JOURNAL OF APPLIED ECONOMETRICS, VOL. 11, 573–593 (1996)

STOCK MARKET VOLATILITY AND THE BUSINESS CYCLE

JAMES D. HAMILTON AND GANG LIN
Department of Economics, University of California at San Diego, La Jolla, CA 92093-0508, USA

SUMMARY

This paper investigates the joint time series behavior of monthly stock returns and growth in industrial production. We find that stock returns are well characterized by year-long episodes of high volatility, separated by longer quiet periods. Real output growth, on the other hand, is subject to abrupt changes in the mean associated with economic recessions. We study a bivariate model in which these two changes are driven by related unobserved variables, and conclude that economic recessions are the primary factor that drives fluctuations in the volatility of stock returns. This framework proves useful both for forecasting stock volatility and for identifying and forecasting economic turning points.

1. INTRODUCTION

Stock returns are difficult to forecast, but squared stock returns are not. Scores of studies have documented significant time variation in the conditional variance of stock returns, meaning that stocks are much riskier investments at some times than others. Considerable effort has gone into modelling the dynamic behaviour of this fundamental measure of risk.[1]

The specification that has been most commonly used in such studies is the univariate GARCH model developed by Engle (1982) and Bollerslev (1986). This describes the conditional variance of stock returns as a linear function of past squared forecast errors. Although the parameters that describe such a process appear to be highly statistically significant, they are not stable over time (Lamoureux and Lastrapes, 1990; Engle and Mustafa, 1992), and the forecasts from the simple GARCH specification can be quite poor (Pagan and Schwert, 1990; Lamoureux and Lastrapes, 1993; Hamilton and Susmel, 1994). Hamilton and Susmel (1994) accounted for these findings by suggesting that stock returns are characterized by different ARCH processes at different points in time, with the shifts between processes determined by the outcome of an unobserved Markov chain.[2] They showed that such a specification can significantly improve forecasts of stock volatility.

An alternative approach to modelling time variation in stock volatility is to allow variables other than lagged stock returns to influence the conditional variance. Certainly a satisfactory

[1] See Bollerslev, Chou, and Kroner (1992) and Engle (1993) for recent surveys of this large literature.
[2] Earlier treatments by Schwert (1989b) and Turner, Startz, and Nelson (1989), in which the variance of stock returns is one of two constants depending on the outcome of an unobserved two-state Markov chain, can be viewed as a special case of this approach. Markov-switching ARCH models have also been used to analyse interest rates (Cai, 1994), inflation (Brunner, 1991; Susmel, 1994b), international stock returns (Susmel, 1994a,b), and exchange rates (Susmel, 1994b).

CCC 0883–7252/96/050573–21

Received May 1995
Revised July 1996

explanation for stock volatility must ultimately be grounded in an analysis of macroeconomic fundamentals. Fluctuations in the level of economic activity are a key determinant of the level of stock returns,[3] and a variable such as GDP is much harder to forecast when the economy is in the middle of a recession (Hamilton, 1989, pp. 377–378; French and Sichel, 1993). Campbell (1987) and Glosten, Jagannathan, and Runkle (1993) discovered that higher short-term interest rates lead one to predict higher stock market volatility, as does a larger spread between the yields on safe and risky bonds (Schwert, 1989a). Attanasio (1991) argued that dividend yields help forecast stock return volatility, while Engel and Rodrigues (1989) found oil prices and the money supply to be important factors in determining the conditional variance of bond returns. Schwert (1989a,b) investigated a number of factors that could potentially influence stock volatility, and identified the level of real economic activity as the most important determinant of the conditional variance of stock returns: confidence intervals for forecasting stock prices are widest during economic recessions. Officer (1973) and Campbell, Kim, and Lettau (1993) provided further strong evidence in support of this conclusion.

The implicit view of early students of the business cycle was that economic recessions represent distinct, identifiable events. The unemployment rate goes far above its usual level during a serious economic downturn, but obviously cannot go much below its typical value during a boom. It is natural to conjecture that the dynamic behaviour of the economy could be quite different during times of excess capacity and a slack labour market than it would be in normal times. Indeed, Schwert's (1989a,b) analysis of the effect of real economic activity on stock market volatility implicitly adopted this traditional dichotomous view of the business cycle. Schwert's strongest evidence came from regressions of measures of stock volatility on a dummy that takes the value of unity when the economy is in a recession (as determined after the fact in dates established by the National Bureau of Economic Research, or NBER) and zero otherwise.

Hamilton (1989) recently proposed a formal statistical representation of the old idea that expansion and contraction constitute two distinct economic phases. He suggested that real output growth may follow one of two different autoregressions, depending on whether the economy is expanding or contracting, with the shift between the regimes governed by the outcome of an unobserved Markov chain. Extensions of this approach and investigations of other data sets have confirmed that this is a promising way to formalize what it means for the economy to go into recession.[4]

In this paper we combine Hamilton's (1989) model of recessions with Hamilton and Susmel's model of changes in the ARCH process characterizing stock returns. We hypothesize the existence of a single latent variable (the 'state' of the economy) which determines both the mean of industrial production growth and the scale of stock volatility. This unobserved variable takes on one of a finite set of values, and is presumed to be determined by an unobserved Markov chain. One advantage of this approach over Schwert's dummy variables is that it offers a fully articulated description of the joint time series properties of economic activity and stock returns, and thus can be used to generate real-time forecasts. By contrast, since NBER dates at which recessions begin and end are not available until well after the events have taken place, a dummy variable regression such as Schwert's cannot be used to generate usable forecasts of stock

[3] See Chen, Roll, and Ross (1986), Keim and Stambaugh (1986), Fama and French (1989), Fama (1990), Schwert (1990), and Chen (1991), among others.

[4] See for example Lam (1990), Jefferson (forthcoming), Ghysels (1993), Filardo (1994), Boldin (1994), Durland and McCurdy (1994), Chauvet (1996), Kim and Yoo (1995), and Diebold and Rudebusch (1996). Less sanguine assessments of this approach include Hansen (1992) and Goodwin (1993).

volatility. The joint specification is further of potential interest in that volatility in the stock market may prove useful in forecasting the future trend in real economic activity.

Section 2 describes the class of models investigated in this study, and empirical results are presented in Section 3. Section 4 briefly concludes.

2. MODELS OF INDUSTRIAL PRODUCTION AND STOCK RETURNS

Let y_t denote the monthly growth rate of an aggregate index of industrial production. Hamilton's (1989) business cycle model takes the form

$$y_t = z_t + \mu_{s_t^\dagger} \tag{1}$$

where

$$z_t = \alpha_1 z_{t-1} + \alpha_2 z_{t-2} + \cdots + \alpha_q z_{t-q} + \varepsilon_t \tag{2}$$

Here $s_t^\dagger$ is an unobserved latent variable that reflects the state of the business cycle. For example, $s_t^\dagger = 1$ could indicate that the economy is in recession and $s_t^\dagger = 2$ that it is in expansion. In the general specification we allow $s_t^\dagger$ to assume one of $K^\dagger$ different values represented by the integers $(1, 2, \ldots, K^\dagger)$. When $s_t^\dagger = 1$, the average growth rate of industrial production is given by the population parameter μ_1 whereas when $s_t^\dagger = 2$, the average growth rate is μ_2. Equations (1) and (2) imply that the deviation of output growth from the phase-specific norm $(z_t = y_t - \mu_{s_t^\dagger})$ follows a qth-order autoregression whose innovation ε_t is assumed to be i.i.d. $N(0, \sigma^2)$.

It is also possible to make the variance of the growth rate of industrial production a function of the state s_t as well, as in Hamilton (1988). This involves replacing equation (2) with

$$z_t = \alpha_1 z_{t-1} + \alpha_2 z_{t-2} + \cdots + \alpha_q z_{t-q} + \sigma_{s_t^\dagger} \varepsilon_t \tag{2$'$}$$

for $\varepsilon_t \sim$ i.i.d. $N(0, 1)$.

Such a model also requires a formulation of the probability law that causes the economy to go into and out of recession. One very simple specification is that the business cycle phase is the outcome of a $K^\dagger$-state Markov chain that is independent of z_τ for all τ:

$$\begin{aligned} \text{Prob}(s_t^\dagger = j \mid s_{t-1}^\dagger = i, s_{t-2}^\dagger = k, \ldots, z_t, z_{t-1}, z_{t-2}, \ldots) \\ = \text{Prob}(s_t^\dagger = j \mid s_{t-1}^\dagger = i) \\ = p_{ij}^\dagger \end{aligned} \tag{3}$$

This specification assumes that if the economy was in expansion last period $(s_{t-1}^\dagger = 2)$, the probability of going into recession now $(s_t^\dagger = 1)$ is a fixed constant $p_{21}^\dagger$ that does not depend on how long the economy has been in expansion or other measures of the strength of the expansion. This assumption does not appear to be a bad representation of historical experience (Diebold and Rudebusch, 1990), though several researchers have suggested that more complicated specifications of the transition probabilities ought to be considered (see Sichel, 1991; Diebold, Rudebusch, and Sichel, 1993; Filardo, 1994; Durland and McCurdy, 1994; Ghysels, 1993). We prefer here to rely on the simplest specification (3) to keep the model parsimonious.

