JIA 117 (1990) 407-413

COMMENTS ON A MODEL OF RETAIL PRICE INFLATION

BY DR A. KITTS

(Department of Social Statistics, University of Southampton)

ABSTRACT

The paper comments on the estimation and sensitivity of the retail price inflation component of the stochastic financial model proposed by Professor Wilkie. Statistical tests provide evidence of non-independence and non-normality of residuals, suggesting non-linearity. However, it is noted that the model is most sensitive to the assumption of long-term mean inflation.

KEYWORDS

Inflation; Model; Simulation; Time-series

1. INTRODUCTION

Stochastic financial models are increasingly being applied by actuaries and financial economists to incorporate the effects of random variation in absolute and relative levels of inflation, interest rates, and investment returns. Unfortunately, while the relative levels of economic and financial variables are broadly consistent, their absolute levels are difficult to predict. Therefore, stochastic financial models are particularly sensitive to assumptions concerning the process governing their absolute levels.

In this paper, the determinant of the absolute levels of variables of one particular stochastic financial model for the United Kingdom is considered—the stochastic investment model for actuarial use, proposed by Wilkie (1984). In this model, essentially, a stochastic model of retail price inflation drives a composite model. Here, the estimation of the retail price inflation component of the model is reconsidered, and its sensitivity is discussed.

2. THE STOCHASTIC INVESTMENT MODEL PROPOSED BY WILKIE

The stochastic investment model proposed by Wilkie uses simulation techniques to generate series of values of retail price inflation, equity dividend yields and dividend values (and hence equity values), and the yields on British Government 2·5% Consols. For each series, values are generated by combining a systematic component (calculated from past values of the series and possibly also past and present values of other series), and a stochastic component (generated from standard normal random variables). For each annual period, first, a value for the Retail Price Index (RPI) is generated by combining past RPI values and a random component. Then, values for the other three variables are generated. Thus, the composite model is essentially driven by the retail price inflation model.

The model proposed for the series of RPI values, $Q[t]$, is (Wilkie, 1984, 1986, 1986a):

"...

$$\nabla \log_e\{Q[t]\} = QMU + QA \cdot (\nabla \log_e\{Q[t-1]\} - QMU) + QSD \cdot QZ[t]$$

where the backwards difference operator ∇ is defined by

$$\nabla X[t] = X[t] - X[t-1]$$

and $QZ[t]$ is a sequence of independent identically distributed standard normal variates.

This model says that the annual rate of inflation follows a first order auto-regressive process, with fixed mean QMU, and a parameter QA such that the expected rate of inflation each year is equal to the mean plus QA times last year's deviation from the mean. Appropriate values for the parameters are:

$$QMU = 0{\cdot}05, \quad QA = 0{\cdot}6, \quad QSD = 0{\cdot}05.$$

..."

The model provides that each expected value of the series will lie between the previous value of the series and a mean value; the model is auto-regressive, an ARIMA (1,1,0) model. Thus, retail price inflation values for adjacent years are assumed to be identically distributed, but not independent. In this sense, the model is more sophisticated, for example, than the model of independent and identically distributed annual interest rates assumed by Boyle (1976).

The resulting series of RPI values generated by the model, with neutral initial conditions, can be described as follows. The retail price inflation generated for a period will have a normal distribution, with constant mean, and standard deviation increasing at a decreasing rate over time towards an upper limit. The RPI value calculated for a period will have a log-normal distribution with mean increasing linearly over time, and standard deviation increasing at a decreasing rate towards an upper limit. For further details, see Box & Jenkins (1976).

It is perhaps worth noting that the expected values generated by the stochastic model are not equal to those produced by the equivalent deterministic model, because of the normal to log-normal transformation.

3. ESTIMATION

The development of the retail price inflation model is based on an examination of data for the period 1661–1980 (Wilkie, 1981). Perhaps the most important feature of the data is that the structure of the process appears to change over time, particularly during the present century. This feature leads to a consideration of shorter sub-periods: 1661–1914, 1896–1980, and 1946–80. The actual estimation of the retail price inflation model is based on data for three more recent periods: 1919–82, 1933–82, and 1946–82 (Wilkie, 1984).

