Statistics O.S

Netherlands School of Economics

ECONOMETRIC INSTITUTE

Report 7016

A BAYESIAN PROCEDURE FOR TESTING REGRESSION DISTURBANCES FOR HETEROSKEDASTICITY AND AUTOCORRELATION

by F.B. Lempers and T. Kloek

July 20, 1970 Preliminary and Confidential

A BAYESIAN PROCEDURE FOR TESTING REGRESSION DISTURBANCES FOR HETEROSKEDASTICITY AND AUTOCORRELATION

by F.B. Lempers and T. Kloek

Contents

Page

1. Introduction 1

2. Derivation of Posterior Probabilities for Alternative Assumptions about the Disturbances 3

3. First-Order Markov-Scheme Models 6

4. Numerical Illustrations 8

5. Some Concluding Remarks 11

References 13

1. INTRODUCTION

In this paper we apply the Bayesian procedure of posterior probabilities of alternative linear models, as discussed by Kloek and Lempers (1970), to compare linear models with alternative nonspherical assumptions about the disturbances. In the simplest version of the linear model $y = X\beta + \varepsilon$, where y is an n-dimensional vector of sample observations, X an $n \times r$ fixed matrix, β an r-dimensional vector of parameters, and ε an n-dimensional vector of disturbances, one often assumes that these disturbances are independent and normally distributed with $E\varepsilon = 0$ and $E\varepsilon\varepsilon^t = \sigma^2 I$. The most obvious generalization of these assumptions is that $E\varepsilon\varepsilon^t = \sigma^2\Omega$, where Ω is a real positive definite symmetric (PDS) matrix. This Ω matrix, however, is not manageable without restrictions, since it contains too many parameters. Therefore, one usually considers simple cases as diagonal or first-order Markov-scheme Ω matrices.

The simple assumption $E\varepsilon\varepsilon^t = \sigma^2 I$ implies that all disturbances have the same variance σ^2. This property is often denoted by the term <u>homoskedastic</u> disturbances. For many purposes this assumption is too restrictive. For example, high-income families show much greater variability in their saving behavior than do low-income families, so that the assumption of common disturbance variance would be inappropriate for a cross-section saving-income relationship. One can deal with this phenomenon by assuming that the standard deviation of the i-th disturbance is proportional with the income (per head or equivalent adult) of the i-th family. More generally, in cross-section analysis one can assume that the standard deviation of the i-th disturbance is proportional to the i-th value of some explanatory variable, or with the i-th element of $X\alpha$ for some given vector α. Similar assumptions may be made in time-series analysis, or (alternatively) in a growing economy one may assume that the standard deviation of the j-th disturbance is proportional with $(1 + g)^j$, where g is an average growth rate. If the disturbances have different variances, they are said to be <u>heteroskedastic</u>. Now we adopt the assumption of heteroskedasticity but retain the independence assumption, so that our basic linear model can be rewritten as

$$y = X\beta + \varepsilon \tag{1.1}$$

with

$$E\varepsilon = 0 \quad \text{and} \quad E\varepsilon\varepsilon^t = \sigma^2\Omega \tag{1.2}$$

where Ω is a diagonal matrix with positive diagonal elements according to a pattern that is chosen in advance. Note that if we want to compare a number of different possibilities, say k, k different matrices Ω_i $(i = 1, \ldots, k)$ are supposed to be specified. One of these matrices, of course, could be the unit matrix. This enables us to compare different assumptions about the heteroskedasticity with each other, and in addition, with the assumption of homoskedasticity.

Another simple alternative to the assumption $E\varepsilon\varepsilon^t = \sigma^2 I$ is the first-order Markov-scheme model which generates autocorrelated disturbances in a simple way. This subject has received much attention from econometricians. The Durbin-Watson (1950) article has become a classic, and Theil (1965) and Koerts (1967) developed the BLUS test. More recently, Abrahamse and Koerts (1969) derived a "new class of disturbance estimators", which are applied to the problem of testing for autocorrelation by Abrahamse and Louter (1969). All these procedures consider two hypotheses, namely, zero autocorrelation against positive (or negative) autocorrelation. If the hypothesis of zero autocorrelation has to be rejected, then the amount of autocorrelation in the model in question is not yet known. Estimation of the parameter of the autoregressive structure can be done with the maximum-likelihood method or, much simpler, with a two-stage estimation procedure as has been suggested by Durbin (1960).