The above equations model a change in the business cycle phase as a shift in the average growth rate. By contrast, Hamilton and Susmel (1994) modelled a change in the volatility phase of the stock market as a shift in the overall scale of the ARCH process followed by the unforecastable component of stock returns. Let r_t denote the monthly excess return on the S&P 500 stocks over the Treasury bill yield. The specification proposed by Hamilton and Susmel is as follows:

$$r_t = \delta_0 + \delta_1 r_{t-1} + \delta_2 r_{t-2} + \cdots + \delta_q r_{t-q} + e_t \tag{4}$$

$$e_t = \sqrt{g_{s_t^*}} \cdot u_t \tag{5}$$

$$u_t = \sqrt{h_t} \cdot w_t \tag{6}$$

$$\begin{aligned} h_t = {} & \zeta + \xi_1 u_{t-1}^2 + \xi_2 u_{t-2}^2 + \cdots + \xi_m u_{t-m}^2 \\ & + \aleph u_{t-1}^2 I_{t-1} \end{aligned} \tag{7}$$

Here w_t is assumed to be i.i.d. N(0, 1), while s_t^* is an unobserved latent variable that represents the volatility phase of the stock market. In the absence of such phases, the parameter $g_{s_t^*}$ would simply equal unity for all t, in which case equations (4)–(7) would describe stock returns with an autoregression whose residual e_t follows an mth-order ARCH-L process. The 'L' in 'ARCH-L' stands for the 'leverage effect', which is the possibility that stock price increases and decreases could have asymmetric effects on subsequent volatility. Black (1976) and Nelson (1991), among others, have provided strong evidence that this asymmetry is an important feature of stock return data. The variable I_{t-1} in equation (7) is defined to be unity if e_{t-1} is negative and is zero otherwise. Thus if the parameter $\aleph > 0$, a stock price decrease has a greater effect on subsequent volatility than would a stock price increase of the same magnitude. This parameterization of the leverage effect was first proposed by Glosten, Jagannathan, and Runkle (1993).

More generally, for $g_{s_t^*}$ not identically equal to unity, the latent ARCH process u_t is multiplied by a scale factor $\sqrt{g_{s_t^*}}$ representing the current phase s_t^* that characterizes overall stock volatility. We normalize $g_1 = 1$, in which case g_2 has the interpretation as the ratio of the average variance of stock returns when $s_t^* = 2$ compared to that observed when $s_t^* = 1$.

The Markov-switching autoregressive and Markov-switching ARCH features could, in principle, be combined into a single univariate specification, though using such a large set of parameters to describe the non-linear dynamics of a single series might pose numerical problems for finding a global maximum of the likelihood function. Moreover, given the limited predictability of stock returns, it is surely a mistake to overparameterize the mean of r_t. By the same token, evidence of ARCH effects in industrial production is rather weak. By contrast, the tendency of stock market *volatility* (as distinct from mean returns) to exhibit episodic variation, and the periodic shifts in mean output growth associated with economic recessions, are fairly significant and well-documented features of these two series. For this reason, the goal of this paper is to explore the nature of the link between a process for industrial production of the form of equations (1)–(3) and a process for stock returns of the form of equations (4)–(7).

Our key task is to describe the connection between the phase of the business cycle ($s_t^\dagger$) and the phase of stock volatility (s_t^*). One hypothesis is that shifts in stock volatility are completely unrelated to the factors driving the business cycle, in which case s_t^* would be the outcome of a Markov chain with transition probabilities p_{ij}^*, where s_t^* is independent of $s_\tau^\dagger$ for all t and τ. We describe this possibility as Model A. A second hypothesis, Model B, is suggested by the analysis of Schwert (1989a,b) and Campbell, Kim, and Lettau (1993). Under this specification, the factor that causes a shift in stock market volatility and the factor that causes the economy to go into a recession represent one and the same event ($s_t^* = s_t^\dagger$). A third possibility (Model C) is that industrial production and stock returns are ultimately driven by the same fundamentals but are not in phase together. For example, since participants in the stock market are forward-looking, an incipient recession may affect the stock market even before industrial production actually starts to fall, which could be modeled by imposing that $s_t^{\dagger} = s_{t-1}^{*}$.

We also evaluate a more general specification (Model D) that nests Models A, B, and C as special cases. Suppose for illustration that $K^{*} = K^{\dagger} = 2$. Imagine constructing a latent variable $\mathscr{s}_t$ from the separate latent processes $s_t^{\dagger}$ and s_t^{*} as follows:

$$\mathscr{s}_t = 1 \quad \text{if } s_t^{*} = 1 \text{ and } s_t^{\dagger} = 1 \tag{8a}$$

$$\mathscr{s}_t = 2 \quad \text{if } s_t^{*} = 2 \text{ and } s_t^{\dagger} = 1 \tag{8b}$$

$$\mathscr{s}_t = 3 \quad \text{if } s_t^{*} = 1 \text{ and } s_t^{\dagger} = 2 \tag{8c}$$

$$\mathscr{s}_t = 4 \quad \text{if } s_t^{*} = 2 \text{ and } s_t^{\dagger} = 2 \tag{8d}$$

If Model A were the correct specification, then $\mathscr{s}_t$ should follow a 4-state Markov chain whose transition probabilities can be found by multiplying together those for the independent chains governing s_t^{*} and $s_t^{\dagger}$; for example,

$$\begin{aligned}\text{Prob}(\mathscr{s}_t = 3 \mid \mathscr{s}_{t-1} = 2) &= \text{Prob}(s_t^{*} = 1 \mid s_{t-1}^{*} = 2)\,\text{Prob}(s_t^{\dagger} = 2 \mid s_{t-1}^{\dagger} = 1) \\ &= p_{21}^{*} p_{12}^{\dagger}\end{aligned}$$

Let $\mathbf{P}$ be a matrix whose row j, column i entry is $\text{Prob}(\mathscr{s}_t = j \mid \mathscr{s}_{t-1} = i)$. Then under Model A,

$$\mathbf{P} = \begin{bmatrix} p_{11}^{*} p_{11}^{\dagger} & p_{21}^{*} p_{11}^{\dagger} & p_{11}^{*} p_{21}^{\dagger} & p_{21}^{*} p_{21}^{\dagger} \\ p_{12}^{*} p_{11}^{\dagger} & p_{22}^{*} p_{11}^{\dagger} & p_{12}^{*} p_{21}^{\dagger} & p_{22}^{*} p_{21}^{\dagger} \\ p_{11}^{*} p_{12}^{\dagger} & p_{21}^{*} p_{12}^{\dagger} & p_{11}^{*} p_{22}^{\dagger} & p_{21}^{*} p_{22}^{\dagger} \\ p_{12}^{*} p_{12}^{\dagger} & p_{22}^{*} p_{12}^{\dagger} & p_{12}^{*} p_{22}^{\dagger} & p_{22}^{*} p_{22}^{\dagger} \end{bmatrix}$$

On the other hand, if Model C is correct, then $s_t^{\dagger} = s_{t-1}^{*}$, in which case the variable $\mathscr{s}_t$ defined in equation (8) would implicitly be characterized by

$$\begin{aligned} \mathscr{s}_t = 1 &\quad \text{if } s_t^{*} = 1 \text{ and } s_{t-1}^{*} = 1 \\ \mathscr{s}_t = 2 &\quad \text{if } s_t^{*} = 2 \text{ and } s_{t-1}^{*} = 1 \\ \mathscr{s}_t = 3 &\quad \text{if } s_t^{*} = 1 \text{ and } s_{t-1}^{*} = 2 \\ \mathscr{s}_t = 4 &\quad \text{if } s_t^{*} = 2 \text{ and } s_{t-1}^{*} = 2 \end{aligned}$$

In this case, $\mathscr{s}_t$ again follows a Markov chain with transition probabilities

$$\mathbf{P} = \begin{bmatrix} p_{11}^{*} & 0 & p_{11}^{*} & 0 \\ p_{12}^{*} & 0 & p_{12}^{*} & 0 \\ 0 & p_{21}^{*} & 0 & p_{21}^{*} \\ 0 & p_{22}^{*} & 0 & p_{22}^{*} \end{bmatrix}$$

For the general Model D, it is assumed that $\mathscr{s}_t$ follows a 4-state Markov chain with transition matrix $\mathbf{P}$ restricted only by the condition that each column sums to unity. The dependencies of the parameters in equations (1) and (5) on $s_t^{\dagger}$ and s_t^{*} are then rewritten as dependencies on the value of $\mathscr{s}_t$; for example, when $\mathscr{s}_t = 3$, the parameter μ_2 is used in equation (1) while g_1 is used in equation (5).

In Models A, B, C, or D, the variance of the innovation ε_t in equation (2) for the growth rate of industrial production is assumed to be the constant σ^2. If instead the variance ($\sigma_{s_t^{\dagger}}^2$) depends on the state $s_t^{\dagger}$ as in equation (2)′, we call the resulting specifications A′, B′, C′, and D′, respectively. It is also possible to allow for dynamic linkage between the stock market and economic activity through more conventional coefficients as well, adding lagged r_{t-j} to the autoregression explaining y_t, for example, and allowing for contemporaneous correlation between ε_t in equations (1) and (2) and w_t in equations (4)–(7). However, we did not find such additional parameters to be statistically significant and only the simpler specifications are reported below.

