POOLING CROSS SECTION AND TIME SERIES DATA IN THE ESTIMATION OF A DYNAMIC MODEL: THE DEMAND FOR NATURAL GAS

BY PIETRO BALESTRA AND MARC NERLOVE

In this paper, we consider two basic aspects of demand analysis, with application to the demand for natural gas in the residential and commercial market. The more fundamental one consists in the formulation of a demand function for commodities—such as natural gas—whose consumption is technologically related to the stock of appliances. We believe that in such markets, the behavior of the consumer can be described best in terms of a dynamic mechanism.

Related to this is the more specific problem of estimating the parameters of the demand function, when the demand model is cast in dynamic terms and when observations are drawn from a time series of cross sections.

Accordingly, this paper is centered around these two major themes, although, as the title suggests, the emphasis is placed on the second one. In Section 1, we present the theoretical formulation of the dynamic model for gas. In Section 2, the results of the estimation of the gas model by ordinary least squares methods are presented. These results, together with more fundamental theoretical considerations, suggest a different approach. The essence of this approach, which is not restricted to the gas model, is discussed in Section 3, while two alternative procedures for estimating the coefficients of the dynamic model in the light of this new approach are proposed in Section 4. It is subsequently shown that the application of these procedures to the gas data produces results that are reasonable on the basis of a priori theoretical considerations.

1. THE DYNAMIC MODEL OF GAS DEMAND

While it is true that natural gas is not a durable commodity—i.e., a commodity that is enjoyed repeatedly over a length of time or that may be stored for future use—yet it is also true that the consumption of gas, at least at the household level, is closely related to the stock of gas appliances in existence and that to a large extent it is governed by such stocks. It follows that the approach to the problem should be a dynamic one and that the demand function should incorporate a stock effect and some assumptions about the adjustment of these stocks over time.

The approach underlying the dynamic model developed in this paper is consideration of the demand in the new market for gas, i.e., the incremental demand for gas (inclusive of demand due to the replacement of gas appliances and the

---

1 The research underlying this paper was performed in part under Grant NSF GS-142 from the National Science Foundation to Stanford University. We are indebted to Sherman H. Clark of Stanford Research Institute for his help in formulating the underlying framework in this study, to Zvi Griliches and L. G. Telser for pointing out a major error in a first draft, and to T. Amemiya, D. Blackwell, E. J. Hannan, H. Konijn, E. Kuh, C. R. Rao, and A. Zellner for helpful comments. Of course, responsibility for remaining errors is our own.

2 Apart from theoretical considerations, actual experimentation with a static model yielded insignificant or even positive price elasticities in most cases.
demand due to net increases in the stock of such appliances). The rationale of this approach is that, given the particular characteristics of the gas market, it is unrealistic to assume that the consumer's choice is determined in the manner suggested by traditional (static) demand theory. In the gas market, for instance, a short-run change in the relative price of gas does not induce many consumers to revise choices once made, because of the high transfer costs involved in the shift to a different type of fuel. More specifically, once a major appliance is installed in a home, there is little or no substitution between the different types of fuels. And because the demand for space heating is probably highly inelastic, we would expect a very low short-run price elasticity of demand from current gas consumers.

In the planning stage, on the other hand, the relative price of the different fuels surely has some effect on the decision making process and, at least at this stage, we should observe a pattern of behavior consistent with traditional demand theory. Therefore, the demand function considered here must be understood in an ex ante sense. It describes the behavior of a consumer not committed by past contracts to any particular form of technique (or type of service).

We assume that the new demand for gas, \( G^* \), is a function of the relative price of gas, \( p \), and the total new requirements for all types of fuel, \( F^* \). In symbols

\[
G^* = f(p, F^*).
\]

Note that equation (1) implies that the price variable has an effect primarily on the rate of change of gas consumption rather than on its absolute levels.

The problem is to define the concepts of new demand and new market and incorporate them into a model expressed in terms of observable variables.

Let \( F_t \) be the demand for all types of fuel in period \( t \). The increment in total fuel consumption between any two periods is given by

\[
\Delta F_t = F_t - F_{t-1}.
\]

The quantity \( \Delta F_t \) represents the change in total fuel demand between period \( t \) and period \( t-1 \), but it does not express the total new demand for fuels. The reason is

The same notion has been applied in the study of investment decisions, a field in many respects similar to the one presently under investigation. See P. J. Dhyre and M. Kurz, "Technology and Scale in Electricity Generation," *Econometrica*, Vol. 32, No. 3 (July, 1964), pp. 287-315.

Two different definitions of the price variable were tried: (1) the price of gas deflated by the consumer price index, and (2) the price of gas deflated by the price of the substitutes. Since the results are quite similar (although a different interpretation must necessarily apply), only those corresponding to the first definition are reported here.

One of the referees has questioned the connection between the demand for fuel and the demand for a particular sort of fuel, namely gas, which is expressed by (1). The rationale for considering the demand for a component of an aggregate category will become clearer as we progress. It is, essentially, that the demand for gas is a derived demand, derived from the demand for space heating. Total fuel demand is a surrogate for the total demand for space heating; equation (1) thus, in more conventional terms, simply expresses the demand for a "factor of production" as a function of its price relative to other factors and "output."
that not all of the demand prevailing in period \((t-1)\) is also committed in period \(t\), as some of the installations that existed in period \((t-1)\) are retired during the course of the year, because of obsolescence or simple wear and tear.

To clarify this matter, consider explicitly the stock of appliances. Let \(W_{t-1}\) be the average stock of appliances in period \((t-1)\) and \(\lambda_{t-1}\) the rate of utilization of these appliances in the same period. Then, by definition

\[
F_{t-1} = \lambda_{t-1} W_{t-1}.
\]

Of the stock of appliances \(W_{t-1}\) (assuming a constant, proportional rate of depreciation \(r\) in the aggregate), only \((1-r)W_{t-1}\) will be present in period \(t\) and, given the rate of utilization \(\lambda_t\) will be associated with a fuel consumption of

\[
\lambda_t(1-r)W_{t-1}.
\]

The quantity in (4) expresses the portion of fuel consumption that in period \(t\) is associated with the stock of appliances already in existence at the beginning of the period (and not yet retired).

In period \(t\), the average stock of appliances will be \(W_t\). This stock of appliances is associated with a total fuel consumption of

\[
F_t = \lambda_t W_t.
\]

The new demand for fuel, \(F^*\), may then be defined as the difference between the total demand for fuel and the "committed" demand for fuel:

\[
F^*_t = \lambda_t W_t - (1-r)\lambda_t W_{t-1}.
\]

The rate of utilization, \(\lambda_t\), is not likely to vary violently from one period to the next. On the contrary, it may be safely assumed for gas, that the rate of utilization is relatively constant, because a high (and therefore constant) efficiency of combustion is easily attained for gas fuels.\(^6\) It may be true for other types of fuel that the rate of utilization is affected by technological progress. It is unlikely, however, for the relatively short time period encompassed in this study that this effect causes appreciable variation in the rate of utilization. We, therefore, assume

\[
\lambda_t = \lambda, \quad \text{all } t.
\]

Equation (6) may now be expressed in terms of fuel variables alone, since

\[
\lambda_t W_{t-1} = \lambda_{t-1} W_{t-1} = F_{t-1}.
\]

Thus, the total new demand for fuel becomes

\[
F^*_t = F_t - (1-r)F_{t-1}.
\]

It is clear that the total new demand for fuel as defined here is larger than the

\(^6\) We, of course, abstract from variations in weather conditions. It should be emphasized in this connection that all fuel variables have been normalized for weather changes in the statistical analyses here reported.
incremental change in fuel consumption, \( F_t - F_{t-1} \). Equation (9) may be rearranged as follows:

\[
F_t^* = (F_t - F_{t-1}) + rF_{t-1} .
\]

The total new demand for fuel then appears as the sum of the incremental change in consumption (the terms in parentheses) and "replacement" demand (expressed by \( rF_{t-1} \)). The quantity \( rF_{t-1} \) represents that portion of the total demand for fuel "freed" by the retirement (and replacement) of old appliances.

By an analogous argument, the new demand for gas, \( G_t^* \), may be defined as

\[
G_t^* = G_t - (1-r_g)G_{t-1} .
\]

The depreciation rate for gas appliances, \( r_g \), is not necessarily the same as the depreciation rate appropriate to fuel consuming appliances generally. Two factors account for this discrepancy. First, the life-time of a given installation (appliance) using a particular type of fuel may be different from the lifetime of another installation using a different type of fuel. Second, the rate of depreciation for an aggregate stock is clearly related to the average age of the stock. The average age of the stock of appliances is certainly lower in a new market (such as gas for instance) than in an older one (such as coal).

Assuming linearity,\(^7\) equation (1) may be rewritten

\[
G_t^* = \beta_0 + \beta_1 p_t + \beta_2 F_t^* ,
\]

or, equivalently,

\[
G_t = \beta_0 + \beta_1 p_t + \beta_2 [F_t - (1 - r)F_{t-1}] + (1-r_g)G_{t-1}.
\]

Equation (13) portrays the basic dynamic mechanism of the model. The parameters \( r \) and \( r_g \) need not be known a priori, since estimates of them are provided by (13) itself.

