Advanced information on the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel
8 October 2003

KUNGL.
VETENSKAPSAKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

Information Department, P.O. Box 50005, SE-104 05 Stockholm, Sweden
Phone: +46 8 673 95 00, Fax: +46 8 15 56 70, E-mail: info@kva.se, Website: ww.kva.se

Time-series Econometrics: Cointegration and Autoregressive Conditional Heteroskedasticity

Time-series Econometrics:
Cointegration and Autoregressive Conditional Heteroskedasticity

1. Introduction

Empirical research in macroeconomics as well as in financial economics is largely based on time series. Ever since Economics Laureate Trygve Haavelmo's work it has been standard to view economic time series as realizations of stochastic processes. This approach allows the model builder to use statistical inference in constructing and testing equations that characterize relationships between economic variables. This year's Prize rewards two contributions that have deepened our understanding of two central properties of many economic time series – nonstationarity and time-varying volatility – and have led to a large number of applications.

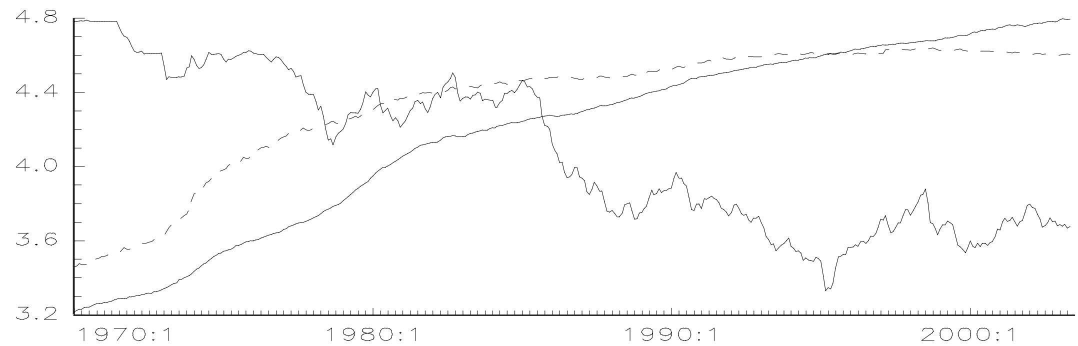

Figure 1.1: Logarithm (rescaled) of the Japanese yen/US dollar exchange rate (decreasing solid line), logarithm of seasonally adjusted US consumer price index (increasing solid line) and logarithm of seasonally adjusted Japanese consumer price index (increasing dashed line), 1970:1 − 2003:5, monthly observations

Nonstationarity, a property common to many macroeconomic and financial time series, means that a variable has no clear tendency to return to a constant value or a linear trend. As an example, Figure 1.1 shows three monthly series: the value of the US dollar expressed in Japanese yen, and seasonally adjusted consumer price indices for the US and Japan. None of these series, of which the price series are considerably smoother than the exchange rate, seems to be stationary, i.e., return to a fixed value or fluctuate around a linear trend (in which case the deviations from trend are stationary). Other aggregate variables, such as gross national product, consumption, employment, and asset prices share this property. It is therefore proper to assume that they have been generated by nonstationary processes and follow stochastic trends.

An important objective of empirical research in macroeconomics is to test hypotheses and estimate relationships, derived from economic theory, among such aggregate variables. The statistical theory that was applied well into the 1980s in building and testing large simultaneous-equation models was based on the assumption that the variables in these models are stationary. The problem was that statistical inference associated with stationary processes is no longer valid if the time series are indeed realizations of nonstationary processes.

This difficulty was not well understood among model builders three decades ago. This is no longer the case, and Clive Granger can be credited with this change. He has shown that macroeconomic models containing nonstationary stochastic variables can be constructed in such a way that the results are both statistically sound and economically meaningful. His work has also provided the underpinnings for modeling with rich dynamics among interrelated economic variables. Granger has achieved this breakthrough by introducing the concept of cointegrated variables, which has radically changed the way empirical models of macroeconomic relationships are formulated today.

The second central property of economic time series, common to many fi-

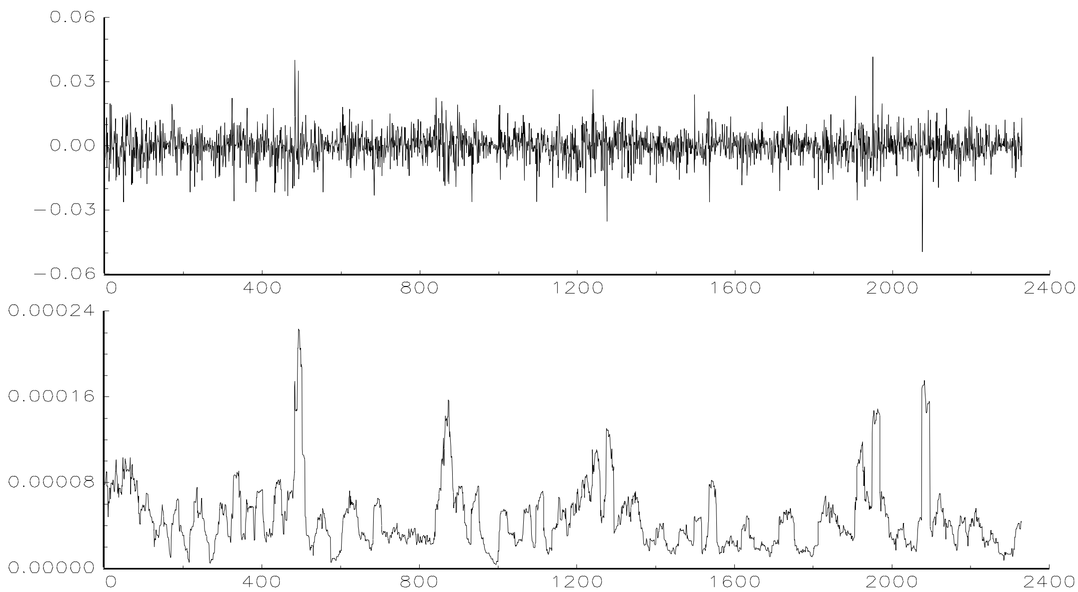

Figure 1.2: Daily returns of the yen/dollar exchange rate and the corresponding 20-day moving average of the squared changes, 1986-1995

nancial time series, is that their volatility varies over time. Consider a financial return series such as the rate of change of a daily exchange rate or stock index. As an example, the upper panel of Figure 1.2 contains the first difference of the series in Figure 1.1 measured at the daily frequency. The lower panel, which displays a 20-day (four trading weeks) moving average of the squared returns, clearly illustrates how high-volatility periods alternate with periods of relative calm.

Volatility of returns is a key issue for researchers in financial economics and analysts in financial markets. The prices of stocks and other assets depend on the expected volatility (covariance structure) of returns. Banks and other financial institutions make volatility assessments as a part of monitoring their risk exposure. Up until the 1980s, both researchers and practitioners in financial markets used models in which volatility was assumed constant over time. As Figure 1.2 illustrates, however, volatility may vary considerably over time: large (small) changes in returns are followed by large (small) changes. The modeling and forecasting of volatility are therefore crucial for financial markets.

Research on volatility models was initiated by Robert Engle who, in the early 1980s, developed a new concept that he termed autoregressive conditional heteroskedasticity, and acronymized ARCH. Since their advent, models built around this concept have become an indispensable tool for financial analysts, bankers and fund managers throughout the world. For two decades, Robert Engle has remained at the forefront of research on volatility modelling and made several outstanding contributions in this area.

2. Cointegrated economic variables

Macroeconomists build time-series models for testing economic theories, for forecasting, and for policy analysis. Such models are constructed and applied by economists at universities, economic research institutes and central banks. There is a long tradition of building large macroeconomic models with hundreds of equations and variables. More recently, small models with only a handful of equations and variables have become more common. Since many of the time series model builders use are best viewed as nonstationary, exploiting such series requires both a new approach and statistical inference different from the traditional inference developed for applications to stationary series.

In this section, we describe Clive Granger's contributions that lead up to the concept of cointegration and its applications. We begin by defining the concept and the statistical theory related to it, including the so-called Granger representation theorem. This is followed by a description of the two-step method used to test for cointegrating relationships and estimate equation systems with cointegrated variables. A number of extensions of the basic concept of cointegration are briefly highlighted. We end by discussing applications. Empirical research on the purchasing power parity (PPP) hypothesis is used to illustrate how cointegration may not only change empirical analysis, but also give it a new dimension.

2.1. Cointegration: basic definition

Returning to the time series in Figure 1.1, it is appropriate to assume that they have been generated by nonstationary stochastic processes. For a long time it was common practice to estimate equations involving nonstationary variables in macroeconomic models by straightforward linear regression. It was not well understood that testing hypotheses about the coefficients using standard statistical inference might lead to completely spurious results. In an influential paper, Clive Granger and his associate Paul Newbold (Granger and Newbold (1974)) pointed out that tests of such a regression may often suggest a statistically significant relationship between variables where none in fact exists. Granger and Newbold reached their conclusion by generating independent nonstationary series, more precisely random walks.[1] They regressed these series on each other and observed the value of the t-statistic of the coefficient estimate calculated under the assumption that the true value of the coefficient equals zero. Despite the fact that the variables in the regression were independent, the authors found that the null hypothesis of a zero coefficient was rejected much more frequently than standard theory predicts. At the same time, they observed that the residuals of the estimated equation displayed extremely strong positive autocorrelation.[2]

These results indicated that many of the apparently significant relationships between nonstationary economic variables in existing econometric models could well be spurious. This work formed an initial step in Granger's research agenda of developing methods for building more realistic and useful econometric models.

