ISSN 0316-5078

INSTITUTE FOR ECONOMIC RESEARCH

QUEEN'S UNIVERSITY

Kingston, Ontario, Canada K7L 3N6

CALIBRATION AS ESTIMATION

Allan W. Gregory* and Gregor W. Smith

Department of Economics
Queen's University
Kingston, Ontario
Canada
K7L 3N6

Discussion Paper Number 700

* Department of Economics
University of Western Ontario
London, Ontario
Canada, N6A 5C2

We are grateful to David Backus, Thomas McCurdy, Simon Power and Stan Zin for helpful comments. The second author acknowledges the support of the Social Sciences and Humanities Council of Canada.

December, 1987

ABSTRACT

One aspect of calibration in macroeconomics is the notion that the free parameters of models should be chosen by matching certain moments of the simulated models with those of actual data. We formally examine this notion by treating the process of calibration as an econometric estimator. A numerical version of the Mehra-Prescott (1985) economy is the setting for an evaluation of calibration estimators via Monte Carlo methods. While these estimators sometimes may have reasonable finite-sample properties they are not robust to mistakes in setting non-free parameters. In contrast, generalized method-of-moments (GMM) estimators have satisfactory finite-sample characteristics, quick convergence, and informational requirements less stringent than those of consistent calibration estimators. In dynamic equilibrium models in which GMM is infeasible we offer some suggestions for improving estimates based on calibration methodology.

1. INTRODUCTION.

This paper examines the properties of a method used by some economists to attach numerical parameters to theoretical models and which has come to be known as calibration. This method was introduced to macroeconomics by Kydland and Prescott (1982) (see also Prescott, 1986) and some aspects of the procedure have been applied by Long and Plosser (1983), Hansen (1985), Kydland (1987), and Greenwood, Hercowitz, and Huffman (1987).[1] For this reason it has become associated with real business cycle models and criticism of the method has not always been distinguished from that of the models. In this paper we make this distinction sharp by constructing artificial economies and embedding in them supposed calibrators. This allows us to provide an (unavoidably model-specific) answer to the following question. If the qualitative form of the model posited by a theorist were true would calibration reveal the correct parameters? In our Monte Carlo analysis we formalize this correctness by treating the process of calibration as an econometric estimator. Although we refer to a particular model we believe that our evidence on calibration applies to a wide class of dynamic economic models.

A subsidiary purpose of this paper is to compare calibration estimators

---

[1] Slutsky (1937, page 119) may be regarded as having calibrated a time series model by comparing the coefficient of variation in its cycle length with this measure in business cycle data. He also (pages 109-110) compared visually the plots from various models with an index of the nineteenth century British business cycle.

with generalized method-of-moments (GMM) estimators which are natural to apply in the model that we consider. We compare the estimated probability density functions of GMM and calibration estimators under various parametric assumptions.

Section 2 describes calibration. While sharing elements and language common to both classical and Bayesian modellers,[2] calibrators for the most part have eschewed formal statistical inference, the calculation of posterior odds, forecasting, and the like in favour of relatively informal model evaluation. As practiced so far, this method also involves particular views on detrending and on the relationship between microeconomic and macroeconomic data. We abstract from these issues later in order to evaluate critically the view that a model's free parameters should be chosen by matching certain properties of the model with those of data.

Section 3 examines calibrators in a numerical version of the Mehra-Prescott (1985) economy as described by Backus, Gregory, and Zin (1986). For the most part these calibrators face a perfect correspondence (with no errors of measurement) between their model and the data generation process. They choose free parameters just as real calibrators do.

Section 4 presents simple Monte Carlo experiments in which the densities of the calibration and GMM estimators are compared and contrasted. Here we

---

[2] For these approaches see McAleer, Pagan, and Volker (1985) and Leamer (1983) respectively.

estimate nonparametrically the density functions.

Section 5 provides a summary of the results and offers specific recommendations for improving the performance of calibration as an estimation strategy.

2. CALIBRATION.

The method we shall examine was introduced by Kydland and Prescott (1982). These authors gave the following necessary criterion for evaluating theories:

> The test of the theory is whether there is a set of parameters for which the model's co-movements for *both* the smoothed series and the deviations from the smoothed series are *quantitatively* consistent with the observed behavior of the corresponding series for the U.S. post-war economy.(1982, page 1359, italics in original)

Series first are smoothed with an arbitrary linear filter. Then a subset of parameter values is imposed based on the long-run or average values of certain variables. As Manuelli (1986) notes, this use of growth (as well as cross-section) observations to calibrate the (cyclical part of the) model does not seem to agree with the view that the same model should explain both growth and fluctuations. For a critique of this first step in the calibration methodology see Singleton (1987).

A second, and larger, set of parameters is selected with reference to

> An added requirement ... that the parameters

> selected not be inconsistent with the relevant micro observations ...(1982, page 1360)

This use of cross-section or survey data in a model of cycles and growth is innovative. As Prescott (1986) notes

> A fundamental thesis of this line of inquiry is that the measures obtained from aggregate series and those from individual panel data must be consistent. After all, the former are just the aggregates of the latter.(page 14)

Other analysts of real business cycles also seek this consistency. For example, Long and Plosser (1983) find values for the parameters in their model by using input-output studies. This approach may be contrasted with that of, say, Taylor (1980) or Grossman and Weiss (1983). These authors treat contract length and transactions frequency respectively as free parameters, estimated in macroeconomic time series data. Their models can be tested by comparing these estimates with actual microeconomic observations. This reverses the method of Kydland and Prescott.

The evaluation of this method using simulation evidence would require a meta-data-generation-process relating cross-section to time series observations.[3] We choose instead to evaluate the third step in the calibration method, namely the selection of free parameters. As Lucas (1987) notes

---

[3] For some criticisms of mixing microeconomic and aggregate evidence, again see Singleton (1987).

> Kydland and Prescott began by estimating as many parameters as possible from a wide variety of out-of-sample evidence ... without even looking at the time series they were attempting to fit. ... [They] then chose values for [the free] parameters so as to make certain low order moments (variances, covariances, autocovariances) predicted by the model "match" the corresponding moments from the collection of time series in the sample they used. The result of this last step completed the estimation, and the matches between the theoretical and actual moments they reported are the only reported 'test' of the model's ability to 'fit' these series. (pages 42-43)

Kydland argues that a good theory is one with few free parameters so that

> the formulation permits the imposition of a great deal of quantitative discipline, and covariance or correlation matrices can easily be computed, both between contemporaneous and lagged variables, and compared with those of the aggregate data. (1987, page 6)

While we evaluate the method of selecting parameter values by matching certain properties of data and models, note that nothing in most calibrated models precludes the use of standard econometrics. Altug (1985) and Eichenbaum, Hansen, and Singleton (1987) study real business cycle models from this perspective.

In describing calibration, Kydland and Prescott wrote

> We chose not to test our model versus the less restrictive vector autoregressive model. This most likely would have resulted in the model being rejected, given the measurement problems and the abstract nature of the model. Our approach is to focus on certain statistics for which the noise introduced by approximations and measurement error is likely to be small relative to the statistic. (page 1360)

This approach involves conducting a grid search over the free parameters so

asto match the correlations of nominal variables with income, variability (root mean-squared percentage deviation from trend), or the number of peaks and troughs in the model with those in the smoothed series. How can the results of this exercise be evaluated? Lucas, a sympathetic critic, writes that "a more detailed comparison of the artificially generated series with their observed counterparts is not so encouraging." (page 40) However, Kydland and Prescott argue that

> The variables in our model do not correspond perfectly to those available for the U.S. economy so care must be taken in making comparisons. A second problem is that there may be measurement errors that seriously bias the estimated correlations and standard deviations. A final problem is that the estimates for the U.S. economy are subject to sampling error. (page 1363)

Likewise Prescott (1986) suggests that cyclical errors in measurement might account for some of the deviation between the data and the predictions of his model. In fact,

> An important part of this deviation could very well disappear if the economic variables were measured more in conformity with theory. That is why I argue that theory is now ahead of business cycle measurement and theory should be used to obtain better measures of the key economic time series. This feedback between theory and measurement is the way mature, quantitative sciences advance. (page 21)

In this paper we abstract from issues concerning measurement error. Again quoting Lucas,

> Whether [the results of calibration] are viewed as 'good' or 'bad' is a difficult question, as is the related question of which comparisons of theoretical to sample moments are most interesting. (page 45)

In what follows, we formally address these questions.

3. CALIBRATION AS ESTIMATION.

In our study of calibration we abstract from two elements in this methodology, namely (1) the method of detrending or pre-filtering practiced by calibrators, and (2) their views on testing theories. We concentrate on the effects of two other elements, namely (3) the division of the parameters of a model into free parameters and those selected on the basis of extra-sample information, and (4) the use of simulation estimators and the choice of particular moments to match. Our concern is to measure the combined effects of these two elements in small data sets.

To begin, suppose that an economic theory predicts that a variable p_t satisfies

$$f(p_t, x_t, \omega_0, \omega_1, \omega_2) = 0 \qquad (1)$$

in which ω_0, ω_1, and ω_2 are parameters and x_t is a state variable which evolves as

$$x_t = g(x_{t-1}, \omega_1, \varepsilon_t) \qquad (2)$$

and ε_t is a drawing from a probability density $G(\varepsilon, \omega_2)$.[4] An *estimator* may be seen as consisting of (i) a partition of the parameter vector (ω_0, ω_1, ω_2), and (ii) rules for choosing estimates in each part. In what follows we assume that the parameter of interest is ω_0 and examine *conditional estimators*, that is rules for inverting the system above to learn ω_0 which are conditional upon rules for choosing ω_1 and ω_2.

Many estimation strategies can be studied in this framework. For example, in calibration the parameters ω_1 and ω_2 are chosen on the basis of microeconomic studies or long-run values of time series.[5] Denote the values so chosen ω_1^c and ω_2^c and the calibrated forcing process $x_t^c = g(x_{t-1}^c, \omega_1^c, \varepsilon_t)$. The free parameters ω_0 are chosen to match certain properties of the series p_t^c generated artificially from

$$f(p_t^c, x_t^c, \omega_0, \omega_1^c, \omega_2^c) = 0 \qquad (3)$$

with those of the actual economic series p_t.

In contrast, GMM estimation involves choosing a value for the parameter ω_0 so that the sample analogue of equation (1) is as close to zero as

---

[4] In what follows we consider the univariate case for simplicity but vector processes for endogenous and state variables could be analysed similarly.

[5] We treat this selection as exogenous since we do not posit a meta-data generation process which gives rise to both time series and cross section data. But we do examine the effects of errors in this selection.

possible. If (1) is a set of equations with more elements than there are parameters to be estimated then over-identifying restrictions could be tested.

These two strategies can differ for several reasons. First, the predictions of a theory may be examined using calibration even when error terms for use in GMM cannot be formulated (see Labadie (1984) for an example). Second, calibrators may match different moments from those restricted in GMM estimation or they may focus on other properties of actual and simulated data. Third, where the two methods adopt the same moments, calibration will be inefficient since it uses simulated rather than actual data.[6] For asymptotic measures of this inefficiency see Ingram and Lee (1987) and also Pakes and Pollard (1986). This distinction between the methods obviously will not apply if the state variable is unobservable. Moreover, measurement errors in the data may give the state variable an element of unobservability which makes the comparison between estimation strategies an interesting one. Fourth, calibration estimators of ω_0 are conditioned on values assigned to the parameters ω_1 and ω_2 whereas GMM estimators normally do not require knowledge of these parameters. Clearly the quality of estimates of the free parameters in calibration may depend on the accuracy with which other parameters are set.

---

[6]The grid search used in calibration may also account for differences in general. Regularity conditions which rule out this effect apply in the examples below.

For these reasons analytical comparisons generally are not possible and hence the comparison of estimation stategies must be undertaken empirically. While we observe one realization of $\{p_t\}$ in history, in the Monte Carlo analysis to follow we may create many such histories by numerical methods in order to evaluate estimators.

This exercise presumes knowledge of the functions f and g and of the density G from which the forcing variable is generated. Without these one cannot possibly obtain estimates of ω_0. Economic theory may give rise to a specific function f but is unlikely to be very precise with respect to the function g or the density G. Our goal here is to investigate the usefulness of the calibration estimator when the economic environment generating the data is known up to a specific parameterization. Thus we examine the properties of calibration estimators for the most part under ideal conditions.

To make our discussion of these estimators more precise, we consider the economy described by Mehra and Prescott (1985). A representative consumer has preferences represented by the utility function

$$E_t \sum_{j=0}^{\infty} \beta^j \, u(c_{t+j}) \tag{4}$$

in which E_t denotes the expectation conditional upon information at time t, β is a positive discount rate, $u(c) = c^{1-\alpha}/(1-\alpha)$ ($\log(c)$ if $\alpha = 1$) and α is a positive, constant coefficient of relative risk aversion. The consumer's endowment y_t evolves exogenously according to

$$y_{t+1} = x_{t+1}\, y_t \tag{5}$$

in which the growth rate, x_t, follows a process which is Markov on a discrete state space Λ. This process has transition probability density Q and is stationary and ergodic.

An equilibrium in this economy is represented by a set of prices for which consumption equals the endowment at all times and in all states. Relative prices are calculated by equating them to marginal rates of substitution. The price of a one-period, riskfree, discount bond which provides one unit of the endowment at time t+1 is given by:

$$p_t = \beta\, E_t\, u'(y_{t+1})/u'(y_t) \tag{6}$$

We shall focus on the series $\{p_t\}_{t=0}^{T}$.

Following Mehra and Prescott let the growth rate take on s (finite) values which we denote λ_i, i = 1,2, ..., s. In principle this is not restrictive (since s may be as large as we wish) and it considerably simplifies computation. The probability structure is discrete-state Markov with transition probabilities

$$q_{ij} = \text{Prob}\,[x_{t+1} = \lambda_j \mid x_t = \lambda_i] \qquad i,j = 1,2, \ldots, s.$$

The equilibrium or unconditional probabilities are given by q_i = Prob $[x_t = \lambda_i]$ $\forall$ t. If the current state is (y_t, λ_i) then the price (relative to one unit now) of one unit of the commodity next period if state j occurs is:

$$q_{ij}\ \beta\ u'(y_{t+1})/u'(y_t) = q_{ij}\ \beta\ u'(\lambda_j\ y_t)/u'(y_t) \tag{7}$$

The mean, variance, and autocovariance can be calculated as follows

$$E[p_t] = \beta \sum_i^s \sum_j^s q_i\ q_{ij}\ \lambda_j^{-\alpha} \tag{8}$$

$$V[p_t] = \beta^2 \sum_i^s \sum_j^s q_i\ q_{ij} \left\{ \sum_j^s q_{ij}\lambda_j^{-\alpha} - \sum_i^s \sum_j^s q_i\ q_{ij}\ \lambda_j^{-\alpha} \right\}^2 \tag{9}$$

$$C[p_t, p_{t-1}] = \beta^2 \sum_i^s \sum_j^s \sum_k^s q_i\ q_{ij}\ q_{jk} \left\{ \left\{ \sum_j^s q_{ij}\ \lambda_j^{-\alpha} - \sum_i^s \sum_j^s q_i\ q_i\ \lambda_j^{-\alpha} \right\} \cdot \left\{ \sum_k q_{jk}\lambda_k^{-\alpha} - \sum_i^s \sum_j^s q_i\ q_{ij}\ \lambda_j^{-\alpha} \right\} \right\} \tag{10}$$

To simplify further we parameterize the transition probabilities as

$$q_{ij} = q_j + \theta\ (\delta_{ij} - q_j) \tag{11}$$

in which $\delta_{ij} = 1$ if $i = j$ and 0 otherwise (see Barton, David, and Fix, 1962). The persistence parameter θ must lie in $(-(s-1)^{-1}, 1)$ to ensure non-negative probabilities. This description not only reduces the number of parameters from $s(s-1)$ to $s+1$ but also allows us directly to control the degree of temporal dependence in the $\{x_t\}$ series by varying θ. If θ is zero, states are independent over time while if θ is positive (negative) growth rates will be positively (negatively) autocorrelated.

