1 Introduction

1.1 Summary

Stochastic calculus is used in finance and econom(etr)ics for instance for solving stochastic differential equations and handling stochastic integrals. This requires stochastic processes. Although stemming from a rather recent area of mathematics, the methods of stochastic calculus have shortly come to be widely spread not only in finance and economics. Moreover, these techniques – along with methods of time series modeling – are central in the contemporary econometric tool box. In this introductory chapter some motivating questions are brought up being answered in the course of the book, thus providing a brief survey of the topics treated.

1.2 Finance

The names of two Nobel prize winners[1] dealing with finance are closely connected to one field of applications treated in the textbook at hand. The analysis and the modeling of stock prices and returns is central to this work.

Stock Prices

Let $S(t)$, $t \geq 0$, be the continuous stock price of a stock with return $R(t) = S'(t)/S(t)$ expressed as growth rate. We assume constant returns,

$$R(t) = c \quad \Longleftrightarrow \quad S'(t) = c\,S(t) \quad \Longleftrightarrow \quad \frac{dS(t)}{dt} = cS(t).$$

[1] In 1997, R.C. Merton and M.S. Scholes were awarded the Nobel prize jointly, "for a new method to determine the value of derivatives" (according to the official statement of the Nobel Committee).

U. Hassler, *Stochastic Processes and Calculus*, Springer Texts in Business and Economics, DOI 10.1007/978-3-319-23428-1_1

This differential equation for the stock price is usually also written as follows:

$$dS(t) = c\,S(t)\,dt\,. \tag{1.1}$$

The corresponding solution is (see Problem 1.1)

$$S(t) = S(0)\,e^{c\,t}\,, \tag{1.2}$$

i.e. if $c > 0$ the exponential process is explosive. The assumption of a deterministic stock price movement is of course unrealistic which is why a stochastic differential equation consistent with (1.1) is often assumed since Black and Scholes (1973) and Merton (1973),

$$dS(t) = c\,S(t)\,dt + \sigma\,S(t)\,dW(t)\,, \tag{1.3}$$

where $dW(t)$ are the increments of a so-called Wiener process $W(t)$ (also referred to as Brownian motion, cf. Chap. 7). This is a stochastic process, i.e. a random process. Thus, for a fixed point in time t, $S(t)$ is a random variable. How does this random variable behave on average? How do the parameters c and σ affect the expected value and the variance as time passes by? We will find answers to these questions in Chap. 12 on stochastic differential equations.

Interest Rates

Next, $r(t)$ denotes an interest rate for $t \geq 0$. Assume it is given by the differential equation

$$dr(t) = c\,(r(t) - \mu)\,dt \tag{1.4}$$

with $c \in \mathbb{R}$ or equivalently by

$$r'(t) = \frac{dr(t)}{dt} = c\,(r(t) - \mu).$$

Expression (1.4) can alternatively be written as the following integral equation:

$$r(t) = r(0) + c\int_0^t (r(s) - \mu)\,ds. \tag{1.5}$$

The solution to this reads (see Problem 1.2)

$$r(t) = \mu + e^{c\,t}\,(r(0) - \mu)\,. \tag{1.6}$$

For $c < 0$ therefore it holds that the interest rate converges to μ as time goes by. Again, a deterministic movement is not realistic. This is why Vasicek (1977) specified a stochastic differential equation consistent with (1.4):

$$dr(t) = c\,(r(t) - \mu)\,dt + \sigma\,dW(t)\,. \tag{1.7}$$

As aforementioned, $dW(t)$ denotes the increments of a Wiener process. How is the interest rate movement (on average) affected by the parameter c? Which kind of stochastic process is described by (1.7)? The answers to these and similar questions will be obtained in Chap. 13 on interest rate models.

Empirical Returns

Looking at return time series one can observe that the variance (or volatility) fluctuates a lot as time passes by. Long quiet market phases characterized by only mild variation are followed by short periods characterized by extreme observations where extreme amplitudes again tend to entail extreme observations. Such a behavior is in conflict with the assumption of normally distributed data. It is an empirically well confirmed law ("stylized fact") that financial market data in general and returns in particular produce "outliers" with larger probability than it would be expected under normality.

