

TEACHING FINANCIAL ECONOMETRICS TO  
STUDENTS CONVERTING TO FINANCE

By

Stan Hurn, Vance Martin, Peter C.B. Phillips and Jun Yu

May 2024

COWLES FOUNDATION DISCUSSION PAPER NO. 2397



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY  
Box 208281  
New Haven, Connecticut 06520-8281

<http://cowles.yale.edu/>

# Teaching Financial Econometrics to Students Converting to Finance

Stan Hurn, Vance Martin, Peter C.B. Phillips and Jun Yu

## Abstract

Financial econometrics is a dynamic discipline that began to take on its present form around the turn of the century. Since then it has found a permanent position as a popular course sequence in both undergraduate and graduate teaching programs in economics, finance, and business schools. Because of the breadth of the subject's foundations, its extensive coverage in applications and because these courses attract a wide range of students with accompanying interests and skill sets that cover diverse areas and technical capabilities, teaching financial econometrics presents many challenges to the university educator. This chapter addresses some of these challenges, provides helpful guidelines to educators, and draws on the combined experience of the authors as teachers and researchers of modern financial econometrics as well as their recent textbook *Financial Econometric Modeling* (Hurn et al., 2021). The focus is on students converting to finance and econometrics with limited technical background.

*Key words:* University education, econometrics, financial theory, financial data, financial industry, software implementation, teaching econometrics

*JEL classification:* A22, A23, C58

---

Queensland University of Technology, Brisbane, Australia, e-mail: [s.hurn@qut.edu.au](mailto:s.hurn@qut.edu.au)  
University of Melbourne, Melbourne, Australia, e-mail: [vance@unimelb.edu.au](mailto:vance@unimelb.edu.au)  
Yale University; Singapore Management University; University of Auckland, Auckland, New Zealand, e-mail: [peter.phillips@yale.edu](mailto:peter.phillips@yale.edu)  
University of Macau, Macao, China, e-mail: [junyu@um.edu.mo](mailto:junyu@um.edu.mo)

## 1 Introduction

The discipline of financial econometrics is a remarkable success story which, in a short period of time, has given rise to its own professional society, a dedicated journal, and to Nobel Prize awards in Economics. It has a flourishing literature which began with simple models to explain asset prices and their volatility, moved on to tackle complex phenomena such as the pricing of financial derivatives, and proceeded through a rich avenue of applications to study recent problems. The primary concerns of this subject bear most closely on all the major financial and wider economic challenges that societies are now facing. Indeed, as the financial universe grows, new challenges emerge for which models need to be created and econometric methodologies developed. Some of these challenges are posed by phenomena such as ultra-high frequency trading and the multidimensional nature of financial market microstructure; others arise from the requirements of high dimensional financial portfolio analysis, the nature of algorithmic trading, and its wider impact on financial markets and financial stability.

While the reasons to study financial econometrics are compelling, teaching financial econometrics is particularly challenging because the audience has diverse backgrounds. The instructor is faced not only with having to keep all these students interested, but also to motivate them to buy into financial econometrics in a deeper sense by learning how existing methods as well as advanced probabilistic concepts in stochastic process theory combine to advance understanding of financial data. Based on these diverse needs, teachers of financial econometrics must provide, *inter alia*, a skilled exposition of theory in conjunction with computational algorithms, the visual display of quantitative information, constructive estimation and inference in both parametric and nonparametric model environments, and recent improvements in filtering and prediction.

In this chapter we focus on designing a course in financial econometrics primarily aimed at masters level conversion students. Our aim is to demonstrate how teaching a single unifying framework, namely maximum likelihood, enables students to appreciate how a supposedly vast array of disparate concepts, models and techniques are related. This will produce a deeper understanding of the techniques being taught and why, as well as introduce these conversion students to the data, methods, and concerns of this modern subject area where the topics, ideas, theories, and findings relate closely to work that is ongoing in the financial industry

## 2 Course Outline and Motivation

At the outset it is useful to clarify prerequisites in terms of prior knowledge. Students in these conversion classes often have little exposure to any of the primary disciplines of economics, finance or econometrics. Nonetheless, some basic knowledge is presumed, including familiarity with basic statistics, variance and covariance rules, summation notation and simple hypothesis testing as would be taught in a typical first year business statistics course. Students would know the concept of a probability distribution, the bell-shaped curve of the normal distribution and have had an introduction to linear regression. Linear algebra is kept simple as far as possible but vector and matrix notation is needed as well as the concept of a determinant when dealing with the formulation and statistical estimation of multivariate financial models. Finally, students are assumed to have a basic understanding of differential and integral calculus at the secondary school level, a requirement that can be demanding for conversion students coming from a predominantly arts background. Remedial work by way of a short tutorial camp can assist in easing the bridgeway into the mainline course.

Integral to our approach in building and motivating such a bridgeway is the early embrace of modern software packages to embed them into the teaching program as a means of learning and reinforcing key concepts. Commercial computing and econometric packages such as EViews and Stata have the twin advantages of easy use and direct access to powerful tools that help confirm verify results obtained from canned routines.<sup>1</sup> The programming language R is also a useful and economically viable option for students adept at coding, although in our experience teaching econometrics, student appreciation of modeling, estimation and inference is usually enhanced when the coding requirement is minimal and instructor advice reinforces the contextual message and guidance of good software manuals.<sup>2</sup>

Our outlined course focuses on financial prices and returns with dividends paying a minor supporting role. This focus is naturally delimiting because it eliminates discussion of discrete financial variables and various microstructure concepts. But this tighter focus than usual is justified because it is beneficial to students whose

---

<sup>1</sup> EViews, for instance, has an object called LOGL which enables students to write out the log likelihood functions for a particular model; and, similarly, Stata has two powerful maximum likelihood tools -mlexp- and -ml- that allow log likelihood functions to be programmed. Fully exploring the capabilities of the software beyond the point of accepting automated tuning parameters and following the guidance from the manual are, in our experience, powerful learning techniques.

<sup>2</sup> All the numerical results and graphical plots in this chapter are reproducible using a single Stata .do file which is available from the authors on request.

exposure to financial concepts is limited. Indeed many students have never seen the computation of the returns to a financial asset. In our experience, display and discussion of the time series trajectories of prices and returns coupled with an introductory coverage of some of the empirical regularities of financial returns are effective ways of engaging students in the core subject matter right from the outset. Indeed, the relationship between simple time series plots and actual market events is easy to trace out, intelligible to all, and particularly appealing to students with little background in econometrics or finance.

Emphasis on empirical characteristics of returns leads naturally to consideration of the unconditional distribution of returns. This discussion is useful in bringing into play the concept of a probability density and in familiarising a novice audience with the forms of the normal and  $t$  distributions. Many conversion students are familiar with the concept of ‘the bell curve’ but in our experience few have actually seen the analytic form of the density. Early introduction of these two continuous distributions turns out to be an important stepping stone to maximum likelihood estimation and inference.

After introducing the notion of financial returns and elaborating some their empirical properties, a little financial theory comes naturally. Two models are standouts in an introductory course: the constant mean model as a basic expository tool; and the single factor capital asset pricing model (CAPM) as a finance fundamental that introduces a number of key concepts such as the distinction between systemic and idiosyncratic risk. Further, both these models introduce the idea of unknown parameters that must be estimated from observed data.

The existence of unknown parameters leads naturally to issues of estimation. Of the many possible approaches to this subject, maximum likelihood (ML) is especially advantageous in providing a synthesis of specification, estimation and testing within its framework, thereby capturing some of the elements of real world experience in financial econometric modeling. It is this strategy that formed the basis of our recent text *Financial Econometric Modeling* (Hurn et al., 2021). The approach requires knowledge of the necessary ingredients for constructing a likelihood and some familiarity with the rules of probability. ML is naturally intuitive with its goal of selecting parameter values that are ‘most likely’ to have generated the data given a specified model; and basic calculus guides students to use differentiation to identify local maxima of the likelihood function. What students typically do not appreciate is that while solving the likelihood equation to find stationary points of maxima using the first and second derivatives leads to the estimator, shape characteristics

from the second derivative can also be used to estimate the standard error. Choice of estimator and standard error are then together governed by the specification of the likelihood. With this understanding established, the ML approach can be illustrated in the constant mean and CAPM models in detail, providing a framework that covers specification, estimation and inference. These applications clearly reveal that use of ML involves distributional input. But students can correspondingly be made aware of potentially wider generality in terms of quasi-maximum likelihood (QML) estimation and its relevance asymptotically.

The use of maximum likelihood as a unifying theme in financial econometrics is by no means universal. Regression methods are commonly adopted in many financial econometric courses and books because of their simplicity and general applicability in estimating and testing financial models that involve linear specifications. But reliance on regression analysis limits applications according to the nature of the underlying financial models and the data, typically eliminating the consideration of more advanced models; and these methods are easily explained as nested within the general ML framework. Another widely used approach in financial applications is Generalized Methods of Moments (GMM) estimation, whose importance in terms of its utilization of financially relevant moment conditions is highlighted for instance in the asset pricing text of [Cochrane \(2001\)](#). Both GMM and ML involve extremum estimation. ML relies on specific distributional assumptions and GMM relies on relevant moment conditions but not full model specification. Both methods have advantages and disadvantages and discussion of these has continued since the publication of [Hansen \(1982\)](#) promoted GMM methods. In many instances the two approaches can be shown to overlap. GMM was developed under conditions that require stationary and ergodic data and suitable instrumentation to ensure the validity of the moment conditions used in estimation, although overidentifying conditions can be tested using methods developed by [Sargan \(1958\)](#) in the original work on instrumental variables. Stationarity and ergodicity impose strong conditions on typical time series and panel data employed in financial applications; and extending GMM asymptotic theory to allow for the types of stochastically trending nonstationary data that are common in financial data on asset prices, dividends and yields has proved difficult, although some progress has been made in the special case of deterministic trends ([Andrews and McDermott, 1995](#)). The performance of GMM relies on suitability of the instruments and inevitably suffers when instruments are weak or irrelevant. A prime argument for the use of GMM over maximum likelihood is the avoidance of full model specification including distributional choice. This can

be an advantage in practical work. But there are also natural reasons for choosing a range of probability distributions to capture and explain financial processes; and from a teaching perspective the discussion of relevant distributional features can play a useful role in a beginning course as illustrated in [Hurn et al. \(2021\)](#).<sup>3</sup>

Once a basic methodological groundwork is established with two simple model illustrations it is possible to introduce more advanced material. From a financial perspective, modeling the variance of financial returns is potentially more interesting and relevant than the mean in many applications because it provides a crucial input into financial decision making. Examples include portfolio management, the construction of hedge ratios, the pricing of options and the pricing of risk in general. Moreover, while mean returns can be challenging to explain empirically because of the dominance of volatility in the short run, the variance of returns is typically easier to explain using historical data on shocks and other observables that reflect changing risk conditions in financial markets. In implementing various strategies to explore market risk factors empirically, practitioners soon realised that an important characteristic of the return variance is that in most cases it is noticeably time-varying. In consequence, the same is true of the square root of the variance which is commonly known as *volatility*. A traditional approach to modeling time-varying variance is to employ a class of models known as generalised autoregressive conditional heteroskedasticity (GARCH) ([Engle, 1982](#); [Bollerslev, 1986](#)). This is a flexible class of volatility models that can capture a wide range of features that characterise time-varying risk. GARCH models are quintessential financial econometric models that provide a convenient vehicle for the use of maximum likelihood. The multivariate case can also be considered and it is convenient at this stage in class discussion to employ the bivariate normal distribution in setting up the dynamic conditional correlation (DCC) model ([Engle, 2002](#)).

