income<$50,000	Income>$50,000
		Tomato	
		Non-outbreak versus September	
Oregon and Washington	8.45%	9.99%	13.50%
California	15.33%	16.23%	19.45%
		Non-outbreak versus November	
Oregon and Washington	8.59%	10.13%	13.64%
California	15.47%	16.37%	19.59%
		Cabbage	
		Non-outbreak versus September	
Oregon and Washington	20.45%	21.99%	25.50%
California	27.33%	28.23%	31.45%
		Non-outbreak versus November	
Oregon and Washington	20.80%	22.34%	25.85%
California	27.68%	28.58%	31.80%
		Lettuce	
		Non-outbreak versus September	
Oregon and Washington	23.95%	25.49%	29.00%
California	30.83%	31.73%	34.95%
		Non-outbreak versus November	
Oregon and Washington	24.37%	25.91%	29.42%
California	31.25%	32.15%	35.37%
		Spinach	
		Non-outbreak versus September	
Oregon and Washington	-11.12%	-13.45%	-13.87%
California	-31.21%	-35.24%	-38.56%
		Non-outbreak versus November	
Oregon and Washington	-10.69%	-13.02%	-13.44%
California	-30.78%	-34.81%	-38.13%
		Onion	
		Non-outbreak versus September	
Oregon and Washington	17.95%	16.49%	15.06%
California	13.95%	15.49%	11.20%
		Non-outbreak versus November	
Oregon and Washington	18.06%	17.60%	16.11%
California	14.06%	13.60%	12.81%

Table 16: Percentage changes of marginal effects of Status on the consumption of salad vegetables during the non-outbreak versus the outbreak periods

Percentage change of Marginal Effect of Status on the consumption of salad vegetables during the non-outbreak versus the outbreak periods			
	Income<$25,000	$25,000<Income<$50,000	Income>$50,000
		Tomato	
		Non-outbreak to September	
Married	14.25%	16.32%	21.26%
Single	25.35%	30.21%	35.56%
		Non-outbreak to November	
Married	14.75%	16.82%	21.76%
Single	31.22%	36.08%	41.43%
		Cabbage	
		Non-outbreak to September	
Married	33.32%	38.18%	43.53%
Single	15.08%	17.15%	22.09%
		Non-outbreak to November	
Married	33.91%	38.77%	44.12%
Single	15.67%	17.74%	22.68%
		Lettuce	
		Non-outbreak to September	
Married	38.32%	43.18%	48.53%
Single	20.08%	22.15%	27.09%
		Non-outbreak to November	
Married	43.12%	47.98%	53.33%
Single	24.88%	26.95%	31.89%
		Spinach	
		Non-outbreak to September	
Married	-22.00%	-25.22%	-31.42%
Single	-33.14%	-35.24%	-38.41%
		Non-outbreak to November	
Married	-21.73%	-24.95%	-31.15%
Single	-32.87%	-34.97%	-38.14%
		Onion	
		Non-outbreak to September	
Married	19.16%	21.59%	24.27%
Single	10.04%	11.08%	13.55%
		Non-outbreak to November	
Married	21.76%	24.19%	26.87%
Single	12.64%	13.68%	16.15%

Table 17: Percentage changes of marginal effects of Children on the consumption of salad vegetables during the non-outbreak versus the outbreak periods

Number of Children	Income<$25,000	$25,000<Income<$50,000	Income>$50,000
Percentage change of Marginal Effect of Children on the consumption of salad vegetables during the non-outbreak versus the outbreak periods			
		Tomato	
		Non-outbreak versus September	
0	2025%	27.51%	29.25%
1	36.22%	38.51%	38.16%
2	35.45%	33.20%	34.98%
≥3	36.21%	36.50%	36.01%
		Non-outbreak versus November	
0	23.21%	27.61%	28.32%
1	38.21%	38.21%	40.18%
2	38.45%	38.22%	38.15%
≥3	38.21%	38.20%	38.57%
		Cabbage	
		Non-outbreak versus September	
0	16.86%	17.02%	17.12%
1	31.25%	32.55%	36.80%
2	32.01%	32.32%	32.60%
≥3	38.21%	39.89%	40.68%
		Non-outbreak versus November	
0	17.75%	18.46%	18.14%
1	32.14%	33.45%	37.49%
2	33.42%	33.22%	33.59%
≥3	38.51%	40.89%	41.37%
		Lettuce	
		non-outbreak versus September	
0	27.86%	28.02%	28.12%
1	35.25%	36.55%	37.80%
2	35.01%	35.32%	38.60%
≥3	33.21%	36.89%	31.68%
		Non-outbreak versus November	
0	28.75%	22.46%	22.14%
1	36.14%	39.45%	39.49%
2	36.42%	40.22%	41.59%
≥3	39.51%	40.89%	42.37%
		Spinach	
		Non-outbreak versus September	

0	-30.07%	-31.22%	-33.31%
1	-41.03%	-37.70%	-46.02%
2	-41.71%	-38.39%	-46.74%
≥3	-41.89%	-39.80%	-46.81%
		Non-outbreak versus November	
0	-31.50%	-31.22%	-35.31%
1	-49.03%	-45.70%	-54.02%
2	-49.71%	-46.39%	-54.74%
≥3	-49.89%	-47.80%	-54.81%
		Onion	
		Non-outbreak versus September	
0	3.05%	2.20%	2.11%
1	3.40%	2.56%	2.52%
2	3.41%	2.62%	2.53%
≥3	3.52%	2.88%	2.75%
		Non-outbreak versus November	
0	3.33%	2.83%	2.33%
1	3.52%	2.62%	2.52%
2	4.10%	2.60%	2.59%
≥3	4.11%	3.61%	3.11%

Table 18: Elasticity of age on the consumption of salad vegetables during the non-outbreak versus the outbreak periods

Elasticity of age on the consumption of salad vegetables during the non-outbreak versus the outbreak periods					
	Tomato	**Cabbage**	**Lettuce**	**Spinach**	**Onion**
		Non-outbreak			
Income<$25,000	1.321	1.421	1.650	0.989	1.532
$25,000<Income<$50,000	1.465	1.321	1.685	0.882	1.632
Income>$50,000	1.365	1.334	1.562	1.131	1.563
		September			
Income<$25,000	1.442	1.542	1.771	1.11	1.653
$25,000<Income<$50,000	1.586	1.442	1.806	1.003	1.753
Income>$50,000	1.486	1.455	1.683	1.252	1.684
		November			
Income<$25,000	1.456	1.556	1.785	1.124	1.667
$25,000<Income<$50,000	1.6	1.456	1.82	1.017	1.767
Income>$50,000	1.5	1.469	1.697	1.266	1.698

Estimation of Censored Regression Models Based on the Minimum Power Diversion Class of Probability Distributions

Faysal Habib.Fahs

School of Economic Sciences
Washington State University
Pullman, WA 99163-4741
(509)-339-3376
Email: faysal74@wsu.edu

ESSAY THREE

Estimation of Censored Regression Models Based on the Minimum Power Diversion Class of Probability Distributions

Abstract

This paper utilizes the Minimum Power Divergence (MPD) class of probability distributions to estimate censored regression models. Based on the minimization of the Cressie-Read (CR) power divergence function, we are able to implement an estimator that requires less priori model structure than conventional parametric models such as the Tobit estimator. Our estimator assumes that the distribution of the noise term is neither based on, nor restricted to, the conventional parametric families (normal, logistic) and suggests a range of CDFs that is based on the MPD principle. The paper pursues two estimation approaches to estimate censored regression model using the MPD principle: 1) Generalized Method of Moments (GMM) and 2) Maximum Likelihood approach (ML). Monte Carlo sampling experiments suggest that the estimators within the CR class will be more robust than conventional methods often used in empirical practice while also producing estimation precision that rivals the tightly specified parametric approaches in the event that the data generating distributional assumptions underlying the parametric specifications are true.