It is crucial, however, that extreme observations occur in clusters (volatility clusters). Even though returns are not correlated over time in efficient markets, they are not independent as there exists a systematic time dependence of volatility. Engle (1982) suggested the so-called ARCH model (see Chap. 6) in order to capture the outlined effects. His work constituted an entire field of research known nowadays under the keyword "financial econometrics", and consequently he was awarded the Nobel prize in 2003.[2]

1.3 Econometrics

Clive Granger (1934–2009) was a British econometrician who created the concept of cointegration (Granger, 1981). He shared the Nobel prize "for methods of analyzing economic time series with common trends (cointegration)" (official statement of the Nobel Committee) with R.F. Engle. The leading example of trending time series he considered is the random walk.

[2] R.F. Engle shared the Nobel prize "for methods of analyzing economic time series with time-varying volatility (ARCH)" (official statement of the Nobel Committee) with C.W.J. Granger.

Random Walks

In econometrics, we are often concerned with time series not fluctuating with constant variance around a fixed level. A widely-used model for accounting for this nonstationarity are so-called integrated processes. They form the basis for the cointegration approach that has become an integral part of common econometric methodology since Engle and Granger (1987). Let's consider a special case – the random walk – as a preliminary model,

$$x_t = \sum_{j=1}^{t} \varepsilon_j \,, \quad t = 1, \ldots, n \,, \tag{1.8}$$

where $\{\varepsilon_t\}$ is a random process, i.e. ε_t and ε_s, $t \neq s$, are uncorrelated or even independent with zero expected value and constant variance σ^2. For a random walk with zero starting value $x_0 = 0$ it holds by definition that:

$$x_t = x_{t-1} + \varepsilon_t \,, \quad t = 1, \ldots, n \,, \qquad \text{with } \operatorname{Var}(x_t) = \sigma^2 \, t \,. \tag{1.9}$$

The increments can also be written using the difference operator Δ,

$$\Delta x_t = x_t - x_{t-1} = \varepsilon_t \,.$$

Regressing two stochastically independent random walks on each other, a statistically significant relationship is identified which is a statistical artefact and therefore nonsense (see Chap. 15). Two random walks following a common trend, however, are called cointegrated. In this case the regression on each other does not only give the consistent estimation of the true relationship but the estimator is even "superconsistent" (cf. Chap. 16).

Dickey-Fuller Distribution

If one wants to test whether a given time series indeed follows a random walk, then equation (1.9) suggests to estimate the regression

$$x_t = \hat{a} \, x_{t-1} + \hat{\varepsilon}_t \,, \quad t = 1, \ldots, n \,.$$

From this, the (ordinary) least squares (LS) estimator under the null hypothesis (1.9), i.e. under $a = 1$, is obtained as

$$\hat{a} = \frac{\sum_{t=1}^{n} x_t \, x_{t-1}}{\sum_{t=1}^{n} x_{t-1}^2} = 1 + \frac{\sum_{t=1}^{n} x_{t-1} \, \varepsilon_t}{\sum_{t=1}^{n} x_{t-1}^2} \,.$$

This constitutes the basic ingredient for the test by Dickey and Fuller (1979). Under the null hypothesis of a random walk ($a = 1$) it holds asymptotically ($n \to \infty$)

$$n\,(\hat{a} - 1) \ \stackrel{d}{\to}\ \mathcal{DF}_a\,, \tag{1.10}$$

where "$\stackrel{d}{\to}$" stands for convergence in distribution and $\mathcal{DF}_a$ denotes the so-called Dickey-Fuller distribution. Corresponding modes of convergence will be explained in Chap. 14. Since Phillips (1987) an elegant way for expressing the Dickey-Fuller distribution by stochastic integrals is known (again, $W(t)$ denotes a Wiener process):

$$\mathcal{DF}_a = \frac{\int_0^1 W(t)\,dW(t)}{\int_0^1 W^2(t)\,dt}\,. \tag{1.11}$$

Note (and enjoy!) the formal correspondence of the sum of squares $\sum_{t=1}^{n} x_{t-1}^2$ in the denominator of $\hat{a} - 1$ and the integral over the squared Wiener process in the denominator of (1.11), $\int_0^1 W^2(t)\,dt$ (this is a Riemann integral, cf. Chap. 8). Just as well the sum $\sum_{t=1}^{n} x_{t-1}\,\varepsilon_t = \sum_{t=1}^{n} x_{t-1}\,\Delta x_t$ resembles the so-called Ito integral $\int_0^1 W(t)\,dW(t)$. But how are these integrals defined, what are they about? How is this distribution (and similar ones) attained? And why does there exist another equivalent representation,

$$\mathcal{DF}_a = \frac{W^2(1) - 1}{2\int_0^1 W^2(t)\,dt}\,, \tag{1.12}$$

of the Dickey-Fuller distribution? We concern ourselves with these questions in connection with Ito's lemma in Chap. 11.

