November 21, 1975

#70

ESTIMATION OF A DYNAMIC DEMAND FUNCTION FOR GASOLINE WITH DIFFERENT SCHEMES OF PARAMETER VARIATION

by

J.S. Mehta, G.V.L. Narasimham and P.A.V.B. Swamy

Estimation of A Dynamic Demand Function for Gasoline with Different Schemes of Parameter Variation

by

J.S. Mehta, G.V.L. Narasimham and P.A.V.B. Swamy*

Temple University, Department of Commerce and Federal Reserve System

1. INTRODUCTION

This paper attempts to study the demand for motor gasoline by private individuals in the U.S. The study has a three-fold objective: (i) to formulate a demand function for gasoline whose consumption is technologically related to the stock of automobiles owned by individuals, (ii) to specify a dynamic equation which captures the effect of the adjustment of these stocks over time on the consumption of gasoline and (iii) to empirically verify the stability of the postulated model across geographical regions in the U.S.

The interesting studies by Houthakker, Verleger and Sheehan (1974), Verleger and Sheehan (1973) and by Ramsey, Rasche and Allen (1974) are allied to the work in this paper. Our study differs from these studies in three important respects. First, we consider the hypothesis that the gasoline purchased on the market by consumers are inputs into the production of automobile services which are the arguments of a household utility function. This hypothesis provides the microeconomic foundations of our aggregate demand function. Second, we treat the aggregation problem explicitly. Third, the parameters of the aggregate demand function are allowed to vary freely across geographical regions.

In Section 2 we present the theoretical formulation of the dynamic model for gasoline. In Section 3 the results of the estimation of the gasoline model from aggregate time series data are presented. In Section 4 we take advantage of recently developed tools to estimate the coefficients of the dynamic model from a time series of cross-sections. Some concluding remarks are given in Section 5.

*/ The views expressed herein are solely those of the authors and do not necessarily represent the views of the Federal Reserve System.

2. A Model of Demand for Gasoline

It seems reasonable to assume that motor gasoline is not desired for it's own sake, but only as an input into the production of automobile services. The quantity of these services, rather than the quantity of gasoline, is an argument of a household utility function. It is, essentially, that the demand for gasoline is a derived demand, derived from the demand for automobile services. We, therefore, assume that the demand for gasoline, G_{it}, by the i^{th} household in the period t is a function of the total demand for automobile services, A^d_{it}, the price of gasoline and the prices of other "goods" that go into the production of automobile services:[1]

$$G_{it} = \phi_i(A^d_{it}, p_{it}) \tag{1}$$

where p_{it} is a vector of prices of goods. The functional form for the demand function in (1) depends on the functional form for the i^{th} household's technology.

Regarding the household production function we assume that it incorporates the "putty-clay" synthesis where the set of <u>ex ante</u> factor combinations is represented by a homogeneous vintage function but <u>ex post</u> operation of a vintage car must maintain a fixed proportion of gasoline to utilized car throughout the effective life of the car. Let S_{it} and S_{it-1} be the stocks of automobiles owned by the i^{th} household at the end and the beginning of the period t, respectively. If δ_{it} denotes a proportional rate of depreciation,

[1] Our terminology follows that of Pollak and Wachter (1975) who use the term "goods" to refer to inputs which are combined with time in a "household production function" to produce "commodities" like automobile services.

then the gross investment in automobiles in the period t is given by $I_i(t,t) = S_{it}-(1-\delta_{it})S_{it-1}$. In terms of these variables a putty-clay production function can be stated as

$$(2) \qquad A^s_{it} = \theta_{it}\lambda_i(1-\delta_{it})S_{it-1}+F_i(G_i(t,t),I_i(t,t),t)$$

where A^s_{it} is the output of automobile services, λ_i is the fixed proportion of gasoline to automobiles of vintage $v < t$[2], θ_{it} is the output-gasoline ratio and F_i is the ex ante production function with three inputs: gasoline, gross investment in automobiles and households' time. We assume that $A^d_{it} = A^s_{it} = A_{it}$.

It follows from the results of Pollak and Wachter (1975) that the demand for the commodity A_{it} is

$$(3) \qquad A^d_{it}-\theta_{it}(1-\delta_{it})\lambda_i S_{it-1} = f_{it}(X_{it},\underset{\sim}{p}_{it})$$

where $\underset{\sim}{p}_{it}$ is a vector of goods prices and X_{it} is the labor and nonlabor income of the i^{th} household in the period t. Changes in the i^{th} households' technology as well as taste changes will take the form of shifts in the demand function f_{it}.

The implication of eq.(3) is that the incremental demand for gasoline due to the replacement of automobiles and net increases in the stock of automobiles is determined in the manner suggested by the household production function model.

For the sake of simplicity we assume that the functions in (1) and (3) are linear. The quantity $\theta_{it}(1-\delta_{it})$ is not likely to vary violently from one period to the next. It may be safely assumed for gasoline that $\theta_{it}(1-\delta_{it}) = \theta_i(1-\delta_i)$ over our sample period. We also assume that the function

---

[2] The assumption of constancy of λ_i can be justified as in Balestra and Nerlove (1966, p. 587).

f_{it} in (3) is stable over our sample period.[3] With these assumptions if we substitute (3) into (1) we obtain

$$(4) \qquad G_{it} = \alpha_i+\gamma_{i1}G_{it-1}+\gamma_{i2}X_{it}+\gamma_{i3}p_{git}+u_{it}$$

where we follow Balestra and Nerlove (1966) in equating $\lambda_i S_{it-1}$ to G_{it-1}, p_g is the gasoline price and u_{it} represents all omitted variables. Unless $\gamma_{i1} = (1-\delta_i)$ the coefficient of G_{it-1} cannot be regarded as a pure depreciation effect. There is no reason to suppose that $\gamma_{i1} = (1-\delta_i)$. Thus our interpretation of γ_{i1} differs from that of Balestra and Nerlove (1966).

Notice that we do not force the coefficients of eq.(4) to be identical for all households; rather, we follow a general approach by allowing the coefficients to vary freely among individual households. Given that we do not know what the true utility function is, for an individual and much less for some large aggregation of individuals, it is doubtful whether any fixed-coefficient demand models integrated into a specific utility-maximization theory can compete with the model in (4) in terms of empirical adequacy. If, as we shall see to be the case, eq. (4) is a good approximation of the relation among observed magnitudes, this must be interpreted to mean that the variation of the coefficients of eq. (4) over time is of only secondary importance. Eq.(4) provides the maintained hypothesis within which the statistical analysis is carried out in this paper.

3 At the expense of some complication it is possible to relax these assumptions, see Swamy and Mehta (1975). However, the statistical procedures for estimating autoregressive equations with time-varying parameters are not fully developed.

3. Estimation from Aggregate Time Series Data

The preceding analysis is microeconomic in nature. Since data on individual households are not available, an explicit treatment of the aggregation problem is in order.

Assumption 1: (a) $u_{it} = \tau_t + \varepsilon_{it}$ where the τ_t are regarded as the individual-invariant time effects which are not accounted for by the included explanatory variables in eq.(4).

(b) The vectors $(\alpha_i, \gamma_{i1}, \gamma_{i2}, \gamma_{i3})'$ with different i subscripts are the realizations of a 4-dimensional variable with the mean vector $(\bar{\alpha}, \bar{\gamma}_1, \bar{\gamma}_2, \bar{\gamma}_3)'$ and a finite symmetric variance-covariance matrix. The variable u_{it} is independent of $(\alpha_i, \gamma_{i1}, \gamma_{i2}, \gamma_{i3})'$.

(c) $-1 < \gamma_{i1} < 1$ with probability $1 \forall i$.

(d) The variables X_{it} and p_{git} are predetermined from the standpoint of the i^{th} household such that $\sup_{1 \le t \le T} |X_{it}|$ and $\sup_{1 \le t \le T} |p_{git}|$ are uniformly, bounded almost surely for all i. $G_{io}, G_{i,-1}, X_{io}$ and p_{gio} are known constants.

(e) $\{u_{it}\}$ is a stationary process such that $E\, u_{it} = 0 \; \forall (i,t)$.

Now aggregate eq.(4) across individuals and divide through by the number of individuals; this gives

$$\frac{1}{n_t}\sum_{i=1}^{n_t} G_{it} = \bar{\alpha} + \bar{\gamma}_1 \frac{1}{n_t}\sum_{i=1}^{n_t} G_{it-1} + \bar{\gamma}_2 \frac{1}{n_t}\sum_{i=1}^{n_t} X_{it} + \bar{\gamma}_3 \frac{1}{n_t}\sum_{i=1}^{n_t} p_{git} + \frac{1}{n_t}\sum_{i=1}^{n_t} e_{it} + \tau_t \tag{5}$$

where n_t is the total number of individuals in a population in the period t, $\alpha_i = \bar{\alpha} + \xi_{io}$, $\gamma_{i\ell} = \bar{\gamma}_\ell + \xi_{i\ell}$ $(\ell = 1,2,3)$ and

$$e_{it} = \xi_{io} + \xi_{i1} G_{it-1} + \xi_{i2} X_{it} + \xi_{i3} p_{git} + \varepsilon_{it} \tag{6}$$

Assumption 1:

(f) Let $n = \min (n_1, \ldots, n_T)$. Then

$$\lim_{n \to \infty} E\left(\frac{1}{n_t} \sum_{i=1}^{n_t} e_{it}\right) = 0 \text{ and } \lim_{n \to \infty} \mathrm{var}\left(\frac{1}{n_t} \sum_{i=1}^{n_t} e_{it}\right) = 0 \quad \forall t$$

One can easily formulate a set of sufficient conditions for the validity of Assumption 1(f). Theil's (1971, pp. 571-2) assumption that the explanatory variables are independent of their own coefficients for "every" individual is sufficient but not necessary for the truth of Assumption 1(f). The basic premise of eq.(4) is that not all individuals behave in the same way when they are faced with rising incomes and rising prices. Gasoline is a necessary good for some individuals while it is a luxury good for some others. The price of gasoline may not have as significant an effect on the demand for gasoline by rich households as it would on the demand for gasoline by poor households. Consequently, it seems reasonable to assume that in eq. (4) the explanatory variables are correlated with their own coefficients for a finite number of individuals, say L, which is functionally independent of n. Under these conditions Assumption 1(f) is true.

