Universitá di Pisa

Galileo Galilei School of Graduate Studies

Ph.D. Course in Physics

XXIst Entrance

Theoretical and Empirical Essays on the Dynamics of Financial and Energy Markets.

Author:

Davide Pirino

Supervisor:

Prof. Roberto Renó

Introduction and Work Plan

This thesis is inspired by two main lines of research. Topics are analyzed in Chapters 1, 3, 4, 5 and 6. Chapter 2 is devoted to help the reader unfamiliar with the concepts of measure theory and stochastic processes.

The first line of research is dedicated to highlight a drawback of the standard economic equilibrium model. We start from a question mainly raised by the ecological problem: is the economic equilibrium consistent with the physical world? The answer seems to be negative. Economic equilibrium theory establishes the optimal level of production and consumption of goods. Consumption is, in fact, a social issue. It depends on what consumers prefer for their own utility. For this reason consumption is not directly related to the laws of physics. However production is unavoidably linked with a physical process: the thermodynamic transformation of basic commodities in elaborated one, useful for consumption. Despite this fact most part of economic models, in the mainstream literature, completely neglect the thermodynamic cycles hidden in every production process.

In the last two decades the ecological problem has gained attention over the scientific community, focusing on the role of thermodynamic efficiency in the conversion of energy into work as a factor of economic growth. In Chapter 1 we propose an analitycal approach to economic equilibrium which takes into account for thermodynamic efficiency. Our idea is tho show that if irreversibility is present the classical economic equilibrium is changed, resulting in a more parsimonious use of energy. Standard economic equilibrium implies that the equilibrium itself remains unchanged if the numeraire adopted to price good is changed, i.e. all numeraires are equivalent. This is a strong and quite controversial result: it is intuitive that, being the conversion of energy into work intrinsically irreversible, energy is a special commodity and it is not equivalent to the other ones. Pricing in terms of energy should not be equivalent to pricing in terms of other goods, which in fact are obtained by energy itself.

Moreover the proposed "thermodynamic-consistent" economy turns back into the classical one if the production process is reversible. In this sense economic theory implicitly assumes that all production processes are reversible.

This assumption conflicts with any real world production process.

The second line of research is independent from the first one and it is mainly devoted to the analysis of discontinuities of assets quoted in financial markets. Several drastic events are known to have influenced and changed the status of the financial markets. In such a situation the uncertainty hidden in financial assets increased rapidly getting the markets into a very turbulent state. Examples of such events are the 1929 crash of Wall Street, Black Monday crisis of 1987 and the 9/11 terrorist attack. In fact these are

rare events of very high intensity. Many discontinuities of smaller amplitude affect the behavior of assets: on the average we can identify, visually, $5-10$ of such abrupt variations per year. Such kind of rapid and intense variations are usually referred as *jumps*. In this context we expect that, after a jump has occurred, the market switches in a new status characterized by an high level of volatility. As a consequence jumps are expected to have a predictive power on the future behaviour of assets. Despite this is a very intuitive fact it has not yet proved in the financial literature. A volatility forecasting model requires the definition of a volatility proxy. The idea is that proxies adopted in the literature for forecasting purposes are, in fact, contaminated by the presence of jumps in finite sample. In this context the forecasting power of jumps on future volatility cannot be revealed.

In order to highlight such a feature it is needed a precise estimate of the jump component. In this spirit we propose in Chapter 3 a powerful jump separation technique and we test its performances on eight markets of electricity. The separation technique we adopt is taken from very recents results of the financial literature and only requires the introduction of a threshold. In Chapter 4 we construct precise volatility estimators using the threshold separation technique. This approach allows for a volatility estimation unaffected by jumps. Moreover in Chapter 5 we show that a jump purified estimate of volatility allow for a better investigation of its memory properties. Finally in Chapter 6 we construct a volatility forecasting model based on the proposed estimators. Being based on an accurate separation of continuous and discontinuous component of volatility, the model reveals the forecasting power of jumps on future volatility. Moreover we find that the forecasting power of jumps extend to at least one month. A dazzling example of the turbulence triggered by discontinuous variations is the recent crisis of markets. In September 2008 a global big crash has occurred in most part of stock exchanges. Since ever markets show an high level of volatility, characterized by large returns of both negative and postivie intensity. In this sense our results are very topical and constitute a basis for further investigations.

Contents

Chapter 1

Economic Equilibrium: The Role of Thermodynamic Efficiency

This chapter is inspired from the paper of Roma and Pirino (2008).

1.1 Overview and Criticism of General Equilibrium Theory

Microeconomics is a discipline devoted to the investigation of individual decision making. Production of goods and consumption is analyzed in the microeconomic theory from an abstract point of view. This section briefly describes the fundamental principles of the theory which will be frequently used throughout all the chapter.

The scenario is composed of two parts: the producer who releases goods in the market and tries to maximize his profits and the consumer who buys goods and maximizes the utility derived from the goods themselves.

At the equilibrium point profits and utilities are maximized simultaneously. Usually the maximum is reached within some constraints which describe maximum expense sustainable by the consumer, producer's maximum production capability and so on.

The general equilibrium depends on the production technology used by the producer. Economic production is usually modelled ignoring the underlying thermodynamic mechanism which brings consumption goods to the market, despite it frequently involves physical processes.

Our goal is to show, focusing on a particular production process, how economic equilibrium is changed when thermodynamic laws are taken into account. We are especially interested on the consequences of thermodynamic irreversibility.

We will show that if the production process is irreversible then the producer is forced to a more parsimonious use of energy, while energy itself is found to play no particular role if the process is reversible.

The theoretical approach described in what follow is taken from Colell et al. (1995).

1.1.1 Preference and Choice

Individual choice behaviour is usually modelled introducing a set of alternatives $\mathbb{X}$. It is an abstract set because alternatives could be everything: take an holiday in Spain, France, Germany etc. It is assumed that the alternative presented by $\mathbb{X}$ are *mutually exclusive*.

The preferences on $\mathbb{X}$ are described by a binary relation between any two elements of $\mathbb{X}$. We indicate this relation by $\succeq$ and we say that "x is at least as good as y " for every x and y in $\in \mathbb{X}$ such that $x \succeq y$.

Other important relations are the *strict preference* relation $x \succ y$ defined by: $x \succeq y$ but not $y \succeq x$, and the *indifference* relation defined by: $x \sim y \Leftrightarrow x \succeq y$ and $y \succeq x$.

We assume that the preference relation $\succeq$ is *rational* i.e. it respects the following properties:

- $\forall x, y \in \mathbb{X}$ we have that $x \succeq y$ or $y \succeq x$ or both.
- $\forall x, y, z \in \mathbb{X}$ if $x \succeq y$ and $y \succeq z$ then $x \succeq z$.

It can be shown that if $\succeq$ is rational then $\succ$ is irreflexive ($x \succ x$ never holds) and transitive (if $x \succ y$ and $y \succ z$ then $x \succ z$) while $\sim$ is reflexive ($\forall x, x \sim x$) and symmetric ($x \sim y \Rightarrow y \sim x$).

In order to represent preferences as real numbers microeconomic theory refers to the concept of *utility functions*:

Definition 1.1.1 *A function $u : \mathbb{X} \to \mathbb{R}$ is a utility function representing preference relation $\succeq$ if $\forall x, y \in \mathbb{X}$ it happens that:*

$$x \succeq y \Rightarrow u(x) \geq u(y). \tag{1.1}$$

Relation (1.1) establishes, in some sense, that the preference is kept unchanged by the utility function in the real line. If $u(x)$ is a utility function then any strictly increasing function $f : \mathbb{R} \to \mathbb{R}$ define a new utility function $v(x)$, which describes the same preferences as $u(x)$, by imposing $v(x) \equiv f(u(x))$.

