Essays of Empirical Studies in Agricultural and Resource Economics

By

Zhihua Shen

B.A. (Fudan University, Shanghai, P. R. China) 1991

M.S. (University of California, Berkeley) 1995

M.A. (University of California, Berkeley) 1998

A dissertation submitted in partial satisfaction of the

requirement of the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Jeffrey Perloff, Chair

Professor Larry Karp

Professor Daniel McFadden

Professor David Zilberman

Spring 2000

UMI Number: 9979805

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Essays of Empirical Studies in Agricultural and Resource Economics

by

Zhihua Shen

To,

My Parents.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I want to thank my advisor Jeffrey Perloff for his support, encouragement, and guidance during my graduate study at Berkeley. I want to thank Larry Karp who provided me with numerous insightful suggestions. I also want to thank David Zilberman for his support and encouragement for conducting innovative research. I want to thank Daniel McFadden for many valuable comments. I want to thank Amos Golan and George Judge who provide generous help in my learning of maximum entropy. I am grateful to Arnold Zellner for his numerous suggestions and new ideas for future research. I want to thank the faculty and students who have participated in the Resource Economics Seminars where I presented my work. I also want to thank the graduate students and staff at ARE as a whole with whom I have been enjoying working.

I want to thank Yuan Ge, my brother-in-law, and my sisters for their support throughout my study at Berkeley. Finally, I want to thank Pu Yang, my wife.

Chapter 1 Overview

This dissertation investigates some econometrics methods with an aim to provide better quantitative results in a few important issues in agricultural and environmental economics. In particular, it focuses on the application of new estimators such as maximum entropy by Golan, Judge, and Miller (1996) as well as minimum expected loss estimator and Bayesian method of moment estimator by Zellner (1978, 1996, 1997).

The first essay proposes a new estimator to solve the problem of estimating system of demand equations with non-negativity constraints. Traditionally, there are two approaches to tackle this problem. The first is the maximum likelihood (ML) approach and the second one is Amemiya's two-stage estimator. These two approaches have some limitations in empirical work. ML is complicated in practice when the number of equations is large and the two-stage estimator is not efficient. We adopted a new approach – maximum entropy estimator that is single-staged, robust, and easy to impose non-negativity as well as other restrictions from economic theory. It is demonstrated to be more efficient than the traditional approaches. We applied that method into an estimation of AIDS model using Mexican household expenditure data.

The second essay aims to improve the supply elasticity estimates from Nerlove's agricultural supply, a is a widely used econometric model. Literature shows that the estimated supply elasticities using the least square method have wide variation. We suggest two new approaches to improve the estimates. The first method is maximum entropy and the second method is Zellner's Bayesian method of moments. Monte Carlo experiment shows that both methods yield more accurate and stable estimates than the least square and minimum expected loss approach.

The third essay studies the willingness-to-pay estimates in contingent valuation studies. Dichotomous choice model has been widely used as contingent valuation tool to assess many non-market resources and public goods such as park, beach, and national forest. The key estimates, WTP, however suffer from biasness and variability in empirical studies. While many papers focus upon such factors as starting-point bias and hypothetical market bias, one important factor has been neglected. The widely used ML estimates of WTP involve ratio of random variables and thus do not possess desirable properties in finite samples. This essay suggests the minimum expected loss (MELO) technique to solve this problem. Given diffuse prior, the MELO estimates are demonstrated to be more accurate and stable than the ML ones. Given non-diffuse prior, the MELO estimates become even more accurate. It is applied into two contingent valuation studies in California and Oregon.

Chapter 2 Estimating a Demand System with Nonnegativity Constraints: Mexican Meat Demand

2.1 Introduction

We present a new approach to efficiently estimate a system of many equations with binding nonnegativity constraints. Using this approach, we estimate both five-equation and six-equation meat demand systems based on data from a cross-section survey of Mexican households. Most of these households did not purchase one or more of these meat products during the survey week. We compare our demand elasticity estimates for Mexico to estimates of those in wealthier countries. We also show how elasticities of demand vary with demographic characteristics.

There have been relatively few previous attempts to estimate demand systems with binding nonnegativity constraints.[1] We know of only one study, Heien and Wessells (1990), that estimates a many-equation demand system with nonnegativity constraints with variable prices. They use a two-stage Amemiya (1974) approach to estimate an Almost Ideal Demand System (AIDS, Deaton and Muelbauer, 1980) for 11 food items with an emphasis on dairy products.

Although such two-step methods are consistent, they are not invariant to the choice of which good is dropped, and they are inefficient and require specific distributional assumptions. Flood and Tasiran (1990) find that the Amemiya two-stage estimator performs poorly compared

[1] Deaton and Irish (1984), Kay, Keen, and Morris (1984), Keen (1986), and Blundell and Meghir (1987) use models based on the discrepancy between observed expenditure and actual consumption. We (and the other papers discussed here) concentrate on actual purchases.

to maximum likelihood (ML) in estimating a system of tobit equations with normal errors and that this inefficiency does not decrease with sample size.[2] Moreover, they find that the ML and the two-stage approaches perform poorly when the errors are not normal.

If the errors are normal (or another known distribution), greater efficiency can be achieved by using full-information ML techniques. Using ML techniques, however, is feasible only for systems consisting of a relatively small number of goods, say three.[3]

Wales and Woodland (1983) use ML techniques to estimate a demand system with nonnegativity constraints based on a random quadratic utility function for three goods (beef, lamb, and other meats) based on Australian data.[4] Because they observe no variation in prices, they estimate only the variation in demand due to differences in demographic characteristics.

Ransom (1987) examined the relationship between the Wales and Woodland method and the Amemiya approach to estimating simultaneous tobits. He showed that the internal consistency condition for the Wales and Woodland model is equivalent to the second-order condition for systems of demand equations without binding quantity constraints. If prices are constant, Wales and Woodland's method and the simultaneous system with limited dependent variables of Amemiya are identical. If prices vary, the error terms are heteroscedastic.

Our objective is to recover the unknown parameters of a censored demand system with many goods where we make no distribution assumption and where the exogenous variables may be correlated. Our approach has its roots in information theory and builds on the entropy-information measure of Shannon (1948), the classical maximum entropy (ME) principle of Jaynes

[2] Their experiments suggest, however, that the Nelson and Olsen two-stage estimator (which does not drop the constrained observations) performs reasonably well compared to ML. In contrast, Lee (1978) shows that, for a system of equations, the Amemiya two-stage estimator is more efficient than the Nelson and Olsen or Heckman two-stage estimators when the normality hypothesis is maintained.

[3] It may be possible to estimate larger systems using a general method of moments estimator, however, this approach has not be used in a demand study.

[4] Lee and Pitt (1986) propose using the dual of Wales and Woodland's method to transform binding nonnegativity constraints into nonbinding constraints based on virtual (shadow) prices. They estimate a three-input energy demand system using a translog cost function.

(1957a, 1957b), which was developed to recover information from underdetermined systems, and the generalized maximum entropy (GME) theory of Golan, Judge, and Miller (1996).

The GME method allows us to consistently and efficiently estimate a demand system with nonnegativity constraints and a large number of goods without imposing restrictions on the error process. The GME estimates are robust even if errors are not normal.

In this paper, we use our GME approach to estimate an Almost Ideal Demand System (AIDS), but our method could be applied to any demand system. In our empirical application, we concentrate on estimating the elasticities of demand and examining how these elasticities vary with demographic characteristics of households.

2.2 AIDS Model

The Almost Ideal Demand System (AIDS) is a flexible, complete demand system: It satisfies the adding up of budget shares, homogeneity, and symmetry. Throughout most of the paper, we assume that meat and all other goods are separable in the utility functions and estimate a five-equation system of demand for meats only.[5] We briefly discuss our complete, six-equation demand system estimates.

The AIDS consists of a set of budget-share equations:

$$s_i = \alpha_i + \sum_{i=1}^{n} \gamma_{i\,j} \ln p_i + \beta_i \ln(E/P), \qquad (1)$$

where s_i (≥ 0) is the budget share of meat product i, p_i is the price of product i, E is the total expenditure on meats, P is a price index, and α_i, γ_{ij}, and β_i are constant parameters. We

[5] Alston and Chalfant (1987) contrast using expenditure or total income to estimate a separable meat demand systems using Australian data. Based on non-nested tests, they favored using expenditure. Their results are mixed on whether separability holds. Moschini, Moro, and Green (1994) find support for separabitity between meat and other foods in a Rotterdam model.

allow the intercept terms to vary with a matrix X of K exogenous demographic and geographic variables, where

$$\alpha_i = \sum_{k=0}^{K} \rho_{ik} X_k \tag{2}$$

and the ρ_{ik} are constant parameters. Consumer theory requires that the equations satisfy adding up, symmetry, and homogeneity constraints:

$$\gamma_{ij} = \gamma_{ji},$$

$$\Sigma_i \alpha_{i0} = 1,$$

$$\Sigma_i \rho_{ik} = 0 \text{ for } k = 1, \ldots, K, \text{ and}$$

$$\Sigma_i \beta_i = 0,$$

$$\Sigma_j \gamma_{ij} = 0.$$

The nonlinear price index is

$$\ln P = \phi + \sum_{i=1}^{n} \alpha_i \ln p_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \ln p_i \ln p_j . \tag{3}$$

Although we use the nonlinear price index, a common practice is to replace the nonlinear price index, Equation 3, with Stone's linear approximation:[6]

$$\ln P^* = \sum_{i=1}^{n} s_i \ln p_i . \tag{4}$$

We follow the standard practice of adding an error term, ε_i, to each budget-share equation. Thus, the model we estimate is

$$s_i = \sum_{k=1}^{K} \rho_{ik} X_k + \sum_{i=1}^{n} \gamma_{ij} \ln p_i + \beta_i \ln(E/P) + \varepsilon_i \quad , \quad \text{for } s_i > 0 \tag{5}$$

[6] In our sample, there is little difference between using the exact, nonlinear price index and the linear approximation.

$$s_i > \sum_{k=1}^{K} \rho_{ik} X_k + \sum_{i=1}^{n} \gamma_{ij} \ln p_i + \beta_i \ln(E/P) + \varepsilon_i \quad , \quad \text{for } s_i = 0 \tag{6}$$

2.3 Estimation Approach

To estimate this system of censored demand equations, we generalize the GME method for estimating a single, censored equation in Golan, Judge, and Perloff (1997). We start by providing some intuition as to how the maximum entropy approach works. Then, we show how to estimate the AIDS using GME. The appendix discusses the properties of this estimator and derives the asymptotic variance matrix.

2.3.1 Maximum Entropy

The traditional maximum entropy (ME) formulation is based on the entropy-information measure of Shannon (1948). It is developed and described in Jaynes (1957a, 1957b), Kullback (1959), Levine (1980), Jaynes (1984), Shore and Johnson (1980), Skilling (1989), Csiszár (1991), and Golan, Judge, and Miller (1996). Shannon's (1948) entropy is used to measure the uncertainty (state of knowledge) we have about the occurrence of a collection of events. Letting x be a random variable with possible outcomes x_s, $s = 1, 2, \ldots, n$, with probabilities δ_s such that $\Sigma_s \delta_s = 1$, Shannon (1948) defined the *entropy* of the distribution $\underline{\delta} = (\delta_1, \delta_2, \ldots, \delta_n)'$, as

$$H \equiv -\sum_s \delta_s \ln \delta_s \ , \tag{7}$$

where $0 \ln 0 \equiv 0$. The function H, reaches a maximum of $\ln(n)$ when $\delta_1 = \delta_2 = \ldots = \delta_n = 1/n$. It is zero when $\delta_s = 1$ for one value of s. To recover the unknown probabilities $\underline{\delta}$ that characterize a given data set, Jaynes (1957a, 1957b) proposed maximizing entropy, subject to available sample-moment information and adding up constraints on the probabilities.

This procedure has an intuitive appeal. Suppose we have a sample of T draws of an identically and independently distributed random variable x that can take n values, $x_1, x_2, \ldots, x_n$, with probabilities $\delta_1, \delta_2, \ldots, \delta_n$. Because the draws are independent, a list of the number of times each value occurs contains all of the information this experiment provides about the random variable (i. e., the order contains no information about the probabilities). We define the *outcome* of the experiment as a vector $f = (f_1, f_2, \ldots, f_n)$, where f_s is the number of times x_s occurs and $\Sigma_s f_s = T$. A particular outcome may be obtained in a number of ways. For example, the outcome $(1, T-1, 0, 0, \ldots, 0)$ can occur in T possible ways because x_1 may be observed in any of the T draws. In contrast, the outcome $(T, 0, 0 \ldots 0)$ can occur in only one way, where x_1 was drawn each time.

Define $v(f)$ as the number of ways that a particular outcome can occur. Suppose we have no information about the draws and are asked which outcome is the most likely. An "intuitively reasonable" response is that the outcome that can occur in the most number of ways, $f^* \equiv \text{argmax } v(f)$, is the most likely outcome. Equivalently, we would consider it more likely to observe the frequency f^*/T than any other frequency. Shannon shows that, in the limit as $T \to \infty$, choosing f to maximize $v(f)$ is equivalent to choosing $\underline{\delta}$ to maximize the entropy measure, $H(\underline{\delta})$.

That is, the frequency that maximizes entropy is an intuitively reasonable estimate of the true distribution when we lack any other information. If we have information from the experiment, such as the sample moments, or non-sample information about the random variable, such as restrictions from economic theory, we want to alter our "intuitively reasonable" estimate. The ME method chooses the distribution that maximizes entropy, subject to the sample and non-sample information. That is, out of all the possible estimates or probability distributions that are consistent with the sample and nonsample data, the ME method picks the one that is most uninformed: closest to a uniform distribution. In this sense, the ME estimator is conservative.

2.3.2 Estimating an AIDS Model using GME

In the traditional maximum entropy (ME) approach, sample information in the form of moment conditions is assumed to hold exactly. In contrast, the generalized maximum entropy approach (Golan, Judge, and Miller, 1996) allows these conditions to hold only approximately by treating them as stochastic restrictions.

Further, the GME uses a flexible, dual-loss objective function: a weighted average of the entropy of the systematic part of the model and the entropy from the error terms. The ME is a special case of the GME where no weight is placed on the entropy of the error terms and where the data are exact moments. By varying the weight in the GME objective, we can improve either our precision or predictions. Here, we use a balanced approach where we give equal weight to both objectives.[7]

To write these two entropy measures, we need to express all the coefficients and errors in Equations 4, 5, and 6 in terms of proper probabilities. To transform γ_{ij}, we start by choosing a set of discrete points, called the support space, $z_j = [z_{j1}, z_{j2}, \ldots, z_{jD}]$ of dimension $D \geq 2$, that are at uniform intervals, symmetric around zero, and span the interval $[-a, a]$. We then introduce a vector of corresponding unknown weights $\boldsymbol{q}_j = [q_{j1}, q_{j2}, \ldots, q_{jD}]$ such that $\Sigma_d\, q_{ijd} = 1$ and $\Sigma_d\, z_{jd}\, q_{jd} = \gamma_{ij}$ for all i and j. For example, if $D = 3$, then $z_j = (-a, 0, a)'$, and there exists q_{j1}, q_{j2}, and q_{j3} such that each $\gamma_{ij} = -aq_{j1} + aq_{j3}$. We index the number of discrete points (dimensions) in the support space for each unknown coefficient with $d = 1, 2, \ldots, D$. Each support space and the associated probability distribution can be of different dimension. We use the same approach for the β, ρ, and ϕ coefficients.

[7] In our empirical application, the results are not very sensitive to the weight. Indeed, raising the weight from 0.5 to 0.9 on the systematic measure causes the estimated coefficients and the correlation between actual and estimated values to changes by less than 1%.

We treat the errors ε_{it} as unknowns and define a transformation matrix V that converts the possible outcomes for ε_{it} to the interval [0, 1]. This transformation is done by defining a set of $H \geq 2$ discrete points $\underline{v} = [v_1, v_2, \ldots, v_H]'$, distributed evenly and uniformly about zero, and a corresponding vector of unknown weights $\boldsymbol{w} = [w_{it1}, w_{it2}, \ldots, w_{itH}]'$ such that $\Sigma_h v_h w_{ith} = \varepsilon_{it}$. No subjective information on the distribution of probabilities is assumed. Substituting these reparameterized terms into the AIDS Equations 5 and 6, we obtain:

$$\begin{aligned} s_{it} = {} & \sum_{k=0}^{K} \sum_{d=1}^{D} z_{ikd}^{\rho} q_{ikd}^{\rho} x_{tk} + \sum_{i=1}^{n} \sum_{d=1}^{D} z_{ijd}^{\gamma} q_{ijd}^{\gamma} \ln(p_{tj}) \\ & + \sum_{d=1}^{D} z_d^{\beta} q_{id}^{\beta} \ln(E_t / P_t^*) + \sum_{h=1}^{H} v_h w_{ith}, \quad \text{for } s_{it} > 0 \end{aligned} \tag{8}$$

$$\begin{aligned} s_{it} > {} & \sum_{k=0}^{K} \sum_{d=1}^{D} z_{ikd}^{\rho} q_{ikd}^{\rho} x_{tk} + \sum_{i=1}^{n} \sum_{d=1}^{D} z_{ijd}^{\gamma} q_{ijd}^{\gamma} \ln(p_{tj}) \\ & + \sum_{d=1}^{D} z_d^{\beta} q_{id}^{\beta} \ln(E_t / P_t^*) + \sum_{h=1}^{H} v_h w_{ith}, \quad \text{for } s_{it} = 0. \end{aligned} \tag{9}$$

The GME estimator maximizes the joint entropy for the signal (ρ, γ, β, ϕ) and the noise (ε), subject to the data, the linear price index Equation 4, and the adding up (of the probabilities), homogeneity, and symmetry conditions.

Letting $\boldsymbol{q} = (\boldsymbol{q}^{\rho\prime}, \boldsymbol{q}^{\gamma\prime}, \boldsymbol{q}^{\beta\prime}, \boldsymbol{q}^{\phi\prime})'$, the GME estimator is

$$\underset{q,w}{Max}\, H = -\, q' \ln q - w' \ln w\ , \tag{10}$$

subject to budget-share Equations 8 and 9, the linear price index Equation 4, the GME adding-up conditions,

$$\sum_{d}^{D} q_{ikd}^{\rho} = \sum_{d}^{D} q_{ijd}^{\gamma} = \sum_{d}^{D} q_{id}^{\beta} = \sum_{d}^{D} q_{d}^{\phi} = \sum_{h}^{H} w_{ith} = 1. \tag{11}$$

and the consumer-theory restrictions

$$\sum_{i=1}^{n} \alpha_{i0} = 1, \tag{12}$$

$$\sum_{i=1}^{n} \rho_{ik} = 0, \quad k = 1, \ldots, K \tag{13}$$

$$\sum_{i=1}^{n} \beta_i = \sum_{i=1}^{n} \gamma_{ij} = \sum_{j=1}^{n} \gamma_{ij} = 0 , \tag{14}$$

$$\gamma_{ij} = \gamma_{ji}. \tag{15}$$

The solution to this maximization problem is unique. Forming the Lagrangean and solving for the first-order conditions yields the optimal solution q and w, from which we derive the point estimates for the AIDS coefficients:

$$\hat{\gamma}_{ij} = \sum_{d=1}^{D} z_{ijd}^{\gamma} \; \hat{q}_{ijd}^{\gamma}. \tag{16}$$

$$\hat{\rho}_{ik} = \sum_{d=1}^{D} z_{ijd}^{\rho} \; \hat{q}_{ijd}^{\rho}. \tag{17}$$

$$\hat{\beta}_i = \sum_{d=1}^{D} z_{id}^{\beta} \; \hat{q}_{id}^{\beta}. \tag{18}$$

$$\hat{\phi} = \sum_{d=1}^{D} z_d^{\phi} \; \hat{q}_d^{\phi}. \tag{19}$$

$$\hat{\varepsilon}_{it} = \sum_{h=1}^{H} v_h \;\; \hat{w}_{ith} . \tag{20}$$

That this GME estimator is consistent follows immediately by extending the proof in Golan, Judge, and Perloff (1997) that a censored GME estimator for a single equation is consistent (see the Appendix). The GME has several other desirable properties (see Golan, Judge, and Perloff 1997 and Golan, Judge, and Miller 1996). The GME approach uses all the data points and does not require restrictive moment or distributional error assumptions. Thus, unlike the ML estimator, the GME is robust for a general class of error distributions. Further, the GME estimator performs well in both well-posed and ill-posed problems. Thus, the GME estimator may be used when the sample is small, there are many covariates, and when the covariates are highly correlated. Moreover, using the GME method, it is easy to impose nonlinear constraints, such as those in the nonlinear price index (Equation 3).

Most important for demand system estimation, the GME produces efficient estimates of a system of many censored equations. Using ML methods, one is practically restricted to estimating only a few equations or using relatively inefficient two-stage methods. The sampling experiments of Golan, Judge, and Perloff (1997) indicate that the balanced single-equation GME estimator is more efficient — has lower empirical mean square error — than the ML tobit estimator in small samples. These results hold even when the true underlying error distribution is normal, as is assumed in the tobit estimator. These experiments indicate that the GME is even more efficient relative to the ML when the error term is not normal.