In order to investigate the adequacy of the model, residuals are calculated as the differences between actual and expected inflation. Two types of test are conducted. First, to investigate whether the residuals can be assumed to be independent, the runs test of the auto-correlation model function (correlation between lagged residuals) is considered. It is reported that, "at a roughly 5% level" the RPI model with rounded parameters "passes these tests on every count". Second, to investigate whether the residuals can be assumed to have a normal distribution, the values and standard errors of the skewness and kurtosis coefficients of the distribution of residuals are examined. It is concluded that the residuals for the period 1919–82 have an approximately normal distribution, "although somewhat negatively skewed and definitely fat-tailed".

In the remainder of this section, the residuals are re-examined. Considering independence, "it is important to realise that the residuals are necessarily correlated even if the true errors are independent" (Chatfield, 1980). This is because the model parameters are (by definition) estimates of the true parameters. Nevertheless, the correlations ought to be close to zero, as the larger their absolute values, the further the estimated parameters are likely to be from the true parameters. For the residuals under consideration, the aggregate tests mentioned previously give little indication of significant correlations. Also, Steven's test for the grouping of signs (see Benjamin & Pollard, 1980) provides

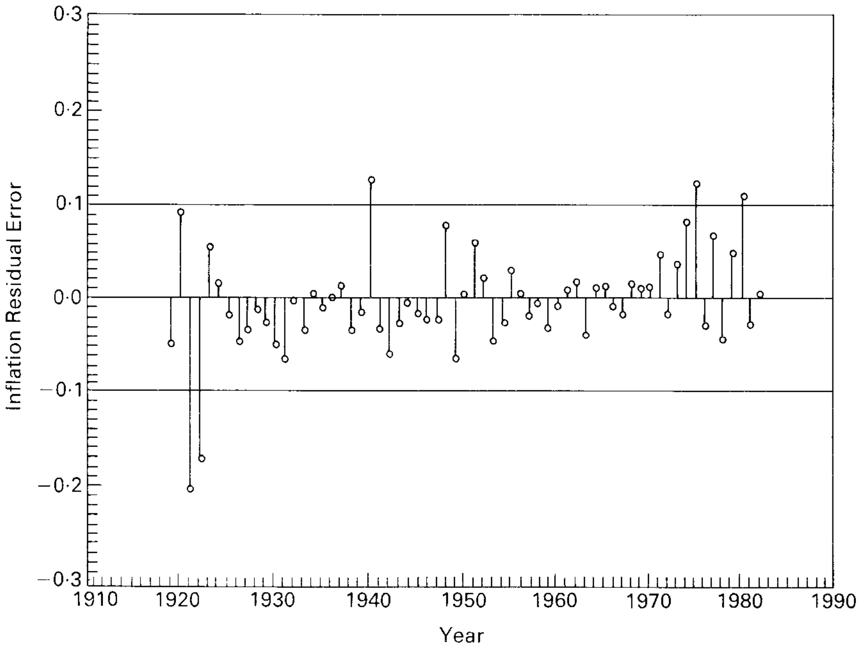

Figure 1. Force of inflation 1919–1982: residual errors (actual-expected). (Source Wilkie, 1984.)

little evidence of correlation. However, an examination of the plot of residuals presented in Figure 1 suggests that there might be significant correlation.

Why, then, are none of the test statistics significant? There are two reasons. First, the power of the tests is limited by the small number of residuals; for example, for the period 1946–82, the correlation coefficient with lag one is $-0{\cdot}09$, but with only 36 observations, this is not significantly different from zero, even at the 20% level. Second, the above tests do not examine the lengths of sequences of residuals of the same sign; for example, there are two periods, 1924–33 and 1940–7, during which there are nine and seven consecutive negative residuals, respectively. It is possible to construct tests to examine whether such sequences of residuals are acceptable. For a sequence of 64 independent random variables from a standard normal distribution, the probability of two or more sequences consisting of seven or more consecutive negative numbers, is estimated as just 0·0175 (estimated standard error 0·0002, from 100,000 simulations). Alternatively, the probability of two or more sequences of seven or more consecutive numbers of the same sign, is 0·0699 (estimated standard error 0·0008, from 100,000 simulations), which is not significant at a 5% level, but is nevertheless uncomfortably low. The method of calculation of these probabilities is given in the Appendix. The effect is that some residual correlation remains. Therefore, the model under-generates sustained periods of retail price deflation and inflation; in particular, sustained periods of extreme deflation or inflation are under-generated.

