Bayesian analysis of heteroscedasticity in regression models

Chowdhury, S.R.; Vandaele, W.H.

Publication date:
1969

Document Version
Publisher's PDF, also known as Version of record

S. R. Chowdhury and W. Vandaele

A bayesian analysis of heteroscedasticity in regression models

Research memorandum

ECONOMIC INSTITUTE TILBURG
DEPARTMENT OF ECONOMETRICS

BAYESIAN ANALYSIS OF HETEROSCEDASTICITY IN REGRESSION MODELS

BY

S. R.CHOWDHURY AND W. VANDAELE

1. Introduction

In this paper we will try to examine heteroscedasticity in the regression model with Bayesian analysis. We will set up two types of models, one linear and the other ratio, and examine the posterior distributions of the unknown variances.

If the form of heteroscedasticity is completely known we can by Aitken's generalised least squares method find out the best linear unbiased estimates of the parameters. But before applying such a method we should first of all be able to test the presence of heteroscedasticity in the orginal model. The usual Bartlett's † test of homogeneity of variances cannot be applied because we have only one sample at our disposal.

Goldfeld, M. and R.E. Quandt { 5 } , have also tackled this problem and have given a parametric and a nonparametric test to compare the ratio and linear models.

† BARTLETT, M.S. "The Problem in Statistics of Testing Several Variances", Proceedings of the Cambridge Philosophical Society. Vol. 30, 1934.

2. Theory

Let us consider the regression model:

$$\forall_t^{1,\ldots,T} y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_n x_{nt} + u_t \qquad (1)$$

where:

- y_t = observation of the dependent variable at time t;
- $x_{1t}, \ldots, x_{nt}$: are the observations on the n explanatory variables at time t; these values are nonstochastic and identical in repeated samples; the first variable is the usual constant and takes the value 1;
- u_t : is the disturbance at time t;
- There are T independent observations on the dependent and explanatory variables;
- $T > n$.

It is assumed that $\forall_{t,t'}^{1,\ldots,T}$

- $E(u_t) = 0$; (2)
- $E(u_t u_{t'}) = 0$, $t \neq t'$; (3a)
- $E(u_t^2) = \phi_t^2 \sigma^2$, where σ^2 is unknown, but ϕ_t^2 is known ; (3b)
- that the error terms are normally and independently distributed. (4)

Bayesian analysis.

To apply the Bayesian analysis, we first make a transformation of (1) :

$$\forall_t^{1,\ldots,T} \frac{y_t}{\phi_t} = \beta_1 \frac{x_{1t}}{\phi_t} + \beta_2 \frac{x_{2t}}{\phi_t} + \ldots + \beta_n \frac{x_{nt}}{\phi_t} + \frac{u_t}{\phi_t} \qquad (5)$$

Model (5) may be called a ratio model.

Now $\forall_t^{1,\ldots,T}$

$$E\left(\frac{u_t}{\phi_t}\right) = 0 \qquad \text{, because of (2)} \tag{6}$$

$$E\left(\frac{u_t^2}{\phi_t^2}\right) = \sigma^2 \qquad \text{, because of (3b)} \tag{7}$$

Because of (7), the model (5) is a homoscedastic one.

For simplicity, we write the ratiomodel (5) as

$$\forall_t^{1,\ldots,T} y_{t,\phi} = \beta_1 x_{1t,\phi} + \beta_2 x_{2t,\phi} + \ldots + \beta_n x_{nt,\phi} + \\ + u_{t,\phi} \tag{8}$$

Under the assumption (4), the likelihood function of the sample is given by

$$\ell(\beta_1,\ldots,\beta_n,\sigma|y) \propto \frac{1}{\sigma^T} \exp\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_{t,\phi} - \\ - \beta_1 x_{1t,\phi} - \ldots - \beta_n x_{nt,\phi})^2\}$$

Or in matrix notation,

$$\ell(\beta,\sigma|y) \propto \frac{1}{\sigma^T} \exp\{-\frac{1}{2\sigma^2}(y_\Phi - X_\Phi\beta)'(y_\Phi - X_\Phi\beta)\} \tag{9}$$

where

$$\beta' = (\beta_1,\ldots,\beta_n)$$

$$y_\Phi' = (y_{1,\phi}, y_{2,\phi},\ldots, y_{n,\phi})$$

$$X_\Phi = \begin{bmatrix} x_{11,\phi} & x_{21,\phi} \cdots & x_{n1,\phi} \\ x_{12,\phi} & & \\ \vdots & & \\ x_{1T,\phi} & \cdots & x_{nT,\phi} \end{bmatrix}$$