2.4 Data

We use data from a cross-sectional Mexico household survey conducted by the National Institute of Statistics, Geography and Informatics (INEGI), an agency of the Ministry of Budgeting and Programming in Mexico, in the last quarter of 1992, which was provided by the World Bank. A stratified and multi-stage sampling method was used to produce a representative sample for the entire population and for urban and rural households. The data cover 31 states and one Federal District.

The data base has detailed information about consumption during a one-week survey period and demographic characteristics by household. The survey recorded 581,027 observations of purchasing events by about 10,500 households.[8] At least 205 types of foods are separately reported.

[8] Own produced and consumed goods are included in quantity measures.

We examine the quantities purchased for five aggregates of meat products: beef, pork, chicken, processed meat, and fish. The corresponding prices are also aggregates. For example, the price of beef is an expenditure weighted average of beef steak, pulp, bone, fillet, special cuts, and ribs and other.

The pricbuy pork; 35%, chicken; 57%, processed meat; and 87%, fish. Because prices are only reported if purchases are made, we need measures of the prices for households that did not make purchases. We assume that those household face the average price level for that product in that particular geographic location: a rural or urban area in a particular state or Federal District.

We experimented with various sample sizes and found that our estimates were not very sensitive to sample size. In the following, our estimates are based on a random sample of 1,000 observations. Table 2-1 shows that the means and standard deviations of our 1,000 observations are virtually the same as for all 7,897 households. Table 2-1 also provides summary statistics for the consumption shares of the five meats, the corresponding prices, expenditures on meats, and the 12 demographic variables we use in our GME LA/AIDS Model.

2.5 Estimation

We obtained our GME estimates by maximizing the joint-entropy objective, Equation 10, which depends on Equations 8 and 9, subject to the nonlinear price index, Equation 3, the GME adding-up restrictions, Equation 11, and the consumer-theory restrictions, Equations 12-15.

We set our support vectors z^l ($l = \rho, \gamma, \beta, \phi$) wide enough to include all the possible outcomes. The natural support vector for the error term is $v = (-1, 0, 1)$, because all the dependent variables are shares that lie between 0 and 1. In a variety of AIDS empirical studies [Heien and Pompelli (1988), Moschini and Meilke (1989), Heien and Wessells (1990), Chalfant *et al.* (1991) and Halbrendt *et al.* (1994)], we found that the estimated coefficients on log prices were within the interval of (-0.2, 0.2) and the intercepts and coefficients on log expenditures were within the interval of (-1, 1). We chose support vectors that are 100 times wider than these intervals: (-20, 20) for the log price coefficients and (-100, 100) for the intercept and log expenditure coefficients. Making a moderately large change in these support vectors, while keeping the center of the support unchanged, has negligible effects on the estimated coefficients and elasticities.

The model was estimated using GAMS (Generalized Algebraic Modeling System), which is a nonlinear-optimization program.[9] Table 2-2 shows our GME estimates of the LA/AIDS demand system. The asymptotic standard errors are calculated using the second method described in the Appendix.

We tested for symmetry and homogeneity. The objective value with both symmetry and homogeneity imposed is 630.1404. The corresponding objective value without symmetry is 630.4464, and the one without homogeneity is 630.1422. Thus, on the basis of log-likelihood tests, we fail to reject both of the hypotheses at the 5% significance level.

2.5.1 Prediction

We can contrast a measure of the predictive power of these estimates to those from the two-step method of Heien and Wessells (1990). Table 2-3 shows the correlation between observed and predicted shares for various estimators including both the nonlinear AIDS model (using Equation 3) and the linear AIDS approximation (Equation 4). Given that the estimates are based on a cross-section of households (with measurement errors in the price data), the correlations for all three estimators are surprisingly high.

---

[9] We are very grateful to Michael Ferris and GAMS Corporation for helping us in convert our primal nonlinear maximization problem to a dual one, thereby substantially decreasing computation time. This method is described in Dirkse and Ferris (1995), Ferris and Munson (1997), and Ferris and Horn (1998).

Except for fish, the predicted shares of the Heien-Wessells two-step estimator are more highly correlated with the actual data than those obtained using single-equation tobit estimates. Apparently the information from the cross-equation restrictions in the two-step estimator more than compensate for the greater efficiency of the tobit ML approach for a single equation.

The GME predicted shares are more highly correlated with the actual data than are the two-step estimator for the entire system and for each type of meat (with two exceptions). Surprisingly, the linear approximation method predicts better than does the nonlinear model for the Heien-Wessells method (though not for the GME method).

2.5.2 Elasticities and Confidence Intervals

Table 2-4 reports the Hicks-compensated price elasticities and expenditure elasticities for each type of meat and corresponding asymptotic standard errors.[10] All the own-elasticities have the expected signs. Most of the elasticities are statistically significantly different from zero at the 0.05 level on the basis of asymptotic t-tests.

2.5.3 Complete Demand System

We also estimated a "complete" AIDS model with the five meats and "all other goods." The consumption of "all other goods" was calculated by using households' income data and normalizing the price of all other goods to one. The complete, six-equation demand system uses income, whereas the meat-only, five-equation demand system uses expenditures on meat.

Table 2-5 shows the correlation between actual and estimated shares in the meat-only and the complete demand systems based on 500 observations. The correlation is much higher for each meat except fish in the complete system.

---

[10] To save space, we do not report the Marshallian price elasticities, though the own elasticities are reported in Table 6.

We obtain income elasticities with the complete model rather than the expenditure elasticities with the meats-only model. We see no obvious pattern between the two models in the expenditure and the income elasticities. The Hicksian price elasticities for the two models (not reported in the table) are close. The Marshallian meat demand elasticities are also close except for fish, where the complete system estimate is five times as large in absolute value due to the large difference between the expenditure and income elasticities.

2.5.4 Comparison with Other Estimates

We can find no other estimates of meat-demand price elasticities for Mexico.[11] Nor have we found meat demand studies for other countries that us cross-sectional data.[12] Table 2-6 compares our estimated Marshallian own-price elasticities to time-series studies of meat demand in wealthier countries.

Our estimated Marshallian demand elasticity for processed meat is -0.78. None of the other studies calculate an elasticity for this meat. The estimated Marshallian elasticities for Mexican beef and chicken lie at the high-end and the pork elasticity at the low-end of the ranges over other countries. The Mexican fish elasticity is much more elastic than the few other existing estimates in other countries.

[11] Heien, Jarvis, and Perali (1989) examine a nine-commodity food demand system (including meat) for Mexico. They also examine poultry, pork, and beef in more detail. As they lack price variation data, however, they cannot estimate elasticities. They describe their estimates as a demographically augmented Engel curve analysis.

[12] Deaton (1988) uses cross-sectional data to estimate price elasticities for beef, other meat, cereal, and starches for the Ivory Coast. He does not, however, break down the other meat category in further detail. Wales and Woodland's (1983) study of meat demand using cross-sectional Australian data shows only how demand varies with demographic characteristics. As with Heien, Jarvis, and Perali (1989), Wales and Woodland could not estimate price effects because they did not observe variations in price.

2.5.5 Demographic Effects

Because our demand system uses demographic variables as well as prices to explain demand, we can examine how changes in various demographic characteristics affect the price and expenditure elasticities. Table 2-7 shows how the share of meat expenditures responds as we change one demographic variable at a time, while holding other demographic variables at the sample means.

The top of the table shows the effect of changing each dummy variable from 0 to 1. The share of beef is 4.9 percentage points higher for a resident of an urban area than a comparable person who lives in a rural area. Urban dwellers also eat relatively more chicken and fish and less pork and processed meats. Female-headed households eat relatively less beef, pork, and chicken, and relatively more processed meats and fish. Households headed by someone with a college degree eat much more processed meats and much less pork than those headed by people who did not complete their elementary education.

The bottom part of the table shows the effect of changing a household's age composition. If a family adds a child under the age of 5 to a family of five, the share of beef falls 1.4 percentage points, the share of por4k falls by -2.4 percentage points, and the other shares change by smaller amounts. In general, changes in age composition have the largest effects for the two youngest and the oldest age groups.

2.6 Conclusions

Our generalized maximum entropy (GME) approach is a practical way to estimate systems of many equations with nonnegativity constraints. The GME approach has several advantages over traditional maximum likelihood (ML) methods.

First, because no assumptions about the error structure need be made to use the GME estimator and because it uses all the data, it is more robust and efficient than are ML estimators. The sampling experiments of Golan, Judge, and Perloff (1997) indicate that the balanced single-equation GME estimator is more efficient — has lower empirical mean square error — than the ML tobit estimator in small samples regardless of whether errors are normal or not. In particular, the predictive power of the GME estimator is greater than that of two-stage estimators or single-equation maximum likelihood estimators for the data set we examined.

Second, imposing inequality (nonnegativity) constraints, equality (various consumer theory) constraints, and nonlinear constraints is straight forward. In general, theoretical and other nonsample information may be directly imposed on the GME estimates much more easily than with classical ML or Bayesian techniques.

Third, GME performs well in both well-posed and ill-posed problems. Fourth, the GME objective can be adjusted to stress either precision or prediction. Fifth, the GME approach can be used with a larger number of censored equations than is practical to estimate with standard full-information maximum likelihood approaches. Finally, the GME is a one-stage procedure that is easy to compute and solve.

We use our GME method to estimate the demand for five types of meat using cross-sectional data from Mexico, where most households did not buy at least one type of meat during the survey week. Our estimates of the Marshallian elasticities of demand for Mexico are similar to estimates based on aggregate, time-series data from other, wealthier countries except for fish, where Mexican demand is more elastic. We believe that this study is the first to show how a system of demands varies across demographic groups.

References

Alston, Julian M., and James A. Chalfant, "Weak Separability and a Test for the Specification of Income in Demand Models with an Application to the Demand for Meat in Australia," *Australian Journal of Agricultural Economics*, 31(1), April 1987:1-15.

Amemiya, T., "Multivariate Regression and Simultaneous Equation Model When the Dependent Variables Are Truncated Normal." *Econometrica*. 42(6). November 1974:999-1012.

Blundell, Richard W., and Carlos Meghir, "Bivariate Alternatives to the Tobit Model," *Journal of Econometrics*, 34(1-2), January-February 1987:179-200.

Capps, Oral, Jr., Reyfong Tsai, Raymond Kirby, and Gary W. Williams, "A Comparison of Demands for Meat Products in the Pacific Rim Region," *Journal of Agricultural and Resource Economics*, 19(1), July 1994:210-24.

Cashin, Paul, "A Model of the Disaggregated Demand for Meat in Australia," *Australian Journal of Agricultural Economics*, 35(3), December 1991:263-83.

Chalfant, James A., "A Globally Flexible, Almost Ideal Demand System," *Journal of Business and Economic Statistics*, 5(2), April 1987:233-42.

Chalfant, James A., Richard S. Gray, and Kenneth J. White, "Evaluating Prior Beliefs in a Demand System: The Case of Meat Demand in Canada," *American Journal of Agricultural Economics*, 73(2), May 1991:476-90.

Csiszár, I., "Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems," *The Annals of Statistics*, 19, 1991:2032-2066.

Dahlgran, Roger A., "Is U.S. Meat Demand In Equilibrium?" *The Economics of Meat Demand*, R. C. Buse, ed., Proceedings of the Conference on The Economics of Meat Demand, October 20-21 1986, Charleston, South Carolina.

Deaton, Angus, "Quality, Quantity, and Spatial Variation of Price," *American Economic Review*, 78(3), June 1988:418-30.

Deaton, Angus, and J. Muellbauer, "An Almost Ideal Demand System," *American Economic Review*, 70(3), June 1980:312-26.

Deaton, Angus, and Margaret Irish, "Statistical Models for Zero Expenditures in Households Budgets," *Journal of Public Economics*, 23(1-2), February-March 1984:59-80.

Dirkse, S. P., and M. C. Ferris (1995), "The {PATH} solver: A Non-Monotone Stabilization Scheme for Mixed Complementarily Problems," *Optimization Methods and Software*, 5, 123-156.

Ferris, M. C., and T. S. Munson (1997), "Interface to {PATH} 3.0: Design, Implementation and Usage," Computer Science Department, University of Wisconsin at Madison.

Ferris, M. C., and J. D. Horn (1998), "{NLP2NMCP} Automatic Conversion of Nonlinear Programs into Mixed Complementary Problems," Computer Science Department, University of Wisconsin at Madison.

Flood, Lennart R., and Ali C. Tasiran, "A Monte Carlo Comparison of System Tobit Estimators," *Economics Letters*, 33(3), July 1990:249-54.

Golan, Amos, and George Judge (1996), "A Maximum Entropy Approach to Empirical Likelihood: Estimation and Inference," Working Paper, U.C. Berkeley. (Presented at the 1997 Summer Meetings of the North America Econometric Society).

Golan, Amos, George Judge, Jeffrey M. Perloff, "Recovering Information from Censored and Ordered Multinomial Response Data," *Journal of Econometrics*, forthcoming 1997.

Golan, Amos, George Judge and Douglas Miller, *Maximum Entropy Econometrics: Robust Estimation With Limited Data*, New York: John Wiley & Sons, 1996.

Halbrendt, Catherine, F. Tuan, C. Gempesaw, and D. Dolk-Etz, "Rural Chinese Food Consumption: The Case of Guangdong," *American Journal of Agricultural Economics*, 76(4), November 1994:794-99.

Hayes, Dermot J., Thomas I. Wahl, and Gary W. Williams, "Testing Restrictions on a Model of Japanese Meat Demand," *American Journal of Agricultural Economics*, 72(3), August 1990:556-566.

Heien, Dale, Lovell S. Javris, and Federico Perali, "Food Consumption in Mexico," *Food Policy*, May 1989:167-80.

Heien, Dale, and G. Pompelli, "The Demand for Beef Products: Cross-Section Estimation of Demographic and Economic Effects," *Western Journal of Agricultural Economics*, 13(1), July 1988:37-44.

Heien, Dale, and Cathy Roheim Wessells, "Demand Systems Estimation with Microdata: A Censored Regression Approach," *Journal of Business & Economic Statistics*, 8(3), July 1990:365-71.

Jaynes, E. T., "Information Theory and Statistical Mechanics," *Physics Review*, 106, 1957a:620-630.

Jaynes, E. T., "Information Theory and Statistical Mechanics, II," *Physics Review*, 108, 1957b:171-190.

Jaynes, E. T., "Prior Information and Ambiguity in Inverse Problems," *Inverse Problems*, D. W. McLaughin, ed., Providence, Rhode Island: American Mathematical Society, 1984.

Kay, J. A., Michael J. Keen, and C. N. Morris, "Estimating Consumption From Expenditure Data," *Journal of Public Economics*, 23(1-2), February-March 1984:169-82.

Keen, Michael J., "Zero Expenditures and the Estimation of Engel Curves," *Journal of Applied Econometrics*, 1(3), July 1986:277-86.

Kullback, J., *Information Theory and Statistics*, New York: John Wiley & Sons, 1959.

Levine, Rafi D. (1980): "An Information Theoretical Approach to Inversion Problems," *Journal of Physics*, A, 13, 91-108.

Lee, Lung-Fei, "Simultaneous Equations Models With Discrete and Censored Dependent Variables," in Charles F. Manski and Daniel. McFadden, eds., *Structural Analysis of Discrete Data With Econometric Applications*, MA: MIT Press, 1978:246-64.

Lee, Lung-Fei, and Mark M. Pitt, "Microeconometric Demand Systems With Binding Nonnegativity Constraints: The Dual Approach," *Econometrica*, 54(5), September 1986:1237-42.

Mittelhammer, R. C., and N. S. Cardell, "On the Consistency and Asymptotic Normality of the Data-Constrained GME Estimator in the GLM," Working Paper, Washington State University.

Moschini, Giancarlo, and Karl D. Meilke, "Modeling the Pattern of Structural Change in U.S. Meat Demand," *American Journal of Agricultural Economics*, 71(2), May 1989:253-61.

Moschini, Giancarlo, Daniele Moro, Richard D. Green, "Maintaining and Testing Separability in Demand Systems," *American Journal of Agricultural Economics*, 76(1), February 1994:61-73.

Owen, A. (1990), "Empirical Likelihood Ration Confidence Regions," *The Annals of Statistics*, 19, 1725-1747.

Qin, J., and J. Lawless (1994), "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22, 300-325.

Ransom, Michael R., "A Comment on Consumer Demand System with Binding Non-Negativity Constraints," *Journal of Econometrics*, 34(3), March 1987: 355-59.

Shannon, C. E., "A Mathematical Theory of Communication," *Bell System Technical Journal*, 27, 1948:379-423.

Shore, J. E., and R. W. Johnson, "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," *IEEE Transactions on Information Theory*, IT-26, 1980:26-37.

Skilling, J., "The Axioms of Maximum Entropy," in J. Skilling, ed., *Maximum Entropy and Bayesian Methods in Science and Engineering*, Dordrecht: Kluwer Academic, 1989:173-87.

Thurman, Walter N., "Have Meat Price And Income Elasticities Changed? Their Connection With Changes In Marketing Channels," *The Economics of Meat Demand*, R. C. Buse, ed., Proceedings of the Conference on The Economics of Meat Demand, October 20-21 1986, Charleston, South Carolina.

Wales, Terence J., and Alan D. Woodland, "Estimation of Consumer Demand Systems with Binding Non-Negativity Constraints," *Journal of Econometrics*, 21(3), April 1983:263-85.

2.7 Appendix: Properties of the Generalized Maximum Entropy Estimator

Given four mild conditions, our GME estimator is consistent and asymptotically normal. These conditions are that (i) the errors' support $\underline{v}$ is symmetric, (ii) z_δ spans the true values for each one of the unknown parameters [δ = (ρ, γ, β, ϕ)], (iii) the errors are independently and identically distributed, and (iv) the operator X is of full rank. The proofs of consistency and asymptotic normality follow immediately from those in Golan, Judge, and Miller (1996), Golan, Judge, and Perloff (1997), and Mittelhammer and Cardell (1996).[13] These asymptotic properties can also be established via the empirical likelihood approach (Owen, 1990, and Qin and Lawless, 1994, Golan and Judge, 1996).

On the basis of these results, the entropy ratio statistic for the different parameters of the unknown distribution generating the data have a limiting χ^2 distribution and are used to obtain confidence intervals. Let $\underline{\lambda} \equiv (\underline{\lambda}, \underline{\lambda}, ..., \underline{\lambda}, \underline{\lambda})'$ be the vector of Lagrange multipliers for the $2n$ share equations where $\underline{\lambda}$ and $\underline{\lambda}$ stand for the n multipliers associated with the equality and inequality constraints (for each good i = 1, 2,..., n) respectively. Let $H_u()$ be the objective (total entropy) value for the complete AIDS-GME model where *none of the parameters* $\underline{\delta}$ = ($\underline{\zeta}$, $\underline{\gamma}$, $\underline{\beta}$, ϕ)' *are constrained*, or similarly, none of the elements of $\underline{\lambda}$ are constrained. Thus, $H_u()$ is just the optimal value of Equation (10). Next, let $H_M(\underline{\lambda}_0)$ be the entropy value of the constrained problem where $\underline{\lambda}$ = $\underline{0}$, or equivalently all the parameters are constrained to be zero (or at the *center* of their supports). Thus, $H_M(\underline{\lambda}_0)$ is the maximum value of the joint entropies (objective function). It can be obtained by maximizing Equation (10) subject to no constraints (except for the requirements

[13] Note that asymptotic normality is *not affected* by nonlinear price equation as the proof is based on the assumptions on the errors and the support spaces and thus the unknown Lagrange parameters are not affected at the limit.

that all distributions are proper). Doing so yields the total entropy value of the two discrete, uniform distributions **q** and **w**,

$$H_M(\underline{\lambda}_0) = K\ln(D) + nT\ln(H)$$

where K is the total number of parameters to be estimated, D is the dimension of the support space for each one of the K parameters (taken here to be the same for all $k = 1,2,...K$, to simplify exposition), n is the number of data equations, and T is the total number of observations.

Then, the entropy ratio statistic for testing the null hypothesis that H_0: $\underline{\delta} = \underline{0}$ is

$$\zeta(\underline{\delta}=\underline{0}) = 2H_M(\underline{\delta}=\underline{0}) - 2H_u(\hat{\underline{\delta}}).$$

Under the four mild assumptions we made above (or the assumptions of Owen, 1990 and Qin and Lawless, 1994), $\zeta(\underline{\delta}=\underline{0}) \to \chi^2_{(K)}$ as $T \to \infty$ when H_0 is true. The approximate α-level confidence intervals for $\underline{\delta}$ are obtained by setting $\zeta(\underline{\delta}) \le C_\alpha$, where C_α is chosen so that $Pr\left[\chi^2_{(K)} < C_\alpha\right] = \alpha$. Similarly, we can test any other hypothesis H_0: $\underline{\delta} = \underline{\delta}_0$ for all, or any subset, of the parameters.