Under the simple persistence decomposition (11) the moments (8)-(10) become:

$$E[p_t] = \beta \sum_{i}^{s} q_i \ \lambda_i^{-\alpha} \tag{12}$$

$$V[p_t] = \beta^2\theta^2 \sum_{i}^{s} q_i \left\{ \sum_{j}^{s} q_j \ (\lambda_i^{-\alpha} - \lambda_j^{-\alpha}) \right\}^2 \tag{13}$$

$$C[p_t, p_{t-1}] = \theta \ V[p_t] \tag{14}$$

These population moments will prove helpful in interpreting the Monte Carlo evidence of section 4. Notice that under this parameterization the mean of the price does not depend on the serial correlation in the growth rate series. The variance and autocovariance increase as (i) the discount rate β increases, (ii) the persistence parameter θ increases in absolute value, (iii) the growth rates λ_i become more different, and (iv) the coefficient of relative risk aversion α increases. There is no price variability when either α or θ is zero.

Note that equations (6) and (5) in this theory correspond to equations (1) and (2) above. Since $\{x_t\}$ is a Markov process p_t is a function only of the current state. The transition probabilities in (11) and the state values λ_i correspond to the parameters ω_1 and ω_2 respectively and are assumed to be chosen in the initial calibration. Given choices of Q and Λ one can estimate α, which corresponds to the free parameter ω_0. This correspondence seems appropriate given the wide range of values attributed to this parameter in

empirical studies. For example, Hansen and Singleton (1983) and Grossman, Melino, and Shiller (1987) find estimates of α that vary between 0 and 400 depending upon the model estimated. Mehra and Prescott allow α to vary in order to determine whether the model can rationalize observed equity premia.

In order to limit the number of experiments we assume that there are only three states (s = 3) with growth rates λ_i = (0.9873 1.0055 1.0177) and equilibrium probabilities q_i = (0.25 0.5 0.25). These growth rates correspond roughly to the mean and plus and minus twice the standard deviation of the observed quarterly rates for per capita U.S. consumption from 1950:1 to 1985:4 (see Backus, Gregory, and Zin). The equilibrium probabilities imply a larger volatility in consumption than that observed over this period. Following the arguments of Prescott (1986), for quarterly data, the discount rate is β = 0.99.

Our investigation of calibration as estimation in this model involves five steps:

(1) Create an artificial economy by choosing α and θ.

(2) Generate R realizations or histories $\{p_t\}$ each of sample size T which we denote $\{p_t^r\}$, t = 1, ..., T and r = 1, ..., R.

(3) Calibrate a further artificial economy by using either

(i) the same θ as in step (1), or

(ii) a different θ, or

(iii) an estimated transition matrix (see the discussion below).

(4) Simulate the second economy over a range (itself based on some theory) of α, denoted A, for each calibration in step (3).[7] Restrict A = [0,16] (with a step size of 0.05) which mimics the inequality restrictions used by calibrators. Select the element of A, $\hat{\alpha}$, which minimizes a certain loss function (discussed below) over this range using history $\{p_t^r\}$ and calibration series $\{p_t^c\}$ for c = i, ii, iii from step (3). For each of the R histories, each calibration (i, ii, or iii) and loss function produces an estimate of α which we denote $\hat{\alpha}^r$, r = 1, ..., R.

(5) Estimate nonparametrically the density function of $\hat{\alpha}$.

Step (4) in this sequence, for a given history, mimics a single, actual calibration exercise. In this framework the calibrator uses the correct growth rates (as set up in step (1)), the correct number of states (λ_i, i= 1, ...,s), the corresponding equilibrium probabilities (q_i), and the discount rate β. To complete the parameterization of the artificial economy values for α and the simple persistence parameter θ are required. Three different calibrations for θ are considered in step (3). The first assumes that the true value of θ is used, the second uses a θ different (incorrect) from the one generating the data, and the third estimates the transition matrix (q_{ij})

---

[7] The sample size at this stage need not correspond to the sample size T of the historical data in step 2. In fact, Ingram and Lee have shown that the efficiency of simulation estimators increases as this simulation sample size gets large relative to T.

from the observed history of the consumer's endowment y_t for each of the R histories.

Studying the first of these calibrators gives us the density of various estimates of α under ideal conditions, in which the supposed calibrator is fortunate enough to set each parameter equal to the truth. We should emphasize that we are *not* assuming that the calibrator knows the true values of the parameters. If this were the case, it would be trivial to calculate *exactly* the α which generated the data. This solution could be based on an examination of a few data points. Instead, we limit the number of possible calibration exercises and hold some parameter settings fixed at their true values. While these conditions may favour calibration more than those found in practice, focussing on a few parameters clarifies our analysis. Moreover, Ingram and Lee (1987) have shown that this simulation estimator is $\sqrt{T}$-consistent and asymptotically normally distributed, subject to identification. Our results on step (3) (i) provide evidence on the small-sample properties of such ideal estimators.

The second calibrator assigns an incorrect value to θ. To be realistic we assume the assigned value is the true value $\pm$ 0.1. With an incorrect θ we should reasonably expect inconsistent estimates of α. A key question concerns the robustness of such estimates.

The third calibrator attempts to learn about α by first estimating the transition matrix (q_{ij}) which generates the time series of observed growth rates. This matrix is estimated by maximizing the log likelihood (see Anderson and Goodman (1957)). This yields estimates $\hat{q}^r_{ij}$, for each of the r =

1, ..., R histories, given by

$$\hat{q}_{ij}^{r} = \frac{\text{number of transitions of growth rates from } \lambda_i \text{ to } \lambda_j \text{ in the } r^{th} \text{ sample}}{\text{number of times the growth rate is in } \lambda_j \text{ in the } r^{th} \text{ sample.}}$$

This estimator is $\sqrt{T}$-consistent for q_{ij} and hence for large T the outcome in step (3) (iii) should mimic that in (i).

In step (4) the criteria are the sample analogues of:

$$(V) \quad \hat{\alpha}_V = \underset{\alpha \in A}{\text{argmin}} \left\{ L_V(\alpha) = \left| \text{var}(p_t^r) - \text{var}(p_t^c(\alpha)) \right| \right\}$$

$$(C) \quad \hat{\alpha}_C = \underset{\alpha \in A}{\text{argmin}} \left\{ L_C(\alpha) = \left| \text{cov}(p_t^r, p_{t-1}^r) - \text{cov}(p_t^c(\alpha), p_{t-1}^c(\alpha)) \right| \right\}$$

$$(M) \quad \hat{\alpha}_M = \underset{\alpha \in A}{\text{argmin}} \left\{ L_M(\alpha) = E\,[p_t^r - p_t^c(\alpha)]^2 \right\}$$

Criterion (V) matches the variances of the two series, criterion (C) matches the autocovariances, and criterion (M) minimizes the mean, squared difference between the two series.[8] With one endogenous variable we cannot study the cross-covariances which might also be used in calibration more generally. Note also that criterion (M) is perhaps least representative of well-known calibration studies (see Prescott 1986, page 16).

---

[8] Note that (M) does not match the mean squared innovations of the two series.

1, ..., R histories, given by

$$\hat{q}_{ij}^{r} = \frac{\text{number of transitions of growth rates from } \lambda_i \text{ to } \lambda_j \text{ in the } r^{th} \text{ sample}}{\text{number of times the growth rate is in } \lambda_j \text{ in the } r^{th} \text{ sample.}}$$

This estimator is $\sqrt{T}$-consistent for q_{ij} and hence for large T the outcome in step (3) (iii) should mimic that in (i).

In step (4) the criteria are the sample analogues of:

$$(V) \quad \hat{\alpha}_V = \underset{\alpha \in A}{\text{argmin}} \left\{ L_V(\alpha) = \left| \text{var}(p_t^r) - \text{var}(p_t^c(\alpha)) \right| \right\}$$

$$(C) \quad \hat{\alpha}_C = \underset{\alpha \in A}{\text{argmin}} \left\{ L_C(\alpha) = \left| \text{cov}(p_t^r, p_{t-1}^r) - \text{cov}(p_t^c(\alpha), p_{t-1}^c(\alpha)) \right| \right\}$$

$$(M) \quad \hat{\alpha}_M = \underset{\alpha \in A}{\text{argmin}} \left\{ L_M(\alpha) = E\,[p_t^r - p_t^c(\alpha)]^2 \right\}$$

Criterion (V) matches the variances of the two series, criterion (C) matches the autocovariances, and criterion (M) minimizes the mean, squared difference between the two series.[8] With one endogenous variable we cannot study the cross-covariances which might also be used in calibration more generally. Note also that criterion (M) is perhaps least representative of well-known calibration studies (see Prescott 1986, page 16).

---

[8] Note that (M) does not match the mean squared innovations of the two series.

An alternative calibration criterion would be to match autocorrelations of the series (as in Greenwood, Hercowitz, and Huffman, 1987) . Using equations (13) and (14) the population autocorrelation coefficient in our model economy is simply θ. Thus the autocorrelation in $\{p_t\}$ is independent of α so that adopting this criterion in a calibration exercise would provide no quantitative discipline. Nevertheless, despite the fact that α cannot be identified from the autocorrelation coefficient, the calibration procedure would select some α that most closely matched the sample autocorrelation coefficient. Unfortunately, the density of such an estimator would be *uniform* over the grid interval [0,16] for the R histories; hence for the calibration exercise any α could be justified. Although more general parameterizations of the transition matrix could imply identification of α from the autocorrelation coefficient this example clearly illustrates the potential drawbacks of blindly 'matching' some properties of artificial data to historical data in order to estimate α.

Finally, we consider two generalized method of moments (GMM) estimators over the same R histories and replace steps (3) and (4) by the following:

$$(G1) \quad \hat{\alpha}_{G1} = \underset{\alpha \in A}{\text{argmin}} \left\{ L_{G1}(\alpha) = \left[E\left(p_{t+1} - \beta\, x_{t+1}^{-\alpha}\right) \right]^2 \right\}$$

$$(G2) \quad \hat{\alpha}_{G2} = \underset{\alpha \in A}{\text{argmin}} \left\{ L_{G2}(\alpha) = \left[E\left[z_t \left(p_{t+1} - \beta\, x_{t+1}^{-\alpha}\right)\right] \right]^2 \right\} ; \; z_t = (1 \; x_t)$$

where for these experiments we set β equal to 0.99.

The estimator $\hat{\alpha}_{G1}$ is based on an orthogonality condition arising from equation (6). The orthogonality conditions are obtained by noting that $p_t = \beta E_t x_{t+1}^{-\alpha}$ from equation (6). Then $p_t - \beta x_{t+1}^{-\alpha} = \nu_t$ where $\nu_t = x_{t+1}^{-\alpha} - E_t x_{t+1}^{-\alpha}$. Therefore $E_t \nu_t = 0$ and the two GMM estimators follow directly. The estimator $\hat{\alpha}_{G2}$ is an instrumental variables estimator with the current growth rate x_t serving as an instrument. This method is adopted by Hansen and Singleton (1982) and Tauchen (1986).[9] Further lags of x also could be used. One common occurrence in GMM estimation is that theory may suggest a number of different restrictions which may yield different estimates (as in Grossman, Melino, and Shiller (1987)). In calibration the choice of moment(s) to match seems arbitrary. Our GMM estimates are obtained by grid search over the same interval as is used for the calibration estimates above. A subsidiary aim of this study is to investigate the small-sample properties of two simple GMM estimators for data generating processes of economic interest arising from an optimization problem.

It is worthwhile to compare the GMM and calibration estimators. The most obvious distinction between the two is based upon their differring informational demands. The GMM estimators seek α simply by orthogonality restrictions directly based on the theory. To identify α in (G1) we require only the value of the discount rate β, about which we have precise prior beliefs. When the instrument set is enlarged as in (G2) we may in principle estimate both α and β. Moreover, the use of further instruments leads to

---

[9] One potentially could gain efficiency by taking into account heteroskedasticity in ν_t as in Hansen (1982).

direct tests of the model through over-identifying restrictions (see Hansen, 1982).

In contrast, the calibration approach requires a complete parameterization of the model. It involves assigning values to all parameters of the model including those describing the evolution of the state variable x_t. As the number of free parameters increases the dimension of the grid search increases as does the computational cost of the procedure. It is natural to ask whether all of the parameters can be identified by matching a small number of moments. Regrettably, the procedure selects parameter values regardless of their identifiability. Our results are of interest precisely because we examine the case with one free parameter, the case in which identification should be easiest.

A more subtle distinction between the GMM and calibration estimators concerns their replication. Note that two calibrators presented with the same historical data series, using identical models, and adopting identical loss functions in principle could arrive at entirely different estimates (we shall provide a detailed example below in section 4).[10] Since the calibration estimates do not come with standard errors (although in principle these could be found by replication) but are themselves realizations of random variables, it is not immediately clear how much confidence to attach to any estimate

---

[10] In this sense the calibration estimators resemble bootstrap estimators. We could reasonably expect that as the number of calibrators becomes large the average estimate for fixed T converges to some (not necessarily true) value.

even if the model were true. In contrast, the GMM estimates are unique given the data set (subject to some identification conditions) and have a well-articulated asymptotic theory, although many finite-sample issues are unresolved.

4. MONTE CARLO EVIDENCE.

Many possible Monte Carlo experiments could be conducted within the model outlined above. Our intent here is simply to provide some insight into the issues and problems that arise when applying the calibration and GMM estimators. We examine the properties of these estimators as we vary (a) sample size ($T = 100, 500$), (b) the relative risk aversion parameter ($\alpha = 2, 4$), and (c) the degree of persistence in the dividend process ($\theta = 0.2, 0.8$). The number of experiments is limited, but we believe that the results convey an accurate impression of the sensitivities of the estimators with respect to changes in the underlying economic environment.

