Least Squares Estimation with Autocorrelated Residuals: A Survey

Hak Yong Rhee*

1. Introduction

Ever since Gauss discussed the least-squares method in 1812 and Bertrand translated Gauss's work in French, the least-squares method has been used for various economic analysis.[1] The justification of the least-squares method was given by Markov in 1912 in connection with the previous discussion by Gauss and Bertrand. The main argument concerned the problem of obtaining the best linear unbiased estimates. In some modern language, the argument can be explained as follow.

Suppose that we have a single linear specification between a variable $\boldsymbol{y}$ and $k-1$ explanatory variables $x_2, x_3, \cdots\cdots, x_k$ and a disturbance term $\boldsymbol{u}$ such that $\boldsymbol{y}=\beta \boldsymbol{x}'+\boldsymbol{u}$, where $\beta=(\beta_1, \beta_2, \cdots\cdots, \beta_k)$, $\boldsymbol{x}=(1, x_2, x_3, \cdots\cdots, x_k)$ is a known vector, and $\boldsymbol{y}$ and $\boldsymbol{u}$ are random variables. If we have n observations, we can write

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{u},$$

where $\boldsymbol{y}$ is a $n \times 1$ vector; $\boldsymbol{X}$ is a $n \times k$ matrix with the first column being all ones; β is a $k \times 1$ vector of coefficients; $\boldsymbol{u}$ is a $n \times 1$ vector of disturbances. We now make the following assumptions:

A1 $\quad E(\boldsymbol{u})=0$

---

* Associate Professor of Economics and Statistics, Yonsei University. The author would like to point out that a part of the survey was done during his graduate study and the comments made by Dr. Suk Bum Yoon were helpful.

1) R.L. Plackett, "A Historical Note on the Method of Least Squares," *Biometrika*, 1949, pp. 459.

A2 $E(\boldsymbol{uu}')=\sigma^2 I_n$

A3 $\boldsymbol{X}$ is a matrix of fixed numbers

A4 $\boldsymbol{X}$ has a rank k, and $k<n$

Then it is well known that a least-squares estimate (LSE) of β is obtained by $\hat{\beta}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$. Furthermore such β is an unbiased, and it has the least variance.[1]

In 1935, A.C. Aitken raised the question whether or not the assumptions A1 and A2 are valid, and he suggested alternative method for the cases in which we cannot make these two assumptions.[2] The first assumption, $E(\boldsymbol{u})=0$, shows that $E(u_i)=0$ for all i, that is, that the u_i's are random variables with zero expections. The second assumption states that

$$E(\boldsymbol{uu}')=\begin{pmatrix} E(u_1^2) & E(u_1u_2)\cdots\cdots E(u_1u_n) \\ E(u_2u_1) & E(u_2^2)\cdots\cdots\cdots E(u_2u_n) \\ \vdots & \vdots \qquad\qquad \vdots \\ E(u_nu_1) & E(u_nu_2)\cdots\cdots E(u_n^2) \end{pmatrix}=\begin{pmatrix} \sigma^2 & 0 & 0 & \cdots\cdots 0 \\ 0 & \sigma^2 & 0 & \cdots\cdots 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots\cdots \sigma^2 \end{pmatrix}$$

This latter assumption has two implications: one is $E(u_i^2)=\sigma^2$ for all i, that is, the u_i have constant variance which is called homoscedasticity; the other is $E(u_t u_{t+s})=0$ for $s\neq 0$, that is, the u_i values are uncorrelated.

The last implication makes it possible to have a diagonal variance-covariance matrix. This assumption of uncorrelated residuals and the homoscedasticity assumption make it possible to show that the estimate $\hat{\beta}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ has the smallest variance, and the assumption A1 makes the estimate unbiased.

If the residuals are serially correlated, the estimate given by least-squares method will not be the best estimate. This problem turned out to be very

1) Johnston, *Econometric Method*, pp. 109~113.

2) A.C. Aitken, "On Least Squares and Linear Combinations of Observations," *Proceedings of the Royal Society of Edinburgh*, Vol. 55, 1935, pp. 42~48.

important because in many cases, particularly in time series, we often have autocorrelated residuals.

A very important problem is how to determine whether or not the autocorrelated residuals are autocorrelated. In 1941, J. Von Neumann calculated the asymptotic properties of the probability distribution of $d=\frac{\sum(\Delta u_i)^2}{\sum u_i^2}$ of a random series u_i.[1] In 1950, J. Durbin and G. S. Watson, using the estimated residuals, calculated the following statistic, $\hat{d}=\frac{\sum(\hat{u}_t-\hat{u}_{t-1})^2}{\sum\hat{u}_t^2}$ as a measure of autocorrelation. If the estimated $\hat{d}$ is lower than the lower bound, the residuals are positively correlated and if the $\hat{d}$ statistic is larger than the upper bound, the series is negatively correlated. They calculated the lower bounds and the upper bounds for different sample sizes and for different numbers of variables for a 5 per cent level of significance.[2] H. Theil and A.L. Nagar developed the Durbin and Watson's approach further to determine the upper and lower bounds for the case in which the difference of the explanatory variables are rather small compared with the range of the corresponding variable itself.[3]

