

Maximum entropy and Bayesian approaches to the ratio problem

Edward Z. Shen*
Jeffrey M. Perloff**

January 2001

Abstract

Maximum entropy and Bayesian approaches provide superior estimates of a ratio of parameters, as this paper illustrates using the classic Nerlove model of agricultural supply. Providing extra information in the supports for the underlying parameters for generalized maximum entropy (GME) estimators or as an analytically derived prior distribution in Zellner's minimum expected loss (MELO) estimators and Bayesian method of moments (BMOM) estimators helps substantially. Simulations illustrate that GME, MELO, and BMOM estimators with "conservative" priors have much smaller mean square errors and average biases than do standard ordinary least squares or MELO and BMOM estimators with uninformative priors. In addition, a new estimator of the structural agricultural supply model provides estimates of parameters that cannot be obtained directly using traditional, reduced-form approaches.

Key Words: agricultural supply response; generalized maximum entropy; Bayesian estimation; MELO; BMOM; simulation; shrinkage

JEL classification: C1; C2; C3; C4; C5

* Senior Decision Scientist, Gator.com, 2000 Bridge Parkway, Suite 100, Redwood City, CA 94065

** Professor and Member of the Giannini Foundation, Department of Agricultural and Resource Economics, 207 Giannini Hall, University of California, Berkeley 94720-3310

We are extremely grateful to Arnold Zellner for bringing this problem to our attention. He made numerous constructive and insightful suggestions on a large number of drafts. He suggested that we investigate the properties of the informative MELO and BMOM estimators as well as the GME estimator. Russell Lamb was very helpful in discussing the experimental design. Amos Golan and George Judge gave us very helpful advice on our GME and Bayesian estimators. Adrian Pagan gave us useful comments. GAMS, Matlab, and Shazam were used to estimate the various models. We thank the Giannini Foundation for support. An earlier version of this paper was part of the first author's dissertation.

Maximum entropy and Bayesian approaches to the ratio problem

1. Introduction

We compare the estimates of a ratio of parameters using traditional methods, generalized maximum entropy (GME, Golan, Judge, and Miller, 1996), two Bayesian approaches, minimum expected loss (MELO; Zellner 1978; Zellner and Park 1979) and the Bayesian method of moments (BMOM; Zellner 1996, 1997). We use simulations to show that two versions of the GME estimator and the Bayesian estimators with informative priors have much smaller mean square errors (MSE) and average biases than do the traditional techniques or Bayesian estimators with uninformative priors.

The problem of estimating the ratio of two parameters appears frequently. The ratio problem arises when one estimates market power (Bresnahan, 1982), the dose-response relationship (Zellner and Rossi, 1984), the Keynesian savings rate in Haavelmo's model of long-run supply response (Bewley and Fiebig 1990), the willingness-to-pay in contingent valuation studies (Shen 1999), and the Nerlove model of agricultural supply response (Nerlove and Addison 1958; Nerlove 1979; Diebold and Lamb 1997).

We compare various estimators using the Nerlove model of agricultural supply response, one of the world's most widely-used econometric models in development economics. Most of the literally hundreds of applications of this model that use ordinary least squares (OLS) or nonlinear least squares suffer from two problems. First, because not all the parameters of the structural model can be estimated using traditional maximum likelihood (ML) estimators, these studies estimate only the reduced-form model. Second, estimates of

the key, supply-response parameter are extremely variable, as the survey by Askari and Cummings (1977) illustrates.

We use GME and Bayesian methods to overcome these two problems. First, we can estimate a structural model using either GME or BMOM, and hence estimate structural parameters that cannot be estimated using standard, reduced-form methods.

Second, we can reduce the variability of OLS estimates of a ratio of parameters. Zellner (1978) showed that the OLS reduced-form estimator of such a ratio possesses infinite moments of all orders and may have a bimodal distribution. For example, in the model below, the OLS reduced-form estimator of α (the slope of the supply curve) is $\hat{\alpha} = \hat{b}_2 / (1 - \hat{b}_3 - \hat{b}_4)$, where the \hat{b}_i , $i = 2, 3, 4$, are least squares estimators for the reduced-form equation coefficients b_i . Unfortunately, $\hat{\alpha}$ possesses infinite moments (hence OLS is inadmissible relative to quadratic and other loss functions). Diebold and Lamb (1997) demonstrate that a MELO estimate of α based on the reduced-form equation using an uninformative prior (MELO-U) has smaller MSE than does OLS. We show similar improvements for BMOM and GME estimators.

The GME approach requires that, for each parameter and realized error, we specify support spaces: ranges within which each estimated parameter and realized error lies. The support space is specified based on economic theory or other prior knowledge. Using Bayesian terminology, imposing a support space means that the prior on a parameter is no longer diffuse. We indicate below how the restrictions from the specification of a support space on a realized error can be imposed with Bayesian estimators. We compare the estimators based on finite support spaces or informative priors (GME-R, GME-S, MELO-I, BMOM-

I) to traditional (OLS) and other noninformative estimators (MELO-U and BMOM-U) to demonstrate the role that prior restrictions play. We also conduct sensitivity analyses to changes in the priors.

In Section 2, we describe the Nerlove agricultural supply model. In the next four sections, we describe the OLS, MELO, BMOM, and GME estimators. In Section 7, we describe the Monte Carlo simulation results. We examine how sensitive the shrinkage estimators are to prior information in Section 8. We draw conclusions in the final section.

2. The Agricultural Supply Response Model

The standard structural Nerlove model is¹

$$A_t^* = \alpha_0 + \alpha P_t^e + u_t, \quad (1)$$

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

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

$$u_t \sim \overset{\text{iid}}{\left(0, \sigma_u^2\right)}, \quad (4)$$

where for period t , 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 Eq. 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. The adaptive expectations Eq. 2 links P^e to P . Muth (1960) shows that these

¹ 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.

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 presumably between zero and one).

The structural model, Eqs. 1-4 cannot be estimated by the usual OLS 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:

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

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. 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 regressors. 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 $1 - \gamma$, the reduced-form disturbance is also approximately white noise.