Any of the models we estimated can be described as a special case of the following general framework. We suppose that there exists an unobserved latent variable s_t which takes on one of the integer values $(1, 2, \ldots, K\}$ and which summarizes the current phase for both the stock market and the business cycle. Under any of Models A, B, C, or D, this variable s_t itself follows a Markov chain. For example, consider Model C′ ($s_t^{\dagger} = s_{t-1}^{*}$) and suppose that $q = m = 1$. Then the relevant state at date t (s_t) is completely summarized by the value of s_t^{*}, s_{t-1}^{*} and s_{t-2}^{*}. Thus for this case we can define

$$s_t = \begin{cases} 1 \text{ if } s_t^{*} = 1,\ s_{t-1}^{*} = 1, \text{ and } s_{t-2}^{*} = 1 \\ 2 \text{ if } s_t^{*} = 2,\ s_{t-1}^{*} = 1, \text{ and } s_{t-2}^{*} = 1 \\ 3 \text{ if } s_t^{*} = 1,\ s_{t-1}^{*} = 2, \text{ and } s_{t-2}^{*} = 1 \\ 4 \text{ if } s_t^{*} = 2,\ s_{t-1}^{*} = 2, \text{ and } s_{t-2}^{*} = 1 \\ 5 \text{ if } s_t^{*} = 1,\ s_{t-1}^{*} = 1, \text{ and } s_{t-2}^{*} = 2 \\ 6 \text{ if } s_t^{*} = 2,\ s_{t-1}^{*} = 1, \text{ and } s_{t-2}^{*} = 2 \\ 7 \text{ if } s_t^{*} = 1,\ s_{t-1}^{*} = 2, \text{ and } s_{t-2}^{*} = 2 \\ 8 \text{ if } s_t^{*} = 2,\ s_{t-1}^{*} = 2, \text{ and } s_{t-2}^{*} = 2 \end{cases}$$

Notice that s_t follows an 8-state Markov chain whose transition probability $\text{Prob}(s_t = j \mid s_{t-1} = i)$ is given by the row j, column i element of the following matrix:

$$\mathbf{P} = \begin{bmatrix} p_{11}^{*} & 0 & 0 & 0 & p_{11}^{*} & 0 & 0 & 0 \\ p_{12}^{*} & 0 & 0 & 0 & p_{12}^{*} & 0 & 0 & 0 \\ 0 & p_{21}^{*} & 0 & 0 & 0 & p_{21}^{*} & 0 & 0 \\ 0 & p_{22}^{*} & 0 & 0 & 0 & p_{22}^{*} & 0 & 0 \\ 0 & 0 & p_{11}^{*} & 0 & 0 & 0 & p_{11}^{*} & 0 \\ 0 & 0 & p_{12}^{*} & 0 & 0 & 0 & p_{12}^{*} & 0 \\ 0 & 0 & 0 & p_{21}^{*} & 0 & 0 & 0 & p_{21}^{*} \\ 0 & 0 & 0 & p_{22}^{*} & 0 & 0 & 0 & p_{22}^{*} \end{bmatrix} \tag{9}$$

Let $\mathbf{x}_t = (y_t, r_t)'$ be a (2×1) vector containing the growth rate of industrial production and the excess return on stocks, and consider the vector process

$$\mathbf{x}_t = \boldsymbol{\theta}_{s_t} + \boldsymbol{\phi}_1 \mathbf{x}_{t-1} + \boldsymbol{\phi}_2 \mathbf{x}_{t-2} + \cdots + \boldsymbol{\phi}_q \mathbf{x}_{t-q} + \mathbf{L}_{t,s_t} \mathbf{v}_t \tag{10}$$

For the special case represented by equations (2) and (4), $\boldsymbol{\phi}_j$ would be a diagonal (2×2) matrix given by

$$\boldsymbol{\phi}_j = \begin{bmatrix} \alpha_j & 0 \\ 0 & \delta_j \end{bmatrix}$$

for $j = 1, 2, \ldots, q$. To obtain Model C′ with $q = m = 1$ as discussed above, the (2×1) vector $\boldsymbol{\theta}_{s_t}$ would have the interpretation given in Table I.

The error term $\mathbf{v}_t$ in equation (10) is presumed to be i.i.d. $N(0, \mathbf{I}_2)$ for $\mathbf{I}_2$ the (2×2) identity

Table I. Meaning of parameters for Model C′

s_t	θ_{s_t}	L_{t,s_t}	u^2_{t-1,s_t}
1	$\begin{bmatrix}(1-\alpha_1)\mu_1\\ \delta_0\end{bmatrix}$	$\begin{bmatrix}\sigma_1 & 0\\ 0 & \sqrt{g_1 h_t}\end{bmatrix}$	e^2_{t-1}/g_1
2	$\begin{bmatrix}(1-\alpha_1)\mu_1\\ \delta_0\end{bmatrix}$	$\begin{bmatrix}\sigma_1 & 0\\ 0 & \sqrt{g_2 h_t}\end{bmatrix}$	e^2_{t-1}/g_1
3	$\begin{bmatrix}\mu_2-\alpha_1\mu_1\\ \delta_0\end{bmatrix}$	$\begin{bmatrix}\sigma_2 & 0\\ 0 & \sqrt{g_1 h_t}\end{bmatrix}$	e^2_{t-1}/g_2
4	$\begin{bmatrix}\mu_2-\alpha_1\mu_1\\ \delta_0\end{bmatrix}$	$\begin{bmatrix}\sigma_2 & 0\\ 0 & \sqrt{g_2 h_t}\end{bmatrix}$	e^2_{t-1}/g_2
5	$\begin{bmatrix}\mu_1-\alpha_1\mu_2\\ \delta_0\end{bmatrix}$	$\begin{bmatrix}\sigma_1 & 0\\ 0 & \sqrt{g_1 h_t}\end{bmatrix}$	e^2_{t-1}/g_1
6	$\begin{bmatrix}\mu_1-\alpha_1\mu_2\\ \delta_0\end{bmatrix}$	$\begin{bmatrix}\sigma_1 & 0\\ 0 & \sqrt{g_2 h_t}\end{bmatrix}$	e^2_{t-1}/g_1
7	$\begin{bmatrix}(1-\alpha_1)\mu_2\\ \delta_0\end{bmatrix}$	$\begin{bmatrix}\sigma_2 & 0\\ 0 & \sqrt{g_1 h_t}\end{bmatrix}$	e^2_{t-1}/g_2
8	$\begin{bmatrix}(1-\alpha_1)\mu_2\\ \delta_0\end{bmatrix}$	$\begin{bmatrix}\sigma_2 & 0\\ 0 & \sqrt{g_2 h_t}\end{bmatrix}$	e^2_{t-1}/g_2

matrix. To implement the example in equations (1), (2)′, and (3)–(7) with ε_t independent of u_t, the (2×2) matrix $\mathbf{L}_{t,s_t}$ would be diagonal with the diagonal elements as indicated in Table I, where h_t is given by

$$h_t = \zeta + \xi_1 u^2_{t-1,s_t} + \aleph\, u^2_{t-1,s_t} I_{t-1}$$

for u^2_{t-1,s_t} as specified in Table I and

$$e_{t-1} = r_{t-1} - \delta_0 - \delta_1 r_{t-2}$$

$$I_{t-1} = \begin{cases} 1 & \text{if } e_{t-1} < 0 \\ 0 & \text{if } e_{t-1} \geqslant 0 \end{cases}$$

Thus any of the specifications for the interaction between the stock market and macroeconomic activity discussed above can be represented in terms of a particular parameterization for the conditional density

$$f(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots, \mathbf{x}_{t-q}, s_t) = (2\pi)^{-1} \left| \mathbf{L}_{t,s_t} \right|^{-1} \exp[-\tfrac{1}{2}\boldsymbol{\xi}'_{t,s_t}\boldsymbol{\xi}_{t,s_t}]$$

where

$$\boldsymbol{\xi}_{t,s_t} = \mathbf{L}^{-1}_{t,s_t}[\mathbf{x}_t - \boldsymbol{\theta}_{s_t} - \boldsymbol{\phi}_1\mathbf{x}_{t-1} - \cdots - \boldsymbol{\phi}_q\mathbf{x}_{t-q}]$$

Given this conditional density and the parameters of the Markov transition matrix **P**, it is possible to evaluate the log likelihood of the observed data,

$$\mathcal{Q} = \sum_{t=1}^{T} \log f(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots, \mathbf{x}_{-q}; \boldsymbol{\gamma}) \tag{11}$$

using the methods described in Hamilton (1994, Chapter 22), where $\boldsymbol{\gamma}$ is a vector of population parameters containing the unknown elements of **P**, $\boldsymbol{\phi}_j$, $\boldsymbol{\theta}_i$, and $\mathbf{L}_{t,i}$. This log likelihood can then be maximized with respect to $\boldsymbol{\gamma}$ numerically, and the resulting maximum likelihood estimates $\hat{\boldsymbol{\gamma}}$ can be used to form an inference about the latent state of the form $\text{Prob}(s_t = 1 | \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1; \hat{\boldsymbol{\gamma}})$. For the special case of Model C described above, this in turn implies an inference about whether the economy was in a recession at date t, that is given by

$$\begin{aligned}\text{Prob}(s_t^* = 1 | \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1; \hat{\boldsymbol{\gamma}}) = {} & \text{Prob}(s_t = 1 | \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1; \hat{\boldsymbol{\gamma}}) \\ & + \text{Prob}(s_t = 3 | \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1; \hat{\boldsymbol{\gamma}}) \\ & + \text{Prob}(s_t = 5 | \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1; \hat{\boldsymbol{\gamma}}) \\ & + \text{Prob}(s_t = 7 | \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1; \hat{\boldsymbol{\gamma}}) \end{aligned} \tag{12}$$

3. EMPIRICAL RESULTS

The output measure y_t used in this analysis was constructed from 100 times the monthly change in the natural logarithm of the Federal Reserve Board's index of industrial production for 1965:1 to 1993:6. All the maximum likelihood estimates reported below condition on the first three values, so that the first observation used for maximum likelihood estimation ($t = 1$) corresponds to the change in industrial production between March and April of 1965. The last observation ($T = 339$) is the change between May and June of 1993. The stock return r_t is 100 times the change in the natural logarithm of the S&P 500 stock index plus the dividend yield on the S&P 500 minus the yield on 3-month Treasury bills, with the latter both quoted at monthly rates. All data were taken from Citibase.