1.1.2 The Consumer's Problem

In our market model the consumer faces the problem of how to choose optimal consumption level for all commodities (and services) that are available for purchase. For the sake of simplicity suppose we have a finite number L of commodities. The consumer must choose for each $i = 1, ..., L$ the optimal level of consumption x_i of the i-th commodity. In fact his choice defines a *commodity vector* (or *commodity bundle*):

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_L \end{pmatrix} \tag{1.2}$$

which, generally, lies in $\mathbb{R}^L$. However the set of feasible values for the commodity vector (1.2) is, in fact, restricted to a smaller set than $\mathbb{R}^L$. This will become clear when budget constraints will be introduced. As a first constraint we require that all consumptions level are non-negative, i.e. the consumer is characterized to have only inflows of commodities: debits or outflows are not taken into account. We highlight that

the entries of vector (1.2) may represent the same commodity at different time: bread today and bread tomorrow are different commodities.

We assume that the price in dollar of all commodities is known. Further it is assumed that the consumer's demand for any commodity is negligible and thus prices are established beyond his influence [1]. The commodity prices are listed in the *price vector*:

$$p = \begin{pmatrix} p_1 \\ \vdots \\ p_L \end{pmatrix} \in \mathbb{R}^L. \tag{1.3}$$

Note that in this context price could be negative. A negative price could be paid by the consumer for the consumption of bad goods like pollution. However in our model prices are restricted to the positive semi-space $\mathbb{R}^{L+}$.

The consumer is supposed to have a finite amount of dollar to be spent on the market. We call this quantity the *total wealth* w.

As a consequence a commodity bundle $x \in \mathbb{R}^{L+}$ is affordable only if its total cost is less or equal than the consumer's total wealth, in formula:

$$p \cdot x = p_1 x_1 + ... + p_L x_L \leq w. \tag{1.4}$$

As a natural consequence, given a price vector p and the total wealth w, we define the *competitive budget set* as: $\mathbb{B}_{p,w} = \left\{x \in \mathbb{R}^{L+} | p \cdot x \leq w\right\}$.

We have now the correct theoretical tools to formalize the consumer's decision making problem. As anticipated consumer's preference are represented by a utility function $u(x)$. Therefore consumer will choose, among all affordable commodity bundles, the one that maximizes his utility. In formula:

$$\textbf{Consumer's Problem}: \quad \max_{x \in \mathbb{B}_{p,w}} \quad u(x). \tag{1.5}$$

1.1.3 The Producer's Problem

On the other side of the economy there is an agent, the producer, who is capable to produce goods to be consumed. This could be, for example, a legally recognized firms. The typical approach consists in modelling the producer as a "black box" able to transform inputs in outputs. This approach completely ignores internal management issues and the underlying process required for the transformation of basic goods in elaborated ones. We do not know how inputs become outputs, but only that a given amount of input is transformed in a known amount of output [2]. The production possibilities are described by a *production vector*:

[1] This is known as the *price-taking assumption.*

[2] This is one of the key point we will criticize. Our approach provide a detailed thermodynamic description of the production process. This indeed allows to investigate how irreversibility impacts classical economic equilibrium.

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_L \end{pmatrix} \in \mathbb{R}^L \tag{1.6}$$

where y_i indicates the amount of the i-th commodity used or produced. We use the convention that negative entries of (1.6) describe input goods while positive quantities indicates output commodities. For example $y = (-2, 1, 0)$ indicates that 2 units of good number one are used to produce 1 unit of the good number two, while commodity three is not used nor produced.

Given a specified production technology not all $y \in \mathbb{R}^L$ are feasible production vectors. In order to understand this limitation one should think to the finite capacity of each production plant and to the minimum inputs required to carry out production. We postulate that there exist a set $\mathbb{Y} \subset \mathbb{R}^L$ such that if $y \in \mathbb{Y}$ then y is possible while any $y \notin \mathbb{Y}$ is not.

We focus on production model with one single output. In this case it is convenient to introduce the *production function* i.e. the maximum level of output produced for a given set of inputs. If $-y_1, ..., -y_{L-1}$, with $y_i \geq 0$, indicate the input quantities and $q = y_L \geq 0$ is the correspondent output then the production function f is such that $q \leq f(y_1, ..., y_{L-1})$. The maximum production is obtained when $q = f(y_1, ..., y_{L-1})$. It is clear that f should describe, in some way, the technology used to carry out production. Suppose that all production inputs are multiplied by a constant factor λ. Three cases may arise:

- $f(\lambda\, y_1, ..., \lambda\, y_{L-1}) = \lambda\, f(y_1, ..., y_{L-1})$, which is referred as *constant return to scale.*
- $f(\lambda\, y_1, ..., \lambda\, y_{L-1}) < \lambda\, f(y_1, ..., y_{L-1})$, which is referred as *decreasing return to scale.*
- $f(\lambda\, y_1, ..., \lambda\, y_{L-1}) > \lambda\, f(y_1, ..., y_{L-1})$, which is referred as *increasing return to scale.*

Once the production technology is specified by the production function the producer tries to maximize its profits:

$$\textbf{Consumer's Problem}: \qquad \max_{y \in \mathbb{Y}} \quad p_1\, y_1 + p_2\, y_2 + ... + p_L\, y_L, \tag{1.7}$$

where $p_1, ..., p_L$ are the price vector components introduced in (1.3). At the equilibrium point we suppose that supply equals demand, this assertion is known as *market clearing condition.* A set of price p that realizes perfect balance between demand and supply is called a *market clearing price.* In formula the clearing condition is written as:

$$\forall i = 1, ..., L \text{ if } y_i \geq 0 \Rightarrow x_i = y_i. \tag{1.8}$$

The condition $y_i \geq 0$ implies that a quantity of the good i is produced and released on the market, therefore equality $x_i = y_i$ resets disparities between demand and supply.

Problems (1.5)-(1.7) must be solved simultaneously with the conditions (1.8).

1.1.4 Two Commodities Models

Suppose we have two commodity ($L = 2$) which are the same physical good produced at two different times, for example a mineral today and the same mineral tomorrow. On the consumer's side we have:

$$\max_{x_1\, p_1 + x_2\, p_2 \leq w} \quad u\left(x_1, x_2\right). \tag{1.9}$$

First order conditions imply that:

$$\frac{\partial u}{\partial x_1} - \lambda\, p_1 \quad = \quad 0, \tag{1.10}$$

$$\frac{\partial u}{\partial x_2} - \lambda\, p_2 \quad = \quad 0, \tag{1.11}$$

where λ is a Lagrange multiplier: $\lambda \neq 0$ when $x_1\, p_1 + x_2\, p_2 = w$ while $\lambda = 0$ when $x_1\, p_1 + x_2\, p_2 < w$. Suppose that the consumer spends all his wealth w on the market, thus $\lambda \neq 0$. From equations (1.10)-(1.11) we find that:

$$p_2 = p_1 \left(\frac{\frac{\partial u}{\partial x_2}}{\frac{\partial u}{\partial x_1}} \right). \tag{1.12}$$

Equation (1.12) states that, at equilibrium, the price of the good 2 is given multiplying the price of good 1 by the factor:

$$B \equiv \frac{\frac{\partial u}{\partial x_2}}{\frac{\partial u}{\partial x_1}}. \tag{1.13}$$

In our case, commodities one and two are the same physical good at different times, thus we can interpret B as a discount factor. On the producer's side of the economy we have:

$$\max_{y \in \mathbb{Y}} \quad p_1\, \left(y_1 + B\, y_2\right), \tag{1.14}$$

having used relations (1.12)-(1.13). Solution of problem (1.14) is independent of p_1 which appears only as a multiplying factor. This is a well-known result of neoclassical economic theory usually referred as *neoclassical paradigm*.