Using Eq. 5, the key parameter α can be expressed as the following function of the reduced-form parameters:

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

where $\delta \equiv 1 - b_3 - b_4$. The usual OLS estimator for α , $\hat{\alpha} = \hat{b}_2/\hat{\delta}$, is the ratio of the least squares estimator, \hat{b}_2 , divided by $\hat{\delta} \equiv 1 - \hat{b}_3 - \hat{b}_4$ least squares estimators using Eq. 5 and thus does not possess finite moments. As Zellner (1978, 1985, 1986), Zellner and Park (1979), Zaman (1981), and Diebold and Lamb (1997) observe, ratios or reciprocals of unbounded random variables do not have finite moments. Marsaglia (1965) shows that, if two random variables are normally and independently distributed, their ratio can be expressed as a function of several bivariate normals or as a Nicholson's V function. He also provides conditions under which the ratio's distribution is bimodal or unimodal (see also Zellner, 1978). Lehmann and Popper Shaffer (1988) show that the distribution of reciprocals or ratios may be bimodal for more general distributions. The nonexistence of moments and the multimodality may contribute to substantial variability in estimates of agricultural supply response.

4. Minimum Expected Loss

To avoid the problems from using ordinary least squares, Diebold and Lamb (1997) suggest using Zellner's (1978, 1986) minimum-expected-loss (MELO) approach to estimate

α .² The MELO estimator for α 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.

The MELO approach minimizes the expectation of a generalized quadratic loss function,

$$L = (b_2 - \delta\hat{\alpha})^2 = \delta^2(\alpha - \hat{\alpha})^2,$$

where $\hat{\alpha}$ is an estimate of the true α . This loss function contrasts with the standard quadratic loss function, $(\alpha - \hat{\alpha})^2$, which specifies the same loss for a given absolute error regardless of the value of α . As Diebold and Lamb show for the Nerlove model, the MELO shrinkage estimate (which minimizes the posterior expected loss) is

$$\hat{\alpha} = \frac{E(b_2)}{E(\delta)} \frac{1 + \text{cov}(b_2, \delta) / [E(b_2)E(\delta)]}{1 + \text{var}(\delta) / E^2(\delta)} = \frac{E(b_2)}{E(\delta)} F, \quad (7)$$

where F is the shrinkage factor, $E(\cdot)$ denotes a posterior expectation, $\text{var}(\cdot)$ indicates a posterior variance, and $\text{cov}(\cdot)$ is the posterior covariance. Thus, the MELO estimate is the product of the posterior mean ratio and the shrinkage factor, which depends on higher posterior moments of b_2 and δ .

To estimate the parameters in Eq. 5, Diebold and Lamb (1997) assumed a normal likelihood function, a diffuse prior, and a reduced-form disturbance that is approximately white noise. Thus, their marginal posterior density for the coefficients in Eq. 5 is a multi-

² 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.

variate Student-t distribution that was employed to evaluate the moments in Eq. 7. Henceforth, we refer to such estimates as MELO-U, where the U indicates that the priors are uninformative or diffuse. Given the above assumptions, the Bayesian posterior conditional density of \underline{b} given the data and σ (the standard deviation) is a multivariate normal density:

$$\{\mathbf{b}|\sigma, X, A\} \sim MVN\left((X'X)^{-1}X'A, (X'X)^{-1/2}\sigma\right), \quad (8)$$

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

$$f(\sigma | X, A) \propto \frac{1}{\sigma^{\nu+1}} \exp\left(-\frac{\nu s^2}{2\sigma^2}\right), \quad (9)$$

where the degrees of freedom [given four regressors in the reduced-form Eq. 5] is $\nu = T - 4$, the variance of the residuals is $s^2 = \hat{e}'\hat{e}/\nu$.

The MELO-I estimates are similar to the MELO-U estimates except that we use informative priors for the parameters $\underline{b} = (b_1, b_2, b_3, b_4)'$. Since the exact posterior distribution of \underline{b} given an informative prior given inequality constraints is difficult to derive analytically, we use Geweke's (1986) numeric method to compute the posterior distribution. To do so, we start by generating values of σ^2 using Eq. 9. We then generate the vector \underline{b} using Eq. 8 conditional on the generated values of σ^2 . Next, we reject the drawn b 's that fail to satisfy the prior restrictions. The remaining observations represent a pseudo-random sample from the actual posterior distribution of \underline{b} that are consistent with the prior restric-

tions. The MELO-I shrinkage estimator is of the same form as Eq. 7 where we evaluate moments in Eq. 7 using the generated values of the b 's.

For comparison with the GME estimators, we sometimes also want to restrict the realized errors to lie within a certain range. In traditional Bayesian approaches, no priors are placed on the realized error terms (except through the choice of a likelihood function). For the purposes of comparison, we use the Geweke method to restrict the realized errors, which are a function of the \underline{b} and of the left-hand-side variable. That is, we generate a vector of the residuals as $\hat{\epsilon}^{(i)} = \hat{A} - X\bar{\mathbf{b}}^{(i)}$ and then reject those observations for which $\hat{\epsilon}$ fail to satisfy the prior restrictions.³

5. Bayesian Method of Moments

We also use Zellner's (1996, 1997) Bayesian method of moments (BMOM) with and without informative priors to obtain estimates of the reduced-form Nerlove model in our experiments. BMOM allows researchers to compute post-data densities for parameters and future observations when not enough information is available to estimate a satisfactory likelihood function. BMOM differs from traditional Bayesian analyses in that it is based on two weak assumptions about moment conditions of data without specifying the prior density and the likelihood function.

³ In an autocorrelation problem where a variable is regressed on its lags, Zellner and Tiao (1964) and Geweke (1986) suggest treating the presample values of the left-hand-side variable as either fixed or as unknown with a diffuse prior so that the form of the posterior distribution remains unchanged. If we could view A in this manner, this method would follow immediately from Geweke's work. As it is, we present these results for comparison purposes with the GME results only. In Section 8, we examine the sensitivity of our results to such restrictions.