Total fuel consumption, however, is not given a priori, as fuel consumption itself is a function of several other variables. On the basis of theoretical considerations and actual experimentation (both in this study and in others) it was found that total fuel consumption may be well approximated by an equation of the form

\[
F_t = \gamma_0 + \gamma_1 N_t + \gamma_2 Y_t ,
\]

where \( N_t \) and \( Y_t \) stand, respectively, for total population and per capita income. Price effects are found to be negligible. Substituting (14) in (13), we obtain

\[
G_t = \alpha_0 + \alpha_1 p_t + \alpha_2 \Delta N_t + \alpha_3 N_{t-1} + \alpha_4 \Delta Y_t + \alpha_5 Y_{t-1} + \alpha_6 G_{t-1} .
\]

Equation (15) is in a form suitable for estimation. The implicit parameter \( r \),

\(^7\) The assumption of linearity is necessary in order to insure appropriate identification of the parameters. Certain nonlinearities, however, might be introduced later in the estimating equation. A few which were tried did not change the results appreciably.
however, is now overidentified, since estimates of it are provided by both ratios $\alpha_3/\alpha_2$ and $\alpha_5/\alpha_4$. A constrained maximization is thus necessary. This may be accomplished by defining

\begin{equation}
N_t^* = N_t - (1-r) N_{t-1},
\end{equation}

\begin{equation}
Y_t^* = Y_t - (1-r) Y_{t-1}.
\end{equation}

Equation (15) becomes

\begin{equation}
G_t = \alpha_0 + \alpha_1 p_t + \alpha_2 N_t^* + \alpha_4 Y_t^* + \alpha_5 G_{t-1}.
\end{equation}

Equation (17) is now estimated for values of $r$ in the admissible interval and the value so chosen as to maximize the likelihood function (which in this case is the same as maximizing the $R^2$).

2. ESTIMATION BY ORDINARY LEAST SQUARES METHODS

Before turning to a discussion of estimation proper, we discuss two subsidiary questions: (1) the type of data to which the model developed is applied; and (2) the question of the identification of the demand equation, which involves a discussion of the conditions under which gas is supplied.

The investigation reported here is based on data by state and covers the period 1950–1962.\footnote{These data were compiled by Stanford Research Institute on the basis of published statistics and other information available to the Institute.} We thus think of the demand relationship as applicable to a "representative" consumer at the state level. All fuel variables are expressed in quantities ($10^{12}$ Btu) and are normalized for weather changes within each state and between states. The price variables are given in cents per million Btu and are deflated by the consumer price index. The population variable is expressed in thousands and, finally, the per capita income variable is given in constant 1961 dollars.

Thirteen observations are available on each state (only 36 states, however, have enjoyed gas consumption over the entire time span). In order to obtain an adequate number of degrees of freedom, estimation by individual states must be abandoned. Instead, all observations are grouped together and estimation (though not necessarily by ordinary least squares) is performed on the combined sample of cross section and time series.\footnote{A similar set of data (a combination of cross section and time series data by states) was also used in a recent study of the demand for electricity by F. M. Fisher and C. Kaysen, The Demand for Electricity in the United States, A Study in Econometrics, Contributions to Economic Analysis, (Amsterdam: North-Holland Publishing Co., 1962).}

The problem of identifying the demand relation when considering a market involving both demand and supply relationships is a serious one in studies of the type reported here. Nonetheless, it may plausibly be argued that regulation of the
gas market by Federal and state commissions coupled with the heavily capitalized nature of the gas industry suggest that a recursive system of Wold’s type may not be an unrealistic representation of the working of the gas market, at least in the residential and commercial sector. Furthermore, because sales of natural gas are made to the industrial sector on an interruptible basis (and thus represent a buffer for possible expansion in the residential-commercial sector) and because of the existence of excess capacity (which is typically generated in the expansion process of a heavily capitalized industry), it is plausible to consider the supply of gas to the residential-commercial sector as nearly perfectly elastic. However, in order to insure a perfect elasticity of supply (which is a property holding only for sufficiently short periods in this market), it is necessary to divide the time period under investigation into two technologically different periods corresponding to two stages of development: an innovating stage and a mature stage. During the latter period (1957–62), all states included in the sample are reasonably homogeneous as far as gas availability is concerned, and the assumption of perfectly elastic supply is approximated. Only the results for this latter period are reported here.

Bearing these considerations in mind, let us examine various estimates of the gas equation. Parameter estimates corresponding to equation (15) are shown in line 1 of Table I, while those corresponding to equation (17) are shown in line 2 (for the maximizing value of i).

In both cases, the estimated coefficient of the lagged endogenous variable is above one, a result that is incompatible with theoretical expectations as it implies a negative depreciation rate for gas appliances. It may be argued that this result is due to partial incorporation of a time trend in the coefficient of Gt - 1. However, the explicit introduction of a time effect into the equation (by either a time trend or dummy variables for years) did not change the results appreciably. Similarly, other attempts to incorporate special factors such as the difference in weather conditions did not yield statistically significant effects. On the other hand, the estimation rate of depreciation for all fuel consuming appliances (11 per cent) is not unreasonable.

Under certain restrictive conditions, one could argue that the coefficient of Gt - 1 may be very close to unity. Such might be the case, for example, because of inertia factors, the relatively young age of the stock of gas appliances, and because gas apparently is not losing its own replacement market. Then, the variable Gt - 1 may be shifted to the left-hand side of the equation. The estimated coefficients for both equations (15) and (17) corresponding to this model are shown in lines 3 and 4 of Table I.11

It should be remembered that the demand function used here refers to the new

---

11 The R² given here are not those for the regressions actually performed but reflect the “explanation” of Gt rather than Gt - Gt - 1. The R² appropriate for the latter variable as dependent were 0.400 and 0.398, respectively.
### TABLE I

**Various Estimates of the Parameters of the Gas Model from the Pooled Sample, 1957–62**

<table>
<thead>
<tr>
<th>Based on</th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\alpha_3$</th>
<th>$\alpha_4$</th>
<th>$\alpha_5$</th>
<th>Maximizing value of $r$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equation (15)</td>
<td>-3.650</td>
<td>-0.451</td>
<td>0.174</td>
<td>0.0111</td>
<td>0.0183</td>
<td>0.00326</td>
<td>1.010</td>
<td>—</td>
</tr>
<tr>
<td>No constraints</td>
<td>(3.316)</td>
<td>(0.0270)</td>
<td>(0.0093)</td>
<td>(0.00041)</td>
<td>(0.0080)</td>
<td>(0.00197)</td>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td>Equation (17)</td>
<td>-2.300</td>
<td>-0.3887</td>
<td>0.1149</td>
<td>—</td>
<td>0.01945</td>
<td>—</td>
<td>1.012</td>
<td>11%</td>
</tr>
<tr>
<td>No constraints</td>
<td>(2.295)</td>
<td>(0.02518)</td>
<td>(0.00302)</td>
<td>(0.00751)</td>
<td>(0.0135)</td>
<td>(0.0135)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equation (15)</td>
<td>-3.654</td>
<td>-0.566</td>
<td>0.0187</td>
<td>0.0135</td>
<td>0.0189</td>
<td>0.00371</td>
<td>1.0</td>
<td>—</td>
</tr>
<tr>
<td>$\alpha_6$ constrained at 1 (3.312)</td>
<td>(0.0221)</td>
<td>(0.0091)</td>
<td>(0.00025)</td>
<td>(0.0079)</td>
<td>(0.00188)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equation (17)</td>
<td>-2.066</td>
<td>-0.0111</td>
<td>0.0128</td>
<td>—</td>
<td>0.0206</td>
<td>—</td>
<td>1.0</td>
<td>12%</td>
</tr>
<tr>
<td>$\alpha_6$ constrained at 1 (2.260)</td>
<td>(0.0212)</td>
<td>(0.0012)</td>
<td>(0.0072)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equation (15)</td>
<td>—</td>
<td>-0.026</td>
<td>-0.135</td>
<td>0.0327</td>
<td>0.0131</td>
<td>0.0044</td>
<td>0.6799</td>
<td>—</td>
</tr>
<tr>
<td>Including Dummy Variables for States</td>
<td>(0.0532)</td>
<td>(0.0215)</td>
<td>(0.0046)</td>
<td>(0.0084)</td>
<td>(0.0101)</td>
<td>(0.0633)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Figures in parentheses are standard errors of the corresponding coefficients above.
demand for gas. Therefore, the average price elasticity, \( \bar{\varepsilon} \), is

\[
\bar{\varepsilon} = \frac{\partial G^*}{\partial p} \frac{\bar{\hat{p}}}{G^*},
\]

where \( G^* = G_t - (1-r_p)G_{t-1} \) is the relevant consumption variable and not \( G_t \). The estimated average elasticity of new gas demand for the various forms listed in Table I ranges from .58 to .69. But the average price elasticity appears to be increasing over time. A higher price elasticity, for instance, is obtained for the period 1957–62 than for the period 1950–56. Furthermore, when only the data for the years 1961–62 are considered, the average price elasticity is above unity.