Bayesian statisticians can handle this problem by assuming an uniform prior distribution about the autocorrelation coefficient (denoted by ρ) and derive a posterior distribution of ρ, given the sample. See, e.g., Zellner and Tiao (1964) and Thornber (1967). In this paper we construct k different models by choosing alternative values for the autocorrelation coefficient ρ. We do this by taking a finite number (k) of ρ values from the open interval (-1, 1). It is needless to say that no continuous posterior distribution of ρ is obtained in this way, but that a continuous distribution may be approximated to any desired degree of accuracy, by taking a sufficiently large value for k.

The order of discussion is as follows. In Section 2 we start to discuss the linear transformations which send nonspherical disturbance vectors into spherical ones and the resulting transformations of the sample statistics. We continue to recall the formula of the posterior probabilities for the case of uniform prior distributions discussed in Kloek and Lempers (1970); we also give a general approach to the problem of obtaining some definite value of the ratio of two normalizing constants of such uniform prior distributions for the cases we consider. We finish this section by deriving the ratio of posterior probabilities of two models with alternative assumptions about the pattern of the heteroskedastic disturbances. In Section 3 we consider the properties of the autocorrelated disturbances and derive their posterior probabilities. We illustrate these results with some numerical examples in Section 4. Finally, some concluding remarks are presented in Section 5.

2. DERIVATION OF POSTERIOR PROBABILITIES FOR ALTERNATIVE ASSUMPTIONS ABOUT THE DISTURBANCES

To avoid notational confusion and to obtain notational correspondence with Kloek and Lempers (1970), we rewrite the linear models we consider as

$$z = W\beta + \eta_i \qquad (i = 1, \ldots, k) \tag{2.1}$$

where η_i is a normally distributed n-dimensional vector with

$$(2.2) \qquad E\eta_i = 0 \quad \text{and} \quad E\eta_i\eta_i^t = h_i^{-1}\Omega_i$$

It can easily be seen that the alternative models only differ by the assumptions made about the disturbances. If we transform such a model by means of a matrix B_i satisfying $B_i^tB_i = \Omega_i^{-1}$, we obtain

$$(2.3) \qquad B_iz = B_iW\beta + B_i\eta_i$$

or

$$(2.4) \qquad y_i = X_i\beta + \varepsilon_i$$

where $y_i = B_iz$, $X_i = B_iW$, $\varepsilon_i = B_i\eta_i$, and conclude that ε_i is normally distributed with $E\varepsilon_i = 0$ and $E\varepsilon_i\varepsilon_i^t = h_i^{-1}I$. Now the likelihood function as discussed in Kloek and Lempers (1970), Section 3, is characterized by

$$(2.5a) \qquad N_i = X_i^tX_i = W^tB_i^tB_iW = W^t\Omega_i^{-1}W$$

$$(2.5b) \qquad b_i = (X_i^tX_i)^{-1}X_i^ty_i = (W^t\Omega_i^{-1}W)^{-1}W^t\Omega_i^{-1}z$$

if the matrix W has full column rank[1] r_i, and

$$(2.5c) \qquad \lambda_i = n - r_i$$

$$(2.5d) \qquad \lambda_iv_i = (y_i - X_ib_i)^t(y_i - X_ib_i) = (z - Wb_i)^t\Omega_i^{-1}(z - Wb_i)$$

$$(2.5e) \qquad |J_i| = |B_i| = |\Omega_i|^{-\frac{1}{2}}$$

where n is the number of observations.