To derive the BMOM estimator of the standard multivariate linear regression model, $y = X\beta + u$, Zellner (1996) made the following two assumptions about moment conditions of the data. The first moment condition is that the error term is uncorrelated with the regressors X given the data $D = (y, X)$, so that $(X'X)^{-1}X'E(u|D) = 0$. The second moment assumption is that the post-data variance-covariance matrix for the realized error vector u given the data and σ^2 , the variance of the parameter, is $Var(u | \sigma^2, D) = \sigma^2 X(X'X)^{-1}X'$.

Zellner (1996) explained how to use maximum entropy (maxent) to compute the BMOM posterior density function given these two moment conditions. The maxent post-data density for $\underline{b} = (b_1, b_2, b_3, b_4)'$ in the Nerlove model given the data and the two moment restrictions is in the following multivariate normal density:

$$\{b | \sigma, X, A\} \sim MVN\left((X'X)^{-1}X'A, (X'X)^{-1/2}\sigma\right).$$

Zellner showed that, when the first moment of σ^2 is used as a side condition, the proper maxent posterior distribution for σ^2 is an exponential distribution with a mean equal to s^2 .⁴

$$f_{BMOM}(\sigma^2) = \frac{1}{s^2} \exp(-\sigma^2/s^2).$$

⁴ When more moments are used, this approach leads to other post-data densities for the coefficients and σ^2 (Zellner and Tobias, 1997).

Our MELO and BMOM estimates use the same objective function but vary due to differences in estimating procedures.⁵ Minimizing the generalized quadratic loss function, we obtain our BMOM-U estimate of α :

$$\alpha = \frac{E(b_2)}{E(\delta)} \frac{1 + \text{cov}(b_2, \delta)/[E(b_2)E(\delta)]}{1 + \text{var}(\delta)/E^2(\delta)}. \quad (10)$$

We use Geweke's (1986) method to impose prior restrictions that \underline{b} and the realized errors lie within certain ranges to obtain the BMOM-I estimates. Alternatively, one could introduce bounds on the parameters and realized error terms — which implies a bound on the σ^2 parameter — and derive maxent densities for parameters that incorporate these bounds as side conditions.)

6. Generalized Maximum Entropy

We now introduce a GME approach to estimate both the structural and reduced-form agricultural supply models. We start by briefly reviewing the maxent approach and then discuss its application in our model.

6.1 Maximum Entropy

The traditional maxent (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

⁵ An alternative BMOM estimator could be obtained by using the balanced loss function (Zellner 1994) instead of the quadratic loss function.

(state of knowledge) we have about the occurrence of a collection of events. Letting x be a random variable with possible outcomes x_s , $s = 1, 2, \dots, n$, with probabilities δ_s such that $\sum_s \delta_s = 1$, Shannon (1948) defined the *entropy* of the distribution $\underline{\delta}$ as

$$H \equiv -\sum_s \delta_s \ln \delta_s, \quad (11)$$

where $0 \ln 0 \equiv 0$. The function H , reaches a maximum of $\ln(n)$ when $\delta_1 = \delta_2 = \dots = \delta_n = 1/n$. 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 sample-moment information and adding up constraints on the probabilities. The intuition behind this approach is that the frequency that maximizes entropy is a reasonable estimate of the true distribution when we lack any other information. If we have information from an experiment, such as the sample moments, or nonsample information about the random variable such as restrictions from economic theory, we use that information to alter this estimate. 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.

6.2 GME Estimator

The GME objective maximizes the joint entropy of the parameters ($\alpha_0, \alpha, \gamma, \theta$ 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). We weight the two entropies equally; however, by varying the weight placed on the entropy of the parameters or the error terms, we can tradeoff

between precision and predictive power. By putting relatively more weight on the entropy from the parameters, we improve the accuracy of estimation (decrease the MSE of the estimates of the coefficients) at the cost of predictive power. If we put more relatively more weight on the error terms entropy, we predict better. The ME estimator is a special case of the GME, in which no weight is placed on the noise component.

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 (rather than the more natural $\alpha, \alpha_0, b_3,$ and b_4) to facilitate comparisons with previous papers.

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 $\underline{z}^i = [z_{i1}, z_{i2}, \dots, z_{iM}]'$ 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 symmetric about zero with "large" negative and positive bounds. We use $M = 3$ because we find little if any gain from using larger M such as 5 or 7.

The parameters are $b_i = \sum_{m=1}^M p_{im} z_{im}$, where $i = 1, \dots, 4$ and the p_{im} are probabilities

that correspond to the M -dimensional support vectors of weights with the restriction that

$\sum_{m=1}^M p_{im} = 1$ for $i = 1, \dots, 4$. Similarly, we reparameterize the errors so that $e_t = \sum_{j=1}^J w_{tj} v_j$,

where $\sum_{j=1}^J w_{tj} = 1$ for all t , and \underline{v} is a support space of dimension $J \geq 2$ and is symmetric

about zero.

Having reparameterized the unknowns, we maximize the sum of the entropies of $\underline{p}' = (p_1', p_2', p_3', p_4')$ and $\underline{w} = (w_1, \dots, w_T)'$:

$$\max_{\underline{p}, \underline{w}} H = - \underline{p}' \ln \underline{p}' - \underline{w}' \ln \underline{w}' \quad (12)$$

subject to⁶

$$\begin{aligned} A_t &= \sum_{m=1}^M p_{1m} z_{1m} + \sum_{m=1}^M p_{2m} z_{2m} P_{t-1} + \sum_{m=1}^M p_{3m} z_{3m} A_{t-1} \\ &\quad + \sum_{m=1}^M p_{4m} z_{4m} A_{t-2} + \sum_{j=1}^J w_{tj} v_{tj} \\ \sum_{m=1}^M p_{1m} &= \sum_{m=1}^M p_{2m} = \sum_{m=1}^M p_{3m} = \sum_{m=1}^M p_{4m} = \sum_{j=1}^J w_{tj} = 1. \end{aligned}$$

Having obtained the optimal values \hat{p} and \hat{w} , we can compute the GME-R coefficient

$$\text{estimates as } \hat{b}_i = \sum_{m=1}^M \hat{p}_{im} z_m.$$

⁶ As LaFrance (1999) notes, these GME conditions are restrictions on the means of the parameters and the error terms. We impose no conditions on higher-order moments (as, for example, may be used in the BMOM approach).