The probability densities are estimated using a quartic kernel and a Parzen window as discussed by Silverman (1986, page 45).[11] In all experiments we set the number of histories, R, equal to 1000. The state variables for generating the R histories and the calibration simulations are generated from

---

[11] We adopt a quartic kernel since the support of all the estimators we consider is finite, obtained from a fixed grid search (see Tapia and Thompson, 1978). Following Silverman, the kernel estimator uses a variable window width given by: $0.698 \min[s, IQR/1.34]$ where s is the sample standard deviation and IQR the inter-quartile range of the estimates over the R histories.

a uniform density between 0 and 1 using a pseudo-random number generator (G05) in the NAG subroutines. States are defined on this interval according to the appropriate transition probabilities. Thus for the R histories (step (1)) the transition probability in equation (11) is employed. The calibration simulations in steps (3)i and (3)ii also use (11) while that in step (3)iii defines the states based on the estimated transition probabilities. In all cases the first observation is determined from states defined from the true equilibrium density.

The R histories of data, p_t^r (r = 1, ..., R), are generated from a preassigned seed so that we may produce identical histories for the different calibration and GMM estimators. Thus the calibrations in 3i, ii, and iii and the two GMM estimators are based upon the same histories. Thus the differences between these estimators are due solely to their differring calibration criteria or orthogonality conditions and *not* to sampling variation. Note that the artificial data used in calibration, p_t^c, *are* generated from an non-replicable seed, a procedure which mimics that used in actual calibration.

Figure 1 shows estimated densities for the calibrated estimators when $\alpha = 4$ and $\theta = 0.2$. In all figures the first row applies to T = 100 and the second row to T = 500. The densities are labelled with the letters corresponding to the loss functions used in calculating them. Figure 2 shows the two GMM estimators (the two are virtually indistinguishable) under the same data generation process. Figures 3 and 4 present the estimated densities under the same conditions except that $\theta = 0.8$.

We shall first discuss results for $\alpha = 4$ and later consider their generality. Figures 1 and 3 illustrate cases in which all parameters but α are set at their true values, so that the parameter vectors in steps (1) and (3) coincide. For these cases, Ingram and Lee show that estimators such as (V) and (C) are $\sqrt{T}$-consistent and asymptotically normally distributed. Our concern is to establish their finite-sample properties. Estimator V (matching variances) gives excellent results with the estimated density centred on the true value of α with a relatively small variance. As sample size increases the density of the V estimator converges rapidly. Tables 1 and 2 record the average biases (B), mean-squared errors (M_s), and approximate 95 per cent confidence intervals (0.95) for all experiments.[12] From table 1, for estimator V with $\alpha = 4$, $T = 100$, and $\theta = 0.2$ the average bias is only -0.02 with an approximate 95 per cent confidence interval of [3.3,5.0]. This interval is compressed to [3.7,4.4] at a sample size of 500. When $\theta = 0.8$ so that the data are more serially correlated the same general pattern for the V estimator emerges except that the estimator is much less efficient. For example, the 95 per cent confidence interval for $\theta = 0.8$ and $T = 100$ is [2.3,7.6] with an average bias of -1.3.

In contrast, estimator C (matching autocovariances) performs poorly for low values of θ and yields reasonably precise estimates of α only when $\theta = 0.8$ and $T = 500$. The confidence interval for $\theta = 0.2$ and $T =$

---

[12] Since the density itself is estimated the true 95 per cent confidence interval would be wider than shown. We ignore this added complication and present an approximate interval.

100 is wide: [1.2,13.7].

Perhaps the poorest calibration estimator is that based on matching mean squared errors (M). Its density is more concentrated than that of estimator C (but less so than V) but it converges to the true α very slowly. On average it underestimates α; with an econometric sample size of 100 the average bias is 0.9 for $\theta = 0.2$ and 2.11 for $\theta = 0.8$. Particularly with high θ the actual data have persistent cycles so that this estimator will minimize mean-squared error by choosing an α which generates cycles of lower amplitude. To be consistent with this smoothing, the estimated α will be less than the true α. While the M estimator is consistent we have found that there can be substantial bias with as many as 5000 observations.

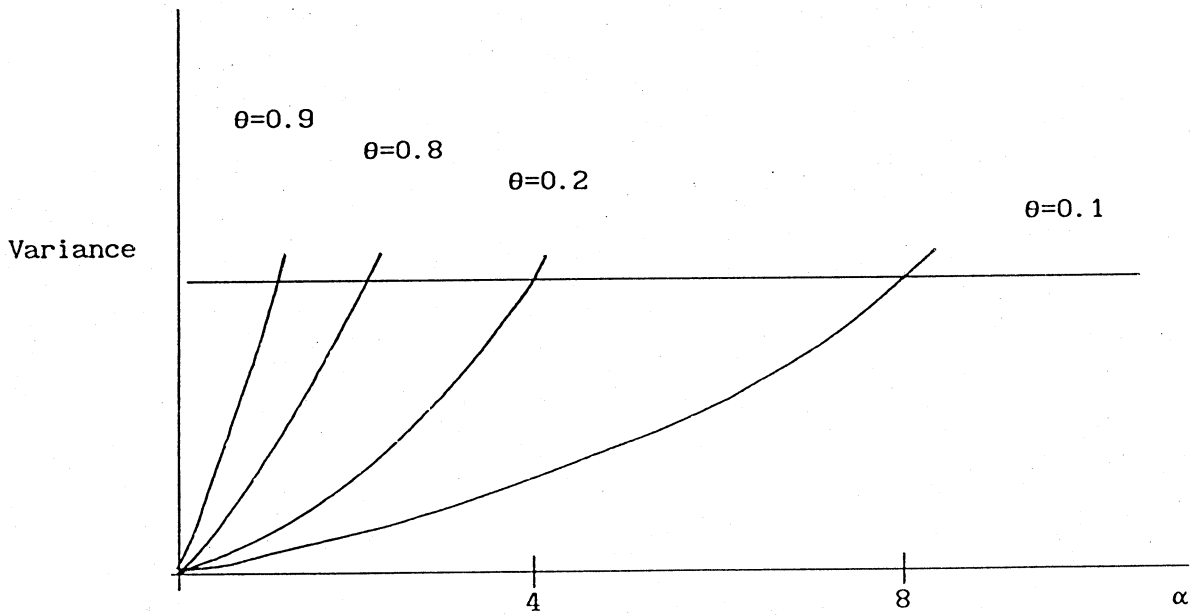
Figures 3 and 4 illustrate the GMM estimators. These are virtually identical, except at $\theta = 0.8$ and T = 500. The estimated densities of these estimators are centred at the true value with small average biases. A comparison of tables 1 and 2 shows that the bias increases with the value of α. This finding accords with those of Tauchen (1986), who considers a DGP similar to ours in which α takes on the values 0.30 and 1.30. The bias also declines with more persistence in the state variable. The confidence intervals for T = 100 are fairly wide and become narrower as θ rises, as might be expected for these estimators. Convergence is relatively rapid; the confidence intervals narrow by 50 per cent across the sample sizes we consider.

In the experiments discussed so far the V estimator outperforms the other calibration estimators and the GMM ones. But this result is misleading

since the calibration estimators are conditioned on correct settings of the other parameters (as in step 3i). Next, we deviate from these ideal conditions and examine steps 3ii and 3iii. In step 3ii an incorrect value of θ is assigned, given by the true value $\pm$ 0.1. For $\alpha = 4$ figure 5 gives the density estimates for the three calibration estimators in the case in which the true $\theta = 0.2$ but the calibrator assumes that $\theta = 0.1$. With this miscalculation the densities of the V and C estimators are translated to the right and centred roughly at $\alpha = 8$. Moreover for large T the densities converge to roughly twice the true α. In figure 7 the true θ is 0.8 but the calibrator now overestimates the persistence parameter and selects a value of 0.9. In this case the effects of the misspecification are much less severe, with a slight underestimate.

Figure 10 gives a numerical example based on the analysis of section 3, which rationalizes some of these findings. The calculations are based on equation (13) above, with the same state values and equilibrium probabilities as in our experiments. Population variances are measured on the vertical axis and the α's deduced from them are given on the horizontal axis. These relations are drawn for various persistence parameters. The solid lines denote variances. The evaluations are based on the equilibrium probabilities and state values given above and the horizontal line gives the variance for $\alpha = 4$ and $\theta = 0.2$.

Figure 10



Clearly the effect of small mistakes in setting θ depends on the data generation process. If θ is small then mistakes in selecting its value will lead to large errors in estimates of α based on matching variances and even larger errors based on matching covariances (recall from equation (14) that $C = \theta V$). Conversely, for large values of θ mistakes in its setting will not lead to large errors in conditional estimates of α.

These results show that there is a trade-off between efficiency and robustness inherent in these calibration estimators.[13] They also prompt us to examine step 3iii in which the transition matrix is estimated from the

---

[13] We have also examined small errors in setting β; these have little effect, as our analytical results would also predict.

history p_t^r. Figure 6 illustrates the case in which $\theta = 0.2$ and figure 8 that in which $\theta = 0.8$. These may be compared with figures 1 and 3 respectively, in which the first stage of calibration is correct. Estimating, as opposed to correctly setting, the transition probabilites gives rise to wider confidence intervals. The discrepancies are smaller the larger the sample size and the larger the value of θ, as one would expect.

While the discussion so far has concerned a true α of 4, our qualitative findings do not depend on this value. A complete set of experimental results for an α of 2 is provided in table 2.

Finally, figure 9 illustrates the distinction between calibration and more conventional estimators mentioned at the end of section 4 above. This distinction is based on the potential irreplicability of calibration estimates. To illustrate it we repeat steps (1) to (5) above except that in step (2) R = 1 so that there is only *one* history. We conduct step (3i) so that the model and DGP coincide. Then in step (4) we model three groups of calibrators with one thousand members in each group. The groups are identified with the three calibration loss functions. Figure 5 gives the results for $\alpha = 4$, $\theta = 0.8$, and an econometric sample size of 100. Note that the densities are those of *estimates* rather than *estimators* since there is only one history. Note also that averaging findings within a group would not lead its members collectively to converge to the true α.

One way to eliminate the variability in estimates based on common data and models is to derive estimates from population moment conditions. Specifically, each of equations (13) and (14) may be used with its observed

sample counterpart for the historical data to solve for α. (Of course, the estimates need not be the same.) Greenwood, Hercowitz, and Huffman (1987) have adopted this strategy in estimating their free parameters. The quality of these estimators again will depend on settings of other parameters. However, if the initial parameter settings are correct (as in step 3i above) then these estimators will be more precise than those which simply match the sample moments of artificial and historical data. We consider such an estimator based on solving for α from the population moments for V and C under step 3i, with all other parameters set at their true values. In this experiment we set $\alpha = 4$, $\theta = 0.8$, and $T = 100$. The estimates from V have an average bias of -0.36 and a 95 per cent confidence interval of 3.0 to 4.7. The corresponding measures for estimates from C are -0.28 and 2.7 to 4.8 respectively. Compared to the earlier results reported in Table 1 for the same setting (where the confidence intervals for V and C are 2.3 to 7.6 and 2.1 to 9.2 respectively) there is a dramatic gain in precision. Clearly estimates based upon population moments can improve efficiency.

5. SUMMARY.

The aim of this paper is to evaluate calibration as an estimation strategy. This evaluation takes place in the Mehra-Prescott model, since this model is simple enough to allow both analytical results which help interpret the simulation evidence and the application of more familiar estimation methods. While our results are specific to this model, we believe that the properties of calibration as estimation discussed here apply to more

general, dynamic, optimizing models.

Important features of a dynamic model often may be revealed by simulating the model over a variety of parameter values. Having simulated the model and generated artificial data it is natural and useful to to compare these with actual time series data and to compare the moments and patterns (turning points for example) of the two series. One can regard the parameters which yield the best match in such comparisons as estimates of the true parameters. Our results show that these estimates are very sensitive to choices of other parameters made in the first stage of calibration. They can be highly misleading if the first stage involves misspecification.

Our Monte Carlo evidence suggests that generalized method-of-moments estimators are the most reliable (though they may lack precision) and converge quite rapidly to the true parameter. Notwithstanding this observation, there are many dynamic equilibrium models for which GMM estimation is impractical or computationally demanding. In these cases in which GMM is not feasible, we offer the following suggestions on calibration which may improve applications of this method.

1. A key difficulty with calibration is that parameters are selected regardless of their identifiability. Therefore, where possible, analytical expressions for the moments (or other properties matched in calibration) should be derived so that potential singularities may be identified. These expressions should suggest which parameter estimates are likely to be sensitive to the choice of moment in calibration. They may also show which moment condition is the most appropriate.

2. When analytical expressions for population moments cannot be found. We recommend that the model be simulated many times (with large sample sizes) to achieve convergence, perhaps by calculating the mean of the estimates across replications.

3. Parameters describing the evolution of the state variable should be consistently estimated whenever possible. Following this suggestion leads to a noisy calibration estimate but a consistent one.

4. Obtaining close matches between simulated and actual moments may not be compelling evidence in favour of a model. Our evidence suggests very close matching may yield misleading estimates. Calibrators should provide evidence on the sensitivity of the moments in the artificial economy to alternative settings of the model (much like describing likelihood functions).

REFERENCES

Altuğ, S. (1985) Estimation and Tests of an Aggregate Equilibrium Model. unpublished Federal Reserve Bank of Minneapolis working paper.

Anderson, T.W. and Goodman, L.A. (1957) Statistical Inference About Markov Chains. *Annals of Mathematical Statistics* 28, 89-110.

Backus, D.K., Gregory, A.W., and Zin, S.E. (1986) Risk Premiums in the Term Structure: Evidence from Artificial Economies. Queen's University Institute for Economic Research Discussion Paper 665.

Barton, E., David, F., and Fix, E. (1962) Persistence in a Chain of Multiple Events when there is Simple Persistence. *Biometrika* 49, 351-357.

Eichenbaum, M., Hansen, L. and Singleton, K. (1987) A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice Under Uncertainty. *Quarterly Journal of Economics* forthcoming.

Greenwood, J., Hercowitz, Z., and Huffman, G.W. (1987) Investment, Capacity Utilization and the Real Business Cycle. mimeo, University of Western Ontario.

Grossman, S. and Shiller, R.J. (1981) The Determinants of the Variability of Stock Market Prices. *American Economic Review* 71, 222-227.

Grossman, S. and Weiss, L. (1983) A Transactions-Based Model of the Monetary Transmission Mechanism. *American Economic Review* 73, 871-880.

Grossman, S., Melino, A., and Shiller, R.J. (1987) Estimating the Continuous-Time Consumption-Based Asset Pricing Model. *Journal of Business and Economic Statistics* 5, 315-327.

Hansen, G. (1985) Indivisible Labor and the Business Cycle. *Journal of Monetary Economics* 16, 309-327.

Hansen, L. (1982) Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50, 1029-1054.

Hansen, L. and Singleton, K. (1983) Stochastic Consumption, Risk Aversion, and the Temporal Behaviour of Asset Returns. *Journal of Political Economy* 91, 249-265.

Ingram, B.F. and Lee, B.-S. (1987) Estimation by Simulation. mimeo, Cornell University.

Kydland, F.E. and Prescott, E.C. (1982) Time to Build and Aggregate Fluctuations. *Econometrica* 50, 1345-1370.

Kydland, F.E. (1987) The Role of Money in a Competitive Theory of Business Fluctuations. mimeo. Carnegie-Mellon University.

Labadie, P. (1984) A Test of Risk Premia Behavior in an Overlapping Generations Model. mimeo, Graduate School of Business, Columbia University.