The recent literature on this subject has been concerned with the method of obtaining the revised estimates after we find that the residuals are autocorrelated. First of all most of the writers assume that there exists a first order Markov scheme among the residuals, that is, the residuals are correlated such that $u_t=\rho u_{t-1}+e_t$, where the e_t has the following properties: $E(e_t)=0$ for all t and $E(e_t e_{t+s})=\sigma_e^2$ if $s=0$: $E(e_t e_{t+s})=0$ if $s\neq 0$. With these assumptions, it has been shown that we can use a transformation on the original data so that we can have a new series which has the uncorrelated residuals. Then we can apply the least-squares method to the transformed data rather than to the original data to get the best linear

1) J. Von Neumann, "Distribution of the Ratio of the Mean Square Successive Difference to the Variance," *Annals of Mathematical Statistics*, 1941, pp. 367~95.

2) J. Durbin and G.S. Watson, "Testing for Serial Correlation in Least Squares Regression," *Biometrica*, 1951, pp. 162.

3) H. Theil and A.L. Nagar, "Testing the Independence of Regression Disturbances," *Journal of the American Statistical Association*, 1961, pp. 794.

unbiased estimates. It has also been shown that the transformation matrix can be derived from the variance-covariance matrix under the assumption of the first order Markov scheme.

Serial correlation problem also arises in simultaneous equation systems. Indeed if ordinary least-squares procedures are to be used on each single equations or on the reduced forms, the absence of serial correlation in the error terms is critical. In other simultaneous equation techniques such as two-stage least squares, indirect least squares and three stage least squares, the same problem can arise.

Fisher has given a considerable treatment of this question.[1] He shows that in the system,

$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{y}_t + \boldsymbol{B}\boldsymbol{y}_{t-1} + \boldsymbol{C}\boldsymbol{z}_t + \boldsymbol{u}_t,$$

$\boldsymbol{u}_t$ is an m-component column vector of disturbances; $\boldsymbol{y}_t$ is an n-component exogenous variables; $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ are constant matrices to be estimated; and $(\boldsymbol{I}-\boldsymbol{A})$ is nonsingular, $\boldsymbol{A}$'s diagonal elements are all zeros. "The model is recursive and does not violate the assumption that in eache quation the dist urbance term is uncorrelated with the variables which appear therein other than the one to be explained by that equation." In addition, $\boldsymbol{u}_t$ should be normally distributed and homoscedastic sf the ordinary least-squares method is to be valid in the use of having maximum likelihood estimator.[2] He fur ther argues that the autocorrelation problem in the simultaneous equation system is a critical problem to get the maximum likelihood estimate.

Then, in principle, it is also possible to use the transformation procedure to get rid of the autocorrelation problem in simultaneous equation technique as well as in the single equation system. The main problem is then how to get the transformation matrix or the variance-covariance matrix of

1) F.M. Fisher, "Dynamic Structure and Estimation in Economywide Econometric Models," in *Econometric Model of the United States*, ed. by J.S. Duesenberry, G. Fromm, L.R. Klein, and E. Kuh, 1965, pp. 580~638.

2) *Ibid.*, pp. 593.

the disturbances. My main attempt in this paper is to present the recent propositions on estimating the transformation matrix.

In the first part of this paper I present a procedure to get revised estimates by using a known variance-covariance matrix of the residuals. In the second part, I investigate the various procedures of estimating the transformation matrix. Finally I present some comments on some of the empirical works in which the revised estimation procedure is used.

2. Least-Squares Estimates with Autocorrelation

Let us assume that we have the single equation, n variable linear model, $\boldsymbol{y}=X\beta+\boldsymbol{u}$, where $\boldsymbol{y}$ is a $n\times1$ vector of observations; X is a $n\times k$ matrix of observations on k explanatory variables; β is a $k\times1$ vector of unknown coefficients; and $\boldsymbol{u}$ is $n\times1$ disturbance terms.

Suppose now that we cannot make the assumption A2. Then we have non-diagonal variance-covariance matrix.

$$\frac{\sigma_\epsilon^2}{1-\rho}\begin{pmatrix} 1 & \rho & \rho^2\cdots\rho^{n-1} \\ \rho & 1 & \rho\ \cdots\rho^{n-2} \\ \vdots & \vdots & \vdots \\ \rho^{n-1} & \rho^{n-2}\cdots & \cdots 1 \end{pmatrix}=V,^{1)} \quad \text{and} \quad V^{-1}=\frac{1}{\sigma_\epsilon^2}\begin{pmatrix} 1 & -\rho & 0\cdots0 & 0 \\ -\rho & 1+\rho^2 & -\rho\cdots0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots\cdots-\rho & 1 \end{pmatrix}$$

Then the next procedure is to use the transformation of T such that $T'T=V^{-1}$ to the original model $\boldsymbol{y}=X\beta+\boldsymbol{u}$. Then the original model becomes $T\boldsymbol{y}=TX\beta+T\boldsymbol{u}$.