Autocorrelation

The assumption of the increments $\Delta x_t = x_t - x_{t-1}$ of economic times series being free from serial (temporal) correlation – as it is true for the random walk – is too restrictive in practice. Thus, we have to learn how the Dickey-Fuller distribution is generalized with autocorrelated (i.e. serially correlated) increments. In practise, so-called ARMA models are used most frequently in order to model autocorrelation. This class of models will be discussed intuitively as well as rigorously in Chap. 3. The so-called spectral analysis translates autocorrelation patterns in oscillation patterns. In Chap. 4 we learn to determine which frequency's or period's oscillations add particularly intensely to a time series' variation. Often economists are refused access to spectral analysis because of the extensive use of complex numbers. Therefore, we suggest an approach that avoids complex numbers. Finally, Chap. 5 introduces a model where the temporal dependence is particularly persistent such that the autocorrelations die out more slowly than in the ARMA case. Such a feature has been called "long memory" and is observed with many economic and financial series.

1.4 Mathematics

Stochastic calculus, which will be applied here, is a rather recent area in mathematics. It was pioneered by Kiyoshi Ito[3] in a sequence of pathbreaking papers published in Japanese starting from the forties of the last century.[4] The Ito integral as a special case of stochastic integration is introduced in Chap. 10.

Ito Integrals

The aforementioned interest rate model by Vasicek (1977) leads to a stochastic process given by an integral constructed as $\int_0^t f(s)\, dW(s)$ where f is a deterministic function and again dW denotes the increments of a Wiener process. Such integrals – being in a sense classical integrals – will be defined as Stieltjes integrals in Chap. 9. Ito integrals are a generalization of these. At first glance, the deterministic function f is replaced by a stochastic process X, $\int_0^t X(s)\, dW(s)$. Mathematically, this results in a considerably more complicated object, the definition thereof being a problem on its own, cf. Chap. 10.

Ito's Lemma

At this point, the idea of Ito's lemma is briefly conveyed. For the moment, assume a deterministic (differentiable) function $f(t)$. Using the chain rule it holds for the derivative of the square f^2:

$$\frac{df^2(t)}{dt} = 2f(t)f'(t)$$

or rather

$$\frac{df^2(t)}{2} = f(t)f'(t)\, dt = f(t)\, df(t)\,. \tag{1.13}$$

[3] Alternative transcriptions of his name into the Latin alphabet, Itô or Itō, are frequently used in the literature and are equally accepted. In this textbook we follow the spelling of Ito's compatriot (Tanaka, 1996).

[4] In 2006, Ito received the inaugural Gauss Prize for Applied Mathematics by the International Mathematical Union, which is awarded every fourth year since then.

Thus, for the ordinary integral it follows

$$\int_0^t f(s)\,df(s) = \int_0^t f(s)f'(s)\,ds = \frac{1}{2}f^2(s)\bigg|_0^t = \frac{1}{2}\left(f^2(t) - f^2(0)\right) .$$

However, among other things, we will learn that the Wiener process is not a differentiable function with respect to time t. The ordinary chain rule does not apply and for the according Ito integral one obtains

$$\int_0^t W(s)\,dW(s) = \frac{1}{2}\left(W^2(s) - s\right)\bigg|_0^t = \frac{1}{2}\left(W^2(t) - W^2(0) - t\right) . \tag{1.14}$$

This result follows from the famous and fundamental lemma by Ito being a kind of "stochastified chain rule" for Wiener processes in its simplest case. Instead of (1.13) for Wiener processes it holds that

$$\frac{dW^2(t)}{2} = W(t)\,dW(t) + \frac{1}{2}\,dt . \tag{1.15}$$

Substantial generalizations and multivariate extensions will be discussed in Chap. 11. In particular, Ito's lemma will enable us to solve stochastic differential equations in Chap. 12, and it will turn out that $S(t)$ solving (1.3) is a so-called geometric Brownian motion. In Chap. 13 we will look in greater detail in models for interest rates as e.g. given by Eq. (1.7).