In many applications multiple financial processes must be modeled simultaneously, as when constructing optimal investment portfolios. The advent of vast financial datasets now accentuates this traditional need and raises further issues of how large bodies of information can be processed to improve financial decision making. While not addressing big data requirements in modeling, the final part of an introductory course can at least consider joint modeling of prices and dividends

---

<sup>3</sup> Another framework that avoids making distributional choices or imposing a specific parametric structure on financial models is the nonparametric approach. See, for example, [Pagan and Ullah \(1999\)](#) and [Martin et al. \(2013\)](#). That route may be considered too demanding for most students in an introductory conversion course to financial econometrics, which is the position taken in our own treatment here.

using the basic present value model (Gordon, 1959). Our approach is to start by exploring short-run dynamics of log returns on financial assets and log dividend returns and then to introduce the notion of a long-run equilibrium relationship between the log levels of the data that constrains the dynamics. This development directly leads to cointegration by means of an error correction mechanism (e.g., Davidson et al. (1978)) and can conveniently sidestep more technical issues of unit root testing (Dickey and Fuller, 1979, 1981; Phillips, 1987; Phillips and Perron, 1988) and tests for cointegration (Johansen, 1988; Phillips and Ouliaris, 1990). The treatment can remain firmly within a maximum likelihood framework without a formal discussion of Johansen-type reduced rank vector autoregression (Johansen, 1988, 1991). This approach is intuitive enough to appeal to a non-technical audience and retains the flavor of error correction dynamic modeling.<sup>4</sup> From a modeling viewpoint the move to multivariate distributions can be a difficult jump for students,<sup>5</sup> but is partly mitigated by avoiding unnecessary technicalities and the fact that the students have already been exposed to a multivariate volatility model.

The empirical data used to illustrate the concepts covered by the course is a monthly dataset for the period January 1962 to December 2022. This data is available from the authors on request and includes the following time series:

- (i) From Robert Shiller's data page:<sup>6</sup> the monthly S&P 500 index, dividend payments on the S&P 500 index and 1-month Treasury Bill returns.
- (ii) From Kenneth French's data page:<sup>7</sup> the returns to 5 industry portfolios, namely Consumer (including consumer durables, nondurables, wholesale, retail, and some services), Manufacturing (including manufacturing, energy, and utilities), Tech (including computing, business equipment, telephone and television transmission), Health (including healthcare, medical equipment, and drugs) and Other (including mines, construction, transport, hotels, business services, entertainment and finance).

---

<sup>4</sup> The emphasis on maximum likelihood in a correctly specified system means that problems of bias in least squares approaches are avoided and methods such as fully modified (Phillips and Hansen, 1990; Phillips, 1995) or dynamic least squares estimation (Phillips and Loretan, 1991; Saikkonen, 1991; Stock and Watson, 1993) are not dealt with.

<sup>5</sup> As indeed it was for statisticians in extending the univariate normal density to the multivariate normal case.

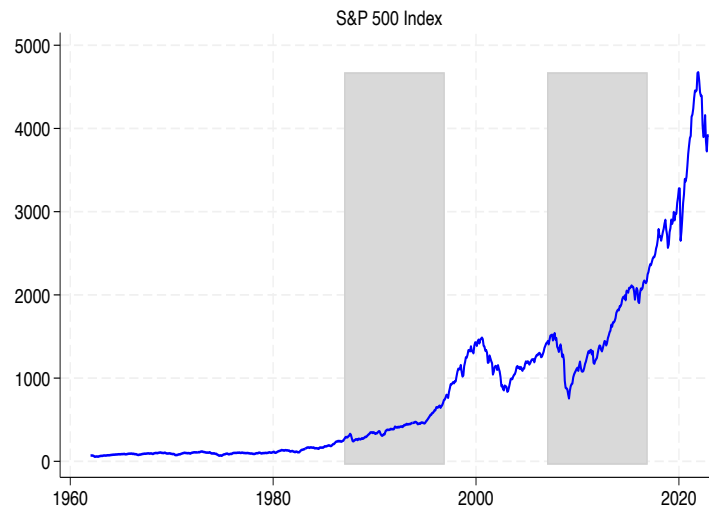
<sup>6</sup> <https://shillerdata.com/>.

<sup>7</sup> [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).



### 3 Financial Asset Prices and Returns

One of the most frequently encountered concept in financial econometrics is the price of a financial asset. The price of an equity security is defined in terms of the dollar (or other currency denomination) amount at which a transaction can occur (a quoted price) or has occurred (an historical transaction price). A major property of financial asset prices is their nonstationarity, which implies that the observed trajectories of prices depend on the time period in which they are examined. This property is highlighted in Figure 1 which plots the monthly S&P 500 Index from January 1962 to December 2022. The time path of equity prices shows long-run growth over this period whose general shape can be roughly captured by an exponential trend. The two shaded areas, January 1987 to December 1996 and January 2007 to December 2016, respectively, demonstrate that the means and variances of the price series in these two periods are very different, indicating time dependence. Although the movement of stock prices over long historical periods is of substantial interest to investors, dealing with nonstationary series poses intriguing and challenging econometric problems which require special techniques that will be touched upon in Section 8.



**Fig. 1** The monthly S&P 500 Index from January 1962 to December 2022. The two shaded areas are from January 1987 to December 1996 and from January 2007 to December 2016.

To capture the trend properties of prices, we can consider filtering the series by taking a log difference, giving

$$r_t = \log P_t - \log P_{t-1}, \quad (1)$$

which generates the log return on the asset. A financial return provides a measure of outcome of the decision to invest in a financial asset. This measure accounts not only for the capital gain or loss due to the price change over the holding period of the asset but also for the cumulative impact of the contractual stream of cash flows that take place over the course of the holding period, although these cash flows in the form of dividends are often ignored.

There are good reasons to work with log returns. First, log returns represent continuously compounded returns. If  $m$  is the compounding period and  $r_t$  the return, then prices evolve according to

$$P_t = P_{t-1} \left(1 + \frac{r_t}{m}\right)^m.$$

Continuous compounding is produced when  $m \rightarrow \infty$ , leading to

$$P_t = P_{t-1} \lim_{m \rightarrow \infty} \left(1 + \frac{r_t}{m}\right)^m. \quad (2)$$

Letting  $s = m/r_t$  this expression may be rewritten as

$$P_t = P_{t-1} \lim_{s \rightarrow \infty} \left[\left(1 + \frac{1}{s}\right)^{s r_t}\right] = P_{t-1} \left[\lim_{s \rightarrow \infty} \left(1 + \frac{1}{s}\right)^s\right]^{r_t} = P_{t-1} e^{r_t}. \quad (3)$$

Taking logarithms of expression (3) and rearranging yields the definition of the log returns given in equation (1).

Second, log returns are particularly useful because of the simplification they allow in dealing with multi-period returns. For example, the 2-period return is given by

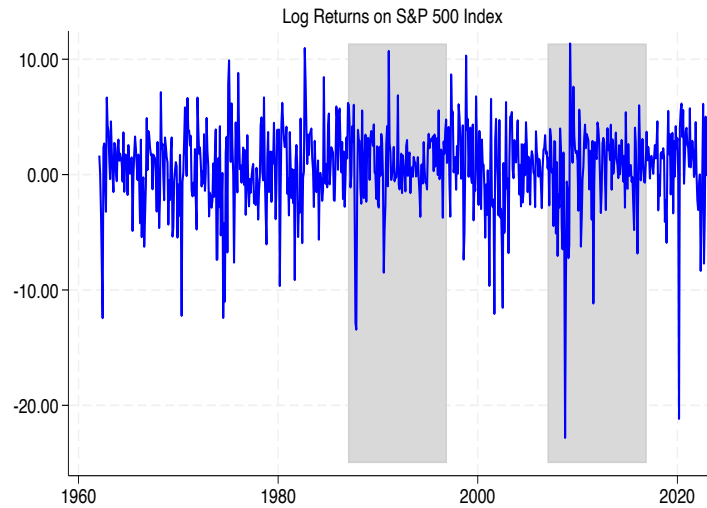
$$r_t(2) = \log P_t - \log P_{t-2} = (\log P_t - \log P_{t-1}) + (\log P_{t-1} - \log P_{t-2}) = r_t + r_{t-1}.$$

By extension, the  $k$ -period return is

$$r_t(k) = r_t + r_{t-1} + \cdots + r_{t-(k-1)} = \sum_{j=0}^{k-1} r_{t-j}.$$

Third, and perhaps most significant from an econometric perspective, the log returns series obtained by filtering prices in this way yields a stationary series. because the behaviour of returns in the two shaded sub-samples in Figure 2 is very

similar. This result is demonstrated in Figure 2 which shows monthly log returns of the S&P500 expressed in percentage terms. The behaviour in the time series in the two shaded windows is now very similar, unlike the situation in Figure 1. This observation is reinforced by Table 1 which reports the summary statistics for the percentage monthly log returns of the S&P500 for the entire sample period and also for the two highlighted sub-periods. It can be seen all the sample statistics are very similar.



**Fig. 2** Percentage monthly log returns on the S&P 500 Index from January 1962 to December 2022. The two shaded areas are from January 1987 to December 1996 and from January 2007 to December 2016, respectively.

**Table 1** Summary statistics for percentage monthly log returns on S&P 500 for the periods January 1962 to December 2022, January 1987 to December 1996 and January 2007 to December 2016.

	1962 - 2022	1987 - 1996	2007- 2016
Mean	0.552	0.913	0.384
Median	0.906	0.939	1.059
Variance	13.245	10.169	16.540
Standard Deviation	3.639	3.189	4.067
Skewness	-1.226	-1.298	-1.865
Kurtosis	8.018	8.575	11.139
Minimum	-22.804	-13.425	-22.804
Maximum	11.352	10.703	11.352

The largest negative monthly returns in the two sub-samples are, respectively, -13.425% (October 1987) and -22.804% (October 2008). These negative returns are associated with important stock market events. Monday 19 October 1987 is known as Black Monday, a day on which the U.S. stock market experienced a large one day decline that triggered a global stock market slump. The stock market crash of October 2008 arose from defaults on consolidated mortgage-backed securities and led to what is now known as the Global Financial Crisis. The occurrence of large negative monthly returns leads naturally to the question of the distributional characteristics of returns. Investors in financial assets are interested in the mean and standard deviation (or riskiness) of the investment, but also in the shape of the returns distribution which is related to the skewness and kurtosis statistics given in Table 1.

#### 4 Distributional Characteristics of Returns

The most commonly used distribution in statistics is the normal distribution in which the probability density of returns has the familiar shape of a symmetric bell-shaped curve about the mean. The probability density function is given by<sup>8</sup>

$$f(r_t; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(r_t - \mu)^2}{2\sigma^2}\right) \quad \text{or} \quad r_t \sim N(\mu, \sigma^2),$$

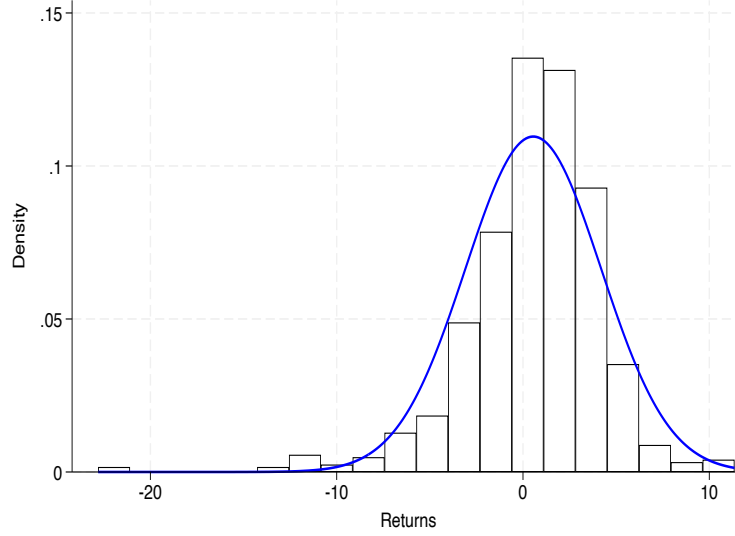
in which  $N(\mu, \sigma^2)$  signifies a normal distribution with mean  $\mu$  and variance  $\sigma^2$ . The normal distribution is symmetric and therefore has zero skewness. The kurtosis of the normal distribution which refers primarily to the tail behaviour<sup>9</sup> of the distribution has the value 3.