Accordingly $\hat\beta=[(TX)'TX]^{-1}(TX)'(T\boldsymbol{y})$

$$=(X'T'TX)^{-1}X'T'T\boldsymbol{y}$$

1) See Johnston, pp. 178.

$$=(X'V^{-1}X)^{-1}X'V^{-1}y$$ [1]

If we take $(n-1)\times n$ matrix, $T=\begin{pmatrix} -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$,

$$T'T=\begin{pmatrix} \rho^2 & -\rho & 0 & 0\cdots\cdots 0 \\ -\rho & 1+\rho^2 & -\rho & 0\cdots\cdots 0 \\ 0 & -\rho & 1+\rho^2 & -\rho\cdots\cdots 0 \\ \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & 0\cdots\cdots 1 \end{pmatrix}.$$

So that $T'T$ becomes V^{-1} except the first element of the first row and that of the first column, and that σ_ε^2 should be unity. To make a simple and approximate estimation, this T matrix transformation is proposed. Now if we take the simple least-squares estimate from this model,

$$\hat{\beta}=[(TX)'(TX)]^{-1}(TX)'(Ty) \text{ and } E[(Tu)(Tu)']=\sigma_u^2 I_{n-1}.$$

Then the next problem is the estimation of ρ where $|\rho|<1.$ [2]

(1) Method of H. Theil and Nagar

Let u_t, $(t=1,\cdots,n)$, denote the residuals from the general linear model which we defined before. Test these residuals whether the residuals are autocorrelated or not by the Durbin-Watson $\hat{d}$ statistics, which is defined as $\hat{d}=\frac{\sum(\hat{u}_t-\hat{u}_{t-1})^2}{\sum\hat{u}_t^2}$. These residuals are estimated by the ordinary least-squares estimate. Suppose that it shows that the residuals are autocorrelated. Then the following procedure can be used to estimate ρ value in the first-order Markov scheme. Take the probability limit

$$\text{Plim } \hat{d}=\frac{\text{Plim } E(u_t-u_{t-1})^2}{\text{Plim } E(u_t)^2}.$$

Since $E(u_t-u_{t-1})=0$, $E(u_t-u_{t-1})^2=\text{Var}(u_t-u_{t-1})$, and $E(u_t)^2=\text{Var}(u_t)$.

1) It is known that $\hat{\beta}=(X'V^{-1}X)^{-1}X'V^{-1}y$ is the best unbiased estimator of β when the residuals are autocorrelated with first order Markov scheme. For the proof of this, See Johnston's *Econometric Methods*, (New York, 1960), pp. 180~184.

2) J. Johnston, *Econometric Methods*, pp. 178.

Therefore $\dfrac{\mathrm{Var}(u_t-u_{t-1})}{\mathrm{Var}(u_t)}$ can be regarded as consistent estimate of d.

Now under the assumption of the first-order Markov scheme,

$$\begin{aligned}\mathrm{Var}(u_t-u_{t-1}) &= E[\{(u_t-u_{t-1})-E(u_t-u_{t-1})\}^2] \\ &= E[(u_t-u_{t-1})^2]=E[u_t^2-2u_tu_{t-1}+u_{t-1}^2] \\ &= E(u_t^2)-2E(u_tu_{t-1})+E(u_{t-1}^2)=2\frac{\sigma_\epsilon^2}{1-\rho^2}-2\rho\sigma_u^2 \\ &= \frac{2\sigma_\epsilon^2}{1-\rho^2}-2\rho\frac{\sigma_\epsilon^2}{1-\rho^2}=\frac{2\sigma_\epsilon^2(1-\rho)}{1-\rho^2}\end{aligned}$$

and $\mathrm{Var}(u_t^2)=E[\{u_t-E(u_t)\}^2]=E(u_t^2)=\sigma_u^2=\dfrac{\sigma_\epsilon^2}{1-\rho^2}$

Then, $\dfrac{\mathrm{Cov}(u_tu_{t-1})}{\mathrm{Var}(u_t)}=\dfrac{2\sigma_\epsilon^2(1-\rho)}{1-\rho^2}\times\dfrac{1-\rho^2}{\sigma_\epsilon^2}=2(1-\rho).$

Therefore, d can be regarded as a consistent estimate of $2(1-\rho)$, *i.e.*, $d=2(1-\hat{\rho})$ or $\hat{\rho}=1-\dfrac{d}{2}$.[1]