Maximum likelihood estimates for the parameters of Models A, B, C, C′, D, and D′ are reported in Table II for the case $q = m = 1$ and $K^{\dagger} = K^* = 2$. In order to guarantee that the variance h_t in equation (7) is positive for all possible realizations of (u_t), we imposed the constraint $\xi_1 \geqslant 0$. In every case (regardless of whether we allowed for the possibility of regime switching), the maximum likelihood estimate converged to the boundary $\xi_1 = 0$. We therefore dropped the parameter ξ_1 from the model, concluding that for this data set, the only ARCH effects are those caused by downward movements in stock prices, as captured by the parameter $\aleph$.[5]

3.1. Univariate Results (Model A)

Consider first the results for Model A, according to which the process followed by stock returns is completely unrelated to industrial production. The parameter estimates for this case were obtained by fitting equations (1)–(3) to the univariate industrial production data, fitting equations (4)–(7) to the univariate stock return data, and simply adding the maximum values obtained for the univariate log likelihoods together. There is substantial evidence of

[5] We tried experimenting with allowing the parameter ξ_1 to be negative, but found the likelihood function to be ill-behaved and difficult to maximize numerically.

non-linearity in the dynamics of both output and stock returns of the kind predicted by the regime-switching model. The output process (equations (1) and (2)) includes a fixed-coefficient Gaussian AR(1) specification as a special case when $\mu_1 = \mu_2$. The value achieved for the univariate log likelihood for output growth by a fixed-coefficient AR(1) specification is $-391{\cdot}13$, compared with a univariate log likelihood of $-377{\cdot}25$ for the process represented by the first six parameters in the Model A column of Table II. The latter specification has three more parameters than the former. The usual likelihood ratio statistic is then $2(-377{\cdot}25 + 391{\cdot}13) = 27{\cdot}76$. If this statistic had the usual $\chi^2(3)$ distribution, the probability of achieving these results if there had in fact been no changes in regime would be 4×10^{-6}.

Table II. Maximum likelihood estimates for two-state models (standard errors in parentheses)

	Model A	Model B	Model C	Model C′	Model D	Model D′
μ_1	−1·98 (0·37)	−0·780 (0·154)	−0·762 (0·137)	−0·662 (0·219)	−0·743 (0·138)	−0·632 (0·214)
μ_2	0·303 (0·054)	0·409 (0·056)	0·417 (0·055)	0·407 (0·049)	0·344 (0·080)	0·345 (0·067)
α_1	0·265 (0·056)	0·220 (0·058)	0·238 (0·055)	0·194 (0·056)	0·208 (0·060)	0·172 (0·061)
σ_1^2	0·478 (0·040)	0·484 (0·041)	0·484 (0·038)	1·033 (0·231)	0·486 (0·039)	1·056 (0·241)
σ_2^2	—	—	—	0·396 (0·036)	—	0·397 (0·036)
$p_{11}^{\dagger}$	0·659 (0·159)	—	—	—	see text	see text
$p_{22}^{\dagger}$	0·988 (0·007)	—	—	—	see text	see text
δ_0	0·322 (0·174)	0·306 (0·169)	0·356 (0·169)	0·359 (0·169)	0·356 (0·164)	0·350 (0·164)
δ_1	0·270 (0·050)	0·271 (0·053)	0·271 (0·050)	0·264 (0·053)	0·265 (0·049)	0·258 (0·050)
g_2	0·255 (0·079)	0·106 (0·051)	0·0881 (0·0412)	0·083 (0·051)	0·108 (0·052)	0·122 (0·055)
ζ	11·6 (1·9)	19·3 (2·2)	19·7 (2·1)	19·5 (2·1)	22·5 (3·6)	22·7 (3·6)
$\aleph$	0·253 (0·146)	0·185 (0·105)	0·159 (0·107)	0·151 (0·114)	0·183 (0·134)	0·182 (0·149)
p_{11}^{*}	0·457 (0·234)	0·854 (0·062)	0·887 (0·049)	0·885 (0·051)	see text	see text
p_{22}^{*}	0·624 (0·218)	0·975 (0·011)	0·979 (0·010)	0·977 (0·011)	see text	see text
log likelihood	−1262·26	−1263·23	−1259·95	−1250·42	−1258·70	−1248·82
Number of parameters	13	11	11	12	13	15

Unfortunately, the usual asymptotic distribution theory does not hold for this case, because the auxiliary parameters $p_{11}^{\dagger}$ and $p_{22}^{\dagger}$ are unidentified under the null hypothesis that $\mu_1 = \mu_2$. Tests which get around this problem are described in the bivariate analysis below. For now, we note that the tiny p-value for the conventional test is fairly convincing evidence of non-linearity, for the following reason. One could instead take as the maintained null hypothesis the existence of two distinct states and form a marginal confidence interval for the parameter μ_1 using the usual Wald t test. Under this maintained hypothesis, all the parameters are identified and the t test has its usual asymptotic N(0, 1) distribution. This confidence interval for μ_1 is far below the average output growth rate during expansions, suggesting that the difference between the regimes is indeed highly statistically significant. This Wald test and the likelihood ratio test are essentially summarizing the same feature of the likelihood surface.

Another indication that an AR(1) specification with constant parameters does not adequately capture the dynamics of the output data is provided by a comparison of the forecasting performance of the linear and non-linear specifications reported in Table III. The in-sample mean squared error (MSE) for the output forecast derived from the Markov-switching model is 0·579, compared with 0·588 for the standard AR(1). A tiny forecasting improvement holds out of sample as well. All out-of-sample forecast evaluations reported in this paper were done as follows. First the parameters of the model were estimated using data from 1965:1 to 1974:10. Then a one-period-ahead forecast was generated for 1974:11. The model was next re-estimated by maximum likelihood using data from 1965:1 to 1974:11, and a one-period-ahead forecast for 1974:12 was generated. This process was continued until the end of the sample (1993:6) was reached. The average squared deviations of these forecasts from the true values over 1974:11 to 1993:6 was 0·626 for the Markov-switching model, slightly better than the MSE of 0·627 achieved by the AR(1) specification.

Stock returns also display their own evidence of non-linearity. Here a standard ARCH-L specification is obtained as a special case of the Markov-switching model in equations (4)–(7) by setting $g_2 = 1$. The estimated value of $\hat{g}_2 = 0{\cdot}255$ means that the variance of stock returns is four times as large when $s_t^* = 1$ compared to the variance when $s_t^* = 2$. Again a 95% confidence interval for $\hat{g}_2$ is very far from unity, and a likelihood ratio test of the switching model

Table III. Mean squared errors for forecasting monthly growth rate of industrial production

Specification	In-sample	Post-sample
Univariate AR(1)	0·588	0·627
Univariate $K = 2$ Markov-switching	0·579	0·626
Univariate $K = 3$ Markov-switching	0·579	0·623
Bivariate VAR(3)	0·550	0·604
Bivariate $K = 2$ (Model B)	0·562	0·611
Bivariate $K = 2$ (Model C)	0·551	0·604
Bivariate $K = 2$ (Model C′)	0·557	0·621

represented by the last seven parameters in column A of Table II against the null of an ARCH-L specification without switching produces a test statistic of 20·18, which would mean a p-value of 0·00016 if the $\chi^2(3)$ distribution were appropriate. Forecasting comparisons reported in Table IV also suggest the practical importance of non-linearities. Equations (4)–(7) imply that

$$E(e_t^2 \mid r_{t-1}, r_{t-2}, \ldots) = h_t[g_1 \operatorname{Prob}(s_t^* = 1 \mid r_{t-1}, r_{t-2}, \ldots) + \cdots + g_{K^*} \operatorname{Prob}(s_t^* = K^* \mid r_{t-1}, r_{t-2}, \ldots)] \tag{13}$$

The difference between the expression on the right side of equation (13) and $\hat{e}_t^2$, where $\hat{e}_t^2 = (r_t - \hat{\delta}_0 - \hat{\delta}_1 r_{t-1})^2$, is then a measure of the error the model made in forecasting stock return volatility for date t. Rather than square this error (which would then involve the fourth moments of the actual return series), we report in Table IV the average absolute value of the difference between $\hat{e}_t^2$ and the forecast described in equation (13). The forecast in equation (13) is compared both with that for a simple AR(1) process for stock returns (for which the forecast is a constant given by the average residual variance) and with the forecast for ARCH-L and GARCH-L specifications without regime-switching.