Price is usually called *numeraire* and it expresses the chosen mean of exchange in our economy. Neoclassical paradigm states that the economic equilibrium is indifferent to the choice of the numeraire. We can choose to price commodity in apples, potatoes or a single currency without changing the equilibrium established between consumers and producers.

1.1.5 Isoquants and Substitutions

Suppose that a production function exists. For the sake of simplicity we further require that the production is based on only two inputs $-y_1$ and $-y_2$. They could be capital and labour or energy and raw material to be refined etc. An isoquant at level K is defined as the lieu of points (y_1, y_2) such that $f\left(y_1, y_2\right) = K$. We wonder whether it is possible to preserve the same level of production K allowing one between y_1 and y_2 to go to zero. This problem arises in concrete situations when one of the two inputs is scarce or even null.

In other words: how much quantity of y_1 is necessary to compensate a loss in y_2 (and vice versa) keeping production at the same level? We will say that y_1 and y_2 are *perfect substitute* when every loss in one of the two variables can be compensate with a proper increase in the other. This idea could be extended to an arbitrary high number of input variables. As an example a linear technology like $f(y_1, y_2) = a\, y_1 + b\, y_2$ allows for perfect substitutability of inputs. In literature, most specifications of production functions allow for a complete or a partial substitution between variables.

1.1.6 Some Criticism of Neoclassical Theory

In this chapter we provide an overview of the current research on substitutability among production inputs and production function modelling. Our goal is to show how substitutability among inputs is limited when a thermodynamic approach is taken. We will focus on the production of a refined mineral which is obtained using as inputs energy and raw material. As the thermodynamic intuition suggests, the substitution between these two inputs is limited by physical constraints. The production function derived from a full thermodynamic model incorporates directly these constraints providing a more realistic description of the production process. We further study the consequences of irreversibility on economic equilibrium. We will show that neoclassical equilibrium is drastically changed when irreversibility is taken into account in the production technology. We also find that, when a finite time thermodynamic model is adopted, energy plays an important economic role. Moreover the decision making strategy is completely changed when the production process is reversible i.e. it is carried out in an infinite time. We will show that reversible technology reproduces the same results one would obtain in the neoclassical theory with a linear technology. In this sense ignoring irreversibility is equivalent to allow for unrealistic perfect substitutability among inputs.

1.2 Introduction

The centrality of the role of energy as an input in the production process of economically valuable goods is difficult to dispute. Yet in economics a consensus on how this should be modeled has not emerged. The resource degradation connected with its use suggests energy warrants a specific treatment as a factor of production. Intensive energy use creates environmental damage through the release of residual heath. As discussed in the seminal contribution of Georgescu-Roegen (1971), the second law of thermodynamics, which governs energy utilization and degradation, is a key source of negative externalities. In the production literature a limited amount of attention has been dedicated to the concept of waste as an unavoidable joint product of any production process (Ayres and Kneese, 1969; Ethridge, 1973; Kummel, 1989). Traditional economic analysis of production generally avoids thermodynamic considerations. Typical production models require substitutability between any inputs none of which, including energy, has a special role. On these premises the economics literature dealing with the use of energy has largely focused on the possibilities for substituting it as a factor of production in the presence of energy price shocks or energy shortages. [3] Factors of production (as well as consumption goods) are interchangeable

[3] Abel (1983) proposes a model for the substitution between energy and capital in a generic production function when the energy price is stochastic. A dynamic model is developed by Pindyck (1983) who assumes that the firm has to optimise

if they provide the same functionality. The complete substitutability among natural resources (including energy), labor and capital leads however to paradoxical consequences. Daly (1997) observes that, if labor and natural resources are substitutes and not complements then it would be possible to make a cake with "[...]only the cook and his kitchen. We do not need flour, eggs, sugar, etc., nor electricity or natural gas, nor even firewood. If we want a bigger cake, the cook simply stirs faster in a bigger bowl and cooks the empty bowl in a bigger oven that somehow heats itself[...]". A dazzling example of this paradox can be found in the standard representation of funds and flows in the Cobb-Douglas production function model, that is:

$$Q = K^{\alpha_1} L^{\alpha_2} R^{\alpha_3}, \tag{1.15}$$

where Q is the output of the process per unit of time, K represents the stock of capital, L the labor supply and R the flow of natural resources. From expression (1.15) it is evident that we can obtain a fixed amount of output Q_0 even if $R \to 0$, it is sufficient to choose an amount of K such that:

$$K = \left(\frac{Q_0}{L^{\alpha_2} R^{\alpha_3}} \right)^{\frac{1}{\alpha_1}} \to +\infty. \tag{1.16}$$

Although it is well understood that the Cobb-Douglas production function (as well as its extension) is only an abstract concept and not an actual description of a production process, this very concept has a pervasive influence on economic modeling. Substitutability plays a critical role in the Neoclassical general equilibrium construction. It underlies the view that there are no real limits in the economic growth as, in the extreme, "the world can, in effect, get along without natural resources" Solow (1974). Incidentally, Neoclassical general equilibrium also implies the substitutability of every good as a numeraire and means of payment. Still this is a useful framework which yields insight on the value and optimal exploitation over time of non-renewable resources (Hotelling, 1931).

It may be sensibly argued, as Ayres and Miller (1980) put it, that there are definite and well-known limits on physical performance in almost every field which derive from the unique role of some specific factor of production. Energy is a case in point. See Cleveland and Ruth (1997) for a survey of the literature

four level of inputs: capital, labor, energy and materials. The effect of unpredictable energy price variations is analysed in a real business cycle model by Kim and Loungani (1992), explicitly incorporating energy as an input in a CES production function. A vast number of papers are dedicated to the empirical analysis of capital-energy substitution, which is of "great importance in predicting economic disruptions arising from energy shortages" (Field and Grebenstein, 1980). The papers of Berndt and Wood (1975) and Fuss (1977) highlight a negative value for the capital-energy substitution elasticity, indicating that they are not substitutes, while Griffin and Gregory (1976) and Pindyck (1979) find a positive value. The discrepancy is fully explained in Field and Grebenstein (1980) who find that the difference between the two results lies in the way the capital input is handled. A paper by Magnus (1979) remarks that energy inputs interact differently with labor (in a substitutable way) and capital (in a complementary way) and this justifies the introduction of energy as an independent input in the production function. The work of Atkenson and Kehoe (1980) sheds light on the fact that, with respect to the energy prices, energy use is inelastic when dealing with time-series data and elastic with cross-sectional data. They develop a putty-putty model and a putty-clay model which reproduce these differences in elasticities. Finally, the difference between cross-sectional data and time series data in capital-energy substitution elasticity is resolved by Thompson and Taylor (1995) using Hiroshima elasticities. Arrow et al. (1961) provide an exhaustive analysis of capital-labor substitutability, concluding that the substitution elasticity between capital and labor in manufacturing may typically be less than one.

covering this alternative point of view. However, such analyses are often of a qualitative nature. The case for (the lack of) substitutability can be often argued either way due to the vagueness of the approach.

As a contribution to this debate we provide a robust theoretical foundation for the lack of substitutability of the energy input in the extraction and refinement of a commodity. Specifically, we propose an analytic model for a mining operator who refines a mineral from its natural concentration to a strike concentration which defines the product. Consistently with actual physical constraints the production process is modelled analytically, following the thermodynamic theory of an irreversible separation process.[4].