(2) Method of Hildreth and Lu

Let $\boldsymbol{y}=X\beta+\boldsymbol{u}$ be the general linear model as before, and define,

$$\boldsymbol{X}^*=\begin{pmatrix} x_{01} & x_{02}\cdots\cdots\cdots x_{0k} \\ x_{11} & x_{12}\cdots\cdots\cdots x_{1k} \\ \vdots & \vdots \qquad\qquad \vdots \\ x_{n-1,1} & x_{n-1,2}\cdots\cdots x_{n-1,k}\end{pmatrix} \quad \boldsymbol{y}^*=\begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1}\end{pmatrix} \quad \boldsymbol{u}^*=\begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_{n-1}\end{pmatrix} \quad \eta=\begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n\end{pmatrix}$$

Then we have $\boldsymbol{y}^*=\boldsymbol{X}^*\beta+\boldsymbol{u}^*$ or $\boldsymbol{u}^*=\boldsymbol{y}^*-\boldsymbol{X}^*\beta$, and the first-order Markov scheme implies $\boldsymbol{u}=\rho\ \boldsymbol{u}^*+\eta$. By substituting $\boldsymbol{u}^*$ and $\boldsymbol{u}$ to the latter equation, we get $\eta=\boldsymbol{u}-\rho\boldsymbol{u}^*=\boldsymbol{y}-X\beta-\rho(\boldsymbol{y}^*-X^*\beta)=\boldsymbol{y}-\rho\boldsymbol{y}^*-X\beta-\rho X^*\beta=(\boldsymbol{y}-\rho\boldsymbol{y}^*)-(X+\rho X^*)\beta$

We know that η is normally distributed with zero mean and σ_η^2. Then

1) H. Theil and A.L. Nagar, "Testing the Independence of Regression Disturbances," *Journal of American Statistical Association*, Dec. 1961, pp. 804.

$$P(\eta)=(2\ \sigma_\eta^2)^{-\frac{n}{2}}\cdot \exp\{-\frac{1}{2\sigma_\eta^2}\eta'\eta\}$$

$$=(2\ )^{-\frac{n}{2}}\sigma_\eta^{-n}\cdot \exp\{-\frac{1}{2\sigma_\eta^2}[(y-\rho y^*)-(X-\rho X^*)\beta]'[(y-\rho y^*)$$
$$-(X-\rho X^*)\beta]\}.$$

When the samples of y, y^*, X, X^* are applied to the above density function, it becomes a likelihood function of ρ, β, and σ_η^2. Therefore, we have,

$$L=(2\Pi\sigma_\eta^2)^{-\frac{n}{2}}\exp\{-\frac{1}{2\sigma_\eta^2}[(y-\rho y^*)-(X-\rho X^*)\beta]'[(y-\rho y^*)-(X-\rho X^*)\beta]\}$$

In order to obtain the maximum likelihood estimates, take logarithm and drop the constant term involving π. Then we have

$$\log L^*=-\frac{n}{2}\log\sigma_\eta^2-\frac{1}{2\sigma_\eta^2}[(y-\rho y^*)-(X-\rho X^*)\beta]'[(y-\rho y^*)$$
$$-(X-\rho X^*)\beta].$$

Now we can readily see that log L^* is a function of ρ, β, σ_η^2 so that we can write,

$$\log L^*(\rho,\beta,{}^2\sigma_\eta^2)=-\frac{n}{2}\log\sigma_\eta^2-\frac{1}{2\sigma_\eta^2}[(y-\rho y^*)-(X-\rho X^*)\beta]'[(y-\rho y^*)$$
$$-(X-\rho X^*)\beta].$$

Now differenciate log L^* with respect to σ_η^2 and set equal to zero. Solve for σ_η^2 and substituting back to log L^*, *i.e.*, log L^* is concentrated.

$$\frac{\partial \log L^*}{\partial\sigma_\eta^2}=-\frac{n}{2}\cdot\frac{1}{\sigma_\eta^2}+\frac{2}{4\sigma_\eta^4}(S)=0,$$

where $S=[(y-\rho y^*)-(X-\rho X^*)\beta]'[(y-\rho y^*)-(X-\rho X^*)\beta].$

Then $-2n\sigma_\eta^2=-2S$ or $\sigma_\eta^2=\frac{S}{n}$ and

$$\log L^*(\rho,\beta)=-\frac{n}{2}\log\left(\frac{S}{n}\right)-\frac{S}{2\left(\frac{S}{n}\right)}=-\frac{n}{2}(\log S-\log n)-\frac{n}{2}$$

$$= \frac{-n\log S}{2} + \frac{n \log n}{2} - \frac{n}{2} = \frac{-n\log S}{2} + \frac{n(\log n - 1)}{2}$$

Now one can see that minimizing S is equivalent to maximizing $\log L^*$ and hence L, and we have

$$\log S(\rho,\beta) = [(\boldsymbol{y}-\rho\boldsymbol{y}^*)-(\boldsymbol{X}-\rho\boldsymbol{X}^*)\beta]'[(\boldsymbol{y}-\rho\boldsymbol{y}^*)-(\boldsymbol{X}-\rho\boldsymbol{X}^*)\beta],$$

which is a function of ρ and β. Howeve, if ρ is given,

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\ \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}, \text{ where } \boldsymbol{V}^{-1} = \boldsymbol{T}'\boldsymbol{T} \text{ and } \boldsymbol{T} = \begin{pmatrix} -\rho & 1 & 0 \cdots\cdots 0 \\ 0 & -\rho & 1 \cdots\cdots 0 \\ \vdots & \vdots & \vdots \quad\quad \vdots \\ 0 & 0 & 0 \cdots\cdots 1 \end{pmatrix}$$

Therefore, $\log S$ is actually function of ρ.