The GME structural-model estimator (GME-S) is handled similarly (Golan, Judge, Miller, 1997). We reparameterize the coefficients of the structural model, Eqs. 1-3, in terms of probabilities. The probabilities q_0 correspond to α_0 , q_α to α , q_γ to γ , q_θ to θ , and $\underline{\omega}_t$ to the realized error terms u_t . Let $\underline{q}' = (q_0', q_\alpha', q_\gamma', q_\theta')$. The supports are defined analogously. We maximize the joint entropies of the signal and noise:

$$\max_{\underline{q}, \underline{\omega}} H = - \underline{q}' \ln \underline{q}' - \underline{\omega}' \ln \underline{\omega}', \quad (13)$$

subject to

$$\begin{aligned} A_t^* &= \sum_{m=1}^M q_{0m} z_{0m} + \sum_{m=1}^M q_{\alpha m} z_{\alpha m} P_t^e + \sum_{j=1}^J \omega_{tj} v_{tj} \\ P_t^e &= P_{t-1}^e + \sum_{m=1}^M q_{\gamma m} z_{\gamma m} (P_{t-1} - P_{t-1}^e) \\ A_t &= A_{t-1} + \sum_{m=1}^M q_{\theta m} z_{\theta m} (A_{t-1}^* - A_{t-1}) \\ \sum_{m=1}^M q_{0m} &= \sum_{m=1}^M q_{\alpha m} = \sum_{m=1}^M q_{\gamma m} = \sum_{m=1}^M q_{\theta m} = \sum_{j=1}^J \omega_{tj} = 1. \end{aligned}$$

Solving the problem in Eq. 13, we obtain our GME-S estimates of α_0 , α , γ , θ , and P_t^e and A_t^* for all t .⁷

⁷ Similarly, using the Bayesian estimation procedures, we obtain densities for the parameters and hence could derive densities for P_t^e and A_t^* .

6.3 Supports and Prior Restrictions

We specify the structural GME support spaces for α_0 , α , γ , and θ . Because we assume no overshooting, we specify that the support of γ and θ ranges from 0 and 1. [If we wanted to allow for overshooting, we would require only that the support be positive.]

In the reduced-form methods, MELO-I, BMOM-I, and GME-R, we place supports (or prior restrictions) on the reduced-form parameters b_2 , b_3 , and b_4 that are consistent with the supports for the structural parameters: $\alpha = b_2/[1 - b_3 - b_4]$. From theory, we know that $b_3 \in (0, 2)$ and $b_4 \in (-1, 0)$.

In the Appendix, we show that, if $b_2 \sim \text{Uniform}(-d, d)$ for some given d , $b_3 \sim \text{Uniform}(0, 2)$, and $b_4 \sim \text{Uniform}(-1, 0)$, the analytical prior density function of $f(\alpha)$, given that b_2 , b_3 , and b_4 are independent is

$$f(\alpha) = \begin{cases} \frac{2}{3d}, & \text{if } -d < 2\alpha < d \\ \frac{1}{2d} \left(\frac{d^2}{\alpha^2} - \frac{1}{3} \frac{d^3}{\alpha^3} \right), & \text{otherwise} \end{cases}$$

Figure 1 shows the prior distribution of α given $d = 10$, the value we use in the following experiments. (It also shows several posterior distributions that are discussed later). Thus, even if all the reduced-form parameter are distributed uniformly, the prior distribution of α is not uniform. For a larger d (a more diffuse prior on b_2), the variance of the prior distribution on α increases. This distribution approaches a uniform, uninformative prior as t becomes infinite.

7. Experiments

To compare the sampling properties of OLS, MELO-U, MELO-I, BMOM-U, BMOM-I, GME-R, and GME-S estimators, we use simulations. 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 MSE and average bias.

7.1 *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:⁸

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

$\rho = 0.5$ and 0.9

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

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

where

⁸ 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 these experiments to save space.

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

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. 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.

We specify the structural GME support spaces as $[-20, 0, 20]$ for α_0 , $[-10, 0, 10]$ for α , and $[0, 0.5, 1]$ for γ and θ . 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. (Additional unreported experiments indicate that allowing these supports to be $[0, 2]$ slightly improves the GME results.) We follow Golan, Judge, and Miller (1996) in

using a "three sigma" rule to determine the support for the error term in each sample. That is, we set the support bounds at $\pm 3\hat{\sigma}$, where $\hat{\sigma}$ is sample standard deviation of the left-hand-side variable (for example, the sample standard deviation for A in our simulations range from 4 to 6). Below, we discuss the sensitivity of our results to this restriction.

In the reduced-form methods, we place supports (or prior restrictions) on the reduced-form parameters b_2 , b_3 , and b_4 that are consistent with the supports for the structural parameters. The prior restrictions in MELO-I, BMOM-I, and GME-R are derived from the structural supports using Eq. 5: $[-20, 20]$ for b_1 , $[-10, 10]$ for b_2 , $[0, 2]$ for b_3 , and $[-1, 0]$ for b_4 .

7.2 Results

The following tables and figure summarize our experimental results.⁹ 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. We also examine the relative out-of-sample predictive power of the four estimates using the correlation between the actual and predicted values of A .

⁹ 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.

Table 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. This table replicates 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.

Tables 2 ($\rho = 0.5$) and 3 ($\rho = 0.9$) 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 $\gamma = 0.5$ and 1 . When γ equals 1 , the OLS approach is appropriate — and it performs better than when $\gamma = 0.5$. As γ increases or T increases, the advantage of MELO-U over OLS in terms of MSE decreases, 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.¹⁰

Figure 1 shows the posterior distribution for the various Bayesian estimators with and without informative priors for one sample drawn from the first experiment in Table 2 ($\alpha = 2$,

¹⁰ These results could reflect overfitting due to using many parameters, as Zellner (1999) notes. If so, adding prior information may improve the fit of the GME and BMOM-I, but could result in worse predictions. However, the following prediction experiments do not indicate an obvious problem.