Leamer, E.E. (1983) Let's Take the Con out of Econometrics. *American Economic Review* 73, 31-43.

Long, J.B. and Plosser, C.I. (1983) Real Business Cycles. *Journal of Political Economy* 91, 39-69.

Lucas, R.E. (1987) *Models of Business Cycles.* Yrjö Jahnsson Lectures. Oxford: Basil Blackwell.

Manuelli, R.E. (1986) Modern Business Cycle Analysis: A Guide to the Prescott-Summers Debate. *Federal Reserve Bank of Minneapolis Quarterly Review.* (Fall) 3-8.

McAleer, M., Pagan, A.R., and Volker, P.A. (1985) What Will Take the Con out of Econometrics? *American Economic Review* 75, 293-306.

Mehra, R and Prescott, E.C. (1985) The Equity Premium: A Puzzle. *Journal of Monetary Economics* 15, 145-161.

Pakes, A. and Pollard, D. (1986) The Asymptotics of Simulation Estimators. University of Wisconsin-Madison SSRI Working Paper 8628.

Prescott, E.C. (1986) Theory Ahead of Business Cycle Measurement. *Federal Reserve Bank of Minneapolis Quarterly Review.* (Fall) 9-22, or Carnegie-Rochester Conference Series on Public Policy 25, 11-44.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Singleton, K. (1987) Econometric Issues in the Analysis of Equilibrium Business Cycle Models. mimeo.

Slutsky, E. (1937) The Summation of Random Causes as the Source of Cyclic Processes. *Econometrica* 5, 105-146.

Tapia, R.A. and Thompson, J.R. (1978) *Nonparametric Probability Density Estimation.* Baltimore: The Johns Hopkins University Press.

Taylor, J.B. (1980) Aggregate Dynamics and Staggered Contracts. *Journal of Political Economy* 88, 1-23.

Tauchen, G. (1986) Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained From Financial Market Data. *Journal of Business and Economic Statistics* 4, 397-416

TABLE 1 $\alpha = 4$

Average Biases, Mean-squared Errors, and Approximate 95% Confidence Intervals

p_t^r	p_t^c	T	V		C		M		G_1		G_2	
			B/M_s	0.95	B/M_s	0.95	B/M_s	0.95	B/M_s	0.95	B/M_s	0.95
θ=0.2		100	-0.02	3.3	-1.0	1.2	0.9	2.2	-0.6	2.4	-0.6	2.4
	θ=0.2		0.14	5.0	8.8	13.7	1.1	4.1	4.9	12.0	4.9	11.9
		500	0.002	3.7	-0.17	2.6	1.03	2.3	-0.08	3.1	-0.08	3.1
			0.03	4.4	0.85	6.1	1.17	3.7	0.41	5.8	0.40	5.8
		100	-4.10	6.7	-4.59	2.6	0.24	2.9				
	θ=0.1		17.33	9.7	31.32	15.3	0.28	4.8				
		500	-4.04	7.4	-4.35	5.4	0.31	3.1				
			16.43	8.8	21.69	12.2	0.18	4.4				
		100	1.32	2.2	0.42	0.8	1.60	1.5				
	θ=0.3		1.81	3.4	5.85	12.1	2.77	3.4				
		500	1.34	2.4	1.22	1.8	1.72	1.7				
			1.81	3.5	1.89	4.2	3.05	3.0				
		100	-0.62	1.6	-0.15	1.6	1.06	1.5				
	$\hat{q}_{ij}$		8.01	13.9	9.94	14.0	1.93	5.2				
		500	-0.21	2.5	-0.23	2.9	1.06	2.0				
			1.54	7.2	0.87	6.2	1.37	4.0				
θ=0.8		100	-0.30	2.3	0.50	2.1	2.11	0.3	0.13	0.2	0.12	0.2
	θ=0.8		1.80	7.6	3.32	9.2	5.54	4.7	2.86	10.9	2.82	10.3
		500	-0.02	3.3	-0.02	3.1	2.86	0.3	-0.22	3.5	-0.21	3.5
			0.18	5.1	0.24	5.1	8.49	2.4	1.35	8.4	1.22	8.0
		100	-0.91	2.6	-1.00	2.4	1.84	0.3				
	θ=0.7		3.03	8.8	3.97	9.7	4.78	5.2				
		500	-0.62	3.7	-0.64	3.6	2.65	0.3				
			0.61	5.7	0.72	5.9	7.38	2.8				
		100	0.15	2.0	0.01	1.9	2.32	0.2				
	θ=0.9		1.31	6.7	2.23	7.4	6.26	4.3				
		500	0.42	2.9	0.42	2.8	3.02	0.2				
			0.32	4.5	0.35	4.6	9.36	2.1				
		100	-0.64	2.6	-0.75	2.5	1.96	0.3				
	$\hat{q}_{ij}$		2.94	8.7	3.65	9.8	5.35	4.7				
		500	-0.05	3.3	-0.08	3.2	2.84	0.2				
			0.16	5.1	0.20	5.1	8.39	2.5				

The first column gives the DGP while the second column gives the calibrated model. In each section the first row corresponds to step 3i, the second and third to step 3ii, and the third to step 3iii. For each estimator and sample size there are two columns of entries. The first gives the average bias (B) and mean squared error (M_s); the second gives the lower and upper boundaries of the approximate 95% confidence interval (0.95).

TABLE 2 $\alpha = 2$

Average Biases, Mean-squared Errors, and Approximate 95% Confidence Intervals

p_t^r	p_t^c	T	V B/M_s	V 0.95	C B/M_s	C 0.95	M B/M_s	M 0.95	G_1 B/M_s	G_1 0.95	G_2 B/M_s	G_2 0.95
θ=0.2		100	-0.004	1.7	-0.81	0.5	0.46	1.1	-0.33	1.2	-0.31	1.2
	θ=0.2		0.03	2.5	5.54	10.9	0.27	2.0	1.83	5.9	1.50	5.9
		500	0.002	1.5	-0.08	1.3	0.49	1.2	-0.03	1.6	-0.03	1.6
			0.007	2.5	0.21	3.2	0.27	1.8	0.09	2.9	0.09	2.9
		100	-2.04	3.3	-3.04	1.3	0.13	1.4				
	θ=0.1		4.29	5.0	16.87	3.5	0.08	2.5				
		500	-2.01	3.8	-2.18	2.6	0.15	1.6				
			4.06	4.2	5.57	6.1	0.05	2.1				
		100	0.67	1.2	-0.04	0.3	0.80	0.7				
	θ=0.3		0.47	1.7	4.5	12.2	0.71	1.7				
		500	0.65	1.2	0.61	0.9	0.82	0.8				
			0.44	1.5	0.49	2.3	0.70	1.4				
		100	-0.39	0.8	-0.93	0.7	0.51	0.7				
	$\hat{q}_{ij}$		3.36	7.4	6.96	12.0	0.47	2.7				
		500	-0.1	1.2	-0.11	1.4	0.50	1.0				
			0.40	3.7	0.26	3.2	0.31	2.1				
θ=0.8		100	-0.12	1.2	-0.23	1.0	1.14	0.1	-0.02	0	-0.02	0
	θ=0.8		0.38	3.7	0.84	4.6	1.61	3.5	1.48	5.9	1.49	5.9
		500	-0.001	1.6	-0.006	1.5	1.48	0.2	-0.11	1.8	-0.10	1.8
			0.05	2.6	0.06	2.6	2.28	1.2	0.56	4.3	0.38	4.3
		100	-0.45	1.3	-0.57	1.2	1.03	0.1				
	θ=0.7		0.83	4.6	1.37	5.2	1.42	3.2				
		500	-0.29	1.8	-0.30	1.8	1.38	0.1				
			0.14	2.6	0.16	2.7	2.01	1.5				
		100	0.1	1.0	0.02	0.9	1.25	0.0				
	θ=0.9		0.57	3.3	0.91	3.8	1.82	2.4				
		500	0.22	1.4	0.21	1.4	1.56	0.0				
			0.08	2.2	0.09	2.3	2.50	1.0				
		100	-0.32	1.2	0.42	1.2	1.04	0.1				
	$\hat{q}_{ij}$		1.07	4.2	1.63	5.2	1.96	4.0				
		500	-0.02	1.6	-0.02	1.6	1.47	0.1				
			0.04	2.5	0.05	2.6	2.25	1.2				

The first column gives the DGP while the second column gives the calibrated model. In each section the first row corresponds to step 3i, the second and third to step 3ii, and the third to step 3iii. For each estimator and sample size there are two columns of entries. The first gives the average bias (B) and mean squared error (M_s); the second gives the lower and upper boundaries of the approximate 95% confidence interval (0.95).

KEY TO FIGURES 1 TO 8

Estimated densities of $\hat{\alpha}$ are labelled according to the loss function used in step 4:

V - matches variances

C - matches autocovariances

M - minimizes mean, squared error of differences between series

G - unconditional GMM (G1)

Ψ - conditional GMM (G2)

Subtitles in the figures give the α, θ, and sample size T in the DGP of step 1. Subtitles also describe steps 3ii and 3iii in which θ is set with error and estimated respectively.

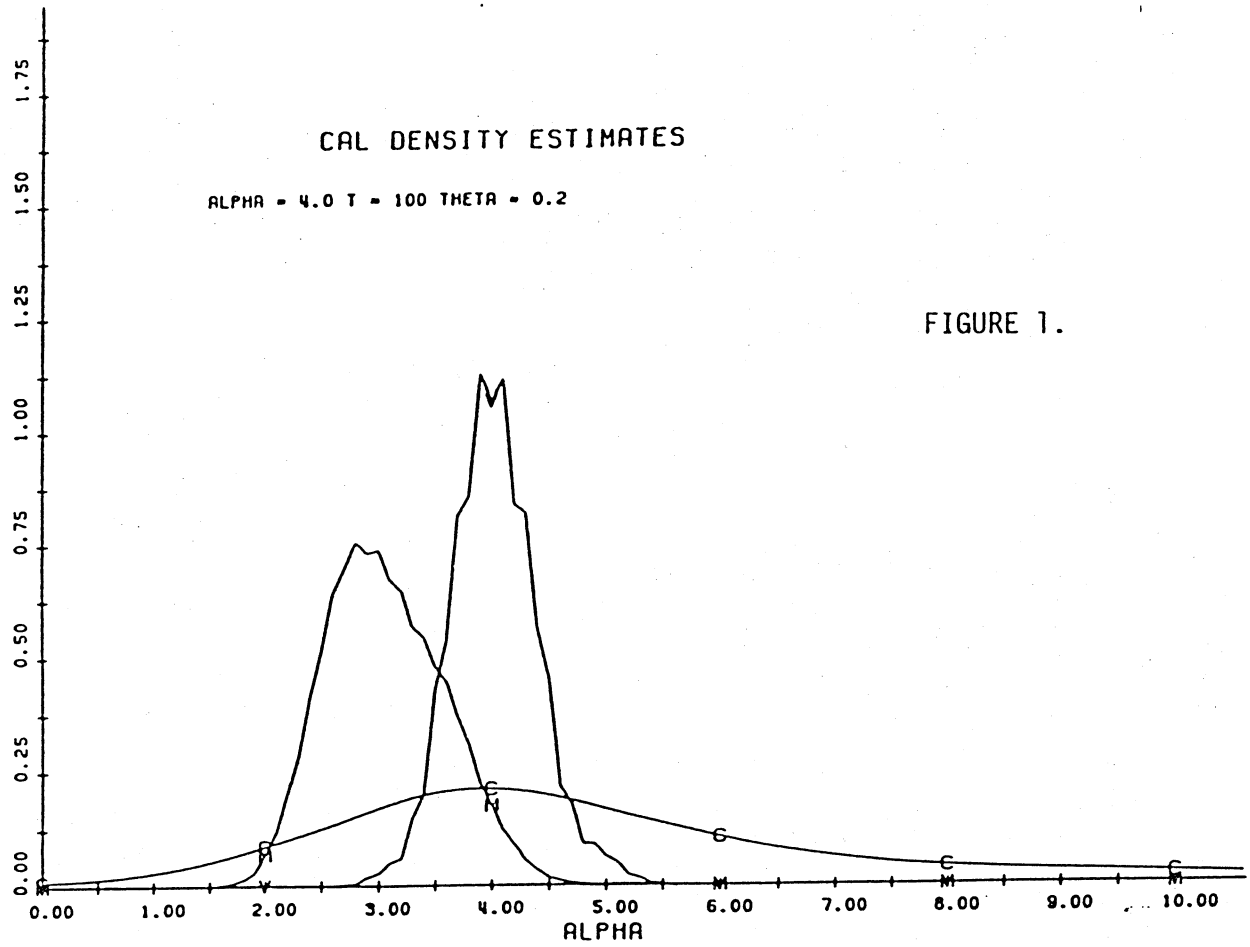
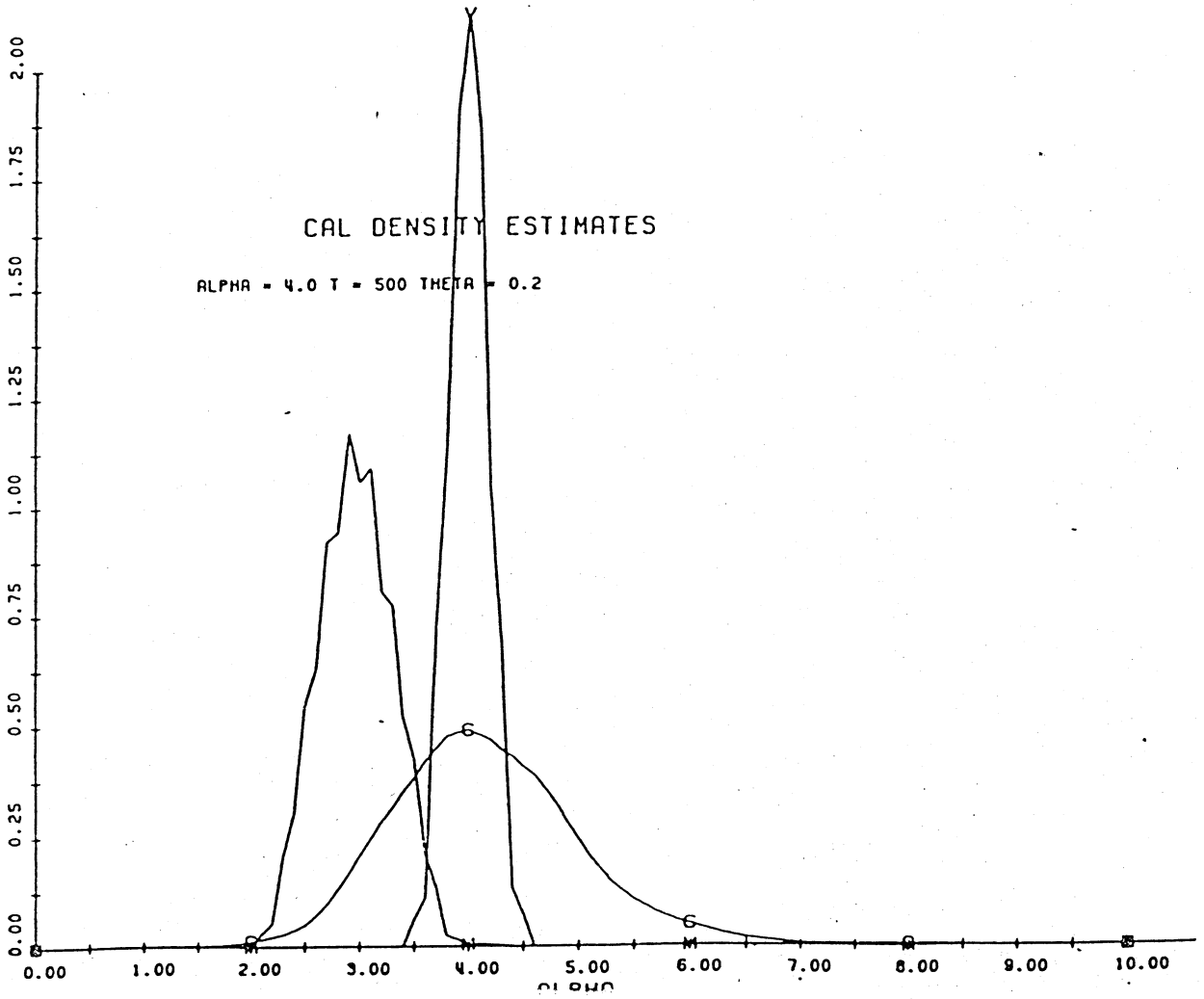

CAL DENSITY ESTIMATES
ALPHA = 4.0 T = 100 THETA = 0.2
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
ALPHA

FIGURE 1.