We recall that our purpose is to compare alternative Ω matrices to obtain the most probable one in a given class with the help of the Bayesian procedure of posterior probabilities of alternative linear models. We shall make use of the results of Kloek and Lempers (1970), Section 5; for an explanation of the symbols used we refer to that paper. We confine ourselves to the case of uniform prior distributions, where β and $\log h$ ($h = \sigma^{-2}$) are uniformly and independently distributed. For such prior distributions the ratio of posterior probabilities of two alternative specifications, say i and j, is given by

[1] If rank [W] < r_i, then b_i denotes any solution of $(W^t\Omega_i^{-1}W)b_i = W^t\Omega_i^{-1}z$.

$$\text{(2.6)}\qquad \frac{p''_{\alpha i}}{p''_{\alpha j}} = \frac{p'_i\gamma'_{\alpha i}\gamma''_{\alpha j}|N_j|^{\frac{1}{2}}(\frac{1}{2}\lambda_i - 1)!(\frac{1}{2}\lambda_j v_j)^{\frac{1}{2}\lambda_j}(2\pi)^{\frac{1}{2}r_i}|J_i|}{p'_j\gamma'_{\alpha j}\gamma''_{\alpha i}|N_i|^{\frac{1}{2}}(\frac{1}{2}\lambda_j - 1)!(\frac{1}{2}\lambda_i v_i)^{\frac{1}{2}\lambda_i}(2\pi)^{\frac{1}{2}r_j}|J_j|}$$

for i, j = 1, ..., k. In the heteroskedastic case, as well as in the first-order Markov-scheme case, we consider a set of models with different Ω matrices, all other assumptions being equal. In both cases we have $r_i = r_j = r$ and $\lambda_i = \lambda_j = \lambda$ (i, j = 1, ..., k), so that the expression (2.6) can be considerably simplified, as follows

$$\text{(2.7)}\qquad \frac{p''_{\alpha i}}{p''_{\alpha j}} = \frac{\gamma'_{\alpha i}\gamma''_{\alpha j}|W^t\Omega_j^{-1}W|^{\frac{1}{2}}v_j^{\frac{1}{2}(n-r)}|\Omega_j|^{\frac{1}{2}}}{\gamma'_{\alpha j}\gamma''_{\alpha i}|W^t\Omega_i^{-1}W|^{\frac{1}{2}}v_i^{\frac{1}{2}(n-r)}|\Omega_i|^{\frac{1}{2}}}$$

where use has been made of (2.5a)-(2.5e) and where it has been assumed that $p'_i = p'_j$ to express the opinion that we consider all models equally probable <u>a priori</u>.

The problem of assigning uniform prior distributions to β and log h, such that we obtain a determinate value for the ratio $\gamma'_{\alpha i}/\gamma'_{\alpha j}$, has a rather natural solution for the models we consider. We observe that there is no difference in the explanatory variables and variable to be explained; compare (2.1). Therefore, since the uniform prior distribution of the parameter β is taken independent of the prior distribution of the precision (h) of a model, we immediately conclude that it is natural to take the uniform prior distributions of the parameter β identical for the models we consider. We now turn to the distributions of the precisions of the models. Our conclusions will be equally simple but they are reached in a less straightforward way. Without loss of generality we restrict ourselves to two models i and j with precisions h_i and h_j. Then, the problem arises how to specify prior distributions of h_i and h_j. A particularly interesting property is the following. The variance matrix $h^{-1}\Omega$ does not change when one replaces Ω by $\phi\Omega$ and at the same time h by ϕh. It is easily verified that the constant $1/(\log h_{u\alpha} - \log h_{\ell\alpha})$ of the marginal prior density of log h is not changed by this transformation and it will be proved later that the posterior probabilities are not changed either. Therefore, we do not see much reason to bother about the ratio

$\gamma'_{\alpha i}/\gamma'_{\alpha j}$ and simply assume that $\gamma'_{\alpha i}/\gamma'_{\alpha j} = 1$ for all α, i and j. In Kloek and Lempers (1970) we derived that $\lim_{\alpha\to\infty} \gamma''_{\alpha i} = 1$ $(i = 1, \ldots, k)$, so that the ratio of the posterior probabilities for $\alpha \to \infty$ can be written as