When Assumption 1(f) is true eq. (5) becomes

$$(7) \qquad G_t = \bar{\alpha} + \bar{\gamma}_1 G_{t-1} + \bar{\gamma}_2 X_t + \bar{\gamma}_3 p_{gt} + \tau_t \quad (t = 1,2,\ldots,T)$$

where

$$G_t = \frac{1}{n_t} \sum_{i=1}^{n_t} G_{it}, \; G_{t-1} = \frac{1}{n_t} \sum_{i=1}^{n_t} G_{it-1}, \; X_t = \frac{1}{n_t} \sum_{i=1}^{n_t} X_{it} \text{ and } p_{gt} = \frac{1}{n_t} \sum_{i=1}^{n_t} p_{git} .$$

Assumption 2: The τ_t are independently and identically distributed with 0 mean and constant variance σ^2.

When we estimated eq. (7) using quarterly data for the U.S. for the period 1963I - 1973IV we obtained the following results.[4]

(8) $$G_t = 26.85 + \underset{(0.12)}{0.65}\, G_{t-1} + \underset{(1.73)}{5.56}\, X_t \underset{(0.55)}{-0.57}\, p_{gt} + \hat{\tau}_t \qquad \bar{R}^2 = 0.99.$$

Figures given in parentheses are the standard errors. The value of Durbin's (1970) h-statistic is -1.4 which is greater than the critical value of -1.96 at the 0.05 level of significance for N(0,1). Hence the null hypothesis that there is no first-order serial correlation among the τ_t's cannot be rejected. The coefficient estimates have the right signs, although the price coefficient is not precisely determined. In assessing the parameter estimates in (8), attention should be given to their economic import, that is, to their interpretation in terms of the income and price responses of the classical theory of consumer demand. It should be remembered that the demand function used here refers to the new demand for gasoline. Therefore, the average income and own-price elasticities are given by

(9) $$\eta_g = \left(\frac{\partial G^*}{\partial X}\right)\frac{\bar{X}}{\bar{G}^*} \text{ and } \eta_{gg} = \left(\frac{\partial G^*}{\partial p_g}\right)\frac{\bar{p}_g}{\bar{G}^*}$$

where $G_t^* = G_t - \bar{\gamma}_1 G_{t-1}$ and $\bar{X}$, $\overline{G^*}$ and $\bar{p}_g$ are the sample means of X_t, G_t^* and p_{gt} respectively. These definitions are borrowed from Balestra and Nerlove (1966, p. 592.)

The income and own-price elasticities have been obtained from our parameter estimates in (8) by evaluating, at the sample means, the formulas in (9). These are:

(10) $$\eta_g = 1.22 \text{ and } \eta_{gg} = -0.38$$

This value of η_g is implausible because we do not believe that the gasoline is a luxury good for majority of people. Since we do not regard the coefficient of G_{t-1}

[4] The source and nature of data are explained in the Appendix of the paper.

as a pure depreciation effect, a value of 0.65 for $\overline{\gamma}_1$ is not implausible.

Two different versions of eq. (8) were tried: (1) Eq. (8) with p_{gt} defined as the price of gasoline deflated by the implicit price index for the personal consumption expenditure, and (2) eq. (8) with X_t defined as the permanent income, $X_t^e = (1-\lambda_1) \sum_{j=0}^{\infty} \lambda_1^j X_{t-j}$ and p_{gt} defined as the "expected" absolute or relative price of gasoline, $p_{gt}^e = (1-\lambda_2) \sum_{j=0}^{\infty} \lambda_2^j p_{gt-j}$. These experiments produced highly implausible estimates for many parameters. Also, the inclusion of auto-price index as an explanatory variable worsened the results. For these reasons and also for reasons mentioned in Pollak and Wachter (1975), we do not include the auto-price index or some other commodity price index in eq. (4) and also, do not deflate p_{gt} by a consumer price index.

Klein (1962, p. 69) puts forward a very convincing argument to show that the assumption of endogeneity of X_t is plausible even when Assumption 1(d) is true. Consequently, our estimates in (8) might suffer from simultaneous equations bias. In the next section we use a cross-section sample which helps to avoid the problems of simultaneous equations bias.

Our study might suffer from the added problem of ignoring the supply side of the market for gasoline. However, the assumption of perfect elasticity of supply of gasoline does not seem to be implausible for our sample period. After examining simultaneously the static demand and supply of gasoline Ramsey et al. (1974) have obtained price and income elasticities which, in our judgement, are very high.

Finally, Assumption 1(b) is unrealistic for the whole population but can still reasonably be applied within geographical regions. The situation is reminiscent of the problems encountered in the Bayesian analysis of finite populations where the assumption of "exchangeability" is unrealistic for the whole population but can still reasonably be applied within strata, see Lindley (1971, p. 40). We now turn to the analysis of time series of cross-sections.

4. Estimation from Temporal Cross-Section Data

We might get improved estimates if we use diaggregated data. Furthermore, the richness of disaggregated data base allows a varied menu of alternative parameter variations. Fortunately, we could get the temporal cross-section data for the U.S. which consisted of quarterly observations drawn from the forty-eight continental states plus the District of Columbia for the period 1963I - 1973IV.[5] In this section we consider the problem of estimating eq. (4) from these data. We shall begin with the case in which the intercept varies in both the time and cross-section dimensions but not the slope coefficients. In the time series analysis of cross-section data, if the inter-individual and inter-temporal parameter variation takes the form of mere shifts in the regression intercept, then the specification of an equation with a shifting intercept and fixed slopes is adequate. We then go on to consider a model in which all parameters vary in a systematic or random fashion.

We may rewrite eq. (4) as

$$G_{ijt} = \alpha_{ij} + \gamma_{ij1} G_{ijt-1} + \gamma_{ij2} X_{ijt} + \gamma_{ij3} P_{gijt} + u_{ijt} \tag{11}$$

where i stands for the i^{th} household, j stands for the j^{th} state and t represents time.

<u>Assumption 3</u>: (a) $u_{ijt} = \tau_t + v_{jt} + \varepsilon_{ijt}$ where the term τ_t reflects the effects that are specific to the quarter but common to all individuals and states and v_{jt} is common to all households within the j^{th} state at a given time.

(b) For given j the vectors $(\alpha_{ij}, \gamma_{ij1}, \gamma_{ij2}, \gamma_{ij3})'$ with different i subscripts are the realizations of a 4-dimensional variable with mean vector $(\bar{\alpha}_j, \bar{\gamma}_{j1}, \bar{\gamma}_{j2}, \bar{\gamma}_{j3})'$

---

[5] The data base and sources are described in detail in the Appendix.

and a finite symmetric variance-covariance matrix Δ_{jj}. The variable u_{ijt} is independent of $(\alpha_{ij},\gamma_{ij1},\gamma_{ij2},\gamma_{ij3})'$.

(c) $-1<\gamma_{ij1}<1$ with probability 1 $\forall$ (i,j).

(d) The variables X_{ijt} and p_{gijt} are predetermined from the standpoint of the i^{th} household in the j^{th} state such that $\sup_{1\le t\le T} |X_{ijt}|$ and $\sup_{1\le t\le T} |p_{gijt}|$ are uniformly bounded almost surely for all i and j. The initial values G_{ijo}, $G_{ij,-1}$, X_{ijo} and p_{gijo} are known constants.

(e) $\{u_{ijt}\}$ is a stationary process such that $EU_{ijt} = 0$ $\forall$ (i,j,t).

On summing eq. (11) over i and dividing through by the number of individuals we obtain

$$(12) \quad \frac{1}{n_{jt}} \sum_{i=1}^{n_{jt}} G_{ijt} = \bar{\alpha}_j + \bar{\gamma}_{j1} \frac{1}{n_{jt}} \sum_{i=1}^{n_{jt}} G_{ijt-1} + \gamma_{j2} \frac{1}{n_{jt}} \sum_{i=1}^{n_{jt}} X_{ijt} + \bar{\gamma}_{j3} \frac{1}{n_{jt}} \sum_{i=1}^{n_{jt}} p_{gitj} + \frac{1}{n_{jt}} \sum_{i=1}^{n_{jt}} e_{ijt} + \tau_t + v_{jt}$$

where n_{jt} is the number of households in the j^{th} state in the period t, $\alpha_{ij} = \bar{\alpha}_j + \xi_{ijo}$, $\gamma_{ij\ell} = \bar{\gamma}_{j\ell} + \xi_{ij\ell}$ $(\ell = 1,2,3)$ and

$$(13) \qquad e_{ijt} = \xi_{ijo} + \xi_{ij1} G_{ijt-1} + \xi_{ij2} X_{ijt} + \xi_{ij3} p_{gijt} + \varepsilon_{ijt}$$

<u>Assumption 3:</u> (f) For given j and t $\lim_{n\to\infty} E(\frac{1}{n_{jt}} \sum_{i=1}^{n_{jt}} e_{ijt}) = 0$ and $\lim_{n\to\infty} \text{var}(\frac{1}{n_{jt}} \sum_{i=1}^{n_{jt}} e_{ijt}) = 0$ where n is the minimum of n_{jt} (j = 1,2,...,49; t = 1,2,...,44). When Assumption 3(f) is true we can approximate eq. (12) by

(14) $$G_{jt} = \bar{\alpha}_j + \bar{\gamma}_{j1}G_{jt-1} + \bar{\gamma}_{j2}X_{jt} + \bar{\gamma}_{j3}p_{gjt} + \tau_t + v_{jt} \quad (j = 1,2,\ldots,m;\ t = 1,2,\ldots,T)$$

where $G_{jt} = \frac{1}{n_{jt}} \sum_{i=1}^{n_{jt}} G_{ijt}$, $G_{jt-1} = \frac{1}{n_{jt}} \sum_{i=1}^{n_{jt}} G_{ijt-1}$, $X_{jt} = \frac{1}{n_{jt}} \sum_{i=1}^{n_{jt}} X_{ijt}$ and

$$p_{gjt} = \frac{1}{n_{jt}} \sum_{i=1}^{n_{jt}} p_{gijt} \ .$$

4.1. Error Components Model

In the special case in which the slope coefficients are independent of j eq. (14) can be written as

(15) $$G_{jt} = \bar{\alpha} + \bar{\gamma}_1 G_{jt-1} + \bar{\gamma}_2 X_{jt} + \bar{\gamma}_3 p_{gjt} + \mu_j + \tau_t + v_{jt} \quad (j = 1,2,\ldots,m;\ t = 1,2,\ldots,T)$$

where $\bar{\alpha}_j = \bar{\alpha} + \mu_j$, μ_j is a time-invariant state effect, τ_t is a state-invariant period effect and the v_{jt} represent the combined effect on G_{jt} of the neglected variables. For our data m = 49 and T = 43.