CAL DENSITY ESTIMATES
ALPHA = 4.0 T = 500 THETA = 0.2
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

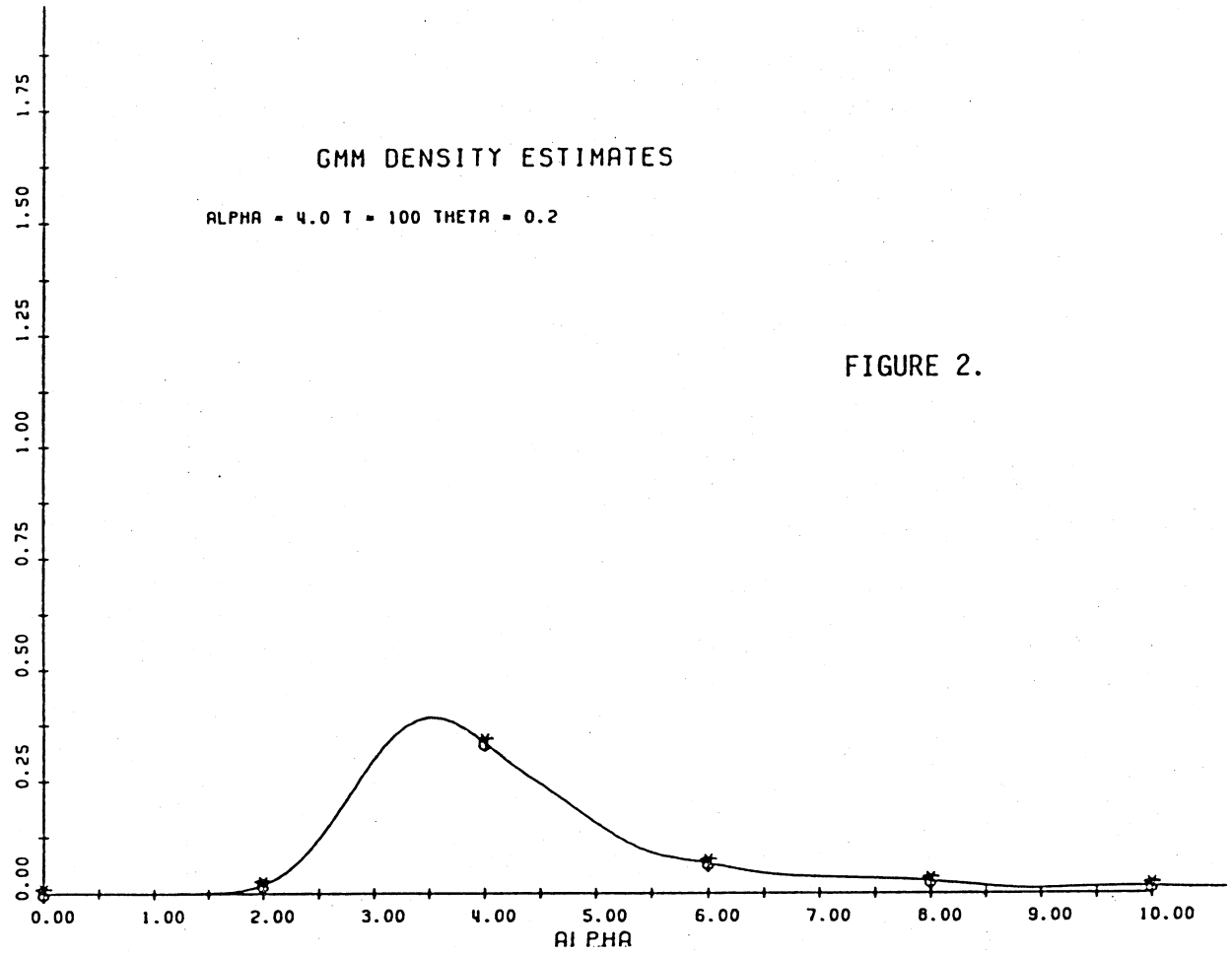
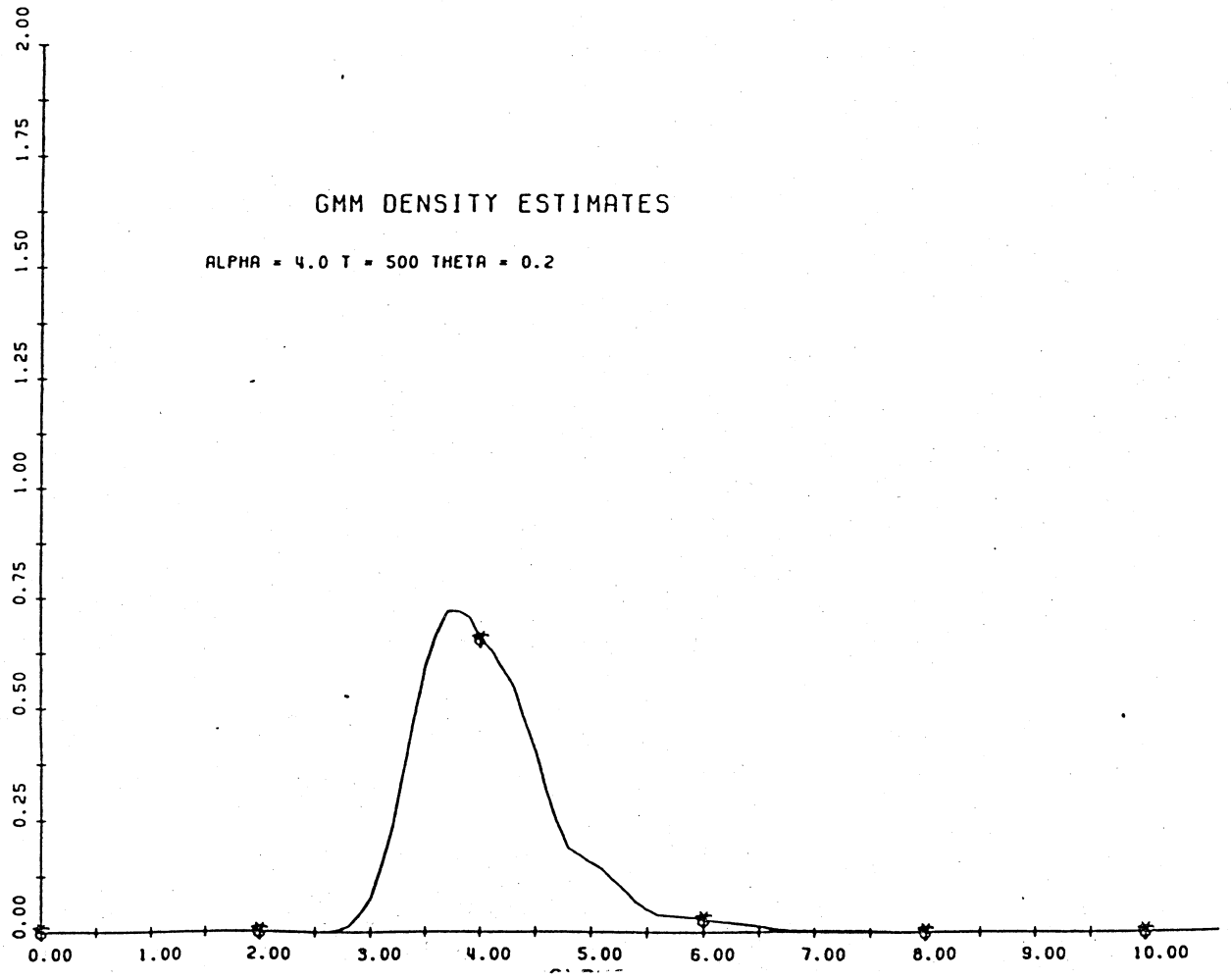

GMM DENSITY ESTIMATES
ALPHA = 4.0 T = 100 THETA = 0.2
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
ALPHA

FIGURE 2.


GMM DENSITY ESTIMATES
ALPHA = 4.0 T = 500 THETA = 0.2
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

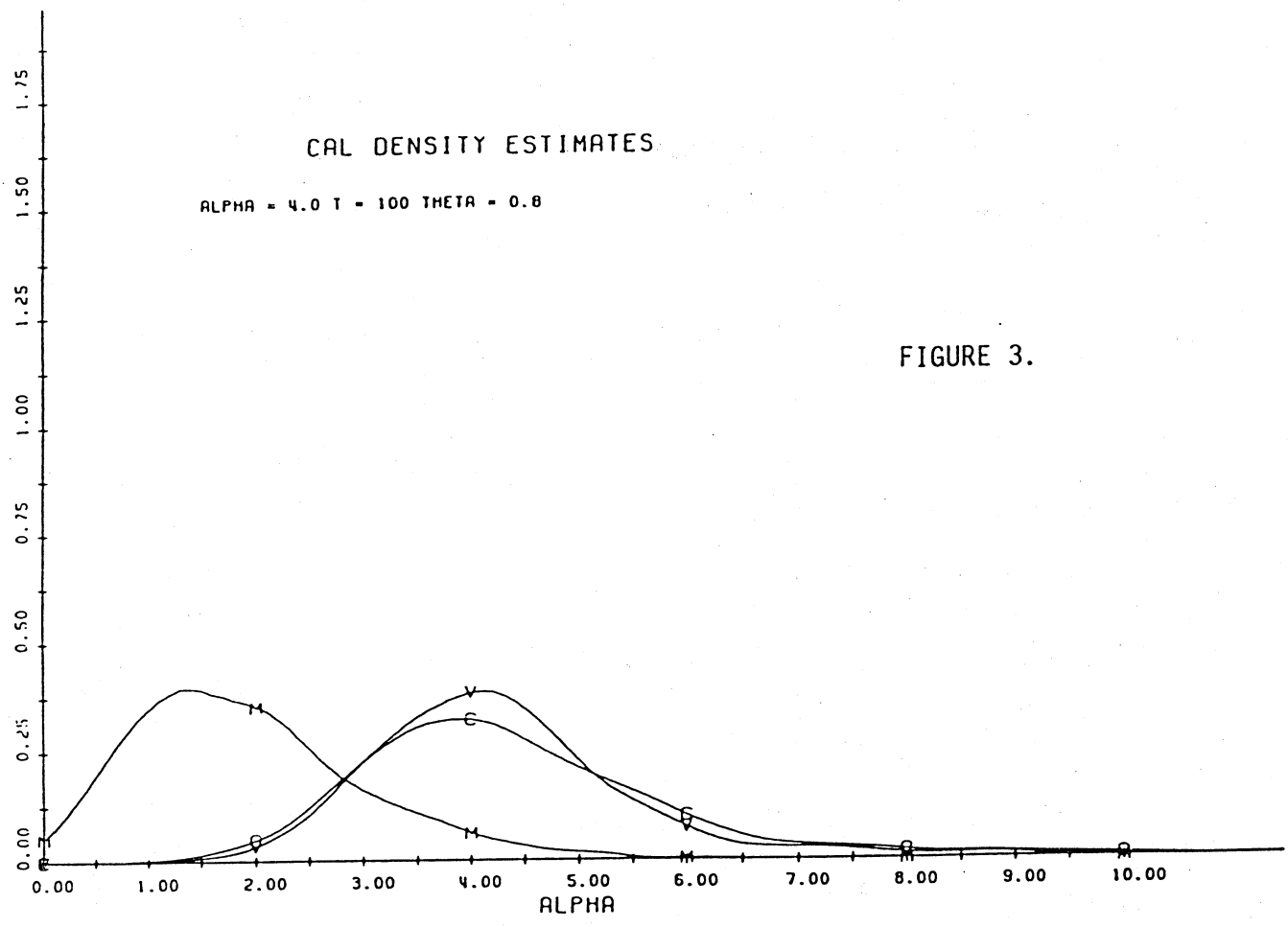
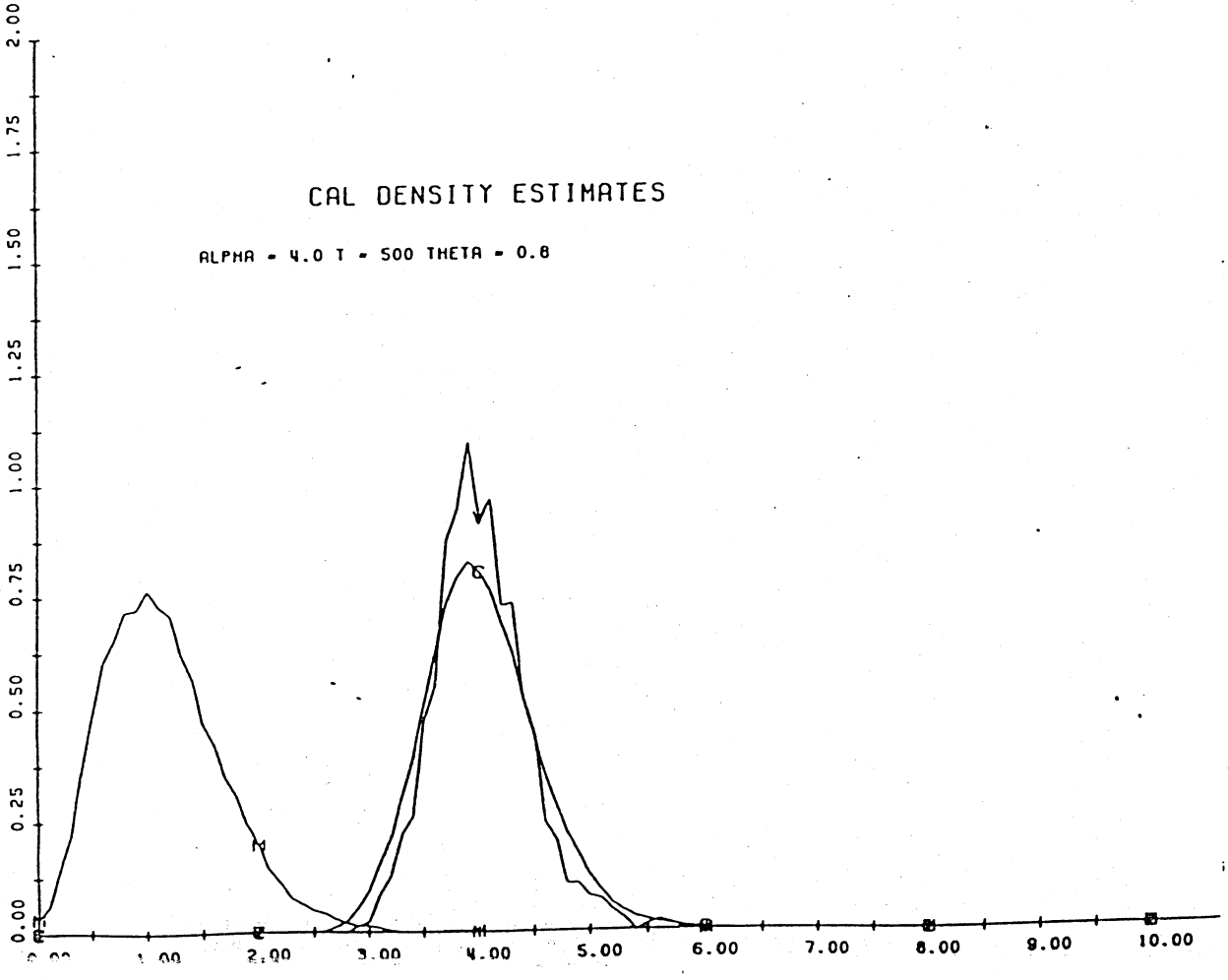

CAL DENSITY ESTIMATES
ALPHA = 4.0 T = 100 THETA = 0.8
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
ALPHA

FIGURE 3.