$$(2.8) \qquad \frac{p''_i}{p''_j} = \lim_{\alpha\to\infty} \frac{p''_{\alpha i}}{p''_{\alpha j}} = \frac{|W^t\Omega_j^{-1}W|^{\frac{1}{2}}\; v_j^{\frac{1}{2}(n-r)}|\Omega_j|^{\frac{1}{2}}}{|W^t\Omega_i^{-1}W|^{\frac{1}{2}}\; v_i^{\frac{1}{2}(n-r)}|\Omega_i|^{\frac{1}{2}}}$$

which easily can be computed. Note that if one of the Ω matrices, say Ω_i, is multiplied by an arbitrary positive number, the posterior probabilities of (2.8) are invariant for this transformation; compare (2.5a)-(2.5e). Finally we note that the application of (2.8) to the case of two heteroskedastic models is straightforward. Its application to the case of two first-order Markov-scheme models will be discussed in the next section.

3. FIRST-ORDER MARKOV-SCHEME MODELS

In this section we consider the properties of the disturbances of the linear model $z = W\beta + \eta$, where η is an n-dimensional vector of disturbances generated by a first-order Markov-scheme

$$(3.1) \qquad \eta(t) = \rho\eta(t-1) + \zeta(t) \qquad (t = 2, \ldots, n)$$

where[2] $|\rho| < 1$, and the $\zeta(t)$ are independent normally distributed random variables with zero mean and variance σ_ζ^2. Hence

$$(3.2) \qquad E\zeta = 0 \quad \text{and} \quad E(\zeta\zeta^t) = \sigma_\zeta^2 I$$

where ζ is a column vector defined by

$$(3.3) \qquad \zeta^t = [\zeta(2) \;\ldots\; \zeta(n)]$$

For simplicity we add some assumptions on $\eta(1)$: $\eta(1), \zeta(2), \ldots, \zeta(n)$ have a joint normal distribution, with moments given by (3.2) and[3]

$$(3.4) \qquad E\eta(1) = 0$$

---

[2] Note that Zellner and Tiao (1964) consider explosive ($|\rho| \geq 1$) as well as non-explosive ($|\rho| < 1$) schemes.

[3] Note that (3.6) is tantamount to assuming that $\eta(1), \ldots, \eta(n)$ are homoskedastic.

(3.5) $$E\eta(1)\zeta = 0$$

(3.6) $$\text{var } \eta(1) = \sigma_\zeta^2/(1 - \rho^2) \equiv \sigma^2$$

It follows that

(3.7) $$\eta(t) = \rho^{t-1}\eta(1) + \sum_{s=0}^{t-2} \rho^s \zeta(t - s)$$

(3.8) $$E\eta(t) = 0$$

(3.9) $$\text{var } \eta(t) = \sigma^2$$

(3.10) $$E\eta(t)\eta(s) = \rho^{|t-s|}\sigma^2$$

The complete variance matrix looks as follows

(3.11) $$E(\eta\eta^t) = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix} = \sigma^2\Omega \equiv h^{-1}\Omega$$

The determinant of Ω is given by $|\Omega| = (1 - \rho^2)^{n-1}$ and its inverse is equal to

(3.12) $$\Omega^{-1} = (1-\rho^2)^{-1} \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1 + \rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1 + \rho^2 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 + \rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}$$

Now, if we transform the model $z = W\beta + \eta$ by means of a matrix B satisfying $B^tB = \Omega^{-1}$, we obtain

$$y = X\beta + \varepsilon \tag{3.13}$$

where $y = Bz$, $X = BW$ and $\varepsilon = B\eta$, and conclude that ε is normally distributed with $E\varepsilon = 0$ and $E\varepsilon\varepsilon^t = h^{-1}I$. Given the results of Section 2 and denoting Ω_i as the Ω matrix for $\rho = \rho_i$, we find that the ratio of posterior probabilities for $\alpha \to \infty$ equals

$$\frac{p_i''}{p_j''} = \lim_{\alpha\to\infty} \frac{p_{\alpha i}''}{p_{\alpha j}''} = \frac{|W^t\Omega_j^{-1}W|^{\frac{1}{2}}\ v_j^{\frac{1}{2}(n-r)}(1-\rho_j^2)^{\frac{1}{2}(n-1)}}{|W^t\Omega_i^{-1}W|^{\frac{1}{2}}\ v_i^{\frac{1}{2}(n-r)}(1-\rho_i^2)^{\frac{1}{2}(n-1)}} \tag{3.14}$$

for $i, j = 1, \ldots, k$.