Minimizing $\log S(\rho)$ is a complicated task. Therefore, Hildreth and Lu suggest applying $(n-1)\times n$ matrix $\boldsymbol{T}$ transformation to the data $\boldsymbol{y}$ and $\boldsymbol{X}$ and successively substituting values of ρ between -1 and 1. This procedure will give the value of $\log S(\rho)$ for different ρ. Choose ρ which gives the lowest $\log S(\rho)$.

(3) Cochrane and Orcutt Method

Define $\boldsymbol{y} = X\beta + \boldsymbol{u}$ as before. By simple least-square method,

we have $\beta = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ Then set $\boldsymbol{y} - \boldsymbol{X}\beta = \boldsymbol{u}$, where $\boldsymbol{u} = \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \vdots \\ \hat{u}_n \end{pmatrix}$

Now define $\boldsymbol{u}_1 = \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_{n-1} \end{pmatrix}$, $\boldsymbol{u}_{11} = \begin{pmatrix} \hat{u}_2 \\ \hat{u}_3 \\ \vdots \\ \hat{u}_n \end{pmatrix}$, and $\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$

By the first-order Markov scheme, we know that $\hat{\boldsymbol{u}}_1 = \rho\boldsymbol{u}_{11} + \epsilon$.

By using simple least-square method, we obtain $\hat{\rho} = (\hat{\boldsymbol{u}}_{11}'\hat{\boldsymbol{u}}_{11})^{-1}\hat{\boldsymbol{u}}'_{11}\hat{\boldsymbol{u}}_1$.

Form the $\boldsymbol{T}$ matrix, $\boldsymbol{T} = \begin{pmatrix} -\hat{\rho} & 1 & 0 \cdots\cdots 0 \\ 0 & -\hat{\rho} & 1 \quad \cdots 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \quad \cdots 1 \end{pmatrix}$

and obtain $\hat{\beta}=(X'T'TX)^{-1}X'T'Ty$. Now put this β to the original form $y=X\beta+u$ and obtain $\hat{u}$. By using $\hat{u}$, we can obtain $\hat{\rho}$ to estimate $\hat{\hat{\beta}}$ and so on. Repeat this procedure until the residuals are independent and therefore no adjustments are necessary.

Cochrane and Orcutt also suggest the use of the first difference on the assumption that ρ is approximately one. The first order Markov scheme, $u_t=\rho u_{t-1}+\epsilon_t$ becomes $u_t=u_{t-1}+\epsilon_t$ if $\rho=1$. We then have $u_t-u_{t-1}=\epsilon_t$. By substituting $u_t=Y_t-\beta_0-\beta_1X_t$ and $u_{t-1}=Y_{t-1}-\beta_0-\beta_1X_{t-1}$, we have,

$$Y_t-\beta_0-\beta_1X_t-Y_{t-1}+\beta_0+\beta_1X_{t-1}=\epsilon_t$$

$$\text{or } Y_t-Y_{t-1}=\beta_1X_t-\beta_1X_{t-1}+\epsilon_t$$

$$\text{or } Y_t-Y_{t-1}=(X_t-X_{t-1})\beta_1+\epsilon_t.$$

Set $Y_t-Y_{t-1}=Y_t^*$, $X_t-X_{t-1}=X_t^*$. Then we have $y_t^*=X_t^*\beta_1+\epsilon_t$, where ϵ_t is independently and normally distributed with mean zero and σ_ϵ^2. We can now use simple least-squares method.

(4) Durbin's Two-Stage Method

Let us denote the general linear model as $Y_t=\sum_{i=1}^{k}\beta_iX_{it}+u_t$, $t=1, 2, \cdots n$. And assume the first-order Markov scheme; $u_t=\rho u_{t-1}+\epsilon_t$. By substituting the latter equation into the first, we have $Y_t=\sum_{i=1}^{k}\beta_iX_{it}+\rho u_{t-1}+\epsilon_t$.