Figure 3 plots the histogram of the percentage monthly return of S&P500 for the full sample period from January 1962 to December 2022 with the best fitting

<sup>8</sup> In the spirit of this chapter, we keep the notation as simple as possible in recognition of the technical background of the target audience. In more theoretical presentation the “ $\sim$ ” symbol is sometimes written as “ $\sim_d$ ” or as “ $\xrightarrow{d}$ ” to emphasise convergence in distribution.

<sup>9</sup> The word kurtosis comes from the Greek root ‘kurtos’ meaning curved or arching. Correspondingly, kurtosis is often associated with the apparent peakedness of a distribution. This interpretation, including the nomenclature, was introduced by Pearson (1905) but has been recently challenged (Westfall, 2014). Since kurtosis is defined in terms of the scaled fourth moment of a distribution a primary effect measured by kurtosis is the extent of outliers in the distribution arising from heaviness in the tails. But as more probability is drawn into the tails, the net effect is less probability elsewhere, although this may not always be at the center of the distribution. The normal distribution has kurtosis with value 3 and is said to have zero excess kurtosis. Distributions with higher kurtosis than 3 have positive excess kurtosis.

normal distribution superimposed.<sup>10</sup> It is immediately apparent that the histogram demonstrates what Table 1 indicates, namely that the returns distribution exhibits a sharper peak and fatter tails (particularly in the left tail consistent with the reported negative skewness) than the best-fitting normal distribution, which is overlaid on the histogram of the returns. Distributions of returns exhibiting these characteristics are referred to as leptokurtic.



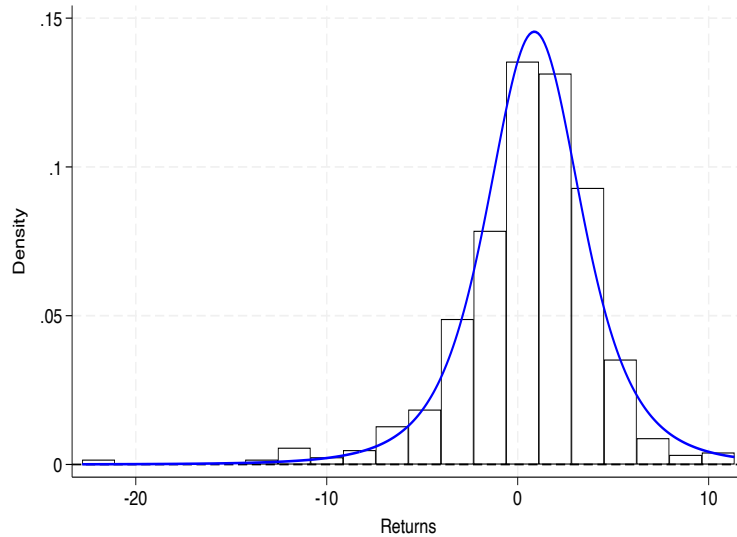
**Fig. 3** The histogram of percentage monthly log returns on the S&P 500 from January 1962 to December 2022 with the best fitting normal distribution superimposed.

To capture potential non-normal asset returns the normal distribution assumption is often replaced by a Student  $t$  distribution whose probability density is given by

$$f(r_t; \mu, \sigma^2, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\sigma^2\nu} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(r_t - \mu)^2}{\sigma^2\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}, \quad (4)$$

<sup>10</sup> The best fitting normal distribution is based on the maximum likelihood estimates of  $\mu$  and  $\sigma^2$  which are dealt with in Section 6.

where  $\nu$  represents the degrees of freedom parameter and  $\Gamma(\cdot)$  is the gamma function<sup>11</sup>. The parameter  $\nu$  provides additional flexibility compared with the normal distribution in modeling empirical distributions. In the special case where  $\nu \rightarrow \infty$ , the  $t$  distribution becomes the normal distribution.



**Fig. 4** The histogram of percentage monthly log returns on the S&P 500 from January 1962 to December 2022 with the best fitting Student  $t$  distribution superimposed.

The value of the parameter  $\nu$  in the Student  $t$  distribution determines the number of integer moments of the distribution that exist. Specifically, integer moments lower than  $\nu$  exist. Thus, the mean or first moment of the  $t$  distribution exists provided  $\nu > 1$ ; the second moment exists if  $\nu > 2$  and in that case the variance is given by  $\sigma^2\nu/(\nu - 2)$ ; the first three moments exist if  $\nu > 3$ ; and for  $\nu > 4$  the first four moments all exist. For the special case where  $\nu = 1$ , no integer moments exist (moments of fractional order less than unity do exist) and the density is known as the Cauchy distribution. The form of the  $t$  distribution given in equation (4) is a version of the distribution in which  $\mu$  characterises central location (the mean provided that  $\nu > 1$ ) and  $\sigma^2$  characterises dispersion (the variance provided that  $\nu > 2$ ).

<sup>11</sup> The gamma function is a generalisation of the factorial operator that accommodates real and complex number arguments, not just integers. The form of this function is not important for current purposes and its values are readily computed using available software.

The plot in Figure 4 illustrates how the Student  $t$  distribution does a slightly better job of capturing the distributional features of returns than does the normal distribution shown in Figure 3, particularly in terms of the peakedness of the distribution (with the higher and somewhat sharper peak than the normal density fit) and the length of the left tail. In this particular case the degrees of freedom parameter  $\nu$  is 3.89, indicating that the first three moments of the distribution exist.

## 5 Models of Returns

Having discussed some empirical facts about financial asset returns, it is now appropriate to introduce some theoretical models which purport to model these returns. The simplest model of financial returns is the constant mean and constant variance model which is given by

$$r_t = \alpha + u_t, \quad u_t \sim N(0, \sigma^2), \quad (5)$$

where  $r_t$  is returns and  $u_t$  is a normally distributed idiosyncratic disturbance term, with zero mean and variance  $\sigma^2$ , that captures unexplained deviations of returns from their mean. The expected return is given by  $\alpha$  and the risk of the asset by  $\sigma$ .

The constant mean model is of little predictive use as it simply summarises the returns distribution. The most long standing model of financial asset returns is the CAPM that encapsulates the risk characteristics of an asset in terms of its so-called  $\beta$ -risk, a quantity that is measured by the ratio

$$\beta = \frac{\text{cov}(r_t - r_{ft}, r_{mt} - r_{ft})}{\text{var}(r_{mt} - r_{ft})}. \quad (6)$$

The  $\beta$ -risk is a measure of the exposure of the returns  $r_t$  on a particular financial asset to movements in market returns,  $r_{mt}$ , relative to a risk-free rate of interest  $r_{ft}$ . Individual stocks, or even the portfolios of stocks, are classified as follows in terms of their degree of  $\beta$ -risk: if  $\beta > 1$  the asset is aggressive in the sense that its returns magnifies movements in the market; if  $0 < \beta < 1$  the asset is defensive and serves to attenuate movements in the market; and if  $\beta = 1$  the asset is a benchmark for the market. Assets with  $\beta = 0$  are uncorrelated with the market, while those with  $\beta = -1$  provide a perfect hedge for movements in the market.

The CAPM is formulated by expressing the relationship between the excess return on the asset  $r_t - r_{ft}$  and the market  $r_{mt} - r_{ft}$  in terms of the following linear equation

$$r_t - r_{ft} = \alpha + \beta(r_{mt} - r_{ft}) + u_t, \quad u_t \sim N(0, \sigma^2), \quad (7)$$

where the excess return on the asset represents the dependent variable and the excess return on the market represents the explanatory variable. The disturbance term  $u_t$  captures additional (idiosyncratic) movements in the dependent variable not predicted by CAPM. Assuming that these additional movements are also independent of  $r_{mt} - r_{ft}$ , then  $E[(r_{mt} - r_{ft})u_t] = 0$ . The model in equation (7) contains three unknown parameters. The first is the intercept parameter  $\alpha$ , which captures the abnormal return to the asset over and above the asset's exposure to the excess return on the market. The second is the slope parameter  $\beta$ , which corresponds to the asset's  $\beta$ -risk as defined in (6). The third parameter is  $\sigma^2$  which captures the idiosyncratic risk of the asset.

The CAPM provides a convenient method of decomposing the total risk of an asset into systematic risk (or risk that is inherent to the entire market) from exposure to movements in the market, and idiosyncratic risk caused by other factors. Formally, this decomposition is achieved by squaring both sides of (7) and then taking expectations, which gives

$$\underbrace{E[(r_t - r_{ft})^2]}_{\text{Total risk}} = \underbrace{E[(\alpha + \beta(r_{mt} - r_{ft}))^2]}_{\text{Systematic risk}} + \underbrace{E(u_t^2)}_{\text{Idiosyncratic risk}},$$

using the zero correlation property  $E[(r_{mt} - r_{ft})u_t] = 0$ . The systematic risk is also known as nondiversifiable risk while the idiosyncratic risk,  $E(u_t^2) = \sigma^2$ , represents the diversifiable risk.

## 6 Maximum Likelihood Estimation

The maximum likelihood estimator of the generic unknown parameter vector  $\theta$  is denoted  $\hat{\theta}$  and is defined to be the parameter value that maximises the logarithm of the probability distribution of the data, which is typically known as the log likelihood function. Put differently, the maximum likelihood estimator of the elements of the unknown parameter vector  $\theta$  is obtained by finding the value of  $\theta$  that is *most likely* to have generated the observed data. Let  $\log L(\theta)$  denote the log likelihood function which, by definition is conditional on the data. In some cases, a closed-form expression for  $\hat{\theta}$  can be obtained. For many interesting cases, numerical methods are required to perform the optimization. The necessary ingredients for constructing a



likelihood require the use of certain rules of probability and rules for summation to manipulate distributional expressions, together with some knowledge of the forms of various distributions.<sup>12</sup>

## 6.1 Specification

Having specified the probability distribution  $f(r_t; \theta)$  of a financial process  $r_t$  at each observation  $t$ , the probabilities associated with each event across the full time series sample  $\{t = 1, 2, \dots, T\}$ , are combined to generate the *average* log likelihood function. It is worth emphasising that the term log likelihood function will be used subsequently without any reference to it being an average.<sup>13</sup> This convention is used, by among others, [Newey and McFadden \(1994\)](#) and [White \(1994\)](#). See also [Martin et al. \(2013\)](#).

There exist four types of log likelihood functions:

$$\log L(\theta) = \begin{cases} \frac{1}{T} \sum_{t=1}^T \log f(r_t; \theta) & \text{[Case 1]} \\ \frac{1}{T} \sum_{t=1}^T \log f(r_t | x_t; \theta) & \text{[Case 2]} \\ \frac{1}{T} \sum_{t=1}^T \log f(r_t | r_{t-1}, \dots, r_{t-K}; \theta) & \text{[Case 3]} \\ \frac{1}{T} \sum_{t=1}^T \log f(r_t | x_t, r_{t-1}, \dots, r_{t-K}; \theta). & \text{[Case 4]} \end{cases}$$

Case 1 represents the simplest case in which the distribution is the same at every observation. In Case 2 the distribution of the data varies over the sample according to changes in a conditioning variable  $x_t$ . Case 3 differs from Case 2 in the sense that the distribution of the data changes over the sample driven by conditioning on past values of  $r_t$ . In the most general case, Case 4, variations in the probability distributions arise from conditioning on both  $x_t$  and past values  $r_{t-1}, \dots, r_{t-K}$ . The specification here assumes that observations at  $r_0, r_{-1}, \dots$  are available. Examples of Case 1 and Case 2 are given in this section and examples of Cases 3 and 4 are given in Sections 7 and 8, respectively.

<sup>12</sup> To facilitate exposition many steps in the derivations are omitted in what follows. Interested readers are referred to [Hurn et al. \(2021\)](#) for the intermediate details.

<sup>13</sup> The definition of the log likelihood function as an average over  $T$  observations is consistent with the notion of convergence to the true average value as  $T \rightarrow \infty$ .