If we arrange the observations on each variable first by state, then according to period, we may depict the observations as

(16) $$\underline{y} = \bar{\alpha}\underline{\iota}_{mT} + Z\underline{\bar{\gamma}} + (\underline{\mu} \otimes \underline{\iota}_T) + (\underline{\iota}_m \otimes \underline{\tau}) + \underline{v}$$

where $\underline{y}$ is a mTx1 vector of observations on G_{jt}, $\underline{\iota}$ denotes a "summer" vector, all of whose elements are 1, the subscripts on $\underline{\iota}$ indicates its order, Z is a mTx3 matrix of observations on G_{jt-1}, X_{jt} and p_{gjt}, $\underline{\bar{\gamma}} = (\bar{\gamma}_1,\bar{\gamma}_2,\bar{\gamma}_3)'$ $\underline{\mu} = (\mu_1,\mu_2,\ldots,\mu_m)'$, $\underline{\tau} = (\tau_1,\tau_2,\ldots,\tau_T)'$, $\underline{v}$ is a mTx1 vector of observations on v_{jt} and $\otimes$ denotes the Kronecker product.

We now make the following assumption.

Assumption 4:

(a) The rank of $(\underline{\iota}_{mT}, Z)$ is 4;

(b) $E\underline{\mu} = \underline{0}$ and $E\underline{\mu}\underline{\mu}' = \sigma_\mu^2 \Omega_1$ where $\Omega_1 = [\rho_1 \underline{\iota}_m \underline{\iota}_m' + (1-\rho_1)I_m]$, $-(m-1)^{-1} \leq \rho_1 < 1$

and $0 < \sigma_\mu^2 < \infty$;

(c) $E\underline{\tau} = \underline{0}$ and $E\underline{\tau}\underline{\tau}' = \sigma_\tau^2 \Omega_2$ where $\Omega_2 = [\rho_2 \underline{\iota}_T \underline{\iota}_T' + (1-\rho_2)I_T]$, $-(T-1)^{-1} \leq \rho_2 < 1$

and $0 < \sigma_\tau^2 < \infty$;

(d) $E\underline{v} = \underline{0}$ and $E\underline{v}\underline{v}' = \sigma_v^2(\Omega_3 \otimes \Omega_4)$ where $\Omega_3 = [\rho_3 \underline{\iota}_m \underline{\iota}_m' + (1-\rho_3)I_m]$,

$-(m-1)^{-1} \leq \rho_3 < 1$, $\Omega_4 = [\rho_4 \underline{\iota}_T \underline{\iota}_T' + (1-\rho_4)I_T]$, $-(T-1)^{-1} \leq \rho_4 < 1$ and $0<\sigma_v^2 \leq \infty$;

(e) $E\underline{\mu}\underline{\tau}' = 0$, $E\underline{\tau}\underline{v}' = 0$ and $E\underline{\mu}\underline{v}' = 0$;

(f) $\underline{\mu}, \underline{\tau}$ and $\underline{v}$ are normally distributed

(g) X_{jt} and p_{gjt} are exogenous.

The implications of Assumption 4 are made clear by Swamy and Mehta (1973) and Hill (1967). Under Assumption 4 the vector $\underline{w} = (\underline{\mu} \otimes \underline{\iota}_T) + (\underline{\iota}_m \otimes \underline{\tau}) + \underline{v}$ is normally distributed with mean vector $\underline{0}$ and variance-covariance matrix of the form

$$\Sigma_1 = \sigma_\mu^2(\Omega_1 \otimes \underline{\iota}_T \underline{\iota}_T') + \sigma_\tau^2(\underline{\iota}_m \underline{\iota}_m' \otimes \Omega_2) + \sigma_v^2(\Omega_3 \otimes \Omega_4) \tag{17}$$

Now it is convenient to split the sample of mT observations into the following four (unequal) independent parts. For this purpose we utilize the orthogonal transformation described in Swamy and Mehta (1973).

(18a) $\bar{y}_1 = \bar{\underline{x}}_1' \bar{\underline{\beta}} + \bar{w}_1$

(18b) $\bar{\underline{y}}_2 = \bar{Z}_2 \bar{\underline{\gamma}} + \bar{\underline{w}}_2$

(18c) $\bar{\underline{y}}_3 = \bar{Z}_3 \bar{\underline{\gamma}} + \bar{\underline{w}}_3$

(18d) $\bar{\underline{y}}_4 = \bar{Z}_4 \bar{\underline{\gamma}} + \bar{\underline{w}}_4$

where $\bar{y}_1 = Q_1\underline{y}$, $\bar{\underline{y}}_\ell = Q_\ell \underline{y}$ $(\ell = 2,3,4)$, $\bar{\underline{x}}_1' = Q_1(\underline{\iota}_{mT}, Z)$, $\bar{\underline{\beta}} = (\bar{\alpha}, \bar{\underline{\gamma}}')'$, $\bar{Z}_\ell = Q_\ell Z$ $(\ell = 2,3,4)$,

$\bar{w}_1 = Q_1\underline{w}$, $\bar{\underline{w}}_\ell = Q_\ell \underline{w} (\ell = 2,3,4)$, $Q_1 = (\underline{\imath}'_m/\sqrt{m} \otimes \underline{\imath}'_T/\sqrt{T})$, $Q_2 = (C_2 \otimes \underline{\imath}'_T/\sqrt{T})$, $Q_3 = (\underline{\imath}'_m/\sqrt{m} \otimes C_1)$ $Q_4 = (C_2 \otimes C_1)$, $0_T = (\underline{\imath}_T/\sqrt{T}, C_1')'$ is an orthogonal matrix of order T, and $0_m = (\underline{\imath}_m/\sqrt{m}, C_2')'$ is an orthogonal matrix of order m.

Swamy and Mehta (1973) have shown that

(19) $E\,\bar{w}_1 = 0$, $E\,\bar{\underline{w}}_\ell = \underline{0}$ $(\ell = 2,3,4)$

(20a) $E\,\bar{w}_1^2 = \sigma_1^2$,

(20b) $E\,\bar{\underline{w}}_2\bar{\underline{w}}_2' = \sigma_2^2\, I_{m'}$, $m' = m-1$,

(20c) $E\,\bar{\underline{w}}_3\bar{\underline{w}}_3' = \sigma_3^2\, I_{T'}$, $T' = T-1$,

(20d) $E\,\bar{\underline{w}}_4\bar{\underline{w}}_4' = \sigma_4^2\, I_{m'T'}$

and

(21) $E\,\bar{w}_1\bar{\underline{w}}_\ell = \underline{0}$ $(\ell = 2,3,4)$ and $E\,\bar{\underline{w}}_\ell\bar{\underline{w}}_{\ell'}' = 0$ for $\ell \neq \ell' = 2,3,4$. The variances $\sigma_1^2, \sigma_2^2, \sigma_3^2$ and σ_4^2 are functions of $\sigma_\mu^2, \sigma_\tau^2, \sigma_\nu^2$, $\rho_\ell (\ell = 1,2,3,4)$, m and T.

Some Point Estimates

Alternative estimators of $\bar{\underline{\gamma}}$ developed by Swamy (1971, Ch. III) are

(22) $$\hat{\underline{\gamma}} = (m'\bar{Z}_2'\bar{Z}_2 + T'\bar{Z}_3'\bar{Z}_3 + m'T'\bar{Z}_4'\bar{Z}_4)^{-1}(m'\bar{Z}_2'\bar{\underline{y}}_2 + T'\bar{Z}_3'\bar{\underline{y}}_3 + m'T'\bar{Z}_4'\bar{\underline{y}}_4)$$

(23) $$\hat{\underline{\gamma}}_2 = (\bar{Z}_2'\bar{Z}_2)^{-1}\bar{Z}_2'\bar{\underline{y}}_2$$

(24) $$\hat{\underline{\gamma}}_3 = (\bar{Z}_3'\bar{Z}_3)^{-1}\bar{Z}_3'\bar{\underline{y}}_3$$

(25) $$\hat{\underline{\gamma}}_4 = (\bar{Z}_4'\bar{Z}_4)^{-1}\bar{Z}_4'\bar{\underline{y}}_4$$

(26) $$\hat{\underline{\gamma}}_{24} = \left(\frac{\bar{Z}_2'\bar{Z}_2}{\hat{\sigma}_2^2} + \frac{\bar{Z}_4'\bar{Z}_4}{\hat{\sigma}_4^2}\right)^{-1}\left(\frac{\bar{Z}_2'\bar{\underline{y}}_2}{\hat{\sigma}_2^2} + \frac{\bar{Z}_4'\bar{\underline{y}}_4}{\hat{\sigma}_4^2}\right)$$

(27) $$\hat{\underline{\gamma}}_{34} = \left(\frac{\bar{Z}_3'\bar{Z}_3}{\hat{\sigma}_3^2} + \frac{\bar{Z}_4'\bar{Z}_4}{\hat{\sigma}_4^2}\right)^{-1}\left(\frac{\bar{Z}_3'\bar{\underline{y}}_3}{\hat{\sigma}_3^2} + \frac{\bar{Z}_4'\bar{\underline{y}}_4}{\hat{\sigma}_4^2}\right)$$

$$(28)\quad \hat{\bar{\underline{\gamma}}}_{234} = \left(\frac{\bar{Z}_2'\bar{Z}_2}{\hat{\sigma}_2^2} + \frac{\bar{Z}_3'\bar{Z}_3}{\hat{\sigma}_3^2} + \frac{\bar{Z}_4'\bar{Z}_4}{\hat{\sigma}_4^2}\right)^{-1} \left(\frac{\bar{Z}_2'\bar{y}_2}{\hat{\sigma}_2^2} + \frac{\bar{Z}_3'\bar{y}_3}{\hat{\sigma}_3^2} + \frac{\bar{Z}_4'\bar{y}_4}{\hat{\sigma}_4^2}\right)$$

where $\hat{\sigma}_2^2 = (\bar{y}_2-\bar{Z}_2\hat{\bar{\gamma}}_2)'(\bar{y}_2-\bar{Z}_2\hat{\bar{\gamma}}_2)/(m-4)$, $\hat{\sigma}_3^2 = (\bar{y}_3-\bar{Z}_3\hat{\bar{\gamma}}_3)'(\bar{y}_3-\bar{Z}_3\hat{\bar{\gamma}}_3)/(T-4)$ and

$\hat{\sigma}_4^2 = (\bar{y}_4-\bar{Z}_4\hat{\bar{\gamma}}_4)'(\bar{y}_4-\bar{Z}_4\hat{\bar{\gamma}}_4)/[(m-1)(T-1)-3]$.