One possible explanation for the results thus far obtained is that, when cross section and time series data are combined in the estimation of a regression equation, certain "other effects" may be present in the data.\(^1\)\(^2\) A natural way to account for these "other effects" is to introduce explicitly into the equation individual shift variables. The rationale of this procedure is that the data contain an additive effect specific to the individual (state). To account for such effects, dummy variables corresponding to the 36 different states may be introduced explicitly into the model. It is moot, however, whether the dummy variable method is appropriate in the case of a dynamic model. The presence of lagged endogenous variables may make it difficult, if not impossible, to separate the individual (state) effects from the effect induced by the lagged variable.

The results of including 36 dummy variables in equation (15) are shown in line 5 of Table I.\(^3\)\(^3\) As expected, the estimated coefficient of the lagged endogenous variable is drastically reduced, but is reduced to such a low level that it implies a depreciation rate of gas appliances of over 30 per cent—highly implausible.\(^4\)\(^4\) The coefficient of the price variable, on the other hand, is increased.

Despite its limitations, the dummy variable experiment suggests that the coefficient of the lagged consumption variable (in the general pooled model) may indeed reflect in part a regional effect rather than a true lag effect. The problem with dummy variables is that they appear to reflect too much and thus to reduce the coefficient of the lagged gas variable to too low a level.\(^5\)\(^5\) The idea of a regional (individual) effect, however, may be introduced in an alternative fashion.


\(^2\) Note that the overall constant term has been omitted, obviating the necessity for constraining the 36 state coefficients to sum to zero.

\(^3\) E. Kuh has pointed out that regarding the coefficient of \( G_{t-1} \) as a superposition of an ordinary distributed lag and the depreciation effect (rather than as a pure depreciation effect as we have done) renders the very low value of the estimated coefficient plausible. Inasmuch as we have expressed the model in terms of the "new" demand for gas, however, it is difficult to see why any additional distribution of lag should arise. Indeed, what we have actually done is to build a model which has a distributed lag effect as a consequence.

\(^4\) Another way of looking at this problem is to say that the use of dummy variables wastes
3. THE RESIDUAL MODEL

The problem of estimating dynamic relationships from a time series of cross sections is more general than the material in the preceding sections would suggest. For this reason it seems best to discuss the problem in a somewhat more general setting. In the present section, therefore, we develop in general terms a model that includes as special cases both ordinary time series and ordinary cross section analysis, along lines suggested originally by E. Kuh. In Section 4 we present two alternative estimation procedures, together with applications to the gas model. A modification of the second procedure is also suggested but not applied to the gas model.

Suppose we have observations on $N$ individuals, $n = 1, 2, ..., N$, taken over $T$ periods of time, $t = 1, 2, ..., T$. Although in this study we are concerned with states, the individuals may be firms, consumers, regions, or any other statistical entity. We denote the variable to be explained by $y_n$. This variable we assume is explained by $K$ truly exogenous variables $z^{(1)}_{nt}, ..., z^{(K)}_{nt}$, including one that is identically one, so that no constant term need be included, and $\theta$ lagged values of the dependent variable. It is, incidentally, the presence of such lagged values which produces the essential difficulty of the problem, and which distinguishes it from the type of problem discussed in recent econometric literature.

To express the relation to be estimated in matrix form, we may write:

$$ y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \\ \vdots \\ y_N \\ \vdots \\ y_{NT} \end{bmatrix} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT} \end{bmatrix}, $$

an $NT \times 1$ vector.

$$ Z = \begin{bmatrix} z^{(1)}_{11} & \cdots & z^{(K)}_{11} \\ \vdots & \ddots & \vdots \\ z^{(1)}_{NT} & \cdots & z^{(K)}_{NT} \end{bmatrix}, $$
degrees of freedom since we are not really interested in the values of their coefficients but only in the coefficients of lagged gas consumption, price, income, etc. What we imply is that boiling this effect down to one parameter as we do in the next section is somehow more efficient. Clearly, however, this is a small sample question, since in large samples when the number of observations increases faster than the number of regions it cannot really matter how many parameters are estimated (as long as there are only a finite number). Unfortunately, there is no small sample theory for stochastic difference equations of the sort estimated here, and it is thus not possible to prove the assertion of greater efficiency. Monte Carlo experiments may, however, yield evidence on the matter and will be attempted by one of the authors at a future date.

an $NT \times K$ matrix of purely exogenous variables.

$$Y_{\theta} = \begin{bmatrix} y_{1-1} & \cdots & y_{1-\theta} \\ \vdots & & \vdots \\ y_{1T-1} & y_{1T-\theta} \\ \vdots & & \vdots \\ y_{NT-1} & \cdots & y_{NT-\theta} \end{bmatrix},$$

an $NT \times \theta$ matrix of lagged values of the dependent variable.

$$u = \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix} = \begin{bmatrix} u_{11} \\ \vdots \\ u_{NT} \end{bmatrix},$$

an $NT \times 1$ vector of residuals. We also let $\alpha$ be a $K \times 1$ vector of constant coefficients of the exogenous variables, and $\beta$ a $\theta \times 1$ vector of coefficients for the lagged endogenous variables. Throughout much of the discussion, we shall further simplify by writing $X = [Z: Y_{\theta}]$, an $NT \times K + \theta$ matrix, and

$$\gamma = \begin{bmatrix} \alpha \\ \beta \end{bmatrix},$$

a $K + \theta \times 1$ vector.

The relationship to be estimated from the combination of time series and cross section data then takes the form:

$$y = Z\alpha + Y_{\theta}\beta + u = X\gamma + u.$$  \hspace{1cm} (19)

Of course, the statistical properties of various estimates, and indeed the real meaning of the model, are determined by what we assume about the properties of the residual vector $u$. We assume that each residual $u_{nt}$ may be decomposed into two statistically independent parts: an individual effect and a remainder. Thus,

$$u_{nt} = \mu_n + \nu_{nt}.$$  \hspace{1cm} (20)

We could also allow for a separate time effect, say $\lambda_t$, but this would greatly complicate the analysis without adding any essential generality. As indicated, the random variables $\mu_n$ and $\nu_{nt}$ are assumed to have zero means, and to be independent.\footnote{The assumption of zero means, especially for $\nu_{nt}$, is a rather important one. If $E\nu_{nt} = m_n \neq 0$, it is necessary to introduce certain shift variables for $N-1$ of the individuals into the matrix $Z$. There should be $n-1$ of these variables, if an overall constant term is allowed, one for all values of $t$ in the $n$th block, and zero otherwise, except for the $N$th block. If such variables are introduced, we may define new residuals with components $v_{nt}$ which do have zero mean. Assuming no shift variables of this sort, however, is not the same thing as the assumption of zero variance, $\sigma^2_{\nu}$, i.e., $\varnothing = 0$. In this case, $\Omega = \sigma^2 I$, and ordinary least squares estimates of $\alpha$ and $\beta$ would be appropriate despite the presence of lagged values of the dependent variable (under certain standard conditions, of course). One might be tempted to assume zero variance, $\sigma^2_{\nu}$, if shift variables had been included in the $Z$ matrix. In Section 3 of this study, however, it was found that such an assumption led to economically implausible results. Hence it seems better to adopt the more general mode.}
from which it follows that

\[(21) \quad E\mu_n \nu_{nt} = 0, \quad \text{all} \ n, t.\]

We further assume that there is no serial correlation among the \(\nu_{nt}\), and that these are independent from one individual to another; thus,

\[(22) \quad E\nu_n \nu_{n't'} = \begin{cases} \sigma_n^2, & n = n' \text{ and } t = t', \\ 0, & \text{otherwise}. \end{cases}\]

Similarly,

\[(23) \quad E\mu_n \mu_{n'} = \begin{cases} \sigma_n^2, & n = n' \\ 0, & \text{otherwise}. \end{cases}\]

Perhaps the most dubious assumption is that \(\mu_n\) are independent of themselves for different individuals. For example, if the individuals are geographical regions with arbitrarily drawn boundaries, as they are here, we would not expect this assumption to be well satisfied. Nonetheless, we shall adopt it in what follows.

Equations (21) imply that the variance-covariance matrix of the residuals \(u\) may be written

\[(24) \quad E(u'u') = \Omega = \sigma^2 \begin{bmatrix} A & 0 & \ldots & 0 \\ 0 & A & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & A \end{bmatrix}\]

where

\[(25) \quad E(u_n'u_n') = \sigma^2 A = \sigma^2 \begin{bmatrix} 1 & \rho & \ldots & \rho \\ \rho & 1 & \ldots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & 1 \end{bmatrix}\]

is a \(T \times T\) matrix, and where

\[(26) \quad \begin{cases} \sigma^2 = \sigma_n^2 + \sigma^2, \\ \rho = \sigma_n^2 / \sigma^2. \end{cases}\]

Note that in a pure cross section, there is only one time period for each individual, say the first, and we have only the upper left-hand corner one of each \(A\); hence, in this case, \(\Omega = \sigma^2 I\), where \(I\) is an \(N \times N\) identity matrix. Furthermore, if there is only one individual, we cannot, in principle, ever distinguish between "individual" and "remainder" effects; hence, no generality is lost by assuming \(\mu_n = 0\) and thus \(\sigma^2 = 0\). In this case, then, too, \(\Omega = \sigma^2 I\), where \(I\) is a \(T \times T\) identity matrix, and ordinary least squares would, under fairly general circumstances, be an appropriate method of estimation.