### Constant mean model

Consider the constant mean model of returns in equation (5), as  $u_t$  is assumed to be normally distributed it follows that  $r_t$  is also normally distributed. The distribution of  $r_t$  is therefore given by

$$f(r_t; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(r_t - \alpha)^2}{2\sigma^2}\right),$$

in which the sample data,  $r_t$ , are independent drawings from a normal distribution with mean  $\alpha$  and variance  $\sigma^2$ . The log likelihood function follows Case 1 and has the form

$$\log L(\theta) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \frac{1}{T} \sum_{t=1}^T (r_t - \alpha)^2,$$

with parameter vector  $\theta = \{\alpha, \sigma^2\}$ .

### CAPM

The CAPM given in equation (7) may be rewritten as

$$y_t = \alpha + \beta x_t + u_t, \quad u_t \sim N(0, \sigma^2), \quad (8)$$

where for notational simplicity  $y_t = r_t - r_{ft}$  is the excess return on the asset and  $x_t = r_{mt} - r_{ft}$  is the excess return on the market. If it is assumed that  $u_t$  once more has a normal distribution with mean 0 and variance  $\sigma^2$ , then it follows that

$$u_t \sim N(0, \sigma^2) \Rightarrow (y_t - \alpha - \beta x_t) \sim N(0, \sigma^2) \Rightarrow y_t \sim N(\alpha + \beta x_t, \sigma^2).$$

Taking  $r_{ft}$  and  $r_{mt}$  as given, the distribution of  $y_t$  is

$$f(y_t; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - \alpha - \beta x_t)^2}{2\sigma^2}\right).$$

The log likelihood function is constructed based on Case 2 and is given by

$$\log L(\theta) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \frac{1}{T} \sum_{t=1}^T (y_t - \alpha - \beta x_t)^2,$$

with  $\theta = \{\alpha, \beta, \sigma^2\}$ .

## 6.2 Estimation

The first derivative of the log likelihood function is known as the gradient  $G(\theta)$  and is defined as

$$G(\theta) = \frac{\partial \log L(\theta)}{\partial \theta}.$$

The second derivative of the log likelihood function is known as the Hessian and is given by the matrix

$$H(\theta) = \frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'}.$$

From the basic rules of calculus, the maximum likelihood estimator  $\hat{\theta}$  is found by solving the likelihood equation

$$G(\hat{\theta}) = 0. \quad (9)$$

The Hessian is integral to estimating a measure of the precision of the maximum likelihood estimator<sup>14</sup> given by the standard error  $\text{se}(\hat{\theta})$ . This property is based on the result that

$$\text{var}(\hat{\theta}) = \frac{1}{T} \left( -H(\hat{\theta})^{-1} \right).$$

It follows therefore that the standard errors of  $\hat{\theta}$  are given by the square roots of the diagonal elements of this matrix.

### Constant mean model

The  $(2 \times 1)$  gradient vector of the constant mean model is given by

$$G(\theta) = \begin{pmatrix} \frac{\partial \log L(\theta)}{\partial \alpha} \\ \frac{\partial \log L(\theta)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2} \frac{1}{T} \sum_{t=1}^T (y_t - \alpha) \\ -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} \frac{1}{T} \sum_{t=1}^T (y_t - \alpha)^2 \end{pmatrix}.$$

The maximum likelihood estimator is found by setting  $G(\hat{\theta}) = 0$  which requires solving the following two equations

---

<sup>14</sup> The Hessian is also used to determine whether  $\hat{\theta}$  does provide a local maximum of  $\log L(\theta)$ . This feature of  $H(\hat{\theta})$  follows from standard calculus rules for determining the shape of a curve at a stationary point. Issues of multiple local maxima versus the absolute maximum of the likelihood function should be mentioned but are not dealt with here. Iterative methods to find extrema in optimization problems are also relevant and are briefly discussed below in Section 6.3.

$$\begin{aligned}\frac{1}{\widehat{\sigma}^2} \frac{1}{T} \sum_{t=1}^T (y_t - \widehat{\alpha}) &= 0 \\ -\frac{1}{2\widehat{\sigma}^2} + \frac{1}{2\widehat{\sigma}^4} \frac{1}{T} \sum_{t=1}^T (y_t - \widehat{\alpha})^2 &= 0,\end{aligned}$$

for  $\widehat{\alpha}$  and  $\widehat{\sigma}^2$ . Solving these expressions yields the maximum likelihood estimators

$$\begin{aligned}\widehat{\alpha} &= \frac{1}{T} \sum_{t=1}^T y_t = \bar{y} \\ \widehat{\sigma}^2 &= \frac{1}{T} \sum_{t=1}^T (y_t - \widehat{\alpha})^2.\end{aligned}$$

The sample mean is therefore the maximum likelihood estimator of  $\alpha$  and the sample variance of  $y_t$ , without any degrees of freedom correction, is the maximum likelihood estimator of  $\sigma^2$ . In other words, the maximum likelihood estimate of the variance uses the full sample size  $T$  in the denominator.<sup>15</sup>

## CAPM

The  $(3 \times 1)$  gradient vector for the CAPM model is given by

$$G(\theta) = \begin{pmatrix} \frac{\partial \log L(\theta)}{\partial \alpha} \\ \frac{\partial \log L(\theta)}{\partial \beta} \\ \frac{\partial \log L(\theta)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2} \frac{1}{T} \sum_{t=1}^T (y_t - \alpha - \beta x_t) \\ \frac{1}{\sigma^2} \frac{1}{T} \sum_{t=1}^T (y_t - \alpha - \beta x_t) x_t \\ -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} \frac{1}{T} \sum_{t=1}^T (y_t - \alpha - \beta x_t)^2 \end{pmatrix}.$$

Setting  $G(\widehat{\theta}) = 0$  yields a linear system of three equations and three unknowns with the following solutions<sup>16</sup>

$$\widehat{\alpha} = \bar{y} - \widehat{\beta} \bar{x}$$

<sup>15</sup> These are the estimators that were used in the construction of the normal approximation to the histogram in Figure 3.

<sup>16</sup> The expressions for  $\widehat{\alpha}$  and  $\widehat{\beta}$  are in fact the ordinary least squares estimators, demonstrating that for this class of models the ordinary least squares estimator is equivalent to the maximum likelihood estimator, at least under the assumption of normality. The expression for  $\widehat{\sigma}^2$  shows that the maximum likelihood estimator is equivalent to the ordinary least squares estimator apart from the degrees of freedom correction.

$$\hat{\beta} = \frac{\sum_{t=1}^T (y_t - \bar{y})(x_t - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}$$

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\alpha} - \hat{\beta}x_t)^2.$$

As an example, consider estimating the CAPM for a portfolio of high technology stocks as compiled by Kenneth French and published on his website. The maximum likelihood estimate of  $\theta = \{\alpha, \beta, \sigma\}$  is

$$\hat{\theta} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\sigma} \end{pmatrix} = \begin{pmatrix} 0.382 \\ 0.905 \\ 4.3475 \end{pmatrix},$$

and the value of the log likelihood function at the maximum is  $-2111.5$ . The gradient vector at the optimum is

$$G(\hat{\theta}) = \begin{pmatrix} \frac{\partial \log L(\theta)}{\partial \alpha} \\ \frac{\partial \log L(\theta)}{\partial \beta} \\ \frac{\partial \log L(\theta)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} 3.218 \times 10^{-7} \\ 9.780 \times 10^{-6} \\ 9.838 \times 10^{-6} \end{pmatrix},$$

indicative that the condition  $G(\hat{\theta}) = 0$  is satisfied to a very close approximation. The covariance matrix of  $\hat{\theta}$  is the symmetric matrix

$$\text{var}(\hat{\theta}) = \begin{pmatrix} 0.0259 & \cdot & \cdot \\ -0.0004 & 0.0019 & \cdot \\ 0.0000 & 0.0000 & 0.0129 \end{pmatrix},$$

and the standard errors of the elements of  $\hat{\theta}$  are given by the square roots of the diagonal elements of this matrix

$$\text{se}(\hat{\theta}) = \begin{pmatrix} \text{se}(\hat{\alpha}) \\ \text{se}(\hat{\beta}) \\ \text{se}(\hat{\sigma}) \end{pmatrix} = \begin{pmatrix} 0.161 \\ 0.044 \\ 0.114 \end{pmatrix}.$$

Estimates of the CAPM for all 5 industry portfolios are presented in the top panel of Table 2. Three of the portfolios, consumer, manufacturing and health, are very defensive portfolios with their  $\beta$ -risk being significantly less than 1. Unsurprisingly the portfolios with the highest  $\beta$ -risk are the technology stocks and the catch-all

portfolio, other. Even these, however, have  $\beta < 1$ . All the portfolios have positive  $\alpha$ 's which means that they earn a return above the risk-free rate of interest even when their exposure to systematic risk is taken into account.

### 6.3 Numerical Optimization: The Robust CAPM

Both the estimation problems considered so far have analytical solutions. Typically numerical methods based on iterative gradient algorithms are employed to maximise the log likelihood function when analytic formulae are not available. Econometric software usually provides for a number of alternative algorithms. The Newton-Raphson (NR) algorithm uses both the first (gradient) and second derivatives (Hessian) of the log likelihood and despite being one of the oldest iterative algorithms available is nonetheless still widely used. Letting  $\theta_{(k)}$  be the value of  $\theta$  at iteration  $k$ , and  $\theta_{(k+1)}$  the updated value, the Newton-Raphson algorithm is given by

$$\theta_{(k+1)} = \theta_{(k)} - H^{-1}(\theta_{(k)})G(\theta_{(k)}).$$

The algorithm begins with a starting value  $\theta_{(0)}$ , with the iterations continuing until there is no change in the estimates of  $\theta$ , that is  $\theta_{(k+1)} \simeq \theta_{(k)}$ .

An example of the use of numerical methods to compute the maximum likelihood estimator is provided by the robust CAPM with disturbances now following a Student  $t$  distribution. The heavy-tailed  $t$  distribution helps to accommodate the presence of outliers in the data. Unfortunately, closed-form expressions for the maximum likelihood estimators of the parameters of the model are no longer available in this case. This model is

$$y_t = \alpha + \beta x_t + u_t, \quad u_t \sim St(0, \sigma^2, \nu),$$

where, as in the simple CAPM,  $y_t = r_t - r_{ft}$  is the excess return on the asset and  $x_t = r_{mt} - r_{ft}$  is the excess return on the market. The notation  $St(0, \sigma^2, \nu)$  represents a standardised form of  $t$  distribution in which the density is given by

$$f(u_t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\sigma^2(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{u_t^2}{\sigma^2(\nu-2)}\right)^{-\left(\frac{\nu+1}{2}\right)}, \quad (10)$$

where the degrees of freedom parameter  $\nu$  is assumed to satisfy  $\nu > 2$ , a condition that ensures the variance is finite as remarked earlier. In fact, the variance of the standardised distribution  $St(0, \sigma^2, \nu)$  is the squared scale coefficient  $\sigma^2$ , which is convenient in many applications. This form of the  $t$  distribution differs slightly from that given earlier in equation (4) because it allows  $u_t$  to have dispersion measured directly by the scale parameter  $\sigma$ . As before in (4), the parameter  $\nu$  determines how heavy the tails of the distribution are, thereby capturing the effects of outliers.

Transforming to the implied distribution of the observations  $y_t$ , gives the density

$$f(y_t | x_t; \theta) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\sigma^2(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(y_t - \alpha - \beta x_t)^2}{\sigma^2(\nu-2)}\right)^{-\left(\frac{\nu+1}{2}\right)}. \quad (11)$$

To derive the maximum likelihood estimator of  $\theta = \{\alpha, \beta, \sigma^2, \nu\}$ , the likelihood function  $\log L(\theta)$  is based on Case 2 and has the form

$$\begin{aligned} \log L(\theta) &= \frac{1}{T} \sum_{t=1}^T \log f(y_t | x_t; \theta) \\ &= \log \Gamma\left(\frac{\nu+1}{2}\right) - \frac{1}{2} \log \sigma^2 - \frac{1}{2} \log(\pi(\nu-2)) \\ &\quad - \log \Gamma\left(\frac{\nu}{2}\right) - \left(\frac{\nu+1}{2}\right) \frac{1}{T} \sum_{t=1}^T \log \left(1 + \frac{(y_t - \alpha - \beta x_t)^2}{\sigma^2(\nu-2)}\right). \end{aligned} \quad (12)$$

Since this expression for the log likelihood function is nonlinear in the parameters, a numerical approach is adopted to compute the maximum likelihood estimator,  $\hat{\theta}$ .