4. NUMERICAL ILLUSTRATIONS

To illustrate the results of the previous section we generated a sample according to the following model

$$z = 2.5 + 0.6w_1 - 0.94w_2 + \eta \tag{4.1}$$

where w_1 refers to the logarithm of real income per head and w_2 to the logarithm of the deflated price of spirits. Both vectors w_1 and w_2 were taken from the well-known spirits example due to Prest (1949) that deals with the demand for spirits in the United Kingdom from 1870 to 1938. The numbers 0.6 and -0.94 have been arbitrarily chosen; they represent the income elasticity and price elasticity, respectively. The disturbances were generated according to $\sigma = 0.15$ and $\rho = 0.6$. This standard deviation is rather large in comparison with the systematic part of (4.1) which is about 1.8, since all explanatory variables are about 2. It is also large in comparison with the estimated standard deviation 0.06 resulting from the application of least squares to the original sample.

We computed the posterior probabilities for 49 alternative values of ρ. These values were determined with the following scheme

(4.2) $$\rho_j = -1 + 0.04j \qquad j = 1, \ldots, 49$$

We also computed these posterior probabilities for different numbers of sample observations, namely 10, 21, 45, and 69. In Figure 1 below a graphical representation of the results can be found. The posterior probabilities for values of ρ smaller than zero are deleted since they were negligible in comparison with the others. Note that for the smallest sample ($n = 10$) no serious conclusion about the value of ρ can be drawn. However, if we choose in advance a smaller standard deviation to generate the disturbances, all posterior probabilities are more clustered around the true value of ρ. We recall that our posterior distributions of ρ are discrete. For clarity of the figure, however, we have connected the points of the mass functions by continuous curves.

Further, in Table 1, one can find the modes, as well as, the (approximated) means of the posterior distributions of ρ. This table also contains Durbin-Watson statistics based on the residuals obtained by applying least squares to the simple model (2.1) with $\rho = 0$ and the transformed model (2.4) with the modal value of ρ, respectively. The fact that the most probable values of ρ are positive is in accordance with the values found for the Durbin-Watson statistics for $\rho = 0$, which are (except for the smallest sample) considerably less than 2. The Durbin-Watson statistics for the modes of the posterior distributions of ρ are close to 2, which suggests that the autocorrelation has been removed; see Table 1.

TABLE 1. MODES AND MEANS OF POSTERIOR DISTRIBUTIONS OF
AND DURBIN-WATSON STATISTICS FOR FOUR SAMPLE SIZES

n	POSTERIOR DISTRIBUTION OF ρ		DURBIN-WATSON STATISTIC	
	MODE	MEAN	$\rho = 0$	MODE OF ρ
10	0.80	0.46	1.8039	2.1586
21	0.64	0.58	1.2143	2.1852
45	0.60	0.62	0.9709	2.1432
69	0.60	0.60	0.9338	1.9150