CAL DENSITY ESTIMATES
ALPHA = 4.0 T = 500 THETA = 0.8
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

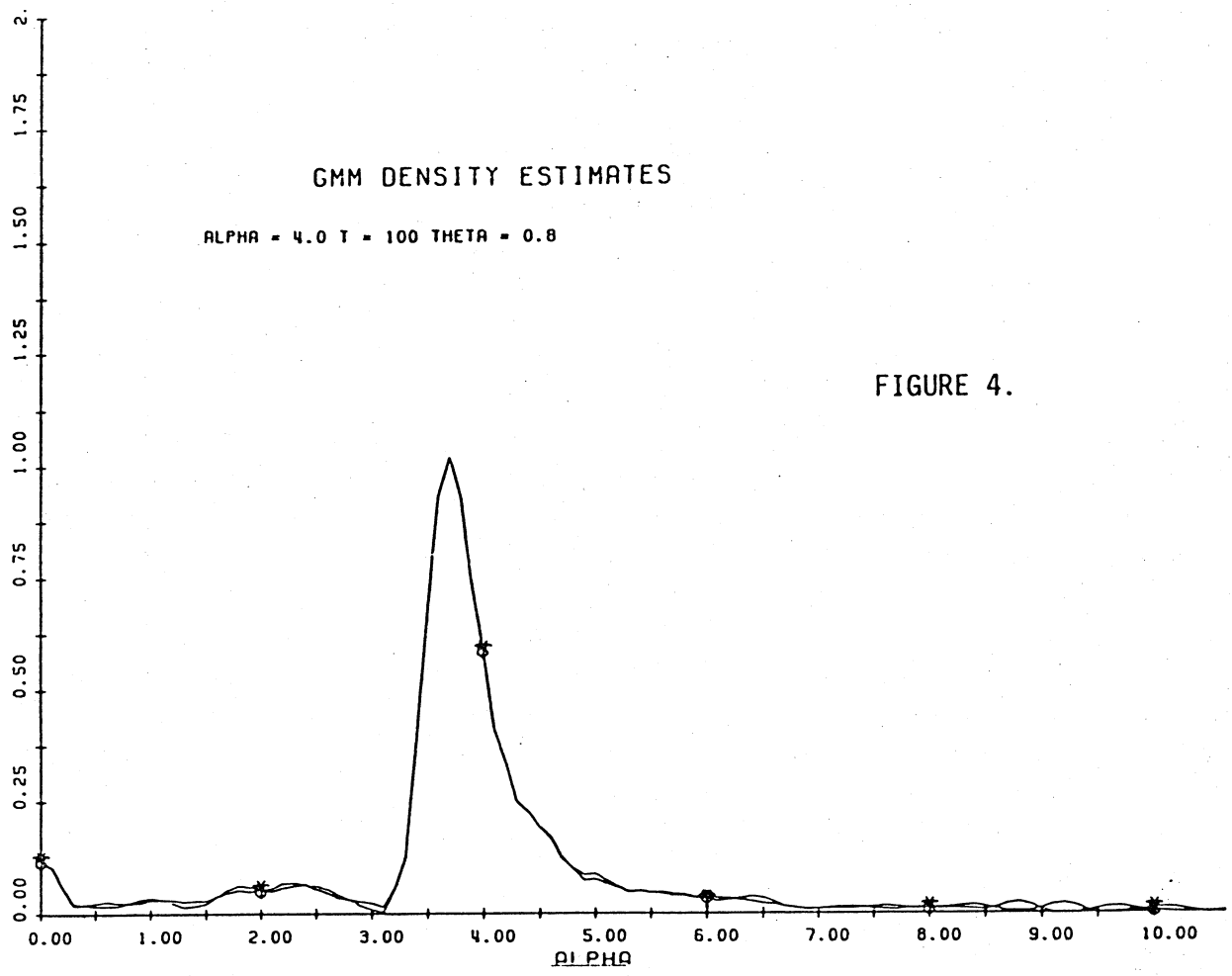
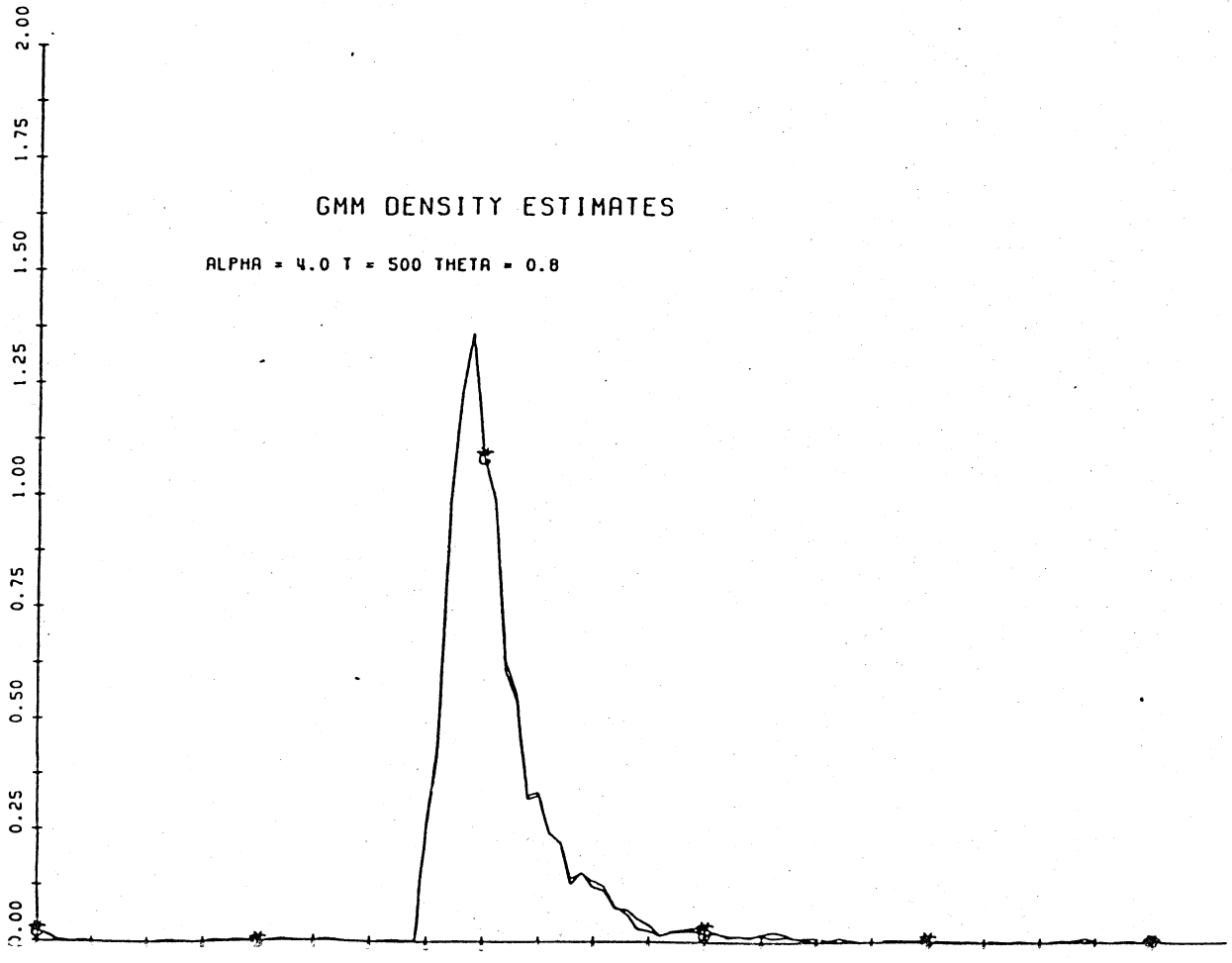

GMM DENSITY ESTIMATES
ALPHA = 4.0 T = 100 THETA = 0.8
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
ALPHA

FIGURE 4.


GMM DENSITY ESTIMATES
ALPHA = 4.0 T = 500 THETA = 0.8
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

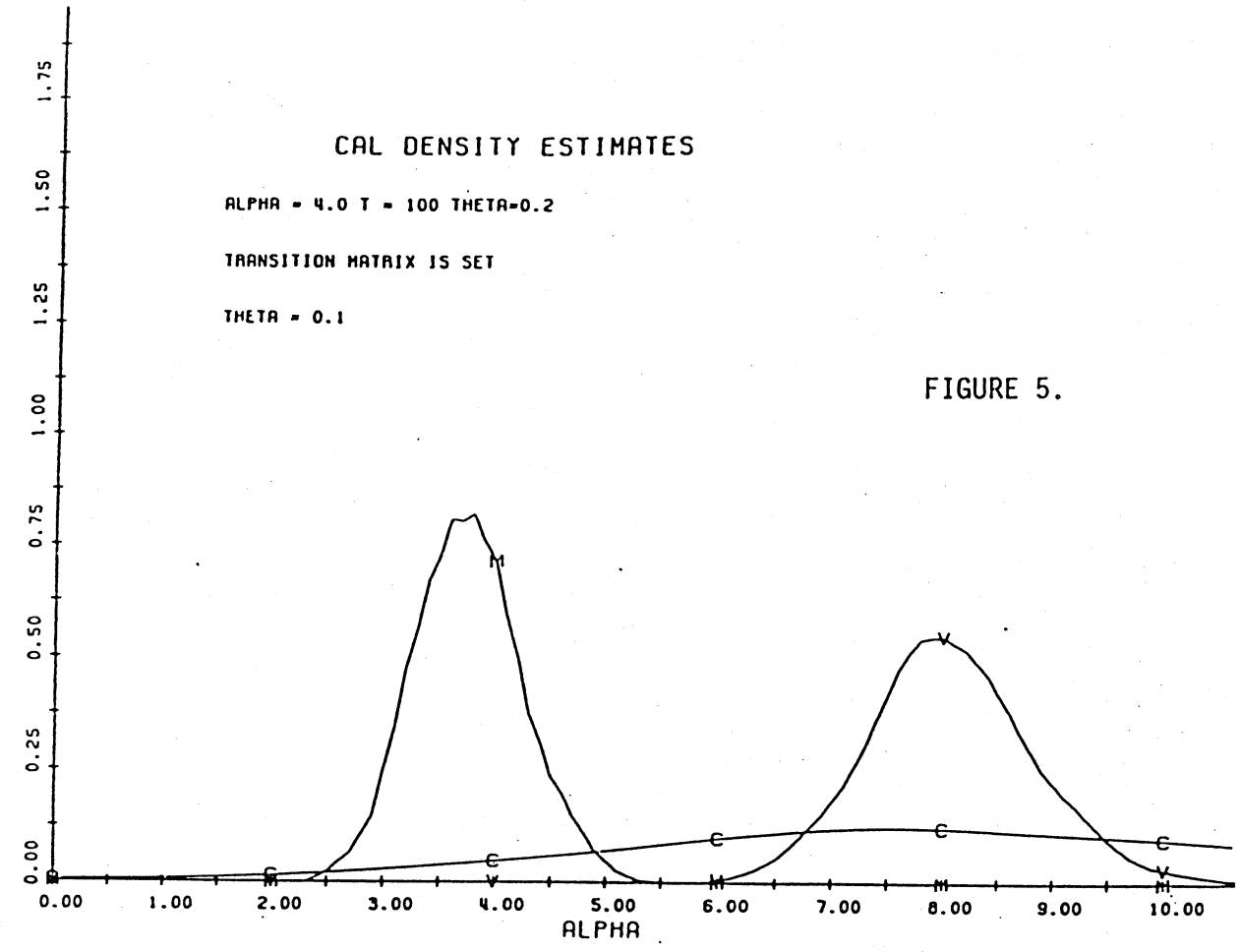
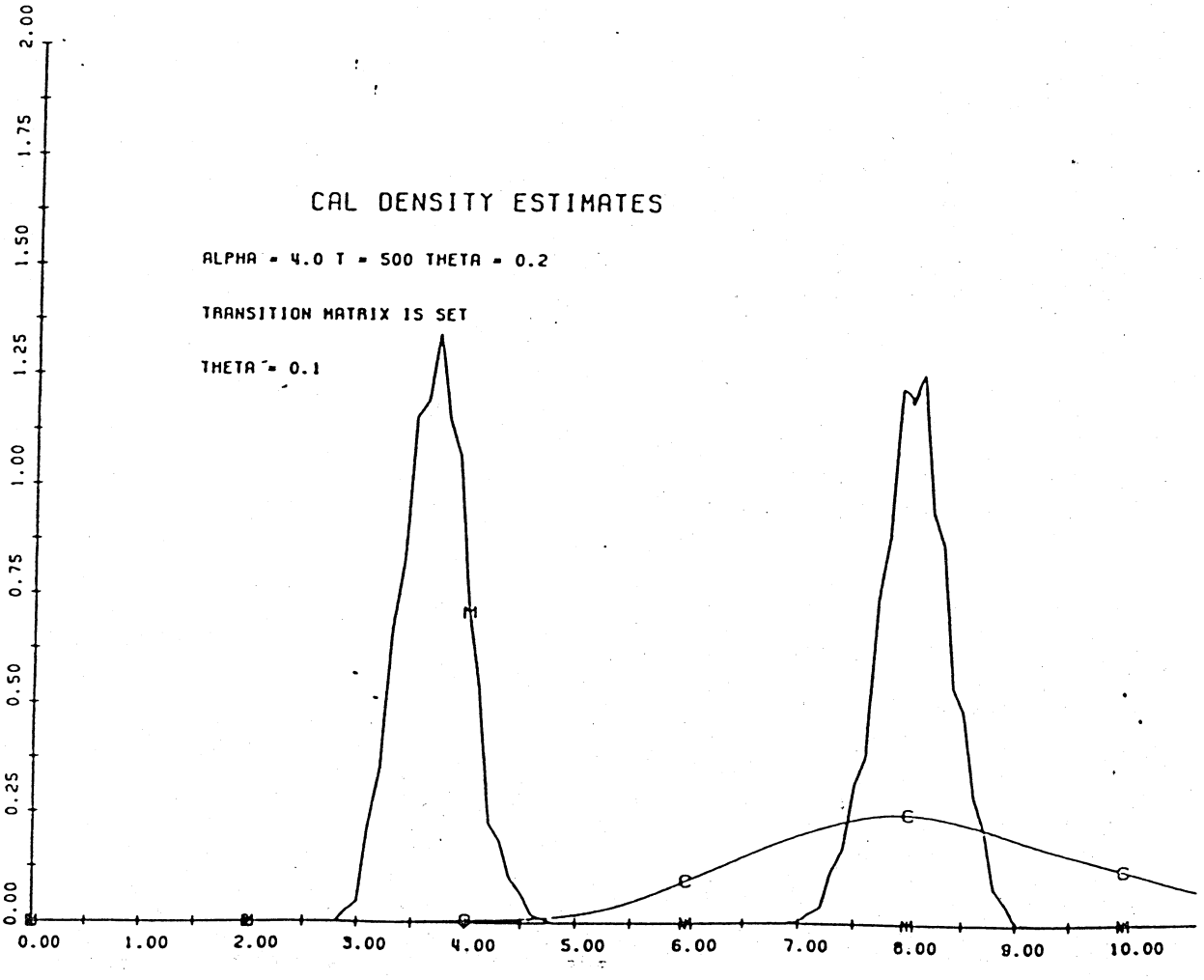