Swamy (1974, p. 148) has shown that in eq s. (18b)-(18d) each value of the lagged dependent variable which appears as an explanatory variable is independent of the contemporaneous and the succeeding values of the disturbances. Consequently, an argument parallel to that of Goldberger (1964, pp. 272-274) and Theil (1971, pp. 408-413) shows that the estimators (25)-(28) are consistent and have the same limiting distribution. Swamy (1971, Chs. II and III) and Arora (1972) have shown that when $\bar{\gamma}_1 = 0$, any of the estimators in (22), (25)-(28) is not uniformly (in the unknown parameter vector) more efficient according to the minimum variance unbiasedness criterion in small samples than the others.

Since the limiting distributions of $\sqrt{mT}(\hat{\bar{\gamma}}-\bar{\gamma})$, $\sqrt{m}(\hat{\bar{\gamma}}_2-\bar{\gamma})$, $\sqrt{T}(\hat{\bar{\gamma}}_3-\bar{\gamma})$, $\sqrt{mT}(\hat{\bar{\gamma}}_{24}-\bar{\gamma})$, $\sqrt{mT}(\hat{\bar{\gamma}}_4-\bar{\gamma})$, $\sqrt{mT}(\hat{\bar{\gamma}}_{34}-\bar{\gamma})$ and $\sqrt{mT}(\hat{\bar{\gamma}}_{234}-\bar{\gamma})$ have finite variance-covariance matrices, the estimated symptotic variance-covariance (EAVC)matrices of the estimators (22)-(28) are[7]

$$(29)\quad \text{EAVC}(\hat{\bar{\gamma}}) = (m'\bar{Z}_2'\bar{Z}_2+T'\bar{Z}_3'\bar{Z}_3+m'T'\bar{Z}_4'\bar{Z}_4)^{-1}(\hat{\sigma}_2^2 m'^2\bar{Z}_2'\bar{Z}_2+\hat{\sigma}_3^2T'^2\bar{Z}_3'\bar{Z}_3+\hat{\sigma}_4^2m'^2T'^2\bar{Z}_4'\bar{Z}_4)\cdot(m'\bar{Z}_2'\bar{Z}_2+T'\bar{Z}_3'\bar{Z}_3+m'T'\bar{Z}_4'\bar{Z}_4)^{-1},$$

$$(30)\quad \text{EAVC}(\hat{\bar{\gamma}}_2) = \hat{\sigma}_2^2(\bar{Z}_2'\bar{Z}_2)^{-1},$$

$$(31)\quad \text{EAVC}(\hat{\bar{\gamma}}_3) = \hat{\sigma}_3^2(\bar{Z}_3'\bar{Z}_3)^{-1},$$

$$(32)\quad \text{EAVC}(\hat{\bar{\gamma}}_4) = \hat{\sigma}_4^2(\bar{Z}_4'\bar{Z}_4)^{-1},$$

$$(33)\quad \text{EAVC}(\hat{\bar{\gamma}}_{24}) = \left(\frac{\bar{Z}_2'\bar{Z}_2}{\hat{\sigma}_2^2} + \frac{\bar{Z}_4'\bar{Z}_4}{\hat{\sigma}_4^2}\right)^{-1},$$

[7] If we do not multiply eqs.(18b)-(18d) by $\sqrt{m'}$, $\sqrt{T'}$ and $\sqrt{m'T'}$ respectively, the variances of the asymptotic distribution of $\sqrt{mT}(\hat{\bar{\gamma}}-\bar{\gamma})$ are not finite, see Wallace and Hussain (1969).

$$(34)\quad \mathrm{EAVC}(\hat{\underline{\gamma}}_{34}) = \left(\frac{\bar{Z}_3'\bar{Z}_3}{\hat{\sigma}_3^2} + \frac{\bar{Z}_4'\bar{Z}_4}{\hat{\sigma}_4^2}\right)^{-1} ,$$

$$(35)\quad \mathrm{EAVC}(\hat{\underline{\gamma}}_{234}) = \left(\frac{\bar{Z}_2'\bar{Z}_2}{\hat{\sigma}_2^2} + \frac{\bar{Z}_3'\bar{Z}_3}{\hat{\sigma}_3^2} + \frac{\bar{Z}_4'\bar{Z}_4}{\hat{\sigma}_4^2}\right)^{-1} .$$

Using the temporal cross-section data, referred to above, we have evaluated the quantities in (22)-(35). The results are listed in Table 1. The sample means are given below:

$$\frac{1}{2107}\sum_{i=1}^{49}\sum_{i=1}^{43} G_{jt} = 384.6068,$$

$$\frac{1}{2107}\sum_{j=1}^{49}\sum_{t=1}^{43} G_{jt-1} = 381.7850,$$

$$\frac{1}{2107}\sum_{j=1}^{49}\sum_{t=1}^{43} X_{jt} = 30.9802,$$

and

$$\frac{1}{2107}\sum_{j=1}^{49}\sum_{t=1}^{43} p_{gjt} = 102.5186.$$

Table 1

Estimates of Parameters appearing in Eq. (15)

Formula used	Quantity Estimated $\bar{\gamma}_1$	$\bar{\gamma}_2$	$\bar{\gamma}_3$	η_g	η_{gg}
(1) $\hat{\bar{\gamma}}$	0.7097 (0.0117)	0.9832 (0.1630)	-0.3201 (0.0622)	0.268	-0.289
(2) $\hat{\bar{\gamma}}_2$	1.0027 (0.0020)	-0.1160 (0.0263)	-0.0059 (0.0516)	-2.009	-0.339
(3) $\hat{\bar{\gamma}}_3$	0.5201 (0.1025)	6.3806 (1.2889)	-0.5078 (0.3725)	1.063	-0.280
(4) $\hat{\bar{\gamma}}_4$	0.5122 (0.0192)	1.5653 (0.2894)	-0.5256 (0.0668)	0.257	-0.285
(5) $\hat{\bar{\gamma}}_{24}$	0.9971 (0.0019)	-0.1326 (0.0260)	-0.1051 (0.0406)	-1.043	-2.738
(6) $\hat{\bar{\gamma}}_{34}$	0.6183 (0.0164)	3.6296 (0.2269)	-0.2874 (0.0621)	0.757	-0.198
(7) $\hat{\bar{\gamma}}_{234}$	0.9978 (0.0019)	-0.1223 (0.0255)	-0.0797 (0.0372)	-1.035	-2.231

The figures in parentheses below each estimated value are the standard errors computed by taking the square root of the diagonal elements of the matrices given in (29)-(35). The income and price elasticities (evaluated at the sample means) are given in the last two columns of Table 1. The estimates of variance functions are

(36) $$\hat{\sigma}_2^2 = 31.745,\ \hat{\sigma}_3^2 = 1450.624 \text{ and } \hat{\sigma}_4^2 = 182.164 .$$

A priori information is needed on the values of ρ_1, ρ_2, ρ_3 and ρ_4 in order to derive the estimates of variance components $\sigma_\mu^2, \sigma_\tau^2$ and σ_v^2 from (36), see Swamy and Mehta (1973). If we employ the a priori values $\rho_\ell = 0 (\ell = 1,2,3,4)$, then $\sigma_2^2 = T\,\sigma_\mu^2 + \sigma_v^2$, $\sigma_3^2 = m\,\sigma_\tau^2 + \sigma_v^2$, and $\sigma_4^2 = \sigma_v^2$. In this case the estimates in (36) imply a negative estimate for σ_μ^2 and a positive estimate for σ_τ^2 which is significantly different from zero. On the other hand, if we employ the a priori values $\rho_1 = \rho_3 = -(m-1)^{-1}$ and $\rho_2 = \rho_4 = -(T-1)^{-1}$, then $\sigma_2^2 = \frac{mT}{(m-1)} \sigma_\mu^2, \sigma_3^2 = \frac{mT}{(T-1)} \sigma_\tau^2, \sigma_4^2 = \frac{mT}{(m-1)(T-1)} \sigma_v^2$ and the estimates of variance components are

$$\hat{\sigma}_\mu^2 = 0.723,\ \hat{\sigma}_\tau^2 = 28.916 \text{ and } \hat{\sigma}_v^2 = 174.296 \ . \tag{37}$$

In the absence of acceptable a priori information about the ρ_ℓ's it is hard to assess the estimates of variance components. The a priori values $\rho_\ell = 0$ which imply a negative estimate for γ are not acceptable to us.

In cases where the ρ_ℓ's are truly 0 one can estimate the variance components by the "fitting-of-constants" method exposited by Fuller and Battese (1974). We do not follow this method for two reasons. (1) Since the individuals in our study are geographical regions with arbitrarily drawn boundaries, we would hardly expect the error components $\mu_1,\ldots,\mu_m$ (or $v_{1t},\ldots,v_{mt}$) to be mutually independent, see Nerlove (1965, p. 160). The assumption that $\rho_1 = \rho_3 = 0$ is of dubious merit or validity. (2) The fitting-of-constants estimators for σ_μ^2 and σ_τ^2 are based on the following residuals.

$$\hat{\underline{\mu}} = (C_2'C_2 \otimes I_T)\{\underline{w} - Z[Z'(C_2'C_2 \otimes I_T)Z]^{-1}Z'(C_2'C_2 \otimes I_T)\underline{w}\}$$

$$\hat{\underline{\tau}} = (I_m \otimes C_1'C_1)\{\underline{w} - Z[Z'(I_m \otimes C_1'C_1)Z]^{-1}Z'(I_m \otimes C_1'C_1)\underline{w}\}$$

where C_1 and C_2 are as defined in (18).