Were no lagged values of the dependent variable \(y\) included among the explanatory variables in (19), i.e., if \(\beta\) were identically zero, ordinary least squares estimates of the coefficients, \(\alpha\), would be unbiased and consistent under the usual assumptions.
They would not, however, be minimum variance or, in general, asymptotically efficient. Furthermore, the standard estimates of the variance-covariance matrix of the estimated coefficients, \( \hat{\sigma} \), say, would be biased and inconsistent. If the true variance-covariance matrix of the residuals \( u, \Omega \), were known up to a multiplicative constant—in this case \( \sigma^2 \)—the minimum-variance, linear, unbiased estimators of \( \alpha \) are given by

\[
\hat{\alpha} = [Z' \Omega^{-1} Z]^{-1} Z' \Omega^{-1} y. 
\]

An unbiased estimate of \( V(\hat{\alpha}) = E(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)' \) is

\[
\hat{\sigma}^2 \Omega^* = \Omega \quad \text{and} \quad \hat{\alpha} = y - Z\hat{\alpha}. 
\]

Provided we assume that the \( u \)'s are distributed according to a multivariate normal distribution with zero means and variance-covariance matrix \( \sigma^2 \Omega^* \), \( \hat{\alpha} \) given by (27) is the maximum-likelihood estimate of \( \alpha \), and

\[
\hat{\theta}^2 = \frac{\hat{\alpha}' \Omega^{*-1} \hat{\alpha}}{NT} 
\]

is the maximum-likelihood estimate of \( \sigma^2 \). Under quite restrictive conditions these estimates are (i) functions of sufficient statistics, (ii) consistent, (iii) asymptotically normal, and (iv) efficient.\(^{19}\)

Unfortunately, however, \( \Omega \) is not known up to a multiplicative constant, and the question is whether it may be possible to derive estimators having some or all of the desirable properties of the least squares, or the maximum-likelihood estimates by estimating \( \Omega \) up to a multiplicative factor.\(^{20}\) In our case, this simply involves estimating the single parameter \( \rho \), namely, the fraction of the overall residual variance accounted for by individual differences, as \( \Omega \) depends only on this parameter and on \( \sigma^2 \), which can be estimated for known \( \rho \) by (29). The case in which no lagged endogenous variables appear in the regression has been discussed by Zellner\(^{21}\) and Telser.\(^{22}\) Zellner's idea is simply to estimate by two stages: first, using the ordinary least squares estimates of \( \alpha \), and under suitable restrictions on


the form of $\Omega$, he obtains estimates of the variances and covariances of the residuals. Since the ordinary least squares estimates of $\alpha$ are consistent under general conditions, the estimates of the variances and covariances of the $u$'s are also. Thus, second, the matrix $\Omega$ may be replaced in (27) by a consistent estimate and new estimates of $\alpha$ derived. Zellner discusses the asymptotic properties of such two-stage estimators and shows that the gain in efficiency depends on the values of the off-diagonal elements in $\Omega$ (in our case, the extent to which $\rho$ differs from zero), and on the correlation of the independent variables for the different individuals (blocks). If the independent variables for each individual are perfectly correlated, Zellner's results show that if no shift variables are included, the asymptotic efficiency of the ordinary least squares estimators is the same as that of the proposed two-stage estimators. Such perfect correlation will rarely be the case, however.

Telser develops alternative computational procedures for estimates of the general type proposed by Zellner, and shows that these are best, asymptotically normal (BAN) estimators in much the same way that we show a related type of estimator is asymptotically maximum likelihood below.\(^{24}\)

The difficulty with the Zellner two-stage procedure and the Telser iterative procedure in connection with the present problem is the necessity of beginning with consistent estimates of the regression coefficients. Only in this way can consistent estimates of the residual variances and covariances be obtained, and such are essential if desirable asymptotic properties of the final estimates are to be achieved. When lagged endogenous variables are included among the explanatory variables of $y$ in (19), it is no longer true that the ordinary least squares estimates of $\gamma = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ are consistent unless there is no serial correlation of any kind. In this case, of course, there is no possibility either of increasing the efficiency of the estimates by any sort of iteration.

There are two possible approaches to this problem, both of which are explored in Section 5. First, we may try to find estimators of the coefficients $\gamma$ in (19) which are consistent despite the presence of lagged endogenous variables. We can then obtain a consistent estimate of $\rho$ and, using $\Omega$ so determined, find

\begin{equation}
\gamma = [X' \Omega^{-1} X]^{-1} X' \Omega^{-1} y,
\end{equation}

which we show below will have desirable properties. Alternatively, since the structure of our problem is much simpler than the more general one considered by

\(^{23}\) Zellner's model is different from the one developed in this study in two important respects. First, Zellner postulates different sets of regression coefficients for each (cross section) equation and zero restrictions on the coefficients occurring in other equations. Second, Zellner does not assume constancy of the parameter $\theta$ for different years.

\(^{24}\) Zellner himself shows that these estimates are asymptotically normal and have the same asymptotic distribution as the minimum variance unbiased estimates. A discussion of BAN estimators and their properties is given in connection with the discussion of our estimates below.
Zellner and Telser by virtue of the fact that $\Omega^*$ depends only on a single parameter $\rho$, we may try to obtain simultaneous maximum-likelihood estimates for $\gamma$, $\sigma^2$, and $\rho$. These will have all the desirable asymptotic properties we seek, and an easily determined asymptotic variance-covariance matrix as well.

4. ESTIMATION PROCEDURES

Maximum-likelihood estimates. We suppose, as is usual in such discussions, that the values of $y_{nt}$ for $t \leq 0$ and all $n$ are fixed. The $z_{nt}^{(2)}$ are assumed to be non-stochastic. If the $u_{nt}$ are distributed according to a multivariate normal distribution with zero means and variance-covariance matrix $\Omega$, their probability density may be written

\[
p(u) = (2\pi)^{-NT/2} |\Omega^{-1}|^{\frac{1}{2}} e^{-\frac{1}{2} u' \Omega^{-1} u}.
\]

Consequently, using $u$ as a shorthand notation for $y - X\gamma$, the logarithmic likelihood function for the parameters, $\gamma$, $\rho$, and $\sigma^2$ may be written

\[
L(\gamma, \rho, \sigma^2) = -\frac{NT}{2} \log 2\pi - \frac{1}{2} \log |\Omega| - \frac{1}{2} u' \Omega^{-1} u.
\]

The maximum of $L$ with respect to $\gamma$ is clearly obtained when the quadratic form

\[
u' \Omega^{-1} u = [y - X\gamma]' \Omega^{-1} [y - X\gamma]
\]

is at a minimum. This occurs for

\[
\hat{\gamma} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y.
\]

$\hat{\gamma}$ is the maximum-likelihood estimate of $\gamma$ for $\Omega$ known up to a scalar multiple. For unknown $\Omega$, (33) is one of a series of equations which must be solved simultaneously to find the maximum-likelihood estimate of $\gamma$, $\rho$, and $\sigma^2$.

The appropriate equation for $\sigma^2$ may be found by writing $\Omega = \sigma^2 \Omega^*$, differentiating $L$ with respect to $\sigma^2$, and setting the result to zero. This yields

\[
\sigma^2 = \frac{[y - X\gamma]' \Omega^{*-1} [y - X\gamma]}{NT}.
\]

It may be verified that the second derivative of $L$ with respect to $\sigma^2$ is negative at this point as required for a maximum, given $\gamma$ and $\rho$.

It is a trifine more difficult to find the final equation of those determining the maximum-likelihood estimates. The derivation of the final expressions for $\rho$ and

---

25 This is a standard result which follows from the fact that $X' \Omega^{-1} X$ is positive definite.

26 Considering a model in which the variance-covariance matrix of the residuals is identical to our $\Omega$, M. Halperin states: "The maximum likelihood equations ... are of such a formidable character that an explicit solution does not appear possible" (p. 574). However, Halperin then shows that, for the case of no lagged endogenous variables, the least squares estimators and certain tests of significance are still valid. In his proofs, Halperin makes use of the same orthogonal transformation as the one adopted below. It should be pointed out, however, that Halperin does
\( \sigma^2 \) given by maximizing the likelihood function is contained in the Appendix to this paper where we find that the maximum-likelihood estimates for \( \rho \) and \( \sigma^2 \) are given by

\[
(35) \quad \rho = \frac{\sum_{n=1}^{N} \left[ \sum_{i=1}^{T} u_{ni} \right]^2 - \sum_{i=1}^{T} u_{ni}^2}{(T-1) \sum_{n=1}^{N} \sum_{i=1}^{T} u_{ni}^2},
\]

\[
(36) \quad \sigma^2 = \frac{\sum_{n=1}^{N} \sum_{i=1}^{T} u_{ni}^2}{NT}.
\]

The reader should bear in mind that the \( u \)'s appearing in (35) are merely shorthand expressions for the appropriate combination of \( \gamma \)'s, \( y \)'s, and \( x \)'s.\textsuperscript{27}

Equations (33) and (35) determine the maximum-likelihood estimates of \( \gamma \), \( \rho \), and \( \sigma^2 \). Unfortunately, they are highly nonlinear in the \( \gamma \)'s and \( \rho \). While there are numerical methods for solving such systems of nonlinear equations, the computational burden is very great. Fortunately the interpretation of the parameter \( \rho \) as the ratio of a component of variance to an overall variance, which forces \( \rho \) to lie in the interval [0,1], suggests a much simpler procedure. For any \( \rho \) lying in this interval we may determine those values of \( \gamma \) and \( \sigma^2 \) which maximize the likelihood function. As indicated, these values will be given by equations (33) and (34).