All real world transformations involving energy are in fact irreversible. A thermodynamic transformation or cycle is said to be reversible if it is carried out by varying a state variable with infinitesimal changes that allow the system to be at rest throughout the entire process (Fermi, 1956). Such a transformation is impossible because it would require an infinite amount of time. All production processes are therefore irreversible in nature, because they have to be carried out in a finite time-period to bring production to the market. They involve then a strictly positive increase in entropy $\Delta S > 0$. An increase in entropy means a reduction in "useful" energy, that is the part of energy that can be converted into work by an engine. Thus an entropy increase can be interpreted as waste or resource degradation. If the production function does not accommodate the real thermodynamic process which leads to the final consumption good, the impact of the producer's choices on resource depletion and the waste released on the environment will not be evident.

The derivation of our production model is driven by the pursuit of the maximum efficiency in the use of available energy so that the least possible amount of it is required. Despite the attempt to minimize its use as an input a positive energy amount is always necessary to extract the natural resource. For the specific case, this lower limit provides an answer to all issues raised by the recent and past literature about substitutability between production process inputs. Although limited substitution of energy is possible, total substitution is impossible because there is a physical energy threshold, for a given quantity of raw material input, below which no production exists. This emphasises the importance of a physics-based approach to production modelling as a correct methodological way of resolving the substitutability issue. A standard Cobb-Douglas model would lead to results in contradiction with the laws of nature (Islam, 1985)[5].

Except for a few papers microeconomic analysis completely avoids a detailed consideration of the physical constraints which are the essence of every production process. [6]. Krysiak and Krysiak (2003) study the impact of the first law of thermodynamics on the standard general equilibrium model of production and consumption. They show that general equilibrium theory is consistent with the mass and energy conservation laws. Krysiak (2006) analyzes the consequences of the second law of thermodynamics on

[4]A number of papers deal with the decisions concerning the exploitation of physical resources and the extraction/production of commodities (Brennan and Schwartz, 1985; Coratzar et al., 1998; Hartwick, 1978; Stiglitz, 1976) but none describes the thermodynamics of extraction.

[5]A list of economic key-objections and economic advantages of the Cobb-Douglas function can be found in Murthy (2002).

[6]Sav (1984) uses a micro-engineering production function derived from physical laws to model the exploitation of solar energy for domestic water heaters. Substitution elasticities between nonrenewable fuel inputs (oil or natural gas) and capital-intensive solar-produced heat is investigated. In this context Sav (1984) finds that a governmental taxation policy for solar energy incentives could paradoxically result in an increase in consumption of nonrenewable energy resources.

economic equilibrium in a general framework, arguing that if irreversibility is taken into account a non-zero level of inputs and a non zero-level of emissions are necessary to sustain a positive level of consumption[7]. A production function derived from finite time thermodynamic constraints can be found in Roma (200), Roma (2006), who proposes a general equilibrium model for the production of a basic good (hot water) based on a fully consistent thermodynamic description of the process. The production process we analyze here from a physical point of view differs from the hot water production process in Roma (2006) in that the final product obtained will be of homogeneous quality and unused energy can be stored.

An important precedent for our commodity extraction problem is Ruth (1995). This paper contains explicitly thermodynamic constraints on production in the form of lower bounds on the inputs to be fed into a Cobb-Douglas production function. The lower energy bound is calculated for a reversible separation process. As mentioned above, this approach does not reflect the reality of production, which involves only irreversible processes. The degradation of energy is far higher in irreversible processes than in the reversible ones and irreversible processes make a greater contribution to entropic increase, as stated by the second law of thermodynamics.

We then turn to the analysis of the rational use of energy from an economics point of view. Berry et al. (1978) highlight the difference between the concepts of thermodynamic optimality of the use of energy in production and overall economic (cost) efficiency. The economic notion of Pareto optimality for resource allocation (where no individual can be better off without making al least one other worse off) will not in general amount to the optimal use of available energy from a thermodynamic point of view. If energy is in aggregate scarce and becomes a limiting factor because of its lack of substitutability (as argued by Ayres and Miller, 1980) the two criteria may indeed coincide.

We are able to carry further the analysis in Berry et al. (1978). They do not model the thermodynamic properties of the technology and assume that substitution of energy by other inputs along a production isoquant amounts to an increase in thermodynamic efficiency. As thermodynamic efficiency cannot be increased beyond the reversible limit full substitutability of energy is prevented, i.e. it is impossible to move beyond a certain point on the isoquant in the energy dimension.

We can analytically derive instead the production isoquants under the assumption of maximum finite time thermodynamic efficiency (a possibility dismissed by Berry et al. (1978) as optimistic, see p.133). Hence in our world engineers make sure that production is achieved in any case with maximum thermodynamic efficiency. This is still compatible with some (quite limited) amount of substitutability between anergy and raw material.

We analyse the problem of the optimal scale of production and the consequent exploitation of natural resources, including energy, under the finite time thermodynamic foundation adopted. Even if the the most efficient thermodynamic technology is used higher production in the same amount of time implies greater deviation from reversible efficiency and higher energy waste. Hence it is the scale of production which determines in the end thermodynamic efficiency.

[7]On the debate about the impact of the entropy law on economic equilibrium see also Young (1991) and Daly (1992) which lead to two completely opposite conclusions. Many authors have investigated Nicholas Georgescu-Roegen's paradigm of ecological economics, again obtaining conflicting conclusions, such as those reported by Khalil (1990), subsequently criticized by Lozada (1991) and finally re-stated by Khalil (1991).

We incorporate the production model in a simple general equilibrium framework in which we analyze production decisions. We derive the equilibrium in a simple economy in which our good is produced and consumed and address in this context the role of energy efficiency in optimal economic decisions. In a neoclassical equilibrium the optimal allocation of available resources will follow from the non-satiated demand for final goods and the efficiency in the use of available energy will not drive on its own economic decisions. Under a Neoclassical model the amount of thermodynamic waste will be irrelevant and natural resources will be fully exploited. On the other hand, the negative externalities associated with the degradation of energy, if taken into account, would lead to a thermodynamic efficiency criterion in the use of this scarce resource. However, thermodynamic efficiency would have to be imposed on the decision makers (by way, for example, of a "green tax"). We find, similarly to Roma (2006), that a competitive economic equilibrium will be twisted towards higher thermodynamic efficiency if energy is forced to be the numeraire and means of payment. This creates a market in which the energy price of production is established. If energy is the only scarce factor the producer will compare the energy cost of production with the energy value of the firm's sales determined by supply and demand. The lack of substitutability between input factors and the decreasing return to scale feature of irreversible technology imply that part of the energy available in the economy will not be used in production. This will in turn decrease energy degradation and thermodynamic waste. The neoclassical solution, where a change in numeraire and means of payment would not alter the neoclassical equilibrium, is finally obtained if the production process is carried out over an infinite time, i.e. if it is reversible. Reversibility is equivalent to constant returns to scale technology.

This chapter is organized as follows: Section 1.3 is dedicated to a description of the physical process underlying the production of the commodity. Here the thermodynamic production function is derived together with the marginal and average costs of production. Isoquants of the computed production function are compared with those derived from a Cobb-Douglas type production function, in order to highlight the differences between the two approaches. Section 1.3.2 briefly describes the idea by Ruth (1995) of a thermodynamic limit on production inputs. The problem of a producer who uses a thermodynamic production function and faces scarcity of energy is presented in Section 1.5. After introducing the concept of Energy Return on Energy Investment (EROEI), energy will be used as the numeraire and means of exchange in our economy. It is shown that, when energy is scarce, and it cannot be substituted, the producer's decision varies drastically if the accounting is made in terms of energy or in terms of another numeraire (such as the product itself). The consumer's problem is also described in this section. Section 1.7 shows that arbitrage arguments can be stated in a perfectly reversible context to compute the price of the commodity in terms of energy, while the same arguments weaken if irreversibility arises. The producer's choices under a reversible technology are discussed in Section 1.8. Finally, in Section 1.9, we present our conclusions.