1. Introduction

In statistical applications, we often encounter a situation where the dependent variable is only observable under certain conditions. Censoring occurs when we observe the independent variable for the entire sample, but for some sample outcomes we observe only limited information about the dependent variable. The classic example of censoring is Tobin's (1958) study of household expenditures, where a consumer maximizes his or her utility by purchasing durable goods under the constraint that total expenditure does not exceed income. Many other examples of censored outcomes can be found: hours worked by wives (Quester and Green, 1982), scientific publications (Stephan and Levin, 1992), extramarital affairs (Fair, 1978), foreign trade and investment (Eaton and Tamura, 1994), austerity protests in Third World countries (Walton and Ragin, 1990), damage caused by a hurricane (Fronstin and Holtmann, 1994) and in addition to a wide range of examples that researchers encounter in both economic and econometric analysis.

Statistical inference for data containing many zero values has been investigated in various ways, and analysts face numerous choices in selecting an appropriate statistical model and associated estimation criterion. One of two estimation methods has been most often followed in the empirical literature when estimating censored regression models. By far the most popular method has been the parametric approach, where specific assumptions are made regarding the distribution underlying data sampling process. For example, in the Tobit model one assumes that the density function of the disturbance in the censored regression model follows a normal distribution. Any misspecification of the parametric assumptions generally leads to inconsistent estimators.

In empirical applications the data sampling process for observations on a censored random variable Y_i is generally based on the behavior of a latent variable defined by[1]:

(1.1) $$Y_i^* = \boldsymbol{x}_{i.}\boldsymbol{\beta} + \varepsilon_i^* \quad for\ i = 1,\ldots,n$$

where Y_i^* is the i^{th} observation on the latent variable latent variables, $\boldsymbol{x}_{i.}$ is a $(1\times k)$ row vector of explanatory variables, $\boldsymbol{\beta}$ is $(k\times 1)$ vector of the unknown parameters , and ε_i^* is the i^{th} noise term. The $\boldsymbol{x}_{i.}\boldsymbol{\beta}$ in equation (1.1) can be replaced (here and elsewhere) by more general functional form $g(\boldsymbol{x}_{i.},\boldsymbol{\beta})$ if the effect of the response variable on the latent variable is thought to be nonlinear. Assuming the censoring occurs at zero, the censored regression model is generally written as the following function of the latent variable outcome:

(1.2) $$Y_i = \begin{Bmatrix} Y_i^* \\ 0 \end{Bmatrix} \ if \begin{Bmatrix} Y_i^* > 0 \\ Y_i^* \le 0 \end{Bmatrix} \Leftrightarrow \begin{Bmatrix} \varepsilon_i^* > -\boldsymbol{x}_{i.}\boldsymbol{\beta} \\ \varepsilon_i^* \le -\boldsymbol{x}_{i.}\boldsymbol{\beta} \end{Bmatrix},\ for\ \ i = 1,\ldots,n$$

Traditionally β has been estimated by parametric methods that prescribe the density function for ε_i^*, almost invariably as normal. If the density of ε_i^* is unknown, semi-parametric estimation of beta (β) becomes of interest. In literature the researchers' objective to find the best estimator of β and to achieve the semi-parametric efficiency bound. This bound is a function of what is assumed about the density of ε_i^* and the relation of $\boldsymbol{x}_{i.}$ to ε_i^*. There will be different bounds if (i) densities of $\boldsymbol{x}_{i.}$ and ε_i^* are independently distributed (Cosslett, 1987), (ii) the density of ε_i^*, conditional upon $\boldsymbol{x}_{i.}$ has

[1] Note that for the entire paper, X or Y are Scalar random, vectors or matrixes are denoted by bold capital letters **X** or **Y**, A subscripted index on a vector indicates particular row or column elements of the vector ($\boldsymbol{X}_{i.}$ denotes the i^{th} row of **X**), Observed outcomes or fixed values are denoted by lower case letters.

zero median (Newey and Powell, 1990b), and (iii) the density of ε_i^*, conditional upon $\boldsymbol{x}_{i.}$, is symmetrically distributed (Powell, 1986b; Newey, 1990a). The most recent consistent estimators, semi-parametric and non-parametric estimation, have been proposed which allow for much weaker restrictions on the error term (ε_i^*), such as constant conditional quantiles (Powell, 1986, 1990; Nawata, 1990; Kahn and Powell, 1999; Buchinsky and Hahn, 1998), conditional symmetry (Powell, 1986b) and independence between the errors and regressors (Horowitz, 1989; Moon, 1989; Honoré and Powell, 1994). Some of the problems of semi-parametric and non-parametric estimation are small sample size and limited inference power and accuracy.

Mittelhammer and Judge (2007) introduced a new and broad class of estimators based on minimization of the Cressie-Read power divergence measure for binary choice models. They focus on information-theoretic methods that account for inherent model and data uncertainty and lead to a wide class of CDF's, with associated estimators that are minimally divergent from the type and level of information known about the data generating process.

Pursuing the estimation of censored regression model and assuming that the error term ε_i^* in equation (1.1) follows Mittelhammer and Judge's MPD class of distributions. We estimated the censored regression model using two different approaches, the Generalized Method of Moments (GMM) and the Maximum Likelihood (ML) estimation methods.

The remainder of the paper is structured as follow: Section 2 is an overview of the Minimum Power Divergence Class of CDFs and estimators for the binary choice models (Mittelhammer and Judge, 2007). Section 3 describes how we utilize the MPD distributions assumptions to estimate censored regression models using the GMM and the ML approach. Section 4 presents the comparisons between standard Tobit models and the new estimation method proposed in this paper under the assumption of normality and under a skewed gamma distribution. The final section presents a summary and conclusions derived from our findings.

2. Overview of MPD Class of CDFs for Binary Choice Model

In the context of binary response models described in Mittelhammer and Judge (2007), it is assumed that on trial $i = 1, 2, \ldots, n,$ one of two alternatives is observed to occur for the independent binary random variables $\{Y_1, \ldots, Y_n\}$ having $p_i, i = 1, \ldots, n,$ as their respective probabilities of success. The data sampling process for the binary random variable Y_i is specified in terms of the latent variable Y_i^*, as

(2.1) $$Y_i^* = \boldsymbol{x}_{i.}\boldsymbol{\beta} + \varepsilon_i^* \quad for\ i = 1, \ldots, n$$

where Y_i^* is the i^{th} observation on the latent variable latent variables, $\boldsymbol{x}_{i.}$ is a $(1 \times k)$ row vector of explanatory variables, β is $(k \times 1)$ vector of the unknown parameters , and ε_i^* is the i^{th} noise term. $Y_i \equiv I(Y_i^* > 0), i = 1, \ldots, n,$ are independent Bernoulli random variables, $I(A)$ is an indicator function that takes the value “1” when condition (A) is true and takes the value “0” otherwise. Given equation (2.1) the Bernoulli probability p_i, is defined as:

(2.2) $$p_i = P(Y_i = 1) = P(\varepsilon_i^* > -\boldsymbol{x}_{i.}\boldsymbol{\beta}) = 1 - G(-\boldsymbol{x}_{i.}\boldsymbol{\beta}) = G_*(-\boldsymbol{x}_{i.}\boldsymbol{\beta}),$$

where $G(\bullet)$ is the cumulative distribution function (CDF) of the noise term (ε_i^*) of the latent variable in equation (1.1), and $G^*(\bullet)$ is the complement of that CDF.