For example, let $2n$ be the number of symmetry requirements on $\underline{\gamma}$ (Equation 15), then H_0, $\underline{\gamma}_{ij} = \underline{\gamma}_{ij}$. The entropy ratio statistic is

$$\zeta(\hat{\gamma}_{ij}=\hat{\gamma}_{ji}) = 2H_u(\hat{\gamma}_{ij}\neq\hat{\gamma}_{ji}) - 2H_u(\hat{\gamma}_{ij}=\hat{\gamma}_{ji}),$$

where we use the symbol "≠" to indicate that the restrictions are *not* imposed. If H_0 is true, $\zeta(\hat{\gamma}_{ij}=\hat{\gamma}_{ji}) \to \chi^2_{(2n)}$ as $T \to \infty$.

We use the same line of reasoning as above (each constraint, or data point, represents additional potential information that may lower the value of the objective function but can never increase it) to derive a "goodness of fit" measure for our estimator:

$$R^* = 1 - \frac{H_u(\hat{\underline{\lambda}})}{H_M(\underline{\lambda}=\underline{0})},$$

where $R^* = 0$ implies no informational value of the data set, and $R^* = 1$ implies perfect certainty or perfect in-sample prediction.

The small-sample approximated variances can be computed in a number of ways. We discuss two approaches here. For simplicity, we discuss the results for a single equation. The first method, is to calculate

$$\hat{\sigma}_i^2 = \frac{1}{T} \sum_t \hat{\varepsilon}_{it}^2,$$

where $\hat{\varepsilon}_{it} \equiv \sum_h v_h \hat{w}_{ith}$ and $\hat{\text{var}}(\hat{\delta}_{ik}) \equiv \hat{\sigma}_i^2 (X'X)^{-1}$ for each parameter δ_k of $\underline{\delta} = (\underline{\rho}, \underline{\gamma}, \underline{\beta}, \underline{\phi})'$.

Because our model is a system of five equations, the elements of the asymptotic variance-covariance matrix, Ω, for the error terms of the entire system are

$$\hat{\sigma}_{ij}^2 = \frac{1}{T} \sum_t \hat{\varepsilon}_{it} \hat{\varepsilon}_{jt},$$

where $i, j = 1, \ldots, 5$ (for the five meats). Thus, the variance matrix of the GME estimated parameters is

$$C - C(RCR')^{-1}C$$

where $C = (Q'\Omega^{-1}Q)^{-1}$, R is a matrix showing the linear restrictions imposed ont he parameters (such as homogeneity and symmetry), and Q is the kronecker product of X and a 5×5 identity matrix.

A second approach for estimating σ makes use of the symmetry of the errors' support $\underline{v}$ around zero (see for example, Mittelhammer and Cardell, 1996 for the linear regression model). Given the symmetry of $\underline{v}$ and , the 's are also identically and independently distributed with a mean of zero. Then, σ is the ratio of $\hat{\sigma}_i^2(\hat{\underline{\lambda}}_i)$ and the expected inverse of the variance of the estimated probabilities $\underline{w}_{it}$ defined over $\underline{v}$. Specifically, let $\sigma_i^2(\underline{\hat{\lambda}}_i) = \frac{1}{2T}\sum_t \hat{\lambda}_{it}^2$ for $(\underline{\lambda}_j \equiv \underline{\lambda}, \underline{\lambda})'$ and be the variance of the distribution w_{it} over the support $\underline{v}$. Then,

$$\tilde{\sigma}_i^2 = \frac{\hat{\sigma}_i^2(\underline{\hat{\lambda}}_i)}{\left(E\left[\hat{\sigma}_{it}^2(\underline{\hat{w}}_{it})\right]^{1}\right)^2} \cdot$$

and $\hat{var}(\hat{\delta}_{ik}) \equiv \tilde{\sigma}_i^2(X'X)^{-1}$. As a practical matter, the differences among the two approaches is negligible, but the second may be slightly preferred for very small samples, while the first approach is much simpler.

Table 2-1 Summary Statistics

	All 7,897 Observations		Random Sample of 1,000 Observations	
	Mean	SD	Mean	SD
Expenditure Share of Beef Consumption	0.388	0.342	0.369	0.335
Expenditure Share of Pork Consumption	0.123	0.237	0.127	0.245
Expenditure Share of Chicken Consumption	0.301	0.322	0.318	0.328
Expenditure Share of Processed Consumption	0.142	0.247	0.136	0.235
Expenditure Share of Fish Consumption	0.046	0.143	0.049	0.150
Natural Log of Price of Beef	9.552	0.272	9.544	0.278
Natural Log of Price of Pork	9.433	0.236	9.432	0.227
Natural Log of Price of Chicken	8.947	0.305	8.952	0.304
Natural Log of Price of Processed Meat	9.484	0.297	9.483	0.295
Natural Log of Price of Fish	9.365	0.467	9.352	0.459
Natural Log of Expenditure on Meats	16.025	0.891	16.030	0.871
Household Lives in Urban Area	0.622	0.485	0.611	0.488
Household Head is Female	0.128	0.334	0.120	0.325
Household Head is in School	0.021	0.142	0.025	0.156
Household Head Attended:				
Primary School	0.516	0.500	0.548	0.498
Secondary School	0.176	0.381	0.173	0.378
Preparatory or Vocational School	0.078	0.268	0.068	0.252
College	0.095	0.294	0.082	0.275
Share of Household Members				
Between 0 and 5 Years Old	0.139	0.170	0.141	0.173
Between 6 and 15	0.212	0.209	0.222	0.214
Between 16 and 28	0.257	0.247	0.260	0.244
Between 29 and 45	0.206	0.204	0.199	0.194
Between 46 and 60	0.104	0.196	0.106	0.185

Table 2-1 GME Estimates of LA/AIDS Meat Demand System

	Beef	Pork	Chicken	Processed Meat	Fish
Intercept	-0.8330*	-0.0435	1.2382*	0.7290*	-0.0907
Beef Price	-0.1348*	-0.0019	0.0728*	0.0372	0.0267
Pork Price	-0.0019	0.0537	-0.0123	-0.0110	-0.0285*
Chicken Price	0.0728*	-0.0123	0.0128	-0.0674*	-0.0059
Processed Meat Price	0.0372	-0.0110	-0.0674*	-0.0241	0.0653*
Fish Price	0.0267	-0.0285*	-0.0059	0.0653*	-0.0576*
Expenditure	0.1126*	0.0190	-0.0812*	-0.0625*	0.0121*
Household is Urban	0.0487*	-0.0453*	0.0038	-0.0203	0.0131
Household Head is Female	-0.0250	-0.0116	-0.0177	0.0837*	-0.0294*
Household Head is in School	-0.0455	-0.0397	0.1016	-0.0560	0.0397
Household Head Attended:					
Primary	0.0388	-0.0577	0.0109	0.0206	-0.0126
Secondary School	0.0059	-0.0820*	-0.0259	0.1043*	-0.0024
Preparatory or Vocational School	0.0452	-0.1120*	0.0055	0.0793*	-0.0181
College	0.0005	-0.1075*	-0.0488	0.1577*	-0.0018
Share of Household Members					
Between 0 and 5 Years Old	-0.2546*	-0.0077	0.1801*	0.0733	0.0089
Between 6 and 15	-0.1572*	0.0177	0.0899	0.0661	-0.0165
Between 16 and 28	-0.0437	0.1403	-0.0861	0.0246	-0.0351
Between 29 and 45	-0.0886	0.0132	-0.0222	0.1189*	-0.0212
Between 46 and 60	-0.0984	-0.0138	0.0290	0.0757	0.0075
Φ			-4.3838		

* The |asymptotic t-statistic| ≥ 1.96 (5% confidence level).
Note: Sample size is 1,000 observations.

Table 2-3 Correlations between Observed and Predicted Shares

	GME		2-Step Estimator Heien and Wessells (1989)		Least Squares*
	Nonlinear	Linear	Nonlinear	Linear	Nonlinear
Beef	0.347	0.339	0.273	0.263	0.214
Pork	0.198	0.199	0.200	0.194	0.193
Chicken	0.286	0.282	0.209	0.202	0.172
Processed Meat	0.261	0.260	0.230	0.308	0.219
Fish	0.163	0.160	0.127	0.158	0.028
Entire Demand System	0.489	0.487	0.315	0.468	0.245

*Nonlinear least squares of the AIDS model: Nonnegativity constraint is ignored.
Note: Sample size is 1,000 observations.

Table 2-4 Estimated Hicks Price and Expenditure Elasticities
(Asymptotic Standard Error)

	Beef	Pork	Chicken	Processed Meat	Fish
Beef Price	-0.596	0.187	0.228	0.015	0.166
	(0.138)	(0.068)	(0.077)	(0.064)	(0.041)
Pork Price	0.550	-0.418	0.081	-0.059	-0.153
	(0.365)	(0.255)	(0.196)	(0.150)	(0.113)
Chicken Price	0.263	0.034	-0.402	0.111	-0.006
	(0.130)	(0.073)	(0.104)	(0.064)	(0.043)
Processed Meat Price	0.041	-0.052	0.255	-0.706	0.462
	(0.316)	(0.163)	(0.176)	(0.168)	(0.093)
Fish Price	1.236	-0.400	-0.034	1.285	-2.088
	(1.132)	(0.436)	(0.725)	(0.330)	(0.280)
Expenditure	1.305	1.149	0.745	0.542	1.247
	(0.037)	(0.082)	(0.043)	(0.072)	(0.130)

Note: The elasticities are calculated at the sample means.

Table 2-5 Comparisons of Meat-Only and Complete Demand Systems

	Correlation		*Elasticities*	
			Expenditure	*Income*
	Meat Only	*Complete*	*Meat Only*	*Complete*
Beef	0.305	0.580	1.131	0.420
Pork	0.228	0.391	1.218	0.794
Chicken	0.255	0.498	0.952	1.115
Processed Meat	0.334	0.559	0.303	0.358
Fish	0.250	0.176	2.013	-0.008
Other Goods	N/A	0.910	N/A	1.045
System	0.484	0.991		

Note: The estimates for both linear models are based on a sample of 500 observations.

Table 2-6 Estimated Marshallian Own-Price Elasticities in Other Studies

	Period	Country	Beef	Pork	Chicken	Fish
GME Model*	1992	Mexico	-1.078	-0.564	-0.638	-2.149
Cashin (1991)	1960-90	Australia	-1.24	-0.83	-0.47	NA
Chalfant, Gray, & White (1991)	1960-88	Canada	-0.96	-0.73	-0.91	-0.20
Hayes, Wahl, & Williams (1990)	1965-86	Japan	-1.89	-0.76	-0.59	-0.70
Capps (1994)	1960-88	S. Korea	-0.939	-0.647	-0.470	NA
Capps (1994)	1968-91	Taiwan	-1.158	-0.919	-0.278	NA
Chalfant (1987)	1947-87	U. S.	-0.37	-0.67	-0.51	-0.23
Moschini & Meilke (1989)	1967-87	U. S.	-1.05	-0.84	-0.10	-0.20
Thurman (1989)	1955-83	U. S.	-0.11	-0.73	-0.41	NA
Dahlgran (1989)	1950-85	U. S.	-0.66	-0.58	-0.60	NA

Notes:

- Only our GME study uses the nonlinear AIDS model, cross-sectional data, and imposes nonnegativity constraints. The other studies are maximum-likelihood, LA/AIDS models based on annual data unless otherwise stated.
- The asymptotic standard errors for the four GME elasticities shown are 0.114, 0.237, 0.124, and 0.234. The GME Marshallian elasticity for processed meat is -0.780 with an asymptotic standard error of 0.194.
- Cashin: Quarterly data. Elasticities are for 1985:4. Elasticities vary little over time. Lamb elasticity is -1.326.
- Chalfant et al.: Estimates of the AIDS model without concavity restriction imposed.
- Hayes et al.: Elasticity for Wagyu, Japanese domestic beef. Imported-beef elasticity is -0.46. Test that the domestic beef is a perfect substitute for the import beef is strongly rejected. Estimation period is not clear from the paper.
- Capps: Rotterdam Demand Model.
- Moschini-Meilke: Post-structural change, time-varying coefficients, based on quarterly data.
- Thurman: Quad-log demand system. Estimated price elasticities are for 1983 observations, with symmetry restriction imposed.
- Dahlgran: Rotterdam Demand Model.

Table 2-7 Percentage Point Change in Share Due to a Change in an Exogenous Variable

	Beef	*Pork*	*Chicken*	*Processed Meats*	*Fish*
Change from 0 to 1					
Urban	4.9	-4.5	0.4	-2.0	1.3
Female	-2.5	-1.2	-1.8	8.4	-2.9
Household Head is in School	-4.6	-4.0	10.2	-5.6	4.0
Primary	3.9	-5.8	1.1	2.1	-1.3
Secondary	0.6	-8.2	-2.6	10.4	-0.2
Preparatory	4.5	-11.2	0.6	7.9	-1.8
College	0.1	-10.8	-4.9	15.8	-0.2
*Increase of 1 Person**					
Age < 5	1.4	-2.4	-0.6	1.3	0.3
5 ≥ Age < 15	-2.8	-0.7	3.1	-0.1	0.4
15 ≥ Age < 28	-1.2	-0.3	1.7	-0.2	0.01
28 ≥ Age < 45	0.7	1.7	-1.2	-0.9	-0.3
45 ≥ Age < 60	-0.04	-0.4	-0.2	0.7	-0.07
Age ≥ 60	1.4	-0.6	0.2	-1.3	0.3

* The typical family has five people. Initially, the family consists of one child less than 5 years old, one child between 5 and 15, two adults between 28 and 45, and one grandparent between 45 and 60 years old. We then calculate the change in shares resulting from adding one more person in one age group.

Chapter 3 Better Estimates of Agricultural Supply Response

3.1 Introduction

We propose a new, generalized maximum entropy (Golan, Judge, and Miller, 1996) approach to estimating the Nerlove model of agricultural supply response (Nerlove and Addison, 1958, Nerlove, 1979). We use simulations to show that two versions of the generalized maximum entropy estimator and two other shrinkage estimators with informative priors have much smaller mean square errors (MSE) and average bias than do the standard techniques used to estimate the Nerlove model. We also use one version of the GME approach to estimate structural parameters that cannot be estimated using standard methods.

The Nerlove model is one of the world's most widely-used econometric models and is frequently employed in development studies. Most of the literally hundreds of applications that use ordinary least squares (OLS) or nonlinear least squares suffer from two problems. First, because these least squares approaches can only estimate the reduced-form model, many structural parameters are not estimated. Second, estimates of the key, supply-response parameter are extremely variable, as the survey by Askari and Cummings (1977) illustrates.

One reason for the variability in OLS estimates is that the key estimated coefficient is a ratio of random variables. Zellner (1978) showed that the OLS reduced-form estimate of this coefficient possesses infinite moments of all orders and may have a bimodal distribution. Diebold and Lamb (1997) demonstrate that Zellner's minimum expected loss estimator (Zellner

1978; Zellner and Park 1979) of the reduced-form equation based on an uninformative prior (MELO-U) has smaller mean square error (MSE) than does OLS.

We compare the OLS and MELO-U estimates to generalized maximum entropy estimates of both the structural (GME-S) and the reduced-form (GME-R) versions of the Nerlove model. We also examine a MELO estimator with informative priors (MELO-I) and Zellner's (1996, 1997) Bayesian method of moments estimators without and with informative priors (BMOM-U and BMOM-I). In our simulation experiments, the four shrinkage approaches — MELO-I, BMOM-I, GME-R, and GME-S produce estimates of the key slope parameter of the Nerlove model that have much smaller average bias and MSE than do OLS or MELO-U or BMOM-U. We illustrate that the superiority of the MELO-I, GME-S, and GME-R estimators is not very sensitive to the amount of prior information provided (the specified support of the parameters and the error).

One advantage of the two GME approaches over the classical and Bayesian methods is that GME do not require distributional assumptions. Further, the GME-S approach provides estimates of coefficients that are unobtainable from the reduced-form approaches.

In Section 2, we describe the OLS, MELO-U, and MELO-I estimators of the Nerlove agricultural supply model. We present our GME estimators in Section 3. In Section 4, we describe the simulation results. We examine how sensitive the shrinkage estimators are to the prior information in Section 5. We draw conclusions in the final section.

3.2 The Agricultural Supply Model

The standard structural Nerlove model is[14]

[14] To facilitate comparison with Diebold and Lamb (1997), we use their specification of the standard model. Indeed, the entire discussion in this section follows the first section of their paper closely.

$$A_t^* = \alpha_0 + \alpha P_t^e + u_t , \tag{1}$$

$$P_t^e = P_{t-1}^e + \gamma \, (P_{t-1} - P_{t-1}^e), \tag{2}$$

$$A_t = A_{t-1} + \theta \, (A_t^* - A_{t-1}), \tag{3}$$

$$u_t \overset{iid}{\sim} (0 , \ \sigma_u^2 \ ), \tag{4}$$

where A is the crop acreage under cultivation, A^* is the desired acreage, P is the crop price, P^e is the expected price, and α_0, α, θ, γ, and σ_u^2 are parameters.

According to Equation (1), the desired acreage is a function of the expected price, where $\alpha \geq 0$, the slope of the desired acreage (supply) curve, is the key parameter that we want to estimate. Eq. (2) is an adaptive-expectations linking P^e to P. Muth (1960) shows that these adaptive expectations are rational if prices follow an integrated moving average process. Eq. (3) is a partial-adjustment mechanism relating A^* to A, where γ and θ are expected to be positive (and probably between zero and one).

The structural model, Eqs. (1) - (4) cannot be estimated by either OLS or MELO techniques because P^e and A^* are not observable. Instead, the reduced-form equation is estimated. This reduced-form specification is obtained by solving Eqs. (1) - (4) for acreage as a function of observable variables:

$$A_t = b_1 + b_2 \, P_{t-1} + b_3 \, A_{t-1} + b_4 \, A_{t-2} + e_t , \tag{5}$$

where

$$b_1 = \alpha_0 \, \gamma \, \theta$$

$$b_2 = \alpha \, \gamma \, \theta$$

$$b_3 = (1 - \gamma \ ) + (1 - \theta \ )$$

$$b_4 = -(1-\gamma)(1-\theta)$$

$$e_t = \theta u_t - [\theta(1-\gamma)] u_{t-1}.$$

3.2.1 Ordinary Least Squares

Traditionally, OLS (or nonlinear least squares) is used to estimate the reduced-form, Eq. (5). Doing so may not be appropriate because of the potentially serially-correlated disturbance and lagged dependent regresors. As OLS is virtually the only approach actually used, we follow Diebold and Lamb (1997) and use the OLS approach as the "straw man" base case. As they note, OLS is appropriate if farmers adapt their expectations quickly (γ is close to one) so that the reduced-form disturbance is approximately white noise. Similarly, if the supply-response equation's disturbance is approximately first-order autoregressive with parameter 1 - γ, the reduced-form disturbance is also approximately white noise.

The OLS approach estimates the key parameter α^o as a function of the reduced-form parameters:

$$\alpha^o = \frac{b_2^o}{\delta^o}, \quad (6)$$

where $\delta^o \equiv 1 - b_3^o - b_4^o$. As Zellner (1978, 1985, 1986), Zellner and Park (1979), Zaman (1981), and Diebold and Lamb (1997) observe, ratios or reciprocals of random variables have Cauchy tails and hence no finite moments. In addition, the distributions of reciprocals or ratios may be multimodal as Zellner (1978) shows for the normal distribution and Lehmann and Popper Shaffer (1988) show for more general distributions. The nonexistence of moments and the multimodality may contribute to substantial variability in estimates of agricultural supply response. In none of our experiments, however, do we observe multimodality.

3.2.2 Minimum Expected Loss Estimators

Zellner's (1978) minimum-expected-loss (MELO) estimator minimizes posterior expected loss. It has finite first and second moments and finite risk with respect to generalized quadratic loss in small and large samples and is consistent, asymptotically efficient, and asymptotically normal. Because of these characteristics, Diebold and Lamb (1979) recommend using MELO-U instead of OLS to estimate α.[15] The MELO-U shrinkage estimator is

$$\alpha^{mu} = \frac{E(b_2^{mu})}{E(\delta^{mu})} \frac{1 + Cov(b_2^{mu}, \delta^{mu}) / [E(b_2^{mu}) E(\delta^{mu})]}{1 + Var(\delta^{mu}) / E^2(\delta^{mu})} = \frac{E(b_2^{mu})}{E(\delta^{mu})} F^{mu}, \tag{7}$$

where F^{mu} is the shrinkage factor and $\delta^{mu} = 1 - b_3^{mu} - b_4^{mu}$.

In the MELO-I approach, we use Geweke's (1986) method to impose our priors (inequality restrictions) and then base the MELO-I estimates on those Bayesian estimates.[16] We use informative priors about the parameters $\mathbf{b} = (b_1, b_2, b_3, \text{and } b_4)'$ and the error term. All the prior restrictions are of the form that coefficients and the residuals lie in a specified range.