Statisticians working with time-series models suggested a simple solution to the "spurious regressions" problem. If economic relationships are specified in first differences instead of levels, the statistical difficulties due to nonstationary variables can be avoided because the differenced variables are usually stationary even if the original variables are not. Economic theories, however, are generally formulated for levels of variables rather than for differences. For example, theories of consumption postulate a relationship between the levels of consumption, income, wealth and other variables – and not their growth rates. A model relating the first differences of these variables would typically not make full use of these theories. An alternative approach would involve removing a linear time trend from the variables and specifying the empirical relationship between them

[1] Assume that $\{\varepsilon_\tau\}, \tau = 0,1,2,...,$ is a sequence of independent stochastic variables with mean zero and variance σ^2 and let $\xi_t = \sum_{\tau=0}^{t} \varepsilon_\tau$. Sequence $\{\xi_t\}, t = 0,1,2,...,$ is then a random walk (without drift).

[2] Granger and Newbold (1974) reached their conclusions by simulation. The asymptotic distribution theory valid for their experiment was worked out more than a decade later by Phillips (1986). A compact presentation of these developments can be found in Granger (2001).

using detrended variables. Removing (separate) time trends assumes, however, that the variables follow separate deterministic trends, which does not appear realistic, given the awkward long-run implications. Dynamic econometric models based on linearly detrended variables may thus be able to characterize short-term dynamics of economic variables but not their long-run relationships. The same is true for models based solely on first differences.

Clive Granger's solution to this problem may be illustrated by the simplest possible regression equation:

$$y_t = \alpha + \beta x_t + \varepsilon_t, \tag{2.1}$$

where y_t is the dependent variable, x_t the single exogenous regressor, and $\{\varepsilon_t\}$ a white-noise, mean-zero sequence. Granger (1981) argues that in order to be meaningful, an equation has to be consistent in the sense that "a simulation of the explanatory right-hand side should produce the major properties of the variable being explained". For example, if y_t is a seasonal variable, then x_t has to be seasonal, if ε_t is to be white noise. To develop the idea further, Granger (1981) defined the concept of *degree of integration* of a variable. If variable z_t can be made approximately stationary by differencing it d times, it is called integrated of order d, or I(d). Weakly stationary random variables are thus I(0). Many macroeconomic variables can be regarded as I(1) variables: if $z_t \sim \mathrm{I}(1)$, then $\Delta z_t \sim \mathrm{I}(0)$. Note that I(1) variables dominate I(0) variables; in a linear combination of variables the variation of the former overwhelms the variation of the latter. To illustrate, if $z_t \sim \mathrm{I}(1)$ and $w_t \sim \mathrm{I}(0)$, then $z_t + w_t \sim \mathrm{I}(1)$.

Consider again equation (2.1) and assume that both $x_t \sim \mathrm{I}(1)$ and $y_t \sim \mathrm{I}(1)$. Then, generally $y_t - \beta x_t \sim \mathrm{I}(1)$ as well. There is, however, one important exception. If $\varepsilon_t \sim \mathrm{I}(0)$, then $y_t - \beta x_t \sim \mathrm{I}(0)$, i.e., the linear combination $y_t - \beta x_t$ has the same statistical properties as an I(0) variable. There exists only one such combination so that coefficient β is unique.[3] In this special case, variables x_t and y_t are called *cointegrated*. More generally, if a linear combination of a set of I(1) variables is I(0), then the variables are cointegrated. This concept, introduced in Granger (1981), has turned out to be extremely important in the analysis of nonstationary economic time series. A generalization to I(d) variables, where d is no longer an integer, is also possible, in which case the linear combination of cointegrated variables has to be $\mathrm{I}(d - d_0)$, where $d_0 > 0$.

[3] Uniqueness can be shown as follows. Suppose there are two cointegrating relations between the I(1) variables y_t and x_t:

$$y_t = \beta_j x_t + \varepsilon_{jt}, j = 1, 2,\ \beta_1 \neq \beta_2.$$

Subtracting the second from the first yields

$$(\beta_2 - \beta_1) x_t = \varepsilon_{1t} - \varepsilon_{2t}.$$

The left-hand side of this equation is I(1) whereas the right-hand side as a difference of two I(0) variables is I(0). This is a contradiction unless $\beta_1 = \beta_2$, in which case $\varepsilon_{1t} = \varepsilon_{2t}$.

The importance of cointegration in the modeling of nonstationary economic series becomes clear in the so-called Granger representation theorem, first formulated in Granger and Weiss (1983). In order to illustrate this result, consider the following bivariate autoregressive system of order p :

$$\begin{aligned} x_t &= \sum_{j=1}^{p} \gamma_{1j} x_{t-j} + \sum_{j=1}^{p} \delta_{1j} y_{t-j} + \varepsilon_{1t} \\ y_t &= \sum_{j=1}^{p} \gamma_{2j} x_{t-j} + \sum_{j=1}^{p} \delta_{2j} y_{t-j} + \varepsilon_{2t}, \end{aligned}$$

where x_t and y_t are I(1) and cointegrated, and ε_{1t} and ε_{2t} are white noise. The Granger representation theorem says that in this case, the system can be written as:

$$\begin{aligned} \Delta x_t &= \alpha_1(y_{t-1} - \beta x_{t-1}) + \sum_{j=1}^{p-1} \gamma_{1j}^{*} \Delta x_{t-j} + \sum_{j=1}^{p-1} \delta_{1j}^{*} \Delta y_{t-j} + \varepsilon_{1t} \\ \Delta y_t &= \alpha_2(y_{t-1} - \beta x_{t-1}) + \sum_{j=1}^{p-1} \gamma_{2j}^{*} \Delta x_{t-j} + \sum_{j=1}^{p-1} \delta_{2j}^{*} \Delta y_{t-j} + \varepsilon_{2t}, \qquad (2.2) \end{aligned}$$

where at least one of parameters α_1 and α_2 deviates from zero. Both equations of the system are "balanced", that is, their left-hand and right-hand sides are of the same order of integration, since $y_{t-1} - \beta x_{t-1} \sim \mathrm{I}(0)$.

Suppose that $y_t - \beta x_t = 0$ defines a dynamic equilibrium relationship between the two economic variables, y and x. Then $y_t - \beta x_t$ is an indicator of the degree of disequilibrium. The coefficients α_1 and α_2 represent the strength of the disequilibrium correction, and the system is now said to be in *error-correction form.* A system characterized by these two equations is thus in disequilibrium at any given time, but has a built-in tendency to adjust itself towards the equilibrium.

Thus, an econometric model cannot be specified without knowing the order of integration of the variables. Tests of the unit root (nonstationarity) hypothesis were developed by Fuller (1976), Dickey and Fuller (1979, 1981), Phillips and Perron (1988), and others.[4] When these tests are applied to each of the three time series in Figure 1.1, the null hypothesis of a unit root cannot be rejected. But a unit root is rejected for the first differences of the series. The series can thus be regarded as realizations of stochastic I(1) variables.

It should be mentioned that linear combinations of nonstationary variables had appeared in dynamic econometric equations prior to Granger's work on cointegration. Phillips (1957), who coined the term "error correction", and Sargan (1964) were forerunners. The well-known consumption equation in Davidson, Hendry, Srba and Yeo (1978), the so-called DHSY model, disseminated the idea

[4]Differencing a variable with a unit root removes the unit root.

among macroeconomists. In their paper, based on quarterly UK series, a lagged difference of $c_t - y_t$, where c_t is the logarithm of private consumption of nondurable goods and y_t private income, represented the error-correction component. These authors did not, however, consider the statistical implications of introducing such components into their models.[5]

2.2. Cointegration: estimation and testing

The concept of cointegration would not have become useful in practice without a statistical theory for testing for cointegration and for estimating parameters of linear systems with cointegration. Granger and Robert Engle jointly developed the necessary techniques in their classical and remarkably influential paper, Engle and Granger (1987), where the theory of cointegrated variables is summed up and extended. The paper contains, among other things, a rigorous proof of the Granger representation theorem.

Engle and Granger (1987) consider the problem of testing the null hypothesis of no cointegration between a set of I(1) variables. They estimate the coefficients of a static relationship between these variables by ordinary least squares and apply well-known unit root tests to the residuals. Rejecting the null hypothesis of a unit root is evidence in favor of cointegration. The performance of a number of such tests is compared in the paper.

More recently, it has become possible to test the null hypothesis that the estimated linear relationship between the I(1) variables is a cointegrating relationship (errors in the regression are stationary) against the alternative of no cointegration (errors are nonstationary). Tests of this hypothesis were developed by Shin (1994), based on a well-known stationarity test in Kwiatkowski, Phillips, Schmidt and Shin (1992), as well as by Saikkonen and Luukkonen (1993), Xiao and Phillips (2002), and others.