The elements of $\hat{\underline{\mu}}$ (or $\hat{\underline{\tau}}$) are not mutually independent even when the ρ_ℓ's are 0 so that each value of the lagged dependent variable is not independent of the contemporaneous and the succeeding values of μ_j (or τ_t). Consequently, the generalized least-squares estimator of $\bar{\underline{\gamma}}$ with the variance-covariance matrix estimated by the fitting-of-constants method is not consistent.

As exposited in Swamy (1971 , pp. 37-38), the vector $\bar{y}_2$ and the matrix $\bar{Z}_2$ in (18b) consist of observations which are proportional to the means of time series observations for different states. For obvious reasons mentioned in Maddala (1971a, p. 348) the correlation between $\bar{y}_2$ and the first column of $\bar{Z}_2$ is nearly perfect. If we regress $\bar{y}_2$ on $\bar{Z}_2$ the resulting estimate of the coefficient of the lagged dependent variable will be biased toward 1 and the resulting estimates of the coefficients of income and price and the error variance will be biased toward 0. This indicates the spuriousness of estimates given in the second row of Table 1. The estimates of the coefficient of income given in the fifth and seventh rows of Table 1 have an incorrect sign. The reason for this is that in the formulae given in (26) and (28) the low value of σ_2^2 gives an unduly large weight to the observations in eq. (18b). The estimates presented in the second, fifth and seventh rows of Table 1 are not acceptable to us.

It has been shown in Swamy (1971, pp. 71-72) that the vector $\bar{y}_3$ and the matrix $\bar{Z}_3$ consist of observations which are proportional to the means of state observations for different quarters. Both the equations in (7) and (18c) depict aggregate time series observations for the same period, but the source of data in (7) is not the same as that of data in (18c). There are apparent discrepencies between these two data sets because they do not give identical results as can be seen from eq. (8) and the third row of Table 1. The estimated income elasticity given in the third row of Table 1 runs above 1, an implausible value.

The covariance estimates $(\hat{\bar{\gamma}}_4)$ given in the fourth row of Table 1 imply a low value for the income elasticity. Since the estimated values of the income coefficient given in the third and fourth rows of Table 1 are very different, it is not clear whether they can be pooled. If we pool them as in (27), we obtain the estimates given in the sixth row of Table 1.

The Likelihood Approach

The likelihood function of $\bar{Y}$ when the transformed observations in (18a)-(18d) and the initial conditions are taken as data, is

(38)
$$L(\bar{Y},\sigma,\sigma_1^2,\bar{\alpha}|\text{data}) = (2\pi)^{-nT/2}(\sigma_1^2)^{-1/2}(\sigma_2^2)^{-\frac{(m-1)}{2}}(\sigma_3^2)^{-\frac{(T-1)}{2}}(\sigma_4^2)^{-\frac{(m-1)(T-1)}{2}}$$
$$\exp\{-\frac{1}{2}[\frac{\{\sqrt{mT}(\bar{\alpha}-\hat{\bar{\alpha}})+\bar{z}_1'(\bar{Y}-\hat{\bar{Y}}(\sigma))\}^2}{\sigma_1^2}$$
$$+\frac{\nu_2\hat{\sigma}_2^2}{\sigma_2^2}+\frac{(\bar{Y}-\hat{\bar{Y}}_2)'\bar{Z}_2'\bar{Z}_2(\bar{Y}-\hat{\bar{Y}}_2)}{\sigma_2^2}$$
$$+\frac{\nu_3\hat{\sigma}_3^2}{\sigma_3^2}+\frac{(\bar{Y}-\hat{\bar{Y}}_3)'\bar{Z}_3'\bar{Z}_3(\bar{Y}-\hat{\bar{Y}}_3)}{\sigma_3^2}$$
$$+\frac{\nu_4\hat{\sigma}_4^2}{\sigma_4^2}+\frac{(\bar{Y}-\hat{\bar{Y}}_4)'\bar{Z}_4'\bar{Z}_4(\bar{Y}-\hat{\bar{Y}}_4)}{\sigma_4^2}]\}$$

where $\sigma = (\sigma_2^2,\sigma_3^2,\sigma_4^2)$, $\hat{\bar{\alpha}} = \frac{1}{\sqrt{mT}}(\bar{y}_1-\bar{z}_1'\hat{\bar{Y}}(\sigma))$; $\bar{z}_1' = Q_1Z$,

$$\hat{\bar{Y}}(\sigma) = [\frac{\bar{Z}_2'\bar{Z}_2}{\sigma_2^2}+\frac{\bar{Z}_3'\bar{Z}_3}{\sigma_3^2}+\frac{\bar{Z}_4'\bar{Z}_4}{\sigma_4^2}]^{-1}[\frac{\bar{Z}_2'\bar{y}_2}{\sigma_2^2}+\frac{\bar{Z}_3'\bar{y}_3}{\sigma_3^2}+\frac{\bar{Z}_4'\bar{y}_4}{\sigma_4^2}],$$

$\nu_2 = m-4$, $\nu_3 = T-4$ and $\nu_4 = (m-1)(T-1)-3$, see Swamy and Mehta (1973).

If $\rho_1 = \rho_3 = -(m-1)^{-1}$ and $\rho_2 = \rho_4 = -(T-1)^{-1}$, then $\sigma_1^2 = 0$ and the likelihood function gets modified as

(39)
$$L(\bar{Y},\sigma,\bar{\alpha}|\text{data}) = (2\pi)^{-(nT-1)/2}(\sigma_2^2)^{-\frac{(m-1)}{2}}(\sigma_3^2)^{-\frac{(T-1)}{2}}(\sigma_4^2)^{-\frac{(m-1)(T-1)}{2}}$$
$$\exp\{-\frac{1}{2}[\frac{\nu_2\hat{\sigma}_2^2}{\sigma_2^2}+\frac{(\bar{Y}-\hat{\bar{Y}}_2)'\bar{Z}_2'\bar{Z}_2(\bar{Y}-\hat{\bar{Y}}_2)}{\sigma_2^2}$$

$$+ \frac{\nu_3 \hat{\sigma}_3^2}{\sigma_3^2} + \frac{(\bar{\gamma}-\hat{\bar{\gamma}}_3)'\bar{Z}_3'\bar{Z}_3(\bar{\gamma}-\hat{\bar{\gamma}}_3)}{\sigma_3^2}$$

$$+ \frac{\nu_4 \hat{\sigma}_4^2}{\sigma_4^2} + \frac{(\bar{\gamma}-\hat{\bar{\gamma}}_4)'\bar{Z}_4'\bar{Z}_4(\bar{\gamma}-\hat{\bar{\gamma}}_4)}{\sigma_4^2}]\}$$

The "likelihood principle" says that if the assumed model is correct all that the data has to tell us about the parameters is contained in the likelihood function. However, the information which the likelihood function contains is not fully expressed by the estimates in Table 1.[8] It is the whole course of the likelihood function which contains the totality of information coming from the data. The elements of $\bar{\gamma}$ are the parameters of our interest and the elements of σ are the nuisance parameters. Kalbfleisch and Sprott (1970) have given a most wide ranging account of the ways of squeezing likelihood information out of a multiparameter situation. Following their method it can be shown that the "conditional" relative likelihood of $\bar{\gamma}$ is

(40)
$$CR(\bar{\gamma}|\text{data}) = L_1(\bar{\gamma}|\text{data}) / \sup_{\bar{\gamma}} L_1(\bar{\gamma}|\text{data})$$

$$L_1(\bar{\gamma}|\text{data}) = \{1+ \frac{(\bar{\gamma}-\hat{\bar{\gamma}}_2)'\bar{Z}_2'\bar{Z}_2(\bar{\gamma}-\hat{\bar{\gamma}}_2)}{\nu_2\hat{\sigma}_2^2}\}^{-(m-2)/2}$$

$$\cdot\{1+ \frac{(\bar{\gamma}-\hat{\bar{\gamma}}_3)'\bar{Z}_3'\bar{Z}_3(\bar{\gamma}-\hat{\bar{\gamma}}_3}{\nu_3\hat{\sigma}_3^2}\}^{-(T-2)/2}$$

[8] Box and Jenkins (1970, pp. 224-226) point out that mechanical application of the maximum likelihood method might produce nonsensical results.

$$\cdot\{1+\frac{(\bar{\gamma}-\hat{\bar{\gamma}}_4)'\bar{Z}_4'\bar{Z}_4(\bar{\gamma}-\hat{\bar{\gamma}}_4)}{\nu_4\hat{\sigma}_4^2}\}^{-[(m-1)(T-1)-1]/2}$$

It can be seen that (40) allows in the exponent for the loss of precision due to the estimation of all the nuisance parameters. If we replace the exponents $-(m-2)/2$, $-(T-2)/2$, $-[(m-1)(T-1)-1]/2$ in $L_1(\bar{\gamma}|\text{data})$ by $-(m-1)/2$, $-(T-1)/2$ and $-[(m-1)(T-1)]/2$ respectively, we obtain the posterior distribution of $\bar{\gamma}$ corresponding to an improper prior distribution employed by Swamy and Mehta (1973). If we can plot (40) we have summarized almost all the information contained in the data about $\bar{\gamma}$. Unfortunately, there are practical difficulties in plotting a three dimensional function. The function in (40) contains available information about $\bar{\gamma}$ in such a way that information about an element of $\bar{\gamma}$ is inextricably mixed up with the other elements of $\bar{\gamma}$. We cannot integrate $L_1(\bar{\gamma}|\text{data})$ with respect to the elements of $\bar{\gamma}$ to obtain "marginal likelihoods", see Box and Tiao (1970, p. 189). Maddala's (1971b, p. 945) definition of the relative average likelihoods does not make any sense. His use of improper priors to integrate out nuisance parameters from a likelihood function clearly violates the "likelihood principle". Bayesians who employ improper prior distributions are in fact proper in their handling of improper integrals, treating them as limits of any reasonable sequence of proper integrals, see Dempster (1973). Improper distributions are devices for approximating a careful exact assessment that would be too costly to carry out. Though the Kalbfleisch and Sprott definition of conditional likelihood is interesting, it is important to note that the conditional likelihoods do not necessarily satisfy the strong "likelihood principle" that observations in different sample spaces leading to proportional likelihood functions yield identical conclusions.