$\sigma_\varepsilon = 1$, $\sigma_u = 5$, $\gamma = 1$, $\theta = 0.5$, $\rho = 0.5$, and $T = 25$). The figure shows that the uninformative posteriors have larger variance than the informative posteriors.

Figure 2 illustrates the superiority of the shrinkage estimators by comparing the histograms of the various empirical distributions of $\hat{\alpha}$ for the first experiment in Table 2. The frequency scale for the GME-R, GME-S, MELO-I, and BMOM-I is compressed relative to that for the other estimators so that all the distributions fit on the same figure. 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 (and similarly a single line is shown for MELO-U and BMOM-U and another for MELO-I and BMOM-I). The MELO-I and BMOM-I distributions are only slightly wider than the GME ones, and much smaller than the OLS, MELO-U, or BMOM-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. The OLS estimates are likely to be much less precise when the denominator, $\gamma\theta$, of the ratio in Eq. 6, $\alpha = b_2/\delta = b_2/(\gamma\theta)$, 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 $\hat{\alpha}$ 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.

Table 4 shows the recovered reduced-form estimates of α_0 and α and the GME-S structural estimates of α_0 , α , γ , and θ . As with α , the shrinkage estimators, MELO-I, BMOM-I, GME-R, and GME-S, have much smaller MSEs for α_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 directly recovered from the reduced-form estimates. In our experiments, the GME-S estimates of γ have smaller MSEs than those for θ . (Especially when γ is close to 1, we get better estimate of γ and θ if we use the support $[0, 1, 2]$ instead of $[0, 0.5, 1]$, as in this table.)

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 5 shows the in-sample 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). However, 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 lower than for the reduced-form models.

The GME-S provides estimates of the predicted price, P_t^e , and desired acreage, A_t^* , as Table 6 shows.¹¹ In our experiments, GME-S does a remarkable job of predicting P_t^e — the correlations range from 0.767 to 0.975 — but is less impressive in its estimates of A_t^* — the correlations range from 0.323 to 0.596.

¹¹ 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).

To examine the performance of these estimated models in predicting out of sample, we constructed a thousand replications of two paired data sets of 25 observations each generated using the parameters in the upper-left cell of Table 2. We estimated parameters using the first sample set and then compared our forecasts to the data in the second sample. The average out-of-sample correlation for the predicted and actual acreage was 0.413 for the OLS, MELO-U, and BMOM-U estimators. All of the reduced-form estimators with informative priors had higher correlations (GME-R, 0.470; MELO-I, 0.487; BMOM-I, 0.487). The structural model estimator, GME-S, did worse, 0.342 (perhaps due to overfitting).

Table 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 χ^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 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 informative estimators.

8. 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 the uninformative prior Bayesian 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 Table 8, we show what happens to the MSE (for 100 replications) as the supports become wider. 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.]

The GME estimators use a support of $[-3\hat{\sigma}, 0, 3\hat{\sigma}]$ for the error term. Table 8 shows the change in 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. 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.

9. Conclusions

Generalized maximum entropy and two Bayesian methods perform substantially better than traditional approaches in estimating ratios of random variables in Monte Carlo experiments involving the Nerlove model of agricultural supply response. Diebold and Lamb (1997) showed that the variability of estimates of the key parameters in the Nerlove model was partially due to the problem of estimating a ratio regression parameters by using the ratio of their respective ordinary least squares estimates, as is commonly done. 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 informative 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 fairly unrestricted priors (little additional information) are used. The informative Bayesian estimators (and to a lesser extent the GME-R) perform particularly well in out-of-sample prediction.

We also use the generalized maximum entropy approach to obtain point estimates of the full structural agricultural supply response model (so that we obtained estimates of more parameters than with reduced-form models). Although we did not do so, one could estimate

the posterior densities for these extra parameters using the Bayesian methods by sampling from the reduced-form coefficients' posterior densities and solving the nonlinear relations connecting the reduced-form and structural parameters for each draw.

The GME has two advantages over classical and Bayesian approaches. First, the GME approaches performed much better than did the other methods when the underlying error distribution was nonnormal in our simulations.¹² Second, imposing inequality restrictions (such as bounding the adjustment parameters between 0 and 1) is easy in the GME approach and in Bayesian analyses (using Geweke's method). Possible future lines of research will show whether there is a gain to using higher-order moments and generalized entropy criterion functionals in GME and BMOM.

¹² With the other estimators, if one suspects nonnormality one can use various transformation to account for departures from normality. One can also add higher-order moment constraints in the BMOM approach to obtain asymmetric densities.

References

- Askari, A. and T. Cummings, 1977, Estimating agricultural supply response with the Nerlove model: A survey, *International Economic Review* 18, 257-329.
- Bewley, R. and D. Fiebig, 1990, Why are long-run parameter estimates so disparate? *Review of Economics and Statistics* 72, 345-349.
- Bresnahan, T. F., 1982, The oligopoly solution concept is identified, *Economics Letters*, 10, 87-92.
- 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.
- 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. McLaughlin, ed., *Inverse problems* (American Mathematical Society, Providence, Rhode Island).
- Kullback, J., 1959, *Information theory and statistics* (John Wiley & Sons, New York).
- LaFrance, J. T., 1999, Inferring the nutrient content of food with prior information," *American Journal of Agricultural Economics* 81:728-734.
- 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.
- Marsaglia, George, 1965, "Ratios of Normal Variables and Ratios of Sums of Uniform Variables," *American Statistical Association Journal* 60, 193-204.
- 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.