1.3 The Physical Model

Thermodynamic energetic limits to industrial processes have been analyzed by a vast number of authors. The survey of Sieniutycz (2003) provides an exhaustive analysis of the most commonly used methods

for computing thermodynamic energy limits to industrial production, such as separation processes, heat pumps, chemical and electrochemical systems, maximum power in thermal engines, etc. Energy limits in industrial processes are investigated from a classical thermodynamic approach by Forland et al. (1988), Kim and Loungani (1985), and Denbigh (1956). Engineers typically face the problem of how to set-up a production plant capable of achieving non-zero production in a finite time and with finite industrial facilities. In most cases this kind of problem is solved using one of two alternative approaches: finite time thermodynamic (Andersen et al., 1977; Berry et al., 2000) or entropy generation minimization (Bejan, 1996; Nummedal et al., 2005; Salamon et al., 1980), which are usually referred as FTT and EGM, respectively. The idea underlying these approaches is that energy is a scarce resource and its use must be optimized.

The physical model adopted for the construction of our production function is developed by Tsirlin and Titova (2004). They compute (via EGM) the minimum energy required to achieve the separation of an ideal mixture of components with a specified output.

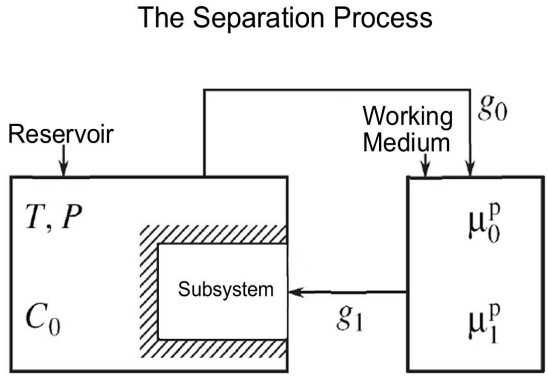

Figure 1.1: The scheme used by Tsirlin and Titova (2004) to model the separation of a finite subsystem from an infinite reservoir. The thermodynamic state of the reservoir is identified by the temperature T, the pressure P and the composition vector C_0. The pumps g_0 and g_1 are the mass-transfer coefficients from the reservoir to the working medium and from the working medium to the reservoir respectively. The working medium has a chemical potential μ_0^P at the contact point with the reservoir and a chemical potential μ_1^P at the contact point with the subsystem.

Figure 1.1 illustrates the scheme used to model the process. A subsystem to be separated out is in contact with an infinite reservoir characterised by a temperature T and a pressure P which do not vary during the transformation. This may be interpreted as a stylized production plant in which a valuable substance with pre-specified purity must be refined and extracted from the original mixture of substances which are found in a natural reservoir. The extracted substance is accumulated in the subsystem.

The mining operator faces an initial mixture of useful mineral and waste rock. We suppose that the mineral is initially a minority presence with respect to the waste rock and the mining operator tries to purify the mixture and extract highly concentrated mineral[8].

[8]A number of papers deal with the decisions concerning the exploitation of physical resources and the extraction/production of commodities (Brennan and Schwartz (1985), Coratzar et al. (1998), Hartwick (1978), Stiglitz (1976))

At time $t = 0$ the composition vector of the reservoir, which describes the (percentage) concentrations of the different mixture components, is $(C_{01}, ..., C_{0n})$. The subsystem is initially at equilibrium with the reservoir and is therefore characterised by the same thermodynamic variables T, P and by the same composition vector $(C_{01}, ..., C_{0n})$ as the reservoir. We indicate with N_0 the number of moles in the subsystem at $t = 0$.

Separation processes and chemical reactions are usually described by means of chemical potentials. We recall that the chemical potential of a substance in a system represents the variation of the system's internal energy when a unit of the substance is added, keeping system's volume and entropy constant.

Each component of the mixture has a chemical potential $\mu_{0i}(T, P)$ *in the reservoir* which depends on its concentration C_{0i} and is given by:

$$\mu_{0i}(T, P) = \mu_0(T, P) + RT \log C_{0i}, \tag{1.17}$$

where R is the universal gas constant and $\mu_0(T, P)$ represents the reservoir chemical potential which does not vary during the transformation. We can interpret $\mu_0(T, P)$ as the ground level from which all other chemical potentials are measured. As it is only a scale variable we could impose $\mu_0(T, P) \equiv 0$. However we choose to leave $\mu_0(T, P)$ unspecified for the sake of generality.

The subsystem and the reservoir are in contact with a third medium, a working medium, which has the chemical potentials μ_0^P and μ_1^P at the contact points with the reservoir and the subsystem respectively. This is a short-hand notation to indicate that the i-th component has a chemical potential μ_{0i}^P at the contact point with the reservoir and μ_{1i}^P at the contact point with the subsystem. A working medium is simply defined as a system whose chemical potentials are under control. When in contact with the reservoir and the subsystem, the working medium generates chemical potential drops between the contact points. Particles tend to diffuse from regions of high chemical potential to regions of low chemical potential, allowing for the mass transfer.

Examples of mixture separation with the formalism of chemical potentials can be found in Wijmans and Baker (1995) and Vallieres et al. (2003). As observed by Wijmans and Baker (1995) a chemical potential variation $d\mu_i$ of the i-th mixture component across a membrane of thickness dx produces a mass flux given by:

$$J_i = -L_i \frac{d\mu_i}{dx}, \tag{1.18}$$

where L_i is a coefficient of proportionality. All driving forces such as gradients in pressure, temperature, concentration and electromotive forces can be reduced to chemical potential gradients and, therefore, they generate mass fluxes which can be expressed in the form (1.18).

Through this mechanism the working medium creates a mass flux g_{0i} of the substance i between the reservoir and the working medium, and a mass flux g_{1i} of the substance i between the working medium and the subsystem. All these fluxes have the dimension of mole per time: $[g] = [\text{mol}]\,[\text{time}]^{-1}$. Following the approach of Onsager (1930), Onsager (1931) and Miller (1994), it is assumed that the system is close to equilibrium. More precisely, as described in Miller (1994), "one divides the system

but none describes the thermodynamics of extraction.

into small subsystems and assumes that each subsystem is in local equilibrium, i.e., it can be treated as an individual thermodynamic system characterized by the small number of equilibrium variables.". This common assumption leads to linear irreversible thermodynamics, in which the relation between chemical drops and mass transfer coefficient is linear (**Onsager's Kinetics**).

At the final instant $t = \tau$ the subsystem has a new specified composition vector $(C_{\tau 1}, ..., C_{\tau n})$ and contains a number of moles N_τ. Following the reasoning in Tsirlin and Titova (2004), the separation work in an isothermic process (i.e. with temperature kept constant) for an adiabatically insulated system (i.e. heat exchanges are not allowed) is given by the Gouy-Stodola formula:

$$E = E_0 + T\,\Delta S, \tag{1.19}$$

where E_0 is the reversible work and ΔS the entropy increment. The reversible work is computed summing the total internal energy variation of the reservoir to that of the subsystem. The internal energy variation of the reservoir is easily computed observing that a number of moles $\Delta N = N_\tau - N_0$ leaves the reservoir. The component concentrations *in the reservoir* do not change during the process, the reservoir being infinitely large. Therefore for each mixture component i we have that a number of moles $\Delta N\, C_{0i}$ leave the reservoir. According to the definition of chemical potential the total internal energy variation of the reservoir ΔU_{Res} is given by:

$$\Delta U_{\text{Res}} = \sum_{i=1}^{n} (-\Delta\, N)\; C_{0i}\, \mu_{0i}\,(T, P)\,. \tag{1.20}$$

Using relation (1.17) and recalling that $\sum_{i=1}^{n} C_{0i} = 1$ we obtain:

$$\Delta U_{\text{Res}} = \left[(-\Delta N)\; \mu_0\,(T, P) + R\,T \sum_{i=1}^{n} (-\Delta N)\; C_{0i}\, \log C_{0i} \right]. \tag{1.21}$$

Within the subsystem the initial number of moles of each component is $N_{0i} = N_0\, C_{0i}$ and after the separation process has operated for a period of time τ the final number of moles becomes $N_{\tau i} = N_\tau\, C_{\tau i}$. At time $t = 0$ the subsystem is in equilibrium with the reservoir, therefore each mixture component has an initial chemical potential given by (1.17). At the final instant τ, the chemical potential of the i-th component in the subsystem will be $\mu_0\,(T, P) + R\,T\, \log C_{\tau i}$.