With some loss of information we can plot second-order likelihood functions defined by Kalbfleisch and Sprott (1970, p. 189). For specified $\bar{\gamma}_2$ and $\bar{\gamma}_3$, the second-order likelihood function of $\bar{\gamma}_1$ is given by $CR(\bar{\gamma}_1|\bar{\gamma}_2,\bar{\gamma}_3$, data). Similarly, for a specified $\bar{\gamma}_1$, the second-order likelihood function of $\bar{\gamma}_2$ and $\bar{\gamma}_3$ is given by $CR(\bar{\gamma}_2,\bar{\gamma}_3|\bar{\gamma}_1$, data). We may use these second-order likelihood functions to quantify the degree of support given to various parameter values by the data. In particular, the quantity $CR(\bar{\gamma}_1|\bar{\gamma}_2,\bar{\gamma}_3$, data) gives a measure of the relative plausibility of a value of $\bar{\gamma}_1$ for any specified pair of values of $\bar{\gamma}_2$ and $\bar{\gamma}_3$. The second-order likelihoods provide much better summary of data than the estimates in Table 1. Our purpose in using likelihood is to give an objective interpretation of the data; whether one believes the implications of the evidence is a subjective judgement, possibly involving many factors such as limitations of the data, specification errors etc.

The second-order likelihoods are evaluated using our data. The function $CR(\bar{\gamma}_1|\bar{\gamma}_2,\bar{\gamma}_3$, data) is plotted in Figure 1 for different values of $(\bar{\gamma}_2,\bar{\gamma}_3)$. It is interesting to note that the shape and location of $CR(\bar{\gamma}_1|\bar{\gamma}_2,\bar{\gamma}_3$, data) are quite insensitive to changes in $(\bar{\gamma}_2,\bar{\gamma}_3)$. The most plausible range of values of $\bar{\gamma}_1$ is 0.5-0.7 which is shorter than the prior interval $0 < \bar{\gamma}_1 < 1$. Thus our data reduced the range of uncertainty regarding the true value of $\bar{\gamma}_1$. Plots of the contours of $CR(\bar{\gamma}_2,\bar{\gamma}_3|\bar{\gamma}_1$, data) are shown in Figure 2. Both the shape and location of these contours change markedly as $\bar{\gamma}_1$ is varied. It can be seen that negative values for the income coefficient and positive values for the price coefficient are definitely plausible with our data. This is in conflict with our theoretical model.

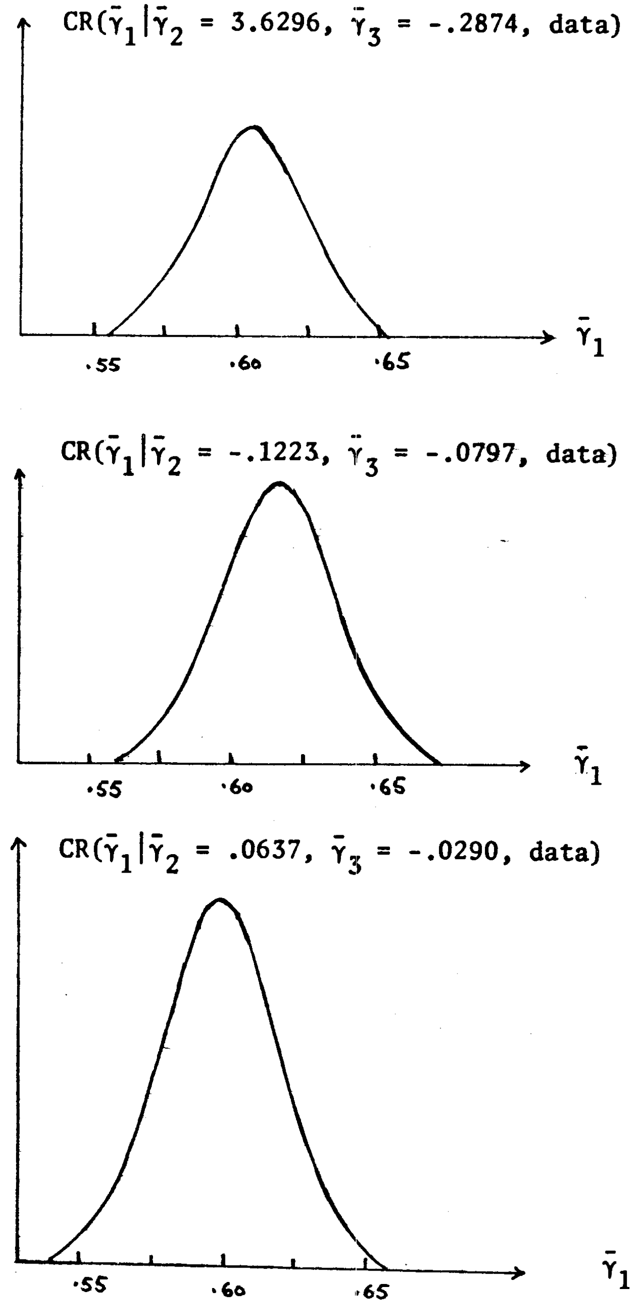

Figure 1: Plots of the second-order likelihood function of $\bar{\gamma}_1$ for different specified values of $\bar{\gamma}_2$ and $\bar{\gamma}_3$[(CR($\bar{\gamma}_1|\bar{\gamma}_2,\bar{\gamma}_3$,data)].

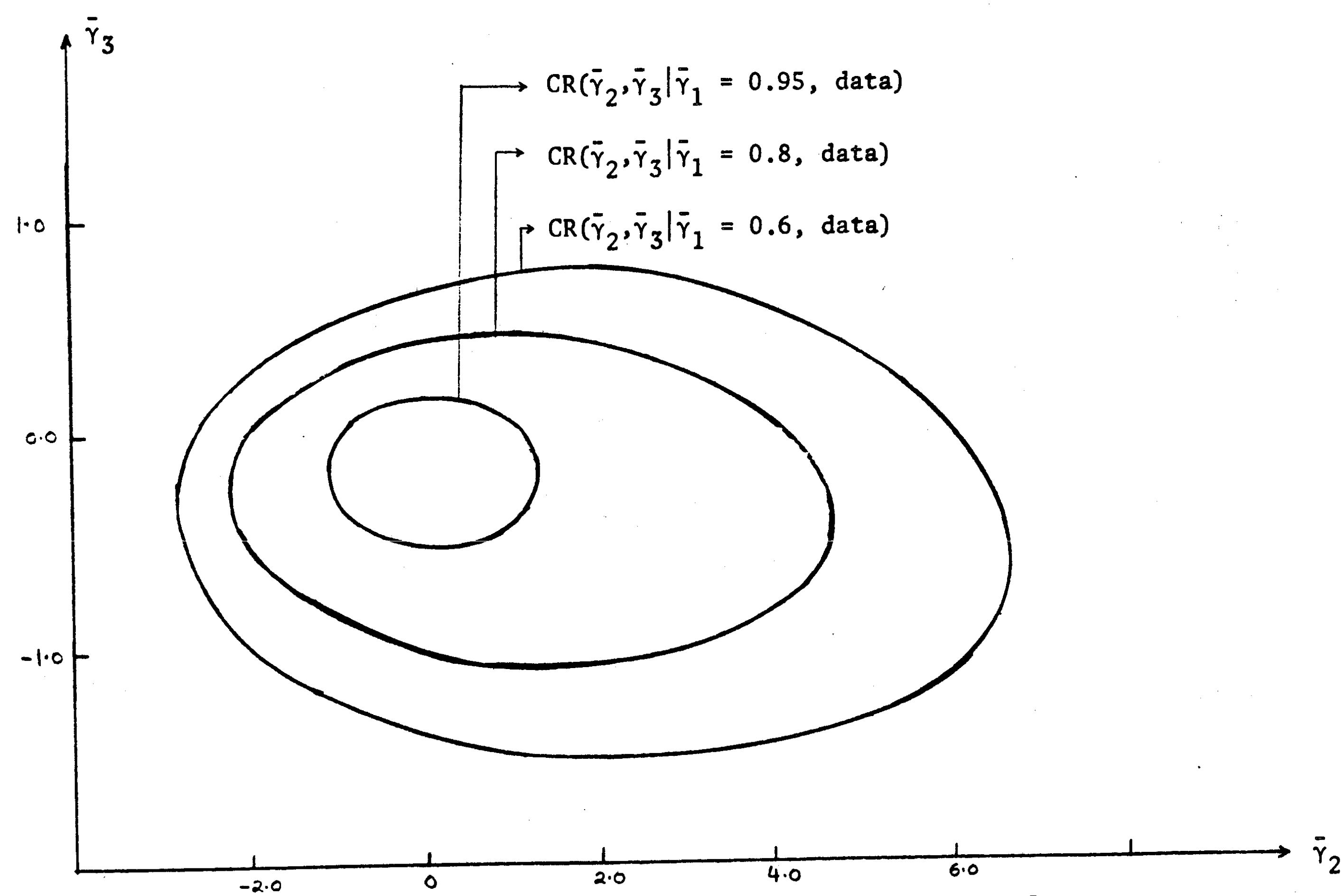

Figure 2: Plots of second-order likelihood function of $\bar{\gamma}_2$ and $\bar{\gamma}_3$ for different specified values of $\bar{\gamma}_1[CR(\bar{\gamma}_2, \bar{\gamma}_3 | \bar{\gamma}_1, data)]$.

4.2 Random Coefficients Model

In this subsection we propose to test explicitly the geographic stability of the slope coefficients of eq.(14). Finding that we cannot accept the hypothesis of geographic stability, we go on to specify and estimate a random coefficients model which permits random variation in all the coefficients across geographic units. Estimates in Table 1 and in Figures 1 and 2 which ignore variations in slope coefficients across states are subject to the type of specification error considered by Zellner (1962a). Hence they are not acceptable to us.