The MELO posterior conditional distribution of **b** given the data and σ follows a multivariate normal distribution:

$$\{b_{MELO} | \sigma, X, A\} \sim MVN\left((X'X)^{-1} X'A,\ (X'X)^{-1/2} \sigma\right), \tag{8}$$

[15] See their article for details on this estimator. Diebold and Lamb also note that one could estimate a reduced-form model with a lagged dependent variable and serially correlated errors using the method of Zellner and Geisel (1970) or BMOM.

[16] In most of our simulations, the initial Bayesian estimates and the MELO-I estimates differ only in the third or fourth place. Consequently, we discuss only the MELO-I except in the one case where the two estimates differ by more.

where X is a $T \times 4$ (where T is the number of observations) matrix of independent variables in the reduced-form regression, Equation 5. The posterior distribution of σ given the data follows an inverted gamma distribution

$$Pr(\sigma | X, A) \propto \frac{1}{\sigma^{v+1}} \exp\left(-\frac{vs^2}{2\sigma^2} \right), \tag{9}$$

where $v = T - 4$ is the degrees of freedom (given four regressors in the reduced-form Equation 5) and s^2 is the variance of the residuals.

To impose the inequality restrictions that coefficients lie within certain ranges, we use Geweke's (1986) technique. The first step is to generate a large number of random samples of **b** and ε using their known posterior distributions. We generate σ^2 using Equation 9. We then generate the vector **b** using Equation 8 conditional of those generated σ^2. Finally, we compute a vector of the residuals as $\varepsilon = A - X\mathbf{b}$. In the second step, we reject those observations for which either **b** or ε fail to satisfy the prior restrictions. The remaining observations represent a truncated posterior distribution of observations for b and that are consistent with the prior restrictions.

The MELO-I shrinkage estimator is

$$\alpha^{mi} = \frac{E(b_2^{mi})}{E(\delta^{mi})} \frac{1 + Cov(b_2^{mi}, \delta^{mi}) / [E(b_2^{mi})\ E(\delta^{mi})]}{1 + Var(\delta^{mi}) / E^2(\delta^{mi})}, \tag{10}$$

where $\delta^{mi} = 1 - b_3^{mi} - b_4^{mi}$.

3.2.3 Bayesian Method of Moments Estimator

We also use Zellner's (1996, 1997) Bayesian method of moments (BMOM) to obtain estimates of the reduced-form Nerlove model in our experiments. BMOM allows researchers to make inverse probability statements regarding a given set of data when they lack information about the forms of the likelihood function and the prior distribution. We can derive the proper maxent (maximum entropy) posterior distribution of BMOM estimates of $\mathbf{b} = (b_1, b_2, b_3, b_4)'$ in the Nerlove model given σ^2 and the data, which Zellner (1996) showed to be a multivariate normal density:

$$\{b_{BMOM} \mid \sigma, X, A\} = MVN\left((X'X)^{-1} X'A, (X'X)^{-1/2}\sigma\right).$$

Zellner also showed that the proper maxent posterior distribution for σ^2 is an exponential distribution with a mean equal to s^2:

$$Pr\left(\sigma^2\right) = \frac{1}{s^2}\exp(-\sigma^2 / s^2).$$

To estimate α using the BMOM-U method, we follow Zellner (1994) and use a balanced loss function (BLF), which minimizes the weighted average of the posterior prediction loss. Our BMOM-U estimate is

$$\alpha^{bu} = 0.5\,\frac{E(b_2^{bu})}{E(\delta^{bu})} + 0.5\frac{E(b_2^{bu})}{E(\delta^{bu})}\frac{1 + Cov(b_2^{bu}, \delta^{bu}) / [E(b_2^{bu})E(\delta^{bu})]}{1 + Var(\delta^{bu}) / E^2(\delta^{bu})}, \qquad (11)$$

where $\delta^{bu} = 1 - b_3^{bu} - b_4^{bu}$.

We again use Geweke's (1986) method to impose prior restrictions to obtain the truncated BMOM-I posterior distributions of **b** and ε. The BMOM-I estimate of α using a BLF is

$$\alpha^{bi} = 0.5\ \frac{E(b_2^{bi})}{E(\delta^{bi})} + 0.5\ \frac{E(b_2^{bi})}{E(\delta^{bi})} \frac{1 + Cov(b_2^{bi}, \delta^{bi}) / [E(b_2^{bi})\, E(\delta^{bi})]}{1 + Var(\delta^{bi}) / E^2(\delta^{bi})}\ , \qquad (12)$$

where $\delta^{bi} = 1 - b_3^{bi} - b_4^{bi}$.

3.3 A Generalized Maximum Entropy Approach

As an alternative, we use a GME approach to estimate both the structural and reduced-form agricultural supply models. We start by providing some intuition as to how the maximum entropy approach works, and then develop the GME estimator.

3.3.1 Maximum Entropy

The traditional maximum entropy (ME) formulation is based on the entropy-information measure of Shannon (1948). It is developed and described in Jaynes (1957a, 1957b), Kullback (1959), Levine (1980), Jaynes (1984), Shore and Johnson (1980), Skilling (1989), Csiszár (1991), and Golan, Judge, and Miller (1996). Shannon's (1948) entropy measure reflects the uncertainty (state of knowledge) we have about the occurrence of a collection of events. Letting x be a random variable with possible outcomes x_s, s = 1, 2, ..., n, with probabilities δ_S such that $\sum_S \delta_S \ln \delta_S$, Shannon (1948) defined the *entropy* of the distribution $\underline{\delta} = (\delta_1, \delta_2, ..., \delta_n)'$, as

$$H \equiv - \sum_s \delta_s \ln\ \delta_s\ , \qquad (13)$$

where 0 ln 0 ≡ 0. The function H, reaches a maximum of ln(n) when $\delta_1 = \delta_2 = \ldots = \delta_n = 1/n$. It is zero when $\delta_s = 1$ for one value of s.

To recover the unknown probabilities $\underline{\delta}$ that characterize a given data set, Jaynes (1957a, 1957b) proposes maximizing entropy, subject to available sample-moment information and adding up constraints on the probabilities. This procedure has an intuitive appeal.

Suppose we have a sample of T draws of the identically and independently distributed random variable x. Because the draws are independent, a list of the number of times each value occurs contains all of the information this experiment provides about the random variable: The order contains no information about the probabilities. We define the *outcome* of the experiment as a vector $\underline{f} = (f_1, f_2, ..., f_n)$, where f_s is the number of times x_s occurs and $\sum_s f_s = T$. A particular outcome may be obtained in a number of ways. For example, the outcome (1, T-1, 0, 0, ..., 0) can occur in T possible ways because x_1 may be observed in any of the T draws. In contrast, the outcome (T, 0, 0 ...0) can occur in only one way, where x_1 is drawn each time.

Define $v(\underline{f})$ as the number of ways that a particular outcome can occur. Suppose we have no information about the draws and are asked which outcome is the most likely. An "intuitively reasonable" response is that the outcome that can occur in the most number of ways, $\underline{f}^* \equiv \text{argmax } v(\underline{f})$, is the most likely outcome. Equivalently, we would consider it more likely to observe the frequency $\underline{f}^*/T$ than any other frequency. Shannon (1948) shows that, in the limit as $T \rightarrow \infty$, choosing $\underline{f}$ to maximize $v(\underline{f})$ is equivalent to choosing $\underline{\delta}$ to maximize the entropy measure, $H(\underline{\delta})$.

Thus, the frequency that maximizes entropy is an intuitively reasonable estimate of the true distribution when we lack any other information. If we have information from the experiment, such as the sample moments, or non-sample information about the random variable, such as restrictions from economic theory, we alter this intuitively reasonable estimate. The ME method chooses the distribution that maximizes entropy, subject to the sample and non-sample information. That is, out of all the possible estimates or probability distributions that are consistent with the sample and nonsample data, the ME method picks the one that is most uninformative: closest to a uniform distribution. In this sense, the ME estimator is conservative.

3.3.2 GME Estimator

The GME objective is a dual-criterion function that depends on the weighted sum of the entropy measures from both the unknown and unobservable coefficients (α_0, α, γ, θ) in the structural model or b_1, b_2, b_3, b_4 in the reduced-form model) and the error terms (u_t, structural, or e_t, reduced form). By increasing the weight on the error component of entropy, we improve the accuracy of estimation (decrease the MSE of the estimates of the coefficients). By increasing the weight on the coefficient component of entropy, we improve prediction. The ME estimator is a special case of the GME, in which no weight is placed on the noise component. In the following, our GME objective weights the coefficient and error entropies equally because we lack any theory that suggests other weights.[17]

We start with a GME estimate of the reduced-form model, Eq. (5), which we call GME-R. We estimate b_1, b_2, b_3, and b_4 to facilitate comparisons with previous papers. [Instead, we could estimate (α_0, α, b_3, and b_4).

Because the arguments of the entropy measures must be probabilities (Golan, Judge, and Miller, 1996; Golan, Judge, and Perloff, 1997), we reparameterize the coefficients to be proper probability distributions that are defined over some support. For example, for each reduced-form coefficient, b_i, we start by choosing a support space, which is a set of discrete points $\underline{z}^i = [\, \underline{z}_1^i, \underline{z}_2^i, \ldots, \underline{z}_M^i \,]'$ of dimension $M \geq 2$, that are at uniform intervals and that span the possible range of the unknown coefficients. Where we do not have knowledge about the coefficients from economic theory, we specify the supports to be uniformly distributed symmetric about zero with "large" negative and positive bounds. For the distribution to be symmetric around zero, M must

[17] We did, however, experiment with various weights ranging between zero and one. The mean square errors hardly vary as we change the weights in the following simulation experiments.

be an odd number. In our simulations, we use $M = 3$ because we find little if any gain from using larger M such as 5 or 7.

Then we let

$$b_i = \sum_{m=1}^{M} p_m^i \; z_m^i , \quad i = 1, \ldots, 4 ,$$

where the p are probabilities that correspond to the M-dimensional support vectors of weights with the property of probabilities that $\sum_{m=1}^{M} p_m^i = 1$ for $i = 1, \ldots, 4$.

Similarly, we treat the errors as unknowns and use the following parametrization. Let each $\underline{e}_t$ be specified as

$$e_t = \sum_{j=1}^{J} w_{tj} \; v_j ,$$

where $\sum_{j=1}^{J} w_{tj} = 1$ for all t, and $\underline{v}$ is a support space of dimension J greater than or equal to two that is symmetric about zero.

Having reparameterized the unknowns, we maximize the dual-loss (objective) function, which is the sum of the joint entropies of the signal and noise in the system. Letting $\underline{p}' = (\underline{p}^{1\prime}, \underline{p}^{2\prime}, \underline{p}^{3\prime}, \underline{p}^{4\prime})$ and $\underline{w} = (\underline{w}_1, \ldots, \underline{w}_T)'$, our problem is

$$\max_{\underline{p}, \underline{w}} H = -\underline{p}' \ln \underline{p}' - \underline{w}' \ln \underline{w}' \tag{14}$$

subject to

$$A_t = \sum_{m=1}^{M} p_m^1 \; z_m^1 + \sum_{m=1}^{M} p_m^2 \; z_m^2 \; P_{t-1} + \sum_{m=1}^{M} p_m^3 \; z_m^3 \; A_{t-1} + \sum_{m=1}^{M} p_m^4 \; z_m^4 \; A_{t-2} + \sum_{j=1}^{J} w_{tj} \; v_{tj}$$

$$\sum_{m=1}^{M} p_m^1 = \sum_{m=1}^{M} p_m^2 = \sum_{m=1}^{M} p_m^3 = \sum_{m=1}^{M} p_m^4 = \sum_{j=1}^{J} w_{tj} = 1.$$

The GME-R coefficient estimates are $\hat{b}_i = \sum_{m=1}^{M} \hat{p}_m^i z_m$. Mittelhammer and Cardell (1997) show that the GME estimates are consistent under standard regularity conditions for the generalized linear model. Thus, given our assumptions, the GME-R estimates are consistent.

The GME structural-model estimator (GME-S) is handled similarly (Golan, Judge, Miller, 1997). We need to reparameterize the coefficients of the structural model, Eqs. (1)-(3), in terms of probabilities. The probabilities $\underline{q}^0$ correspond to α_0, $\underline{q}^\alpha$ to α, $\underline{q}^\gamma$ to γ, $\underline{q}^\gamma$ to γ, and $\underline{\omega}$ to the error term u_t. Let $\underline{q}' = (\underline{q}^{0\prime}, \underline{q}^{\alpha\prime}, \underline{q}^{\gamma\prime}, \underline{q}^{\theta\prime})$. The supports are defined analogously. We maximize the dual-loss objective function (the sum of the joint entropies of the signal and noise in the system),

$$\max_{\underline{q},\, \underline{\omega}} H = -\underline{q}'\ln \underline{q}' - \underline{\omega}'\ln \underline{\omega}' \tag{15}$$

subject to

$$A_t^* = \sum_{m=1}^{M} q_m^0 z_m^0 + \sum_{m=1}^{M} q_m^\alpha z_m^\alpha P_t^e + \sum_{j=1}^{J} \omega_{tj} v_{tj}$$

$$P_t^e = P_{t-1}^e + \sum_{m=1}^{M} q_m^\gamma z_m^\gamma (P_{t-1} - P_{t-1}^e)$$

$$A_t = A_{t-1} + \sum_{m=1}^{M} q_m^\theta z_m^\theta (A_{t-1}^* - A_{t-1})$$

$$\sum_{m=1}^{M} q_m^0 = \sum_{m=1}^{M} q_m^\alpha = \sum_{m=1}^{M} q_m^\gamma = \sum_{m=1}^{M} q_m^\theta = \sum_{j=1}^{J} \omega_{tj} = 1.$$

The normality restriction in Eq. (4) is unnecessary for the GME approach, so we do not impose it. If we want to impose the restriction from theory that α is positive, we specify its support as being only in the positive range. Because we assume no overshooting, we specify the support of γ and θ to include the range from 0 and 1. [If we wanted to allow for overshooting, we would require only that the support be positive.]

Solving the problem in Eq. (15), we obtain our GME-S estimates of α_0, α, γ, and θ. In additional also obtain estimates of P and A^* for all t.

3.4 Experiments

To compare the sampling properties of OLS, MELO-U, MELO-I, BMOM-U, BMOM-I, GME-R, and GME-S estimators, we use Monte Carlo experiments. We confirm Diebold and Lamb's result that the MELO-U estimator has smaller MSE than OLS in small samples. We also show that the shrinkage estimators based on informative priors, MELO-I, BMOM-I, GME-R, and GME-S, dominate OLS, MELO-U, and BMOM-U in terms of mean square error and average bias.

3.4.1 Experimental Design

To facilitate comparison with Diebold and Lamb (1997), we use their experimental design. We generate 1,000 samples of data for each of various sets of parameters. In all experiments, $\theta = 0.5$ (moderate adjustment speed), $\alpha = 2$, $\alpha_0 = 0.25$ (subsistence farming). We vary the other parameters:[18]

$\gamma = 0.5$ and 1 (where OLS is appropriate)

[18] Diebold and Lamb (1997) report some additional intermediate parameter values. As the results tend to vary smoothly in those parameters, we only report the extreme cases to save space. We also tried varying α between 1 and 10 and found results that are qualitatively identical to our other experiments with $\alpha = 2$, so we do not report them to save space.

$\rho = 0.5$ and 0.9

$\sigma_u = 1, 2, 3$, or 5

$\sigma_\varepsilon = 1, 2, 3$, or 5,

where

$$(P_t - 100) = \rho(P_{t-1} - 100) + \varepsilon_t, \quad \varepsilon_t \sim IID\ N(0, \sigma_\varepsilon^2), \quad t = 1, 2, \ldots, T,$$

is the price generating process. The initial conditions are set at their expected values: $P_0 = 100$, $A_0 = A_{-1} = \alpha_0 + \alpha E(P)$, which equals 200.25, where $\alpha = 2$. Because $E(P) = 100$, the supply elasticity is approximately 1.

Unless otherwise stated, the sample size is $T = 25$, which is typical of most empirical work that uses annual time series of acreage and price. For a sample size of $T = 25$, we generate 27 observations and use the last 25 observations to estimate the reduced-form Eq. (5), which has a right-side variable with two lags. The structural model, Eqs. (1) - (3), involves only a single lag, so we use the last 26 observations. Thus, one advantage of using the structural model is that we gain an observation. The difference in performance between GME-S and OLS or MELO-U, however, has little to do with this extra observation. As we show below, the performance of GME-S is not very sensitive to the number of observations and not substantially different from GME-R, which only uses 25 observations.

We use the GAMS software program to obtain the GME estimates and Matlab to estimate the MELO and BMOM models using the same randomly generated samples. Each estimate takes only a few seconds of computer time. Our GME support spaces are [-20, 0, 20] for α_0, [-10, 0, 10] for α, and [0, 0.5, 1] for γ and θ. We follow Golan, Judge, and Miller (1996) in using a 3σ rule to determine the support for the error term in each sample (the standard deviation for A ranges from 4 to 6). The prior restrictions in MELO-I and BMOM-I are derived from the structural support: [-20, 20] for b_1, [-10, 10] for b_2, [0, 2] for b_3, and [-1, 0] for b_4.

The supports for α and α_0 contain the range of the OLS estimates in most of our experiments. We force γ and θ to be elements of [0, 1] based on economic theory.[19]

3.4.2 Results

The following tables and figure summarize our experimental results.[20] To examine the relative efficiency of the seven estimators OLS, MELO-U (uninformative), MELO-I (informative), BMOM-U (uninformative), BMOM-I (informative), GME-R (reduced form), and GME-S (structural), we use the mean-squared error (MSE) criterion. Choosing the MSE as our loss function favors MELO in the sense that it is designed to minimize that particular loss function. We also examine the relative predictive power of the four estimates using the correlation between the actual and predicted values of A.

Table 3-1 compares the MSEs of the seven estimators for $T = 25$, $\alpha = 2$, $\gamma = 1$, and various σ_ε and σ_u. Although OLS has infinite MSE in the population, its sample MSE is, of course, finite. We confirm Diebold and Lamb's finding that the MSEs for MELO-U are smaller than for OLS. The estimators based on informative priors, MELO-I, BMOM-I, GME-R, and GME-S have MSEs on the order of 10^{-2} to 10^{-5}, whereas the MSEs of OLS and MELO-U are many times larger (on the order of 10^{0} or 10^{-1}). The structural shrinkage estimator, GME-S, tends to have slightly smaller MSEs than do the reduced-form shrinkage estimators, MELO-I, BMOM-I, and GME-R.

---

[19] Allowing their supports to be larger, [0, 1, 2], would slightly improves the GME results in the following experiments.

[20] These tables replicate the qualitative results of Diebold and Lamb for the OLS and MELO estimators. Our quantitative results, however, differ from theirs. Russell Lamb graciously discussed this issue at length with us. Even after that discussion, however, we are unsure why our results differ as much as they do. Presumably, much or all of the difference is due to how the random numbers were generated. We use random numbers generated by GAMS or Shazam (which produced virtually identical results). Diebold and Lamb used a C program to generate their random numbers.

[21] See Zellner (1978) for a justification of the squared error loss function.

Tables 3-2 (ρ = 0.5) and 3-3 (ρ = 0.9) describe the empirical distributions of the estimated supply response coefficient, α, for $\alpha = 2$; $\sigma_\varepsilon = 1$; $\sigma_u = 5$; T = 25, 50, and 100; and γ = 0.5 and 1. When γ equals 1, the OLS approach is appropriate — and it performs better than when γ = 0.5. As γ increases or T increases, the advantage of MELO-U over OLS in terms of MSE shrinks, as Diebold and Lamb noted. As T increases, the MSE and the average bias for both OLS and MELO fall. As MELO-I, BMOM-I, GME-S, and GME-R estimates of α are virtually "perfect" — having almost no bias and little variance — even with small samples, there is little room for improvement with larger sample sizes.

An alternative way to illustrate the superiority of the shrinkage estimators is to compare the histograms of the various empirical distributions of , which we do Figure 3.1, where $\alpha = 2$, $\sigma_\varepsilon = 1$, $\sigma_u = 5$, $\gamma = 1$, $\theta = 0.5$, $\rho = 0.5$, and $T = 25$. As the distributions for GME-S and GME-R are virtually identical mass points at the true value of $\alpha = 2$, the figure shows a single distribution for both of them. The MELO-I and BMOM-I distributions are only slightly wider than the GME distributions, and much smaller than the OLS or MELO-U distributions.