We also estimated univariate specifications with three states ($K^\dagger = K^* = 3$), and achieved a value for the joint log likelihood of $-1257{\cdot}15$. In each case three of the estimated transition probabilities were zero; for example, $s_{t-1}^* = 1$ seems never to have been followed by $s_t^* = 3$. With these zeros imposed, the three-state specification required 17 parameters. The likelihood ratio test is then $-2(-1257{\cdot}15 + 1262{\cdot}26) = 10{\cdot}22$, whose p-value would be 0·04 if the $\chi^2(4)$ distribution were valid. We take this as unconvincing statistical evidence that the univariate data are really characterized by three separate regimes.

3.2. Bivariate Evidence that Stock Volatility is Driven by Economic Activity

Model A implies that industrial production and stock returns are completely unrelated processes, an assumption that seems unlikely *a priori*. It is interesting to compare Model A with a standard linear representation of the possible dependence between the series in the form of a fixed-

Table IV. Mean absolute errors for forecasting squared residual of monthly stock return

Specification	In-sample	Post-sample
Univariate AR(1)	13·11	13·72
Univariate ARCH-L	13·03	12·97
Univariate GARCH-L	12·86	12·60
Univariate $K = 2$ Markov-switching	13·05	11·29
Univariate $K = 3$ Markov-switching	13·22	14·22
Bivariate $K = 2$ (Model B)	12·68	11·48
Bivariate $K = 2$ (Model C)	12·87	10·98
Bivariate $K = 2$ (Model C′)	13·02	11·14

coefficient qth-order vector autoregression, or VAR(q):

$$\mathbf{x}_t = \boldsymbol{\theta} + \boldsymbol{\phi}_1 \mathbf{x}_{t-1} + \cdots + \boldsymbol{\phi}_q \mathbf{x}_{t-q} + \mathbf{L}\mathbf{v}_t$$

where $\mathbf{v}_t \sim$ i.i.d. $N(\mathbf{0}, \mathbf{I}_2)$, $\mathbf{x}_t = (y_t, r_t)'$, $\boldsymbol{\theta}$ is a (2×1) vector of constants, and $\boldsymbol{\phi}_j$ and $\mathbf{L}$ are (2×2) matrices of parameters. Unlike Model A, a VAR allows for interdependence between output and stock returns through off-diagonal elements of $\mathbf{L}$ and $\boldsymbol{\phi}_j$. A likelihood ratio test rejects the null hypotheses that the number of lags is equal to 1 or 2 in favour of the alternative that $q = 3$. The null hypothesis that $q = 3$ is accepted against the alternative that $q = 4$. Hence we take the third-order VAR to be the standard of comparison. This VAR(3) achieves a value for the log likelihood of $-1275{\cdot}40$. Note that the VAR(3) requires 17 parameters (compared with only 13 for the two-state version of Model A) and further allows for dependence between r_t and y_τ (whereas Model A allows none). Nevertheless, the value achieved for the log likelihood by the VAR(3) is not nearly as good as for Model A. We conclude that the evidence for univariate non-linearities for each series (as predicted by the independent regime-switching formulation) is stronger than the evidence of linear interdependence between the two series.

Table II also reports parameter estimates for Model B, which posits that the state governing the output series ($s_t^\dagger$) is the same as the state governing stock volatility (s_t^*). Note that Models A and B are non-nested. Model B has fewer parameters than Model A, since B imposes the restriction that $p_{ij}^\dagger = p_{ij}^*$, and for this reason might be expected to worsen the fit. On the other hand, Model B allows for dependence between r_t and y_t through their common dependence on the unobserved state s_t^*. This relaxes the assumption of independence maintained by Model A, and therefore could end up improving the fit, though the only dependence between the series that is allowed by Model B is the dependence captured by the common non-linearities associated with regime-switching. Even with two fewer parameters, Model B is able to achieve almost as good a value for the likelihood as Model A, suggesting that the non-linear dependence is potentially quite important. Model C, in which the shift in regime affects stock returns one month before it affects industrial production, has an even better fit, and achieves a much better value for the likelihood compared to Model A despite having two fewer parameters. We also estimated a model in which stock returns are affected two months before industrial production, with poorer results—the log likelihood for this specification was only $-1264{\cdot}55$.

Table II also reports maximum likelihood estimates for Model D, which nests Models A and C as special cases. This specification allows s_t^* and $s_t^\dagger$ to be related but does not impose any relation *a priori*. The matrix of transition probabilities for the latent variable s_t was estimated for Model D to be (standard errors in parentheses):

$$\mathbf{P} = \begin{bmatrix} \begin{matrix}0{\cdot}876\\(0{\cdot}056)\end{matrix} & 0{\cdot}000 & 1{\cdot}000 & 0{\cdot}000 \\ \begin{matrix}0{\cdot}124\\(0{\cdot}056)\end{matrix} & \begin{matrix}0{\cdot}044\\(0{\cdot}248)\end{matrix} & 0{\cdot}000 & \begin{matrix}0{\cdot}282\\(0{\cdot}322)\end{matrix} \\ 0{\cdot}000 & 0{\cdot}000 & 0{\cdot}000 & \begin{matrix}0{\cdot}026\\(0{\cdot}109)\end{matrix} \\ 0{\cdot}000 & \begin{matrix}0{\cdot}956\\(0{\cdot}248)\end{matrix} & 0{\cdot}000 & \begin{matrix}0{\cdot}692\\(0{\cdot}329)\end{matrix} \end{bmatrix}$$

Although no constraints were imposed on $\mathbf{P}$, the maximum likelihood estimates of several of the transition probabilities turn out to be zero. In such cases a value of zero was then imposed *a priori* for purposes of calculating standard errors with respect to the remaining parameters.[6]

For comparison, if Model C were correct, the parameter values for the above matrix would be

$$\mathbf{P} = \begin{bmatrix} \begin{matrix}0\cdot887\\(0\cdot049)\end{matrix} & 0\cdot000 & \begin{matrix}0\cdot887\\(0\cdot049)\end{matrix} & 0\cdot000 \\ \begin{matrix}0\cdot113\\(0\cdot049)\end{matrix} & 0\cdot000 & \begin{matrix}0\cdot113\\(0\cdot049)\end{matrix} & 0\cdot000 \\ 0\cdot000 & \begin{matrix}0\cdot021\\(0\cdot010)\end{matrix} & 0\cdot000 & \begin{matrix}0\cdot021\\(0\cdot010)\end{matrix} \\ 0\cdot000 & \begin{matrix}0\cdot979\\(0\cdot010)\end{matrix} & 0\cdot000 & \begin{matrix}0\cdot979\\(0\cdot010)\end{matrix} \end{bmatrix}$$

The unconstrained estimates (Model D) strongly suggest the pattern imposed in the constrained Model C. In particular, in the unconstrained specification D, the single most common state is $s_t = 4$, which has an ergodic probability of 0·64. In other words, in two-thirds of the months we observe the economy expanding at the same time that the stock market is calm, even though Model D imposed no relation between theses two events. Similarly, in Model D the state $s_t = 1$ occurs 13% of the time, and is characterized by the economy being in recession at the same time that the stock market is volatile. Thus even a completely unconstrained specification leads to the conclusion that stock market volatility and the level of economic activity are typically driven by the same events. Furthermore, compared to Model C, the more general Model D achieves a fairly trivial improvement in the likelihood function. We conclude that Model C provides a reasonable description of the comovement of stock returns and economic activity.

3.3. Testing the Null Hypothesis of No Regime-switching

We also consider a formal statistical test of the importance of the non-linearities documented here. Model C includes a simple linear-ARCH process with no regime-switching as a special case when $\mu_1 = \mu_2$ and $g_1 = g_2$. As noted earlier, a likelihood ratio test of this null hypothesis does not have the usual limiting χ^2 distribution because the parameters p_{ij}^* are unidentified under the null. Hansen (1992, 1996) has suggested a test that gets around this problem. Let $\boldsymbol{\alpha} \equiv (\mu_1 - \mu_2, g_1 - g_2, p_{11}^*, p_{22}^*)'$ denote the regime-switching parameters of Model C and let $\boldsymbol{\lambda}$ denote the remaining parameters:

$$\boldsymbol{\lambda} \equiv (\mu_1, \alpha_1, \sigma^2, \delta_0, \delta_1, \zeta, \aleph)'$$

Let $l_t(\boldsymbol{\alpha}, \boldsymbol{\lambda})$ denote the conditional log likelihood of the tth observation when evaluated at the parameter values represented by $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$:

$$l_t(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = \log f(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots, \mathbf{x}_{-q}; \boldsymbol{\alpha}, \boldsymbol{\lambda})$$

Under the null hypothesis of no regime-switching, the first two elements of $\boldsymbol{\alpha}$ are zero and the remaining two elements of $\boldsymbol{\alpha}$ have no effect on the likelihood function. Thus we can represent the null hypothesis as $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ where

$$\boldsymbol{\alpha}_0 = (0, 0, 1, 0)'$$

[6] This is necessary because when a maximum likelihood estimate falls on the boundary of the allowable parameter space (as when $\hat{p}_{13} = 0$), the regularity conditions that motivate the usual formula for asymptotic standard errors fail to hold. However, when $p_{13} = 0$ is simply imposed *a priori*, there is no problem with respect to the regularity conditions for the remaining free parameters.