Turning to a consideration of normality, the residuals for the period 1919–82 do not have an even approximately normal distribution; they fail a Filliben (1975) test for normality at the 0·5% level (Filliben Test Correlation Coefficient 0·945). Further, it is noted that an examination of the skewness and kurtosis of the residuals for the period 1933–82 and 1946–82 suggests that the distribution of residuals is becoming less normal. The non-normality of the residuals is illustrated in Figure 2, which shows the normal distributions estimated from the residuals superimposed upon a histogram of the residuals themselves. It can be seen that the normal distribution under-estimates the proportion of large and small negative residuals, but over-estimates the proportion of medium negative residuals; also, it over-estimates the proportion of large positive residuals.

It is interesting to compare these results with the study of 40 daily financial time-series conducted by Taylor (1986). Finding residual correlation over a wide range of lags, the study concludes that linear models (including ARIMA models) are inadequate because of changes in either the unconditional or conditional variances of returns, and proceeds by fitting non-linear models.

4. SENSITIVITY

The non-independence and non-normality of residuals outlined in Section 3 indicate that, on the one hand, in the short-term, the model cannot be expected to reflect the process of retail price inflation particularly closely. Further, the

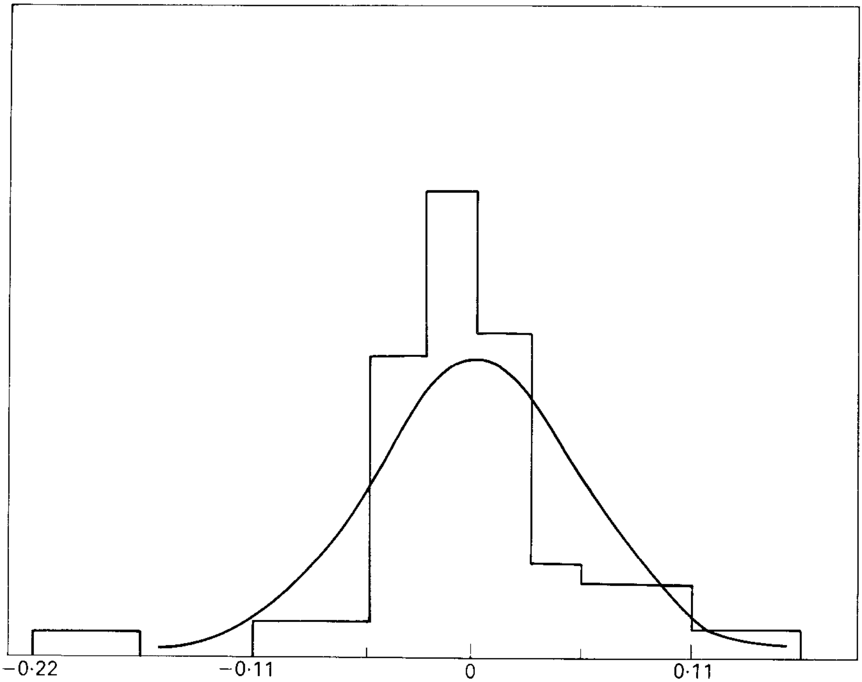

Figure 2. Force of inflation 1919–1982: distribution of residuals (actual and estimated normal).

problems are amplified because the stochastic component of the model is large relative to the systematic component. For example, the probability of deflation in a particular year approaches 21%.