Therefore the total internal energy variation ΔU_{Sub} of the subsystem is given by:

$$\Delta U_{\text{Sub}} = \sum_{i=1}^{n} \left[N_{\tau i}\, \mu_{\tau i} - N_{0i}\, \mu_{0i} \right] \quad = \tag{1.22}$$

$$\sum_{i=1}^{n} \left[N_\tau\, C_{\tau i}\; (\mu_0\,(T, P) + R\,T\, \log C_{\tau i}) - N_0\, C_{0i}\; (\mu_0\,(T, P) + R\,T\, \log C_{0i}) \right] \quad = \tag{1.23}$$

$$\mu_0\,(T, P)\; \Delta N + R\,T\, \left[N_\tau \sum_{i=1}^{n} C_{\tau i}\, \log C_{\tau i} - N_0 \sum_{i=1}^{n} C_{0i}\, \log C_{0i} \right], \tag{1.24}$$

where we have used the conditions $\sum_{i=1}^{n} C_{0i} = \sum_{i=1}^{n} C_{\tau i} = 1$. The sum $\Delta U_{\text{Res}} + \Delta U_{\text{Sub}}$ returns the total reversible energy required for the separation process:

$$E_0 = \Delta U_{\text{Res}} + \Delta U_{\text{Sub}} = N_\tau\, R\,T \sum_{i=1}^{n} \left[C_{\tau i}\, \log C_{\tau i} - C_{0i}\, \log C_{0i} \right], \tag{1.25}$$

and it is independent of N_0. Note that, as expected for a state-function, formula (1.25) only depends from the initial and final states of the system, $C_{\tau i}, C_{0i}$ i.e. it is *path-independent*[9]. For this reason a minimum entropy increment ΔS corresponds to the minimum value for E in (1.19). The entropy increment corresponding to the mass transfer coefficients g_{0i} is given by:

$$\Delta S = \frac{1}{T}\int_0^\tau \sum_{i=1}^{n} \left[g_{0i}\ \left(\mu_{0i} - \mu_{0i}^P\right) + g_{1i}\ \left(\mu_{1i}^P - \mu_{1i}\right)\right] dt = \frac{1}{T}\int_0^\tau \sum_{i=1}^{n} \left[g_{0i}\,\Delta\mu_{0i} + g_{1i}\,\Delta\mu_{1i}\right] dt, \tag{1.26}$$

where μ_{1i} is the chemical potential of the $i-th$ component in the subsystem. The parameters of the working medium do not change within a cycle, therefore:

$$i = 1, ..., n. \qquad \int_0^\tau g_{0i} dt = \int_0^\tau g_{1i} dt. \tag{1.27}$$

Moreover the mass balances impose that:

$$i = 1, ..., n. \qquad \int_0^\tau g_{1i} dt = N_\tau\, C_{\tau i} - N_0\, C_{0i}. \tag{1.28}$$

The problem of minimizing the energy (1.19) is then reduced to the following one:

$$\begin{aligned} \min_{g_{0i}, g_{1i}} \quad & \frac{1}{T}\int_0^\tau \sum_{i=1}^{n} \left[g_{0i}\,\Delta\mu_{0i} + g_{1i}\,\Delta\mu_{1i}\right] dt, \\ sub \quad & \int_0^\tau g_{ji} dt = N_\tau\, C_{\tau i} - N_0\, C_{0i}, \\ & j = 0, 1. \quad i = 1, ..., n. \end{aligned} \tag{1.29}$$

In the case of an Onsager's kinetics (see discussion above) we have that $g_{ij} = \alpha_{ij}\Delta\mu_{ij}$, and it can be shown that (Tsirlin and Kazakov, 2004; Tsirlin et al., 1998) optimal solutions of problem (1.29), under this additional assumption, are constant mass transfer coefficients. From conditions (1.27)-(1.28) we immediately obtain:

$$j = 0, 1. \quad i = 1, ..., n. \qquad g_{ji} = \frac{N_\tau\, C_{\tau i} - N_0\, C_{0i}}{\tau}. \tag{1.30}$$

In order to put formulas in a simpler form we define the total mass variation of the i-th component:

$$\Delta\left(N\, C_i\right) \stackrel{\text{def}}{\equiv} N_\tau\, C_{\tau i} - N_0\, C_{0i}, \tag{1.31}$$

and we introduce the equivalent mass-transfer coefficient:

$$\bar{\alpha}_i \stackrel{\text{def}}{\equiv} \frac{\alpha_{0i}\,\alpha_{1i}}{\alpha_{0i} + \alpha_{1i}}. \tag{1.32}$$

Substituting optimal solutions (1.30) in expression (1.26) we obtain the entropy variation. It follows that the minimum energy required to carry out the transformation $[N_0, (C_{01}, ..., C_{0n})] \rightarrow [N_\tau, (C_{\tau 1}, ..., C_{\tau n})]$ in a finite time τ is:

[9] See Fermi (1956).

$$E_{min} = R\,T\,N_\tau \sum_{i=1}^{n} [C_{\tau i}\,\log C_{\tau i} - C_{0i}\,\log C_{0i}] + \frac{1}{\tau}\sum_{i=1}^{n}\frac{[\Delta\,(N\,C_i)]^2}{\bar{\alpha}_i}. \tag{1.33}$$

In the simplest case the valuable substance may be assumed to be mixed with only one other, so the initial vector of concentrations is $(C_0, 1-C_0)$. For example, C_0 can be the concentration of a mineral in the mineral ore and thus $1-C_0$ can be interpreted as the waste rock concentration.

Differently from Ruth (1995), who adopts the extreme assumption that the mineral can be completely separated from the waste rock, we generalize the definition of the final product and define it more realistically as mineral extracted with a given strike concentration $k > C_0$ but still mixed with a residual concentration $1-k$ of waste rock. With $n = 2$, $(C_{01}, C_{02}) = (C_0, 1-C_0)$ and $(C_{\tau 1}, C_{\tau 2}) = (k, 1-k)$. Therefore (1.33) takes the form:

$$\begin{aligned} E = R\,T\,N_\tau\,[k\,\log k + (1-k)\,\log(1-k) - C_0\,\log C_0 - (1-C_0)\,\log(1-C_0)] + \\ + \frac{1}{\tau\,\alpha}\left[(N_\tau\,k - N_0\,C_0)^2 + (N_\tau\,(1-k) - N_0\,(1-C_0))^2\right], \end{aligned} \tag{1.34}$$

where we have imposed, for simplicity, $\bar{\alpha}_1 = \bar{\alpha}_2 \equiv \alpha$. The second term on the right hand side of expression (1.34) denotes the irreversible use of energy. It involves the square of the difference $N_\tau\,k - N_0\,C_0$ between the refined moles extracted and the moles of pure material in the subsystem, and the corresponding difference $N_\tau\,(1-k) - N_0\,(1-C_0)$ for the waste material.