We investigated a grid containing 6000 possible values for $\boldsymbol{\alpha}$ to represent the alternative hypothesis, with $\mathscr{A}$ denoting the set consisting of these 6000 possibilities considered.[7] For any $\boldsymbol{\alpha}$, let $\hat{\boldsymbol{\lambda}}(\boldsymbol{\alpha})$ denote the value of $\boldsymbol{\lambda}$ that maximizes the likelihood with respect to $\boldsymbol{\lambda}$ taking $\boldsymbol{\alpha}$ as given. Define

$$q_t(\boldsymbol{\alpha}) \equiv l_t[\boldsymbol{\alpha}, \hat{\boldsymbol{\lambda}}(\boldsymbol{\alpha})] - l_t[\boldsymbol{\alpha}_0, \hat{\boldsymbol{\lambda}}(\boldsymbol{\alpha}_0)]$$

and let $\bar{q}(\boldsymbol{\alpha})$ denote the sample mean of this variable:

$$\bar{q}(\boldsymbol{\alpha}) = T^{-1} \sum_{t=1}^{T} q_t(\boldsymbol{\alpha})$$

The likelihood ratio test of the null hypothesis $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ against the specific alternative represented by $\boldsymbol{\alpha}$ could then be represented as $T\bar{q}(\boldsymbol{\alpha})$.

Hansen (1996) suggested calculating the following 'standardized' likelihood ratio statistic:

$$\hat{LR} = \max_{\boldsymbol{\alpha} \in \mathscr{A}} T\bar{q}(\boldsymbol{\alpha}) \div \left\{ \sum_{t=1}^{T} [q_t(\boldsymbol{\alpha}) - \bar{q}(\boldsymbol{\alpha})]^2 \right\}^{1/2} \tag{14}$$

He showed that, if the null hypothesis that $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ is true, then for large samples the probability that $\hat{LR}$ would exceed a critical value z is less than the probability that the following statistic would exceed the same value z:

$$\max_{\boldsymbol{\alpha} \in \mathscr{A}} (1 + M)^{-1/2} \sum_{k=0}^{M} \sum_{t=1}^{T} [q_t(\boldsymbol{\alpha}) - \bar{q}(\boldsymbol{\alpha})] u_{t+k} \div \left\{ \sum_{t=1}^{T} [q_t(\boldsymbol{\alpha}) - \bar{q}(\boldsymbol{\alpha})]^2 \right\}^{1/2}$$

where u_t is an i.i.d. $N(0, 1)$ sequence generated by Monte Carlo and M is a parameter corresponding to the maximum order of autocorrelation allowed in $q_t(\boldsymbol{\alpha})$; Hansen (1992) assumed $M = 0$. We generated Hansen's statistic for Model C with M equal to 0–4 and found that the probability of obtaining as large a value for equation (14) as is observed in the data is less than 3% for all values of M. Recalling that this test only gives an upper bound on the true p-value, we conclude that the null hypothesis of no regime-switching should be rejected.

3.4. Tests for Structural Stability

As another check on the validity of Model C as a description of these data, we investigated the test for parameter stability proposed by Andrews (1993, equation (4.4)):[8]

$$LM(\pi) = \frac{T}{\pi(1 - \pi)} [\mathbf{m}_1(\pi)]' \hat{\mathbf{H}} [\mathbf{m}_1(\pi)] \tag{15}$$

[7] The set $\mathscr{A}$ contains all values of $\boldsymbol{\alpha}$ such that p_{11}^* or p_{22}^* were elements of $\{0{\cdot}2, 0{\cdot}39, 0{\cdot}58, 0{\cdot}77, 0{\cdot}96\}$, $(\mu_1 - \mu_2) \in \{0{\cdot}1, 0{\cdot}3, \ldots, 3{\cdot}1\}$, and $(g_1 - g_2) \in \{0{\cdot}1, 0{\cdot}3, 0{\cdot}5, 0{\cdot}7, 0{\cdot}9\} \cup \{1{\cdot}0, 2{\cdot}5, 5{\cdot}0, 6{\cdot}5, \ldots, 14{\cdot}5\}$.

[8] Andrews actually proposed our equation (15) with $\hat{\mathbf{H}}$ replaced by

$$\hat{\mathbf{H}}^* = \hat{\mathbf{S}}^{-1} \hat{\mathbf{M}} (\hat{\mathbf{M}}' \hat{\mathbf{S}}^{-1} \hat{\mathbf{M}})^{-1} \hat{\mathbf{M}}' \hat{\mathbf{S}}^{-1}$$

where

$$\hat{\mathbf{S}} = T^{-1} \sum_{t=1}^{T} [\mathbf{q}_t(\hat{\gamma})][\mathbf{q}_t(\hat{\gamma})]'$$

$$\hat{\mathbf{M}} = T^{-1} \sum_{t=1}^{T} \left. \frac{\partial \mathbf{q}_t(\gamma)}{\partial \gamma'} \right|_{\gamma = \hat{\gamma}}$$

If the model is correctly specified, then given the information matrix equality (e.g., Hamilton, 1994, p. 429), the above matrix $\hat{\mathbf{H}}^*$ has the same plim as the matrix $\hat{\mathbf{H}}$ given in the text.

where

$$\mathbf{m}_1(\pi) = T^{-1}\sum_{t=1}^{T\pi}\mathbf{q}_t(\hat{\gamma})$$
$$\mathbf{q}_t(\hat{\gamma}) = \left.\frac{\partial \log f(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots, \mathbf{x}_{-q}; \gamma)}{\partial \gamma}\right|_{\gamma=\hat{\gamma}} \tag{16}$$
$$\hat{\mathbf{H}} = \left[-T^{-1}\sum_{t=1}^{T}\left.\frac{\partial \mathbf{q}_t(\gamma)}{\partial \gamma'}\right|_{\gamma=\hat{\gamma}}\right]^{-1}$$

The statistic $LM(\pi)$ was calculated for all $\pi \in [0{\cdot}15, 0{\cdot}85]$ such that $T\pi$ is an integer. The largest value for $LM(\pi)$ over this interval for Model C was 29·82, which is just above the 5% critical value of 28·55 reported in Table 1 of Andrews. We thus find some suggestion of instability in the parameters of Model C. Tested individually, the only element of γ for which there is evidence of instability over time is the parameter μ_2, which gives the growth rate of industrial production during expansion. Andrews's test may thus be detecting some indication of a slowdown in the growth rate since 1965 that is not captured by this model.

3.5. Regime-switching in Both the Mean and Variance of Industrial Production (Models C′ and D′)

Table II also reports results for specifications C′ and D′ in which both the mean ($\mu_{s_t^\dagger}$) and the variance ($\sigma^2_{s_t^\dagger}$) of industrial production growth change with the state $s_t^\dagger$. The added parameter σ_2^2 is highly statistically significant, and Andrews's test statistic dropped to 29·69, which is no longer statistically significant at the 5% level. However, this last finding appears primarily due to the decrease in power that results from estimating an additional parameter, as an individual test of stability of the parameter μ_2 is still highly significant. Furthermore, Model C′ does somewhat worse than Model C in terms of forecasting (see Tables III and IV). For this reason, our discussion of results below focuses primarily on Model C rather than C′.

3.6. Further Discussion and Interpretation of Results for Model C

According to the maximum likelihood estimates for Model C reported in Table II, industrial production tends to fall by $\frac{3}{4}$ of a per cent per month as long as the economy remains in regime 1 ($\hat{\mu}_1 = -0{\cdot}762$). The unforecastable component of stock returns (the residual e_t in equation (4)) has a variance that is over ten times as large in regime 1 as it is in regime 2 ($\hat{g}_2 = 0{\cdot}0881$). Regime 1 tends to last on average for $1/(1-\hat{p}^*_{11}) = 9$ months, whereas episodes of regime 2 typically persist for $1/(1-\hat{p}^*_{22}) = 4$ years.

The raw data for output and stock returns used in this analysis are presented in the top two panels of Figure 1. The bottom panel plots the smoothed probability $\text{Prob}(s_t^* = 1 \mid \mathbf{x}_T, \mathbf{x}_{T-1}, \ldots, \mathbf{x}_1)$ for $\mathbf{x}_t = (y_t, r_t)'$, which probability represents the inference based on Model C about how likely it is that the economy was in the high stock volatility regime at any historical date t; this is, of course, equivalent in this framework to the probability that industrial production would be in the falling output regime at date $t+1$. This inference is based on the full sample of data obtained through 1993:6. The figure also displays vertical lines at the NBER business cycle peak and trough dates (taken from the *Survey of Current Business*). The correspondence between the econometric inference and the NBER dating of economic recessions is remarkable.