Inserting the values so obtained in the likelihood function, we obtain a partially maximized likelihood, say \( \hat{L} \), which we may write purely as a function of \( \rho \):

\[
(36) \quad \hat{L}(\rho) = -\frac{NT}{2} \left[ 1 + \log 2\pi \right] - \frac{NT}{2} \log \delta^2(\rho) - \frac{N}{2} \log \left\{ (1 - \rho)^{T-1} \left[ (1 - \rho) + T\rho \right] \right\}
\]

where \( \delta^2(\rho) \) is given by (34) when \( \gamma \) is chosen, as in (33), for a particular value of \( \rho \). It has been written with a hat and as a function of \( \rho \) to remind us that it is both an optimized value and dependent on \( \rho \). Since \( \rho \) varies over the closed interval, \( \hat{L}(\rho) \) must reach a maximum within the interval or on the boundary, provided it is continuous throughout the interval. Except, possibly, on a set measure of zero, \( \hat{L}(\rho) \) is indeed continuous on the half-open interval [0,1); at \( \rho = 1 \), however, \( \Omega \) is singular and \( \hat{L}(\rho) \) is therefore undefined.

It follows from the above remarks that if \( \hat{L}(\rho) \) reaches a maximum within the

\textsuperscript{27} The expression for \( \sigma^2 \) in (35) differs from that in (34) because the maximum-likelihood estimate of \( \sigma^2 \), given \( \gamma \), has been inserted.
interval \( [0,1] \), we can find the value of \( \rho \) for which this occurs by computing \( \hat{L}(\rho) \) numerically for a sufficient number of points \( \rho \) within the interval. Unfortunately, however, there is no guarantee, for any particular observed \( y \) and \( X \), that the maximum of \( \hat{L}(\rho) \) so defined actually lies within \([0,1] \). It is clear, too, from (35) that there is no guarantee either that \( \rho \) as given there in terms of the \( u_n \) will lie in the interval \([0,1] \), although \( \sigma^2 \) must be positive, as it should be.

We have thus come to the following conclusion: if, for given observations on \( y \) and \( X \), the likelihood function \( \hat{L}(\gamma, \rho, \sigma^2) \), defined in (32) has a maximum with respect to \( \gamma \), \( \rho \), and \( \sigma^2 \) such that the maximizing value of \( \rho \) lies between zero and one, the procedure of maximizing \( \hat{L}(\rho) \), defined in (36), will yield that maximum.\(^{28}\)

If, however, the maximum of \( \hat{L}(\gamma, \rho, \sigma^2) \) does not occur for a value of between zero and one, the maximum-likelihood procedure outlined in this section is inapplicable, and alternative methods of estimation must be employed.

The application of the suggested procedure to the gas data reveals, indeed, that the ML method may lead in practice to inadmissible results. In Table II, the column labelled \( L^* \) gives the values of \( \hat{L}(\rho) \)—omitting the constant term—for 20 values of \( \rho \). The figures show that the likelihood function does not reach a maximum in the interval of the \( a \) priori admissible values of \( \rho \). Hence, the ML method should be abandoned and an alternative procedure adopted.

It is interesting to note, however, the pattern displayed by the coefficient of the lagged consumption variable. This coefficient gradually declines as \( \rho \) increases; for a sufficiently high value of \( \rho \) it becomes smaller than unity.

**Alternative Estimates.** The desirable properties of maximum-likelihood estimates, apart from the intrinsic appeal of the method and the fact that they are functions of sufficient statistics, are all asymptotic: consistency, asymptotic normality, and asymptotic efficiency. But as we have just seen by example, “Blind adherence to the principle of maximum likelihood ... may lead to more difficult computations and still yield less accurate estimates than other methods of estimation.”\(^{29}\)

\(^{28}\) The asymptotic variance-covariance matrix of the maximum-likelihood estimators, \( \hat{\gamma} \), \( \hat{\rho} \), and \( \hat{\sigma}^2 \) satisfying (33) and (35) simultaneously is given by the inverse of the matrix of second derivatives of \( L \), defined in (32), evaluated at \( \gamma = \hat{\gamma} \), \( \sigma^2 = \hat{\sigma}^2 \), \( \rho = \hat{\rho} \):

\[
\begin{bmatrix}
\frac{\partial^2 L}{\partial \gamma^2} & \frac{\partial^2 L}{\partial \gamma \partial \rho} & \frac{\partial^2 L}{\partial \gamma \partial \sigma^2} \\
\frac{\partial^2 L}{\partial \rho \partial \gamma} & \frac{\partial^2 L}{\partial \rho^2} & \frac{\partial^2 L}{\partial \rho \partial \sigma^2} \\
\frac{\partial^2 L}{\partial \sigma^2 \partial \gamma} & \frac{\partial^2 L}{\partial \sigma^2 \partial \rho} & \frac{\partial^2 L}{\partial \sigma^2^2}
\end{bmatrix}^{-1}.
\]

See Kendall and Stuart, op. cit., p. 55.

assessed the same desirable asymptotic properties as maximum-likelihood estimates: consistency, asymptotic normality, and efficiency, but more optimal computational or small sample properties. He called estimates with these large sample properties Best Asymptotically Normal (BAN) estimates.

### TABLE II

**Estimated Value of the Likelihood Function for 20 Values of \( \theta \), 1957–62**

<table>
<thead>
<tr>
<th>Assumed Value of ( \theta )</th>
<th>( L^* )</th>
<th>Const.</th>
<th>( P_t )</th>
<th>Regression Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( N_{t-1} )</td>
</tr>
<tr>
<td>.00</td>
<td>-58.663</td>
<td>-3.650</td>
<td>-.0451</td>
<td>.00111</td>
</tr>
<tr>
<td>.05</td>
<td>-58.786</td>
<td>-3.582</td>
<td>-.0480</td>
<td>.00119</td>
</tr>
<tr>
<td>.10</td>
<td>-58.918</td>
<td>-3.517</td>
<td>-.0509</td>
<td>.00127</td>
</tr>
<tr>
<td>.15</td>
<td>-59.052</td>
<td>-3.455</td>
<td>-.0536</td>
<td>.00136</td>
</tr>
<tr>
<td>.20</td>
<td>-59.185</td>
<td>-3.396</td>
<td>-.0564</td>
<td>.00145</td>
</tr>
<tr>
<td>.25</td>
<td>-59.319</td>
<td>-3.339</td>
<td>-.0591</td>
<td>.00155</td>
</tr>
<tr>
<td>.30</td>
<td>-59.453</td>
<td>-3.286</td>
<td>-.0617</td>
<td>.00165</td>
</tr>
<tr>
<td>.35</td>
<td>-59.587</td>
<td>-3.240</td>
<td>-.0644</td>
<td>.00177</td>
</tr>
<tr>
<td>.40</td>
<td>-59.724</td>
<td>-3.200</td>
<td>-.0670</td>
<td>.00190</td>
</tr>
<tr>
<td>.45</td>
<td>-59.863</td>
<td>-3.172</td>
<td>-.0697</td>
<td>.00204</td>
</tr>
<tr>
<td>.50</td>
<td>-60.007</td>
<td>-3.161</td>
<td>-.0724</td>
<td>.00220</td>
</tr>
<tr>
<td>.55</td>
<td>-60.157</td>
<td>-3.174</td>
<td>-.0751</td>
<td>.00238</td>
</tr>
<tr>
<td>.60</td>
<td>-60.316</td>
<td>-3.227</td>
<td>-.0779</td>
<td>.00258</td>
</tr>
<tr>
<td>.65</td>
<td>-60.487</td>
<td>-3.338</td>
<td>-.0807</td>
<td>.00282</td>
</tr>
<tr>
<td>.70</td>
<td>-60.674</td>
<td>-3.545</td>
<td>-.0837</td>
<td>.00311</td>
</tr>
<tr>
<td>.75</td>
<td>-60.884</td>
<td>-3.913</td>
<td>-.0860</td>
<td>.00347</td>
</tr>
<tr>
<td>.80</td>
<td>-61.128</td>
<td>-4.568</td>
<td>-.0902</td>
<td>.00393</td>
</tr>
<tr>
<td>.85</td>
<td>-61.427</td>
<td>-5.794</td>
<td>-.0942</td>
<td>.00460</td>
</tr>
<tr>
<td>.90</td>
<td>-61.824</td>
<td>-8.361</td>
<td>-.0996</td>
<td>.00571</td>
</tr>
<tr>
<td>.95</td>
<td>-62.446</td>
<td>-15.464</td>
<td>-.1101</td>
<td>.00827</td>
</tr>
</tbody>
</table>

* Constant \((-N\theta/2) \log 2\pi\) suppressed.

Although there are a number of methods currently available for deriving BAN estimates, in this context the simplest procedure is to find a set of estimates asymptotically equivalent to the maximum-likelihood estimates (AML). It will then follow that such estimates have all the desirable asymptotic properties of

---


maximum-likelihood estimates except perhaps efficiency. We continue to assume, as before, that the residuals \( u_m \) are distributed according to (31).