As a final remark note that the optimal coefficients g_{0i}, g_{1i} define the speed at which the extraction plant must be operated in order to waste the minimum possible amount of useful energy. This type of efficiency is clearly desirable for any production plant. However it clearly limits the speed of production, although it is conceivable to operate the plant faster if energy were not a scarce resource. On the other hand the scale of production, which is given by N_τ , is a free choice. The energy required for the separation is a monotonically increasing function of N_τ. For a given level of output N_τ the production technique defined by (1.34) leaves a choice between extracting the actual amount at the desired level of concentration or, where possible, "averaging up" the concentration of an amount N_0 of the original mixture already contained in the subsystem. This is the source of the limited substitutability between N_0 and E. As a special case of (1.33) the minimum energy required to separate completely the two components contained in N_0 moles of a binary mixture is (Tsirlin and Titova, 2004):

$$E_m^{bin} = -R\,T\,N_0\,[C_0\,\ln C_0 + (1-C_0)\,\ln(1-C_0)] + \frac{N_0^2}{\tau}\left(\frac{C_0^2}{\bar{\alpha}_1} + \frac{(1-C_0)^2}{\bar{\alpha}_2}\right), \tag{1.35}$$

where $\bar{\alpha}_1$ and $\bar{\alpha}_2$ are the parameters defined in (1.32).

1.3.1 The reversible limit

The reversible increment in the internal energy of the system which is required for production can be computed taking the limit for $\tau \to +\infty$ of equation (1.34):

$$E_m^R = R\,T\,N_R\,[k\,\log k + (1-k)\,\log(1-k) - C_0\,\log C_0 - (1-C_0)\,\log(1-C_0)], \tag{1.36}$$

where N_R are the number of moles produced reversibly. Equation (1.36) says that the transformation $[C_0, 1-C_0] \rightarrow [k, 1-k]$ requires a positive amount of energy for the pairs (C_0, k) such that:

$$\xi_0 \equiv k \log k + (1-k) \log (1-k) - C_0 \log C_0 - (1-C_0) \log (1-C_0) > 0. \tag{1.37}$$

We assume that our mineral is initially available at a minority concentration, namely $C_0 < \frac{1}{2}$. We try to purify the substance in order to obtain a strike concentration k such that $1 - C_0 < k$, which is greater than the initial concentration of the waste rock. In this framework $\xi_0 > 0$ and the number of moles produced when $\tau \rightarrow +\infty$ is:

$$N_R = \frac{E}{R\,T\,\xi_0}. \tag{1.38}$$

1.3.2 Ruth (1995) Model of Mining Technology

Ruth (1995) presents the problem of the mining operator as the maximisation of a value function essentially defined as the time-integral of production growth minus the growth in energy expenditure for production. Hence his model assumes that the energy input is the only cost of production. To model production, Ruth (1995) uses a Cobb-Douglas type function given by:

$$Y_t = \left(\frac{J_t - J_t^*}{J_A - J_A^*} \right)^{\gamma_1} \left(\frac{E_t - E_t^*}{E_A - E_A^*} \right)^{\gamma_2}, \tag{1.39}$$

where J_t is the raw material input, E_t is the energy input and the starred quantities J_t^* and E_t^* are the minimum material input and minimum energy input required for the production of Y_t, respectively. With J_A and E_A (J_A^* and E_A^*), Ruth (1995) indicates the same quantities in a base year.

The Cobb-Douglas production function (1.39) tries to capture the limits to the substitution between energy and raw material that are necessary to produce the quantity of the final good Y_t. The lower limit E_t^* is derived as the reversible limit of (1.35) for $\tau \rightarrow +\infty$ (and time varying concentration C_0). Still, E_t^* and J_t^* may be partially substituted.

Our approach will be completely different: we will derive the production function Y_t from finite thermodynamic considerations alone. In this approach the limits to the substitution between energy and raw material are represented by an analytical thermodynamic constraint. This production function has very different properties from a Cobb-Douglas function.

1.4 The Production Model

Equation (1.34) implicitly defines a production function $N_\tau\,(N_0, E)$, where the inputs are energy, E, and raw material, N_0, and where production time, τ, also determines the amount of final product that can be obtained. We can re-arrange expression (1.34) as:

$$N_\tau^2 \frac{1}{\tau\,\alpha} \left[k^2 + (1-k)^2 \right] + N_\tau \left[R\,T\,\xi_0 - 2\,\frac{N_0\,\psi_0}{\tau\,\alpha} \right] + \frac{N_0^2}{\tau\,\alpha} \left[C_0^2 + (1-C_0)^2 \right] - E = 0, \tag{1.40}$$

where the constants $\xi_0 \equiv k \log k + (1-k) \log (1-k) - C_0 \log C_0 - (1-C_0) \log (1-C_0)$ and $\psi_0 \equiv k\,C_0 + (1-k)\,(1-C_0)$ depend only on the initial and final concentrations. The solutions of (1.40) are:

$$N_\tau^{\pm} = \frac{-RT\,\xi_0 + 2\,Q_\tau\,N_0\,\psi_0 \pm \sqrt{(RT\,\xi_0)^2 - 4\,Q_\tau^2\,(k-C_0)^2\;N_0^2 - 4\,RT\,Q_\tau\,N_0\,\psi_0\,\xi_0 + 4\,G_k\,Q\,E}}{2\,Q_\tau\,G_k}, \quad (1.41)$$

where we have introduced the notation $Q_\tau = \frac{1}{\tau\,\alpha}$ and $G_x = x^2 + (1-x)^2$. A solution to the problem exists if, and only if, the quantity under the square root of equation (1.41) is greater than or equal to zero:

$$\delta = (RT\,\xi_0)^2 - 4\,Q_\tau^2\,(k-C_0)^2\;N_0^2 - 4\,RT\,Q_\tau\,N_0\,\psi_0\,\xi_0 + 4\,G_k\,Q\,E \geq 0. \quad (1.42)$$

This condition implies that there is a limited substitutability between process inputs and defines a lower bound on the energy input:

$$E \geq \frac{1}{4\,Q_\tau\,G_k}\left[N_0^2\,4\,Q_\tau^2\;(k-C_0)^2 + N_0\,(4\,RT\,\xi_0\,\psi_0\,Q_\tau) - (RT\,\xi_0)^2\right]. \quad (1.43)$$

Inspecting equation (1.43), we can see that if $k \neq C_0$ and $N_0 > 0$, a lower bound on energy always exists for a high enough positive value of N_0. In our model of production $k > C_0$, because we want to improve the quality of the valuable substance, and $N_0 > 0$ because we assume that our basin is not empty at the initial time[10]. This means that we start production only if the energy is over the threshold given by equation (1.43). As pointed out by Islam (1985), such restrictions on production inputs drastically modify the isoquants with respect to those of a Cobb-Douglas function.

Only the solution of (1.41), which increases in the energy input, has a physical sense. We then define our production function as:

$$N_\tau\,(E, N_0) \equiv \frac{-RT\,\xi_0 + 2\,Q_\tau\,N_0\,\psi_0 + \sqrt{(RT\,\xi_0)^2 - 4\,Q_\tau^2\,(k-C_0)^2\;N_0^2 - 4\,RT\,Q_\tau\,N_0\,\psi_0\,\xi_0 + 4\,G_k\,Q\,E}}{2\,Q_\tau\,G_k}. \quad (1.44)$$

Expression (1.44) explicitly gives the number of moles of the highly refined mixture $(k, 1-k)$ we obtain starting from N_0 moles of raw material and spending energy E for production in a finite time τ.