CAL DENSITY ESTIMATES
ALPHA = 4.0 T = 100 THETA=0.2
TRANSITION MATRIX IS SET
THETA = 0.1
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
ALPHA

FIGURE 5.


CAL DENSITY ESTIMATES
ALPHA = 4.0 T = 500 THETA = 0.2
TRANSITION MATRIX IS SET
THETA = 0.1
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

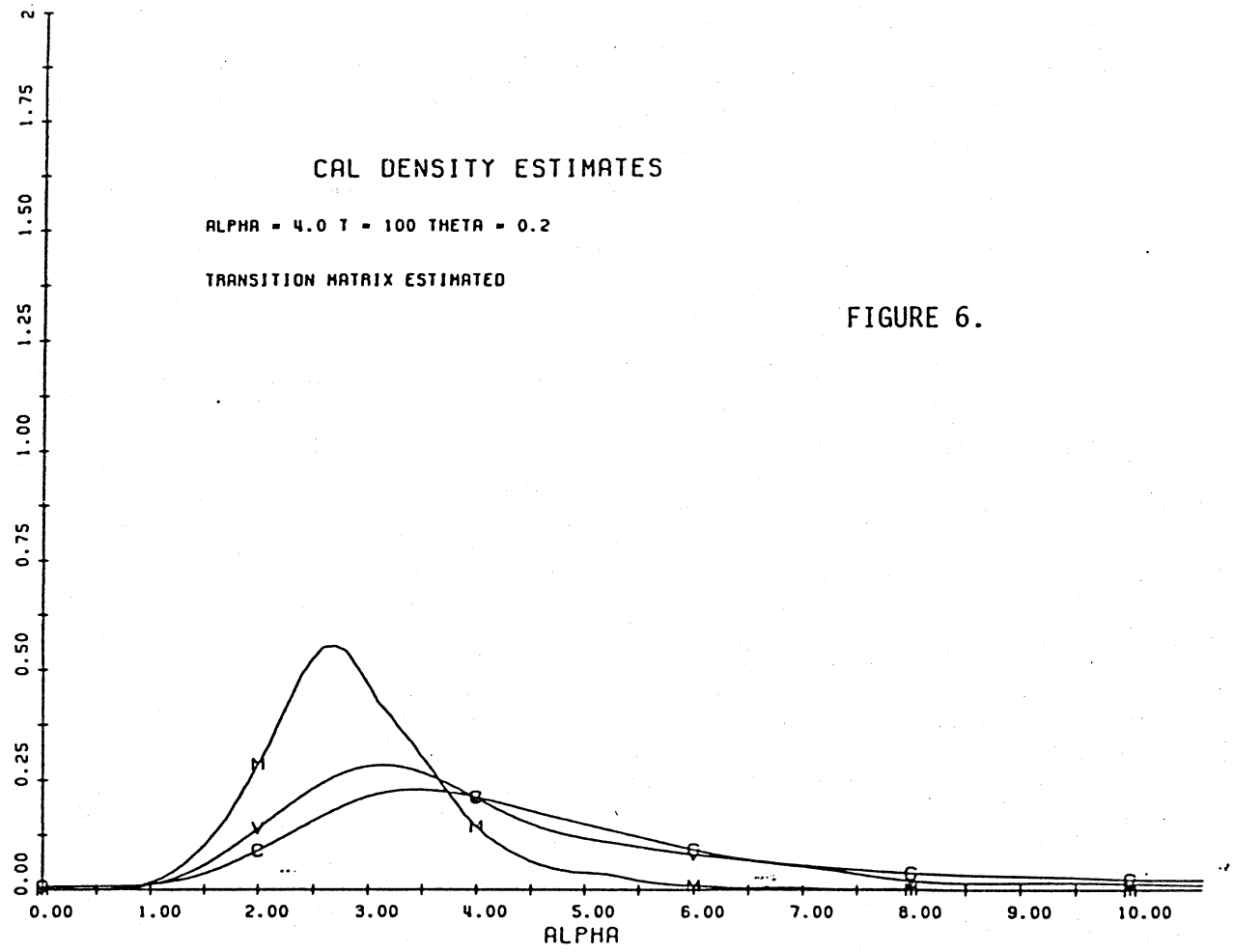
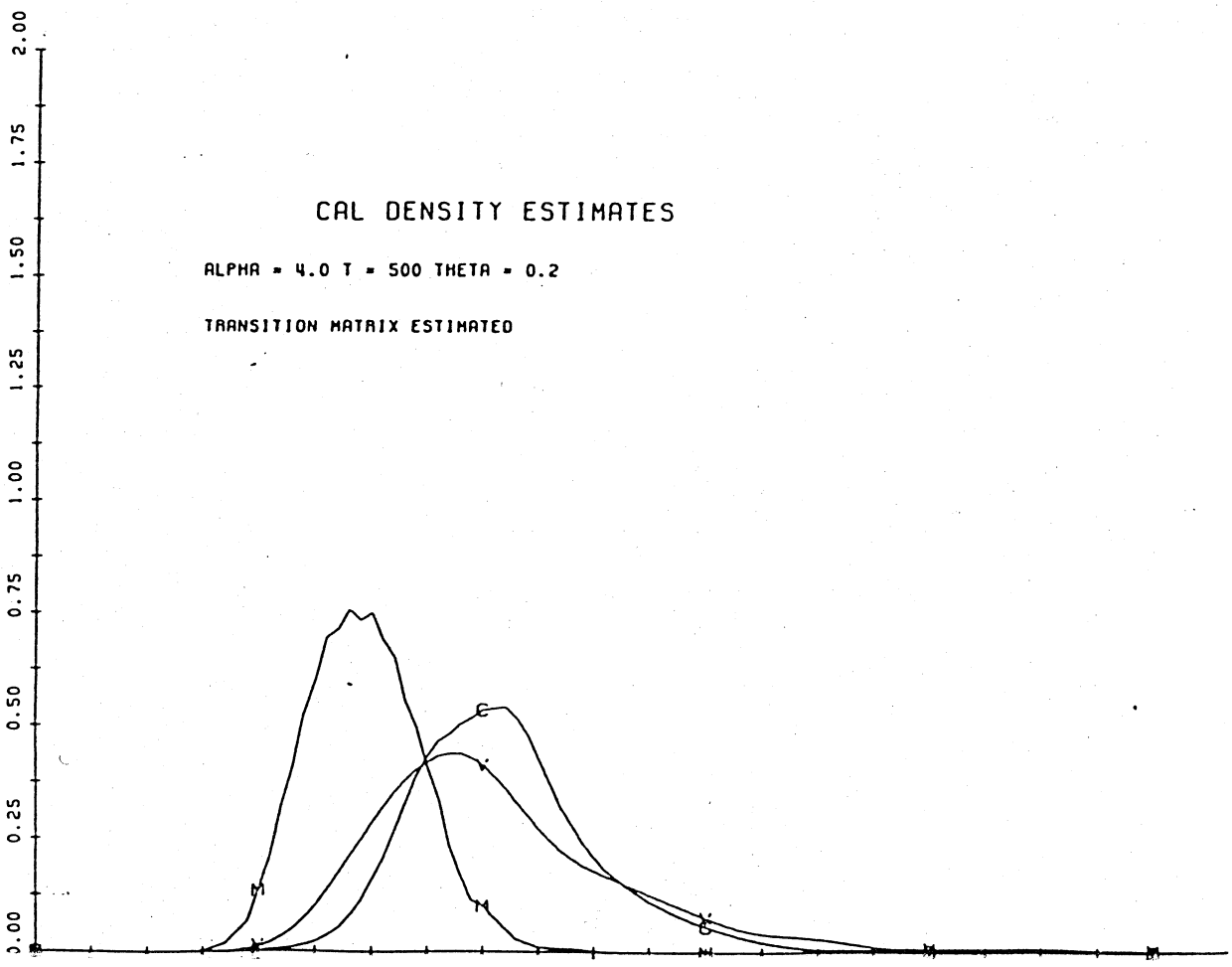

CAL DENSITY ESTIMATES
ALPHA = 4.0 T = 100 THETA = 0.2
TRANSITION MATRIX ESTIMATED
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
ALPHA

FIGURE 6.


CAL DENSITY ESTIMATES
ALPHA = 4.0 T = 500 THETA = 0.2
TRANSITION MATRIX ESTIMATED
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

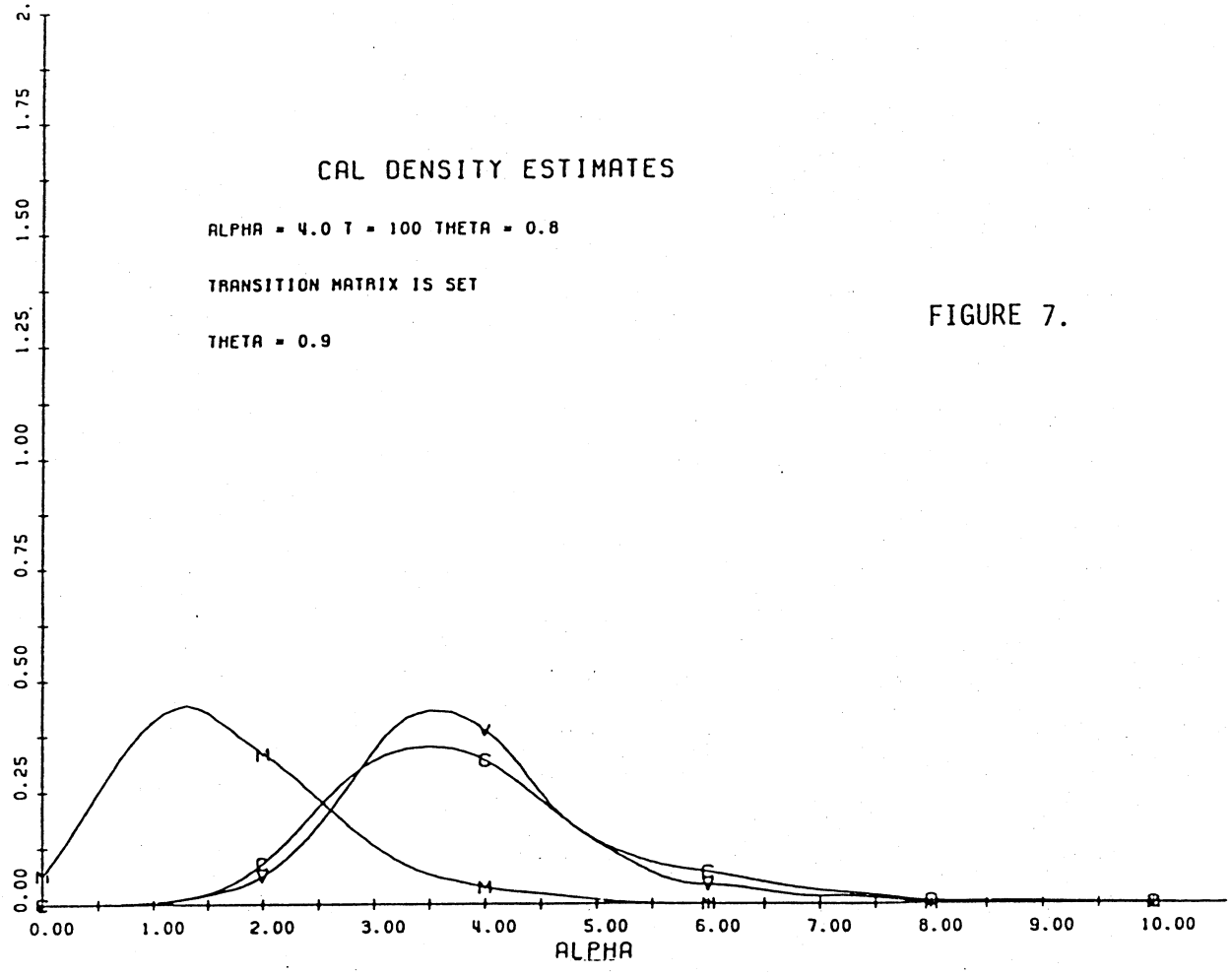
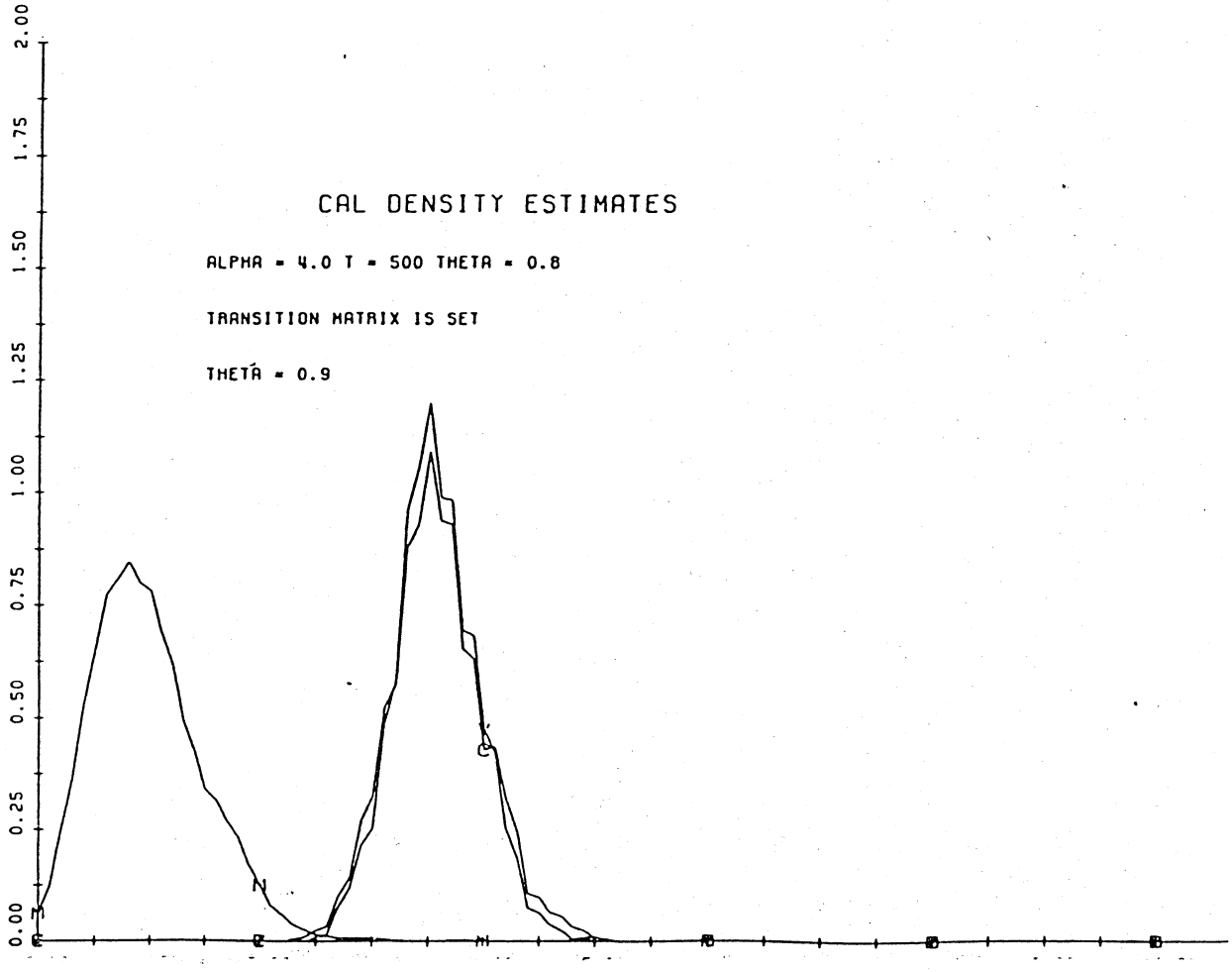