Throughout this paper we shall use the symbol $Q(\beta, \alpha, A)$ to denote a quadratic form in variables β centred at α and with matrix A, namely

$$Q(\beta, \alpha, A) \equiv (\beta - \alpha)' A(\beta - \alpha)$$

We can now write (9) as :

$$\ell(\beta,\sigma|y) \propto \frac{1}{\sigma^T} \exp\{-\frac{1}{2\sigma^2}[Q(\beta,\hat{\beta},Z) + \nu S^2]\} \qquad (10)$$

where

$$Z = (X_\Phi' X_\Phi)$$

$$\hat{\beta} = Z^{-1} X_\Phi' y_\Phi$$

$$\nu = T - n$$

$$S^2 = \frac{(y_\Phi - X_\Phi\hat{\beta})'(y_\Phi - X_\Phi\hat{\beta})}{\nu}$$

It can be seen that $\hat{\beta}$, and S^2 are the usual least squares estimates of β and σ^2.

Using the Bayes's theorem, the likelihood function in (10) is combined with a prior distribution $p(\beta,\sigma)$ of the parameters β and σ to yield a joint posterior distribution $p(\beta,\sigma|y)$ for these parameters, that is

$$p(\beta,\sigma|y) = Kp(\beta,\sigma)\ell(\beta,\sigma|y) \qquad (11)$$

where :

$$K^{-1} = \int_R p(\beta,\sigma)\,\ell(\beta,\sigma|y)\,d\beta d\sigma$$

Clearly the form of our posterior distribution β and σ will depend on what prior distribution we adopt.

Jeffreys, H. ({ 6 }, pp. 179-192), Savage, L.J.† and Box, G.E.P. & G.C. Tiao†† suggested that in situations where little is known about β and σ, the prior distributions of β and log σ should be taken as locally independent and uniform. In the literature this type of prior is usually known as an "Noninformative prior".

In our case, we also adopt such prior distributions††† : that is :

$$p(\beta) \propto k_1 \qquad -\infty < \beta < \infty$$

$$p(\log \sigma) \propto k_2 \quad \text{or} \quad p(\sigma) \propto 1/\sigma \qquad 0 < \sigma < \infty$$

the joint prior distribution of β and σ is

$$p(\beta,\sigma) \doteq p(\beta)p(\sigma)$$

$$p(\beta,\sigma) \propto 1/\sigma \qquad 0 < \sigma < \infty \qquad -\infty < \beta < \infty \tag{12}$$

† SAVAGE, L.J. "Bayesian Statistics". In <u>Decision and Information Processes</u>. New York, Macmillan and Co., 1962.

†† BOX, G.E.P. and G.C. TIAO. "A further look at robustness via Bayes theorem", <u>Biometrika</u>. Vol. 49, 1962, nr 3/4, pp. 419-433.

††† The case where β has an informative prior e.g. a multivariate normal distribution, will be examined in future.

Substituting (10) and (12) in (11), the joint posterior distribution of β and σ is :

$$p(\beta,\sigma|y) \propto \frac{1}{\sigma^{T+1}} \exp\left\{ -\frac{1}{2\sigma^2}\left[Q(\beta,\hat{\beta},Z) + \nu S^2 \right]\right\}$$

Integrating this joint density function over β, by the properties of multivariate normal distribution, we get the marginal posterior distribution of σ.

$$p(\sigma|y) = \int_{-\infty}^{+\infty} p(\beta,\sigma|y)\,d\beta$$

$$\cancel{\sigma^{-(\nu+1)} \exp\left\{-\frac{1}{2\sigma^2} S^2\right\}} \propto \sigma^{-(\nu+1)} \exp\left\{-\frac{1}{2\sigma^2} \nu S^2\right\} \qquad (13)$$

It is to be noted that the expression νS^2 is just the residual sum of squares in the least squares regression.

It can be seen that the marginal posterior distribution of σ, i.e., $p(\sigma|y)$ in (13) is an Inverted-gamma-2 normalized density function (Raiffa, H. and R. Schlaiffer { 9 } p. 228) :

$$f(\sigma|S,\nu) \equiv \frac{2\, e^{-\frac{1}{2}\nu S^2/\sigma^2} \left(\frac{1}{2}\nu S^2/\sigma^2\right)^{\frac{1}{2}\nu+\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\nu\right) \left(\frac{1}{2}\nu S^2\right)^{\frac{1}{2}}} \qquad \begin{array}{l} \sigma \geqslant 0 \\ S,\nu > 0 \end{array} \qquad (14)$$

Its first two moments are .