These experiments are fairly benign ones where OLS does not perform extremely badly. The OLS sample distribution is single peaked and only a small fraction of α estimates are negative (which is inconsistent with the economic theory). The OLS estimates are likely to be much less precise when the denominator, $\gamma\theta$, of the ratio in the expression for α from the reduced form, $\alpha = b_2/\delta = b_2/(\gamma\theta)$, is nearly zero. We experimented with values of γ and θ where $\gamma\theta$ was close to zero and found that the variance of the empirical distributions of for OLS and MELO-U were very large in absolute value, as we would expect, but the MELO-I, BMOM-I, GME-R, and GME-S estimates were still tightly bunched around the true parameter value.[22]

We recover $\hat{\alpha}_0 = \hat{b}_1 / (1 - \hat{b}_3 - \hat{b}_4)$ and $\hat{\alpha}$ from the reduced-form estimators, and $\hat{\alpha}_0$, $\hat{\alpha}$, $\hat{\gamma}$, and $\hat{\theta}$ from the GME-S structural estimator, as Table 3-4 shows. As with $\hat{\alpha}$, the shrinkage estimators, MELO-I, BMOM-I, GME-R, and GME-S, have much smaller MSEs for $\hat{\alpha}_0$ than do OLS, MELO-U, or BMOM-U. The structural estimator, GME-S, has the smallest MSEs by orders of magnitude. This table also shows the GME-S estimates for γ and θ, which cannot be estimated using the reduced-form approaches. In our experiments, the GME-S estimates of $\hat{\gamma}$ have smaller MSEs than those for $\hat{\theta}$.[23]

All of the methods predict the crop acreage under cultivation, A. In addition, the GME-S approach provides estimates of the expected acreage, A^*, and the expected price, P^e. Table 3-5 shows the correlations between the predicted and actual values for each method where $\alpha = 2$, $\sigma_\varepsilon = 1$, $\sigma_u = 5$, $\gamma = 0.5$ or 1, and $\rho = 0.5$ or 0.9. We expected that the maximum likelihood techniques, OLS, MELO-I, MELO-U, would predict acreage slightly better than do BMOM-U, BMOM-I, GME-R, and GME-S (where the objective is a balance between prediction and accuracy of estimation). MELO-I and BMOM-I perform as well or better than do OLS, MELO-U, and BMOM-U. GME-R does almost as well as the other estimators; however, the correlations for GME-S are slightly lower than for the reduced-form models.

Unlike the reduced-form methods, GME-S provides estimates of the predicted price, P, and desired acreage, A^*, as Table 3-6 shows.[24] In our experiments, GME-S does a remarkable job of predicting the expected price — the correlations range from 0.767 to 0.975 — but is less impressive in its estimates of the expected acreage — the correlations range from 0.323 to 0.596.

Table 3-7 shows how the MSEs change if the distribution of the errors is nonnormal. In particular, we draw the errors u and ε from a λ^2 distribution or from a t distribution with either five or seven degrees of freedom. As the GME estimators do not depend on assumptions about

---

[22] Here, the MELO-I did perform better than the standard, informative Bayesian estimator.

[23] Especially when γ is close to 1, we get better estimate of $\hat{\gamma}$ and $\hat{\theta}$ if we use support [0, 1, 2] instead of [0, 0.5, 1], as in this table.

[24] The BMOM approach could be used to obtain estimates of the full structural model. One could also use the moments from the GME-S estimates to derive a MELO estimator (though we expect that the two estimates would be virtually identical in our experiments).

these distributions, we expected the GME to perform better. Indeed, the MSEs for the GME estimators are orders of magnitude smaller than for even the MELO-I or BMOM-I estimators. The MSEs of OLS, MELO-U, and BMOM-U are substantially larger than those of the other estimators.

3.5 Prior Information in Shrinkage Estimators

Our experiments indicate that the shrinkage estimators, MELO-I, BMOM-I, GME-R, and GME-S have much smaller MSEs than do the OLS and MELO-U estimators. In general, we know that prior information can lower the MSE compared to standard techniques provided the "prior information incorporated in the decision rule forecasts is at least reasonably accurate" (Zellner 1963).

One might ask, however, how sensitive our results are to our choice of prior information. All four shrinkage estimators specify the supports for the coefficients and for the error terms. [In addition, MELO-I requires an assumption about the error distribution.]

In Table 3-8, we show what happens to the MSE as the supports become wider (on the basis of 100 replications). If the supports on the reduced-form coefficients increase substantially (up to 40 times), the MSEs from MELO-I and GME-R tend to increase but remain below those of OLS and MELO-U. [Because the GME-S estimates different coefficients and has more supports, we cannot directly compare the effect of increasing supports to the reduced-form estimators, though the qualitative effects are similar.]

In our experiments, we use the sample standard deviation of A, $\hat{s}$, to set our supports for the error term. The GME estimators use a support of $[-3\hat{s}, 0, 3\hat{s}]$ for the error term. Table 3-8 shows what happens to the overall MSE if we widen the supports for the GME estimators or the prior restrictions for MELO-I and BMOM-I by a multiple of sigma. Hence, the effects of this experiment is different for the three estimators. The MSEs for MELO-I, GME-R, and GME-S tend to increase. The three-sigma rule for the GME estimators works reasonable well in obtaining a low MSE. Even with much wider supports (such as 200σ), MELO-I, GME-R, and GME-S have substantially smaller MSEs than do OLS or MELO-U.

Based on these results, we conclude that the MSEs are relatively insensitive to changes in the error support, and are only moderately sensitive to the support of the coefficients. Even with very large supports — for example, the largest support on α is [-400, 400] — these shrinkage estimators perform substantially better than do the estimators that are not based on informative priors. Therefore, one could impose "conservative" priors and still benefit significantly from using shrinkage estimators.

3.6 Conclusions

The traditional approach to estimating the Nerlove model of agricultural supply response using ordinary least squares produces highly variable estimates. Diebold and Lamb (1997) showed that this variability was partially due to the problem estimating a ratio using OLS. They further demonstrated that Zellner's minimum expected loss approach with an uninformative prior (MELO-U) produces less variable estimates. We show that Zellner's Bayesian method of moments estimator with an uninformative prior (BMOM-U) performs similarly to MELO-U.

Our simulation experiments demonstrate that the shrinkage estimators — the minimum expected loss approach with an informative prior (MELO-I), the Bayesian method of moments estimator with an informative prior (BMOM-I), and our generalized maximum entropy reduced-form and structural approaches (GME-R and GME-S) — have much lower average bias and mean square errors in small samples than do OLS, MELO-U, or BMOM-U. Moreover, in our experiments, the benefit of these informative shrinkage estimators occurs even when relatively little additional information.

Another advantage is that we can use the generalized maximum entropy approach to estimate the full structural agricultural supply response model, whereas we can only estimate a reduced-form Nerlove model using traditional methods. Consequently, GME allows us to recover parameters that cannot be estimated using reduced-form approaches.

The GME has two other advantages over classical and Bayesian approaches. First, the GME does not require us to make distributional assumptions, unlike the other approaches. In our simulations, the GME approaches performed much better than did the other methods when the underlying error distribution was nonnormal. Second, imposing inequality restrictions (such as bounding adjustment parameters between 0 and 1) is relatively easy. A similar GME approach will allow estimation of structural models instead of reduced-form equations in other problems besides that of agricultural supply.

References

Askari, A. and T. Cummings, 1977, Estimating agricultural supply response with the Nerlove model: A survey, International Economic Review 18, 257-329.

Csiszár, I., 1991, Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems, The Annals of Statistics 19, 2032-2066.

Diebold, F. X. and R. L. Lamb, 1997, Why are estimates of agricultural supply response so variable? Journal of Econometrics 76, 357-373.

Geweke, J., 1986, Exact inference in the inequality constrained normal linear regression model, Journal of Applied Econometrics, 1, 127-141.7

Golan, A., G. Judge, and D. Miller, 1996, Maximum Entropy Econometrics: Robust Estimation with Limited Data (John Wiley & Sons, New York).

Golan, A., G. Judge, and D. Miller, 1997, "Information recovery in simultaneous equation models, in A. Ullah and D. Giles, eds., Handbook of Applied Economic Statistics (Marcel Dekker, New York).

Golan, A., G. Judge, J. M. Perloff, 1997, Recovering information from censored and ordered multinomial response data, Journal of Econometrics 79, 1997:23-51.

Jaynes, E. T., 1957a, Information theory and statistical mechanics, Physics Review 106, 620-630.

Jaynes, E. T., 1957b, Information theory and statistical mechanics, II, Physics Review 108, 171-190.

Jaynes, E. T., 1984, Prior information and ambiguity in inverse problems, in D. W. McLaughin, ed., Inverse problems (American Mathematical Society, Providence, Rhode Island).

Kullback, J., 1959, Information theory and statistics (John Wiley & Sons, New York).

Lehmann, E. L. and J. Popper Shaffer, 1988, Inverted distributions, American Statistician 42, 191-194.

Levine, R. D., 1980, An information theoretical approach to inversion problems, Journal of Physics A 13, 91-108.

Muth, J., 1960, Optimal properties of exponentially weighted forecasts, Journal of the American Statistical Association 66, 299-305.

Nerlove, M., 1979, The dynamics of supply: Retrospect and prospect, American Journal of Agricultural Economics 61, 874-888.

Nerlove, M. and W. Addison, 1958, Statistical estimator of long-run elasticities of supply and demand, Journal of Farm Economics 40, 861-880.

Shannon, C. E., 1948, A mathematical theory of communication," Bell System Technical Journal 27, 379-423.

Shore, J. E., and R. W. Johnson, 1980, Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, IEEE Transactions on Information Theory, IT-26, 26-37.

Skilling, J., 1989, The axioms of maximum entropy, in J. Skilling, ed., Maximum Entropy and Bayesian Methods in Science and Engineering, (Kluwer Academic, Dordrecht), 173-87.

Zaman, A., 1981, Estimators without moments, Journal of Econometrics 15, 289-298.

Zellner, A., 1963, Decision rules for economic forecasting, Econometrica 31, 111-130.

Zellner, A., 1978, Estimator of functions of population means and regression coefficients including structural coefficients: A minimum expected loss approach, Journal of Econometrics 8, 127-158.

Zellner, A., 1985, Bayesian econometrics, Econometrica 53, 253-269.

Zellner, A., 1986, Further results on Bayesian minimum expected loss estimates and posterior distributions for structural coefficients, in D. J. Slottje, ed., Advances in econometrics (JAI Press, Greenwich, CT) 171-182.

Zellner, A., 1994, Bayesian and non-Bayesian estimation using balanced loss functions, in S. S. Gupta and J. O. Berger (eds), Statistic Decision Theory and Related Topics V (Springer-Verlag, New York, NY) 377-390.

Zellner, A., 1996, Bayesian method of moments/instrumental variables (BMOM/IV) analysis of mean and regression models, in J. Lee, W. Johnson, and A. Zellner (eds), *Modeling and Prediction: Honoring Seymour Geisser* Springer-Verlag.

Zellner, A., 1997, The finite sample properties of simultaneous equations estimates and estimators: Bayesian and non-Bayesian approaches, *Bayesian Analysis in Econometrics and Statistics: The Zellner View and Papers.* Edward Elgar Publishing Limited, Cheltenham, United Kingdom.

Zellner, A. and M. S. Geisel, 1970, Analysis of distributed lag models with applications to consumption function estimation, Econometrica 38, 865-888.

Zellner, A. and S. B. Park, 1979, Minimum expected loss (MELO) estimators for functions of parameters and structural coefficients of econometric models, Journal of the American Statistical Association 74, 185-193.

Figure 3.1

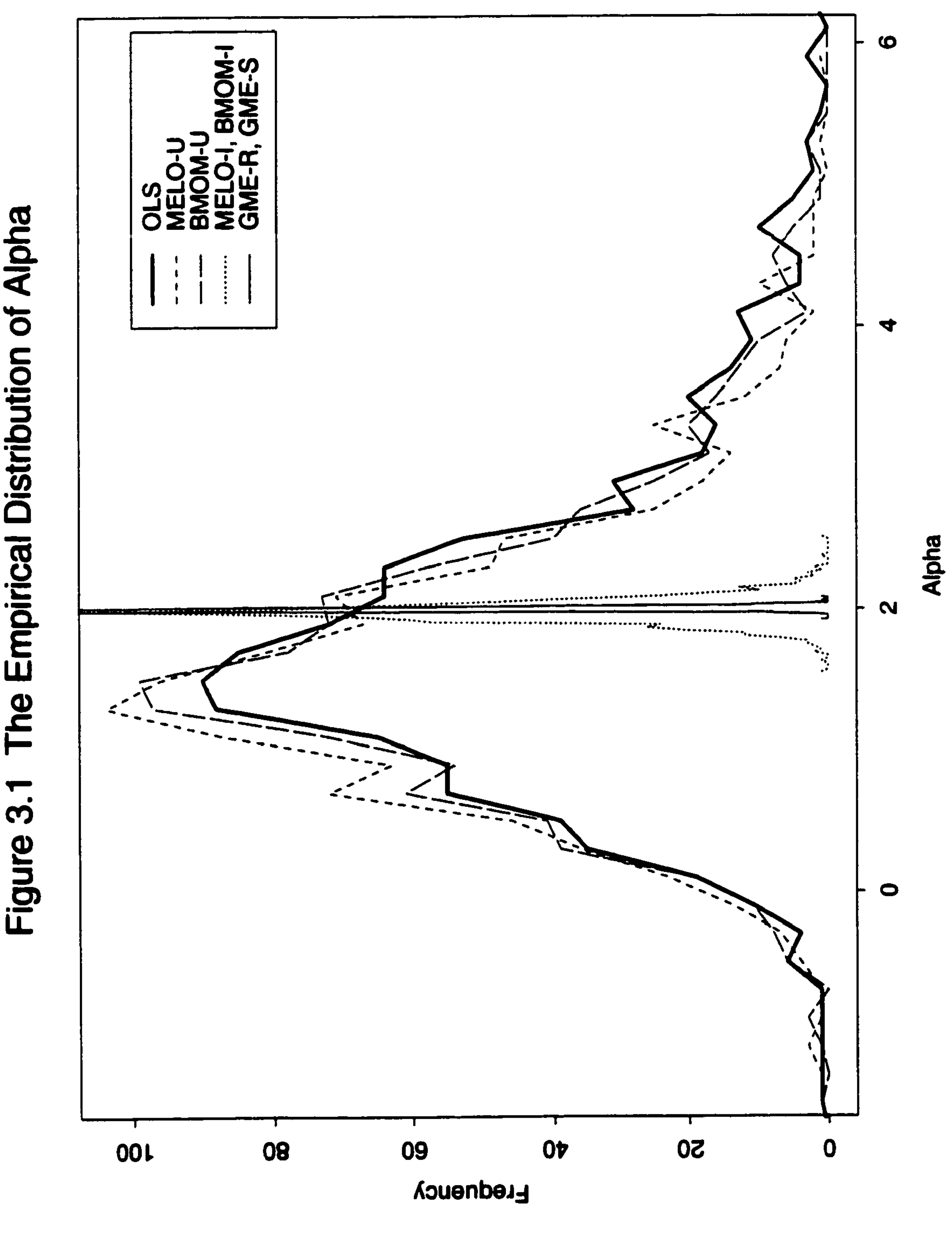
Figure 3.1 The Empirical Distribution of Alpha
OLS
MELO-U
BMOM-U
MELO-I, BMOM-I
GME-R, GME-S
Alpha
Frequency
0
2
4
6
0
20
40
60
80
100

Table 3-1 Mean Square Error of the Estimated *α* from Experimental Data,
$T = 25$, $\alpha = 2$, $\gamma = 1$, number of samples = 1,000

					ρ = 0.5			
σ_ε	σ_u	OLS	MELO-U	MELO-I	BMOM-U	BMOM-I	GME-R	GME-S
1	5	3.9299	1.1394	0.0081	2.0407	0.0136	0.00028	0.00016
3	3	0.1461	0.1271	0.0188	0.1353	0.0194	0.00007	0.00008
2	3	0.3077	0.2647	0.0147	0.2819	0.0165	0.00012	0.00009
3	5	0.4040	0.2928	0.0143	0.3294	0.0160	0.00019	0.00016
5	5	0.1653	0.1352	0.0224	0.1476	0.0243	0.00014	0.00014
					ρ = 0.9			
σ_ε	σ_u	OLS	MELO-U	MELO-I	BMOM-U	BMOM-I	GME-R	GME-S
1	5	1.7559	1.1284	0.0080	1.3468	0.0174	0.00033	0.00015
3	3	0.1542	0.1291	0.0186	0.1402	0.0195	0.00007	0.00007
2	3	0.3385	0.2469	0.0154	0.2808	0.0165	0.00012	0.00009
3	5	0.3897	0.2925	0.0156	0.3286	0.0178	0.00021	0.00018
5	5	0.1544	0.1306	0.0220	0.1414	0.0220	0.00015	0.00015

Table 3-2 Empirical Distributions of the Estimated Supply Response Coefficient α

$\alpha = 2$, $\sigma_\varepsilon = 1$, $\sigma_u = 5$, $\rho = 0.5$

		$\gamma = 0.5$						
T		*OLS*	*MELO-U*	*MELO-I*	*BMOM-U*	*BMOM-I*	*GME-R*	*GME-S*
25	Mean	1.38	1.06	1.97	1.23	1.98	1.99	2.00
	SD	3.34	1.01	0.12	1.93	0.15	0.03	0.01
	MSE	11.55	1.90	0.02	4.33	0.02	0.001	0.0002
	Min	-48.49	-2.43	0.91	-24.55	1.30	1.58	1.92
	Max	74.36	7.09	2.81	37.27	2.96	2.48	2.08
	Bias	1.37	1.16	0.09	1.25	0.11	0.02	0.01
50	Mean	1.33	1.24	1.93	1.28	1.95	1.98	2.00
	SD	0.82	0.75	0.11	0.78	0.14	0.04	0.01
	MSE	1.12	1.14	0.02	1.13	0.02	0.002	0.0002
	Min	-1.46	-1.33	1.31	-1.40	1.45	1.30	1.84
	Max	7.47	6.80	2.45	7.15	2.49	2.04	2.03
	Bias	0.88	0.90	0.10	0.89	0.11	0.03	0.01
100	Mean	1.37	1.32	1.88	1.34	1.90	1.97	2.00
	SD	0.55	0.53	0.10	0.54	0.14	0.04	0.02
	MSE	0.70	0.74	0.02	0.72	0.03	0.003	0.0003
	Min	-0.35	-0.36	1.55	-0.35	1.44	1.29	1.88
	Max	3.38	3.28	2.34	3.33	2.30	2.04	2.21
	Bias	0.71	0.74	0.13	0.72	0.14	0.04	0.01

Table 3-2 (Continued)

T		γ = 1.0						
		OLS	*MELO-U*	*MELO-I*	*BMOM-U*	*BMOM-I*	*GME-R*	*GME-S*
25	Mean	1.79	1.56	1.98	1.69	1.99	1.99	2.00
	SD	1.17	0.96	0.09	1.05	0.11	0.02	0.01
	MSE	1.41	1.12	0.01	1.21	0.01	0.0004	0.0002
	Min	-3.24	-1.24	1.32	-2.29	1.61	1.76	1.96
	Max	7.34	5.87	2.32	6.31	2.38	2.04	2.04
	Bias	0.91	0.86	0.07	0.87	0.09	0.01	0.01
50	Mean	1.93	1.83	1.97	1.88	1.98	2.00	2.00
	SD	0.76	0.70	0.08	0.73	0.11	0.02	0.01
	MSE	0.59	0.52	0.01	0.55	0.01	0.0003	0.0001
	Min	-0.44	-0.55	1.63	-0.49	1.58	1.89	1.94
	Max	5.43	5.14	2.42	5.29	2.37	2.04	2.03
	Bias	0.59	0.56	0.06	0.58	0.08	0.01	0.01
100	Mean	1.97	1.92	1.97	1.95	1.97	2.00	2.00
	SD	0.54	0.53	0.10	0.53	0.12	0.02	0.02
	MSE	0.29	0.28	0.01	0.29	0.02	0.001	0.0002
	Min	0.78	0.77	1.64	0.77	1.54	1.89	1.90
	Max	4.22	4.07	2.39	4.15	2.56	2.06	2.06
	Bias	0.43	0.43	0.08	0.43	0.10	0.02	0.01

Table 3-3 Empirical Distributions of Estimated Supply Response Coefficient α
$\alpha = 2$, $\sigma_\varepsilon = 1$, $\sigma_u = 5$, $\rho = 0.9$

T		$\gamma = 0.5$						
		OLS	*MELO-U*	*MELO-I*	*BMOM-U*	*BMOM-I*	*GME-R*	*GME-S*
25	Mean	1.46	1.49	1.95	1.48	1.97	1.99	2.00
	SD	5.28	0.89	0.11	2.83	0.15	0.03	0.01
	MSE	28.10	1.05	0.01	8.29	0.02	0.001	0.0002
	Min	-129.95	-2.14	1.45	-66.07	0.88	1.68	1.95
	Max	13.44	5.66	2.42	9.02	2.68	2.03	2.04
	Bias	1.10	0.80	0.09	0.95	0.11	0.02	0.01
50	Mean	1.77	1.71	1.92	1.75	1.94	1.99	2.00
	SD	0.57	0.54	0.12	0.56	0.13	0.02	0.01
	MSE	0.38	0.38	0.02	0.37	0.02	0.001	0.0002
	Min	-0.93	-0.79	1.47	-0.86	1.42	1.84	1.95
	Max	4.10	3.80	2.35	3.96	2.41	2.04	2.03
	Bias	0.47	0.47	0.11	0.47	0.11	0.02	0.01
100	Mean	1.84	1.82	1.89	1.83	1.91	1.99	2.00
	SD	0.33	0.33	0.13	0.33	0.14	0.03	0.02
	MSE	0.14	0.14	0.03	0.14	0.03	0.001	0.0003
	Min	0.74	0.73	1.60	0.74	1.44	1.88	1.92
	Max	3.62	3.38	2.35	3.50	2.42	2.06	2.04
	Bias	0.29	0.30	0.14	0.29	0.14	0.02	0.01

Table 3-3 (Continued)

T		$\gamma = 1.0$						
		OLS	*MELO-U*	*MELO-I*	*BMOM-U*	*BMOM-I*	*GME-R*	*GME-S*
25	Mean	1.87	1.75	1.97	1.81	1.99	2.00	2.00
	SD	0.88	0.77	0.09	0.82	0.11	0.02	0.01
	MSE	0.79	0.65	0.01	0.71	0.01	0.0003	0.0002
	Min	-1.07	-0.73	1.39	-0.90	1.63	1.87	1.95
	Max	6.80	5.22	2.55	6.07	2.57	2.04	2.04
	Bias	0.67	0.63	0.07	0.65	0.08	0.01	0.01
50	Mean	1.96	1.92	1.97	1.94	1.98	2.00	2.00
	SD	0.49	0.48	0.10	0.48	0.12	0.02	0.01
	MSE	0.24	0.23	0.01	0.24	0.01	0.0004	0.0002
	Min	0.06	0.04	1.63	0.06	1.58	1.89	1.94
	Max	3.81	3.61	2.31	3.71	2.38	2.04	2.03
	Bias	0.37	0.37	0.08	0.37	0.09	0.01	0.01
100	Mean	1.98	1.96	1.96	1.97	1.97	1.99	2.00
	SD	0.29	0.28	0.12	0.29	0.12	0.02	0.02
	MSE	0.08	0.08	0.02	0.08	0.02	0.0004	0.0003
	Min	0.98	0.96	1.64	0.97	1.68	1.92	1.92
	Max	3.27	3.25	2.32	3.26	2.37	2.05	2.04
	Bias	0.22	0.22	0.10	0.22	0.11	0.02	0.01

Table 3-4 Estimates of the Structural Coefficients

$\sigma_\varepsilon = 1$, $\sigma_u = 5$, $\alpha = 2$, $\rho = 0.5$, $\gamma = 0.5$, $\theta = 0.5$, $T = 25$

	α_0			α		
	Mean	*SD*	*MSE*	*Mean*	*SD*	*MSE*
True	0.25			2		
OLS	62.67	333.45	115,077	1.38	3.34	11.55
MELO-U	94.39	100.66	18,657	1.06	1.01	1.90
MELO-I	3.24	11.94	143.91	1.97	0.12	0.02
BMOM-U	77.62	192.93	4,290	1.23	1.93	4.33
BMOM-I	2.06	15.06	226.28	1.98	0.15	0.02
GME-R	1.56	3.04	10.98	1.99	0.03	0.00125
GME-S	0.14	0.53	0.29	2.00	0.01	0.00017

	γ			θ		
	Mean	*SD*	*MSE*	*Mean*	*SD*	*MSE*
True	0.5			0.5		
GME-S	0.39	0.12	0.0272	0.89	0.04	0.1548

Note: OLS, MELO-U, MELO-I, and GME-R do not provide estimates of γ and θ.