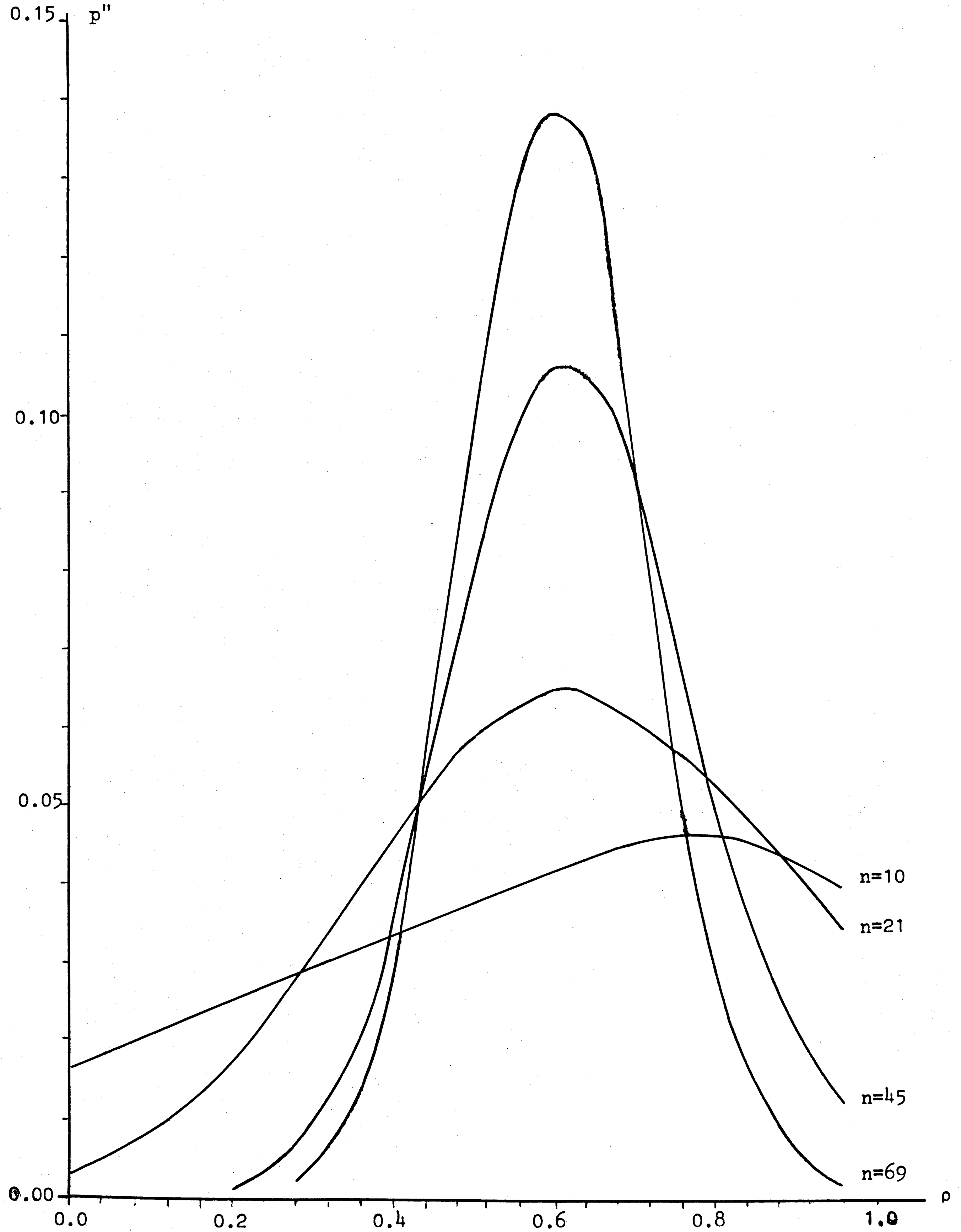

Fig. 1. Posterior probabilities for positive values of ρ and different numbers of sample observations for the spirits example with artificial data.

Finally, we also applied the procedure of posterior probabilities to the actually observed sample of the spirits example. The Durbin-Watson statistic for $\rho = 0$ is equal to 0.2488 which indicates a high positive autocorrelation. After refinement of the alternative values of ρ up to three decimal places, the highest posterior probability was obtained for ρ is 0.989, with (approximated) mean 0.979; see Figure 2. For these values the Durbin-Watson statistics are 2.0619 and 2.0344 respectively. Using $\rho = 0.989$ instead of $\rho = 0$, the posterior means of the income elasticity and price elasticity and price elasticity are 0.609 and -0.932, respectively; these results are more plausible than -0.120 and -1.228 for $\rho = 0$. It is worthwhile to note that Theil and Nagar (1961) obtained an estimate of ρ equal to 0.879. The posterior probability for this value, however, is negligible in comparison with the posterior probability for $\rho = 0.989$.