Mean : $$\mu_1 = S\sqrt{\tfrac{1}{2}\nu}\, \frac{\Gamma\left(\frac{1}{2}\nu-\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\nu\right)} \quad , \nu > 1 \qquad (15)$$

Variance : $$\mu_2 = S^2 \frac{\nu}{\nu-2} - \mu_1^2 \quad , \nu > 2 \qquad (16)$$

The mode is : $$S\sqrt{\frac{\nu}{\nu+1}} \qquad (17)$$

The theory that is given in the previous sections will now be applied to analyse the heteroscedasticity.

In most of the econometric problems heteroscedasticity is usually due to the variances of the disturbance terms being dependent on the explanatory variables. The most frequent form of heteroscedasticity results from standard deviation being proportional to the values of one of the explanatory variables (Fisher, G. { 2 }, p. 156; Glejser, H. { 3 }, p. 3; Goldberger, A. { 4 }, p. 245; Johnston, J. { 7 } p. 210).

In view of the above, we take for ϕ_t's in (5) the values of one of the n explanatory variables.† The posterior distribution of σ, its mean and variance are calculated. This procedure is repeated for all the explanatory variables. The n posterior distributions, their means and variances can be analysed from the point of heteroscedasticity. Theoretically, we can conclude that the model with the sharpest posterior distribution $p(\sigma|y)$ is the least heteroscedastic in comparison to the other n-1 posterior distributions, and one will naturally choose that one if the criterion is homoscedasticity. It is to be noted that our original model is obtained by taking each of the ϕ_t's equal to unity, and is included within the n models. In the next section, as illustrations, we will analyse two numerical examples.

---

† Of course the ratio model is meaningless if any x_{nt}-value is zero.

3. Illustrations

3.1 As a first illustration, we choose a Rate of inventory formation, equation based on United Kingdom figures from 1951-1966 :

$$N_t = \beta_1 + \beta_2(\tilde{N}_{-1}/\tilde{V}'_{-1})_t + \beta_3 H_t + \beta_4 \tilde{V}'_t + \beta_5 K_t + u_t$$

Explanation of Symbols†

$N = \Delta \tilde{N}/\tilde{V}'_{-1}$: the dependent variable represents the inventory changes, their rate of change being expressed as a percentage of lagged total expenditure less inventory changes and net invisibles ;

$\tilde{N}$ = inventory changes ;

H = labour cost per unit ;

$\tilde{V}'$ = total expenditure less inventory changes and net invisibles ;

$\tilde{K}$ = gross profits per unit of output ;

$K = \Delta \tilde{K} \times 100$.

There are in this equation five explanatory variables (constant included). So there will be five posterior distributions which are indexed by positive integers 1.1 to 1.5. The posterior distribution indexed by 1.1 is based on the orginal equation (ϕ_t's = 1). Other posterior distributions are derived by dividing by the values of the explanatory variables e.g. the posterior distribution indexed by 1.2, is derived from the ratio model where ϕ_t's take the values of $(\tilde{N}_{-1}/\tilde{V}'_{-1})_t$, at different time periods.

---

† Symbols without special indication refer to relative changes. Absolute quantities are indicated by ∿.

The calculated results are given in table 1.

The values of the Poisson distributions are used to determine the posterior density functions $p(\sigma|y)$. This was possible because cumulative Inverted gamma-2 distribution is related to cumulated Poisson distribution (Raiffa, H. and R. Schlaifer { 9 }, p. 228). The whole work has been done on a IBM 1620^{II} computer.

From table 1. it is seen that the posterior distribution nr 1.4 has the lowest variance (posterior variance).

In figure 1. the graphs of the posterior distributions are given. From the figure we see that the posterior distribution nr 1.4 is the sharpest one, which as we expected.

The posterior variance of the orginal model is 38 times the variance of the sharpest distribution :

$$\frac{\text{posterior variance of } p(\sigma|y) \text{ nr } 1.1}{\text{posterior variance of } p(\sigma|y) \text{ nr } 1.4} = \frac{.0106551}{.0002805} = 37.986.$$

We can safely infer from the above that the orgi-nal model is significantly heteroscedastic in comparison with the model with the sharpest distribution. We note that, it may also turn out by the above kind of analysis, that our orginal model is the least heteroscedastic in comparison to the other ratio models.