Table 3-5 Correlations between A and $\hat{A}$
$\alpha = 2, \sigma_\varepsilon = 1, \sigma_u = 5, \rho = 0.5, \gamma = 0.5, \theta = 0.5, T = 25$

	γ	ρ	*OLS, MELO-U, BMOM-U**	*MELO-I*	*BMOM-I*	*GME-R*	*GME-S*
25	0.5	0.5	0.578	0.614	0.655	0.546	0.528
		0.9	0.703	0.726	0.747	0.687	0.692
	1.0	0.5	0.623	0.633	0.665	0.608	0.572
		0.9	0.736	0.744	0.767	0.725	0.709
50	0.5	0.5	0.593	0.614	0.648	0.582	0.500
		0.9	0.766	0.770	0.785	0.764	0.736
	1.0	0.5	0.644	0.649	0.663	0.640	0.558
		0.9	0.796	0.796	0.808	0.794	0.764
100	0.5	0.5	0.608	0.620	0.649	0.603	0.459
		0.9	0.814	0.813	0.822	0.813	0.775
	1.0	0.5	0.656	0.654	0.665	0.654	0.523
		0.9	0.835	0.834	0.839	0.834	0.795

* The OLS and MELO-U correlations are very close but not identical.

Table 3-6 GME-S Correlations between P^e and P^e and between A^* and A^*
$\sigma_\varepsilon = 1$, $\sigma_u = 5$, $\rho = 0.5$, $\gamma = 0.5$, $\theta = 0.5$, $T = 25$

T	γ	ρ	P^e and P^e	A^* and A^*
25	0.5	0.5	0.748	0.347
		0.9	0.834	0.472
	1.0	0.5	0.737	0.388
		0.9	0.838	0.485
50	0.5	0.5	0.858	0.337
		0.9	0.938	0.528
	1.0	0.5	0.805	0.389
		0.9	0.903	0.549
100	0.5	0.5	0.900	0.320
		0.9	0.966	0.578
	1.0	0.5	0.844	0.378
		0.9	0.941	0.595

Table 3-7 MSE When the Distributions of σ_u and σ_ε are Not Normal
$T = 25$, $\alpha = 2$, $\gamma = 0.5$, $\rho = 0.5$, number of samples = 100

	χ^2_5	χ^2_7	t_5	t_7
OLS	0.120	0.162	1.96	0.096
MELO-U	0.107	0.113	0.101	0.083
MELO-I	0.026	0.041	0.030	0.027
BMOM-U	0.109	0.130	0.643	0.083
BMOM-I	0.029	0.041	0.032	0.032
GME-R	0.00005	0.00013	0.00005	0.00006
GME-S	0.00007	0.00006	0.00004	0.00003

Table 3-8 MSE for Various Supports
$\sigma_\varepsilon = 1, \sigma_u = 5, \rho = 0.5, \gamma = 0.5, \theta = 0.5, T = 25$

Changes in the support for only the reduced-form coefficients

	MELO-I	*BMOM-I*	*GME-R*
original	0.009	0.020	0.007
5× larger	0.453	0.492	0.142
10× larger	0.911	0.959	0.469
20× larger	1.075	1.195	0.898
40× larger	1.041	0.980	1.160

Change of the support for only the errors

	MELO-I	*BMOM-I*	*GME-R*	*GME-S*
σ	0.0094	0.0202	0.00071	0.00017
2σ	0.0169	0.0546	0.00011	0.00011
3σ	0.0260	0.0377	0.00010	0.00011
4σ	0.0269	0.0650	0.00010	0.00011
5σ	0.0430	0.1000	0.00010	0.00011
10σ	0.0353	0.4831	0.00009	0.00011
20σ	0.0424	1.4460	0.00009	0.00011
100σ	0.0490	1.7309	0.0099	0.0116
200σ	0.0499	1.6573	0.115	0.132

Note: MSE of OLS is 1.283, and the MSE of MELO-U is 1.098.

Chapter 4 MELO Estimates of Willingness-to-Pay in Dichotomous Choice Contingent Valuation

4.1 Introduction

This paper introduces a new approach that provides superior willing-to-pay (WTP) estimates in discrete or dichotomous choice (DC) contingent valuation (CV). The DC model has emerged as a new tool in CV studies to evaluate non-market resources or public goods. Recently economists have identified a large number of biases (systematic over- or under-estimates of true WTP) in this model[25]. While studying the bias and variability problems, many papers focus upon such factors as distributional assumptions, choice of bid vehicles, starting-point bid values, hypothetical market bias, and strategic bias (which occurs when respondents believe that their answers may influence environmental policies). Yet one fundamental factor has been neglected. Because the ML WTP estimators involve ratios of estimated coefficients, they often do not possess finite sample moments and are therefore subject to large errors.

The objective of the paper is to find alternative ways to improve WTP estimates given correct model specification and no hypothetical question bias. This paper adopts a Bayesian approach using Zellner (1978)'s minimum expected loss estimator (MELO), which has been employed in a wide range of estimation problems including structural coefficient system equations and reciprocals and ratios of regression coefficients. Diebold and Lamb (1997) and

[25] Some papers include Boyce et al (1985), Mitchell and Carson (1989), Bateman and Willis (1995), and Kanninen (1995).

Perloff and Shen (1998) used MELO to improve the estimates of the Nerlove agricultural supply model. Bewley and Fiebig (1990) applied it to estimate long-run responses in dynamic models. To better approximate the expected consumer surplus of recreational demand, which involves an inverse of a parameter, Bockstael and Strand (1987, pp. 17, equation [14]) and Hanemann (1982) suggest a formula that is similar to the MELO approach.

Compared with the maximum likelihood approach that is extensively used in CV studies, the MELO estimator has finite moments and provides less biased and more precise WTP estimates. It is easy to implement and the efficiency gain is more significant when the bid questions are ill-designed. When economic-theory-based restrictions are utilized, the MELO framework can further improve the WTP estimates.

The rest of the paper is organized as follows. Section 2 outlines the single-bounded DC framework and the ML approach in estimating WTP. Section 3 discusses problems with the ML approach and introduces the MELO estimator. Section 4 presents Monte Carlo experiments that illustrate the performance of MELO estimates in simulated data sets. In Section 5, MELO is applied in two CV studies. Section 6 discusses the procedure to handle non-diffuse information in the MELO framework. And finally Section 7 presents conclusions.

4.2 WTP Model

The single-bounded (SB) DC WTP model, originally proposed by Bishop and Heberlein (1979) and refined by Hanemann (1984), has been widely used in contingent valuation studies. Let us consider the case of assessing the recreational value of environmental goods such as visits to a national park, a fishing site, or a state beach. In the DC CV framework, participants in the survey are given a scenario of policy that provides public goods or otherwise affects the natural environment. Each participant of the CV study is asked a randomly selected bid question, say, "Are you willing to pay ten dollars to enter a national park for one day?" The respondent can answer "Yes" or "No" to the bid question. The bid values B are randomly selected from a possible range, depending upon the "prior" knowledge of WTP distributions. The probability of saying "Yes" and "No" can be represented respectively by,

$$\pi^Y = 1 - G(B) \qquad (1)$$
$$\pi^N = G(B).$$

$G(.)$ is the cumulative density function for answering "No", which in practice often takes either a logistic or probit form. Given the logistic form, $G(.)$ is

$$G(B,\alpha,\beta) = \frac{1}{1+e^{\alpha-\beta B}} \qquad (2)$$

where α and β are parameters to be estimated.[26] Figure 4.1 illustrates the model where the x-axis represents the bid question and the y-axis represents the probability of "Yes" answers. The curve represents the underlying relationship $G(.)$ and the dots represent sample frequency of "Yes" at each bid value. Suppose N individuals answer I_i^Y ($I_i^Y = 1$ if "Yes" and 0 if "No", $i = 1, \ldots, N$) to a randomly chosen positive bid value B_i. Given the sample, the maximum likelihood (ML) method estimates of α and β are obtained by maximizing the joint log-likelihood

$$\ln L(\alpha,\beta) = \sum_{i=1}^{N} (I_i^Y \ln(\pi_i^Y(B_i)) + (1 - I_i^Y) \ln(1 - \pi_i^Y(B_i))) \qquad (3)$$

[26] The logistic form is here used as a bench model but the major results hold for the probit form as well.

Under the assumption of correct specification and independent observations, the ML produces both consistent and asymptotically efficient estimates $\hat{\alpha}$ and $\hat{\beta}$ [27]. The CV literature defines the median WTP as

$$W\hat{T}P = \frac{\hat{\alpha}}{\hat{\beta}} \quad (4)$$

The corresponding estimate of the truncated (restricted) mean WTP is given by Hanemann (1989) as

$$W\hat{T}P^{+} = \frac{ln(1 + \hat{e}^{\alpha})}{\hat{\beta}} \quad (5)$$

4.3 MELO Approach

The ML approach has been recognized as the most widely used statistical technique in empirical CV studies for finding WTP point estimates in (4) and (5), which are then converted to population total value figures. However, there are problems in (4) and (5) that need to be addressed. This section focuses upon the media WTP (4) because of its simplicity in form, but the discussions can easily extended to the truncated mean WTP (5) as well.

Since both $\hat{\alpha}$ and $\hat{\beta}$ are estimators or random variables, their ratio possesses some undesirable properties. It is well known that ratios of random variables have Cauchy tails and hence no finite moments and risk relative to quadratic and other loss functions (Bergstrom [1962], Zellner [1978, 1986], Zellner and Park [1979], and Zaman [1981]). Moreover, they may generally display multi-modal distributions (Lehmann and Shaffer [1988]). As shown in the

[27] ML yields biased estimates for $\hat{\alpha}$ and $\hat{\beta}$ in small samples, but the bias vanishes as sample size goes to infinity.

following experimental results, the non-existence of first and second moments and multi-modality would contribute to variability in the WTP estimates. To solve these problems, Zellner (1978) suggests a minimum expected loss (MELO) procedure that minimizes the posterior expectation of a generalized quadratic loss function, that is, $\min_{\hat{\theta}} E\beta^2(\frac{\alpha}{\beta} - \hat{\theta})^2$, where *E()* is the posterior expectation and $\hat{\theta}$ is some estimate of $\frac{\alpha}{\beta}$. Given any given prior likelihood and data, the MELO estimate of WTP is given by

$$\hat{WTP}_{MELO} = \frac{E(\alpha)}{E(\beta)} S \tag{6}$$

$$S = \frac{1 + Cov(a, \beta) / [E(\alpha) E(\beta)]}{1 + Var(\beta) / E^2(\beta)}$$

where *E()*, *Var()*, and *Cov()* are posterior mean, variance and covariance respectively, and *S* is the shrinkage factor. MELO has finite first and second moments as well as finite risk with respect to quadratic loss in both small and large samples. In addition, it is consistent, asymptotically efficient, and asymptotically normal (Zellner [1978] and Zellner and Park [1979]). Thus the MELO and ML estimators have very different finite sample property; however as sample size get large, their large sample distribution becomes identical.

Suppose we do not have any prior knowledge regarding the likelihood of the parameters. Given the diffuse (uninformative) prior, the posterior (α, β) is approximated normally distributed with mean $(\hat{\alpha}, \hat{\beta})$ and covariance matrix $Cov(\hat{\alpha}, \hat{\beta})$ where $(\hat{\alpha}, \hat{\beta})$ are ML estimates and $Cov(\hat{\alpha}, \hat{\beta})$ are ML variance-covariance matrix. This approximate is a special case of Jeffrey's

(1967) general result on the large sample normality of posterior pdfs (Zellner and Rossi [1984] and Clogg, et al [1991). Therefore equation (6) can be re-written as[28]

$$\hat{WTP}_{MELO} = \frac{\hat{\alpha}}{\hat{\beta}} S \tag{7}$$

$$S = \frac{1 + Cov(\hat{\alpha}, \hat{\beta}) / \hat{\alpha}\hat{\beta}}{1 + Var(\hat{\beta}) / \hat{\beta}^2}.$$

It can be seen that MELO estimator (7) is the product of $\frac{\hat{\alpha}}{\hat{\beta}}$ and S, the shrinkage factor that is not equal to one in small samples. S is greater (less) than one when $\frac{Cov(\hat{\alpha}, \hat{\beta})}{\hat{\alpha}\hat{\beta}} > (<) \frac{Var(\hat{\beta})}{\hat{\beta}^2}$, or in other words, $\frac{Cov(\hat{\alpha}, \hat{\beta})}{Var(\hat{\beta})} > (<) \frac{\hat{\alpha}}{\hat{\beta}}$. When either $\hat{\beta}$ is smaller or is subject to larger error, S is more likely to be less than one. This is because the errors in denominator have larger impacts upon a ratio when the denominator is closer to zero. Therefore MELO estimates are more stable than the ML estimates especially when the estimated parameters are subject to larger errors or the denominator is more likely to be zero. In large samples, S becomes one as all elements of the variance-covariance matrix goes to zero and hence the MELO estimates are the same as the corresponding ML ones.

Essentially the MELO approach uses the variance-covariance matrix of the estimated parameters to improve point estimates. MELO may also provide less biased estimates than the ML approach does, as shown in the experiments in Section 4. The reason is that equation (7) can be viewed as a second-order Taylor approximation of the ratio under certain conditions (see Appendix A). These conditions are that $\frac{Var(\hat{\beta})}{E(\hat{\beta})^2}$ and $\frac{Cov(\hat{\alpha}, \hat{\beta})}{E(\hat{\alpha})E(\hat{\beta})}$ are small, or in other words, that

[28] It should be noted that the normal approximation might not be close to the finite estimator property in small samples. Numerical analysis such as Newton-raphon routine can be used to compute exact finite sample posterior distribution of coefficients.

both $\hat{\alpha}$ and $\hat{\beta}$ are statistically significant. However, in practice the MELO estimator does not guarantee reduction in WTP bias because both $\hat{\alpha}$ and $\hat{\beta}$ are biased estimators in finite samples (Griffiths [1987]), so that the WTP estimator will still be biased. For instance Diebold and Lamb (1997) show a trade-off between lower MSE and a more biased estimate of agricultural supply elasticity under MELO.

Analogous to that of median WTP, the MELO estimate of the truncated mean WTP given an un-informative (diffuse) prior is

$$W\hat{T}P^{+}_{MELO} = \frac{\hat{\alpha} + \frac{e^{-\hat{\alpha}}}{2(1+e^{-\hat{\alpha}})^2} Var(\hat{\alpha})}{\hat{\beta}} S^{+} \tag{8}$$

$$S^{+} = \frac{1 + \left(\left(\frac{e^{\hat{\alpha}}}{1+e^{\hat{\alpha}}} \right)^2 Var(\hat{\alpha}) Cov(\hat{a}, \hat{\beta}) \right) \Big/ \left(\left(\hat{\alpha} + \frac{e^{-\hat{\alpha}}}{2(1+e^{-\hat{\alpha}})^2} Var(\hat{\alpha}) \right) \hat{\beta} \right)}{1 + Var(\hat{\beta}) / \hat{\beta}^2}.$$

See Appendix B for detailed derivations. The derivation of equation (8) closely follow that of equation (7). The delta method (propagation of error) is adopted in the derivation to approximate the posterior mean $E(ln(1+e^{\alpha}))$ and posterior variance $Var(ln(1+e^{\alpha}))$.

The MELO estimators of WTP (7) and (8) are relatively simple in form and easy to compute. Given a diffuse prior, we first obtain ML estimates and then compute the shrinkage factor S using the variance-covariance matrix of the ML estimates. Section 6 will discuss obtaining MELO estimates given a non-diffuse prior in which the parameter signs are known from the random utility theory of contingent valuation.

4.4 Monte Carlo Experiments

This section presents Monte Carlo (MC) analysis to illustrate the performance of both MELO and ML estimators in the data generated under three experimental designs with various parameters and bid sets. All these experiments assume a correct model specification, a logistic cdf function, and independent observations. They all demonstrate that MELO provides better WTP estimates than the ML approach in terms of MSE and standard errors and that the efficiency gain could be larger when the bid questions inadequately cover the WTP distribution. The simulations and estimations were implemented using Shazam and Matlab programs.

4.4.1 The First MC Experiment

In this experiment, data are drawn randomly from the same underlying logistic cdf specification where $\alpha = 2$, and $\beta = 0.2$, median $WTP = \frac{\alpha}{\beta} = 10$, and $WTP^{+} = \frac{ln(1+e^{\alpha})}{\beta} = 10.635$. Bid values B are generated from a uniform distribution, $B \sim Unif(0,20)$, and the probability π^{Y} under each B is computed using equations (2) and (1). Given each B and its probability π^{Y}, one Bernoulli number I^{Y} is drawn. 500 samples of size $N = 30$ are generated and estimated using both the ML and MELO approaches. Columns two and three of Table 4-1 compare the empirical distributions of both estimates using the above-generated samples. The empirical MSE is 4.69 under MELO in contrast to 16.92 under the ML approach.

The above experiment is well balanced since the bid values cover most (12% to 88%) of the WTP distribution and are symmetric around the median WTP. But empirical WTP knowledge is usually limited in practice. One of the sources of WTP bias in CV is the choice of starting bid points, which has been extensively discussed in many other papers. These papers include Boyle, Bishop, and Welsh (1985), Kristrom (1990), Cooper and Loomis (1992), McFadden and Leonard (1992), Kanninen and Kristrom (1993), and Cooper and Loomis (1993). Most papers find that the mean and median WTP estimates are sensitive to survey design. Adding or omitting bid amounts in either the upper or lower end of the distribution could significantly affect the estimates.

Table 4-1 gives the results under two more experiments: the second and third bid designs. The second design $B \sim Unif(0,10)$ allows only bids between 12% and 50% of the WTP distribution, and the third design $B \sim Unif(10,20)$ covers only the upper portion of the distribution. Not surprisingly, the ML procedure does a worse job in both experiments, sometimes yielding large positive or negative WTP values and generally understating the true median WTP. On the other hand, the MELO estimates cluster around the true WTP, substantially reducing standard errors, MSE, and bias. The third design shows the greatest improvement. The MSE under the MELO approach is 9.79 compared to 2864.8 under the ML one.