5. SOME CONCLUDING REMARKS

The numerical results of the previous section suggest that the procedure of posterior probabilities gives satisfactory results.To compare these results with earlier work in this field we refer to the already mentioned article of Zellner and Tiao (1964). They derived, under slightly different assumptions about the autocorrelated disturbances, a continuous posterior distribution of ρ given the sample z. We proceed to derive such a posterior distribution along their lines, but with the assumptions about the disturbances as presented in our Section 3. Under these assumptions the kernel of the likelihood function is given by

$$(5.1) \qquad \ell(z \mid \beta, h, \rho) \propto |\Omega|^{-\frac{1}{2}} h^{\frac{1}{2}n} \exp\left[-\tfrac{1}{2}h(z - W\beta)^t \Omega^{-1}(z - W\beta)\right]$$

Then, accepting the locally uniform and independent prior distributions

$$(5.2) \qquad D'(\beta) \propto k_1; \qquad D'(h) \propto h^{-1}; \qquad D'(\rho) \propto k_2$$

the following joint posterior distribution can be obtained

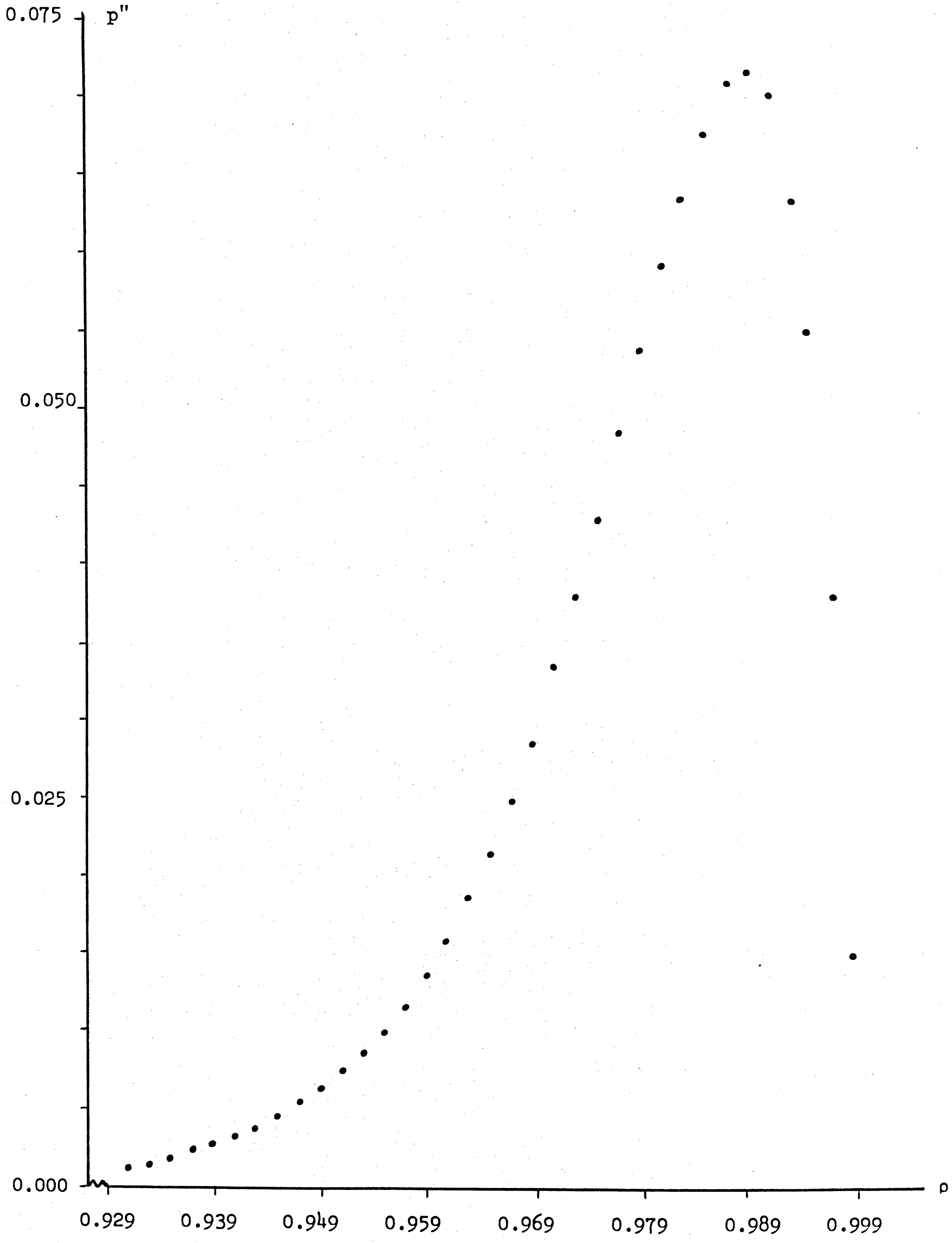

Fig. 2. Posterior Probabilities for alternative values of ρ for the spirits example with observed data.