Since $u_{t-1}=Y_{t-1}-\sum_{i=1}^{k}\beta_iX_{it-1}$,

$$Y_t=\sum_{i=1}^{k}\beta_iX_{it}+\rho(Y_{t-1}-\sum_{i=1}^{k}\beta_iX_{it-1})+\epsilon_t$$

$$\text{or } Y_t-\rho Y_{t-1}=\sum_{i=1}^{k}\beta_iX_{it}-\rho\sum_{i=1}^{k}\beta_iX_{it-1}+\epsilon_t=\sum_{i=1}^{k}\beta_i(X_{it}-\rho X_{it-1})+\epsilon_t.$$

Now set $V_t=Y_t-\rho Y_{t-1}$, $W_{it}=X_{it}-\rho X_{it-1}$.

First compute the least-squares regressions of Y_t on Y_{t-1} and that of X_{it} on X_{it-1} for all i. This gives us the values of $-\rho$, and with that one can compute V_t and W_t for all i.

Then form $V_t=\sum_{i=1}^{k}\beta_i W_{it}+\epsilon_t$ and obtain β_i by the least-squares method.

It is very interesting to see how this autocorrelation problem arises in the distributed lag models. Z. Griliches made an interesting study on this subject.[1] He considers the simple model first:

$$y_t=\nu y_{t-1}+u_t,$$

$$u_t=\rho u_{t-1}+e_t,$$

where e_t is an independently distributed random variable with mean zero. In other words, we have autocorrelated residuals in the model as specified in the second equation.

Then we have $y_t=(\nu+\rho)y_{t-1}-\nu\rho y_{t-2}+e_t$.

Now suppose that we have $y_t=Cy_{t-1}+u_t$ and estimate the C without considering autocorrelation problem. Then the impact of autocorrelation is the effect of an omitted variable y_{t-2} on the coefficient of the included variable y_{t-1}. Then he derives

$$E(C-\nu)=\frac{\rho(1-\nu)}{1+\nu\rho}$$ [2]

It is obvious from this result that ν will be over estimated if ρ is positive and vice versa. The result of bias for different ν and ρ is given by the writer as follow:

ν	ρ	Approximate bias:
.1	.1	.09
.2	.5	.44

1) Zvi Griliches, "A Note on Serial Correlation Bias in Estimates of Distributed Lags," *Econometrica*, Vol. 29, 1961, pp. 65~73.
2) *Ibid.*, pp. 66.

.2	.8	.66
.5	.8	.43
.5	.5	.30
.5	.1	.07
.8	.1	.03
.8	.5	.13
.8	.8	.18

3. Empirical Works on Time Series with Autocorrelated Residuals

(1) Jorgensen and Eisner

Even though the absence of serial correlation is critical to most estimation procedures used in economics, few studies have given systematic treatment to the subject. A recent volume, *Econometric Model of the United States,* has been examined to see how the writers have treated the question of serial correlation of the residuals. Only two writers' treatments are given below.

Jorgensen formulates the theory of investment behavior as $I_t = IE_t + IR_t$, where I_t is the total investment; IE_t is the investment for capital expansion; IR_t is the investment for replacement. He further specifies $IE_t = u_o IN_t + u_1 IN_{t-1} + \ldots$, where IN_t is the level of projects initiated in period t. Then he sets $IE_t = u(\theta) K_t^e - K_{t-1}^e$, where K is the desired investment and $u(\theta) = u_o + u_1\theta + u_2\theta^2 + \ldots$, being the lag operator. He also formulates $IR_t = \delta K_{t-1}$, where the δ is the rate of depreciation and K_t is the actual capital stock. Therefore the total investment function is,

$$I_t = u(\theta) K_t^e - K_{t-1}^e + \delta K_{t-1}.^{1)}$$

He compares his investment function I_t and the 'naive' models $I_t = I_{t-1}$, and $IA_{t+2} = IA_{t+1}$, where I_t, IA_{t+2}, and IA_{t+1} represent actual investment, anticipated investment two quarters hence, and anticipated investment one quarter hence respectively. He concludes that his model gives less autocorrelation of the residuals than the naive models, saying that the Von Neumann

1) D.W. Jorgensen, "Anticipations and Investment Behavior," in *Econometric Model of the United States*, ed. by J.S. Duesenberry, G. Fromm, L.R. Klein, and E. Kuh, 1965, pp. 52.

ratios of his model is "clearly within the region of acceptance for the null hypothesis of zero autocorrelation; for the naive models, the Von Neumann ratios give very clear evidence of high positive autocorrelation." In fact he is saying that for his model there is no autocorrelation problem.

Eisner suggests that the error in the prediction given by his model may be due to the autocorrelation among the residuals. He says that the Durbin-Watson ratios frequently reveal that the residuals from the estimating equations are positively correlated. He suggests that the revised estimation procedure will give more accurate prediction, but as many other empirical workers, he does not use the more complicated procedure.