The estimates in Table 4-1 are also displayed in Figures 4.2 through 4.3. The MELO estimates have much smaller variances compared to the ML ones particularly in Figures 4.2 and 4.3. The instability of the traditional estimates is largely due to thick tails or the presence of "outliers", but the MELO estimator eliminates these tails. Table 4-2 displays the empirical distributions of WTP^{+} estimates (equation [3]) where the MC experiments are the same as in Table 4-1 ($\alpha = 2$, $\beta = 0.2$, $WTP^{+} = 10.635$, and $N = 30$). The same conclusion can be reached from Table 4-2 as from Table 4-1.

4.4.2 The Second and Third Monte Carlo Experiments

To demonstrate that the improvements of the MELO estimates are not limited to specific parameters, two additional MC experiments, the second and the third ones, are selected from the current environmental economics literature. The simulated data are then estimated by both the

ML and MELO procedures. Both experiments have results that are similar to those found in Section 4.1.

The second MC experiment is adopted from Table 4-1 of Kanninen (1995) that studies the bias and variance of WTP using five different bid designs (A through E). The parameters (α, β) are set at (1.8, 0.009). The base set bid (A) has twenty bids that range from $3 to $700 and cover 14.5% and 98.9% of the distribution. For a more detailed discussion of the experimental design, please see Kanninen (1995). The MC results are reported in Table 4-3. The MELO estimator always reduces MSE and the reduction is particularly significant in case E where bids cover the upper tail of WTP only -- MSE is 3932.9 under the MELO compared to 6190.7 under the ML approach. Table 4-3 does not indicate a significant reduction in bias.

The third MC experiment is from Cooper and Loomis (1993) (their Table 4-3) who analyze the effects upon WTP estimates of omitting middle or outer bids. This experiment is generalized here by introducing a random white noise term ε with mean zero and standard deviation γ, which creates more "noise" in the data. The rationale behind having an error term is that many other factors such as taste, experience, and culture affect WTP decisions but are not observed during the survey. Under the assumption of logistic form, the probability of "Yes" answer can be written as

$$\pi^{Y} = 1 - G(B, \alpha, \beta, \theta, \gamma) = 1 - \frac{1}{1 + e^{\alpha - \beta B + \theta D + \varepsilon}} \tag{9}$$

$$\varepsilon \sim N(0, \gamma)$$

where ε is assumed to be *i.i.d.* A higher γ implies greater uncertainty in the simulated data sets, and under γ = 0, the experiment becomes that of Cooper and Loomis. The parameters are $\alpha = 1.4136$, β = 0.008561, θ = 0.00362, and γ = 0 (no noise), 1, and 2. Cooper and Loomis created D as a "Deer-seen" variable, which is randomly drawn from a normal distribution with mean 95.78 and standard deviation 119.226. Given these parameters, the truncated mean WTP is 206.7.

The experimental analyses are presented in Table 4-4, which yield the same qualitative results as those in Cooper and Loomis (1993). As expected, a higher γ leads to less stable WTP estimates. But MSE and standard errors are always smaller in MELO except for one case (Design B, $\gamma = 0$). The gain of MELO over the conventional approach is particularly large in Design C when outer bid values are excluded in the simulated data sets.

4.5 Applications of MELO

The MELO procedure is used to provide basis of comparison with the results using the ML approach in two empirical contingent valuation studies that are adopted from the current literature. This first study is Hanemann, Loomis, and Kanninen (1991) and the second study is Riddle and Loomis (1998). Both applications reveal that the MELO estimates of WTP in SB model are more likely to be precise than the ML ones.

4.5.1 Hanemann and et al (1991)'s study

Hanemann and et al evaluated the fish and wildlife resources in San Joaquin, California. Conducted for the Interagency San Joaquin Valley Drainage Program, the survey focuses on WTP for protecting wildlife and wetlands habitant. The CV technique was designed to evaluate the economic value of fish and wildlife in the Valley. Both mail and telephone methods were used to survey randomly selected households. The survey was carried out in 1989, and more than 1,000 people participated in it. A more detailed description can be found in Jones and Stoke Associates (1990).

The data are fit into both the SB as well as double-bounded (DB) discrete choice models. The DB model is similar to the SB one except that each participant is presented with two bids where the level of the second bid is contingent upon the response to the first one. Those respondents who say "No" in the first bid are presented with a lower bid and those respondents who say "Yes" to the first bid are presented with a higher bid. Hanemann et al have showed that the DB model is asymptotically more efficient than the SB model. Their study reveals that the WTP is over-estimated in the SB model.

Table 4-5 lists the WTP estimates in the two models, by both the ML and MELO approaches. Under the ML approach, the SB model estimate ($250) exceeds the DB model estimate ($151) by a factor of more than 60%. The MELO estimates, in the second row of Table 4-5, reduce their difference: the MELO estimates are $257 and $217.60 in the SB and DB models repetitively. There is small difference between the two approaches in the DB model because the variance-covariance of the estimated parameters is small. Similar results are obtained for the truncated WTP estimates.

4.5.2 Riddle and Loomis (1998)'s Study

Similar findings exist during the MELO's application in Riddle and Loomis's study, which evaluates WTP for three programs (California/Oregon Program combined, Oregon Program only, and California Program Only) designed to reduce fire hazard in California and Oregon's spotted Owl habitat. The data were obtained from a CV survey conducted in 1995, and the total number of respondents who completed the survey is 353. A complete discussion of the CV design and the empirical results can be found in Riddle and Loomis.

The data are fit into the SB as well as DB models and the estimated median WTPs for each of the three programs are reported in Table 4-6. Under the ML approach, WTP estimates in the SB model generally exceed these in the DB model (except for the Oregon Program). The estimates of the two models are more similar under the MELO approach. Take the California/Oregon combined program for example, the median WTP estimates are $32.16 in the

SB model and \$25.24 in the DB model. The MELO approach lowers the SB WTP estimate to \$26.71, which is closer the corresponding DB WTP estimate of \$25.05.[29]

4.6 Use of Prior Information in MELO

In addition to the efficiency gain under a diffuse prior, the MELO approach allows us to use prior restrictions and hence further improve WTP estimates. A common belief in CV economics is that environmental goods contain positive values and people are willing to pay a certain amount in exchange for better environmental quality. This implies restrictions in the dichotomous choice models: $\alpha > 0$, $\beta > 0$, WTP > 0, and $WTP^{+} > 0$.

MELO can use this prior information to gain more efficiency in the value estimates (called MELO-I). The above information implies the restriction that the coefficients lie within certain ranges (here the restrictions are $\alpha > 0$ and $\beta > 0$). To impose the restriction within a Bayesian framework, Geweke's (1986) technique is utilized. The first step of the technique is to generate a large number (J) of samples of α and β using their known posterior distributions. In the second step, those observations for which either α or β fails to satisfy the restrictions are rejected. The remaining observations α and β constitute a truncated posterior distribution α' and β' that are consistent with the prior restrictions. Finally, the MELO-I estimate of median WTP is

[29] Riddle and Loomis presented a technique that can jointly estimate WTP for multiple scenarios within the survey, which is more efficient than estimating equation separately when error correlation exist. Similarly, MELO procedure can be used to jointly estimate system equations.

$$\hat{WTP}_{MELO-I} = \frac{E(\alpha^I)}{E(\beta^I)} S^I \tag{10}$$

$$S^I = \frac{1 + Cov(\alpha^I, \beta^I) / E(\alpha^I)E(\beta^I)}{1 + Var(\beta^I) / E(\beta^I)^2}$$

where *E()*, *Cov()*, and *Var()* are posterior mean and variance-covariance respectively and can be evaluated numerically.

Table 4-8 compares the median WTP estimates under MELO with MELO-I in which restrictions $\alpha > 0$ and $\beta > 0$ are imposed. Given reasonable prior information, MELO-I improves WTP estimates. The improvement indicated in Table 4-7 is particularly significant when bid questions are ill-balanced. Although little gain in MSE is achieved in the first bid design, in the second and third design (Column 4-5 and 6-7) MELO-I drastically cuts the MSEs by more than half compared with those under MELO. Information can greatly increase the reliability of estimates provided that the information is reasonably accurate (Zellner [1963]).

4.7 Conclusions

While contingent valuation is becoming an increasingly useful tool in environmental, resource, and other studies, many papers have shown that there are bias and precision problems in the WTP estimates. Therefore it is important to examine this issue in more detail. This paper finds that some of the problems are attributed to the non-existence of moments under the ML approach and suggests an alternative estimator -- minimum expected loss (MELO). The MELO procedure is more efficient for a given sample size, and the efficiency gain is larger when the bid values are ill-posed. This approach can be applied to the DC, travel cost, and other models where key estimates involve ratios or inverses of regression parameters. The MELO approach is likely to be of great use in determining the economic values of environmental and public goods for which it is difficult to obtain market data.

References

Bateman, I. And K. Willis (eds), *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method*, Oxford University Press, Oxford, 1995.

Bergstrom, A.R., "The Exact Sampling Distributions of Least Squres and Maximum Likelihood Estimators of the Marginal Propensity to Consume," *Econometrica* Vol. 30 (1962), pp. 480-490.

Bewley, Ronald, and Denzil Fiebig, "Why are Long-Run Parameter Estimates Do Disparate?" *Review of Economics and Statistics*, Vol. 72, No.2, (May 1990), pp. 345-349.

Bishop, Richard C. and Thomas A. Heberlein, "Measuring Values of Extra-market Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics* Vol. 5, No. 5, (1979), pp. 926-930.

Bockstael, Nancy, and Ivar E. Strand, "The Effect of Common Sources of Regression on Benefit Estimates." *Land Economics*, Vol. 63, No. 1, (February 1987), pp. 11-20.

Boyle, Kevin, Richard Bishop, and Michael Welsh, "Starting Point Bias in Contigent Valuation Bidding Games," *Land Economics*, Vol. 612 (1985), pp. 188-194.

Clogg, C., D. Rubin, N. Schenker, B. Schultz, and L. Weidman, "Multiple Imputationof Industry and Occupation Codes in Census Public-use Samples Using Bayesian Logistic Regression," *Journal of American Statistical Association*, Vol. 86, No. 413, (March 1991), pp. 68-78.

Cooper, Joseph, and John Loomis, "Sensitivity of Willing-to-Pay Estimates to Bid Design in Dichotomous Choice Valuation Models: Reply," *Land Economics*, Vol. 69, No. 2, (May 1993), pp. 203-208.

Diebold, F. X. and R. L. Lamb, "Why are Estimates of Agricultural Supply Response So Variable?" *Journal of Econometrics*, Vol. 76, (1997), pp. 357-373.

Geweke, J., "Exact Inference in the Inequality Constrained Normal Linear Regression Model," *Journal of Applied Econometrics*, Vol. 1, (1986), pp. 127-141.

Griffiths, W. E., R. C. Hill, and P. J. Poper, "Small Sample Properties of Probit Model Estimators," *Journal of American Statistics Association*, Vol. 82, (1987), pp. 929-937.

Hanemann, Michael, "Some Further Results on Exact Consumer Surplus." Working paper, University of California, Berkeley, 1982.

Hanemann, Michael, "Welfare Evaluations in Contingent Valuation Experiments withDiscrete Response", *American Journal of Agricultural Economics*, Vol. 66, (1984), pp. 332-341.

Hanemann, Michael, "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses Data: Reply," *American Journal of Agricultural Economics*, Vol. 71 (1989), pp. 1057-1061.

Hanemann, Michael, John Loomis, and Barbara Kanninen, "Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation" *American Journal of Agricultural Economics*, Vol. 73 (November 1991), pp. 1255-1263.

Jeffreys, H., *Theory of Probability*, 3rd rev. ed., Oxford University Press, 1967.

Jones and Stokes Associates, *Environmental Benefits Study of San Joaquin Valley's Fish and Wildlife Resources*. JSA 87-150, final report, prepared J. B. Loomis, W.M. Hanemann, and T.C. Wegge. Sacramento, CA 1990.

Kanninen, Barbara, "Bias in Discrete Response Contingent Valuation, " *Journal of Environmental Economics and Management*, Vol. 28, (1995), pp. 114-125.

Kristrom, B., "A Non-parametric Approach to the Estimation of Welfare measures in Discrete Response Valuation Studies," *Land Economics*, Vol. 66, (1990), pp. 135-139.

Lehmann, E. L. and J. Popper Shaffer, "Inverted distributions," *American Statistician* Vol. 42, (1988) pp. 191-194.

Mitchell, R. and R. Carson, *Using Surveys to Value Public Goods, the Contingent Valuation Method*, Resources for the Future, Washington, D.C., 1989.

McFadden, Daniel, and Gregory Leonard, "Issues in the Contingent Valuation of Environmental Goods: Methodologies for Data Collection and Analysis," in Hausman, J.A., ed. *Contingent Valuation: A Critical Assessment*, North-Holland, New York, pp. 165-215.

Perloff, Jeffrey and Zhihua Shen, "Better Estimates of Agricultural Supply Response", Working Paper, University of California, Berkeley, 1998

Park, Timothy, John Loomis, and Michael Creel, "Confidence Intervals for Evaluating Benefit Estimates from Dichotomous Choice Contingent Valuation Surveys," *Land Economics*, Vol. 67, No. 1, (February 1991), pp. 64-73.

Riddle, Mary, and John Loomis, "Joint Estimation of Multiple CVM Scenarios under a Double Bounded Questioning Format," *Environmental and Resource Economics*, Vol. 12, No. 1, (July 1998), pp. 77-98.

Zellner, A., "Decision Rules for Economic Forecasting," *Econometrica* Vol. 31, (1963), pp. 111-130.

Zellner, A., "Estimator of Functions of Population Means and Regression Coefficients Including Structural Coefficients: A Minimum Expected Loss Approach," *Journal of Econometrics* Vol. 8, (1978), pp. 127-158.

Zellner, A. and S. B. Park, "Minimum Expected Loss (MELO) Estimators for Functions of Parameters and Structural Coefficients of Econometric Models," *Journal of the American Statistical Association* Vol. 74, (1979), pp. 185-193.

Zaman, A. "Estimators without Moments", *Journal of Econometrics* Vol. 15, (1981), pp. 289-298.

Zellner, A., and Peter E. Rossi, "Bayesian Analysis of Dichotomous Quantal Response Models," *Journal of Econometrics*, Vol. 25, (1984), pp. 365-393.

Zellner, A., "Further Results on Bayesian Minimum Expected Loss Estimates and Posterior Distributions for Structural Coefficients," in D. J. Slottje, ed., *Advances in Econometrics* (JAI Press, Greenwich, CT), 1986, pp. 171-182.

4.8 Appendix

4.8.1 The MELO and Second-order Taylor Approximation

To gain further insight on $\frac{\hat{\alpha}}{\hat{\beta}}$, it is decomposed into a second-order Taylor approximation around $E(\hat{\alpha})$ and $E(\hat{\beta})$,

$$\frac{\hat{\alpha}}{\hat{\beta}} \approx \frac{E(\hat{\alpha})}{E(\hat{\beta})} + \frac{1}{E(\hat{\beta})}(\hat{\alpha} - E(\hat{\alpha})) - \frac{E(\hat{\alpha})}{E(\hat{\beta})^2}(\hat{\beta} - E(\hat{\beta})) + \frac{E(\hat{\alpha})}{E(\hat{\beta})^3}(\hat{\beta} - E(\hat{\beta}))^2 - \frac{1}{E(\hat{\beta})^2}(\hat{\alpha} - E(\hat{\alpha}))(\hat{\beta} - E(\hat{\beta})) \tag{A.1}$$

Take expectation on both sides of equation (A.1) and simplify it using $E(\hat{\alpha} - E(\hat{\alpha})) = 0$, $E(\hat{\beta} - E(\hat{\beta})) = 0$, and the definitions of variance-covariance

$$E(\frac{\hat{\alpha}}{\hat{\beta}}) \approx \frac{E(\hat{\alpha})}{E(\hat{\beta})} + \frac{1}{E(\hat{\beta})}E(\hat{\alpha} - E(\hat{\alpha})) - \frac{E(\hat{\alpha})}{E(\hat{\beta})^2}E(\hat{\beta} - E(\hat{\beta})) \tag{A.2}$$

$$\frac{E(\hat{\alpha})}{E(\hat{\beta})^3}E(\hat{\beta} - E(\hat{\beta}))^2 - \frac{1}{E(\hat{\beta})^2}E(\hat{\alpha} - E(\hat{\alpha}))(\hat{\beta} - E(\hat{\beta})) = \frac{E(\hat{\alpha})}{E(\hat{\beta})} + \frac{E(\hat{\alpha})}{E(\hat{\beta})^3}Var(\hat{\beta}) - \frac{1}{E(\hat{\beta})^2}Cov(\hat{\alpha},\hat{\beta})$$

$$= \frac{E(\hat{\alpha})}{E(\hat{\beta})}\left[1 + \frac{Var(\hat{\beta})}{E(\hat{\beta})^2} - \frac{Cov(\hat{\alpha},\hat{\beta})}{E(\hat{\alpha})E(\hat{\beta})}\right]$$

Further simplifying the second part of the right side of equation (A.2) by applying the formula $1 + x - y \approx \frac{1+x}{1+y}$ (when both x and y are small, $x = \frac{Var(\hat{\beta})}{E(\hat{\beta})}$, and $y = \frac{Cov(\hat{\alpha},\hat{\beta})}{E(\hat{\alpha})E(\hat{\beta})}$) yields

$$E\left(\frac{\hat{\alpha}}{\hat{\beta}}\right) \approx \frac{E(\hat{\alpha})}{E(\hat{\beta})}\left[\frac{1+\frac{Var(\hat{\beta})}{E(\hat{\beta})^2}}{1+\frac{Cov(\hat{\alpha},\hat{\beta})}{E(\hat{\alpha})E(\hat{\beta})}}\right] \tag{A.3}$$

Small $\frac{Var(\hat{\beta})}{E(\hat{\beta})}$ and $\frac{Cov(\hat{\alpha},\hat{\beta})}{E(\hat{\alpha})E(\hat{\beta})}$ imply that both $\hat{\alpha}$ and $\hat{\beta}$ are statistically significant[30].

Equation (A.3) shows that $\frac{\hat{\alpha}}{\hat{\beta}}$ is a biased estimator of $\frac{E(\hat{\alpha})}{E(\hat{\beta})}$ by a factor of $\left[\frac{1+\frac{Var(\hat{\beta})}{E(\hat{\beta})^2}}{1+\frac{Cov(\hat{\alpha},\hat{\beta})}{E(\hat{\alpha})E(\hat{\beta})}}\right]$. This bias problem can be partially corrected by multiplying $\frac{\hat{\alpha}}{\hat{\beta}}$ by the inverse of that factor, which is the shrinkage factor $S = \left[\frac{1+\frac{Cov(\hat{\alpha},\hat{\beta})}{E(\hat{\alpha})E(\hat{\beta})}}{1+\frac{Var(\hat{\beta})}{E(\hat{\beta})^2}}\right]$. On average,

$$E\left(\frac{\hat{\alpha}}{\hat{\beta}}S\right) \approx \frac{E(\hat{\alpha})}{E(\hat{\beta})}\left[\frac{1+\frac{Var(\hat{\beta})}{E(\hat{\beta})^2}}{1+\frac{Cov(\hat{\alpha},\hat{\beta})}{E(\hat{\alpha})E(\hat{\beta})}}\right]\left[\frac{1+\frac{Cov(\hat{\alpha},\hat{\beta})}{E(\hat{\alpha})E(\hat{\beta})}}{1+\frac{Var(\hat{\beta})}{E(\hat{\beta})^2}}\right] = \frac{E(\hat{\alpha})}{E(\hat{\beta})}$$

Therefore, the MELO solution is the same as Taylor second-order approximation under the condition that both $\frac{Var(\hat{\beta})}{E(\hat{\beta})^2}$ and $\frac{Cov(\hat{\alpha},\hat{\beta})}{E(\hat{\alpha})E(\hat{\beta})}$ are small, or in other words, when both $\hat{\alpha}$ and $\hat{\beta}$ are statistically significant. On the other hand when either $\frac{Var(\hat{\beta})}{E(\hat{\beta})^2}$ or $\frac{Cov(\hat{\alpha},\hat{\beta})}{E(\hat{\alpha})E(\hat{\beta})}$ is large, these two approaches are not the same, and the MELO approach does not necessarily reduce bias.