The authors were able to identify a class of CDFs that satisfy basic and generally applicable conditions relating to the binary response model. Three conditions were identified as:

1) A generally applicable nonparametric statistical model specification of the Bernoulli outcomes reflecting signal and noise components.

They assumed that the vector of Bernoulli random variables, **Y**, can be modeled by the very general representation:

(2.3) $$\mathbf{Y} = \mathbf{p} + \boldsymbol{\varepsilon}$$

where $E(\boldsymbol{\varepsilon}) = 0$ and $\mathbf{p} \in \times_{i=1}^{n}(0,1),$ and the expectation of **Y** is some mean vector **p**.

2) An orthogonality condition between the response variables and the noise component.

In the context of equation (2.3) the Bernoulli probabilities involve known covariate information in the form of associated response variables, **X** (in our case), with dimension of $(n \times k)$. If the probabilities **p** could be given an explicit parametric functional form, $\mathbf{p} = \mathbf{G(x\boldsymbol{\beta})}$ with $\mathbf{G(\bullet)}$ being some CDF, then the moment equations can be specified as $\mathrm{E}\left[\mathbf{X'}\left(\mathbf{Y} - \mathbf{G(X\boldsymbol{\beta})}\right)\right] = \mathbf{0}$ and the empirical moments as

(2.4) $$n^{-1}\left[\boldsymbol{x}'\left(\boldsymbol{y} - \mathbf{p}(\boldsymbol{x})\right)\right] = 0$$

The number of unknowns in (2.4) is greater than the estimating equations thus the system of equations is undetermined regarding a unique interior solution for the probability vector **p**.

3) A minimum distance divergence measures between members of the CDF class and reference distributions for the Bernoulli probabilities underlying the binary response model.

In order to estimate the underdetermined **p**-vector in (2.4), the authors adopted the Cressie-Read (CR) family of power divergence measures (Cressie and Read (1984); Read and Cressie (1988); Mittelhammer, et al., (2000)) as their estimation objective function

(2.5) $$\min_{pij's}\left\{\sum_{i=1}^{n}\left(\frac{1}{\gamma(\gamma+1)}\sum_{j=1}^{2}p_{ij}\left[\left(\frac{p_{ij}}{q_{ij}}\right)^{\gamma}-1\right]\right)\right\}$$

subject to:

(2.6) $$\sum_{j=1}^{2}\mathrm{p}_{ij}=1,\mathrm{p}_{ij}\geq 0,\ \forall i,j$$

(2.7) $$\sum_{j=1}^{2}\mathrm{q}_{ij}=1,\mathrm{q}_{ij}\geq 0,\ \forall i,j$$

(2.8) $$n^{-}(\boldsymbol{x}'(\boldsymbol{y}-\mathbf{p}))=0$$

where the Bernoulli process underlying the binary outcomes for each observation is characterized by the probabilities $\mathbf{p}_{i.}=[\mathrm{p}_{i1}\ \mathrm{p}_{i2}]$, the $\left(\frac{1}{\gamma(\gamma+1)}\sum_{j=1}^{2}p_{ij}\left[\left(\frac{p_{ij}}{q_{ij}}\right)^{\gamma}-1\right]\right)$ represents the (CR) power divergence of the of the Bernoulli probabilities from a reference distributions $\mathbf{q}_{i.}=[q_{i1}\ q_{i2}]$. The summation term can be interpreted as defining the conditional empirical expectation of the divergence between the power ratio $\left(\frac{p}{q}\right)^{\gamma}$ and the value of 1. The choice of $\gamma\in(-\infty,\infty)$ determines an entire family of measures of divergence between **p** and **q** probability distributions. The constraints (2.6) and (2.7) are necessary conditions required for the p_{ij}'s and q_{ij}'s to be interpreted as probabilities. Constraint (2.8) is the empirical implementation of the moment condition $\mathrm{E}\left[\mathbf{X}'(\mathbf{Y}-\mathbf{p})\right]=\mathbf{0}$.

Defining $p_i \equiv p_{i1}$ *and* $q_i \equiv q_{i1}$, they set the following optimization problem as

(2.9) $$L(\mathbf{p},\boldsymbol{\lambda})=\sum_{i=1}^{n}\left[\frac{1}{\gamma(\gamma+1)}\left[p_i\left(\frac{p_i}{q_i}\right)^{\gamma}+(1-p_i)\left(\frac{1-p_i}{1-q_i}\right)^{\gamma}-1\right]\right]+\boldsymbol{\lambda}'\boldsymbol{x}'(\boldsymbol{y}-\mathbf{p})$$

subject to

$$0\le p_i,q_i\le 1,\ \forall i.$$

where the premultiplier n^{-1} on the moment constraints is suppressed. The representation of the p_i's as functions of the response variables and Lagrange multipliers can be defined by solving the first order conditions with respect to **p** that are adjusted by complementary slackness conditions of Kuhn-Tucker theory in the event that inequality constraint are binding. The $p_i(\boldsymbol{x}_i\boldsymbol{\lambda})$ is represented as:

(2.13)

$$p_i(\boldsymbol{x}_i\boldsymbol{\lambda})=\arg_{pi}\left[\left(\left(\frac{p_i}{q_i}\right)^{\gamma}-\left(\frac{1-p_i}{1-q_i}\right)^{\gamma}\right)=\boldsymbol{x}_i\boldsymbol{\lambda}\gamma\right]\quad \textit{for } \gamma<0 \textit{ and } \boldsymbol{x}_i\lambda\in\mathbf{R}$$

$$=\arg_{pi}\left[\left(\ln\left(\frac{p_i}{q_i}\right)-\ln\left(\frac{1-p_i}{1-q_i}\right)\right)=\boldsymbol{x}_i\boldsymbol{\lambda}\right]\quad \textit{for } \gamma=0 \textit{ and } \boldsymbol{x}_i\lambda\in\mathbf{R}$$

$$=\arg_{pi}\left\{\begin{matrix}1\\ \left[\left(\left(\frac{p_i}{q_i}\right)^{\gamma}-\left(\frac{1-p_i}{1-q_i}\right)^{\gamma}\right)=\boldsymbol{x}_i\boldsymbol{\lambda}\gamma\right]\\ 0\end{matrix}\right\}\quad \textit{for } \gamma>0 \textit{ and } \boldsymbol{x}_i\boldsymbol{\lambda}\in\left\{\begin{matrix}\ge\gamma^{-1}q_i^{-\gamma}\\ \left(-\gamma^{-1}(1-q_i)^{-\gamma},\gamma^{-1}q_i^{-\gamma}\right)\\ \le-\gamma^{-1}(1-q_i)^{-\gamma}\end{matrix}\right\}$$

A unique solution for $p_i(\boldsymbol{x}_i\boldsymbol{\lambda})$ necessarily exists by the strict monotonicity of either

$$\eta(p_i)=\left(\left(\frac{p_i}{q_i}\right)^{\gamma}-\left(\frac{1-p_i}{1-q_i}\right)^{\gamma}\right) \textit{ or } \eta(p_i)=\ln\left(\frac{p_i}{q_i}\right)^{\gamma}-\ln\left(\frac{1-p_i}{1-q_i}\right)^{\gamma} \textit{ in } p\in(0,1),\ \textit{for}\ \gamma\neq 0 \textit{ or } \gamma=0,$$

respectively. Note that because the p_i is monotonically increasing function of the $\boldsymbol{x}_i\boldsymbol{\lambda}$. This implies that the $p_i\left(\boldsymbol{x}_i\boldsymbol{\lambda}\right)$ functions can be legitimately interpreted as cumulative distribution functions CDFs on the respective supports $\boldsymbol{x}_i\boldsymbol{\lambda}$. Given the first order condition with respect to **p** and given that the probability model is specified correctly, the authors were able to prove that $\lambda=\beta$.