(2) Cochrane and Orcutt

They argue that the systematic residuals may arise because of a faulty choice of the form of relationship assumed to exist between economic variasble. The residuals may be autocorrelated due to the omission of variables, and simply because the most important economic time series are autocorrelated. Whatever the reasons are, Cochrane and Orcutt conclude that most current formulations of economic relations are highly autocorrelated such that it is not desirable to use the simple least-squares method of estimation.[1)]

They tested the empirical works done by Lawrence R. Klein, M.A. Girshick and T. Haavelmo, and R. Stone to see whether the residuals are autocorrelated or not. Cochrane and Orcutt used the equations, $\rho=1-\frac{1}{2}d$ where $d=\frac{\alpha^2}{s^2}$, $\alpha^2=\frac{1}{N-1}\ \Sigma\ (X_{t+1}-X_t)^2$, $s^2=\frac{1}{N}\ \Sigma\ (X_t-\bar{X})^2$. This equation is developed inthe earlier part of this paper.[2)]

The probability distribution of d^2/s^2 for a random series has been tabulated by J. von Neumann and B.S. Hart.[3)]

1) D. Cochrane and G.H. Orcutt, "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," *Journal of the American Statistical Association.* Vol. 44, 1949, pp. 36.

2) See page 5 of this paper.

3) B.S. Hart and J. von Neumann, "The Tabulation of the Probability for the Ratio of the Mean Square Successive Difference to the Variance," *Annals of Mathematical Statistics*, Vol. 13, pp. 207-214.

Using this probability distribution, Cochrane and Orcutt tests whether the economic residuals are random or correlated each other. The results of the test at five per cent level of significance and 2. 5 percent level of significance are as follow:

Source of residuals	Number of years	Number of parameters				Total
		3	4	5	6	
Klein-Econometrica[1]	22	2	7	2	1	12
Klein-Mimeographed study[2]	20	1	7	1	—	9
Girshick and Haavelme[3]	20	2	2	1	—	5
Stone[4]	19	4	6	6	1	17
Total		9	22	10	2	43
$P(d^2/s^2>k)=0.025$		7	5	4	—	16
$P(d^2/s^2>k)=0.05$		8	10	4	—	22

These results indicate that out of 43 series 16 are significantly different from a random series at the 2. 5 per cent significant level of test and 22 series are significantly different from random series at 5 per cent level. Therefore they conclude that in many cases the assumption of random error terms is not justified.

(3) Johnston's Empirical Work

He used the following two time seires to show the existence of autocorrelation and the use of transformation to get the revised estimates.
By ordinary least-squares method, he gets $\hat{Y}=7.0+0.9025X$,
where $0.9025=\frac{\sum x_i y_i}{\sum x^2}$ and $7.0=Y-0.9025\bar{X}$.
Then we have

$\hat{u}=Y-\hat{Y}$ and $\Delta\hat{u}=\hat{u}_{t+1}-u_t$.

1) L.R. Klein, "The Use of Econometric Models as a Guide to Economic Policy," *Econometrica*,
2) His unpublished work distributed by the Cowles Commission.
3) Girshick and Haavelmo, "Statistical Analysis of the Demand for Food: Examples of Simultaneous Estimation of Structural Equation," *Econometrica*, 1947, pp. 79-110.
4) R. Stone, "The Analysis of Market Demand," *Journal of Rogal Statistical Society* 1945, pp. 289-301.

Personal Disposable Income and Personal Consumpiton of U.S.
(Billions of Dollars, constant 1954 prices)

Year	Consumption	Income
	Y	X
1948	199	212
1949	204	214
1950	216	231
1951	218	237
1952	224	244
1953	235	255
1954	238	257
1955	256	273
1956	264	284
1957	270	290

Now we can compute the Durbin-Watson d statistic, $d=\frac{\sum(\Delta\hat{u})^2}{\sum\hat{u}^2}=1.07.$ Since the expected value of d is 2.11,[1] Johnston concludes that 1.07 is rather low and therefore indicate of positive autocorrelation.

Assuming that there exists the first-order Markov scheme among the residuals, *i.e.*, $\hat{u}_t=\rho\hat{u}_{t-1}+e_t$, we can estimate $\hat{\rho}$ by ordinary least-squares method.

Then $\hat{\rho}=\frac{\sum_{t=2}^{n}\hat{u}_t\hat{u}_{t-1}}{\sum_{t=2}^{n}\hat{u}_{t-1}^2}=0.457.$

Then this leads us to define the transformed variable as

$$Y_t'=Y_t-0.457Y_{t-1}$$

$$X_t'=X_t-0.457X_{t-1}.$$

Applying the ordinary least-squares to the transformed variables, Y_t' and X_t', Johnston gets $Y_t'=2.6+0.9114X_t'$. Then we also have the transformed residuals, $\hat{u}_t'$. The following shows all the computation results.[2]

1) Johnston, pp. 198.
2) *Ibid.*, pp. 199.