For given j, let $\underline{y}_j$ be a 43 x 1 vector of observations on G_{jt}, Z_j be a 43 x 3 matrix of observations on G_{jt-1}, X_{jt} and p_{gtj}, $\bar{\underline{\gamma}}_j$ be a 3 x 1 vector of slope coefficients and v_j be a 43 x 1 vector of disturbances. For all the time series observations on the j^{th} state together, eq.(14) can be written as

$$\underline{y}_j = \underline{\iota}_T \bar{\alpha}_j + Z_j \bar{\underline{\gamma}}_j + \underline{v}_j \qquad (j = 1,2,\ldots,49) \tag{41}$$

In eq. (14) we set $\tau_t \equiv 0$. Alternatively, we could subsume τ_t into v_{jt} and assume, as in Zellner (1962b), that v_{jt} and $v_{j't}$ for $j \neq j'$, are correlated. If we make this assumption the computational formulae require a matrix inversion of rank 49 x 43, an infeasible computation. We, therefore, assume that v_{jt} and $v_{j't}$ for $j \neq j'$ are independent.

For individual states the least squares estimates of $\bar{\alpha}_j$ and $\bar{\gamma}_j$ are obtained by using only that state's data. The income and price elasticities have been obtained from these parameter estimates by evaluating, at the corresponding sample mean point, the elasticity formulae in (9). All estimates of income coefficients are positive and are significantly different from zero on the convention adopted here, namely a t-ratio greater than two in absolute value. Income elasticities run below 1 for 29 states and above 1 for 20 states. They range from 0.39 to 1.40. Of the 49 estimates of price coefficients, 15 have an incorrect positive sign. Only one of these 15 is significantly different from zero. All price elasticities lie between 0.44 to -0.48. The elasticities show considerable variation across states, uniformity being particularly lacking for the price elasticities. Of the 49 estimates of the coefficient of the lagged dependent variable, 3 have a negative sign in conflict with the theoretical model. These 3 are among the 9 whose values are not significantly different from zero. The remaining 46 estimates which are positive lie between 0.13 and 0.72.

To test for the presence of first-order serial correlation among the disturbances we computed Durbin's (1970) h-statistic for each state. For 3 states h-statistic is not defined. The computed h-statistic exceeds the critical value $\pm$ 1.645 for 17 states, indicating the presence of significant first-order serial correlation. For these 17 states the least squares estimates of $\bar{\alpha}_j$ and $\bar{\underline{\gamma}}_j$ are inconsistent. We cannot use the least squares estimates of ρ_j because they are biased towards zero, see Malinvaud (1970, p.558). To eliminate these inconsistencies we assume that for given j the disturbances are generated by

$$v_{jt} = \rho_j v_{jt-1} + \zeta_{jt} \qquad (42)$$

where $-1 < \rho_j < 1$ and $\zeta_{j1}, \ldots, \zeta_{jT}$ are independently and identically distributed with zero mean and finite variance σ_{jj}. Now we can write eq.(41) as

$$\underline{y}_j = \underline{\iota}_T(1-\rho_j)\bar{\alpha}_j + \rho_j\underline{y}_{j-1} + (Z_j - \rho_j Z_{j-1})\bar{\underline{\gamma}}_j + \underline{\zeta}_j \qquad (43)$$

where $\underline{y}_{j-1}$ is a T x 1 vector of values taken by the dependent variable lagged one period Z_{j-1} is a T x 3 matrix consisting of values taken by the explanatory variables lagged one period and $\zeta_j = (\zeta_{j1},\ldots,\zeta_{jT})'$.

Eq.(43) forms a regression model with nonlinear constraints. We utilized Marquardt's (1963) nonlinear estimation technique to estimate eq.(43) from time series on each of forty-nine states. The following are the results of our nonlinear least squares (NL-LS) estimation.

All of the income coefficients are estimated to be positive. Only two of these are not significantly different from zero. All price coefficients should be negative. But 16 of the 49 are estimated to be positive; all of these 16 are not significantly different from zero. The remaining 33 have the correct negative sign, but only 4 of them are significantly different from zero. The estimate of the coefficient of the lagged dependent variable is less than 1 in absolute value for all states but it turned out to be negative for 24 states. Only 9 of these negative estimates are significantly different from zero. Negative sign for the coefficient of the lagged dependent variable is in conflict with the theoretical model. Of the remaining 25 estimates which are positive, 17 are significantly different from zero. Thirty seven of the 49 estimates of the ρ_j's are significantly different from zero.

The nonlinear least squares (NL-LS) and linear least squares (L-LS) results might be compared to determine whether one of the two sets of results has a clear advantage in terms of significance of estimates and correct signs. Both the NL-LS and L-LS estimates of 15 ρ_j's are significantly different from zero. The NL-LS estimates of 21 ρ_j's are significantly different from zero(i.e., the estimates are larger in absolute value than twice their standard errors) whereas the L-LS estimates of the same 21 ρ_j's are not significantly different from zero on "h" test. The L-LS estimates of 2 ρ_j's are significantly different from zero on h test but the NL-LS estimates of the same 2 ρ_j's are less in absolute value than twice their standard errors. The NL-LS estimates of 24 coefficients of the lagged dependent variable have an incorrect sign and the L-LS estimates of only 3 coefficients of the lagged dependent variable have an incorrect sign. The problem of distinguishing between equation dynamics and autocorrelation of the error term remains challenging.

Let $\hat{\rho}_j$ and s_{jj} denote the NL-LS estimates of ρ_j and σ_{jj} respectively. If $\bar{\alpha}_j = \bar{\alpha}$ and $\bar{\underline{\gamma}}_j = \bar{\underline{\gamma}}$ for every j, an approximate generalized least squares estimator of $\bar{\underline{\beta}} = (\bar{\alpha}, \bar{\underline{\gamma}}')'$ is

$$(44) \qquad \hat{\underline{\beta}} = \left[\sum_{j=1}^{49} \frac{(X_j' \hat{\Omega}_{jj}^{-1} X_j)}{s_{jj}} \right]^{-1} \left[\sum_{j=1}^{49} \frac{(X_j' \hat{\Omega}_{jj}^{-1} \underline{y}_j)}{s_{jj}} \right] ,$$

where $X_j = [\iota_T, Z_j]$ is a T_j x 4 matrix, $\hat{\Omega}_{jj} = \frac{1}{(1-\hat{\rho}_j^2)} \{\hat{\rho}_j^{|t-t'|}\} (t,t' = 1,2,\ldots,T_j)$.

In (44) we set $\hat{\rho}_j$ at 0 if the NL-LS estimate of ρ_j is not significantly different from zero.

In order to examine the restriction that the 49 equations in (43) are homogeneous in so far as their coefficient vectors are concerned, the NL-LS estimates of $\bar{\alpha}_j$ and $\bar{\underline{\gamma}}_j$ derived from 49-state models can be combined with the estimate in (44) to test the hypothesis that

$$(45) \qquad H_0: \ \bar{\alpha}_j = \bar{\alpha} \ \text{ and } \ \bar{\underline{\gamma}}_j = \bar{\underline{\gamma}} \ \text{ for } \ j = 1,2,\ldots,49.$$

Now let $\hat{\underline{\beta}}_j = (X_j' \hat{\Omega}_{jj}^{-1} X_j)^{-1} X_j' \hat{\Omega}_{jj}^{-1} \underline{y}_j$ where X_j is T_j x 4 matrix. An argument parallel to that of Zellner (1962b) and Durbin (1970) shows that, under the null hypothesis, the distribution of

(46) $$\frac{1}{(m-1)4}\sum_{j=1}^{m}(\hat{\bar{\beta}}_j - \hat{\bar{\beta}})'\frac{X_j'\hat{\Omega}_{jj}^{-1}X_j}{s_{jj}}(\hat{\bar{\beta}}_j - \hat{\bar{\beta}})$$

is approximately F with $(m-1)4$ and $(T-4)m$ degrees of freedom where T is the minimum of $T_1,\ldots,T_m$. For our temporal cross-section data the value of the statistic in (46) is 454.86. This value is well above the 5 percent significance point 1.17 of F(192,1862) leading to a rejection of the hypothesis in (45).

We now estimate 49-state equations in (43) without forcing the coefficient vector to be the same for all states. Suppose that the finite set of vectors $(\bar{\alpha}_1,\bar{\gamma}_1')',\ldots,(\bar{\alpha}_{49},\bar{\gamma}_{49}')$ is equivalent to a random sample from a hyperpopulation. This assumption is usually made in applying asymptotic distribution theory to samples from finite populations, see Hajek (1960). The justification of this assumption in terms of "exchangeability" is given in Lindley (1971,pp.38-40).

<u>Assumption 5</u>:

(a) Let $\bar{\beta}_j = (\bar{\alpha}_j,\bar{\gamma}_j')'$. Then $\bar{\beta}_1,\ldots,\bar{\beta}_m$ are independently and identically distributed with mean vector $\bar{\bar{\beta}}$ and positive finite variance-covariance matrix $\bar{\Delta}$;

(b) $\bar{\beta}_j$ is independent of $y_{j'}$ and $Z_{j'}$ for every $j,j' = 1,2,\ldots,m$.

(c) The ρ_j and σ_{jj} are fixed parameters.

Now we follow the procedure outlined in Swamy (1974, p.157) to estimate $\bar{\bar{\beta}}$ and $\bar{\Delta}$ with our data. The following results were obtained.