As a result of the above three conditions the authors were able to generate to a wide class of CDFs, with associated estimators, that are minimally divergent from the type and level of information known about the data generating process. This class of CDFs subsumes the logistic distribution as a special case.

In general the authors showed that the MPD class of distributions contained a wide rage of symmetric and skewed probability density functions (when changing γ). Each family of distributions contained within the MPD class exhibited a reflective property around the origin. As for the MPD estimator of the binary choice model, it was shown that if the model distribution was correctly specified then the MPD estimator will consistently estimate β. If the model distribution was not specified correctly then the MPD estimator will generally be inconsistent. This result is similar to the case of a misspecified ML problem. Beside consistency the authors were able to attain the asymptotic normality of β.

3. Estimation of Censored Regression Models Using the MPD Class of CDFs

For estimating the censored regression model in equation (1.1), we assumed that the error term ε_i^* to be one of the members of the large and varied class of MPD distributions.

Analogous to Mittelhammer and Judge (2007), the probability for of observing a positive value for Y_i, in the context of censored regression model (1.1), was defined as:

(3.1) $$p_i = P(Z=1) = P\left(\varepsilon_i^* > -\boldsymbol{x}_{i.}\beta\right) = 1 - G(-\boldsymbol{x}_{i.}\beta) = G_*\left(-\boldsymbol{x}_{i.}\beta\right),$$

where Z is an indicator variable that takes the value of 1 (Z=1 if $Y_i > 0$) and the value of 0 (Z=1 if $Y_i = 0$). $G(\bullet)$ is the cumulative distribution function (CDF) of the noise term, ε_i^* the latent variable in equation (1.1), and $G^*(\bullet)$ is the complement of that CDF.

3.1 GMM Estimation of the Censored Regression Model Assuming MPD distributions

Assuming that the distribution of ε_i^* in equation (1.1) to be one of the members of the large and varied class of MPD distributions. We begin by identifying the expected value of the limited dependent variable as:

(3.1.1) $$\begin{aligned} E(Y_i) &= P(y_i = 0)\bullet E(y_i \mid y_i = 0) + P(y_i > 0)\bullet E(y_i \mid y_i > 0) \\ &= P(y_i = 0)\bullet 0 + P(y_i > 0)\bullet E(y_i \mid y_i > 0) \\ &= P(y_i > 0)\bullet E(y_i \mid y_i > 0) \end{aligned}$$

The conditional expectation of the dependent variable is represented as:

(3.1.2) $$E(y_i \mid y_i > 0) = (X_i\beta) + E\left(\varepsilon_i^* \mid \varepsilon_i^* > -X_{i.}\beta\right)$$

The probability of the dependent variables being positive can be represented as:

(3.1.3) $$P(y_i > 0) = P\left(\varepsilon_i^* > -X_{i.}\beta\right) = P(Z=1) = p_i = F\left(X_i\beta; q, \gamma\right)$$

Thus equation (3.1.1) can be rewritten as:

(3.1.4) $$E(Y_i) = P\left(\varepsilon_i^* > -X_{i.}\beta\right)\bullet\left(X_{i.}\beta\right) + P\left(\varepsilon_i^* > -X_{i.}\beta\right)\bullet E\left(\varepsilon_i^* \mid \varepsilon_i^* > -X_{i.}\beta\right)$$

or

(3.1.5) $$E(Y_i) = F\left(X_{i.}\beta; q, \gamma\right)\bullet\left(X_{i.}\beta\right) + F\left(X_{i.}\beta; q, \gamma\right)\bullet E\left(\varepsilon_i^* \mid \varepsilon_i^* > -X_{i.}\beta\right)$$

where $F\left(X_i\beta; q, \gamma\right)$ represents the CDF complement of the noise term.

A necessary condition for using GMM in the current problem context is that $E\left(\varepsilon_i^* \mid \varepsilon_i^* > -X_{i.}\beta\right)$ should be expressible as a function of $\{\beta, q, \gamma\}$, where γ *and q* are the two parameters of the MPD class of distributions. We can specify the conditional expectation of the error term as:

(3.1.6) $$E\left(\varepsilon_i^* \mid \varepsilon_i^* > -X_{i..}\beta\right) = \frac{\int_{-X_{i.}\beta}^{\infty} \varepsilon_i^* f\left(\varepsilon_i^*\right) d\varepsilon_i^*}{\int_{-X_{i.}\beta}^{\infty} f\left(\varepsilon_i^*\right) d\varepsilon_i^*} = \frac{\int_{-X_i\beta}^{\infty} w f(w) dw}{F\left(X_{i.}\beta; q, \gamma\right)}$$

Mittelhammer and Judge (2007) have shown that the general functional representation of the PDFs contained in the Minimum Power Divergence Class of distributions is given by:

(3.1.7) $$f(w; q, \gamma) = \frac{1}{q^{-\gamma} F(w; q, \gamma)^{\gamma-1} + (1-q)^{-\gamma}(1 - F(w; q, \gamma))^{\gamma-1}} \quad for\ w \in \Upsilon(q, \gamma).$$

Consider the solution for $\int_{-\infty}^{X_i\beta} w f(w) d(w)$. We utilize the PDF of equation (3.1.7) to represent the integral as:

(3.1.8) $$\int_{-\infty}^{X_i\beta} w\,f\left(w\right)d(w)=\int_{-\infty}^{X_i\beta}\frac{w}{q^{-\gamma}F(w,q,\gamma)^{\gamma-1}+(1-q)^{-\gamma}(1-F(w,q,\gamma)^{\gamma-1}}dw\ .$$

Incorporate a change of variable in equation (3.1.8) via the transformation $p=F(w,q,\gamma)$ so that

$w=F^{-1}(p,q,\gamma)$ and $\frac{\partial w}{\partial p}=\frac{\partial F^{-1}\left(p;q,\gamma\right)}{\partial p}$, where $F^{-1}\left(p;q,\gamma\right)$ denotes the inverse function associated with the CDF.

Given $\gamma\neq 0$ we can represent the inverse CDF as:

(3.1.9) $$w=F^{-1}\left(p;q,\gamma\right)=\gamma^{-1}\left(\left(\frac{p_i}{q_i}\right)^{\gamma}-\left(\frac{1-p_i}{1-q_i}\right)^{\gamma}\right)\ for\ p\in\left(0,1\right).$$

It follows that

(3.1.10) $$\int_{-\infty}^{X_{i.}\beta} w\,f\left(w\right)d(w)=\int_{0}^{F\left(X_{i.}\beta;q,\gamma\right)} F^{-1}\left(p;q,\gamma\right)dp$$

and the conditional expectation can be represented as:

(3.1.11) $$E\left(\varepsilon_i^*\,|\,\varepsilon_i^*>-X_i\beta\right)=-\frac{\int_0^{F(X_i\beta;q,\gamma)}\gamma^{-1}\left(\left(\frac{p}{q}\right)^{\gamma}-\left(\frac{1-p}{1-q}\right)^{\gamma}\right)dp}{F(X_i\beta;q,\gamma)}$$

then the solution of (3.1.11) is

(3.1.12) $$E\left(\varepsilon_i^*\,|\,\varepsilon_i^*>-X_{i.}\beta\right)=\frac{-\frac{\left(F(X_{i.}\beta;q,\gamma)\right)^{(\gamma+1)}}{\gamma\left(\gamma+1\right)q^{\gamma}}+\frac{1}{\gamma\left(1-q\right)^{\gamma}}\left(\frac{1}{\gamma+1}-\frac{\left(1-F(X_{i.}\beta;q,\gamma)\right)^{(\gamma+1)}}{\gamma+1}\right)}{F(X_{i.}\beta;q,\gamma)}$$

where $F(X_i\beta;q,\gamma)$ is some member of the Class of Minimum Power Divergence distributions, parameterized by the values $\{\gamma,q\}$. The Distributions in the MPD-Class with $\gamma \le -1$ do not have moments defined of any order as shown in Appendix 2.