Another fundamental contribution in Engle and Granger (1987) is the two-stage estimation method for vector autoregressive (VAR) models with cointegration. Consider the following VAR model of order p:

$$\Delta \mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \sum_{j=1}^{p-1} \mathbf{\Gamma}_j \mathbf{\Delta x}_{t-j} + \boldsymbol{\varepsilon}_t \qquad (t = 1, ..., T) \tag{2.3}$$

where $\mathbf{x}_t$ is an $n \times 1$ vector of I(1) variables, $\boldsymbol{\alpha}\boldsymbol{\beta}'$ is an $n \times n$ matrix such that the $n \times r$ matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ have rank r, $\mathbf{\Gamma}_j$, $j = 1, ..., p-1$, are $n \times n$ parameter matrices, and $\boldsymbol{\varepsilon}_t$ is an $n \times 1$ vector of white noise with a positive definite covariance matrix. If $0 < r < n$, the variables in $\mathbf{x}_t$ are cointegrated with r cointegrating relationships $\boldsymbol{\beta}'\mathbf{x}_t$. Stock (1987) had shown that under certain regularity conditions, the least squares estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is consistent and converges

[5] For a discussion of the DHSY model using cointegration analysis, see Hendry, Muellbauer and Murphy (1990).

to the true value at the rapid rate T^{-1} (T is the number of observations); for this reason $\widehat{\boldsymbol{\beta}}$ is called superconsistent. Using this result, Granger and Engle showed that the maximum likelihood estimator of the remaining parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\Gamma}_j$, $j = 1, ..., p-1$, obtained by replacing $\boldsymbol{\beta}$ by $\widehat{\boldsymbol{\beta}}$, has the same asymptotic distribution as the estimator based on the true value $\boldsymbol{\beta}$.

If the variables in $\mathbf{x}$ are cointegrated, the parameters of (2.3) can thus be estimated in two stages: begin by estimating $\boldsymbol{\beta}$ or, more precisely, the cointegrating space ($\boldsymbol{\beta}$ up to a multiplicative constant) using a form of least squares. Then, holding that estimate fixed, estimate the remaining parameters by maximum likelihood. The estimators of $\boldsymbol{\alpha}$ and $\boldsymbol{\Gamma}_j$, $j = 1, ..., p-1$, are consistent and asymptotically normal. Hypotheses involving these parameters and their values can be tested using standard statistical inference.

The results in Engle and Granger (1987) opened the gates for a flood of applications. They enhanced the popularity of VAR models developed by Sims (1980) to offer an alternative to simultaneous-equation models. Sims had emphasized the use of unrestricted VAR models as a means of modelling economic relationships without unnecessary assumptions. On the other hand, a VAR model with cointegration is often based on the idea of a "long-run", or moving equilibrium, defined by economic theory and characterized by vector $\boldsymbol{\beta}'\mathbf{x}_{t-1}$ in (2.3) . The short-term dynamics represented by the parameter matrices $\boldsymbol{\Gamma}_j$, $j = 1, ..., p-1$, are free from restrictions. So is the strength-of-adjustment matrix $\boldsymbol{\alpha}$ that describes the contribution of the degree of long-run disequilibrium to the adjustment process towards the moving target or equilibrium.

Engle and Granger's two-step method represented a decisive breakthrough in the modeling of economic relationships using nonstationary cointegrated time series. Among later developments, the work of Johansen (1988, 1991) deserves special mention. Johansen derived the maximum likelihood estimator of $\boldsymbol{\beta}$ or, more precisely, the space spanned by the r cointegrating vectors in (2.3) using reduced rank regression.[6] He also derived sequential tests for determining the number of cointegrating vectors. Johansen's method can be seen as a second-generation approach, in the sense that it builds directly on maximum likelihood estimation instead of partly relying on least squares.

2.3. Extensions to cointegration

Granger and Engle, with various co-authors, have also extended the concept of cointegration to seasonally integrated variables. In applied work, it is quite common to render a time series with strong seasonal variation stationary by seasonal differencing. For example, if x_t is a nonstationary quarterly series, its seasonal difference $\Delta_4 x_t = x_t - x_{t-4}$ may be I(0). If two nonstationary seasonal

[6] In practice, vector $\boldsymbol{\beta}$ is normalized by fixing one of its elements so that its estimate, as well as the estimate of $\boldsymbol{\alpha}$, are unique.

series x_t and y_t can be made I(0) by seasonal differencing and there exists a linear combination $y_t - \beta x_t \sim$ I(0), then the two series are called seasonally cointegrated. Such series were first analyzed by Hylleberg, Engle, Granger and Yoo (1990).

Granger and Lee (1990) defined the concept of multicointegration, which can be a useful tool when modeling stock-flow relationships. Suppose x_t and y_t are cointegrated. Then the cumulative sum of deviations from the cointegrating relation $y_t - \beta x_t = 0$ is necessarily an I(1) variable. If this new variable is cointegrated with one of the original cointegrated variables, x_t or y_t, then the latter two are multicointegrated.[7]

In many cases, deviations from equilibrium are explained by transaction and information costs. Granger and Swanson (1996) demonstrate how such costs may be incorporated into models with cointegrated variables and how this may give rise to a nonlinear error correction model. Adjustment costs are often lumpy, however, and an adjustment will not occur unless the deviation from the equilibrium or desired value exceeds a certain threshold value. Granger and Swanson show how such mechanisms, analyzed for instance in (S, s) models for adjusting inventories and in menu cost models for price adjustment, can be incorporated into models with cointegrated variables. Statistical theory for this type of cointegration was first worked out by Balke and Fomby (1997), who called it threshold cointegration. For recent developments in this area, see Lo and Zivot (2001).

2.4. Areas of application

Cointegration has become a common econometric tool for empirical analysis in numerous areas, where long-run relationships affect currently observed values: current consumption is restricted by expected future income, current long-term interest rates are determined by expected short-term rates, and so on. In such areas, potential cointegrating relationships can be derived from economic theory, tested, and – if there is indeed cointegration – incorporated into econometric models.

The wealth-consumption relationship is an example where the interplay between theory and practice is changing our view of the real world. The traditional view in many textbooks is that an increase in wealth causes a rise in consumption roughly in proportion to the real interest rate. This magnitude might also appear reasonable in terms of the so-called life-cycle model of consumption and savings. If it were true, the fluctuations on stock and housing markets would have a very strong impact on consumption.

The traditional view, however, relies on econometric studies and simulated

[7] For example, sales and production in an industry may be I(1) and cointegrated, in which case their difference, the change in inventory, is an I(0) variable. Then, the level of inventory (initial level plus cumulated changes) will be I(1). With a target level of inventory, defined as a fixed proportion of sales, inventory and sales would be cointegrated, This, in turn, makes the original variables, sales and production, multicointegrated.

models that do not distinguish sufficiently between temporary and permanent perturbations in wealth. A very recent study by Lettau and Ludvigson (2003) shows that consumption, labor income and wealth have to be cointegrated if households observe an intertemporal budget constraint in their consumption behavior. After ascertaining that this theoretical starting-point agrees with their data, the authors estimate an error-correction model that gives two results. First, a majority of perturbations in wealth are temporary and related mainly to fluctuations in the stock market and second, such temporary perturbations have little effect on consumption, both in the short and long run.

Well-known cointegration studies based on economic theory include Campbell and Shiller (1987) who study bubbles in asset prices, Campbell and Shiller (1988), Cochrane (1994) and Lettau and Ludvigson (2001) who investigate the predictability of stock prices, Campbell (1987) who tests the hypothesis that consumption is determined by permanent income, King, Plosser, Stock and Watson (1991), who consider the role of permanent productivity shocks in the postwar US economy, Johansen and Juselius (1990) who study demand for money, and Hall, Anderson and Granger (1992) who consider the term structure of interest rates.

The study of exchange rates and prices allows us to illustrate, in a simple way, how cointegration may transform empirical analysis and reveal new ways of investigating old problems. A basic proposition, formulated long ago by Cassel (1922), is that exchange rates adjust so as to maintain purchasing power parity (PPP): the price of a bundle of goods, expressed in common currency, should be the same across countries. (See Froot and Rogoff (1995) and Sarno and Taylor (2002) for surveys of the literature on PPP, exchange rates and prices.) Assuming two countries, we have

$$S_t P_t^* = P_t \tag{2.4}$$

where S_t is the exchange rate between the domestic and the foreign currency, and P_t and P_t^* are the price levels of domestic and foreign goods in local currency. Put differently, the real exchange rate $S_t P_t^*/P_t$ is assumed to be constant and equal to one.

At first, this proposition seems completely at odds with the observed huge swings in exchange rates and relative stability of inflation rates; see e.g., the series in Figure 1.1. The figure indicates that in the 1970s, contrary to the simple PPP hypothesis, the yen gained strength against the dollar while the Japanese inflation rate was higher than the US rate. Developments have been less contradictory thereafter, but the relation between the fluctuating exchange rate and the stable relative price levels nevertheless appears rather weak.

The lack of a clear-cut one-to-one relationship between the exchange rate and relative prices should not seem surprising. First, in the short run, the exchange rate is primarily affected by expectations and capital movements. Consequently, it may take time for trade to smooth out deviations from PPP: at best, PPP is a long-run relationship. Second, deviations from PPP could be explained by some goods being nontradable, as no obvious mechanism would eliminate price differences between such goods. Third, transaction and shipping costs require price differences to be of a certain minimum size before they are removed by trade.

Empirical studies of the PPP hypothesis initially focussed on Cassel's (1922) simple formulation. PPP was tested by using regression models of the form

$$s_t = a + b(p_t - p_t^*) + \varepsilon_t \tag{2.5}$$

where s_t is the logarithm of the exchange rate between the home and foreign currency, and p_t and p_t^* are the logarithms of the price levels of home and foreign goods, respectively. As seen from (2.4) , the null hypothesis of PPP is equivalent to $a = 0$ and $b = 1$ in (2.5). Frenkel (1978, 1981), and many others, estimated equations of this type. In these articles, equation (2.5) was estimated by ordinary least squares and the hypothesis $b = 1$ tested using the standard t-statistic. Strong rejection of the PPP hypothesis was a common result, at least for data from industrialized countries during the post-Bretton Woods period with floating exchange rates. Moreover, the rejections showed no clear pattern: estimates of b deviated from 1 both upwards and downwards. These tests are flawed, however, as they do not take into account the prospective nonstationarity of exchange rates and price levels, i.e., the possibility that $\varepsilon_t \sim \mathrm{I}(1)$. Furthermore, static equations of type (2.5) cannot separate (possibly strong) short-run deviations from PPP and long-run adjustments of the exchange rate towards an equilibrium.