- Shen, Zhihua, 1999, *Essays of Empirical Studies in Agricultural and Resource Economics*, Ph.D. dissertation, University of California, Berkeley.
- 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. Reprinted in A. Zellner, *Bayesian analysis in econometrics and statistics*, Elgar, 1997.
- 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., 1999, New information-based econometric methods in agricultural economics: Discussion, *American Journal of Agricultural Economics* 81:742-746.
- 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.
- Zellner, A. and P. E. Rossi, 1984, Bayesian analysis of dichotomous quantal response models, *Journal of Econometrics* 25, 365-393.
- Zellner, A. and G.C. Tiao, 1964, Bayesian analysis of the regression model with autocorrelated errors, *Journal of the American Statistical Association* 59, 763-778.
- Zellner, A. and J. Tobias. Forthcoming. Further results on Bayesian method of moments analysis of the multiple regression model. *International Economic Review*.

Appendix: Derivation of the Prior Density of α

We derive the prior density of $\alpha \equiv b_2/(1 - b_3 - b_4) \equiv b_2/\delta$, where b_2 , b_3 , and b_4 are independent, uniformly distributed random variables: $b_2 \sim \text{Uniform}(-d, d)$, $b_3 \sim \text{Uniform}(0, 2)$, and $b_4 \sim \text{Uniform}(-1, 0)$. Using a Jacobian transformation, the density of α is

$$f_\alpha(\alpha) = \int_{-\infty}^{\infty} |x| f_{b_2}(x\alpha) f_\delta(x) dx, \quad (\text{A1})$$

where $f_{b_2}(\bullet)$ and $f_\delta(\bullet)$ are density functions for the numerator and denominator of α respectively. These densities are:

$$f_{b_2}(x) = \begin{cases} \frac{1}{2d} & -d < x < d \\ 0 & \text{otherwise} \end{cases} \quad (\text{A2})$$

$$f_\delta(x) = \begin{cases} 0 & x < -1 \\ x + 1 & -1 \leq x < 0 \\ 1 & 0 \leq x < 1 \\ 2 - x & 1 \leq x < 2 \\ 0 & x \geq 2 \end{cases} \quad (\text{A3})$$

The density function $f_\delta(\cdot)$, $\delta = 1 - b_3 - b_4$, can be derived using Theorem 1 of Marsaglia (1965). It is symmetric around 0.5.

To derive $f_\alpha(\cdot)$, we substitute Eq. A2 and A3 into A1:

$$f_\alpha(\alpha) = \begin{cases} \int_{-1}^0 \frac{-x(x+1)}{2d} dx + \int_0^1 \frac{x}{2d} dx + \int_1^2 \frac{x(2-x)}{2d} dx, & |\alpha| < \frac{d}{2} \\ \int_{-1}^0 \frac{-x(x+1)}{2d} dx + \int_0^1 \frac{x}{2t} dx + \int_1^{d/\alpha} \frac{x(x-\alpha)}{2d} dx, & \frac{d}{2} \leq \alpha < d \text{ or } -d < \alpha \leq \frac{d}{2} \\ \int_{-d/\alpha}^0 \frac{-x(x+1)}{2d} dd + \int_0^{d/\alpha} \frac{x}{2d} dx, & |\alpha| \geq d \end{cases}$$

Simplifying this last expression, we obtain

$$f_\alpha(\alpha) = \begin{cases} \frac{2}{3d}, & |\alpha| < \frac{d}{2} \\ \frac{1}{2d} \left(\frac{d^2}{\alpha^2} - \frac{1}{3} \frac{d^3}{\alpha^3} \right), & \text{otherwise.} \end{cases}$$

Thus, $f_\alpha(\alpha)$ is symmetric around 0, is uniform between $(-d/2, d/2)$, and asymptotically approaches 0 as α goes to positive or negative infinity outside of that range.

Table 1
Mean Square Error of the estimated α from experimental data; $T = 25$, $\alpha = 2$, $\gamma = 1$, number of samples = 1,000

		$\rho = 0.5$						
σ_ε	σ_u	<i>OLS</i>	<i>MELO-U</i>	<i>MELO-I</i>	<i>BMOM-U</i>	<i>BMOM-I</i>	<i>GME-R</i>	<i>GME-S</i>
1	5	3.9299	1.1394	0.0081	1.1561	0.0130	0.00028	0.00016
3	3	0.1461	0.1271	0.0188	0.1280	0.0201	0.00007	0.00008
2	3	0.3077	0.2647	0.0147	0.2664	0.0169	0.00012	0.00009
3	5	0.4040	0.2928	0.0143	0.2947	0.0169	0.00019	0.00016
5	5	0.1653	0.1352	0.0224	0.1364	0.0259	0.00014	0.00014

		$\rho = 0.9$						
σ_ε	σ_u	<i>OLS</i>	<i>MELO-U</i>	<i>MELO-I</i>	<i>BMOM-U</i>	<i>BMOM-I</i>	<i>GME-R</i>	<i>GME-S</i>
1	5	1.7559	1.1284	0.0080	1.1439	0.0175	0.00033	0.00015
3	3	0.1542	0.1291	0.0186	0.1303	0.0203	0.00007	0.00007
2	3	0.3385	0.2469	0.0154	0.2488	0.0170	0.00012	0.00009
3	5	0.3897	0.2925	0.0156	0.2951	0.0183	0.00021	0.00018
5	5	0.1544	0.1306	0.0220	0.1319	0.0236	0.00015	0.00015

Table 2

Empirical distributions of estimated supply response coefficient α ; $\alpha = 2, \sigma_\varepsilon = 1, \sigma_u = 5, \rho = 0.5$