Given the above information we can state sample moment conditions as:

$$(3.1.13)\qquad g\left(Y,X,\beta,q,\gamma\right) \equiv n^{-1}X'\left[Y-\left(X_{i.}\beta+E\left(\varepsilon_i^* \mid \varepsilon_i^* > -X_i\beta\right)\right)\odot F\left(X_{i.}\beta,q,\gamma\right)\right]=0$$

And under the GMM approach the parameter vector is chosen so that the sample moment conditions are as close to the zero value as possible. The following weighted Euclidean distance is used as a measure of closeness:

$$(3.1.14)\qquad \begin{bmatrix}\hat{\beta}_{GMM}\\ \hat{q}_{GMM}\\ \hat{\gamma}_{GMM}\end{bmatrix}=\left[\hat{\theta}_{GMM}\right]=\underset{\beta,q,\gamma}{\arg\min}\left\{g\left(Y,X,\beta,q,\gamma\right)'W_{opt}\left(Y,X,\beta,q,\gamma\right)\right\}$$

where W is a conformable positive definite symmetric weight matrix. In order to find the optimal weight matrix, W_{opt}, we set $W = I$ and by calculating the $\hat{\theta}\left(I\right)$ in (3.1.14), we are able to calculate $\left(E\left(\hat{g}\hat{g}'\right)\right)^{-1}=\hat{W}_{OPT}$. The sample estimator of the optimal weighting matrix $\hat{W}_{opt}$ is substituted into (3.1.14) leading to the estimated optimal GMM defined by $\hat{\theta}_{GMM}(\hat{W}_n)$.

As a summary, after obtaining the sample moment, we used the GMM method to estimate the parameters of the censored regression model by minimizing the objective function (3.1.14). We were able to characterize the appropriate minimum power divergence distribution which identifies the noise distribution. It should be noted that the conditional expectation of the noise term will not exist for MPD distributions associated with $\gamma \le -1$, since the integrals defining moments of any order are non-convergent. Thus, for the regression moment approach underlying the GMM formulation to be implementable, the MPD class of distributions must be restricted to those for which $\gamma > -1$.

3.2 ML Estimation of the Censored Regression Model Assuming MPD distributions

An alternative method of estimating the censored regression problem can be defined through the use of the Maximum Likelihood principle. In order to specify the likelihood function of the mixed continuous-discrete process underlying the observation of the censored outcomes of Y_i, we need to specify the appropriate probability density function underlying the noise term. The general functional representation of the PDFs contained in the Minimum Power Divergence Class of distributions is given by:

(3.2.1) $$f(w;q,\gamma) = \frac{1}{q^{-\gamma} F(w,q,\gamma)^{\gamma-1} + (1-q)^{-\gamma} (1-F(w,q,\gamma)^{\gamma-1}} \quad for\ w \in \Upsilon(q,\gamma)$$

where $\Upsilon(q,\gamma)$ denotes the appropriate support of the density , which depends on the values of $\{q,\gamma\}$.

Using the reflexive property of the MPD class of distributions, it follows that the likelihood function associated with the sample outcomes of $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ can be written as:

(3.2.2) $$L(\beta,q,\gamma \mid Y) = \prod_{y_i's=0} F(-X_{i.}\beta;q,\gamma) \bullet \prod_{y_i's>0} f(y_i - X_{i.}\beta;q,\gamma)$$

Substituting into the PDF of equation (3.2.1) the explicit form of $w = y_i - X_{i.}\beta$ the likelihood function can be represented as:

$$(3.2.3)\quad L(\beta,q,\gamma\,|\,Y) = \prod_{y_i's=0} F(-X_{i.}\beta;q,\gamma)\bullet \prod_{y_i's>0}\left[q^{-\gamma}F(y_i - X_{i.}\beta,q,\gamma)^{\gamma-1} + (1-q)^{-\gamma}(1-F(y_i - X_{i.}\beta,q,\gamma)^{\gamma-1}\right]^{-1}$$

where $F(y_i - X_i\beta, q, \gamma)$ is the CDF of the error term evaluated at positive outcomes and $F(-X_i\beta, q, \gamma)$ is the CDF of the error term evaluated at the zero outcomes.

Maximizing the likelihood function (3.2.3) with respect to $\{\beta, q, \gamma\}$, results in the ML estimates of the unknowns, and in the estimation process, we can identify the appropriate MPD distribution which characterizes the noise distribution. Unlike the GMM approach, this method of estimating the unknowns does not rely on the existence of the conditional expectation of the noise term, and thus the choice of the γ value can be unrestricted. Further, the estimation procedure based on (3.2.3) is implemented by taking the negative log likelihood of the objective function and minimizing using an appropriate algorithm, which will be discussed in section 4 ahead.

4. Monte Carlo Sampling Experiments and Results

This section is divided into two parts. The first part presents results relating to a Monte Carlo simulation involving the GMM and ML procedures based on the MPD class of probability distributions. The second part compares both GMM and ML methods to the standard Tobit model first assuming the data generating process follows a normal distribution, and then a skewed Gamma (alpha=4, beta= 2) distribution.

The ML and GMM estimators based on the MPD class of distributions use the Neadler-Meade Polytope direct search method of optimization in calculating solutions, which is an easily implemented direct search method that only requires objective function evaluations for optimization. As a result it is robust to non-differentiabilities and it is useful for functions whose derivative cannot be easily or accurately calculated or approximated, or that are ill-conditioned.

4.1 Monte Carlo Simulation of GMM and ML Using MPD Principle

Based on model (1.1), we begin the sampling experiment by defining the parameters of the model. In both the GMM and ML approaches we set the parameters to $\beta_0 = 0.1$ and $\beta_1 = 0.2$, where β_0 is the intercept and β_1 is the slope of the regression. In order to generate the latent variable we sample X_i and ε_i^* in a way that takes into consideration the properties of the Minimum Power Divergence class of probability distribution functions.

We begin generating a size *n* sample of the X_i's, where *n=1000*, and we define:

(4.1.1) $$p_i = P(y_i > 0) = F(\beta_0 + \beta_1 x_i) \;\; for \;\; i = 1, \cdots, n.$$

where $F(\bullet)$ is a cumulative distribution. The X_i's are chosen such that $X_i = \frac{\left(F^{-1}\left(P\left(y_i > 0\right)\right) - \beta_0\right)}{\beta_1}$ and $F^{-1}\left(P\left(y_i > 0\right)\right)$ is the inverse cumulative distribution that includes three MPD distributions associated with different choices of γ parameter. If the choice of $\gamma \neq 0$ .i.e. $\gamma > 0$ and $\gamma < 0$ then we can represent the inverse CDF as:

(4.1.2) $$F^{-1}\left(P\left(y_i > 0\right)\right) = \frac{\left(\frac{p_i}{q_i}\right)^{\gamma} - \left(\frac{1-p_i}{1-q_i}\right)^{\gamma}}{\gamma}$$

If $\gamma = 0$ then the inverse CDF is represented as

(4.1.3) $$F^{-1}\left(P\left(y_i > 0\right)\right) = \ln\left(\frac{p_i}{q_i}\right) - \ln\left(\frac{1-p_i}{1-q_i}\right).$$

The subject probabilities p_i's in equations (4.1.2) and (4.1.3) are generated randomly from a Beta distribution *B (a, b)* that has *a=3* and *b=3* such that $0 < p < 1$. The beta distribution is symmetric and has a mean equal to 0.5. The *q* in all the experiments performed was set to 0.5. Generating the p_i's and setting *q=0.5* we are able to generate the X_i's

For the noise term ε_i^*, we follow the same process as used in generating the x_i's , if the choice of $\gamma \neq 0$, then the CDF can be represented as (4.1.2) and if $\gamma = 0$ then the CDF will be as (4.1.3).