The next generation of studies explicitly treat PPP as a long-run relationship, implying that deviations from PPP should follow a stationary process. This amounts to assuming $a = 0$ and $b = 1$ in (2.5) and checking whether the residual, that is, the deviation from a fixed real exchange rate, is an I(0) process. When such tests were applied to post-Bretton Woods time series, they did not reject the unit root hypothesis; see e.g., Meese and Rogoff (1988), and thus did not support PPP even as a long-run relationship. An explanation offered in the literature was that the time series had not been long enough for the unit-root tests to have power. When the series are extended backwards a century or more, the unit-root hypothesis is indeed generally rejected; see for example Frankel (1986).

If PPP is valid in the long run, how quick is the adjustment towards it? Edison (1987) examined this question, applying an error-correction framework to series covering a century or more. This meant estimating equations of the type

$$\Delta s_t = a + b(s_{t-1} - p_{t-1} + p_{t-1}^*) + g(\Delta p_{t-1} - \Delta p_{t-1}^*) + \varepsilon_t, \tag{2.6}$$

where b measures the rate of convergence towards PPP. Typical results suggest half-lives of deviations from PPP of between 3 and 7 years. The short-run dynamic behavior of the exchange rate can then be explained as a combination of transitory shocks and gradual adjustment towards PPP.

These second-generation studies assume that there exists a cointegrating relationship between s_t, p_t and p_t^* with cointegrating vector $\boldsymbol{\beta} = (1, -1, 1)$, obtained from the simple form of PPP. It seems realistic, however, to allow for a trending real exchange rate, due to different trends in the relative prices of traded vs. nontraded goods (the so-called Balassa-Samuelson effect).[8] The relevant cointegrating vector in such a situation is $\boldsymbol{\beta} = (1, -\mu, \mu^*)'$ where parameters μ and μ^*, if they deviate from unity, may reflect different price trends for traded and nontraded goods.[9] To consider this extension, begin by running the regression

$$s_t = \alpha + \mu p_t - \mu^* p_t^* + \varepsilon_t \tag{2.7}$$

and testing nonstationarity (or stationarity) of the errors. Stationary errors support the more general version of PPP. Next, if the errors are found to be stationary, the whole dynamic system can be estimated as discussed in Engle and Granger (1987). This procedure gives an idea of whether and how the exchange rate fluctuates around the PPP equilibrium.

A number of third-generation studies using this approach have been conducted since the late 1980s. A common finding is that the null hypothesis of no cointegration is rejected more frequently than the null of a unit root in $s_t - p_t + p_t^*$. The results are still sensitive to the observation period: post-Bretton Woods data tend to produce estimates of μ and μ^* far from unity, while data from the entire twentieth century produce values closer to one.

The three time series in Figure 1.1 serve to demonstrate how the validity of the PPP hypothesis may be examined using cointegration analysis. Estimating the parameters of equation (2.7) yields

$$s_t = 6.63 + 0.44 p_t - 1.07 p_t^* + \widehat{\varepsilon}_t. \tag{2.8}$$

The estimated residuals from (2.8) are graphed in Figure 2.1. They appear to fluctuate around zero, which suggests stationarity, but display considerable persistence, which suggests the opposite. Formal scrutiny is required.

Culver and Papell (1999) tested the PPP hypothesis for a number of countries, with cointegration as the null hypothesis against the alternative of nonstationarity, using the test of Shin (1994). Following their approach here, cointegration is

[8] According to the Balassa-Samuelson effect, poor countries grow faster than rich countries, and consumer prices in the former increase faster than they do in the latter. A reason is that faster growth in poor economies is primarily due to productivity growth in the tradable goods sector where prices tend to be equal across all countries. Strong growth in productivity leads to large wage increases in the non-tradables sector as well; in this sector, however, productivity can be expected to grow at about the same rate in all countries. Large wage increases then imply large increases in prices of non-tradables in poor countries. The latter price increases, in turn, raise the growth rate of the aggregate consumer price index in these countries.

[9] This requires the assumption that the foreign price level p_t^* is exogenous in the sense that there is no feedback to it from the other two variables. This restricts the considerations to a system with two equations, so that there can be at most one cointegrating vector.

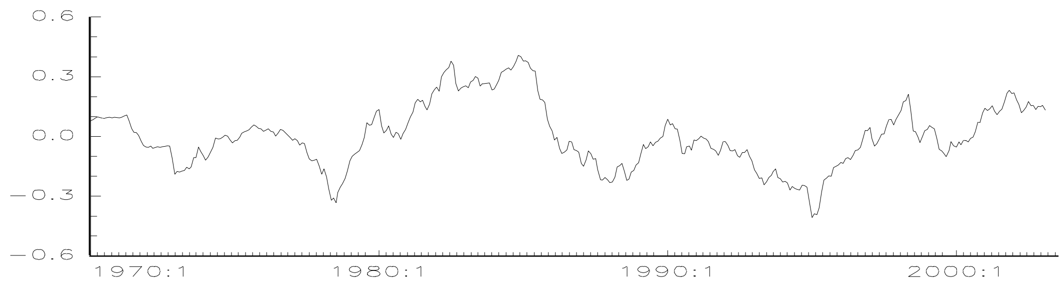

Figure 2.1: Residuals of regression (2.8)

not rejected at the 5% level. If we conclude that s_t, p_t and p_t^* can be regarded as cointegrated, the vector $\widehat{\boldsymbol{\beta}} = (1, -0.44, 1.07)$ estimated from (2.7) is a cointegrating vector. It is not close to $(1, -1, 1)$ as would be required for the restricted version of PPP, but the data provide weak support for a general version of PPP between Japan and the US during the period 1975-2003.[10]

3. Modeling volatility

Many financial economists are concerned with modeling volatility in asset returns. Portfolio-choice theory attempts to derive optimal portfolios as functions of variances and covariances of returns. The capital asset pricing model (CAPM) and other asset-pricing models show how investors are rewarded for taking systematic risks, i.e., risks related to the covariance between their own and the market portfolio or other non-diversifiable factors. Option-pricing formulas give prices of options and other derivative instruments in terms of volatility of the underlying asset. Banks and other financial institutions apply so-called value-at-risk models to assess risks in their marketable assets. For all these purposes, modeling volatility or, in other words, the covariance structure of asset returns, is essential.

Financial economists have long since known that volatility in returns tends to cluster and that the marginal distributions of many asset returns are leptokurtic, which means that they have thicker tails than the density of the normal distribution with the same mean and variance. Even though the time clustering of returns was known to many researchers, returns were still modeled as independent and identically distributed over time. Examples include Mandelbrot (1963) and Mandelbrot and Taylor (1967) who used so-called stable Paretian distributions to characterize the distribution of returns.[11] Robert Engle's modelling

[10] The results are sensitive to the choice of null hypothesis, however. Augmented Dickey-Fuller tests of the unit-root hypothesis, see Dickey and Fuller (1981), give t-values from -2.3 to -2.5, which are not small enough to reject the null hypothesis at the 0.1 significance level (the critical value equals -2.58).

[11] For a thorough account of stable Paretian distributions and their use in finance and econo-

of time-varying volatility by way of autoregressive conditional heteroskedasticity (ARCH) thus signified a genuine breakthrough.

We begin this section by introducing Engle's basic ARCH model, some of its generalizations and a brief application. This is followed by a discussion of the so-called ARCH-in-mean (ARCH-M) model where conditional first moments – in applications, usually expected asset returns – are systematically related to conditional second moments modeled by ARCH. Multivariate generalizations of ARCH models are also considered, as well as new parameterizations of ARCH. We conclude by considering value-at-risk analysis, where ARCH plays an important role.

3.1. Autoregressive conditional heteroskedasticity

In time series econometrics, model builders generally parameterize the conditional mean of a variable or a vector of variables. Suppose that at time t we observe the stochastic vector $(y_t, \mathbf{x}_t')'$ where y_t is a scalar and $\mathbf{x}_t$ a vector of variables such that some of its elements may be lags of y_t. This implies the following predictive model for variable y_t:

$$y_t = \mathsf{E}\{y_t|\mathbf{x}_t\} + \varepsilon_t, \tag{3.1}$$

where the conditional mean $\mathsf{E}\{y_t|\mathbf{x}_t\}$ typically has a parametric form,[12] $\mathsf{E}\{\varepsilon_t|\mathbf{x}_t\} = \mathsf{E}\varepsilon_t = 0$, and $\mathsf{E}\varepsilon_t\varepsilon_s = 0$ for $t \neq s$, that is, $\{\varepsilon_t\}$ is a sequence of uncorrelated random variables with mean zero. When estimating the parameters in $\mathsf{E}\{y_t|\mathbf{x}_t\}$, it is typically assumed that the unconditional variance of the error term ε_t is constant or time-varying in an unknown fashion. (Indeed, we assumed constant variances and covariances in our presentation of estimation of cointegrated systems.)

Engle (1982) considered the alternative assumption that, while the unconditional error variance – if it exists – is constant, the *conditional* error variance is time-varying. This revolutionary notion made it possible to explain systematic features in the movements of variance over time and, *a fortiori*, to estimate the parameters of conditional variance jointly with the parameters of the conditional mean. The literature is wholly devoid of earlier work with a similar idea.