The results of estimating this robust version of the CAPM by maximum likelihood using a numerical optimisation algorithm are given in the bottom panel of Table 2. The estimate of the degrees of freedom parameter,  $\hat{\nu}$ , implies the presence of significant outliers in the returns series. On the other hand, the fact that the estimates of  $\alpha$ -risk and  $\beta$ -risk are qualitatively similar to those obtained using the assumption of normal disturbances implies that the estimates are robust to these outliers.

## 6.4 Testing

The likelihood function provides a simple and intuitive approach to inference, the most common form of which is to test whether the parameter of the model has a certain hypothesised value,  $\theta_0$  say, resulting in the null and alternative hypotheses

**Table 2** Parameter estimates of the CAPM model based on the normal distribution (top panel) and the  $t$  distribution (bottom panel) using monthly excess log returns on five United States industry portfolios for the period January 1962 to December 2022. Standard errors are in parentheses.

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\nu}$	$\log L$
Consumer	0.465 (0.133)	0.792 (0.036)	3.592 (0.094)		-1971.9
Manufacturing	0.415 (0.128)	0.758 (0.035)	3.452 (0.090)		-1942.9
Tech	0.382 (0.161)	0.905 (0.044)	4.347 (0.114)		-2111.5
Health	0.545 (0.157)	0.667 (0.043)	4.251 (0.111)		-2095.1
Other	0.392 (0.153)	0.926 (0.042)	4.128 (0.108)		-2073.7
Consumer goods	0.395 (0.128)	0.755 (0.036)	-3.600 (0.127)	7.408 (1.923)	-1960.3
Manufacturing	0.339 (0.121)	0.754 (0.035)	3.471 (0.132)	6.452 (1.529)	-1928.5
Tech	0.358 (0.146)	0.903 (0.042)	4.364 (0.193)	5.084 (0.938)	-2081.4
Health	0.506 (0.150)	0.662 (0.044)	4.230 (0.146)	7.798 (2.053)	-2079.6
Other	0.363 (0.143)	0.909 (0.040)	4.153 (0.163)	6.058 (1.358)	-2056.7

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0.$$

Let  $\hat{\theta}_0$  and  $\hat{\theta}_1$  represent the estimates of  $\theta$  under the null and alternative hypotheses, respectively. There exist three testing strategies within the maximum likelihood framework: the Likelihood Ratio test (LR), the Wald test (W), and the Lagrange Multiplier test (LM). All the test statistics are a function of the distance between the parameters under the null and alternative hypotheses. An important property of tests based on the likelihood is that they encompass many of the test statistics commonly used in financial econometrics.

### Likelihood ratio test

The LR statistic measures the distance between  $\hat{\theta}_0$  and  $\hat{\theta}_1$  in terms of their log likelihood values, that is,  $\log L(\hat{\theta}_0)$  and  $\log L(\hat{\theta}_1)$ . The test is given by

$$LR = -2T(\log L(\hat{\theta}_0) - \log L(\hat{\theta}_1)).$$



A small difference between the log likelihood values indicates the null hypothesis is supported by the data. Whereas, a large difference between the log likelihood values indicates that the data favour the less restrictive hypothesis. The LR test requires that the model be estimated both under the null and under the alternative hypotheses.

### Wald test

The Wald statistic measures the distance between  $\hat{\theta}_1$  and the value of the parameter under the null hypothesis,  $\theta_0$ , where the weight is given by the inverse of the covariance matrix of the estimate evaluated at  $\hat{\theta}_1$ ,  $\Omega(\hat{\theta}_1)$ . The test is given by

$$W = T(\hat{\theta}_1 - \theta_0)'(\Omega(\hat{\theta}_1))^{-1}(\hat{\theta}_1 - \theta_0).$$

The Wald test requires that the model be estimated under the alternative hypothesis only. The widespread application of  $t$ -tests and  $F$ -tests to financial models are all based on the Wald test principle.

### Lagrange multiplier tests

The LM statistic measures the distance between  $\hat{\theta}_0$  and  $\hat{\theta}_1$  in terms of the gradients of the two log likelihood functions, that is,  $G(\hat{\theta}_0)$  and  $G(\hat{\theta}_1)$ . Since  $\hat{\theta}_1$  maximises  $\log L(\theta)$ ,  $G(\hat{\theta}_1) = 0$ , which implies that the LM test statistic does not depend on the maximum likelihood estimator of  $\theta$  under the alternative hypothesis. The test is given by

$$LM = TG(\hat{\theta}_0)'(\Omega(\hat{\theta}_0))^{-1}G(\hat{\theta}_0),$$

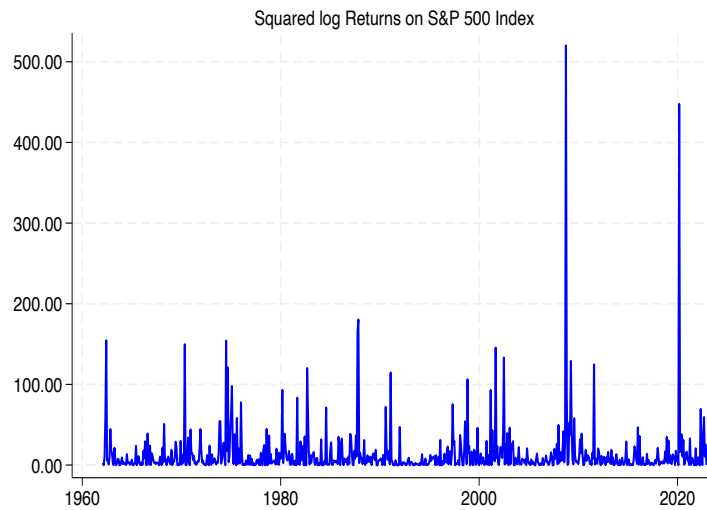
where  $\Omega(\hat{\theta}_0)$  is the disturbance covariance matrix evaluated under the null hypothesis. It follows that the LM test requires that the model be estimated under the null hypothesis.

Under the null hypothesis all three test statistics have chi-squared asymptotic distributions with degrees of freedom equal to the number of restrictions imposed under the null hypothesis. In view of this property the choice of which test statistic to use is often based on expediency. For example, the Likelihood ratio test requires estimation of the parameters of the model under both the null and alternative hypothesis. On the other hand, the Wald test requires estimation of the parameters only under the alternative hypothesis, while the LM test requires estimation of the parameters only under the null hypothesis.



## 7 Modelling Volatility

Although Table 1 suggests the three variances reported for the full and sub-samples of log returns are nearly identical, it would be inappropriate to assume that volatility does not change over time. One of the most documented features of financial asset returns is the tendency for large changes in the log returns on financial assets to be followed by further large changes or reversals (market turmoil), or for small changes in log returns to be followed by further small changes (market tranquility).



**Fig. 5** Squared percentage monthly log returns on the S&P 500 Index from January 1962 to December 2022.

Consider Figure 5 which plots the squared percentage monthly returns on the S&P 500 index for the period January 1962 to December 2022. It is apparent that there are sustained periods of market volatility, for example around the DotCom bubble of the early 2000s and during the subprime mortgage crisis of 2007/2008. This phenomenon is known as volatility clustering which highlights the property that the variation in financial returns is not constant over time but often appears to come in bursts of higher and lower variation. There are also periods of relative tranquility when the magnitude of movements in the returns is relatively small.

## 7.1 The GARCH(1,1) model

The simplest but most important volatility models are the autoregressive conditional heteroskedasticity (ARCH) class (Engle, 1982) and the generalised autoregressive conditional heteroskedasticity (GARCH) variant (Bollerslev, 1986). In particular, the GARCH(1,1) is given by the following equations

$$\begin{aligned}
 r_t &= \mu + u_t && \text{[Mean]} \\
 h_t &= \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} && \text{[Variance]} \\
 u_t &\sim N(0, h_t) && \text{[Distribution]}
 \end{aligned} \tag{13}$$

The first component of the model is the mean (or conditional mean) specification where the parameter  $\mu$  allows for returns to have a non-zero mean (or conditional mean). The second component of the model is the conditional variance specification whereby

$$h_t = E_{t-1}(u_t^2) = E_{t-1}(r_t - E_{t-1}(r_t))^2,$$

is a function of the lag of the squared disturbance term,  $u_t^2$ , and the lag of the conditional variance,  $h_{t-1}$ . The third component of the model is the choice of the distribution of the disturbance which at the moment is simply taken to have zero mean and variance  $h_t$ . The parameter  $\beta_1$  determines how past shocks affect the conditional variance at time  $t$ . The larger is  $\beta_1$  the longer is the memory of the shock. At the other extreme where  $\beta_1 = 0$ , the length of time a shock affects  $h_t$  is finite, with a memory of just 1-period. For this special case the model is known as the ARCH(1) model.

Estimation of the parameters of the GARCH(1,1) model requires that the distribution of  $u_t$  be specified. There are two common choices.

### Normal distribution

If  $u_t$  is normally distributed with mean 0 and variance  $h_t$  then it follows that the distribution of  $r_t$  is

$$f(r_t; \theta) = \frac{1}{\sqrt{2\pi h_t}} \exp\left(-\frac{(r_t - \mu)^2}{2h_t}\right),$$

with  $\theta = \{\mu, \alpha_0, \alpha_1, \beta\}$ . Based on this distribution the log likelihood function is given by

$$\log L(\theta) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \log h_t - \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \frac{(r_t - \mu)^2}{h_t},$$

where  $h_t$  is defined in (13). This is an example of the log likelihood function belonging to Case 3 as a result of  $h_t$  being a function of  $r_{t-1}, r_{t-2}, \dots$ .

An important property required of GARCH models is that the conditional variance  $h_t$ , defined in equation (??), needs to be positive at each point in time. Failure to meet this requirement, even for just one observation, will mean that the log likelihood function becomes undefined as a result of taking the logarithm of a negative number.

To estimate the GARCH model using an iterative optimisation algorithm, starting values are needed for the parameters together with some initial values for computing the conditional variance. The specification at observation  $t = 1$  is

$$h_1 = \alpha_0 + \alpha_1 u_0^2 + \beta_1 h_0,$$

so that starting values for  $u_0$  and  $h_0$  are required in order to compute  $h_1$ . One possible choice of starting values is to set  $u_0 = 0$  and to set  $h_0$  equal to an estimate of the unconditional variance of  $r_t$ .

### Student $t$ distribution

Now assume that  $u_t$  follows the standardised  $t$  distribution given in equation (10) with degrees of freedom parameter  $\nu$  so that  $\theta = \{\mu, \alpha_0, \alpha_1, \beta_1, \nu\}$  and  $\nu > 2$  is assumed. The variance of the conditional distribution is conveniently given directly by  $h_t$ . The log likelihood function is then

$$\begin{aligned} \log L(\theta) = & \log \left( \Gamma \left( \frac{\nu+1}{2} \right) \right) - \frac{1}{2} \log (\pi (\nu-2)) - \log \left( \Gamma \left( \frac{\nu}{2} \right) \right) \\ & - \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \log h_t - \left( \frac{\nu+1}{2} \right) \frac{1}{T} \sum_{t=1}^T \log \left( 1 + \frac{(r_t - \mu)^2}{h_t (\nu-2)} \right), \end{aligned}$$

with  $h_t$  defined in (13). As before, an iterative optimisation algorithm is needed to estimate the parameters of the model by maximum likelihood.