After generating x_i and ε_i^* we can simply generate the latent dependent variable Y_i^*. Here the Y_i^* contains both negative and positive values, and the censored regression model can be represented as:

$$Y_i = \begin{Bmatrix} Y_i^* \\ 0 \end{Bmatrix} \; if \; \begin{Bmatrix} \varepsilon_i^* > -X_i\beta \\ \varepsilon_i^* \leq -X_i\beta \end{Bmatrix}, i = 1,\ldots,n.$$

The results of the above sampling experiment are presented in Appendix 1 and it is based on 1000 repetitions, where both GMM and ML results are reported. Table 1 presents a comparison between the GMM and ML approach using the MPD principle, where the reference distribution *q* is set to 0.5 and the *Gamma* to 0.5. The root means square errors (RMSE) of the ML-MPD are smaller than the RMSEs of the GMM-MPD, which suggest that the ML is more efficient than the GMM under the assumption of this arbitrary MPD distribution. These results are expected because of the asymptotic efficiency of maximum likelihood, also the maximum likelihood has an advantage over the GMM since there are no restriction imposed on the gamma parameter.

Table 2 compares the GMM and ML approach using the MPD that is identical to the standard logit distribution and it is defined at $q=0.5$ and $\gamma=0$. Again the results show that the RMSEs of the ML is smaller the RMSEs of the GMM and suggest that ML is more efficient when assuming the logit MPD distribution.

The above results show that the ML estimator is consistent, asymptotically efficient, and is not subjected to any restriction on its parameters. While the GMM estimator is consistent, not fully efficient and is subjected to restrictions on its parameters.

4.2 Comparison between GMM, ML using MPD to Standard Tobit Model

In order to compare the performance of the ML-MPD and GMM-MPD to the standard Tobit model under the assumption of normality we perform two Monte Carlo experiments with different sample sizes $n=1,000$, and $n=100$, where the data generating process for the three models is assumed to be standard normal.

The Tobit model is similar to model (1.1); the only difference being that the distribution of the error term is assumed to be standard normal. The ML estimator has been the most popular estimation procedure for the Tobit model. The estimators are consistent if the assumption of normality is correct, in which case it is also asymptotically efficient. Table 3 in the Appendix shows the comparison between the three models under the assumption of normality for the large sample size ($n = 1000$). It is not surprising that the correctly specified Tobit-ML out performs ML-MPD and GMM-MPD because in large samples the efficiency of the Tobit-ML should increase, and this is in fact evident in the RMSE results. Furthermore, the ML-MPD out performs the GMM-MPD, which is also expected given that the latter is not fully asymptotically efficient, but rather is efficient only with respect to the moment information used in estimation.

To examine the three models in smaller sample size, Table 4 shows the comparison of the models at sample size *n=100*. The results in this case show that RMSE (ML-MPD) <RMSE (GMM-MPD) <RMSE (Tobit), which suggests that estimation based on the MPD class of distributions in small sample size can outperform the Tobit-ML, even when the latter is correctly specified.

Finally, we compare the ML and GMM estimators that are based on the MPD class of distributions with the Tobit model, assuming a highly skewed distribution of the error term. We choose a Gamma distribution, which is skewed to the right and has the following parameters alpha = 4 and beta = 2. Figure 1 in the Appendix illustrates a graphical representation of Gamma (4, 2) distribution compared to the Standard normal distribution, where it is evident that the Gamma distribution is highly skewed to the right. The results are presented in Table 5 for sample size (*n=1,000*). The RMSEs suggest that ML-MPD is more efficient than the GMM-MPD under the assumption of the MPD that is similar to Gamma (4, 2) distribution, while the Tobit model becomes inconsistent and inefficient when the distribution is not correctly specified. Figure 2 shows the histogram of the Gamma (4, 2) distribution compared to MPD with q=0.1 and γ=1.54, where both graphs show approximately similar distributions. This finding suggest that estimating censored regression model, where the distribution in not correctly specified, The ML-

MPD and the GMM-MPD work reasonably well compared to the Tobit estimator, where the latter becomes inconsistent and inefficient.

5. Conclusions

In this paper, we represent sample information underlying continuous-discrete outcomes through general moment conditions, $E\left[X^{'}\left(Y - p\left(x\right)\right)\right] = 0$. We then use the CR family of divergence measures, CR (γ) for $\gamma \in \left(-\infty, \infty\right)$, to identify a wide and flexible class of CDFs underlying the sampling distribution of censored responses. Utilizing this class of CDFs we are able to estimate censored regression models with two different estimation procedures, ML and GMM.

ML and GMM estimation methods using the MPD class of distributions provide promising and alternative ways of estimating censored regression models, as Monte Carlo results suggest. Consistent and asymptotically efficient estimators are provided by the ML-MPD approach, but the GMM method is not fully efficient, although efficiency can be made to increase by adding more moment equation information to the model. The advantage of these models compared to conventional estimation procedures is that less stringent parametric assumptions are required, leading to robustness, consistency, and full asymptotic efficiency (in the case of ML) across a substantially wider class of distributions for the underlying data sampling process. Future research is being directed towards extending ML-MPD ML-GMM to higher dimensional models, where one can estimate a system of censored regression models having wide applicability in applications of microeconomic theory.

References

Buchinsky, M. and Hahn, J., (1998). "An alternative estimator for the censored quantile regression model". Econometrica 66, pp. 653–672.

Cosslett, S.R. (1983), "Distribution-Free Maximum Likelihood Estimation of the Binary Choice Model". Econometrica 51, 765-782

Cosslett, S.R. (1987), "Efficiency bounds for distribution-free estimators of the binary choice and censored regression models", Econometrica , 55, 559-585 (1987).

Cox, D.R., Snell, E.J., (1979), "On sampling and the estimation of rare errors" Biometrika 66, 125–132

Cressie, N. and Read, T.,(1984), "Multinomial Goodness of Fit Tests". Journal of the Royal Statistical Society, Series B 46, 440-464

Judge, G., Mittelhammer R., and Miller, D., (2004), "Estimating the link function in multinomial response models under endogeneity", Chapter in Chavas, Jean-Paul, ed., Volume in Honor of Stanley Johnson, University of California Press, in progress

Honoré B.E. and Powell, J.L., (1994). "Pairwise difference estimators of censored and truncated regression models". Journal of Econometrics **64**, pp. 241–278

Horowitz, J.L., (1988). Semiparametric "M-estimation of censored linear regression models". Advances in Econometrics **7**, pp. 45–83.

Green, W., (1993), "Econometric analysis" Second edition Prentice Hall, Englewood Cliffs, NJ

Khan , S., Powell, J.L., (1999). "Two-step quantile estimation of the censored regression model". Manuscript, University of Rochester

Klein, R.W. and Spady, R.H., (1993), "An Efficient Semiparametric Estimator for Binary Response Models" Econometrica 61(2), 387-421.