Engle parameterized the conditional variance of ε_t in model (3.1) such that large positive or negative errors ε_t were likely to be followed by another large error of either sign and small errors by a small error of either sign. More formally, he assumed that ε_t can be decomposed as $\varepsilon_t = z_t h_t^{1/2}$, where $\{z_t\}$ is a sequence of iid random variables with zero mean and unit variance, and where

$$h_t = \mathrm{var}(\varepsilon_t|\mathcal{F}_t) = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2. \tag{3.2}$$

metrics, see Rachev and Mittnik (2000).

[12] Alternatively, the conditional mean is a nonparametric function of $\mathbf{x}_t$, in which case the focus is on estimating its functional form.

In (3.2) , $\varepsilon_t = y_t - \mathsf{E}\{y_t|\mathbf{x}_t\}$, and the information set $\mathcal{F}_t = \{\varepsilon_{t-j} : j \geq 1\}$, $\alpha_0 > 0$, and $\alpha_j \geq 0$, $j = 1, ..., q$.

Equation (3.2) defines the ARCH model introduced in Engle (1982), where the conditional variance is a function of past values of squared errors. In this classic paper, Engle developed the estimation theory for the ARCH model, gave conditions for the maximum likelihood estimators to be consistent and asymptotically normal, and presented a Lagrange multiplier test for the hypothesis of no ARCH (equal conditional and unconditional variance) in the errors ε_t.

3.2. Extensions and applications

In practice, ε_t^2 tends to have a relatively slowly decaying autocorrelation function among return series of sufficiently high observation frequency, such as daily or weekly series. An adequate characterization of this stylized fact requires an ARCH model with a long lag q. But if the right-hand side of (3.2) is modified by adding lags of the conditional variance h_t (one lag is often enough), the resulting model can be formulated with only a small number of parameters and still display a slowly decaying autocorrelation function for ε_t^2. Not long after the publication of the ARCH paper, Engle's graduate student Tim Bollerslev introduced such a model and called it the generalized ARCH (GARCH) model. This model has the following form, see Bollerslev (1986),

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}. \tag{3.3}$$

The first-order ($p = q = 1$) GARCH model, also suggested independently by Taylor (1986), has since become the most popular ARCH model in practice. Compared to Engle's basic ARCH model, the GARCH model is a useful technical innovation that allows a parsimonious specification: a first-order GARCH model contains only three parameters. However, it does not add any conceptually new insight.

The application in Engle (1982) involved macroeconomic series such as the inflation rate, but Engle quickly realized that the ARCH model was useful in financial economics, as well. Indeed, when considering time series of price changes, Mandelbrot (1963) had already observed that "... large changes tend to be followed by large changes – of either sign – and small changes tend to be followed by small changes ...", but he did not go on to model the returns as time dependent. If the conditional mean is assumed constant, then the ARCH model can be used to characterize return series that contain no linear dependence but do display clustering of volatility. As the ARCH model is also suitable for modeling leptokurtic observations, it can be used to forecast volatility. This, in turn, may be crucial for investors who want to limit the riskiness of their portfolio.

The upper panel of Figure 3.1 shows the daily logarithmic returns (first differences of the logarithms of daily closing prices) of the Standard and Poor 500

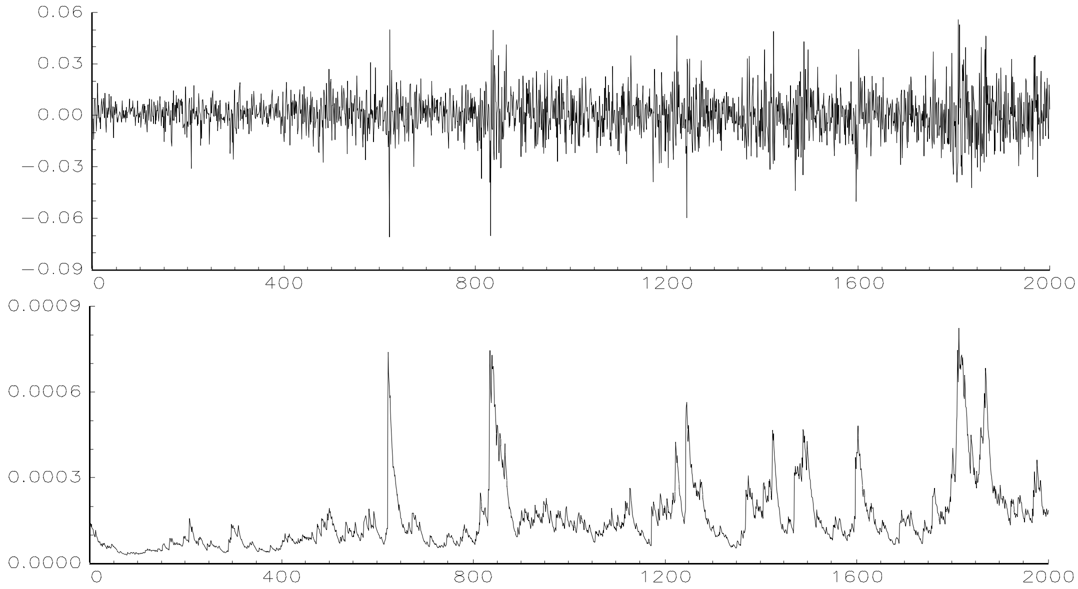

Figure 3.1: Upper panel: Logarithmic daily returns of the Standard and Poor 500 stock index, May 16, 1995 – April 29, 2003. Lower panel: Estimated conditional variance for the same index from the first-order GARCH model (3.4).

stock index from May 16, 1995, to April 29, 2003 – 2000 observations in all. Volatile periods do in fact alternate with periods of relative calm, as in the case of the daily exchange rate series in Figure 1.2.

Fitting a first-order GARCH model to the series in Figure 3.1 under the assumption that the errors z_t are normal (and ε_t thus conditionally normal) yields

$$h_t = 2 \times 10^{-6} + 0.091\varepsilon_{t-1}^2 + 0.899h_{t-1}. \tag{3.4}$$

The sum of $\widehat{\alpha}_1 + \widehat{\beta}_1$ is close to one, which is typical in applications. Condition $\alpha_1 + \beta_1 < 1$ is necessary and sufficient for the first-order GARCH process to be weakly stationary, and the estimated model (3.4) satisfies this condition. The lagged conditional variance h_{t-1} has coefficient estimate 0.9, meaning that 90% of a variance shock remains the next day, and the half-life equals six days. The lower panel of Figure 3.1 shows the estimated conditional variance h_t over time. Spikes in the graph have a relatively slowly decreasing right-hand tail, which shows that volatility is persistent. Another noteworthy observation is that during turbulent periods, conditional variance is several times higher than its basic level. This suggests that the turbulence will have practical implications for investors when forecasting of the volatility of a stock index or a portfolio. For a recent survey on volatility forecasts, see Poon and Granger (2003).

What is parameterized in (3.2) and (3.3) is conditional variance. An alternative proposed by Schwert (1990) is to parameterize the conditional standard deviation $h_t^{1/2}$. This is important in forecasting with GARCH models. If the loss function of the forecaster is based on the mean absolute error instead of the more common root mean square error, it is natural to use a model that parameterizes the conditional standard deviation and not the conditional variance. A more general model would contain both alternatives as special cases. Such a model, called the asymmetric power GARCH model, was introduced by Ding, Granger and Engle (1993). They parameterized a general conditional moment h_t^δ where $\delta > 0$. When Ding et al. (1993) estimated a power-GARCH(1,1) model for a long daily return series of the S&P 500 index, they obtained $\widehat{\delta} = 1.43$, and both null hypotheses, $\delta = 1$ and $\delta = 1/2$, were rejected. The GARCH(1,1) model has nevertheless retained its position as the overwhelmingly most popular GARCH model in practice.

3.3. ARCH-in-mean and multivariate ARCH-in-mean

ARCH and GARCH models are efficient tools for estimating conditional second moments of statistical distributions – i.e., variances and covariances. A great deal of financial theory deals with the connection between the second moments of asset returns and the first moments (expected asset returns). It seemed self-evident to extend the ARCH model to explicitly characterize this connection. Such a model, Engle's first application of ARCH to finance, can be found in Engle, Lilien and Robins (1987). Engle and his co-authors consider a two-asset economy with a risky asset and a risk-free asset. They assume that risk is measured as a function of the conditional variance of the risky asset. As a result, the price offered by risk-averse agents fluctuates over time, and the equilibrium price determines the mean-variance relationship. This suggests including a positive-valued monotonically increasing function of conditional variance in the conditional mean equation. In its simplest form this yields

$$r_t = \beta + g(h_t) + \varepsilon_t, \tag{3.5}$$

where r_t is the excess return of an asset at time t, $g(h_t)$ is a function of the conditional variance h_t, and h_t is defined as in (3.2). Engle et al. (1987) chose $g(h_t) = \delta h_t^{1/2}$, $\delta > 0$, that is, a multiple of the conditional standard deviation of ε_t. Equations (3.5) and (3.2) jointly define an ARCH-in-mean model. The authors applied the model to explain monthly excess returns of the six-month US Treasury bill. Assuming the risk-free asset was a three-month Treasury bill, they found a significant effect from the estimated risk component $\widehat{\delta h_t^{1/2}}$ on the excess return r_t.

In financial theory, however, the price of an asset is not primarily a function of its variance but rather of its covariance with the market portfolio (CAPM) and other non-diversifiable risk factors (Arbitrage Price Theory). To apply the ARCH-in-mean asset pricing model to the pricing of several risky assets thus implied modelling conditional covariances. Instead of the standard CAPM where agents have common and constant expectations of the means and the variances of future returns, this generalization leads to a conditional CAPM, where the expected returns are functions of the time-varying covariances with the market portfolio.