The degrees of freedom parameter for the  $t$  distribution indicates that the normal distribution may be a less appropriate choice the assumption of normally distributed disturbances will therefore result in a misspecification of the log likelihood function. Provided that the conditional mean and variance specifications are not misspecified, estimates of the model's parameters are still consistent despite the fact that the shape

**Table 3** Parameter estimates of the GARCH(1,1) model for the monthly log returns on the S&P 500 index for period January 1962 to December 2022 for log likelihood functions based on the normal and Student  $t$  distributions. Standard errors are in parentheses.

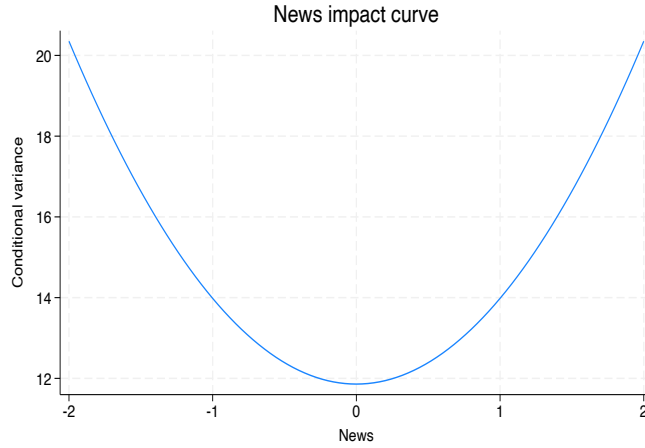
Distribution	$\hat{\mu}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}$	$\hat{\nu}$	$\log L$
Normal	0.686 (0.133)	1.586 (0.561)	0.153 (0.029)	0.740 (0.063)		-1953.8
Student $t$	0.884 (0.109)	1.359 (0.612)	0.140 (0.046)	0.767 (0.073)	4.496 (0.729)	-1907.0

of the distribution is incorrectly represented. But in this case standard errors need to be computed by means of a combination of the Hessian and the outer product of the gradient of the log likelihood function. In the context of GARCH models these standard errors are known as Bollerslev-Wooldridge standard errors (Bollerslev and Wooldridge, 1992).

## 7.2 Univariate Extensions

The GARCH model has been extended in numerous ways but perhaps the most important of these extensions concerns the effects of shocks that embody the effects of news relevant to financial markets that arrived in period  $t - 1$ . In no-news days, good and bad news balance and  $u_{t-1} = 0$ , whereas positive (negative) values of  $u_{t-1}$  represent good (bad) news. An important property of this GARCH(1,1) specification is that shocks of the same magnitude, positive or negative, result in the same increase in volatility  $h_t$ . That is, positive news with  $u_{t-1} > 0$  has the same effect on the conditional variance as negative news  $u_{t-1} < 0$  because it is only the absolute size of the news that matters since it is the squared function,  $u_{t-1}^2$ , that enters the equation. To illustrate the GARCH news impact curve (NIC),  $h_t$  is plotted against  $u_{t-1}$  in Figure 6. The major point to note is that the NIC of the simple GARCH model is symmetric.

In the case of stock markets, an asymmetric response to news is supported by theory, in which case negative shocks,  $u_{t-1} < 0$ , have a larger effect on the conditional variance. The heuristic explanation is that a negative shock raises the debt-equity ratio, thereby increasing leverage and consequently risk. It is this (aptly-named) leverage effect that suggests bad news leads to a greater increase in conditional



**Fig. 6** News impact curve for the GARCH(1,1) model based on the normal distribution for the percentage monthly log returns on the S&P 500 index for the period January 1962 to December 2022.

variance than good news. There are two popular specifications in the GARCH class that relax the restriction of a symmetric response to the news.

### Threshold GARCH (TARCH)

The TARCH(1,1) specification (Zakoian, 1994; Glosten et al., 1993) of the conditional variance is

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \lambda u_{t-1}^2 I_{t-1},$$

where  $I_{t-1}$  is an indicator variable defined as

$$I_{t-1} = \begin{cases} 1 & : u_{t-1} \geq 0 \\ 0 & : u_{t-1} < 0. \end{cases}$$

The leverage effect in equity markets suggests  $\lambda < 0$ , so that negative news,  $u_{t-1} < 0$ , is associated with a higher effect on volatility than positive news of the same magnitude.

### Exponential GARCH (EGARCH)

The EGARCH(1,1) specification Nelson (1991) of the conditional variance is

$$\log h_t = \alpha_0 + \alpha_1 \left| \frac{u_{t-1}}{\sqrt{h_{t-1}}} \right| + \lambda_1 \frac{u_{t-1}}{\sqrt{h_{t-1}}} + \beta_1 \log h_{t-1}.$$

An important advantage of the EGARCH specification is that the conditional variance is guaranteed to be positive at each point in time. This result follows from the fact that the variance is expressed in terms of  $\log h_t$  so that the actual variance is obtained by exponentiation. The parameter  $\alpha_1$  captures any potential asymmetry in the effect of  $u_{t-1}$  on  $\log h_t$ . It is expected that  $\alpha_1 < 0$ , so that negative news is associated with a higher effect than positive news of the same magnitude.

### 7.3 Forecasting

Despite these numerous extensions and refinements, in practice the GARCH(1,1) has remained a workhorse in many financial econometric applications. Perhaps its enduring popularity is due to the fact that the forecasts generated by the GARCH(1,1) model are difficult to beat in practice (Hansen and Lunde, 2005). For a GARCH(1,1) model, forecasts of  $h_t$  from the model are generated by replacing the unknown parameters  $\alpha_0$ ,  $\alpha_1$  and  $\beta_1$  and the unknown quantities  $u_T^2$  and  $h_T$  by their respective sample estimates. The forecasts are computed recursively starting with

$$\widehat{h}_{T+1|T} = \widehat{\alpha}_0 + \widehat{\alpha}_1 \widehat{u}_T^2 + \widehat{\beta}_1 \widehat{h}_T.$$

Given this estimate,  $\widehat{h}_{T+2|T}$  is computed as

$$\widehat{h}_{T+2|T} = \widehat{\alpha}_0 + (\widehat{\alpha}_1 + \widehat{\beta}_1) \widehat{h}_{T+1|T},$$

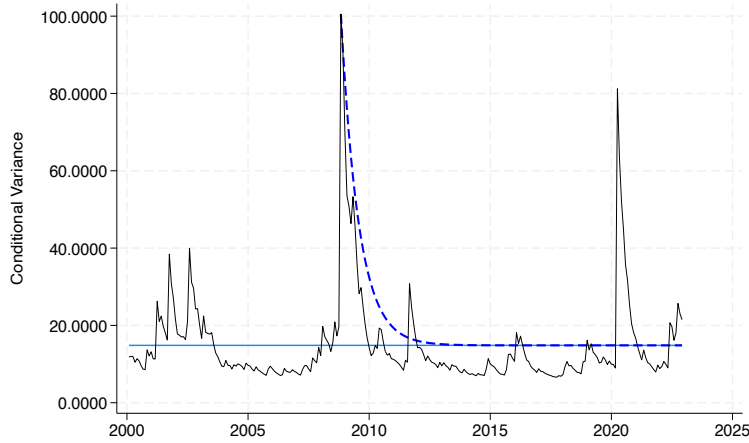
which, in turn, is used to compute  $\widehat{h}_{T+3|T}$  and so on. To forecast higher order GARCH models the same recursive approach is adopted.

The forecast from a GARCH(1,1) model converges relatively quickly to the estimate of the long-term average volatility implied by the model, which is given by

$$\widehat{h} = \frac{\widehat{\alpha}_0}{1 - \widehat{\alpha}_1 - \widehat{\beta}_1}.$$

The estimate unconditional variance  $\widehat{h}$  is defined so long as  $\widehat{\alpha}_1 + \widehat{\beta}_1 < 1$ . Figure 7 demonstrates this convergence for S&P 500 returns. Based on the estimates of the GARCH(1,1) model in the top two rows of Table 3, out-of-sample predictions are made starting in October 2008. The forecast converges to the long-term mean despite





**Fig. 7** Forecast (dashed line) and estimated (solid line) of the conditional variance of S&P 500 returns obtained from a GARCH(1,1) model. Also shown is the horizontal line representing the estimated unconditional variance implied by the model. The forecast starts at the beginning of October 2008 at the beginning of the global financial crisis but quickly converges to the long-term mean.

the fact it starts well above the long-term mean. The fact that the actual estimated conditional variance series drops off a lot more quickly than the forecast indicating that GARCH forecasts are quite persistent.

#### 7.4 Multivariate GARCH

Univariate GARCH models focus on the variance of a financial return. Multivariate GARCH models focus jointly on the variances of financial returns as well as their covariances. Consider, for example, a portfolio containing  $N = 2$  assets with log returns  $r_{1t}$  and  $r_{2t}$ . The optimal weights of the minimum variance portfolio are (Hurn et al., 2021)

$$w_1 = \frac{\text{var}(r_{2t}) - \text{cov}(r_{1t}, r_{2t})}{\text{var}(r_{1t}) + \text{var}(r_{2t}) - 2\text{cov}(r_{1t}, r_{2t})}, \quad w_2 = 1 - w_1, \quad (14)$$

where  $w_1$  is the optimal weight allocated to asset 1 in the portfolio and  $w_2$  is the corresponding weight on asset 2. The weights  $w_1$  and  $w_2$  are assumed to be constant over time as they are a function of the two unconditional variances and the unconditional covariance. Restricting the portfolio weights to be constant suggests that there is no

need to rebalance the portfolio to account for financial shocks. To relax this assumption and allow for time-varying portfolio weights the unconditional expectations in (14) are replaced by conditional expectations. As with the time-varying model of  $\beta$ -risk, the conditional variances can be modelled using a multivariate extension of the GARCH model (MGARCH).

In the simplest bivariate case there are two assets whose means are

$$\begin{aligned} r_{1t} &= \mu_1 + u_{1t} \\ r_{2t} &= \mu_2 + u_{2t}. \end{aligned} \quad (15)$$

The conditional variances and covariance of the disturbances are given by the matrix

$$H_t = E(u_t u_t') = \begin{pmatrix} h_{11t} & h_{12t} \\ h_{21t} & h_{22t} \end{pmatrix}. \quad (16)$$

The conditional variances are located down the main diagonal which are now defined as  $h_{iit}$ . The conditional covariances are located in the off-diagonal terms which satisfy the symmetry restriction  $h_{ijt} = h_{jit}$ . A natural way to proceed is to assume that the disturbances follow a bivariate normal distribution<sup>17</sup>

$$f(u_t) = (2\pi)^{-1} |H_t|^{-1/2} \exp\left(-\frac{1}{2} u_t' H_t^{-1} u_t\right), \quad (17)$$

There are two issues which make the estimation of MGARCH models difficult. The first of these problems relates to the requirement that conditional variances need to be positive at all points in time. The problem is highlighted in equation (17) in which the determinant of the covariance matrix,  $|H_t|$ , be positive at each point in time.<sup>18</sup> The second, and no less acute problem, is that in fully specified multivariate GARCH models the number of parameters to be estimated increases exponentially with the number of assets in the system,  $N$ .