CAL DENSITY ESTIMATES
ALPHA = 4.0 T = 100 THETA = 0.8
TRANSITION MATRIX IS SET
THETA = 0.9
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
ALPHA

FIGURE 7.


CAL DENSITY ESTIMATES
ALPHA = 4.0 T = 500 THETA = 0.8
TRANSITION MATRIX IS SET
THETA = 0.9
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

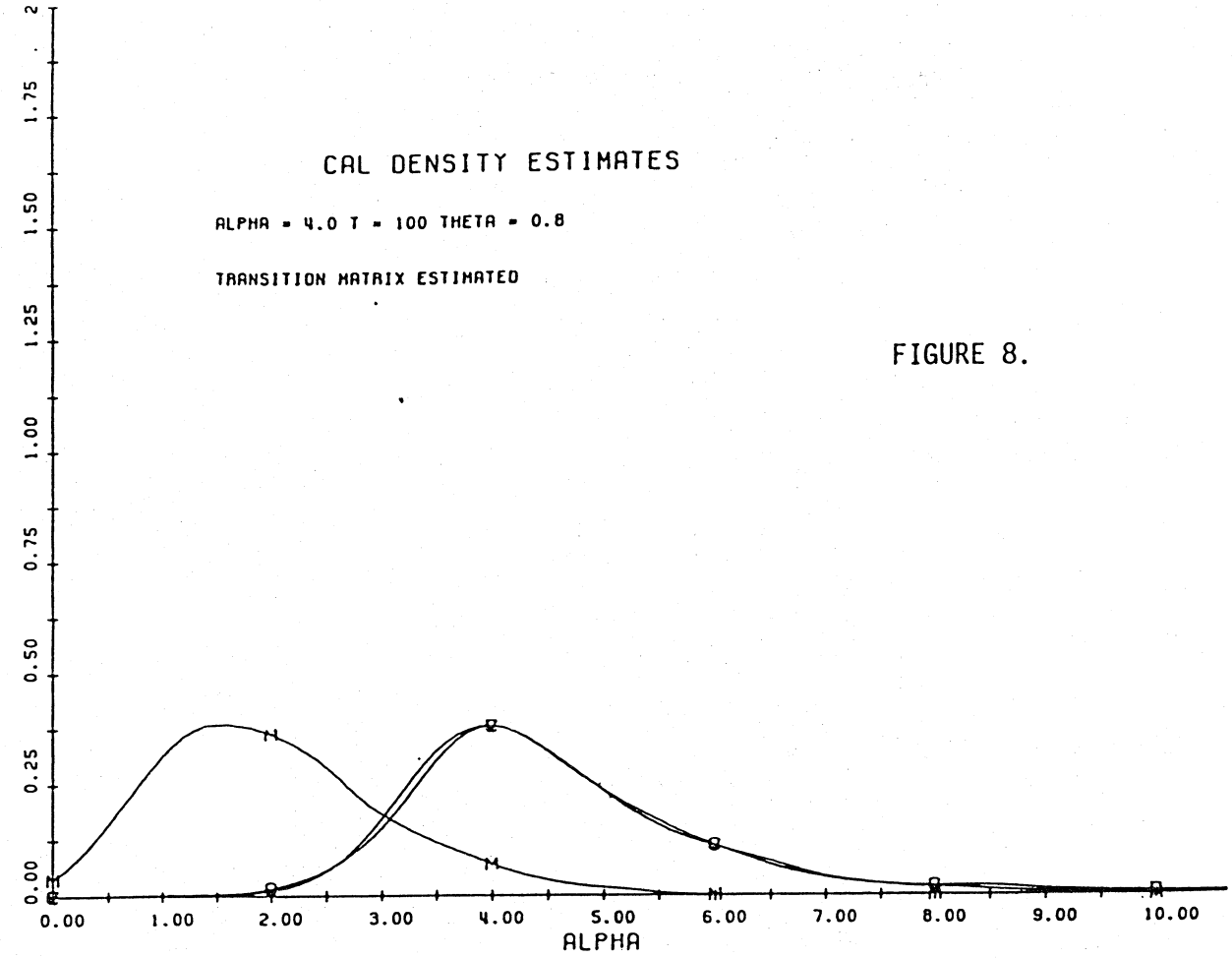
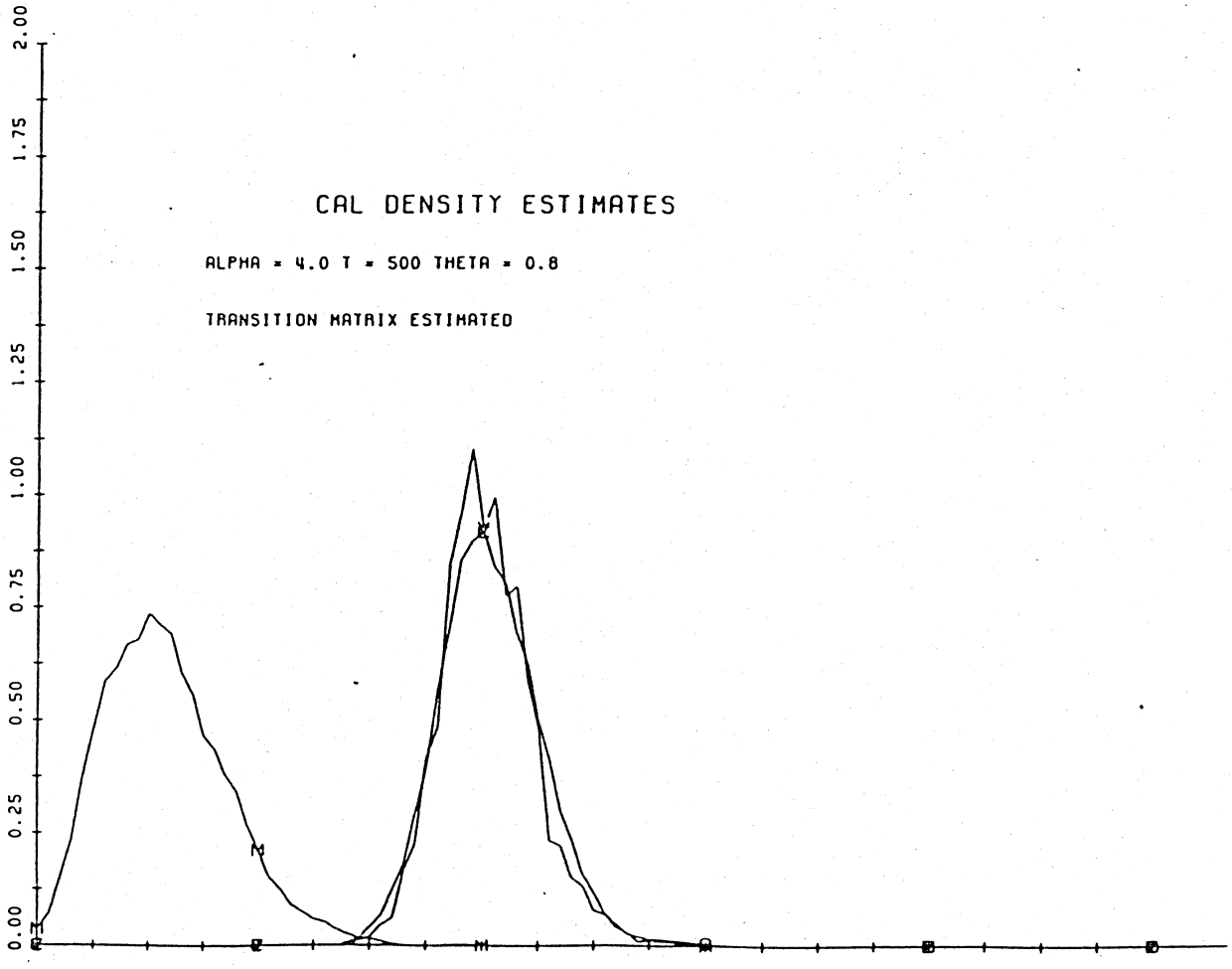

CAL DENSITY ESTIMATES
ALPHA = 4.0 T = 100 THETA = 0.8
TRANSITION MATRIX ESTIMATED
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
ALPHA

FIGURE 8.


CAL DENSITY ESTIMATES
ALPHA = 4.0 T = 500 THETA = 0.8
TRANSITION MATRIX ESTIMATED
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

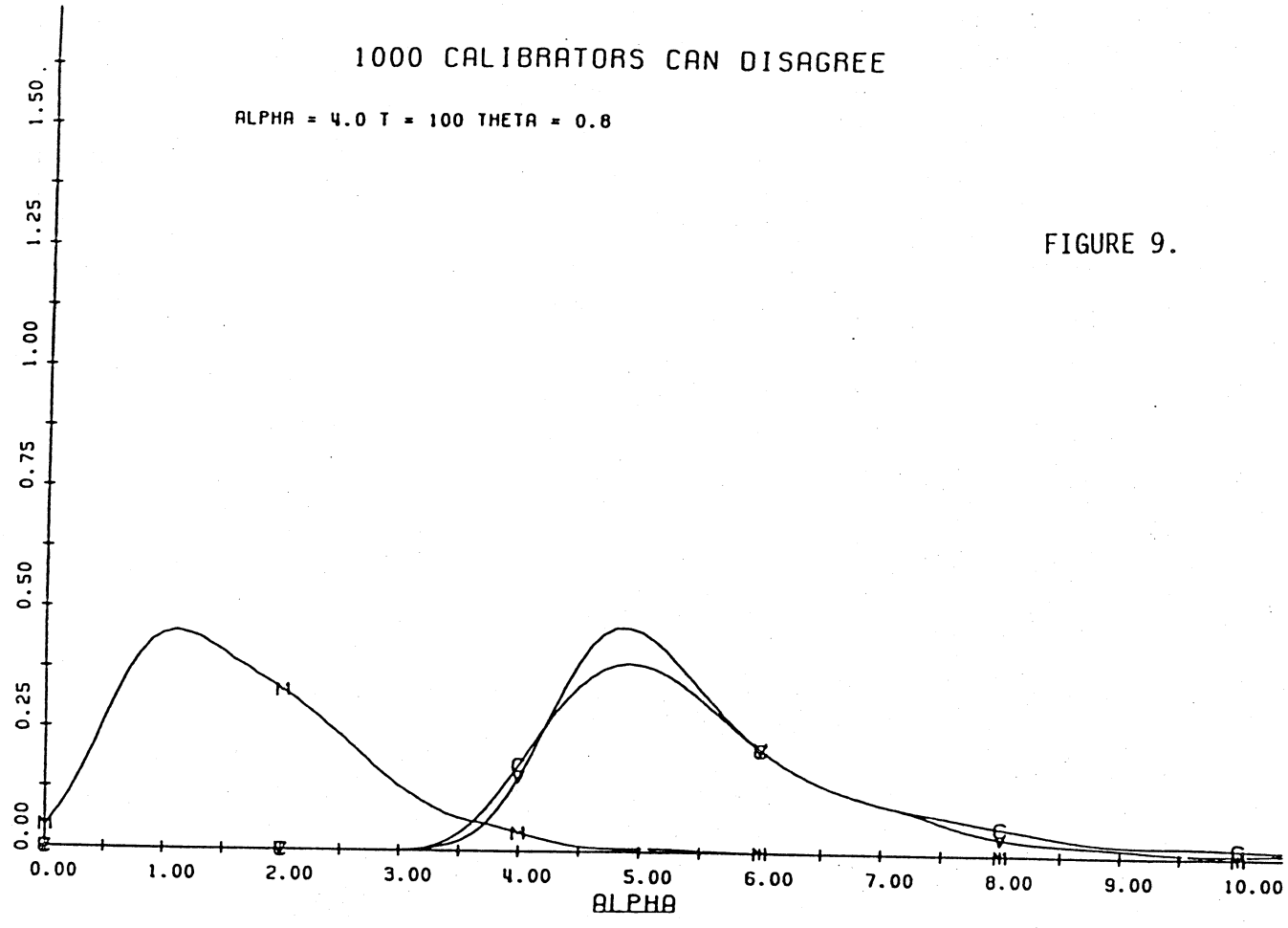

1000 CALIBRATORS CAN DISAGREE
ALPHA = 4.0 T = 100 THETA = 0.8
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
ALPHA

FIGURE 9.