Bollerslev, Engle and Wooldridge (1988) constructed such a multivariate GARCH model. Let $\mathbf{r}_t$ be the $n \times 1$ vector of real excess returns of assets at time t and $\boldsymbol{\omega}_t$ the corresponding vector of value weights. According to the CAPM, the conditional mean vector of the excess returns is proportional to the covariance between the assets and the market portfolio:

$$\boldsymbol{\mu}_t = \delta \mathbf{H}_t \boldsymbol{\omega}_{t-1}, \tag{3.6}$$

where $\mathbf{H}_t = [h_{ijt}]$ is the $n \times n$ conditional covariance matrix, h_{ijt} is the conditional covariance between asset i and asset j at time t and δ is a constant. According to (3.6) , expected returns of assets change over time with variations in the covariance structure. In other words, the so-called β-coefficient in CAPM is time-varying. Matrix $\mathbf{H}_t$ is parameterized in such a way that each conditional variance and covariance has its own equation. As postulated by Bollerslev et al. (1988), this leads to the following multivariate GARCH-in-mean model:

$$\mathbf{r}_t = \boldsymbol{\alpha}_0 + \delta \mathbf{H}_t \boldsymbol{\omega}_{t-1} + \boldsymbol{\varepsilon}_t$$

and[13]

$$\text{vech}(\mathbf{H}_t) = \boldsymbol{\alpha} + \sum_{j=1}^{q} \mathbf{A}_j \text{vech}(\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}'_{t-j}) + \sum_{j=1}^{p} \mathbf{B}_j \text{vech}(\mathbf{H}_{t-j}). \tag{3.7}$$

With three assets ($n = 3$), system (3.7) consists of six equations, for three conditional variances and three conditional covariances. To keep down the number of parameters in (3.7) , Bollerslev et al. (1988) made the simplifying assumptions that $p = q = 1$ and that $\mathbf{A}_1$ and $\mathbf{B}_1$ are diagonal matrices. They then applied the model to quarterly data for three assets: bills (six-month Treasury bill), bonds (twenty-year Treasury bond) and stocks (NYSE index including dividends). The results show, among other things, that the conditional variances and covariances are strongly autoregressive. The hypothesis that the conditional covariance matrix $\mathbf{H}_t$ is constant over time is clearly rejected, which implies that the vector of beta coefficients in the CAPM should be time-varying.

[13] The vech-operator chooses the observations in every column that lie above or on the main diagonal and stacks them on top of each other, beginning with the first column. For instance, let

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Then $\text{vech}(\mathbf{A}) = (a_{11}, a_{12}, a_{22})'$.

In order to mitigate the practical problem inherent in estimating a large number of parameters, Engle (1987) suggested a model that was later applied in Engle, Ng and Rothschild (1990). In this factor-ARCH model the vector of asset returns $\mathbf{r}_t$ has the following definition:

$$\mathbf{r}_t = \mathbf{B}\boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_t, \tag{3.8}$$

where $\mathbf{r}_t$ has a factor structure. In equation (3.8) , $\mathbf{B}$ is an $n \times k$ matrix of factor loadings, that is, parameters to be estimated, $\boldsymbol{\xi}_t$ is a $k \times 1$ vector of unobservable factors, and k is expected to be much smaller than n. This implies that the dynamic behavior of a large number of asset returns is characterized by a small number of common factors. Assume now that the errors $\boldsymbol{\varepsilon}_t$ have a constant conditional covariance matrix $\boldsymbol{\Psi}$ and that the factors have a diagonal conditional covariance matrix $\boldsymbol{\Lambda}_t$. Assuming that $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\xi}_t$ are uncorrelated leads to the following conditional covariance matrix of $\mathbf{r}_t$:

$$\text{cov}(\mathbf{r}_t|\mathcal{F}_{t-1}) = \boldsymbol{\Psi} + \mathbf{B}\boldsymbol{\Lambda}_t\mathbf{B}' = \boldsymbol{\Psi} + \sum_{j=1}^{k} \lambda_{jt}\boldsymbol{\beta}_j\boldsymbol{\beta}_j', \tag{3.9}$$

where $\boldsymbol{\Lambda}_t = \text{diag}(\lambda_{1t}, ..., \lambda_{kt})$ and $\boldsymbol{\beta}_j$ is the jth column of $\mathbf{B}$. If each λ_{jt} is assumed to have an ARCH-type structure, the parameters of this factor-ARCH model can be estimated. Estimation is simplified by assuming that the portfolios (their value weights) are known, since this implies knowing the elements of $\boldsymbol{\beta}_j$. Estimation and implications of the factor-ARCH model are discussed in Engle et al. (1990). They applied a one-factor model ($k = 1$ in (3.9)) to the pricing of Treasury bills of different maturities.

Independently of Engle and his co-authors, Diebold and Nerlove (1989) developed a similar model that was successfully applied to modeling commonality in volatility movements of seven daily exchange rate series. Factor models of this type have also been used to study linkages between international stock markets.

3.4. Other developments

Since its inception, the statistical theory of ARCH models has been extended and applications abound. Hundreds of applications to financial time series had already been listed in a survey by Bollerslev, Chou and Kroner (1992), and the number has continued to grow steadily. Several authors have contributed to the estimation theory for these models and derived conditions for consistency and asymptotic normality of maximum likelihood estimators in both univariate and multivariate ARCH and GARCH models.

Robert Engle himself has enhanced the expanding literature. One of the problems in multivariate GARCH modeling was to ensure that the matrix $\mathbf{H}_t$ of conditional covariances be positive definite for every t. Engle and Kroner (1995) defined a parsimonious GARCH model where this assumption is satisfied that has become popular among practitioners. Engle (2002a) has suggested another multivariate GARCH model with this property, the so-called dynamic conditional correlation GARCH model. Independently of Engle, a similar model was developed by Tse and Tsui (2002). This model is an extension of the constant conditional correlation GARCH model of Bollerslev (1990).

Engle and Ng (1993) devise misspecification tests for GARCH models, which is an important development. They also introduce a new concept, the *news impact curve*. The idea is to condition at time t on the information available at $t-2$ and thus consider the effect of the shock ε_{t-1} on the conditional variance h_t in isolation. Different ARCH and GARCH models can thus be compared by asking how the conditional variance is affected by the latest information, "the news". For example, the news impact curve of the GARCH(1,1) model has the form

$$h_t = A + \alpha_1 \varepsilon_{t-1}^2$$

where $A = \alpha_0 + \beta_1 \sigma^2$ (σ^2 is the unconditional variance of ε_t). This curve is symmetric with respect to $\varepsilon_{t-1} = 0$. Other GARCH models have asymmetric news impact curves; see Engle and Ng (1993) and Ding et al. (1993) for examples and discussion. According to such models, a positive and an equally large negative piece of "news" do not have the same effect on the conditional variance.

Engle's original idea has also spawned different parameterizations. The most commonly applied is the exponential GARCH model of Nelson (1991), where the logarithm of conditional variance has a parametric form. This was the first asymmetric GARCH model. While ordinary GARCH models require parameter restrictions for conditional variance to be positive for every t, such restrictions are not needed in the exponential GARCH model.

Another model that deserves mention in this context is the autoregressive stochastic volatility model. It differs from GARCH models in that the logarithm of the conditional variance is itself a stochastic process. A first-order autoregressive stochastic volatility process, first suggested by Taylor (1982), has the form

$$\ln h_t = \alpha + \beta \ln h_{t-1} + \eta_t$$

where h_t is a positive-valued "conditional variance variable" and $\{\eta_t\}$ is a sequence of independent, identically distributed random variables with mean zero and constant variance. The stochastic volatility model of ε_t has an inherent technical complication. It does not have a closed form because it contains two unobservable random processes: $\{z_t\}$ due to the decomposition $\varepsilon_t = z_t h_t^{1/2}$, and $\{\eta_t\}$. Recently, stochastic volatility models have attracted considerable attention along with the development of effective numerical estimation methods for their parameters; see the surveys by Ghysels, Harvey and Renault (1996) and Shephard (1996).

Building on the theory of ARCH models, Engle recently considered new models for the empirical analysis of market microstructure. The idea is to apply a

GARCH-like model to transaction data on durations between trades, which is feasible because duration is a positive-valued variable in the same way as the squared error ε_t^2 in the ARCH model. In two contributions, Engle and Russell (1998) and Engle (2000), defining the so-called autoregressive conditional duration (ACD), Engle initiated a new literature to clarify the behavior of individual agents in stock markets. These papers have generated a remarkable amount of interest and new papers on ACD models have appeared in rapid succession.

3.5. Application to value at risk

In addition to their use in asset pricing, ARCH and GARCH models have also been applied in other areas of financial economics. The pricing of options and other derivatives, where the variance of the underlying asset is a key parameter, is an obvious area of application; see Noh, Engle and Kane (1994).

ARCH and GARCH models have also become popular and indispensable tools in modern risk management operations. Nowadays banks, other financial institutions and many large companies use so-called value-at-risk analysis. Value-at-risk models are also used to calculate capital requirements for market risks according the so-called Basle II rules; see, for example, Basle Committee on Banking Supervision (1996). To understand the concept, consider an investor with an asset portfolio. The investor wants to predict the expected minimum loss, $L_{\min}$, on this portfolio that will occur at a given, small probability α over the holding period. The predicted value of $L_{\min}$, the value at risk, measures the riskiness of the portfolio. Turning this around, the prediction is that the loss will be no greater than $L_{\min}$ with probability $1-\alpha$. This concept is a natural measure for risk control, for example in cases where a bank regulator wants to ensure that banks have enough capital for the probability of insolvency within, say, the next month not to exceed α.