A solution to both these problems is the DCC model of Engle (2002). This model is now one of the most widely adopted MGARCH specifications in empirical work. The DCC model addresses the joint issues of positive definiteness and parameter dimensionality by specifying the conditional covariance matrix as

---

<sup>17</sup> For this bivariate model, the matrix determinant is  $|H_t| = h_{11t}h_{22t} - h_{12t}^2$ . The matrix inverse is

$$H_t^{-1} = \frac{1}{h_{11t}h_{22t} - h_{12t}^2} \begin{pmatrix} h_{22t} & -h_{12t} \\ -h_{21t} & h_{11t} \end{pmatrix}.$$

<sup>18</sup> In matrix notation, the covariance matrix is required to be positive definite.

$$H_t = \begin{pmatrix} h_{11t} & h_{12t} \\ h_{12t} & h_{22t} \end{pmatrix} = S_t R_t S_t, \quad (18)$$

where  $S_t$  is a diagonal matrix of conditional standard deviations on the diagonal and zero elsewhere, and  $R_t$  is a conditional correlation matrix with a unity diagonal and correlations in the off-diagonal terms. The  $S_t$  matrix is obtained by estimating  $N$  univariate GARCH models for the returns on each asset. The  $R_t$  matrix is constructed by defining

$$R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}, \quad (19)$$

with the matrix  $Q_t$  evolving according to the GARCH specification

$$Q_t = (1 - \lambda_1 - \lambda_2) Q + \lambda_1 z_{t-1} z'_{t-1} + \lambda_2 Q_{t-1}. \quad (20)$$

Equation (19) ensures that the positive definiteness property is satisfied, while by choosing the parameters  $\lambda_1$  and  $\lambda_2$  as scalars ensures the parameter dimensionality issue is addressed as there are just two parameters capturing all covariance dynamics regardless of the dimension  $N$ .

Estimation of the DCC parameters of the MGARCH model is achieved by maximum likelihood methods. For two assets,  $N = 2$ , and assuming bivariate normality, the log likelihood function is given by

$$\log L(\theta) = -\log 2\pi - \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \log \begin{vmatrix} h_{11t} & h_{12t} \\ h_{12t} & h_{22t} \end{vmatrix} - \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}' \begin{pmatrix} h_{11t} & h_{12t} \\ h_{12t} & h_{22t} \end{pmatrix}^{-1} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix},$$

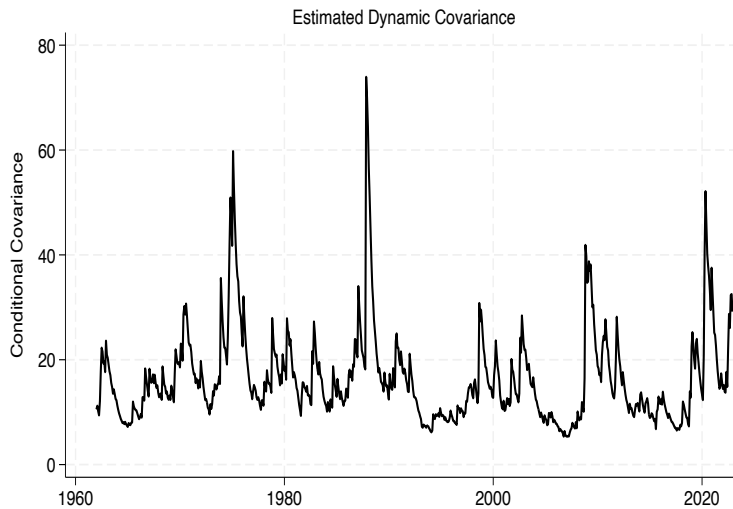
where  $u_{1t}$  and  $u_{2t}$  are explicit functions of the observed data as defined in (15), and  $H_t$  is defined in (18). Note that  $\theta$  is expanded to include all of the MGARCH parameters, including any additional parameters in the means of  $r_{1t}$  and  $r_{2t}$ . This is a nonlinear problem that requires a numerical algorithm. Most software packages routinely implement this estimation.

Maximum likelihood estimation of the model is illustrated using monthly percentage log returns on  $N = 2$  United States industry portfolios, consumer and manufacturing, for the period January 1962 to December 2022. The DCC model is estimated using both the multivariate normal and multivariate standardised Student  $t$  distributions with the parameter estimates reported in Table 4.

Figure 8 plots the predicted conditional covariance of the consumer and manufacturing portfolios obtained from the DCC model with Student  $t$  errors. The adjustment parameters reported in Table 4 are statistically significant suggesting that there is

**Table 4** Coefficient estimates for DCC models based on the normal and Student  $t$  distributions. The data are percentage monthly log returns on the consumer and manufacturing industry portfolios. The sample period is January 1962 to December 2022. Standard errors are in parentheses.

Distribution	Asset	Mean		Variance		Covariance		DOF
		$\hat{\mu}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\nu}$
Normal	Consumer	1.042 (0.149)	1.117 (0.353)	0.118 (0.021)	0.837 (0.028)	0.081	0.866	
	Manufacturing	0.993 (0.139)	0.955 (0.313)	0.102 (0.019)	0.852 (0.026)	(0.018)	(0.032)	
Student $t$	Consumer	1.099 (0.143)	1.221 (0.453)	0.115 (0.026)	0.838 (0.035)	0.077	0.885	6.395
	Manufacturing	1.039 (0.132)	1.062 (0.389)	0.090 (0.021)	0.855 (0.033)	(0.017)	(0.025)	(1.043)



**Fig. 8** Estimated dynamic conditional covariance of the consumer and manufacturing portfolios obtained from the DCC model. Estimates are obtained from the DCC model with Student  $t$  distributed errors. The sample period is January 1962 to December 2022.

*prima facie* evidence to support the claim that the covariance are time varying.<sup>19</sup> This conclusion is mirrored by the plot of the evolution of the conditional correlation.

<sup>19</sup> Testing for dynamic correlation in multivariate GARCH models is discussed in (Harvey and Thiele, 2016; Silvennoinen and Teräsvirta, 2016).

## 8 Multivariate Models of Returns, Prices and Dividends

An important class of multivariate models in finance is the vector autoregression (VAR) wherein each variable in the system is expressed as a linear function of its own lags as well as the lags of all of the other variables in the system. This type of model was first explored systematically in a pioneering article by [Mann and Wald \(1943\)](#) and reinvigorated by the work of [Sims \(1972\)](#). One limitation of VARs is that they can potentially fail to capture long-run relationships between the variables in the system. Vector error correction models (VECMs) impose cross-equation restrictions on the parameters of a VAR that allow these long-run or equilibrium relationships to be captured.

### 8.1 Vector autoregressions

Consider a bivariate model for log equity returns,  $r_{1t}$ , obtained in the now familiar way of applying the first difference filter to the log of equity prices and log dividend returns,  $r_{2t}$ , obtained by applying the first difference filter to the log of dividend payments. The VAR is written as

$$r_{1t} = \delta_1 + \gamma_{11}r_{1t-1} + \gamma_{12}r_{2t-1} + v_{1t} \quad (21)$$

$$r_{2t} = \delta_2 + \gamma_{21}r_{1t-1} + \gamma_{22}r_{2t-1} + v_{2t}, \quad (22)$$

where  $v_{1t}$  and  $v_{2t}$  are disturbance terms capturing unexplained movements in  $r_{1t}$  and  $r_{2t}$ , respectively, and the  $\gamma_{ij}$  are unknown parameters. In this model current changes in the variables are functions of delayed own effects and previous movements in the other variable in the system. Self evidently the VAR model is easily extended to allow for returns on multiple assets, and longer lag structures.

An important feature of the model is that it can be used to study the effects of shocks in  $r_{1t}$  arising from  $v_{1t}$ , and shocks in  $r_{2t}$  arising from  $v_{2t}$ , on current and future movements in the two returns. A further property of the model is that the shocks do not need to be independent as the covariance matrix of the two shocks is given by

$$H = E(v_t v_t') = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix}, \quad (23)$$

where  $h_{11}$  and  $h_{22}$  are the variances of the shocks, and  $h_{12} = h_{21}$  is the covariance between the two shocks.

Estimation of the parameters of the VAR in equations (21) to (23) is accomplished using maximum likelihood methods based on specifying a multivariate distribution for  $v_t = (v_{1t}, v_{2t})'$ . As in Section 7, consider the bivariate case of two assets,  $N = 2$ , where  $v_t$  follows a bivariate normal distribution with mean zero and covariance matrix  $H$  given by

$$f(v_t) = (2\pi)^{-1} |H|^{-1/2} \exp\left(-\frac{1}{2} v_t' H^{-1} v_t\right). \quad (24)$$

The difference between this equation and the bivariate normal distribution given in (17) is that  $H$  is now no longer time varying. The log likelihood function is

$$\log L(\theta) = -\log 2\pi - \frac{1}{2} \log \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix} - \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}' \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix}^{-1} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}, \quad (25)$$

where  $v_{1t}$  and  $v_{2t}$  are explicit functions of the data and are defined in (22). The parameter vector is

$$\theta = (\delta_1, \delta_2, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, h_{11}, h_{22}, h_{12}).$$

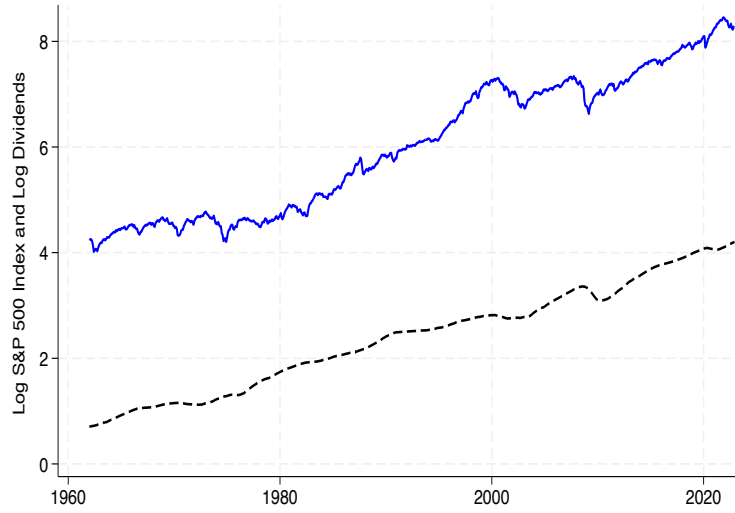
and the log likelihood function is to be maximised with respect to  $\theta$  in the usual manner.<sup>20</sup>

## 8.2 VECM Models

While the VAR is a useful multivariate model to analyse short-run dynamics, it does not necessarily capture long-run behaviour arising from potential relationships between the variables in the system. Figure 9 shows the logarithms of monthly equity prices,  $p_t = \log P_t$ , and the log of monthly dividend payments,  $d_t = \log D_t$ , over the period January 1962 to December 2022. It is also apparent from Figure 9 that although equity prices and dividends are nonstationary, there is a strong comovement in the two series over time.

The observation that prices and dividends are linked is consistent with the present value model of equity prices in which price is related to the future stream of dividends associated with the asset. This long-run relationship is expressed as

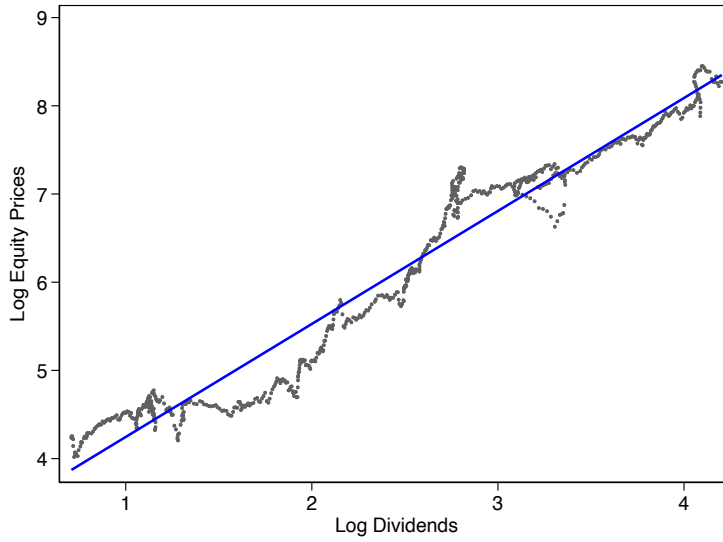
<sup>20</sup> In fact, the maximum likelihood estimation of the VAR parameters reduces to applying ordinary least squares to each equation of the VAR separately. This property represents the multivariate least squares analogue of the result given in footnote 3.