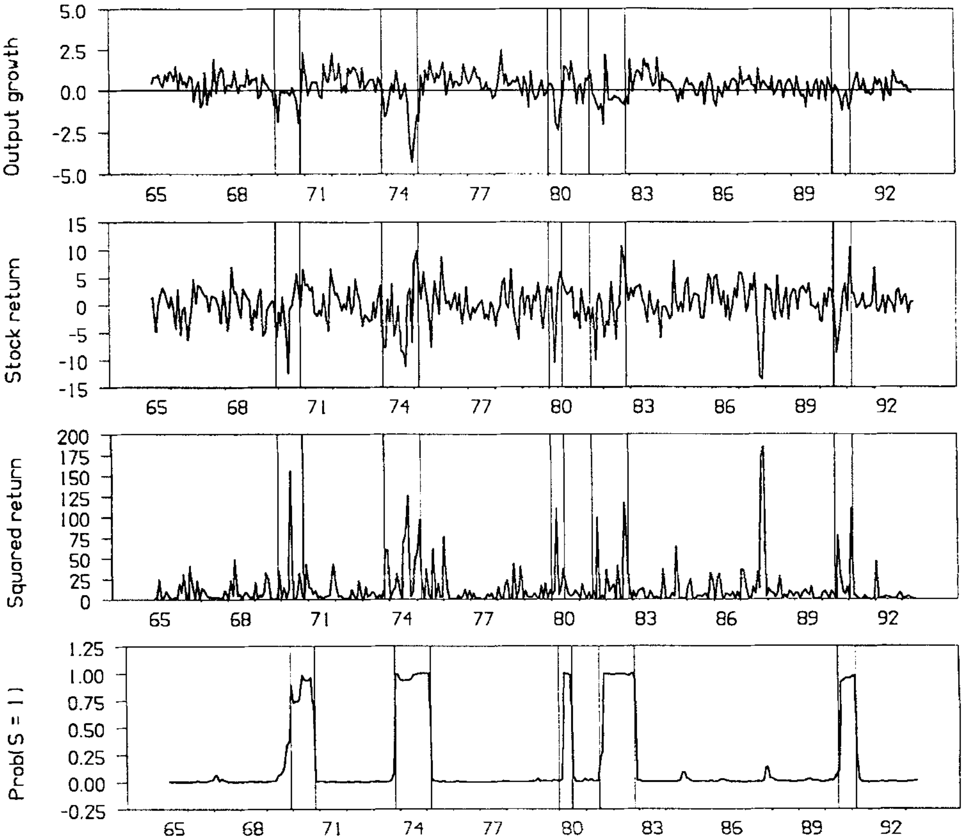

Figure 1. (a) rate of growth of industrial production, quoted at a monthly rate. (b) excess return on S&P 500 stock price index relative to 3-month Treasury bills, quoted at a monthly rate. (c) square of series in second panel. (d) smoothed probability calculated from Model C that economy is in the high-volatility, low-output growth state at date t (based on information available through the end of the sample). Vertical lines denote business cycle peak and trough dates as determined by NBER. All data are monthly 1965:4 to 1993:6

The parameter estimates suggest that economic recessions are a very important cause of stock volatility. From equations (5) and (6) the unconditional MSE in forecasting stock returns can be written as

$$\begin{aligned} E(e_t^2) &= E(h_t g_{s_t^*} w_t^2) \\ &= E(h_t)E(g_{s_t^*}) \\ &= E(h_t)[g_1 \operatorname{Prob}(s_t^* = 1) + g_2 \operatorname{Prob}(s_t^* = 2)] \end{aligned} \tag{17}$$

where the second line follows from the independence of s_t^* and w_τ while the expression '$\operatorname{Prob}(s_t^* = 1)$' denotes the unconditional or ergodic probability that the economy will be in state 1 at any given date. The maximum likelihood estimates for Model C show this latter probability to be

$$\operatorname{Prob}(s_t^* = 1) = \frac{1 - p_{22}^*}{2 - p_{11}^* - p_{22}^*} = 0{\cdot}16$$

If there were no recessions ($s_t^* = 2$ for all t), then the unconditional MSE for forecasting stock returns would simply be $E(h_t)g_2$. Thus, in the absence of recessions, stocks would exhibit much less than half their average observed postwar variability:

$$\frac{E(h_t)g_2}{E(h_t)[g_1 \cdot \mathrm{Prob}(s_t^* = 1) + g_2 \cdot \mathrm{Prob}(s_t^* = 2)]} = \frac{0{\cdot}088}{(1)(0{\cdot}16) + (0{\cdot}088)(0{\cdot}84)} = 0{\cdot}38$$

One implication is that aggregate economic variables such as the index of industrial production should be quite useful in forecasting stock price volatility. Indeed, Table IV indicates that Model C offers substantially better forecasts of stock volatility than any of the univariate ARCH-type models considered.

Conversely, our model also implies that stock volatility may be useful for forecasting the direction of aggregate economic activity, particularly since stocks have been found here to be a leading indicator of the state of the business cycle. Table III confirms that Model C offers a substantial improvement over univariate specifications for forecasting industrial production growth, although the bivariate VAR gives a comparable improvement. The VAR forecasts industrial production as a linear function of the stock return, while Model C forecasts industrial production as a nonlinear function of the squared stock return. The results suggest that the best forecasts of industrial production might combine features of both, for example, by allowing nonzero elements for the (1, 2) element of $\boldsymbol{\phi}_j$ in equation (10).

The specification considered here is particularly well suited for anticipating the state of the business cycle. The bottom panel of Figure 2 plots for each date t the filter probabilities associated with Model C, or the real-time inference about the current state of the business cycle:

$$\mathrm{Prob}(s_t^* = 1 \mid \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1)$$

These do not lead the business cycle downturn, and so cannot signal the occurrence of a recession in advance. However, the filter probabilities typically do give a fairly reliable contemporaneous indication that a recession has started or ended. The one false alarm comes from the October 1987 stock market crash, on the basis of which our model would have concluded that there was a 67% probability that the economy had entered a recession. However, this conclusion would have been reversed quickly after observing the two months of industrial production gains that followed. The model also gave a premature signal of a likely end to the 1973 recession on the basis of the production gains of May 1974, though this again was immediately reversed after seeing production resuming its fall in June.

To put a quantitative measure on Model C's ability to forecast the state of the business cycle, let D_t be a dummy variable equal to unity if the NBER has determined that the economy was in a recession at date t and equal to zero if the economy was in expansion. The average value of D_t over 1965:5 to 1993:6 was 0·172, meaning that the economy was in recession 17·2% of the time. A naive forecast would be that the probability that the economy will be in recession next month is constant ($\hat{D}_t = 0{\cdot}172$), regardless of current economic conditions. The MSE of this naive forecast is

$$(T-1)^{-1} \sum_{t=2}^{T} (D_t - \hat{D}_t)^2 = 0{\cdot}142$$

while the log probability score (LPS) popularized by Diebold and Rudebusch (1989) is

$$(T-1)^{-1} \sum_{t=2}^{T} [D_t \log(\hat{D}_t) + (1 - D_t)\log(1 - \hat{D}_t)] = 0{\cdot}458$$

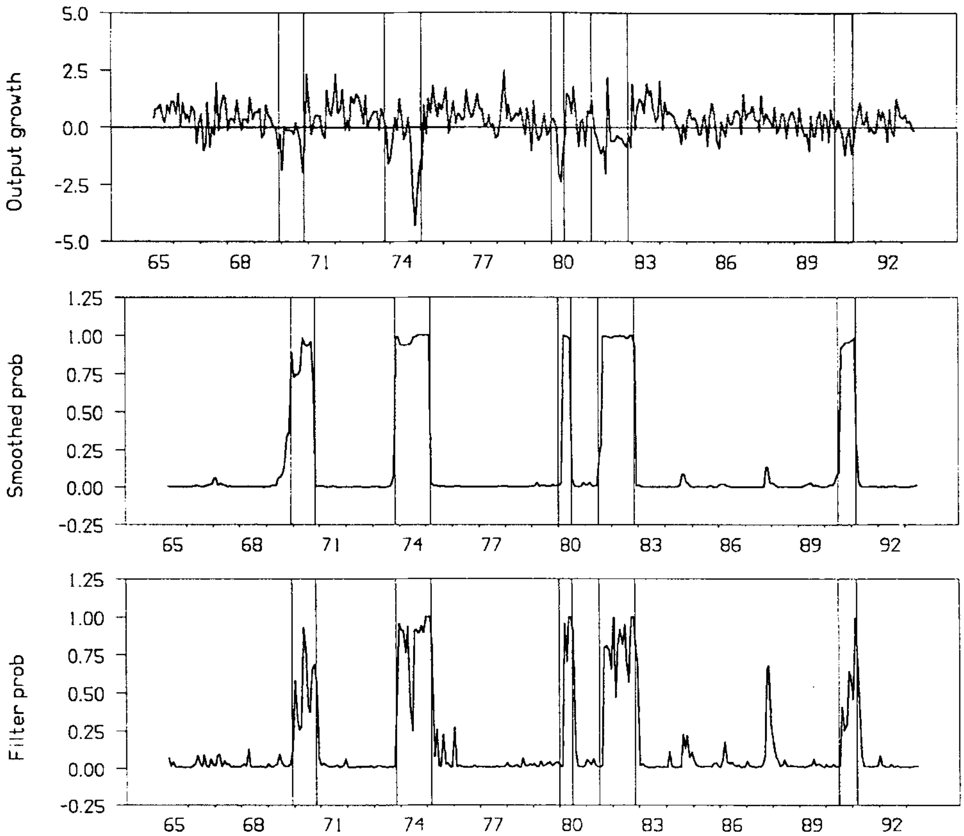

Figure 2. (a) rate of growth of industrial production, quoted at a monthly rate. (b) smoothed probability calculated from Model C that economy is in the high-volatility, low-output growth state at date t (based on information available through the end of the sample). (c) filter probability calculated from Model C that economy is in the high-volatility low-output growth state at date t (based on information available at date t). Vertical lines denote business cycle peak and trough dates as determined by NBER. All data are monthly 1965:4 to 1993:6

By contrast, note that the filter inference provided by Model C offers a forecast of the state of industrial production for the following period:

$$\hat{D}_t = \text{Prob}(s^*_{t-1} = 1 \mid \mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \ldots, \mathbf{x}_1)$$

The MSE of this forecast turns out to be 0·0501. Model C thus has an 'R^2' of

$$1 - (0{\cdot}0501/0{\cdot}142) = 0{\cdot}64$$

in its ability to forecast the future state of the economy. Similarly, the LPS for Model C is 0·184, or a 60% improvement over the naive forecast.