Starting point for all the considerations outlined is the Wiener process – often also called Brownian motion. Before turning to it and its properties, general stochastic processes need to be defined and classified beforehand. This is done – among other things – in the following chapter on basic concepts from probability theory.

1.5 Problems and Solutions

Problems

1.1 Solve the differential equation (1.1), i.e. obtain the solution (1.2).

1.2 Verify that $r(t)$ from (1.6) solves the differential equation (1.4).

1.3 Consider a simple regression model,

$$y_i = \alpha + \beta\,x_i + \varepsilon_i\,, \quad i = 1, \ldots, n\,,$$

with OLS estimator $\hat{\beta}$. Check that:

$$\hat{\beta} - \beta = \frac{\sum_{i=1}^{n} (x_i - \overline{x})\, \varepsilon_i}{\sum_{i=1}^{n} (x_i - \overline{x})^2}\,,$$

with arithmetic mean $\overline{x}$.
Hint: $\sum_{i=1}^{n} (x_i - \overline{x}) = 0$.

1.4 Let f^n denote n-th power of a function f with derivative f', $n \in \mathbb{N}$. Show that:

$$f^n(t) = f^n(0) + n \int_0^t f^{n-1}(s) f'(s)\, ds\,.$$

Hint: Chain rule as in (1.13).

Solutions

1.1 Using equation (1.1) we get by integration

$$\int_0^t \frac{S'(r)}{S(r)}\, dr = \int_0^t c\, dr = ct.$$

Since[5]

$$\frac{d \log(S(t))}{dt} = \frac{S'(t)}{S(t)}$$

this implies

$$\log(S(t)) - \log(S(0)) = ct\,,$$

or

$$\begin{aligned} S(t) &= e^{\log(S(0))}\, e^{ct} \\ &= S(0)\, e^{ct}\,, \end{aligned}$$

which is the required solution.

[5]By "log" we denote the natural logarithm to the base e.

1.2 Taking the derivative of (1.6) yields:

$$\begin{aligned}\frac{dr(t)}{dt} &= c\,e^{ct}\,(r(0)-\mu) \\ &= c\,(r(t)-\mu)\,,\end{aligned}$$

where again the given form of $r(t)$ was used. By purely symbolically multiplying by dt the equation (1.4) is obtained. Hence, the problem is already solved.

1.3 It is well known that the OLS estimator is given by "covariance divided by variance of the regressor", i.e. it holds that:

$$\hat{\beta} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}\,.$$

Because of $\sum_{i=1}^{n}(x_i-\bar{x}) = 0$ this simplifies to

$$\hat{\beta} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

Assuming the model to be correct and substituting $y_i = \alpha + \beta\,x_i + \varepsilon_i$, one obtains

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(\alpha+\beta\,x_i+\varepsilon_i)}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\,.$$

Again applying the argument $\sum_{i=1}^{n}(x_i-\bar{x}) = 0$ yields

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{i=1}^{n}(x_i-\bar{x})(\beta(x_i-\bar{x})+\varepsilon_i)}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \\ &= \beta + \frac{\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.\end{aligned}$$

This was exactly the claim.

1.4 We address the problem in a slightly more general way. Let g be a differentiable function with derivative g'. By the fundamental theorem of calculus it holds that[6]

$$\int_0^t g'(s)ds = g(t) - g(0),$$

[6]For an introduction to calculus we recommend Trench (2013); this book is available electronically for free as a textbook approved by the American Institute of Mathematics.

or

$$g(t) = g(0) + \int_0^t g'(s)ds.$$

If g describes a process over time, this last relation can be interpreted the following way: The value at time t is made up by the starting value $g(0)$ plus the sum or integral over all changes occurring between 0 and t. Now, choosing in particular $g(t) = f^n(t)$ with

$$g'(t) = nf^{n-1}(t)f'(t)\,,$$

we obtain the required result.

References

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy, 81*, 637–654.

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

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica, 50*, 987–1008.

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

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

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science, 4*, 141–183.

Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica, 55*, 277–301.

Tanaka, K. (1996). *Time series analysis: Nonstationary and noninvertible distribution theory*. New York: Wiley.

Trench, W. F. (2013). *Introduction to real analysis*. Free Hyperlinked Edition 2.04 December 2013. Downloaded on 10th May 2014 from http://digitalcommons.trinity.edu/mono/7.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics, 5*, 177–188.