$$(5.3) \qquad D''(\beta, h, \rho \mid z) \propto |\Omega|^{-\frac{1}{2}} h^{\frac{1}{2}n-1} \exp [-\tfrac{1}{2}h\{\lambda v + (b - \beta)^t N(b - \beta)\}]$$

where $N = W^t\Omega^{-1}W$. Upon integrating out the scale parameter h from (5.3), we obtain the following posterior distribution

$$(5.4) \qquad D''(\beta, \rho \mid z) \propto |\Omega|^{-\frac{1}{2}}[\tfrac{1}{2}\{\lambda v + (b - \beta)^t W^t\Omega^{-1}W(b - \beta)\}]^{-\frac{1}{2}n}$$

Then, the marginal posterior distribution of ρ is

$$(5.5) \qquad D''(\rho \mid z) \propto |\Omega|^{-\frac{1}{2}} v^{-\frac{1}{2}\lambda}|W^t\Omega^{-1}W|^{-\frac{1}{2}}$$

where $\lambda = n - r$, so that we conclude that the ratio of posterior probabilities with alternative values of ρ [see (3.14)] is equal to the ratio of the corresponding values of the density (5.5).

The procedure of posterior probabilities can be applied to cases where the set of alternatives does not allow of using a density function. We have seen, however, that if such a prior density exists, a straightforward Bayesian posterior distribution and posterior probabilities give identical results. In other words, the procedure of posterior probabilities is a more general approach than the existing ones mentioned above, but not essentially different.

REFERENCES

Abrahamse, A.P.J., and J. Koerts (1969), "New Estimators of Disturbances in Regression Analysis", Report 6906 of the Econometric Institute, Netherlands School of Economics. To be published in the Journal of the American Statistical Association.

Abrahamse, A.P.J., and A.S. Louter (1969), "On a New Test for Autocorrelation in Least-Squares Regression", Report 6920 of the Econometric Institute, Netherlands School of Economics. To be published in Biometrika.

Durbin, J., and G.S. Watson (1950), "Testing for Serial Correlation in Least-Squares Regression", Biometrika, 37, pp. 409-428.

Durbin, J., (1960) "Estimation of Parameters in Time-series Regression Models", Journal of the Royal Statistical Society, ser. B, 22, pp. 139-153.

Kloek, T., and F.B. Lempers (1970), "Posterior Probabilities of Alternative Linear Models", Report 7006 of the Econometric Institute, Netherlands School of Economics.

Koerts, J. (1967), "Some Further Notes on Disturbance Estimates in Regression Analysis", Journal of the American Statistical Association, 62, pp. 169-183.

Prest, A.R. (1949), "Some Experiments in Demand Analysis", The Review of Economics and Statistics, 31, pp. 33-49.

Raiffa, H., and R. Schlaifer (1961), Applied Statistical Decision Theory, Graduate School of Business Administration, Harvard University, Boston.

Theil, H., and A.L. Nagar (1961), "Testing the Independence of Regression Disturbances", Journal of the American Statistical Association, 56, pp. 793-806.

Theil, H. (1965), "The Analysis of Disturbances in Regression Analysis", Journal of the American Statistical Association, 60, pp. 1067-1079.

Thornber, H. (1967), "Finite Sample Monte Carlo Studies: An Autoregressive Illustration", Journal of the American Statistical Association, 62, pp. 801-818.

Zellner, A., and G.C. Tiao (1964), "Bayesian Analysis of the Regression Model with Autocorrelated Errors", Journal of the American Statistical Association, 59, pp. 763-778.