Although these results are highly encouraging, additional evidence on their usefulness in a real-time forecasting exercise would be helpful. Diebold and Rudebusch (1991a,b) have emphasized the importance of *ex-post* data revisions in series such as the industrial production index. Forecasts based on the preliminary estimates actually available to an applied forecaster might be significantly worse than those based on the revised data analysed here.

4. CONCLUSIONS

Why is the stock market much more volatile at some times than others? This paper has arrived at the same answer provided by Schwert (1989a, 1989b)—economic recessions are the single largest factor, accounting for over 60% of the variance of stock returns. We go beyond Schwert's analysis in proposing a time series model that can be used to form forecasts based on this relation. We found the model to be useful both for forecasting stock volatility and for identifying and forecasting economic turning points.

ACKNOWLEDGEMENTS

Research support from the National Science Foundation under grant SBR93-08301 is gratefully acknowledged. Data and software used in this study can be obtained at no charge by writing James D. Hamilton, Department of Economics 0508, UCSD, La Jolla, CA 92093-0508, USA.

REFERENCES

Andrews, D. W. K. (1993), 'Tests for parameter instability and structural change with unknown change point', *Econometrica*, **61**, 821–856.

Attanasio, O. (1991), 'Risk, time-varying second moments and market efficiency', *Review of Economic Studies*, **58**, 479–494.

Black, F. (1976), 'Studies of stock market volatility changes', *Proceedings of the American Statistical Association*, Business and Economic Statistics Section, 177–181.

Boldin, M. D. (1994), 'Does the link between money and output differ over the business cycle?' Working Paper, Federal Reserve Bank of New York.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics*, **31**, 307–327.

Bollerslev, T., R. Y. Chou and K. F. Kroner (1992), 'ARCH modeling in finance: a review of the theory and empirical evidence', *Journal of Econometrics*, **52**, 5–59.

Brunner, A. D. (1991), 'Testing for structural breaks in U.S. post-war inflation data', Mimeograph, Board of Governors of the Federal Reserve System, Washington, DC.

Cai, J. (1994), 'A Markov model of unconditional variance in ARCH', *Journal of Business and Economic Statistics*, **12**, 309–316.

Campbell, J. Y. (1987), 'Stock returns and the term structure', *Journal of Financial Economics*, **18**, 373–399.

Campbell, J. Y., S. Kim and M. Lettau (1993), 'Dispersion and volatility in stock returns: an empirical investigation', Mimeograph, Princeton University.

Chauvet, M. (1996), 'An econometric characterization of business cycle dynamics with factor structure and regime switching', Working Paper 96-01, University of California at Riverside.

Chen, Nai-fu (1991), 'Financial investment opportunities and the macroeconomy', *Journal of Finance*, **46**, 529–554.

Chen, Nai-fu, R. Roll and S. A. Ross (1986), 'Economic forces and the stock market', *Journal of Business*, **56**, 383–403.

Diebold, F. X. and G. Rudebusch (1990), 'A nonparametric investigation of duration dependence in the American business cycle', *Journal of Political Economy*, **98**, 596–616.

Diebold, F. X. and G. Rudebusch (1991a), 'Forecasting output with the composite leading index: a real-time analysis', *Journal of the American Statistical Association*, **86**, 603–610.

Diebold, F. X. and G. Rudebusch (1991b), 'Turning point prediction with the composite leading index: an *ex ante* analysis', in K. Lahiri and G. H. Moore (eds), *Leading Economic Indicators: New Approaches and Forecasting Records*, Cambridge University Press, Cambridge.

Diebold, F. X. and G. Rudebusch (1996), 'Measuring business cycles: A modern perspective', *Review of Economics and Statistics*, **78**, 67–77.

Diebold, F. X., G. Rudebusch and D. E. Sichel (1993), 'Further evidence on business-cycle duration dependence', in J. H. Stock and M. W. Watson (eds), *Business Cycles, Indicators, and Forecasting*, University of Chicago Press, Chicago.

Durland, J. M. and T. H. McCurdy (1994), 'Duration-dependent transitions in a Markov model of U.S. GNP growth', *Journal of Business and Economic Statistics*, **12**, 279–288.
Engel, C. M. and A. P. Rodrigues (1989), 'Tests of international CAPM with time-varying covariances', *Journal of Applied Econometrics*, **4**, 119–138.
Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', *Econometrica*, **50**, 987–1007.
Engle, R. F. (1993), 'Statistical models for financial volatility', *Financial Analysts Journal*, **49**, 72–78.
Engle, R. F. and C. Mustafa (1992), 'Implied ARCH models from options prices', *Journal of Econometrics*, **52**, 289–311.
Engle, R. F. and V. K. Ng (1993), 'Measuring and testing the impact of news on volatility', *Journal of Finance*, **48**, 1749–1777.
Fama, E. F. (1990), 'Stock returns, expected returns, and real activity', *Journal of Finance*, **45**, 1089–1108.
Fama, E. F. and K. R. French (1989), 'Business conditions and expected returns on stocks and bonds', *Journal of Financial Economics*, **25**, 23–49.
Filardo, A. J. (1994), 'Business cycle phases and their transitional dynamics', *Journal of Business and Economic Statistics*, **12**, 299–308.
French, M. W. and D. F. Sichel (1993), 'Cyclical patterns in the variance of economic activity', *Journal of Business and Economic Statistics*, **11**, 113–119.
Ghysels, E. (1993), 'A time series model with periodic stochastic regime switching', Mimeograph, University of Montreal.
Glosten, L. R., R. Jagannathan and D. Runkle (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *Journal of Finance*, **48**, 1779–1801.
Goodwin, T. H. (1993), 'Business cycle analysis with a Markov-switching model', *Journal of Business and Economic Statistics*, **11**, 331–39.
Hamilton, J. D. (1988), 'Rational–expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates', *Journal of Economic Dynamics and Control*, **12**, 385–423.
Hamilton, J. D. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica*, **57**, 357–384.
Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, NJ.
Hamilton, J. D. and R. Susmel (1994), 'Autoregressive conditional heteroskedasticity and changes in regime', *Journal of Econometrics*, **64**, 307–333.
Hansen, B. E. (1992), 'The likelihood ratio test under non-standard conditions: testing the Markov trend model of GNP', *Journal of Applied Econometrics*, **7**, S61–S82.
Hansen, B. E. (1996), 'Erratum: The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP', *Journal of Applied Econometrics*, **11**, 195–198.
Jefferson, P. N. (forthcoming), 'Inference using qualitative and quantitative information with an application to the conduct of monetary policy', *Economic Inquiry*.
Keim, D. B. and R. F. Stambaugh (1986), 'Predicting returns in the stock and bond markets', *Journal of Financial Economics*, **17**, 357–390.
Kim, M. J. and J.-S. Yoo (1995), 'New index of coincident indicators: a multivariate Markov switching factor model approach', *Journal of Monetary Economics*, **36**, 607–630.
Lam, P. (1990), 'The Hamilton model with a general autoregressive component: estimation and comparison with other models of economic time series', *Journal of Monetary Economics*, **26**, 409–32.
Lamoureux, C. G. and W. D. Lastrapes (1990), 'Persistence in variance, structural change and the GARCH model', *Journal of Business and Economic Statistics*, **8**, 225–234.
Lamoureux, C. G. and W. D. Lastrapes (1993), 'Forecasting stock return variance: toward an understanding of stochastic implied volatilities', *Review of Financial Studies*, **5**, 293–326.
Nelson, D. (1991), 'Conditional heteroskedasticity in asset returns: a new approach', *Econometrica*, **59**, 347–370.
Officer, R. R. (1973), 'The variability of the market factor of the New York Stock Exchange', *Journal of Business*, **46**, 434–453.
Pagan, A. and G. W. Schwert (1990), 'Alternative models for conditional stock volatility', *Journal of Econometrics*, **45**, 267–290.
Schwert, G. W. (1989a), 'Why does stock market volatility change over time?' *Journal of Finance*, **44**, 1115–1153.

Schwert, G. W. (1989b), 'Business cycles, financial crises, and stock volatility', *Carnegie-Rochester Conference Series on Public Policy*, **31**, 83–126.

Schwert, G. W. (1990), 'Stock returns and real activity: a century of evidence', *Journal of Finance*, **45**, 1237–1257.

Sichel, D. E. (1991), 'Business cycle duration dependence: a parametric approach', *Review of Economics and Statistics*, **71**, 254–260.

Susmel, R. (1994a), 'Switching volatility in international equity markets', Working Paper, University of Houston.

Susmel, R. (1994b), 'Switching volatility and persistence: empirical evidence and simulation results', Working Paper, University of Houston.

Turner, C. M., R. Startz and C. R. Nelson (1989), 'A Markov model of heteroskedasticity, risk, and learning in the stock market', *Journal of Financial Economics*, **25**, 3–22.