We may also presume that the β coefficients in the least heteroscedastic model are more accurately estimated.

Table 1.

Posterior distribution $p(\sigma \mid y)$	Mean	Variance	$\overline{R}$ †	Disturbance variance S^2	Values of the regression coefficients β_1	β_2	β_3	β_4	β_5
example 1: Rate of inventory formation									
Nr. 1.1	.4323	.0106	.9763	.1616	-1.1008	-1.5007	.1876	.3584	.6304
1.2	.6070	.0210	.9503	.3186	-1.6772	-1.3304	.2631	.3969	.6345
1.3	.1926	.0021	.8981	.0321	-1.7023	-1.6460	.4197	.3780	1.1690
1.4	.0701	.0002	.9554	.0042	-1.2344	-1.5450	.2346	.3655	.7546
1.5	1.1982	.0818	.9760	1.2417	-1.9477	-1.4946	.3339	.4019	.8810
example 2: Induced investment									
2.1	2.8434	.4138	.8634	7.0823	15.5816	-1.1585	.2732	-.1220	-1.0369
2.2	.5455	.0152	.9798	.2606	17.5651	-1.4527	.3168	-.1304	-1.3249
2.3									
2.4	.1767	.0015	.9789	.0273	14.7125	- .8898	.2916	-.0363	-1.1560
2.5	1.4530	.1080	.9882	1.8495	12.0189	- .6294	.2336	-.1478	- .3059

† Multiple correlation coefficient, adjusted for degrees of freedom.