The attraction of value at risk is that it reduces the market risk associated with a portfolio of assets to an easily understandably number. The loss can be calculated by assuming that the marginal distribution of returns is constant over time, but – in view of the evidence – this does not seem realistic. If the return distribution is time-varying, however, a model is required to predict the future values of the conditional moments characterizing the distribution. If the latter is assumed to be conditionally normal, then the first two moments, the mean and the variance, completely characterize the distribution. GARCH models are widely used for estimating the variance of the conditional return distribution required to calculate the expected loss (their use can be extended to the non-normal case as well). Practitioners often use an exponentially weighted moving average

$$h_t = (1-\beta_1)\varepsilon_{t-1}^2 + \beta_1 h_{t-1},\ 0 < \beta_1 < 1,$$

which is a special case of (3.3) and, more precisely, of the so-called integrated GARCH model introduced by Engle and Bollerslev (1986).

Manganelli and Engle (2001) survey the many approaches to computing the value at risk. Variants of GARCH models have been important components of this development. For instance, Engle and Manganelli (1999) have introduced a so-called conditional autoregressive value at risk model, which is based on the idea of directly modelling the quantile (α) of the distribution that is of interest.

As a simple example of the use of GARCH models in value-at-risk analysis, consider an investor with an S&P 500 index – the series in Figure 3.1 – portfolio of one million dollars. Assume that she applies the estimated GARCH(1,1) model (3.4) with normal errors. The investor wants to estimate the amount below which her loss will remain with probability 0.99 the next day the stock exchange is open if she retains her portfolio. Consider two points in time: September 1, 1995 (Friday), when the conditional variance estimated from (3.4) attains its minimum and July 31, 2002, when it obtains its maximum.[14] The maximum loss predicted from the GARCH model for September 5, 1995 (Tuesday, after Labor Day), equals \$ 12,400, whereas the corresponding sum for August 1, 2002, is \$ 61,500. The difference between the sums illustrates the importance of time-varying volatility and of ARCH as a tool in the value-at-risk analysis.

4. Other contributions

Both Engle and Granger have made valuable contributions in several areas of time-series econometrics. In addition to collaborating closely with Granger to develop tests for cointegration and estimation techniques for models with cointegrated variables, Engle has also done important work on exogeneity, a key concept in econometric modeling (Engle, Hendry and Richard (1983) and Engle and Hendry (1993)). Granger has left his mark in a number of areas. His development of a testable definition of causality (Granger (1969)) has spawned a vast literature. He has also contributed to the theory of so-called long-memory models that have become popular in the econometric literature (Granger and Joyeux (1980)). Furthermore, Granger was among the first to consider the use of spectral analysis (Granger and Hatanaka (1964)) as well as nonlinear models (Granger and Andersen (1978)) in research on economic time series. His contributions to the theory and practice of economic forecasting are also noteworthy. Granger and Morgenstern (1970) is an early classic in this area, while Granger and Bates (1969) may be regarded as having started the vast literature on combining forecasts.

[14] This example is, of course, artificial in the sense that the GARCH model is estimated for a period ending April 29, 2003. In practice, the investor could only use a GARCH model estimated for the observations available at the time the Value at Risk is calculated.

5. Summary and hints for further reading

Since its inception, cointegration has become a vast area of research. A remarkable number of books and articles dealing with theoretical aspects as well as applications have been published. Cointegration has become a standard topic in econometrics textbooks. Engle and Granger (1991) is a collection of key articles, including some of the references in this paper. Books by Banerjee, Dolado, Galbraith and Hendry (1993), Johansen (1995) and Hatanaka (1996) consider the statistical theory underlying cointegration analysis. Watson (1994) is a survey that has a general vector autoregressive model with nonstationary variables as its starting-point. A broad technical overview of the statistical theory of nonstationary processes including cointegration can be found in Tanaka (1996).

Autoregressive conditional heteroskedasticity has, like cointegration, generated an extensive literature, and most time series and econometrics texts nowadays include an exposition of the topic. Engle (1995) is a collection of key papers. Surveys include Bollerslev, Engle and Nelson (1994), Diebold and Lopez (1995), Palm (1996) and Shephard (1996). The book by Gouriéroux (1996) discusses both statistical theory and financial applications. Finally, Engle (2002b) offers a glimpse into the future.

The work of Clive Granger on nonstationary time series and that of Robert Engle on time-varying volatility has had a pervasive influence on applied economic and financial research. Cointegration and ARCH, and the methods the two scholars have developed around these concepts, have indelibly changed the way econometric modeling is carried out.

References

Balke, N. and Fomby, T. B.: 1997, Threshold cointegration, *International Economic Review* **38**, 627–645.

Banerjee, A., Dolado, J., Galbraith, J. W. and Hendry, D. F.: 1993, *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press, Oxford.

Basle Committee on Banking Supervision: 1996, *Amendment to the Capital Accord to Incorporate Market Risks*, Basle.

Bollerslev, T.: 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307–327.

Bollerslev, T.: 1990, Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model, *Review of Economics and Statistics* **72**, 498–505.

Bollerslev, T., Chou, R. Y. and Kroner, K. F.: 1992, ARCH modeling in finance: A review of the theory and empirical evidence, *Journal of Econometrics* **52**, 5–59.

Bollerslev, T., Engle, R. F. and Nelson, D. B.: 1994, ARCH models, *in* R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, North-Holland, Amsterdam, pp. 2959–3038.

Bollerslev, T., Engle, R. F. and Wooldridge, J.: 1988, A capital asset-pricing model with time-varying covariances, *Journal of Political Economy* **96**, 116–131.

Campbell, J. Y.: 1987, Does saving anticipate declining labor income? An alternative test of the permanent income hypothesis, *Econometrica* **55**, 1249–1273.

Campbell, J. Y. and Shiller, R. J.: 1987, Cointegration and tests of present value models, *Journal of Political Economy* **95**, 1062–1088.

Campbell, J. Y. and Shiller, R. J.: 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* **1**, 195–228.

Cassel, G.: 1922, *Money and Foreign Exchange After 1914*, MacMillan, New York.

Cochrane, J. Y.: 1994, Permanent and transitory components of GNP and stock prices, *Quarterly Journal of Economics* **109**, 241–266.

Culver, S. E. and Papell, D. H.: 1999, Long-run purchasing power parity with short-run data: Evidence with a null hypothesis of stationarity, *Journal of International Money and Finance* **18**, 751–768.

Davidson, J. E. H., Hendry, D. F., Srba, F. and Yeo, S.: 1978, Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in United Kingdom, *Economic Journal* **88**, 661–692.

Dickey, D. A. and Fuller, W. A.: 1979, Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* **74**, 427–431.

Dickey, D. A. and Fuller, W. A.: 1981, Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica* **49**, 1057–1072.

Diebold, F. X. and Lopez, J. A.: 1995, Modelling volatility dynamics, *in* K. D. Hoover (ed.), *Macroeconometrics. Developments, Tensions, and Prospects*, Kluwer, Boston, pp. 427–466.

Diebold, F. X. and Nerlove, M.: 1989, The dynamics of exchange rate volatility: A multivariate latent factor ARCH model, *Journal of Applied Econometrics* **4**, 1–21.

Ding, Z., Granger, C. W. J. and Engle, R. F.: 1993, A long memory property of stock market returns and a new model, *Journal of Empirical Finance* **1**, 83–106.

Edison, H.: 1987, Purchasing power parity in the long run: A test of the dollar/pound exchange rate 1890-1978, *Journal of Money, Credit, and Banking* **19**, 376–387.

Engle, R. F.: 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**, 987–1007.

Engle, R. F.: 1987, Multivariate GARCH with factor structures-cointegration in variance, *unpublished paper*, Department of Economics, University of California, San Diego.

Engle, R. F.: 2000, The econometrics of ultra-high-frequency data, *Econometrica* **68**, 1–22.

Engle, R. F.: 2002a, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroscedasticity models, *Journal of Business and Economic Statistics* **20**, 339–350.

Engle, R. F.: 2002b, New frontiers for ARCH models, *Journal of Applied Econometrics* **17**, 425–446.

Engle, R. F. and Bollerslev, T.: 1986, Modeling the persistence of conditional variances, *Econometric Reviews* **5**, 1–50.

Engle, R. F. (ed.): 1995, *ARCH. Selected Readings*, Oxford University Press, Oxford.

Engle, R. F. and Granger, C. W. J.: 1987, Co-integration and error-correction: Representation, estimation and testing, *Econometrica* **55**, 251–276.

Engle, R. F. and Granger, C. W. J. (eds): 1991, *Long-Run Economic Relationships. Readings in Cointegration*, Oxford University Press, Oxford.

Engle, R. F. and Hendry, D. F.: 1993, Testing super exogeneity and invariance, *Journal of Econometrics* **56**, 119–139.

Engle, R. F., Hendry, D. F. and Richard, J.-F.: 1983, Exogeneity, *Econometrica* **51**, 277–304.

Engle, R. F. and Kroner, K. F.: 1995, Multivariate simultaneous generalized ARCH, *Econometric Theory* **11**, 122–150.

Engle, R. F., Lilien, D. M. and Robins, R. P.: 1987, Estimating time-varying risk premia in the term structure: The ARCH-M model, *Econometrica* **55**, 391–407.