		$\gamma = 0.5$						
T		<i>OLS</i>	<i>MELO-U</i>	<i>MELO-I</i>	<i>BMOM-U</i>	<i>BMOM-I</i>	<i>GME-R</i>	<i>GME-S</i>
25	Mean	1.38	1.06	1.97	1.08	1.98	1.99	2.00
	SD	3.34	1.01	0.12	1.03	0.15	0.03	0.01
	MSE	11.55	1.90	0.02	1.90	0.02	0.001	0.0002
	Min	-48.49	-2.43	0.91	-2.44	1.37	1.58	1.92
	Max	74.36	7.09	2.81	7.61	2.91	2.48	2.08
	Bias	1.37	1.16	0.09	1.16	0.11	0.02	0.01
50	Mean	1.33	1.24	1.93	1.24	1.95	1.98	2.00
	SD	0.82	0.75	0.11	0.75	0.15	0.04	0.01
	MSE	1.12	1.14	0.02	1.14	0.02	0.002	0.0002
	Min	-1.46	-1.33	1.31	-1.34	1.46	1.30	1.84
	Max	7.47	6.80	2.45	6.82	2.53	2.04	2.03
	Bias	0.88	0.90	0.10	0.90	0.12	0.03	0.01
100	Mean	1.37	1.32	1.88	1.32	1.90	1.97	2.00
	SD	0.55	0.53	0.10	0.53	0.14	0.04	0.02
	MSE	0.70	0.74	0.02	0.74	0.03	0.003	0.0003
	Min	-0.35	-0.36	1.55	-0.36	1.31	1.29	1.88
	Max	3.38	3.28	2.34	3.28	2.42	2.04	2.21
	Bias	0.71	0.74	0.13	0.74	0.13	0.04	0.01
		$\gamma = 1.0$						
T		<i>OLS</i>	<i>MELO-U</i>	<i>MELO-I</i>	<i>BMOM-U</i>	<i>BMOM-I</i>	<i>GME-R</i>	<i>GME-S</i>
25	Mean	1.79	1.56	1.98	1.58	1.98	1.99	2.00
	SD	1.17	0.96	0.09	0.97	0.12	0.02	0.01
	MSE	1.41	1.12	0.01	1.12	0.01	0.0004	0.0002
	Min	-3.24	-1.24	1.32	-1.34	1.45	1.76	1.96
	Max	7.34	5.87	2.32	5.94	2.46	2.04	2.04
	Bias	0.91	0.86	0.07	0.86	0.09	0.01	0.01
50	Mean	1.93	1.83	1.97	1.83	1.98	2.00	2.00
	SD	0.76	0.70	0.08	0.70	0.11	0.02	0.01
	MSE	0.59	0.52	0.01	0.52	0.01	0.0003	0.0001
	Min	-0.44	-0.55	1.63	-0.54	1.49	1.89	1.94
	Max	5.43	5.14	2.42	5.15	2.57	2.04	2.03
	Bias	0.59	0.56	0.06	0.57	0.09	0.01	0.01
100	Mean	1.97	1.92	1.97	1.92	1.98	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	0.28	0.01	0.001	0.0002
	Min	0.78	0.77	1.64	0.77	1.57	1.89	1.90
	Max	4.22	4.07	2.39	4.07	2.34	2.06	2.06
	Bias	0.43	0.43	0.08	0.43	0.09	0.02	0.01

Table 3**Empirical distributions of estimated supply response coefficient α ; $\alpha = 2$, $\sigma_\varepsilon = 1$, $\sigma_u = 5$, $\rho = 0.9$**

<i>T</i>		$\gamma = 0.5$						
		<i>OLS</i>	<i>MELO-U</i>	<i>MELO-I</i>	<i>BMOM-U</i>	<i>BMOM-I</i>	<i>GME-R</i>	<i>GME-S</i>
25	Mean	1.46	1.49	1.95	1.50	1.97	1.99	2.00
	SD	5.28	0.89	0.11	0.90	0.15	0.03	0.01
	MSE	28.10	1.05	0.01	1.06	0.02	0.001	0.0002
	Min	-129.95	-2.14	1.45	-2.20	0.88	1.68	1.95
	Max	13.44	5.66	2.42	5.76	2.63	2.03	2.04
	Bias	1.10	0.80	0.09	0.80	0.11	0.02	0.01
50	Mean	1.77	1.71	1.92	1.72	1.93	1.99	2.00
	SD	0.57	0.54	0.12	0.54	0.13	0.02	0.01
	MSE	0.38	0.38	0.02	0.38	0.02	0.001	0.0002
	Min	-0.93	-0.79	1.47	-0.80	1.41	1.84	1.95
	Max	4.10	3.80	2.35	3.81	2.40	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	1.82	1.90	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	0.73	1.45	1.88	1.92
	Max	3.62	3.38	2.35	3.39	2.42	2.06	2.04
	Bias	0.29	0.30	0.14	0.30	0.14	0.02	0.01
<i>T</i>		$\gamma = 1.0$						
		<i>OLS</i>	<i>MELO-U</i>	<i>MELO-I</i>	<i>BMOM-U</i>	<i>BMOM-I</i>	<i>GME-R</i>	<i>GME-S</i>
25	Mean	1.87	1.75	1.97	1.76	1.99	2.00	2.00
	SD	0.88	0.77	0.09	0.77	0.10	0.02	0.01
	MSE	0.79	0.65	0.01	0.66	0.01	0.0003	0.0002
	Min	-1.07	-0.73	1.39	-0.73	1.63	1.87	1.95
	Max	6.80	5.22	2.55	5.34	2.56	2.04	2.04
	Bias	0.67	0.63	0.07	0.63	0.08	0.01	0.01
50	Mean	1.96	1.92	1.97	1.92	1.97	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	0.23	0.01	0.0004	0.0002
	Min	0.06	0.04	1.63	0.04	1.59	1.89	1.94
	Max	3.81	3.61	2.31	3.62	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	0.28	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	0.96	1.68	1.92	1.92
	Max	3.27	3.25	2.32	3.25	2.36	2.05	2.04
	Bias	0.22	0.22	0.10	0.22	0.10	0.02	0.01

Table 4

Estimates of the structural coefficients; $\sigma_\varepsilon = 1$, $\sigma_u = 5$, $\alpha = 2$, $\alpha_0 = 0.25$, $\rho = 0.5$, $\gamma = 0.5$, $\theta = 0.5$, $T = 25$

	α_0			α		
	<i>Mean</i>	<i>SD</i>	<i>MSE</i>	<i>Mean</i>	<i>SD</i>	<i>MSE</i>
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	1.08	1.03	1.90
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		
	<i>Mean</i>	<i>SD</i>	<i>MSE</i>	<i>Mean</i>	<i>SD</i>	<i>MSE</i>
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 5