4.8.2 Derivation of MELO Estimate of Truncated Mean WTP⁺

The derivation of equation (3.3) is analogous to that of equation (3.1). First, following equation (3.1), the MELO estimate of the truncated WTP^+ can be written as

$$W\hat{T}P^{+}_{MELO} = \frac{E(\ln(1+e^{\alpha}))}{E(\beta)} S^{+} \tag{B.1}$$

$$S^{+} = \frac{1+\frac{cov(ln(1+e^{\alpha}),\beta)}{E(ln(1+e^{\alpha}))E(\beta)}}{1+\frac{var(\beta)}{E(\beta)^2}}$$

To approximate the posterior mean and covariance of $\ln(1+e^{\alpha})$ is S^{+}, the Delta method is applied. Taking the second-order Taylor expansion of $\ln(1+e^{\alpha})$ around $E(\alpha)$,

$$\ln(1+e^{\alpha}) \approx E(\alpha) + \frac{e^{\alpha}}{1+e^{\alpha}}(\alpha - E(\alpha)) + \frac{e^{-E(\alpha)}}{2(1+e^{-E(\alpha)})}(\alpha - E(\alpha))^2 \tag{B.2}$$

Taking expectations on both sides of equation (B.2) yields the approximated expression for posterior mean of $ln(1+e^{\alpha})$

$$E\left(ln(1+e^{\alpha})\right) \approx E(\alpha) + \frac{e^{-E(\alpha)}}{2(1+e^{-E(\alpha)})} Var(\alpha) \tag{B.3}$$

Similarly, the delta method (that linearlizes around $E(\alpha)$) is used to approximate the posterior covariance,

---

[30] When $\hat{\beta}$ is statistically significant for example, $\frac{\hat{\beta}}{std(\hat{\beta})} > 2.0$, and therefore, $\frac{Var(\hat{\beta})}{\hat{\beta}^2} < 0.25$.

$$Cov(ln(1+e^{\alpha}),\beta)=(ln(1+e^{\alpha}))'Cov(\alpha,\beta)=\frac{e^{E(\alpha)}}{1+e^{E(\alpha)}}Cov(\alpha) \qquad \text{(B.4)}$$

Finally, substituting equations (B.3) and (B.4) into S^{+} in equation (B.1) and using $E(\alpha)=\hat{\alpha}$, $E(\beta)=\hat{\beta}$, $Cov(\alpha,\beta)=Cov(\hat{\alpha},\hat{\beta})$, and $Var(\alpha)=Var(\hat{\alpha})$ under diffuse prior, we get equation (3.3).

Similarly, $\frac{Cov(\hat{\alpha},\hat{\beta})}{E(\hat{\alpha})E(\hat{\beta})}<0.25$ since correlation is less than 1 in absolute value.

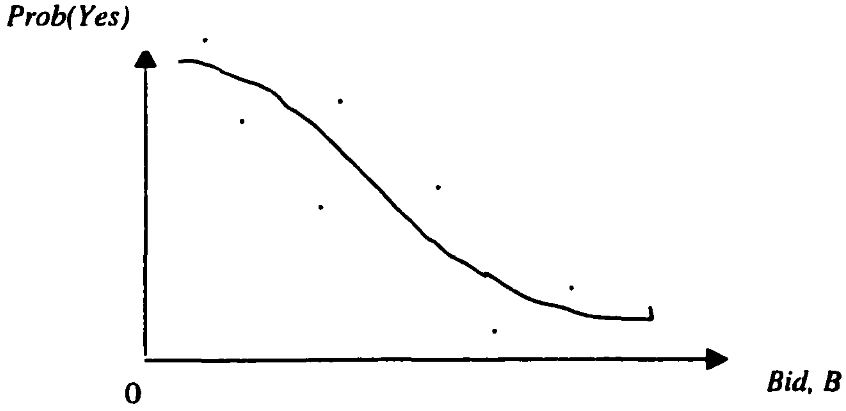

Figure 4.1 Dichotomous Choice (DC) Willingness-to-pay (WTP) Model

Figure 4.2

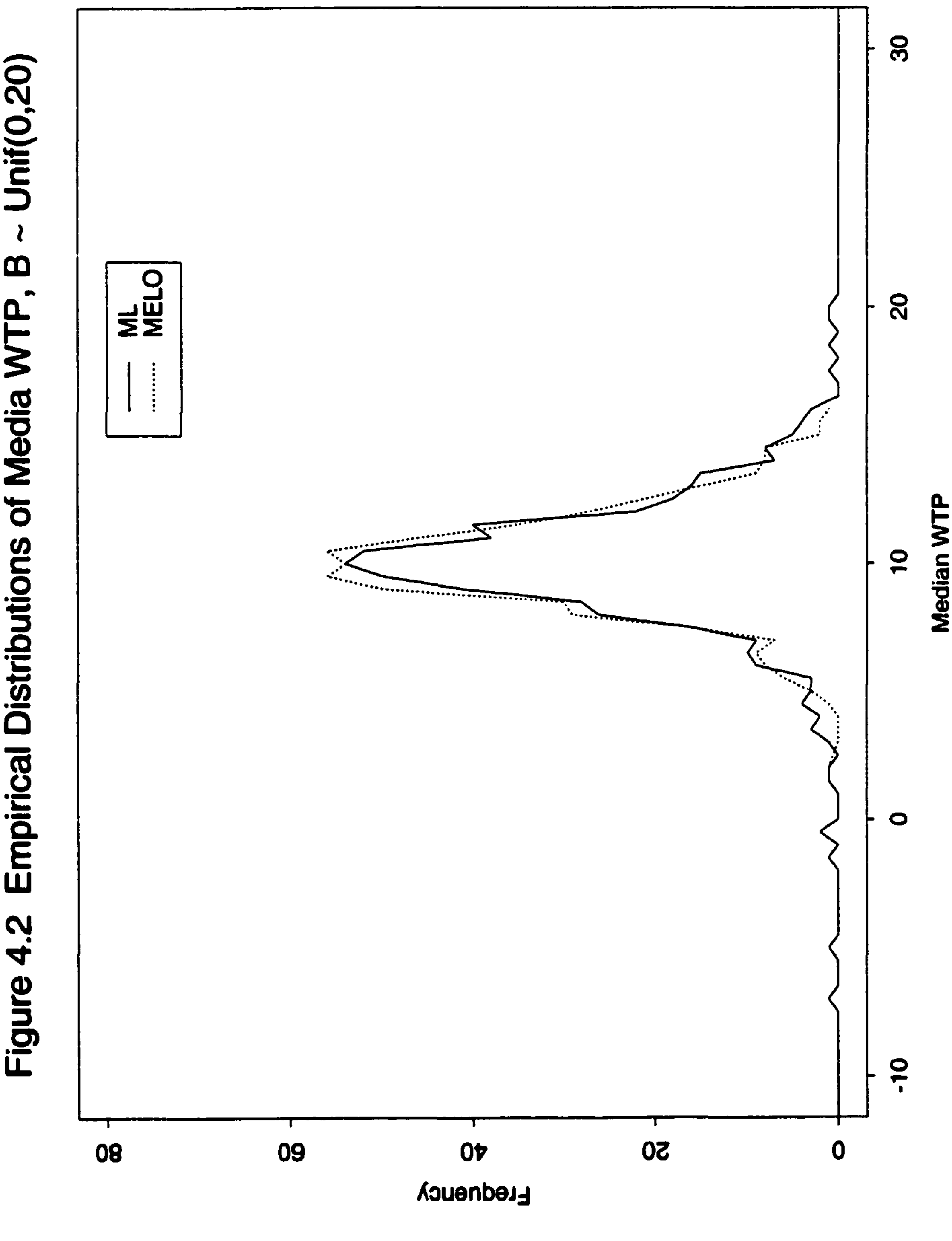
Figure 4.2 Empirical Distributions of Media WTP, B ~ Unif(0,20)
ML
MELO
Frequency
0
20
40
60
80
Median WTP
-10
0
10
20
30

Figure 4.3

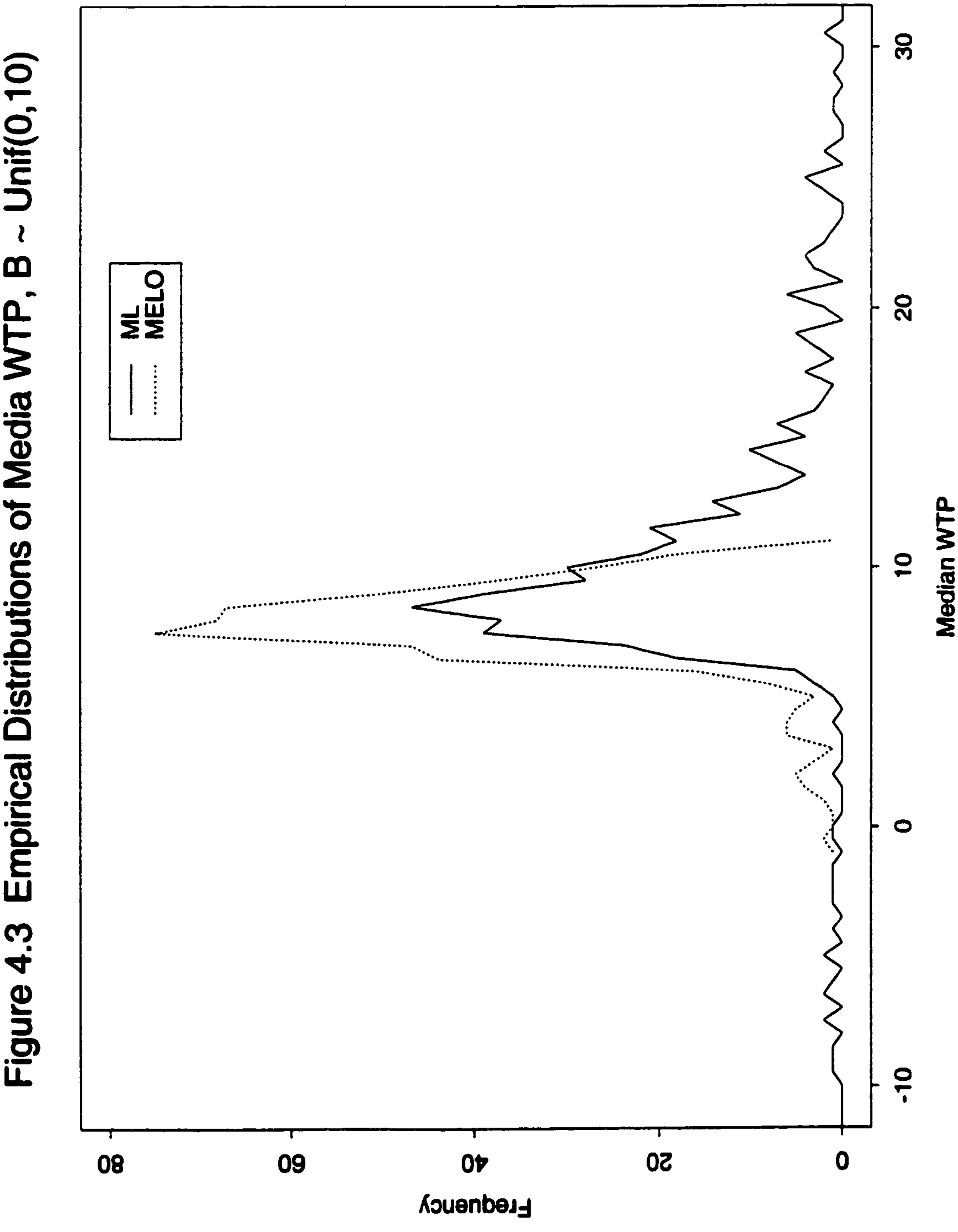

Figure 4.3 Empirical Distributions of Media WTP, B ~ Unif(0,10)
ML
MELO
Median WTP
Frequency
-10
0
10
20
30
0
20
40
60
80

Figure 4.4

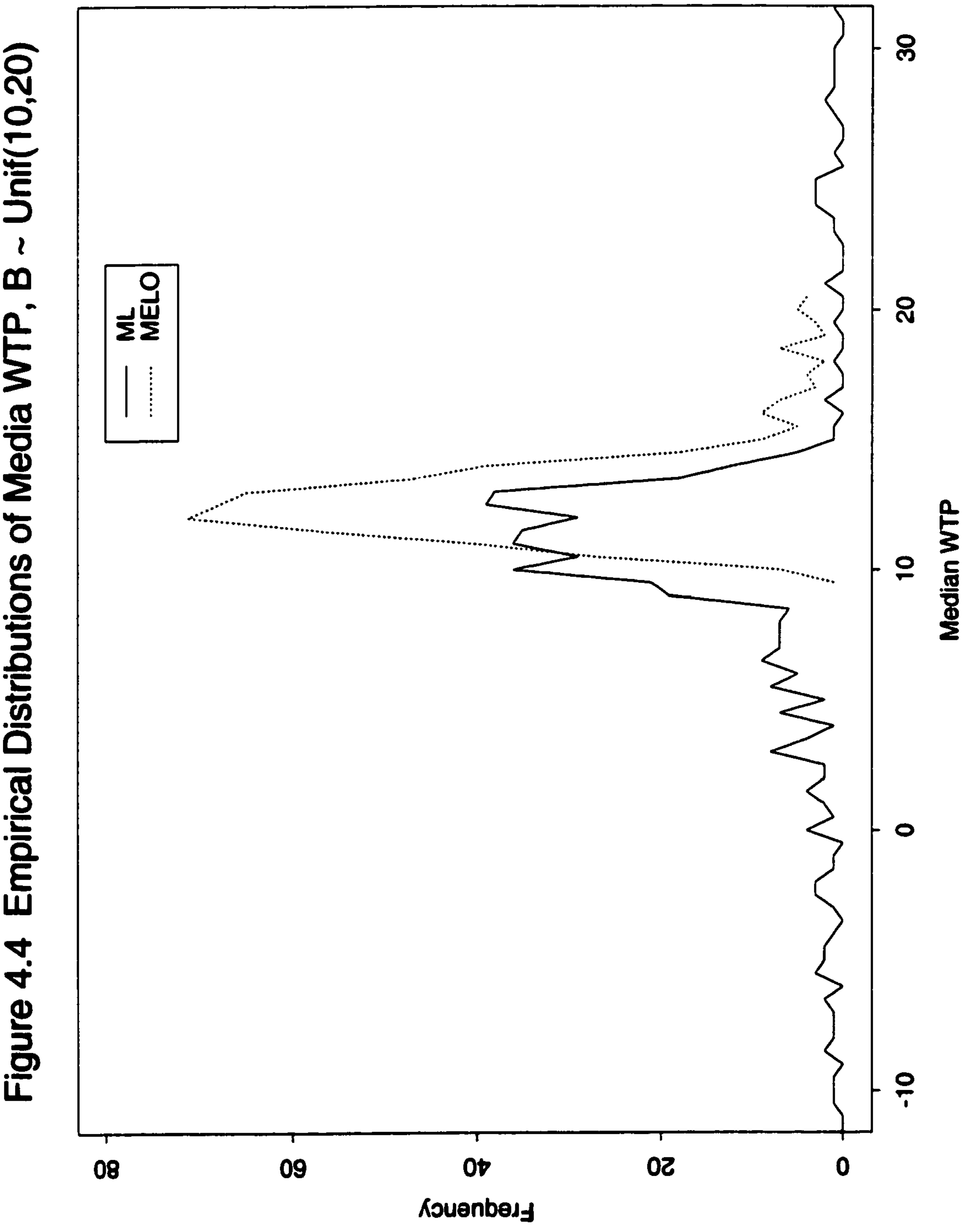
Figure 4.4 Empirical Distributions of Media WTP, B ~ Unif(10,20)
ML
MELO
Frequency
0
20
40
60
80
Median WTP
-10
0
10
20
30

Table 4-1 Empirical Distributions of the Estimated Median WTP.

	$B \sim Unif(0,20)$		$B \sim Unif(0,10)$		$B \sim Unif(10,20)$	
	ML	MELO	ML	MELO	ML	MELO
Minimum	-13.89	5.10	-113.82	-0.40	-361.30	9.71
Maximum	16.13	15.79	34.03	9.71	48.17	17.50
Mean	9.38	9.83	7.01	7.31	3.07	12.67
MSE	16.92	4.69	371.04	11.81	2864.8	9.79

Note: $\alpha = 2$, $\beta = 0.2$, median $WTP = \frac{\alpha}{\beta} = 10$, $N = 30$, and number of experiments = 500.

Table 4-2 Empirical Distribution of the Estimated Truncated Mean WTP⁺.

	$B \sim Unif(0,20)$		$B \sim Unif(0,10)$		$B \sim Unif(10,20)$	
	ML	MELO	ML	MELO	ML	MELO
Minimum	6.65	6.32	-831.75	-2.26	-100.04	-0.34
Maximum	21.56	13.86	124.45	10.81	20.38	15.05
Mean	11.16	10.18	-4.99	7.06	7.99	10.10
MSE	8.47	3.08	14,619	20.34	282.10	15.83

Note: $\alpha = 2$, $\beta = 0.2$, $WTP^+ = \frac{ln(1+e^{\alpha})}{\beta} = 10.635$, $N = 30$, and number of experiments = 500.

Table 4-3 Replication of the Monte Carlo Experiments in Kanninen's (1995)

	ML Approach		*MELO Approach*	
	Mean	MSE	Mean	MSE
[A]Base bid range	198.4	1120.3	194.7	1075
[B]Extended bid range	205.5	1984.9	196.0	1828.6
[C]Middle Only	197.2	380.4	197.7	301.0
[D]Mostly middle, some tail	205.2	708.5	204.5	637.4
[E]Upper tail only	202.4	6190.7	230.1	3932.9

Note $\beta = 0.009$, true WTP = 200, *N*=100, and number of experiments = 500.

[A]: The bid value are 3, 5, 10, 20, 40, 50, 60, 70, 80, 90, 100, 120, 200, 250, 300, 400, 500, and 700, five of each bid to total 100 observations.

[B]: The bid values are 3, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 500, 700, 1000, 1500, and 2000, five each.

[C]: Bids are 106, 155, 200, 245, and 294, 20 each.

[D]: Bids are 15 each at 106, 155, 200, and 294 and 5 each at 3, 5, 500, 700, 10000.

[E]: Bids are 300, 400, 500, 700, and 1000, 20 each.

Table 4-4 Replication of the Monte Carlo Experiments in Cooper and Loomis (1993)

		Minimum	*Maximum*	*Mean*	*MSE*
		A. All bids included in the Simulated Data Sets			
$\gamma=0$	ML	145.1	298.7	210.1	1270.1
	MELO	142.9	292.0	206.0	1178.8
$\gamma=1$	ML	116.7	295.5	209.5	1480.1
	MELO	115.9	285.7	204.5	1303.8
$\gamma=2$	ML	123.3	429.9	212.7	2648.8
	MELO	121.8	382.6	206.3	2073.1
		B. Middle Bids Excluded from the Simulated Data Sets			
$\gamma=0$	ML	115.3	275.0	205.5	1111.4
	MELO	111.9	271.6	203.5	1119.8
$\gamma=1$	ML	136.3	295.9	208.5	1491.1
	MELO	134.4	293.3	206.4	1469.1
$\gamma=2$	ML	125.9	312.9	207.0	1717.7
	MELO	125.3	310.6	204.9	1644.9
		C. Outer Bids Excluded in the Simulated Data Sets			
$\gamma=0$	ML	-4910.1	669.0	173.0	0.271E+06
	MELO	81.9	283.1	197.2	1171.6
$\gamma=1$	ML	-5360.2	7004.5	279.6	0.810E+06
	MELO	108.1	272.4	189.5	1231.0
$\gamma=2$	ML	-1429.4	9316.5	267.0	0.917E+06
	MELO	32.1	245.3	163.5	3217.5

Note $\beta = 0.0{,}8561$, $\theta = 0.00372$, median WTP = 206.7, N =100, and number of experiments = 500. Equal numbers of each bid value are assigned to each simulated data set.
A: the bid values are 2.5, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90. 100, 120, 150, 200, 250, 300, 400, and 700.
B: the middle bids ($50 through $250) are excluded.
C: only the middle bids ($50 through $250) are included.

Table 4-5 Re-etimates of Median WTP in Hanemann, Loomis, and Kanninen (1991)

	Single-Bounded WTP Model	*Double-Bounded WTP Model*
ML (Hanemann et al)	$250	$151
MELO	$211.67	$151.58

Note: The ML approach is used to estimate both SB and DB WTP Model in Hanemann et al (1991)

Table 4-6 Re-estimates of Median WTP in Riddle and Loomis (1998)

	Single-Bounded WTP Model	*Double-Bounded WTP Model*
	California/Oregon Combined Program	
ML	$32.16	$25.24
MELO	$26.71	$25.05
	Oregon Program Only	
ML	$16.52	$16.55
MELO	$15.03	$16.30
	California Program Only	
ML	$25.78	$19.90
MELO	$18.24	$19.37

Table 4-7 Estimated Median WTP under Diffuse and Informative Prior.

	$B \sim Unif(0,20)$		$B \sim Unif(0,10)$		$B \sim Unif(10,20)$	
	MELO	MELO-I	MELO	MELO-I	MELO	MELO
Minimum	2.68	4.17	-1.22	4.09	7.67	8.48
Maximum	16.78	17.93	11.02	13.94	13.59	13.50
Mean	9.89	9.94	7.36	8.68	12.31	11.66
MSE	3.87	3.87	10.63	4.40	8.17	3.96

Note: MELO assumes diffuse prior. MELO-I has the informative prior restriction that $\alpha > 0$ and $\beta > 0$. Parameters are set at $\alpha = 2$, $\beta = 0.2$, median $WTP = 10$, $N = 30$, number of experiments = 500, and $J = 1000$.