Suppose that \( m = NT \to \infty \), and that if \( N \to \infty \) then it does so in such a way that \( N/T \to 0 \). Let us further suppose we are given a consistent estimate \( S_m \) of the residual variance-covariance matrix \( \Omega \). Thus, as \( T \to \infty \), \( S_m \to \Omega \) in probability, and \( S_m^{-1} \to \Omega^{-1} \). However, \( \Omega^{-1} \) depends upon \( T \) and can be shown to be block diagonal with blocks of the form

\[
C' \text{ diag } \left[ \frac{1}{\xi}, \frac{1}{\eta}, \ldots, \frac{1}{\eta} \right] C = \begin{bmatrix} \frac{1}{\xi} & 0 & \cdots & 0 \\ 0 & \frac{1}{\eta} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\eta} \end{bmatrix} \begin{bmatrix} e' \sqrt{T} \\ C' \\ C_i' \\ 1/\eta \end{bmatrix}
\]

\[
= \frac{ee'}{\xi T} + \frac{C_i' C_i}{\eta} = \frac{ee'}{\xi T} + \frac{1}{\eta} \left( I - \frac{ee'}{T} \right)
\]

\[
\to \frac{1}{\eta} I = \frac{1}{\sigma^2 (1 - \rho)} I,
\]

as \( m \to \infty \), by equation (A.6) of the Appendix. The matrices \( C_i, C_i' \) and the variables \( \xi, \eta \) are also defined there.

It follows that replacing \( \Omega \) by a consistent estimate of itself in (30) leads asymptotically to the ordinary least squares estimates based on the pooled sample and that these are the same in the limit as the maximum-likelihood estimates.\(^{32}\) The estimates so obtained may very well have more desirable small sample properties than the direct maximum-likelihood estimates although we have not, other than empirically, succeeded in demonstrating these. Furthermore, \( [X' S_m^{-1} X]^{-1} \) will be a consistent estimate of the asymptotic variance-covariance matrix of these estimates. It follows then that our only problem is to obtain a consistent estimate of the residual variance-covariance matrix \( \Omega \). Once this is obtained, we can proceed to a "second round" estimate,

\[
(37) \quad \gamma = [X' S_m^{-1} X]^{-1} X' S_m^{-1} y.
\]

If there were no lagged values of \( y \) present in the \( X \) matrix (i.e., if \( \beta = 0 \)), moments of the calculated residuals, \( \hat{u} = y - Z(Z' Z)^{-1} Z' y \), from the ordinary least squares regression would be consistent estimates of the corresponding elements of the matrix \( \Omega \). Hence, as Telsier has shown, under appropriate restrictions on the form of \( \Omega \) sufficient to enable one to compute estimates of its elements from moments of the calculated residuals, the second-stage estimates are BAN.\(^{33}\) Unfortunately,

\(^{32}\) The assumption of normality, of course, makes the proof that our estimates are AML quite trivial. It is possible, however, to derive AML estimates in cases where normality is not assumed but then rather tedious regularity conditions must be checked. In essence one must show asymptotic normality; we here assume normality from the start.

\(^{33}\) Telsier, op. cit.
however, there will generally be lagged values of \( y \) present among the independent variables of the regression equation (19); under these circumstances, the ordinary least squares estimates of the coefficients are inconsistent estimates of the elements of \( \Omega \) from moments of the calculated residuals from the ordinary least-squares regression.

Suppose, on the other hand, that consistent estimates of \( \gamma \) were available. The calculated residuals from (19) could then be used to compute, for example by (35), consistent estimates of \( \rho \) and \( \sigma^2 \). These in turn would form the matrix \( S_m \) to be used in a second round producing estimates of \( \gamma \). To ensure, however, that the estimated \( \rho \) would lie between zero and one, we might wish to depart somewhat from the estimates given in (35). The following, for example, would also lead to consistent estimates of \( \rho \):

\[
(38) \quad \rho = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T} a_{nt} a_{nt'} - \frac{1}{N} \left( \sum_{n=1}^{N} \sum_{t=1}^{T} a_{nt} \right)^2}{NT^2 \sigma^2}
\]

where \( \sigma^2 \) is given by the second equation in (35) and \( a_{nt} \) is a residual calculated from (19) by inserting consistent estimates for \( \gamma \).

There are a variety of ways in which one may obtain consistent estimates of \( \gamma \). We shall discuss only two here, one of which is inspired by Theil’s two-stage least squares, the second method has been suggested by E. J. Hannan. Only the first of the two, however, can be applied to the gas model.

Equation (20) suggests that the residual \( u_{nt} \) may be divided into two parts: one, \( \mu_n \), represents the individual effect common, for a given individual \( n \), to all time periods for which we observe him. The other, \( v_{nt} \), represents the remainder. The reason that ordinary least squares estimates are inconsistent when lagged variables are included is that these variables are correlated with the current values of the residuals \( u_{nt} \) since they are determined to the same degree as the current value of the dependent variables by \( \mu_n \). Precisely the same sort of difficulty arises in the estimation of one of a system of structural equations involving more than one endogenous variable of the system. In this case, if one of the endogenous variables is chosen as dependent, and the rest are treated as if independent in a least squares regression, the results are inconsistent because of the correlation between the

---

\( ^{34} \) Note that the corresponding estimate from (35) is really quite similar to (38), for it can be written:

\[
\frac{\sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{t'=1}^{T} \hat{a}_{nt} \hat{a}_{nt'} - \sum_{n=1}^{N} \sum_{t=1}^{T} \hat{a}_{nt}^2}{T(T-1) \sigma^2}
\]

Thus, it is mostly a question of the appropriate adjustment to

\[
\sum_{n=1}^{N} \left( \sum_{t=1}^{T} \hat{a}_{nt} \right)^2 = \sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{t'=1}^{T} \hat{a}_{nt} \hat{a}_{nt'}
\]
endogenous variables treated as independent and the residual of the equation. One solution to this difficulty is to use as instrumental variables a sufficient number of other exogenous or (in the absence of serially correlated residuals) lagged endogenous variables appearing elsewhere in the system in the formation of the "normal" equations so that the current endogenous variables in the equation need not be used for this purpose. The difficulty, of course, is that there are usually more than enough predetermined variables for this purpose, and a choice must be made among them. One of Theil's contributions in the development of two-stage least squares was to show how such a choice could be avoided by selecting as instrumental variables those linear combinations of all predetermined variables most highly correlated with the current endogenous variables whose values they replaced in forming normal equations.\(^{35}\)

We do not, of course, deal here with one of a system of simultaneous equations containing several endogenous variables. It is thus not readily apparent how additional exogenous variables should be obtained. The key to the solution is to be found in the idea that the lagged values of the dependent variables are determined in a sense by other equations, although these are just lagged versions of the equation we are trying to estimate. Thus, under certain restrictions, the \(r\)th of equations (19) for the \(n\)th individual,

\[
y_{nt} = \sum_{k=1}^{K} \alpha_k z_{nt}^{(k)} + \sum_{t=1}^{\theta} \beta_t y_{n(t-\tau)} + u_{nt}
\]

(39)

has the solution

\[
y_{nt} = \sum_{k=1}^{K} \sum_{t=0}^{\infty} \gamma_t^{(k)} z_{nt-\tau} + w_{nt}.
\]

(40)

Equations (40) for \(n = 1, ..., N\), and \(t = 1, ..., T\), may be thought of as the reduced form for equation (39), \(n = 1, ..., N\), \(t = 1, ..., T\). If some of the lagged \(z\)'s are linear combinations of each other (as, for example, when some were shift variables), the solutions would only involve independent variables. Because the \(z\)'s are non-stochastic, they are independent of the current and past values of the residuals \(u_{nt}\). The residuals \(w_{nt}\) are merely linear combinations of the current and lagged values of \(u_{nt}\); hence, in principle it would be proper to estimate the coefficients \(\gamma_t^{(k)}\) by ordinary least squares. Of course, an infinite past history of \(z\)'s is not available so that truncation of the infinite series of them in (40) must occur at some point in practice. Since we are generally very short of time periods in a time series of cross sections, it is suggested that only the current \(z\)'s be used to estimate the coefficients \(\gamma_t^{(k)}\) and the remaining terms in (40) be neglected.