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

T	γ	ρ	<i>OLS, MELO-U, BMOM-U*</i>	<i>MELO-I</i>	<i>BMOM-I</i>	<i>GME-R</i>	<i>GME-S</i>
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 6**GME-S correlations between P^c and \hat{P}^c and between A^* and \hat{A}^* ; $\sigma_\varepsilon = 1$, $\sigma_u = 5$, $\rho = 0.5$, $\gamma = 0.5$, $\theta = 0.5$, $T = 25$**

T	γ	ρ	P^c and \hat{P}^c	A^* and \hat{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 7

MSE for α 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_5^2	χ_7^2	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	0.106	0.114	0.106	0.082
BMOM-I	0.032	0.051	0.032	0.034
GME-R	0.00005	0.00013	0.00005	0.00006
GME-S	0.00007	0.00006	0.00004	0.00003

Table 8**MSE for α 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

	<i>MELO-I</i>	<i>BMOM-I</i>	<i>GME-R</i>
original	0.009	0.020	0.007
5× larger	0.453	0.492	0.142
10× larger	0.911	0.958	0.469
20× larger	1.075	1.185	0.898
40× larger	1.041	0.986	1.160

Change of the support for only the errors

	<i>MELO-I</i>	<i>BMOM-I</i>	<i>GME-R</i>	<i>GME-S</i>
σ	0.0094	0.0199	0.00071	0.00017
2σ	0.0169	0.0280	0.00011	0.00011
3σ	0.0260	0.0437	0.00010	0.00011
4σ	0.0269	0.0768	0.00010	0.00011
5σ	0.0430	0.1214	0.00010	0.00011
10σ	0.0353	0.6173	0.00009	0.00011
20σ	0.0424	1.6837	0.00009	0.00011
100σ	0.0490	1.9794	0.0099	0.0116
200σ	0.0499	1.9257	0.115	0.132

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

Figure 1 Prior and Posteriors of α

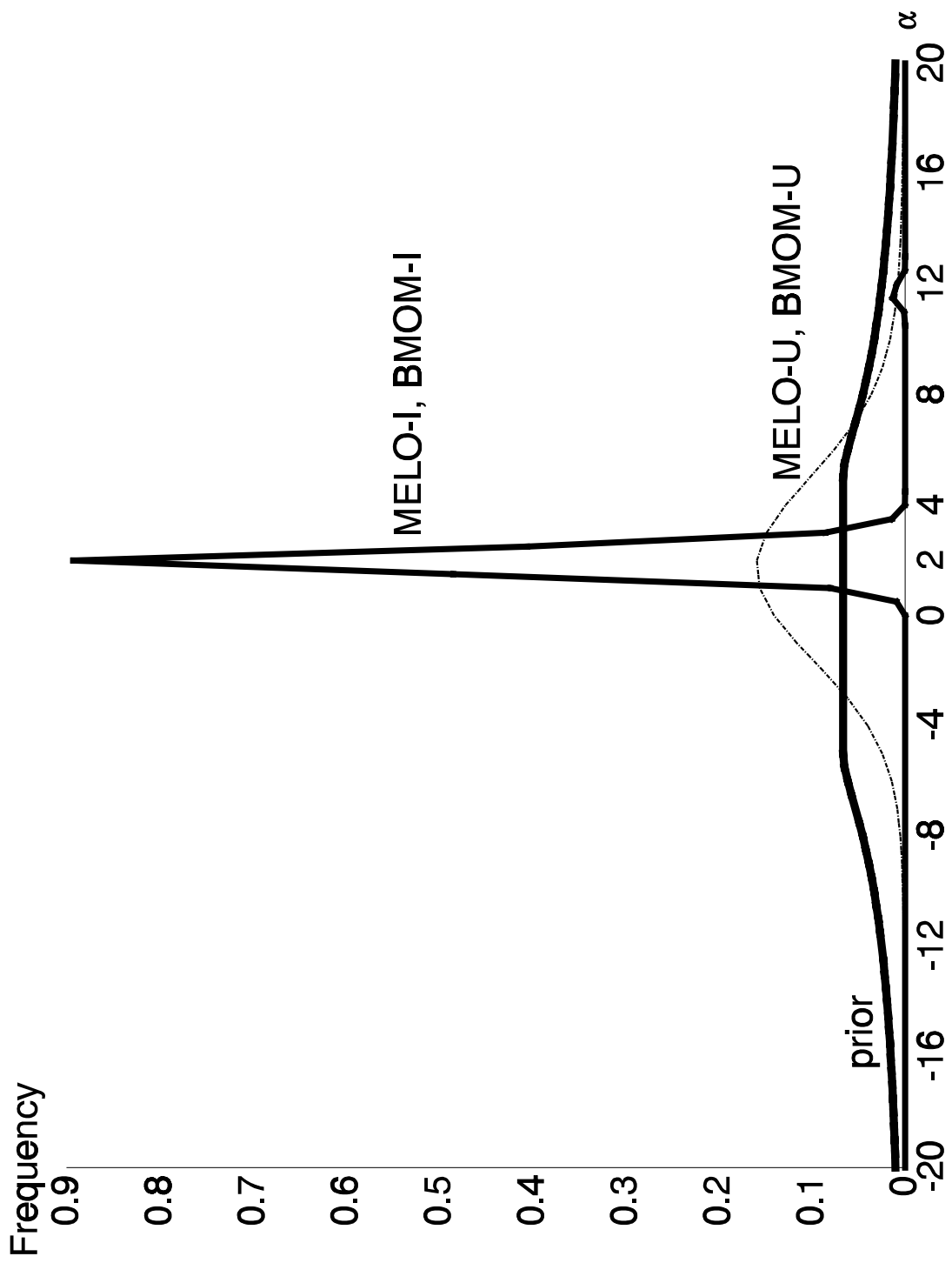


Figure 2 Empirical Distributions of α