(47) $$\hat{\bar{\bar{\beta}}} = \left[\sum_{j=1}^{m}\{\hat{\bar{\Delta}} + s_{jj}(X_j'\hat{\Omega}_{jj}^{-1}X_j)^{-1}\}^{-1}\right]^{-1}$$

$$\cdot\left[\sum_{j=1}^{m}\{\hat{\bar{\Delta}} + s_{jj}(X_j'\hat{\Omega}_{jj}^{-1}X_j)^{-1}\}^{-1}\hat{\bar{\beta}}_j\right].$$

(48) $$\hat{\bar{\bar{\beta}}} = \begin{bmatrix} \hat{\bar{\bar{\alpha}}} & \hat{\bar{\bar{\gamma}}}_1 & \hat{\bar{\bar{\gamma}}}_2 & \hat{\bar{\bar{\gamma}}}_3 \\ 56.6909 & 0.1460 & 9.2093 & -0.1414 \\ (14.9282) & (0.0450) & (0.5810) & (0.0914) \end{bmatrix}'$$

where $\bar{\bar{\alpha}}$ is the mean of $\bar{\alpha}_j$, $\bar{\bar{\gamma}}_1$ is the mean of $\bar{\gamma}_{j1}$, $\bar{\bar{\gamma}}_2$ is the mean of $\bar{\gamma}_{j2}$ and $\bar{\bar{\gamma}}_3$ is the mean of $\bar{\gamma}_{j3}$. The figures in parentheses are the asymptotic standard errors computed by taking the square root of the diagonal elements of the matrix

(49) $$\left[\sum_{j=1}^{m} \{ \hat{\bar{\Delta}} + s_{jj} (X_j' \hat{\Omega}_{jj}^{-1} X_j)^{-1} \}^{-1} \right]^{-1}.$$

Here

(50) $$\hat{\bar{\Delta}} = \frac{1}{m-1} \left[\sum_{j=1}^{m} \hat{\bar{\beta}}_j \hat{\bar{\beta}}_j' - \frac{1}{m} \sum_{j=1}^{m} \hat{\bar{\beta}}_j \sum_{j=1}^{m} \hat{\bar{\beta}}_j' \right]$$

$$\hat{\bar{\Delta}} = \begin{bmatrix} \bar{\alpha}_j & \bar{\gamma}_{1j} & \bar{\gamma}_{2j} & \bar{\gamma}_{3j} \\ 10185.23 & 5.6877 & -154.5582 & -20.5043 \\ & 0.0921 & -0.6559 & -0.0164 \\ & & 14.9178 & -0.4183 \\ & & & 0.2863 \end{bmatrix}$$

The estimates in (48) have the expected sign, yielding the following income and price elasticities (evaluated using (9)).

(51) $$\hat{\eta}_g = 0.8675 \qquad \hat{\eta}_{gg} = -0.0441$$

An estimate of -0.04 for the price elasticity appears to be reasonable. An estimate of income elasticity of 0.87 appears to be high. This may be due to the inclusion of tax payments in the personal incomes, since we could not get data on personal disposable income by states. The income and price elasticities obtained by Houthakker et al. (1974) are not strictly comparable to those of the present study in so far as different measures of income and price were used in the two studies.

Houthakker et al.(1974) obtained the estimates of personal disposable income by states by multiplying each state's personal income by the ratio of U.S. disposable to U.S. personal income. This procedure does not give accurate measures of personal disposable income for different states. Finally, Houthakker et al. deflated the gasoline prices by a consummer price index. For reasons mentioned in Pollak and Wachter (1975) this should not be done.

For given j the correlations between the regressors and the ζ_{jt} in eq.(43) may not be as strong as those between the regressors and the disturbances of the aggregate equation in (7). If this is true the inconsistencies of the estimates in (47) will not be as great as those of the estimates in (8). Further, since $\hat{\bar{\underline{\beta}}}$ in (47) is a matrix weighted average of the $\hat{\bar{\underline{\beta}}}_j$'s, it may happen that the inconsistency of $\hat{\bar{\underline{\beta}}}$ is smaller than that of any individual $\hat{\bar{\underline{\beta}}}_j$. The estimates of each $\bar{\underline{\beta}}_j$ might be further improved by taking explicit account of the possibility of "simultaneous equation" complications. In any case, lack of inconsistency is not necessarily a very important desirable property of an estimator in small sample situations.

5. Conclusions and Discussion

We have found that the aggregate estimates in (10) and the estimates in Table 1 and in Figures 1 and 2 are based on certain assumptions which are not valid for our data. In particular, Assumption 1(b) and the assumption that the slope coefficients of eq.(14) are the same for all states are unrealistic. The estimates of policy-relevant elasticities presented in (51) are based on a general set of assumptions which are considerably weaker than those adopted either in sections 3 and 4.1 of the present study or in other studies on the demand for gasoline.

The results of this paper caution against the customary practice of committing ab initio to one set of assumptions concerning the variability of coefficients across states. One should not lock himself into a committed position and let data do some work in finding good specifications. It is important to examine every aspect of a model as thoroughly as possible. For this purpose, procedures which squeeze the data almost dry should be used. Assumptions concerning the stability of structural parameters across geographic regions are just as important as assumptions concerning the dynamic and autocorrelation structures of a model. Much attention is given to the latter but not to the former in the econometric literature. Good procedures for analyzing the dynamic and autocorrelation structures of a model are given in Zellner and Palm (1974). The present study shows that a combination of these procedures with those in Swamy (1971,1974) yields fruitful results.

Appendix

Sources of Aggregate Data

G_t : percapita consumption of gasoline by individuals in gallons obtained from, DATA RESOURCES, INC., LEXINGTON, MASSACHUSETTS. Data is aggregated from individual State Data for the continental United States.

X_t : per capita personal income in 1967 dollars obtained from BUREAU OF ECONOMIC ANALYSIS , U.S. DEPARTMENT OF COMMERCE, WASHINGTON, D.C.

p_{gt}: price indices for motor gasoline with 1967 = 100 obtained from, DATA RESOURCES INC., LEXINGTON, MASSACHUSETTS. Computed as a weighted average of price indices be individual States in the Continental United States Sources of Temporal Cross-Section data.

G_{jt}: per capita consumption of gasoline by individuals in different states in gallons obtained from DATA RESOURCES INC., LEXINGTON, MASSACHUSETTS.

X_{jt}: per capita personal income by states in 1967 dollars obtained from DATA RESOURCES INC., LEXINGTON, MASSACHUSETTS

P_{gtj}: price indices for motor gasoline with 1967 = 100 obtained from DATA RESOURCES, INC., LEXINGTON, MASSACHUSETTS.

References

Balestra, P. and M. Nerlove (1966): "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas,"Econometrica, 34,pp. 585-612.

Box, G.E.P. and G.M. Jenkins (1970): Time Series Analysis, Forecasting and Control, San Francisco: Holden-Day.

_______________ and G.C. Tiao(1973): Bayesian Inference in Statistical Analysis, Reading Massachusetts: Addison Wesley Publishing Company.

Dempster, A.P. (1973): "Discussion on the paper by Dr. David, Professor Stone and Dr. Zidek," Journal of the Royal Statistical Society, Series B, 35, pp. 222-3.

Durbin, J. (1970): "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors are Lagged Dependent Variable," Econometrica, 38, pp. 410-421.

Fuller, W.A. and G.E. Battese (1974): "Estimation of Linear Models with Cross-Error Structure," Journal of Econometrics, 2, pp. 67-78.

Goldberger, A.S.(1964): Econometric Theory, New York: John Wiley & Sons.

Hajek, J.(1960): "Limiting Distributions in Simple Random Sampling From a Finite Population, "Publication of the Mathematical Institute of Hungarian Academic Science, 5,pp.361-374.

Hill, B.M. (1969): "Correlated Errors in the Random Model," Journal of the American Statistical Association, 62, pp.1387-1400.

Houthakker, H.S., P.K. Verleger, Jr. and D. Sheehan(1974): "Dynamic Demand Analysis For Gasoline and Residential Electricity," American Journal of Agricultural Economics, 56, pp.412-418.

Kalbfleisch, J.D. and Sprott, D.A. (1970): "Application of Likelihood Methods to Models Involving Large Number of Parameters," Journal of the Royal Statistical Society, Series B, 32,pp.175-208.

Klein, L.R. (1962): An Introduction to Econometrics, Englewood Cliffs, New Jersey: Prentice-Hall,Inc.

Lindley, D.V. (1971): Bayesian Statistics, A Review, Philadelphia: SIAM.

Maddala, G.S. (1971a): "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," Econometrica, 39, pp.341-358.

_______________(1971b): "The Likelihood Approach to Pooling Cross-Section and Time Series Data," Econometrica, 39, pp.939-954.

Malinvaud, E. (1970)" Statistical Methods of Econometrics, Second revised Edition, Amsterdam: North-Holland Publishing Company.

Marquardt, D.W. (1963): "An Algorithm for Least Squares Estimation of Non-Linear Parameters," Journal of the Society for Industrial and Applied Mathematics, 11, pp.431-441.

Nerlove, M.(1965): Estimation and Identification of Cobb-Douglas Production Functions, Chicago: Rand McNally & Company.

Pollak, R.A. and M.L. Wachter (1975): "The Relevance of the Household Production Function and its Implications for the Allocation of Time," Journal of Political Economy, 83, pp. 255-278.

Ramsey, J., R. Rasche and B. Allen (1974): "An analysis of the Private and Commercial Demand for Gasoline," Department of Economics, Michigan State University, mimeographed.

Swamy, P.A.V.B. (1971): Statistical Inference in Random Coefficient Regression Model, Berlin, Heidelberg, New York: Springer-Verlag.

____________(1974): "Linear Models with Random Coefficients," Chapter V in P. Zarembka, ed., Frontiers in Econometrics, New York: Academic Press.

____________and S. S. Arora (1972): "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models," Econometrica, 40, pp. 261-275.

____________and J.S. Mehta (1973): "Bayesian Analysis of Error Components Regression Models," Journal of the American Statistical Association, 68, pp.648-658.

____________ and J.S. Mehta (1975): "Bayesian and Non-Bayesian Analysis of Switching Regressions and of Random Coefficients Regression Model," Journal of the American Statistical Association, forthcoming.

Theil, H. (1971): Principles of Econometrics, New York: John Wiley & Sons.

Verleger, Jr., P.K. and D. Sheehan (1973): A Study of the Quarterly Demand for Gasoline and Impacts of Alternative Gasoline Taxes, A Report prepared for the Environmental Protection Agency and the Council on Enviromental Quality, Data Resources, Inc., Lexington, Massachusetts.

Wallace T.D. and Hussain, A. (1973): "The Use of Error Components Models in Combining Cross-Section with Time Series Data," Econometrica, 37, pp.55-72.

Zellner, A. (1962a): "Estimation of Cross-Section Relations: Analysis of a Common Specification Error," Metroeconomica, 14, pp. 11-117.

____________(I962b): "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Agregation Bias," Journal of the American Statistical Association, 57, pp. 348-368.

____________and F. Palm (1974): "Time Series Analysis and Simultaneous Equation Econometric Models," Journal of Econometrics, 2, pp.17-54.