When estimates $\hat{\gamma}_0^{(k)}$ and perhaps others, $\hat{\gamma}_k^{(k)}$, $r > 0$, have been obtained, we have certain linear combinations of exogenous variables:

\[(41) \quad y_{nt}^* = \sum_{k=1}^{K} \hat{\gamma}_0^{(k)} z_{at}^{(k)} + \ldots ;\]

the lagged values of which may be used as instrumental variables in generating normal equations for the estimation of $\alpha$ and $\beta$ in (19). In effect, we replace the $r$th equation for the $m$th individual (39) by

\[(42) \quad y_{nt} = \sum_{k=1}^{K} \alpha_k x_{at}^{(k)} + \sum_{r=1}^{\theta} \beta_r y_{nt-r}^* + u_{nt};\]

\[= \sum_{k=1}^{K} \alpha_k x_{at}^{(k)} + \sum_{r=1}^{\theta} \sum_{k=1}^{K} \beta_r \hat{\gamma}_0^{(k)} z_{at}^{(k)} + \ldots + u_{nt},\]

in which it is appropriate to estimate the coefficients $\alpha_k$ and $\beta_r$ by ordinary least squares, provided the moment matrix of the $x$'s is nonsingular and satisfies certain other regularity conditions. In this case, the estimates of $\alpha_k$ and $\beta_r$ so obtained will be consistent and enable us to construct consistent estimates $S_m$ of $\Omega$ to be used in the second stage of this estimation of $\gamma$ given in (37).\(^{36}\)

The application of the first stage of this suggested estimating procedure to the gas data yields an estimate of $\rho$ of 0.7667. It is interesting to note that the value of $\rho$ so obtained is considerably larger than the corresponding value obtained by Kuh in his analysis of investment functions.\(^{37}\) This high value of $\rho$ is not due, as one may be tempted to think, to the exclusive influence of weather differences among states, since an explicit variable accounting for these differences was introduced in (39).

When this value of $\rho$ is used in the second stage, plausible estimates of the coefficients of both the price variable and lagged gas consumption are obtained (see Table III). For several reasons, the rate of depreciation in the gas market should not be expected to be very high. From the estimate of the coefficient of $G_{t-1}$ in Table III we may derive an estimated depreciation rate of approximately 4.5%.

The high standard errors associated with one income variable and one population variable would suggest that the estimating procedure is quite sensitive to collinear-

\(^{36}\) It may be shown directly that the second-round estimates are consistent, i.e., that

\[\text{plim } X'\Omega^{-1} u = \text{plim } \sum_{n} X_n' A^{-1} u_n = 0\]

where $X_n = [Z_n : Y_{0n}]$ is the $n$th block of the matrix $X = [Z : Y_0]$. While comforting, however, the demonstration is unnecessary as consistency is a property of AML estimates.

\(^{37}\) "... The individual firm effects are important, being in the present instance one-quarter to one-third as large as the time varying errors." E. Kuh, "The Validity of Cross-Sectionally Estimated Behavior Equations in Time Series Applications," *op. cit.*, p. 201.
TABLE III
AML Estimates of Gas Equation, 1957-62
(Standard errors in parentheses)
Coefficient of

<table>
<thead>
<tr>
<th>Const.</th>
<th>$P_t$</th>
<th>$N_{t-1}$</th>
<th>$\Delta N_t$</th>
<th>$Y_{t-1}$</th>
<th>$\Delta Y_t$</th>
<th>$G_{t-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.091</td>
<td>-0.0879</td>
<td>0.00360</td>
<td>-0.00122</td>
<td>0.00354</td>
<td>0.0170</td>
<td>0.9546</td>
</tr>
<tr>
<td>(11.544)</td>
<td>(.0468)</td>
<td>(.00129)</td>
<td>(.0190)</td>
<td>(.00622)</td>
<td>(.0080)</td>
<td>(.0372)</td>
</tr>
</tbody>
</table>

TABLE IV
AML Estimates: Constrained Case, 1957-62
(Standard errors in parentheses)
Coefficient of

<table>
<thead>
<tr>
<th>Const.</th>
<th>$P_t$</th>
<th>$N_t^*$</th>
<th>$Y_t^*$</th>
<th>$G_{t-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.148</td>
<td>-0.0695</td>
<td>0.02163</td>
<td>0.01801</td>
<td>0.9786</td>
</tr>
<tr>
<td>(5.250)</td>
<td>(.0441)</td>
<td>(.00789)</td>
<td>(.00794)</td>
<td>(.0321)</td>
</tr>
</tbody>
</table>

ity. To circumvent such difficulties, one might use the estimated depreciation rate for all fuel using appliances of 11 per cent (obtained in Section 3) to construct the variables $N_t^*$ and $Y_t^*$ and then proceed to the estimation of a constrained version of the gas equation. The results of such computation are shown in Table IV. All coefficients have small standard errors. As compared to the unconstrained model, this new method yields a smaller coefficient for the price variable and a higher coefficient for the lagged consumption variable.

Application of the residual model to the gas data produces estimates that are in agreement with the theoretical expectations. It may be emphasized that the estimated coefficient of lagged gas consumption (as given by the residual model) is smaller than unity. This result lends support to the basic hypothesis embodied in the dynamic model of gas demand.

Unfortunately, the method just outlined breaks down if there are no truly exogenous variables in the problem, apart from constant terms, trends, or shift variables, for the lagged values of these are linear combinations of their current values and the moment matrix of independent variables in (42) will be singular. It is therefore useful to develop an alternative approach that does not require the presence of true exogenous variables, although we do not need it in the particular case of the gas model. (Furthermore, given the limited number of observations available over time, this new approach is inapplicable to the gas model.)

Originally it was thought that the first-difference transformation, suggested
earlier by A. Walters, could be used to obtain consistent estimates needed for the first round of the AML procedure. Regrettably, however, this proved not to be the case. Nonetheless, provided a sufficient number of observations over time on each individual is available, consistent estimates may be found. The six years of gas data, our second period, are unfortunately insufficient considering the number of parameters to be estimated; nevertheless, the method may be useful in other contexts and is as follows:

Rewrite (39) as

\[(43) \quad y_{nt} = (\alpha_1 + \mu_n) + \sum_{k=1}^{K} \alpha_k z_{nt}^{(k)} + \sum_{t=1}^{\theta} \beta_t y_{n,t-\tau} + v_{nt}.\]

Since the \(v_{nt}\) are serially independent, provided sufficient observations over time are available, the ordinary least squares estimates in (43) for each individual may be found and are certainly consistent. Let these be denoted by \(\hat{\alpha}_k\) and \(\hat{\beta}_t^{(n)}\), the superscript indicating the estimates refer to the \(n\)th individual. The following series of estimates of parameters in (39) may be shown to be consistent:

\[
\begin{align*}
\hat{\alpha}_k &= \frac{1}{N} \sum_{n=1}^{N} \hat{\alpha}_k^{(n)}, \quad k = 2, \ldots, K, \\
\hat{\beta}_t &= \frac{1}{N} \sum_{n=1}^{N} \hat{\beta}_t^{(n)}, \quad \tau = 1, \ldots, \theta, \\
\hat{\alpha}_1 &= \frac{1}{N} \sum_{n=1}^{N} \left\{ \bar{y}_n - \sum_{k=2}^{K} \alpha_k \bar{x}_n^{(k)} - \sum_{t=1}^{\theta} \beta_t \bar{y}_{n,t-\tau} \right\}, \\
\text{where} \quad \bar{y}_{n,t-\tau} &= \frac{1}{T} \sum_{t=1}^{T} y_{n,t-\tau}, \quad \tau = 0, \ldots, \theta, \quad \text{and} \quad \bar{x}_n^{(k)} = \frac{1}{T} \sum_{t=1}^{T} z_{nt}^{(k)}, \\
\hat{\sigma}_n &= \frac{1}{N} \sum_{n=1}^{N} \left[ \left( \bar{y}_n - \sum_{k=2}^{K} \hat{\alpha}_k \bar{x}_n^{(k)} - \sum_{t=1}^{\theta} \hat{\beta}_t \bar{y}_{n,t-\tau} \right) - \hat{\alpha}_1 \right]^2, \\
\hat{\sigma}^2 &= \frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} \left\{ y_{nt} - \alpha_1 - \sum_{k=2}^{K} \hat{\alpha}_k z_{nt}^{(k)} - \sum_{t=1}^{\theta} \hat{\beta}_t y_{n,t-\tau} \right\}^2, \\
\hat{\rho} &= \frac{\hat{\sigma}_n^2}{\hat{\sigma}_n^2 + \hat{\sigma}^2}.
\end{align*}
\]


\[\text{39} \quad \text{As was pointed out by Z. Griliches and L. G. Telser.}\]

\[\text{40} \quad \text{If we let} \]

\[\mu_{nt} = y_{nt} - \alpha_1 - \sum_{k=2}^{K} \hat{\alpha}_k z_{nt}^{(k)} - \sum_{t=1}^{\theta} \hat{\beta}_t y_{n,t-\tau}\]

\[\text{then } \varrho \text{ may be written suggestively as}\]
The final estimate of $\rho$ may be used in a second round of estimates in the manner suggested to obtain AML estimates.

5. CONCLUSIONS

In this paper we have developed a dynamic model of the demand for natural gas in the residential and commercial sector. Point estimates of the parameters of the model based on a pooled sample of 36 states over six years suggest an implausible, negative rate of depreciation on gas appliances. Introduction of state shift variables results in an estimated rate of 30% per year—also implausible. After several attempts to incorporate other variables (not reported here), we concluded that time invariant, but perhaps unobservable state effects were responsible for biasing the coefficient of lagged gas consumption in the demand equation.

A model suggested by E. Kuh was explored in some detail and several different methods for estimating its parameters in the presence of lagged endogenous variables were proposed. Of these, the maximum-likelihood method was shown to lead, in the case of gas demand, to an unacceptable boundary solution. An alternative method, however, was proposed which led to theoretically plausible results. The final results obtained suggested that time invariant regional effects account for about three-quarters of the total residual variance in the gas demand equation. The estimated rate of depreciation on gas appliances (mainly furnaces) is in the order of 5% per year. The estimated net long-run price and income elasticities of new gas demand are .63 and .62, respectively, in the unconstrained case, and .63 and .44, when the depreciation rate is assumed to be 11% for all fuel consuming appliances. In this case, the depreciation on gas appliances is estimated at slightly more than 2% per year.