Lin, J. (1991), "Divergence measures based on the Shannon entropy". IEEE Trans. Inform. Theory, 37, 145-151

McFadden, D., (1984), "Qualitative Response Models," In Z. Griliches and M. Intriligator, eds.Handbook of Econometrics 2, Amsterdam, North Holland, pp 1395-1457

McFadden, D., (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in P.Zarembka, ed., Frontiers of Econometrics, New York: Academic Press, pp. 105-142

Maddala, G.S., (1983), “Limited Dependent and Qualitative Variables in Econometrics”. Econometric Society Monographs, Cambridge University Press

McCullough, P. and Nelder, **J.A.,** (1995), “Generalized Linear Models”, New York: Chapman and Hall

Mittelhammer, R. and Judge, G., (2007), “A minimum power divergence class of CDFs and estimators for the Binary Choice mode” (Working paper) Washington State University

Mittelhammer, R., Judge, G. and Miller, D.,(2000), Econometric Foundations, New York: Cambridge University Press

Nawata, K., (1990). “Robust estimation based on group-adjusted data in censored regression models”. Journal of Econometrics **43**, pp. 337–362

Newey, K. W., (1990a), “Semiparametric efficiency bounds” Journal of Applied Econometrics, 5, 99-135

Newey, K. W., and Powell J. L., (1990b), “Efficient estimation of type I censored regression models under conditional quantile restrictions” Econometric Theory, 6, 295-317

Moon,C.-G., (1989).”A Monte Carlo comparison of semiparametric Tobit estimators”. Journal of Applied Econometrics 4, pp. 361–382

Owen, A. (1988), “Empirical likelihood ratio confidence intervals for a single functional”. Biometrika 75, 237–249

Owen, A. (2001), “Empirical Likelihood”. Chapman & Hall, London

Parado, L. (2006), “Statistical Inference Based on Divergence Measure”, Boca Raton, Chapman and Hall

Powell, J.L., (1986). “Symmetrically trimmed least squares estimation of Tobit models”. Econometrica 54, pp. 1435–1460.

Powell , J.L., (1986). “Censored regression quantiles”. Journal of Econometrics 32, pp. 143–155.

Qin, J., (2000), “Combining Parametric and Empirical Likelihood Data”. Biometrika 87, 484-490

Read, T.R. & N.A. Cressie, (1988), “Goodness of Fit Statistics for Discrete Multivariate Data”, New York: Springer Verlag.

Song Xi Chena, Jing Qinb, (2003), "Empirical likelihood-based confidence intervals for data with possible zero observations" Statistics & Probability Letters 65, 29–37

Tamura, H., (1988), "Estimation of rare errors using expert judgment". Biometrika 75, 1–9

Train, K., (2003), "Discrete choice Methods with Simulation" New York ,Cambridge University Press

Tobin, J., (1958), "Estimation of relationships for limited dependent variables". Econometrica 26, 24–36

Ullah, A., (1996), "Entropy, Divergence and Distance Measures with Econometric Applications" Journal of Statistical Planning and Inference 49, 137-162.

Appendix 1

Table 1: Comparison between the GMM-MPD and ML-MPD, where $\left(\gamma = 0.5 \,\&\, q = 0.5\right)$, Sample size=1,000, Repetition=1,000

Sample size=1,000, Repetition=1,000					
Parameters	**True Values**	**GMM-MPD Parameter Estimates**	**GMM-MPD RMSE**	**ML-MPD Parameter Estimates**	**ML-MPD RMSE**
γ	0.5	0.5059	0.0356	0.5019	0.0124
q	0.5	0.5025	0.0425	0.5005	0.0345
β_0	0.1	0.1033	0.0251	0.1001	0.0101
β_1	0.2	0.2064	0.0321	0.2031	0.0276

Table 2: Comparison between the GMM-MPD and ML-MPD, where $\left(\gamma = 0 \,\&\, q = 0.5\right)$,

Sample size=1,000, Repetition=1,000

Sample size=1,000, Repetition=1,000					
Parameters	**True Values**	**GMM-MPD Parameter Estimates**	**GMM-MPD RMSE**	**ML-MPD Parameter Estimates**	**ML-MPD RMSE**
γ	0	0.0003	0.0342	1.84×10^{-08}	0.0231
q	0.5	0.5030	0.0552	0.4999	0.0421
β_0	0.1	0.1041	0.0342	0.1009	0.0321
β_1	0.2	0.2012	0.0401	0.2009	0.0392