**Fig. 9** Time series plots of the logarithms of monthly United States equity prices (solid line) and dividend payments (dashed line) for the period January 11962 to December 2022.

$$p_t = \beta_0 + \beta_d d_t + u_t, \quad (26)$$

where  $p_t$  is the log equity price,  $d_t$  is the log dividend,  $u_t$  is a disturbance term and  $\beta_0$  and  $\beta_d$  are unknown parameters. This relationship is clearly illustrated in Figure 10. Superimposed on the scatter is the best fit of equation (26). What this diagram shows is that any deviation from the relationship encapsulated by the linear model is stationary. The natural conclusion is that the disturbances  $u_t$  are transient shocks and the system reacts in a way so as to restore equilibrium after the impact of a shock. It is this tendency for changes in log equity prices (log equity returns,  $r_{1t}$ ) and changes in log dividends (log dividend returns,  $r_{2t}$ ) to restore equilibrium that is missing from the VAR model. In econometric parlance, this represents a cointegrating relationship where  $\beta_0$  and  $\beta_d$  are the cointegrating parameters. If there were no cointegration between the variables the scatter diagram would have points much more evenly scattered about the two dimensional plane. Cointegration has the effect of an attractor amongst nonstationary series, in the present case compressing equity prices and dividends close to a one dimensional relationship. This suggests that for cointegration to be satisfied  $p_t$  and  $d_t$  must be nonstationary while  $u_t$  is in fact stationary.



**Fig. 10** Scatter plot of the logarithms of monthly United States equity prices and dividends for the period January 1962 to December 2022, with the best fitting estimate of equation (26) superimposed.

To understand the forces at work within the present value model in response to a shock  $u_{t-1}$  at time  $t$  the dynamics of adjustment need to be specified. The change in the log equity price,  $p_t - p_{t-1} = r_{1t}$  and log dividends  $d_t - d_{t-1} = r_{2t}$ , may be represented by the following two adjustment equations

$$p_t - p_{t-1} = \delta_1 + \alpha_1 u_{t-1} + v_{1t}$$

$$d_t - d_{t-1} = \delta_2 + \alpha_2 u_{t-1} + v_{2t},$$

or alternatively, by using equation (26),

$$p_t - p_{t-1} = \delta_1 + \alpha_1(p_{t-1} - \beta_0 - \beta_d d_{t-1}) + v_{1t} \quad (27)$$

$$d_t - d_{t-1} = \delta_2 + \alpha_2(p_{t-1} - \beta_0 - \beta_d d_{t-1}) + v_{2t}, \quad (28)$$

where the  $v_{it}$  are disturbance terms capturing additional movements in prices and dividends unrelated to the shock captured in  $u_t$ . Consider the effect of a positive shock,  $u_{t-1} > 0$ , that moves the system to a point above the solid line in Figure 10. Given that equity prices need to adjust downwards in this scenario to restore equilibrium, the adjustment parameter  $\alpha_1$  satisfies the restriction  $\alpha_1 < 0$ . Similarly, dividends need to increase to restore equilibrium in this scenario, the adjustment



parameter  $\alpha_2$  satisfies the restriction  $\alpha_2 > 0$ . The relative strength of the movements in equity prices and dividends is determined by the relative magnitudes of the adjustment parameters  $\alpha_1$  and  $\alpha_2$ . The adjustment parameters  $\alpha_1$  and  $\alpha_2$  are known as the error correction parameters as they control the relative strengths of the adjustments in the dependent variables with respect to the (lagged) equilibrium error.

The model represented by equations (27) and (28) is known as a vector error correction model (VECM). Once additional short-run dynamics are allowed the model may be respecified as

$$\begin{aligned}\Delta p_t &= \delta_1 + \alpha_1(p_{t-1} - \beta_0 - \beta_d d_{t-1}) + \gamma_{11}\Delta p_{t-1} + \gamma_{12}\Delta d_{t-1} + v_{1t} \\ \Delta d_t &= \delta_2 + \alpha_2(p_{t-1} - \beta_0 - \beta_d d_{t-1}) + \gamma_{21}\Delta p_{t-1} + \gamma_{22}\Delta d_{t-1} + v_{2t}.\end{aligned}$$

The strength of these additional dynamics are captured by the  $\gamma_{11}$ ,  $\gamma_{12}$ ,  $\gamma_{21}$  and  $\gamma_{22}$  parameters which measure the magnitude and direction of influence of the lagged dependent variables. The multivariate nature of a VECM means that all lagged dependent variables appear in both equations.

The model can be written in terms of log returns as

$$r_{1t} = \delta_1 + \alpha_1(p_{t-1} - \beta_0 - \beta_d d_{t-1}) + \gamma_{11}r_{1t-1} + \gamma_{12}r_{2t-1} + v_{1t} \quad (29)$$

$$r_{2t} = \delta_2 + \alpha_2(p_{t-1} - \beta_0 - \beta_d d_{t-1}) + \gamma_{21}r_{1t-1} + \gamma_{22}r_{2t-1} + v_{2t}, \quad (30)$$

making it obvious that the VECM is simply a VAR model of log returns with an added term that imposes the long-run equilibrium on the system in the form of cross-equation restrictions.

The log likelihood function of the bivariate VECM is given by

$$\log L(\theta) = -\log 2\pi - \frac{1}{2} \log \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix} - \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}' \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix}^{-1} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}, \quad (31)$$

where  $v_{1t}$  and  $v_{2t}$  are the disturbances in the VECM equations (29) and (30). Despite the nonlinearity introduced by the long-run restrictions, Johansen (1988) showed that a convenient non-iterative maximum likelihood algorithm is available to maximise the log likelihood function. For this reason the maximum likelihood estimator is often referred to as the Johansen estimator, although it is nonetheless generically a maximum likelihood estimator. Furthermore, it is a special case of what is known as a reduced rank regression because of the linear cointegrating linkage factor  $(p_{t-1} - \beta_0 - \beta_d d_{t-1})$  that appears as a restriction in the two equations (29) and (30). This factor reduces the effective nonstationarity of the individual variables

$\{p_{t-1}, d_{t-1}\}$  to accord with the stationarity of the differenced endogenous dependent variables  $\{\Delta p_t, \Delta d_t\}$  and the equation errors, thereby balancing the fitted equations. Such multivariate linear regression models that rely on reduced rank restrictions were originally examined in the work of [Anderson \(1951\)](#). The primary advance of the econometric literature in the context of such restricted multiple time series models was the rigorous treatment of nonstationarity that opened up the development of a valid asymptotic theory of estimation and inference ([Phillips and Durlauf, 1986](#); [Phillips, 1988a,b](#); [Johansen, 1988](#)).

## 9 Conclusion

The teaching perspective adopted throughout this chapter is that maximum likelihood represents a useful framework in which to achieve a synthesis of specification, estimation and testing in financial econometrics. In our view this framework enables a good blend of data description, finance theory, introductory econometric methodology, and empirical implementation that is most likely to engage the interest of students with little or no finance or econometric backgrounds, while at the same time providing just enough technical detail to maintain the attention of students with greater background strength.

With this simple beginning where might one look to build a second semester that offers more advanced material? Evidently, this first course adopts an intuitive approach that mixes ideas and methods with data and empirical implementation. So there is much detail to be provided if a rigorous foundation is to be built. These additions could be achieved by working backwards to strengthen technicalities or by moving forward in a way that introduces new concepts and material that enhances motivation. Keeping with a maximum likelihood approach, one nice extension is to consider order statistics of returns and how these lead naturally to extreme value distributions, giving new insights about extreme returns and introducing a group of useful new distributions. It is then a short step forward to copulas and extreme return linkages. With these advances toward the financial frontier, some of the earlier topics from the first course can be explored in more detail. Multivariate GARCH modeling and cointegration methods can be revisited at a more general and higher technical level, opening up the vast arena of econometric analysis and treatment of nonstationarity complete with cointegration testing in both VAR and triangular system models. Beyond these topics and closer to the frontiers of financial

econometrics as well as the financial industry itself lies a wealth of new models arising out of market microstructure considerations, many aspects of which are treatable within the maximum likelihood framework.

## References

- Anderson TW (1951) Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics* pp 327–351
- Andrews DW, McDermott CJ (1995) Nonlinear econometric models with deterministically trending variables. *The Review of Economic Studies* 62(3):343–360
- Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31:307–327
- Bollerslev T, Wooldridge J (1992) Quasi-maximum likelihood estimation and inference in a dynamic model with time-varying covariances. *Econometric Reviews* 11:143–172
- Cochrane J (2001) *Asset Pricing*. Princeton University Press, Princeton, New Jersey
- Davidson J, Hendry D, Srba F, Yeo S (1978) Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal* 88:661 – 692
- Dickey D, Fuller W (1979) Distributions of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74:427–431
- Dickey D, Fuller W (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49:1057–1072
- Engle R (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50:987–1008
- Engle R (2002) Dynamic conditional correlation. a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20:339–350
- Glosten L, Jagannathan R, Runkle D (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48:1779–1801
- Gordon M (1959) Dividends, earnings and stock prices. *Review of Economics and Statistics* 41:99–105

- Hansen L (1982) Large sample properties of generalised method of moments estimators. *Econometrica* 50:1029–1054
- Hansen P, Lunde A (2005) A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* 20:873–889
- Harvey A, Thiele S (2016) Testing against changing correlation. *Journal of Empirical Finance* 38:575–589
- Hurn A, Martin V, Phillips PCB, Yu J (2021) *Financial Econometric Modeling*. Oxford University Press, New York
- 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
- Mann H, Wald A (1943) On the statistical treatment of linear stochastic difference equations. *Econometrica* 11:173–220
- Martin V, Hurn A, Harris D (2013) *Econometric Modelling with Time Series: Specification, Estimation and Testing*. Cambridge University Press, New York
- Nelson D (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59:347–370
- Newey W, McFadden D (1994) Large sample estimation and hypothesis testing. In: Engle R, McFadden D (eds) *Handbook of Econometrics, Volume 4*, Elsevier
- Pagan A, Ullah A (1999) *Nonparametric Econometrics*. Cambridge University Press, New York
- Pearson K (1905) The error law and its generalizations by fechner and pearson. A Rejoinder *Biometrika* 4(1/2):169–212
- Phillips PCB (1987) Time series regression with a unit root. *Econometrica* 55:277–301
- Phillips PCB (1988a) Multiple regression with integrated processes. In: Prabhu NU (ed) *Statistical Inference from Stochastic Processes, Contemporary Mathematics*, vol 80, Taylor & Francis, pp 79–106
- Phillips PCB (1988b) Regression theory for near-integrated time series. *Econometrica: Journal of the Econometric Society* 56(5):1021–1043
- Phillips PCB (1995) Fully modified least squares and vector autoregression. *Econometrica* 63:1023–1078
- Phillips PCB, Durlauf SN (1986) Multiple time series regression with integrated processes. *The Review of Economic Studies* 53(4):473–495

- Phillips PCB, Hansen B (1990) Statistical inference in instrumental variables regressions with I(1) errors. *Review of Economic Studies* 57:99–125
- Phillips PCB, Loretan M (1991) Estimating long-run equilibria. *Review of Economic Studies* 58:407–436
- Phillips PCB, Ouliaris S (1990) Asymptotic properties of residual based tests for cointegration. *Econometrica* 58:165–193
- Phillips PCB, Perron P (1988) Testing for a unit root in time series regression. *Biometrika* 75:335–346
- Saikkonen P (1991) Asymptotically efficient estimation of cointegration regressions. *Econometric theory* 7(1):1–21
- Sargan J (1958) The estimation of economic relationships using instrumental variables. *Econometrica* 26:393–415
- Silvennoinen A, Teräsvirta T (2016) Testing constancy of unconditional variance in volatility models by misspecification and specification tests. *Studies in Nonlinear Dynamics and Econometrics* 20:347–364
- Sims C (1972) Money, income, and causality. *American Economic Review* 62:540–552
- Stock JH, Watson MW (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* pp 783–820
- Westfall PH (2014) Kurtosis as peakedness, 1905–2014. *rip. The American Statistician* 68(3):191–195
- White H (1994) *Estimation, Inference and Specification Analysis*. Cambridge University Press, New York
- Zakoïan J (1994) Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18:931–944