Engle, R. F. and Manganelli, S.: 1999, CAViaR: Conditional autoregessive value at risk by regression quantiles, *Working paper 99-20*, Department of Economics, University of California, San Diego.

Engle, R. F. and Ng, V. K.: 1993, Measuring and testing the impact of news on volatility, *Journal of Finance* **48**, 1749–1777.

Engle, R. F., Ng, V. K. and Rothschild, M.: 1990, FACTOR-ARCH covariance structure: Empirical estimates for treasury bills, *Journal of Econometrics* **45**, 213–237.

Engle, R. F. and Russell, J. R.: 1998, Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica* **66**, 1127–1162.

Frankel, J.: 1986, International capital mobility and crowding out in the U.S. economy: Imperfect integration of financial or goods markets?, *in* R. Hafer (ed.), *How Open is the US Economy*, Lexington Books, Lexington.

Frenkel, J.: 1978, Quantifying capital mobility in the 1980s, *in* D. Bernheim and J. Shoven (eds), *National Saving and Economic Performance*, University of Chicago Press, Chicago.

Frenkel, J.: 1981, The collapse of purchasing power parity during the 1970s, *European Economic Review* **16**, 145–165.

Froot, K. A. and Rogoff, K.: 1995, Perspectives on PPP and long-run real exchanger rates, *in* G. Grossman and K. Rogoff (eds), *Handbook of International Economics*, Vol. 3, Elsevier, Amsterdam, pp. 1647–1688.

Fuller, W. A.: 1976, *Introduction to Statistical Time Series*, Wiley, New York.

Ghysels, E., Harvey, A. C. and Renault, E.: 1996, Stochastic volatility, *in* G. S. Maddala and C. R. Rao (eds), *Handbook of Statistics 14: Statistical Methods in Finance*, Elsevier, Amsterdam.

Gouriéroux, C.: 1996, *ARCH Models and Financial Applications*, Springer, Berlin.

Granger, C. W. J.: 1969, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* **37**, 424–438.

Granger, C. W. J.: 1981, Some properties of time series data and their use in econometric model specification, *Journal of Econometrics* **16**, 121–130.

Granger, C. W. J.: 2001, Spurious regressions in econometrics, *in* B. H. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Blackwell, Oxford, pp. 557–561.

Granger, C. W. J. and Andersen, A. P.: 1978, *Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprecht, Göttingen.

Granger, C. W. J. and Bates, J.: 1969, The combination of forecasts, *Operations Research Quarterly* **20**, 451–468.

Granger, C. W. J. and Hatanaka, M.: 1964, *Spectral Analysis of Economic Time Series*, Princeton University Press, Princeton, NJ.

Granger, C. W. J. and Joyeux, R.: 1980, An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis* **1**, 15–30.

Granger, C. W. J. and Lee, T.-H.: 1990, Multicointegration, *in* G. F. Rhodes, Jr and T. B. Fomby (eds), *Advances in Econometrics: Cointegration, Spurious Regressions and Unit Roots*, JAI Press, New York, pp. 17–84.

Granger, C. W. J. and Morgenstern, O.: 1970, *Predictability of Stock Market Prices*, Heath, Lexington, MA.

Granger, C. W. J. and Newbold, P.: 1974, Spurious regressions in econometrics, *Journal of Econometrics* **2**, 111–120.

Granger, C. W. J. and Swanson, N. R.: 1996, Further developments in the study of cointegrated variables, *Oxford Bulletin of Economics and Statistics* **58**, 374–386.

Granger, C. W. J. and Weiss, A. A.: 1983, Time series analysis of error-correction models, *in* S. Karlin, T. Amemiya and L. A. Goodman (eds), *Studies in Econometrics, Time Series and Multivariate Statistics, in Honor of T.W. Anderson*, Academic Press, San Diego, pp. 255–278.

Hall, A. D., Anderson, H. M. and Granger, C. W. J.: 1992, A cointegration analysis of treasury bill yields, *Review of Economics and Statistics* **74**, 116–126.

Hatanaka, M.: 1996, *Time-Series-Based Econometrics. Unit Roots and Cointegrations*, Oxford University Press, Oxford.

Hendry, D. F., Muellbauer, J. N. J. and Murphy, A.: 1990, The econometrics of DHSY, *in* J. D. Hey and D. Winch (eds), *A Century of Economics. 100 Years of the Royal Economic Society and the Economic Journal*, Blackwell, Oxford, pp. 298–334.

Hylleberg, S., Engle, R. F., Granger, C. W. J. and Yoo, B. S.: 1990, Seasonal cointegration, *Journal of Econometrics* **44**, 215–238.

Johansen, S.: 1988, Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* **12**, 231–254.

Johansen, S.: 1991, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* **59**, 1551–1580.

Johansen, S.: 1995, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

Johansen, S. and Juselius, K.: 1990, Maximum likelihood estimation and inference on cointegration - with application to the demand for money, *Oxford Bulletin of Economics and Statistics* **52**, 169–210.

King, R. G., Plosser, C. I., Stock, J. H. and Watson, M. W.: 1991, Stochastic trends and economic fluctuations, *American Economic Review* **81**, 819–840.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, T.: 1992, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic series have a unit root?, *Journal of Econometrics* **54**, 159–178.

Lettau, M. and Ludvigson, S.: 2001, Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* **56**, 815–850.

Lettau, M. and Ludvigson, S.: 2003, Understanding trend and cycle in asset values: Reevaluating the wealth effect on consumption, *Working paper 9848*, National Bureau of Economic Research.

Lo, M. C. and Zivot, E.: 2001, Threshold cointegration and nonlinear adjustment to the law of one price, *Macroeconomic Dynamics* **5**, 533–576.

Mandelbrot, B.: 1963, The variation of certain speculative prices, *Journal of Business* **36**, 394–419.

Mandelbrot, B. and Taylor, H.: 1967, On the distribution of stock price differences, *Operations Research* **15**, 1057–1062.

Manganelli, S. and Engle, R. F.: 2001, Value at risk models in finance, *Working Paper 75*, European Central Bank.

Meese, R. and Rogoff, K.: 1988, Was it real? The exchange rate differential relation over the modern floating exchange rate period, *Journal of Finance* **43**, 933–948.

Nelson, D. B.: 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* **59**, 347–370.

Noh, J., Engle, R. F. and Kane, A.: 1994, Forecasting volatility and option pricing of the SP 500 index, *Journal of Derivatives* **2**, 17–31.

Palm, F. C.: 1996, GARCH models of volatility, *in* G. Maddala and C. Rao (eds), *Handbook of Statistics 14: Statistical Methods in Finance*, Vol. 14, Elsevier, Amsterdam, pp. 209–240.

Phillips, A. W.: 1957, Stabilization policy and the time forms of lagged responses, *Economic Journal* **67**, 265–277.

Phillips, P. C. B.: 1986, Understanding spurious regressions, *Journal of Econometrics* **33**, 311–340.

Phillips, P. C. B. and Perron, P.: 1988, Testing for a unit root in a time series regression, *Biometrika* **75**, 335–346.

Poon, S.-H. and Granger, C. W. J.: 2003, Forecasting volatility in financial markets: A review, *Journal of Economic Literature* **41**, 478–539.

Rachev, S. and Mittnik, S.: 2000, *Stable Paretian Models in Finance*, Wiley, Chichester.

Saikkonen, P. and Luukkonen, R.: 1993, Testing for a moving average unit root in autoregressive integrated moving average models, *Journal of the American Statistical Association* **88**, 596–601.

Sargan, J. D.: 1964, Wages and prices in the United Kingdom: A study in econometric methodology, *in* P. E. Hart, G. Mills and J. N. Whittaker (eds), *Econometric Analysis for National Economic Planning*, Butterworths, London.

Sarno, L. and Taylor, M. P.: 2002, Purchasing power parity and the real exchange rate, *IMF Staff Papers* **49**, 65–105.

Schwert, G. W.: 1990, Stock volatility and the crash of '87, *Review of Financial Studies* **3**, 77–102.

Shephard, N. G.: 1996, Statistical aspects of ARCH and stochastic volatility, *in* D. R. Cox, D. V. Hinkley and O. E. Barndorff-Nielsen (eds), *Time Series Models. In Econometrics, Finance and Other Fields*, Chapman and Hall, London, pp. 1–67.

Shin, Y.: 1994, A residual-based test of the null of cointegration against the alternative of no cointegration, *Econometric Theory* **10**, 91–115.

Sims, C. A.: 1980, Macroeconomics and reality, *Econometrica* **48**, 1–48.

Stock, J. H.: 1987, Asymptotic properties of least squares estimators of cointegrating vectors, *Econometrica* **55**, 1035–1056.

Tanaka, K.: 1996, *Time Series Analysis. Nonstationary and Noninvertible Distribution Theory*, Wiley, New York.

Taylor, S. J.: 1982, Financial returns modeled by the product of two stochastic processes - a study of the daily sugar prices 1961-75, *in* O. D. Anderson (ed.), *Time Series Analysis: Theory and Practice*, North-Holland, Amsterdam, pp. 203–226.

Taylor, S. J.: 1986, *Modelling Financial Time Series*, Wiley, Chichester.

Tse, Y. K. and Tsui, A. K. C.: 2002, A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations, *Journal of Business and Economic Statistics* **20**, 351–362.

Watson, M. W.: 1994, Vector autoregressions and cointegration, *in* R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, North-Holland, Amsterdam, pp. 2844–2915.

Xiao, Z. and Phillips, P. C. B.: 2002, A CUSUM test for cointegration using regression residuals, *Journal of Econometrics* **108**, 43–61.