On the other hand, in the long-term, the model will, of course, reproduce the overall rate of retail price inflation inherent in the data. However, it must then be borne in mind that the estimate of this rate is itself based on what amounts to very limited observations. Indeed, the period on which the estimate is based (1919–82) is selected from the longer period for which the data are available (1661–1982) because the structure of the process generating the data appears to change over time. For example, the period 1919–82 includes a maximum of four non-overlapping 15-year periods, with average annual rates of retail price inflation of about −2·2%, 3·6%, 3·5% and 10·8%, respectively.

5. CONCLUSION

In conclusion, it can be seen that the model of retail price inflation is relatively insensitive to the effects of non-independence and non-normality, in so far as it is ultimately more sensitive to the effects of other assumptions—particularly the validity of long-term parameter estimates. It is possible that the development of non-linear or dynamic time-series models might provide a closer fit to retail price inflation data; but, ultimately, such models will suffer the limitation of all financial time-series models—the inability to anticipate future changes in the process which generated historic data.

On this final note, however, it is appropriate to return to Wilkie's own comments on the model. He concludes that "... inflation over a long period will possibly be high and is certainly uncertain ..." (1981), and argues for the indexation of financial contracts.

6. ACKNOWLEDGMENTS

This paper was prepared as part of a programme of research in Actuarial Studies supported by Skandia Life Assurance Company Limited. The author would like to thank Phil Cooper and Professor Tim Holt for comments and discussions during its preparation. Also, thanks are extended to Peter Ford, Mike Sherris, Professor David Wilkie, and particularly the referee, for comments on earlier drafts.

REFERENCES

BENJAMIN, B. & POLLARD, J. H. (1980). *The Analysis of Mortality and Other Actuarial Statistics*. Second Edition. Heinemann, London.

BOX, G. E. P. & JENKINS, G. M. (1976). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.

BOYLE, P. P. (1976). Rates of Return as Random Variables. *Journal of Risk and Insurance*, **43:** 693–713.

CHATFIELD, C. (1980). *The Analysis of Time Series: An Introduction*. Second Edition. Chapman and Hall, London.

ENGLE, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, **50:** No. 4: 987 1007.

FILLIBEN, J. J. (1975). The Probability of Plot Correlation Coefficient Test for Normality. *Technometrics*, **17**(1): 111–17.

TAYLOR, S. (1986). *Modelling Financial Time Series*. Wiley, Chichester.

WILKIE, A. D. (1981). Indexing Long Term Financial Contracts. *J.I.A.* **108:** 299–360; and, (1982). *T.F.A.* **38:** 55.

WILKIE, A. D. (1984). 'Steps Towards a Comprehensive Stochastic Investment Model'. O.A.R.D. 36, Institute of Actuaries.

WILKIE, A. D. (1986). A Stochastic Investment Model for Actuarial Use. *T.F.A.* **39:** 341 73.

WILKIE, A. D. (1986a). Some Applications of Stochastic Investment Models. *J.I.A.S.S.* **29:** 25 51.

APPENDIX

DETAILS OF ESTIMATION OF ILLUSTRATIVE SEQUENCE PROBABILITIES

The probability, for a sequence of 64 independent random variables from a standard normal distribution, of two or more sequences consisting of seven or more consecutive negative variables, is calculated as a quarter of the probability of two or more sequences consisting of seven or more consecutive variables of the same sign. This latter probability is estimated by simulation. As a computational simplification, instead of using random variables from a standard normal distribution, and testing whether they are negative, random variables from a uniform (0,1) distribution are tested for being less than 0·5, which is theoretically equivalent.

A simple BASIC program is used to simulate 100,000 sequences of 64 random variables, and to calculate the probability estimate, and the associated standard error:

```
10 C=0
20 FOR S=1 TO 100000
30 P=RND: L=1: F7=0
40 FOR I=2 TO 64
50 U=RND
60 IF((U<0·5 AND P>=0·5) OR (U>=0·5 AND P<0·5) OR I=64)
   THEN 90
70 L=L+1
80 GOTO 110
90 IF L>=7 THEN F7=F7+1
100 P=U: L=1
110 NEXT I
120 IF F7>=2 THEN C=C+1
130 NEXT S
140 PRINT C/100000, SQR((C/100000)*((100000-C)/100000)/100000)
```

The output from this program is:

0·06986, 0·0008061