Figure 1: Graphical representation of the scaled gamma (4,2) distribution

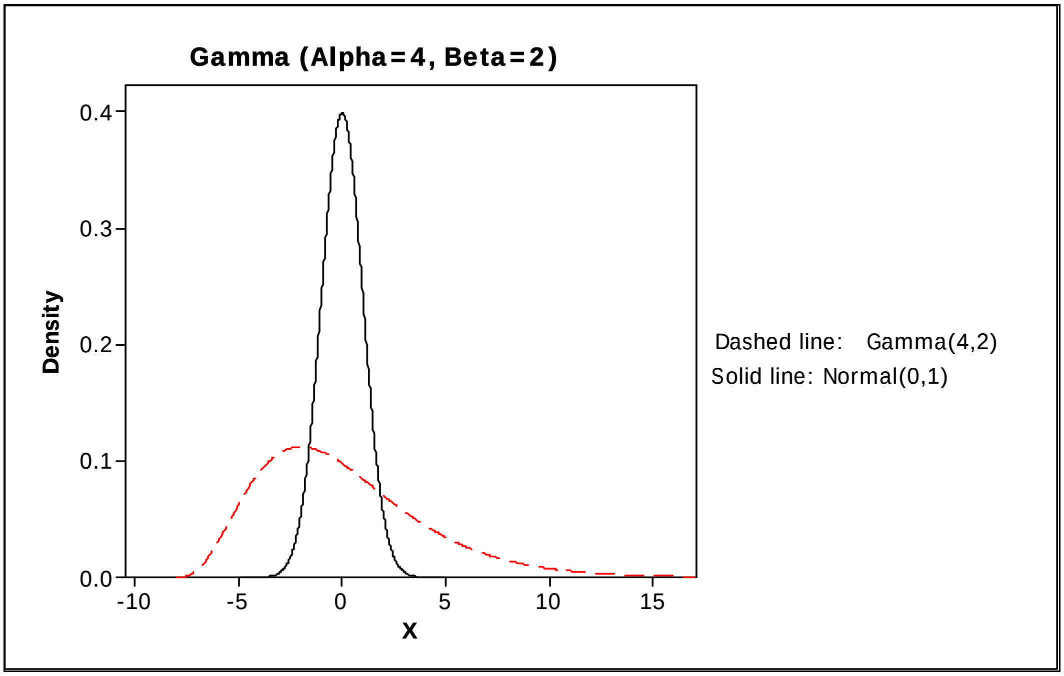


Figure 2: Comparison between Gamma (4, 2) and MPD (q=0.1, γ=1.54) distributions

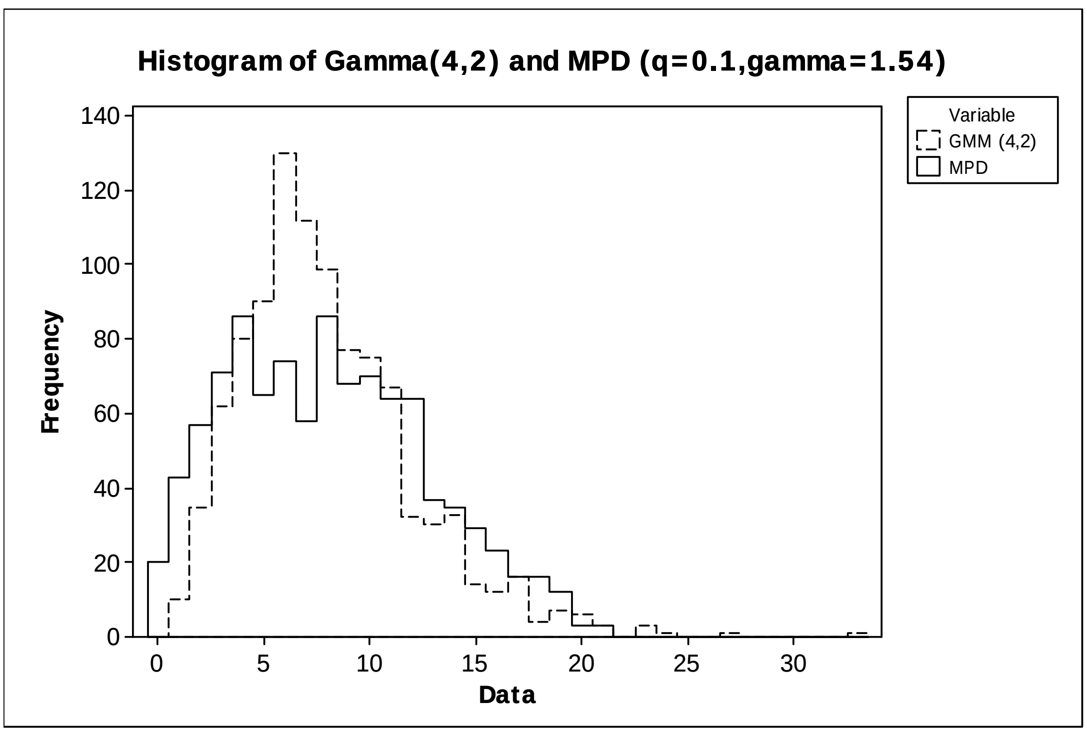


Table 3: Large sample size comparison of the GMM and ML using MPD principle with the classical standard Tobit, under the assumption of Normality

Sample size=1,000, Repetition=1,000							
Parameters	**True Values**	**GMM-MPD Parameter Estimates**	**GMM-MPD RMSE**	**ML-MPD Parameter Estimates**	**ML-MPD RMSE**	**Tobit Parameter Estimates**	**Tobit RMSE**
β_0	0.1	0.0898	0.0201	0.0988	0.0101	0.1021	0.0095
β_1	0.2	0.2210	0.0605	0.2105	0.0408	0.1999	0.0121

Table 4: Small sample size comparison of the GMM and ML using MPD principle with the classical standard Tobit, under the assumption of Normality

Sample size=100, Repetition=1,000							
Parameters	**True Values**	**GMM-MPD Parameter Estimates**	**GMM-MPD RMSE**	**ML-MPD Parameter Estimates**	**ML-MPD RMSE**	**Tobit Parameter Estimates**	**Tobit RMSE**
β_0	0.1	0.0984	0.0321	0.1004	0.0221	0.0954	0.0382
β_1	0.2	0.1985	0.0646	0.2007	0.0582	0.1933	0.0686

Table 5: Large sample size comparison of the GMM and ML using MPD principle with the classical standard Tobit, under the assumption of Gamma (4,2) distribution

Sample size=1,000, Repetition=1,000							
Parameters	**True Values**	**GMM-MPD Parameter Estimates**	**GMM-MPD RMSE**	**ML-MPD Parameter Estimates**	**ML-MPD RMSE**	**Tobit Parameter Estimates**	**Tobit RMSE**
β_0	0.1	0.0876	0.0082	0.0984	0.0023	-0.1292	0.0554
β_1	0.2	0.2287	0.0075	0.2190	0.0019	0.2690	0.0385

Appendix 2

Given the following conditional moment for the error terms

$$E\left(\varepsilon_i^* \mid \varepsilon_i^* > -X_i\beta\right) = -\frac{\int_0^{F(X_i\beta;q,\gamma)} \left(\gamma^{-1}\left(\left(\frac{p}{q}\right)^{\gamma} - \left(\frac{1-p}{1-q}\right)^{\gamma}\right)\right)dp}{F(X_i\beta;q,\gamma)}$$

We want to examine the integration part in the above equation, that is

$$\int_0^{F(X_i\beta;q,\gamma)} \left(\gamma^{-1}\left(\left(\frac{p}{q}\right)^{\gamma} - \left(\frac{1-p}{1-q}\right)^{\gamma}\right)\right)dp$$

1. Assuming $\gamma = -1$

Then $\int \left(\gamma^{-1}\left(\left(\frac{p}{q}\right)^{\gamma} - \left(\frac{1-p}{1-q}\right)^{\gamma}\right)\right)dp = \int \left((-1)\left(\left(\frac{p}{q}\right)^{-1} - \left(\frac{1-p}{1-q}\right)^{-1}\right)\right)dp$

$$= (-1)\left(q\ln(p) + (1-q)\ln(1-p)\right) + C$$

2. Assuming $\gamma \neq -1$

$$\int \left(\gamma^{-1}\left(\left(\frac{p}{q}\right)^{\gamma} - \left(\frac{1-p}{1-q}\right)^{\gamma}\right)\right)dp = \gamma^{-1}\left(\left(\frac{p^{\gamma+1}}{q^{\gamma}(\gamma+1)}\right) + \left(\frac{(1-p)^{\gamma+1}}{(1-q)^{\gamma+1}(\gamma+1)}\right)\right) + C$$

In order to decide on convergence or divergence of the above two integrals, we need to consider the following cases:

- $\gamma = -1$

$$\int_0^{F(X_i\beta;q,\gamma)}\left(\gamma^{-1}\left(\left(\frac{p}{q}\right)^{\gamma}-\left(\frac{1-p}{1-q}\right)^{\gamma}\right)\right)dp=\lim_{c\to 0}\int_0^{c}\left(\gamma^{-1}\left(\left(\frac{p}{q}\right)^{\gamma}-\left(\frac{1-p}{1-q}\right)^{\gamma}\right)\right)dp$$
$$=\lim_{c\to 0}\Big[(-1)\big(q\ln(c)+(1-q)\ln(1-c)\big)+C\Big]=+\infty$$

- $\gamma<-1$

$$\int_0^{F(X_i\beta;q,\gamma)}\left(\gamma^{-1}\left(\left(\frac{p}{q}\right)^{\gamma}-\left(\frac{1-p}{1-q}\right)^{\gamma}\right)\right)dp=\lim_{c\to 0}\int_0^{c}\left(\gamma^{-1}\left(\left(\frac{p}{q}\right)^{\gamma}-\left(\frac{1-p}{1-q}\right)^{\gamma}\right)\right)dp$$
$$=\lim_{c\to 0}\left[\gamma^{-1}\left(\left(\frac{c^{\gamma+1}}{q^{\gamma}(\gamma+1)}\right)+\left(\frac{(1-c)^{\gamma+1}}{(1-q)^{\gamma+1}(\gamma+1)}\right)\right)+C\right]=+\infty$$

- $\gamma>-1$

$$\int_0^{F(X_i\beta;q,\gamma)}\left(\gamma^{-1}\left(\left(\frac{p}{q}\right)^{\gamma}-\left(\frac{1-p}{1-q}\right)^{\gamma}\right)\right)dp=\lim_{c\to 0}\int_0^{c}\left(\gamma^{-1}\left(\left(\frac{p}{q}\right)^{\gamma}-\left(\frac{1-p}{1-q}\right)^{\gamma}\right)\right)dp$$
$$=\lim_{c\to 0}\left[\gamma^{-1}\left(\left(\frac{c^{\gamma+1}}{q^{\gamma}(\gamma+1)}\right)+\left(\frac{(1-c)^{\gamma+1}}{(1-q)^{\gamma+1}(\gamma+1)}\right)\right)+C\right]$$
$$=\gamma^{-1}\left(\frac{1}{(1-q)^{\gamma+1}(\gamma+1)}\right)$$

Therefore the integral will be divergent if $\gamma<-1$ & $\gamma=-1$ and convergent if $\gamma>-1$.