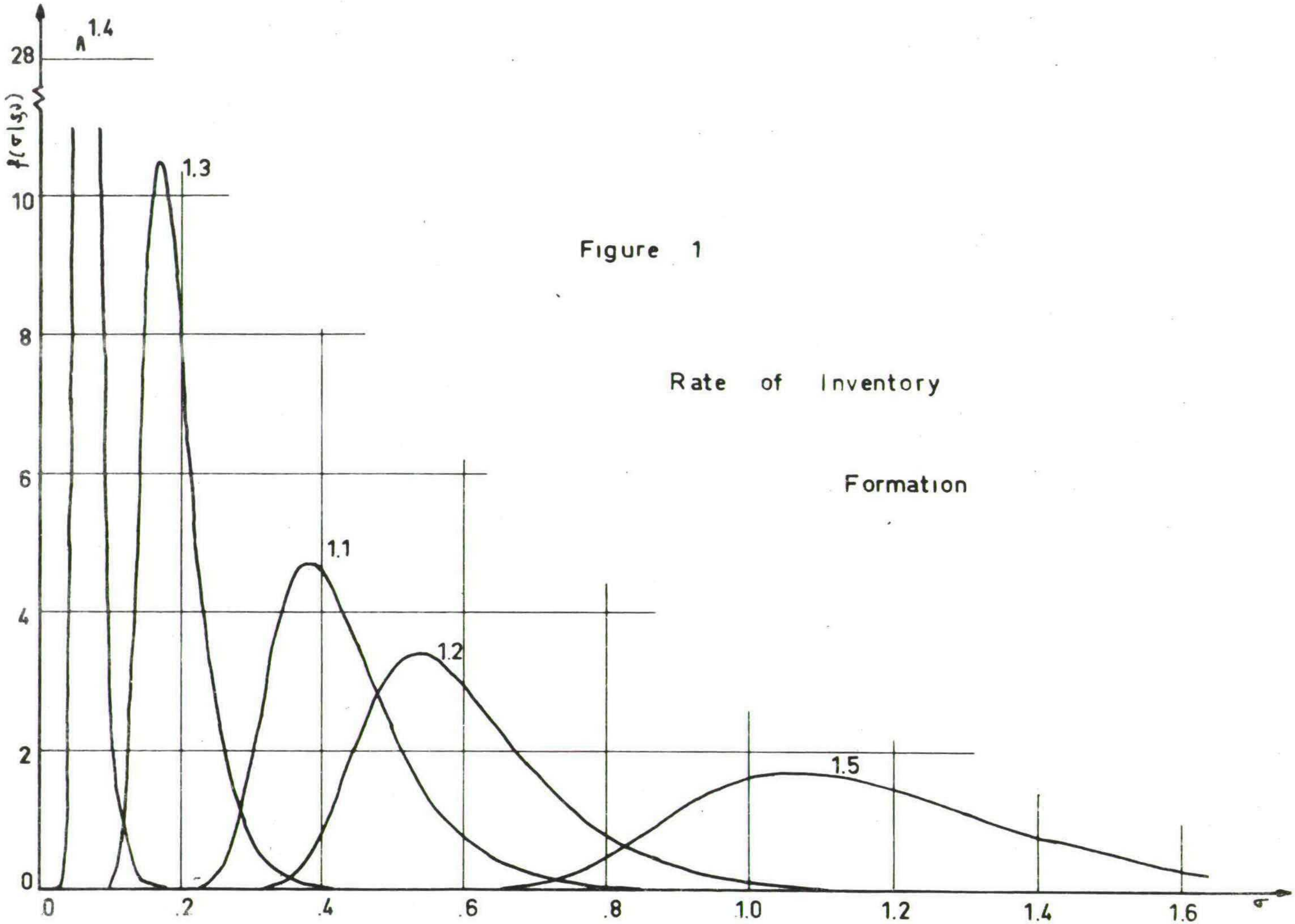

Figure 1
Rate of Inventory
Formation
$f(\sigma|s_j)$
28
10
8
6
4
2
0
1.4
1.3
1.1
1.2
1.5
0
.2
.4
.6
.8
1.0
1.2
1.4
1.6
σ

3.2 Induced investment equation.

This equation is based on United Kingdom figures from 1950-1966 :

$$I_t = \beta_1 + \beta_2(Z_{-1} - T_z'')_t + \beta_3 \, r_{-1/2,t} + \beta_4 \, W_t + \beta_5 \, P_{ai-1,t} + u_t$$

Explanation of Symbols†

I	=	induced investment ;
Z	=	non-labour income ;
$T_z'' = \Delta \ (T/Z)$	:	change of tax rate on non-labour income ;
$(Z_{-1} - T_z'')$	=	difference in change of tax rate on non-labour income and the change in lagged non-labour income itself ;
r	=	Bank rate ;
W	=	registered wholly unemployed ;
P_{ai}	=	price index of autonomous investment.

In this example the posterior distribution indexed by 2.3, cannot be derived because one of the values of $r_{-1/2}$ is exactly zero.

The results of the calculations are also given in above table 1., and the posterior distributions are plotted in figure 2.

Here we find that the $p(\sigma|y)$ indexed by 2.4 is the sharpest.

The posterior variance of the orginal model is now 259 times the variance of the sharpest distribution.

---

† Same remarks as with the symbols in the first example.