Université de Fribourg
and
Yale University

APPENDIX

DERIVATION OF THE MAXIMUM-LIKELIHOOD ESTIMATES OF $\rho$ AND $\sigma^2$

Rather than differentiate directly with respect to $\rho$, which will involve us in the

$$
\varrho = \frac{\sum_{n=1}^{N} \left( \sum_{i=1}^{T} y_{it} \right)^2}{T \sum_{n=1}^{N} \sum_{t=1}^{T} y_{it}^2}.
$$

The perceptive reader will note that this is the same as $\varrho$ in (38) apart from rounding errors.
explicit evaluation of $\Omega^{-1}$ and $|\Omega|$, it is useful to reparameterize the problem by making a particular kind of orthogonal transformation of the variables. Since the Jacobian of an orthogonal transformation is unity, the probability density of the transformed variables is given simply by (31) with appropriate substitutions, and thus the likelihood function of the new parameters will be given by making the corresponding substitutions in (32).

Let $C$ be a $T \times T$ orthogonal matrix, every element of the first row of which is $1/\sqrt{T}$. If $e$ is a $T \times 1$ vector consisting entirely of ones,

$$e = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix},$$

we may write $C$ as

$$C = \begin{bmatrix} e' / \sqrt{T} \\ C_1 \end{bmatrix}$$

where $C_1$ is a $T-1 \times T$ matrix. There is an infinite number of such matrices; for example, one is

$$C = \begin{bmatrix}
\frac{1}{\sqrt{T}} & \frac{1}{\sqrt{T}} & \frac{1}{\sqrt{T}} & \cdots & \frac{1}{\sqrt{T}} \\
\frac{1}{\sqrt{1\times2}} & -1 & 0 & \cdots & 0 \\
\frac{1}{\sqrt{2\times3}} & \frac{1}{\sqrt{2\times3}} & -2 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{1}{\sqrt{(T-1)T}} & \frac{1}{\sqrt{(T-1)T}} & \frac{1}{\sqrt{(T-1)T}} & \cdots & -(T-1) / \sqrt{(T-1)T}
\end{bmatrix}.$$ 

It is unnecessary, however, to specify which one of all those we choose, as only the fact that all the elements in the first row are $1/\sqrt{T}$ and the orthogonality of the matrix as a whole are ever used. Because $C$ is orthogonal, we have

$$CC' = \begin{bmatrix} e' / \sqrt{T} & C_1 \end{bmatrix} \begin{bmatrix} e' / \sqrt{T} \\ C_1' \end{bmatrix} = \begin{bmatrix} 1 & e' C_1' \\ C_1 e & C_1 C_1' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & I \end{bmatrix},$$

so that

$$C_1 e = 0, \quad C_1 C_1' = I$$

(A.2)
where $I$ is a $T-1 \times T-1$ identity matrix. In a similar way we can show that $C' C_1 = I - e e' / T$, where $I$ is $T \times T$.

Let $C^*$ be an $NT \times NT$ orthogonal matrix defined as

$$
C^* = \begin{bmatrix}
    C & 0 & \ldots & 0 \\
    0 & C & \ldots & 0 \\
    \vdots & & & \\
    0 & 0 & \ldots & C
\end{bmatrix}.
$$

The transformation we wish to make is from $u$ to $v$ (implying corresponding transformations in $y$ and $X$), where

(A.3) \quad v = C^* u.

Because $C^*$ is orthogonal, we also have, of course, $C^* v = u$.

By equation (25),

(A.4) \quad \sigma^2 A = \sigma^2 \{ (1-\rho) I + \rho \, ee' \}.

Hence, by (A.2), we deduce

(A.5) \quad \sigma^2 C A C' = \sigma^2 \{ (1-\rho) C C' + \rho \, C \, e e' \, C' \}

\[ = \sigma^2 \left\{ (1-\rho) I+\rho \begin{bmatrix}
    T/\bar{T} C_1 e' e C_1 & \\
    0 & 0 & \ldots & 0
\end{bmatrix} \right\} \]

\[ = \sigma^2 \left\{ (1-\rho) I+\rho \begin{bmatrix}
    T & 0 & \ldots & 0 \\
    0 & 0 & \ldots & 0 \\
    \vdots & & & \\
    0 & 0 & \ldots & 0
\end{bmatrix} \right\}.

Thus, we let

(A.6) \quad \begin{cases} 
    \zeta = \sigma^2 \{ (1-\rho) + T \rho \}, \\
    \eta = \sigma^2 (1-\rho).
\end{cases}

Note that there is a one-to-one correspondence between the parameters, $\zeta$ and $\eta$, and the original parameters, $\sigma^2$ and $\rho$:

(A.7) \quad \begin{cases} 
    \rho = \frac{\zeta - \eta}{\zeta + \eta (T-1)}, \\
    \sigma^2 = \frac{\zeta + \eta (T-1)}{T}.
\end{cases}

Since maximum-likelihood estimates have the property that the estimate of a one-to-one function of a parameter is the function of the maximum-likelihood estimate, it follows that such estimates of the parameters $\rho$ and $\sigma^2$ may be obtained by estimating $\zeta$ and $\eta$ instead. Actually, since $\sigma^2$ is estimated so readily by (34) once $y$ and $\rho$ are known, we will need only to obtain $\rho$ from (41), and $\zeta$ and $\eta$.

The next step in our analysis must be to rewrite the likelihood function in terms of the parameters $\zeta$ and $\eta$. First, evaluate $|\Omega|$ as
Next, evaluate \( u'\Omega^{-1}u \):

(A.9) \[ u'\Omega^{-1}u = v'C^*\Omega^{-1}C^*v = v'[C^*\Omega C^*]'v. \]

Finally \( C^*\Omega C^* = \text{diag} \{ \xi, \eta, ..., \eta, ..., \xi, \eta, ..., \eta \}, \)
so that

(A.10) \[ [C^*\Omega C^*]'^{-1} = \text{diag} \left\{ \frac{1}{\xi}, \frac{1}{\eta}, ..., \frac{1}{\eta}, ..., \frac{1}{\xi}, \frac{1}{\eta}, ..., \frac{1}{\eta} \right\}. \]

Thus, reparameterized, the likelihood function is

(A.11) \[ L(\gamma, \xi, \eta) = -\frac{NT}{2} \log 2\pi - \frac{N}{2} \log \xi - \frac{N(T-1)}{2} \log \eta - \frac{1}{2} \left( \frac{M_1(\gamma)}{\xi} + \frac{M_2(\gamma)}{\eta} \right) \]

where
\[ M_1(\gamma) = \sum_{n=1}^{N} v_{n1}^3, \]
\[ M_2(\gamma) = \sum_{n=1}^{N} \sum_{t=2}^{T} v_{nt}^2. \]

If we now differentiate with respect to \( \xi \) and \( \eta \) and set the resulting expressions to zero, we obtain
\[ \xi = \frac{M_1(\gamma)}{N}, \]
\[ \eta = \frac{M_2(\gamma)}{N(T-1)}. \]

One may readily verify that at this point the second-order conditions for a maximum of \( L \) will generally be satisfied if \( \gamma \) is given. If \( \xi \) and \( \eta \) from (A.13) are inserted in (A.7), the maximum-likelihood equations for \( \rho \) and \( \sigma^2 \) are obtained.

\[
\begin{align*}
\rho &= \frac{(T-1)M_1(\gamma) - M_2(\gamma)}{(T-1)[M_1(\gamma) + M_2(\gamma)]}, \\
\sigma^2 &= \frac{M_1(\gamma) + M_2(\gamma)}{NT}.
\end{align*}
\]

To complete the derivation of the equations for the determination of the maximum-likelihood estimate it is necessary to express \( M_1(\gamma) \) and \( M_2(\gamma) \) of (A.14) in terms of the vector \( u = y - X\gamma \) appearing in (33) and (34). From (A.3) and partitioning \( v \) conformably with the earlier partition of \( u \), we have

\[
v_n = \begin{bmatrix} v_{n1} \\ \vdots \\ v_{nT} \end{bmatrix} = Cu_n = \begin{bmatrix} \epsilon^\prime \sqrt{T} \\ C_1 \end{bmatrix} u_n = \begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{nt} \\ C_1 u_n \end{bmatrix},
\]

so that
\[ M_1(\gamma) = \sum_{n=1}^{N} v_{n1}^2 = \frac{1}{T} \sum_{n=1}^{N} \sum_{t=1}^{T} u_{nt} u_{nt'}, \]
\[ M_2(\gamma) = \sum_{n=1}^{N} \sum_{t=2}^{T} v_{nt}^2 = \sum_{n=1}^{N} u_{n1}^2 C_1 C_1 u_n = \sum_{n=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} u_{nt}^2 - \frac{\left( \sum_{t=1}^{T} u_{nt} \right)^2}{T} \right), \]

since \( C_1 C_1 = I - ee' \). Equations (35) of the text follow from (A.14) by inserting these results.