Year	Y_t'	X_t'	$\hat{Y}_t'$	u_t'	$\Delta u_t'$
1949	113.1	117.1	109.3	3.8	
1950	122.8	133.2	124.0	−1.2	−5.0
1951	119.3	131.4	122.4	−3.1	−1.9
1952	124.4	135.7	126.3	−1.9	1.2
1953	132.6	143.5	133.4	−0.8	1.1
1954	130.6	140.5	130.7	−0.1	0.7
1955	147.2	155.6	144.4	2.8	2.9
1956	147.0	159.2	147.7	−0.7	−3.5
1957	149.4	160.2	148.6	0.8	1.5

Now compute the d value based on the $\hat{u}_t'$and $\Delta\hat{u}_t'$.

$$d' = \frac{\sum(\Delta\hat{u}_t)^2}{\sum\hat{u}_t'^2} = 1.41.$$

Therefore d' is much closer to the expected value of a random residuals than the original d. This then indicates that the transformed residuals are less indicative of positive autocorrelation than the residuals from the original variables; it is much safer to estimate the coefficients from the transformed data than the original data.

(4) Hildreth and Lu's work

As we studied Hildreth and Lu's method of estimating ρ in the second part of this paper (pp. 120), they have contributed a significant improvement of this subject. They studied a number of empirical works done by various writers in econometric studies. The following summary is one of their studies.

In 1955, K.W. Meinken made an econometric study of the wheat industry relating the world price to the world supply. They obtained $P_w = 142 - 0.036 S_w + 1.1 I_w$, where P_w is the average wholesale price of wheat at Liverpool, England, per bushel, converted to U.S. currency at par, in cents; S_w is world production of wheat plus stocks, about August; I_w is the index of wholesale prices of 45 raw materials in England (1910−40=100).[1]

Hildreth and Lu made the Durbin and Watson's test of the residuals; they obtained the lower and upper bounds for 10 per cent significance level as 0.95 and 1.54. They calculated $d=0.8922$, and therefore rejected the hypothesis of independent residuals. Then they tested different value of ρ ranging from 1 to −1 to find out the ρ which gives the lowest log $S(\rho)$.[2] They found out that $\rho=0.64$ gives the lowest log $S(\rho)$ and therefore they set $\hat{\rho}=0.64$. Then they used the tranformation procedure and obtained the new coefficients, the result being $P_w=170.2971-0.0454398\ S_w+1.312387 I_w$.

REFERENCES

[1] Aitken, A. C. "On Least Squares and Linear Combinations of Observations," *Proceedings of the Royal Society of Edinburgh*, Vol. 55 (1935), pp.42∼48.

[2] Anderson, T. W., "On the Theory of Testing Serial Correlation," *Annals of Mathematical Statistics*, Vol. 21 (1950), pp. 59∼81.

[3] Bartlett, M. S., "Some Aspects of the Time-Correlation Problem in Regard to Tests of Significance," *Journal of the Royal Statistical Society*, Vol. 98 (1935), pp. 536∼43.

[4] ———, "On the Theoretical Specification and Sampling Properties of Autocorrelated Time Series," *Journal of the Royal Statistical Society Supplement*, Vol. 8 (1946), pp. 27∼41.

[5] Cochrane, D. and Orcutt, T. H. "Application of Least-Square Regressions to Relationships Containing Autocorrelated Error Terms," *Journal of the American Statistical Association*, Vol. 44 (1949), pp. 32∼61.

[6] Durbin J. "Estimation of Parameters in Time Series Regression Models," *Journal of the Royal Statistial Society*, BV. 22 (1960), pp. 39∼53.

1) Hildreth and Lu, pp. 30.

2) See page 11 of this paper.

〔7〕 Durbin, J. and Watson, G. S., "Testing for Serial Correlation in Least Squares Regression," *Biometrika*, Vol. 52 (1951), pp. 157～77.

〔8〕 Girshick and Haavelmo, "Statistical Analysis of the Demand for Food: Examples of Simultaneous Estimation of Structural Equation," *Econometrica*, Vol. 15 (1947), pp. 79～110.

〔9〕 Hildreth and Lu, J. Y., "Demand Relations with Autocorrelated Disturbances," *Technical Bulletin*, No. 276, Agricultural Experimental Station, Michigan State University, 1960.

〔10〕 Hart, B. S. and von Neumann, J., "The Tabulation of the Probabilities for the Ratio of the Mean Square Successive Differences to the Variance," *Annals of Mathematical Statistics*, Vol. 13, pp. 207～214.

〔11〕 Johnston, J., *Econometric Methods*, New York: McGraw-Hill, 1960.

〔12〕 Plackett, R. L., "A Historical Note on the Method of Least Squares," *Biometrika*, Vol. 40 (1949), pp. 458～60.

〔13〕 Theil, H. and Nagar, A. L., "Testing the Independence of Regression Disturbances," *Journal of the American Statistical Association*, Vol. 56 (1961), pp. 793～806.

〔14〕 Yule, G. U., "On the Time-Correlation Problem," *Journal of the Royal Statistical Society*, Vol. 84 (1921), pp. 496～537.