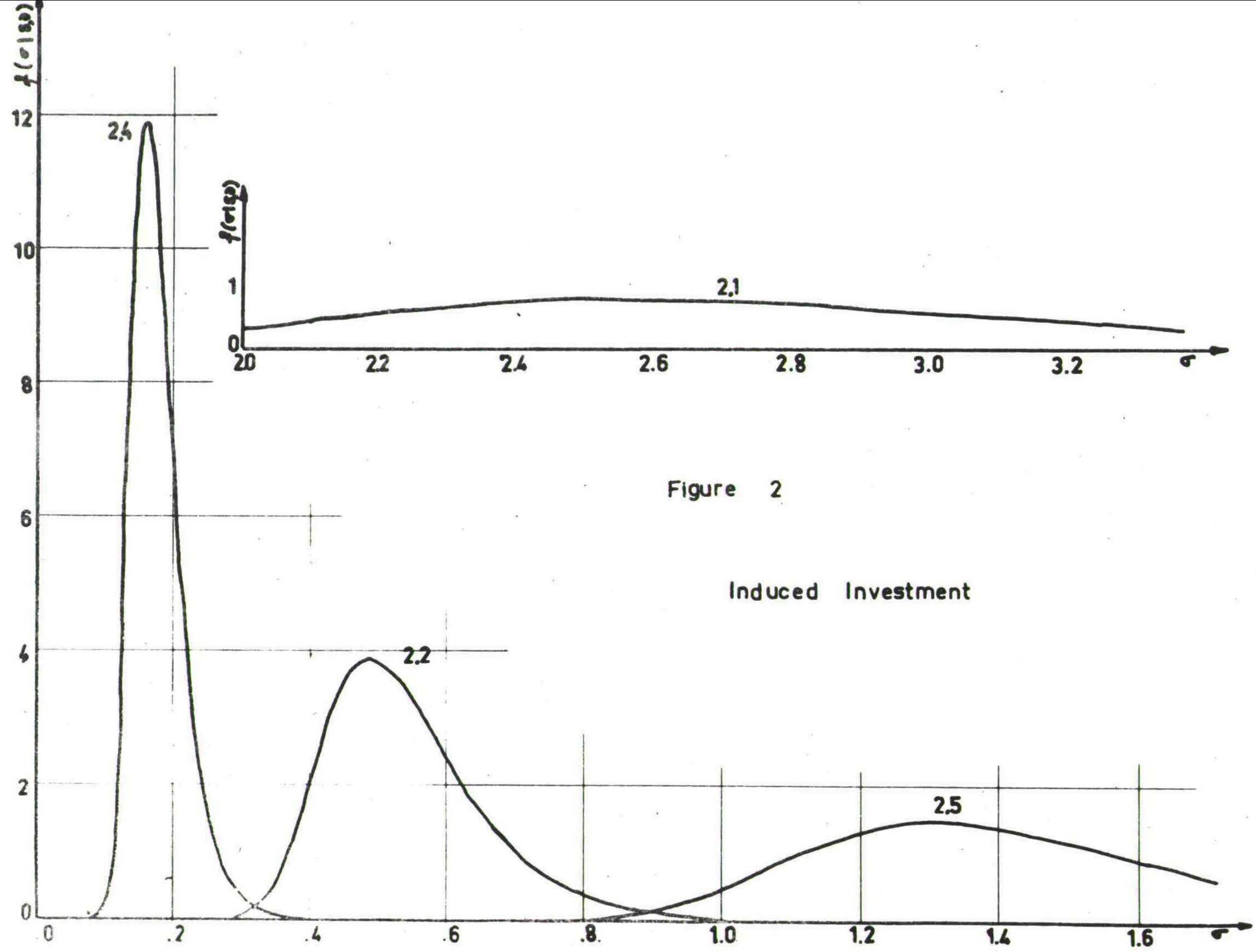

Figure 2

Induced Investment

4. Conclusion

The analysis given in the preceding pages is a simple way of detecting and correcting heteroscedasticity where heteroscedasticity is in the form of standard deviations being proportional to one of the explanatory variables.

It is a comparative procedure based on the posterior distributions, and not a direct test. Nevertheless it is quite reasonable to make such analysis for heteroscedasticity. With high-speed computer this type of analysis will not involve much additional work.

5. References

{ 1 } ANDERSON, T.W. An Introduction to Multivariate Statistical Analysis. New York, John Wiley & Sons, 1958, 374 pp.

{ 2 } FISHER, Gordon R. "Iterative Solutions and Heteroscedasticity in Regression Analysis", Review of the International Statistical Institute. Vol. 30, 1962, nr 2, pp. 153-159.

{ 3 } GLEJSER, H. Testing Heteroscedasticity in regression disturbances. Warsaw, Mimeographed paper presented at the joint European meeting of the Econometric Society and the Institute of Management Sciences, september 1966, 13 pp.

{ 4 } GOLDBERGER, Arthur S. Econometric Theory. New York, John Wiley & Sons, 1964, pp. 231-246.

{ 5 } GOLDFELD, Stephen M. and Richard E. QUANDT. "Some test for Homoscedasticity", Journal of the American Statistical Association. Vol. 60, june 1965, nr 310, pp. 539-559.

{ 6 } JEFFREYS, H. Theory of Probability. Oxford, Clarendon Press, 1961, 3rd edition, 459 pp.

{ 7 } JOHNSTON, J. Econometric Methods. London, McGraw Hill Book, 1963, pp. 207-211.

{ 8 } MALINVAUD, Edmond. Méthodes Statistiques de l'Econométrie. Paris, Dunod, 1964, pp. 261-265.

{ 9 } RAIFFA, Howard and Robert SCHLAIFFER. Applied Statistical Decision Theory. Boston, Division of Research, Graduate School of Business Administration, Harvard University, 1961, 356 pp.

{ 10 } ROTHENBERG, T. A Bayesian analysis of simultaneous equations systems. Rotterdam, Econometric Institute Report 6315, 1963, 20 pp.

{ 11 } RUTEMILLER, Herbert C and David A BOWERS. "Estimation in a Heteroscedastic Regression Model", Journal of the American Statistical Association. Vol. 63, june 1968, nr 322, pp. 552-557.

{ 12 } SAVAGE, Leonard J. The Foundations of Statistics. New York, John Wiley and Sons, 1954, 294 pp.

{ 13 } TIAO, George C. and Arnold ZELLNER. "Bayes's theorem and the use of prior knowledge in regression analysis", Biometrika. Vol. 51, 1964, nrs 1/2, pp. 219-230.

E.T.I. 1969