Figure 1.2 shows the isoquants of $N_\tau\,(E, N_0)$ for a chosen specification of the model parameters (reported in Table 1.1). The chosen values have no particular significance, they have been chosen only for a graphical representation.

The isoquants of (1.44) display an "economically inefficient" portion, where the same level of production is obtained with more N_0 than necessary. We can disregard this portion.

Isoquants derived from a Cobb-Douglas type production function are also reported as thin dotted lines. The plotted Cobb-Douglas function has the form $N^{C-D}\,(E, N_0) = A\,\left(\frac{E}{RT}\right)^{\gamma_1}\;N_0^{\gamma_2}$ and A, γ_1 and γ_2 are chosen to obtain nearly the same levels of $N_\tau\,(E, N_0)$ isoquants.

Using $N^{C-D}\,(E, N_0)$ as a production function would generate perfect substitutability between the two production inputs, that is given a point $\left(N_0^{(1)}, E^{(1)}/\,(RT)\right)$ which belongs to an isoquant $N^{C-D}\,(N_0, E) = K_1$ it is possible to obtain the same level of production K_1, arbitrarily reducing the quantity of energy entering the process and increasing the number of moles N_0 of raw material.

[10] As we will show later, energy-efficient production requires a strictly positive value of N_0 which is proportional to the level of production (see equation 1.45).

Parameter	Value
T	$1250\,K$
k	80%
C_0	30%
τ	$3600\,sec$
α	$10^{-4}\,mol^2/\,(J\,sec)$

Table 1.1: A set of chosen parameters.

This property disappears when the production function is obtained via the real thermodynamics process. The constraint (1.43) does not allow perfect substitutability between inputs. This gives a more realistic description of the process and recognizes the different nature of energy as a production input.

Expression (1.44) provides a theoretically consistent production function for the extraction of refined mineral from a low grade mixture. It naturally incorporates a lower bound on the energy input in the spirit of Ruth (1995) and, as anticipated, makes the lack of substitutability between inputs completely clear.

However, we find that the lower bound on the quantity of low grade mixture N_0 is zero. In what follows, we assume that there is an unlimited quantity of the raw material N_0 and it is therefore freely available to the producer, while energy is scarce. In this we follow Ayres and Miller (1980), who argue that basic materials will always be available in some concentration in the earth's crust, but the availability of energy (work) needed to extract them is, in fact, the only limit to economic growth[11].

[11]Georgescu-Roegen (1979) argued that the degradation of quality materials is, in fact, the real limit to the economic growth, rather than the scarcity of energy itself. This approach is criticised by Ayres and Miller (1980), who argue that technical progress can overcome the scarcity of physical resources. They point out that a finite quantity of resources must always be embodied in capital and a limit on economic growth and technical efficiency always exists, due to the finite availability of renewable resources. Their point is that all resources, no matter how they are distributed on the earth, can be extracted if enough energy is available. Thus they conclude that energy is, in fact, the only resource that could ultimately limit economic growth.

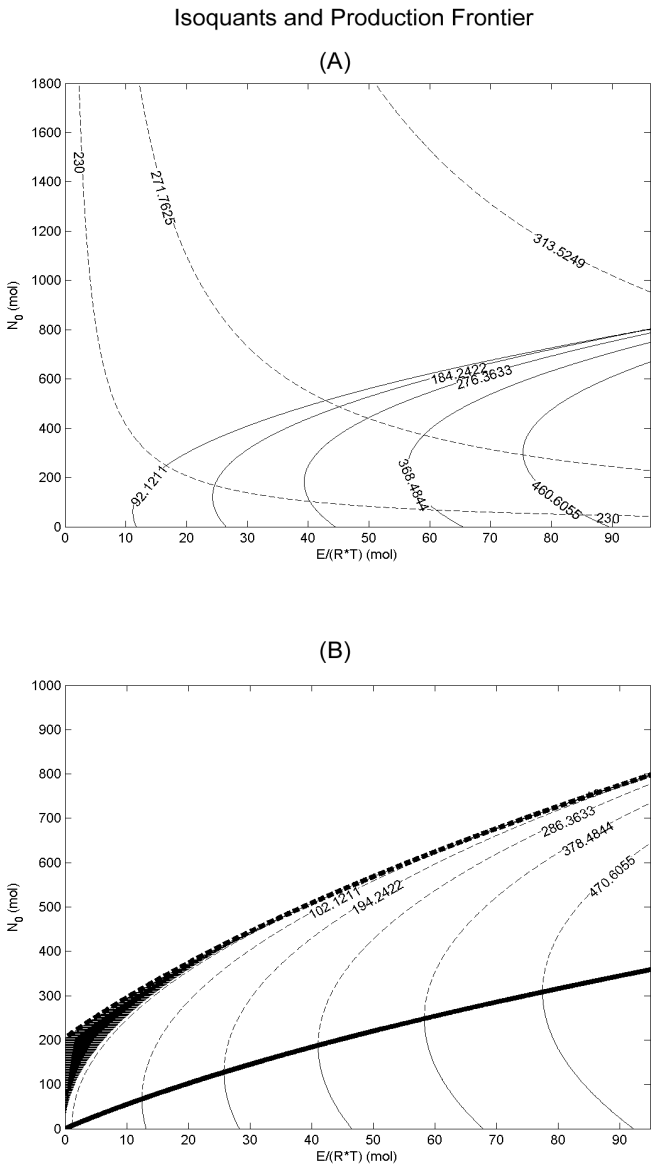

Figure 1.2: (A) The dotted lines are isoquants of a Cobb-Douglas type production function $N^{C-D}(N_0, E) = A\left(\frac{E}{RT}\right)^{\gamma_1} N_0^{\gamma_2}$, with $A = 100$ and $\gamma_1 = \gamma_2 = 0.1$. Continuous lines represent isoquants of the proposed production function $N_\tau(E, N_0)$, where the model parameters are those reported in Table 1.1. (B) The plot shows the efficient portion of the production function $N_\tau(E, N_0)$ isoquants as black continuous lines. The parameters chosen are those reported in Table 1.1. The thin dotted lines are the economically inefficient isoquants of $N_\tau(E, N_0)$. The bold dotted line represents the production frontier (given by equation 1.43), which defines the minimum energy required to carry out production for a given value of N_0. The continuous bold line plots the optimal choice of the freely available raw material N_0 for each value of the energy E (see equation 1.45). The intersection between an isoquant and this line gives the number of moles N_0 that should be used to achieve the chosen level of production with the minimum energy expenditure. The area colored in black is a zone where, despite energy being over the depicted threshold, it is not enough to have positive solutions of equation (1.40).

1.4.1 Minimum Energy Production Frontier

If energy is the only scarce input, it is rational to minimize its use and substitute it with the freely available raw material as far as possible. Inspecting figure (1.2) it can be seen that, for a given level of output, there is a value of the input raw material which minimizes the energy required to obtain such a level of production. This value can be computed analytically. From equation (1.34) it can easily be seen that the level of N_0 which minimizes the energy input for a fixed level N_τ of production is given by:

$$N_0\left(N_\tau\right) \quad = \quad N_\tau \frac{1 - C_0 - k + 2\,C_0\,k}{1 - 2\,C_0 + 2\,{C_0}^2}. \tag{1.45}$$

By substituting (1.45) into the equation (1.34) and solving it with respect to N_τ, we obtain the production frontier where every level of production is achieved at an absolute minimum energy cost:

$$N_\tau\left(E\right) \quad = \quad \frac{\sqrt{4\,E\left(C_0 - k\right)^2 Q_\tau\, G_{C_0} + \left(G_{C_0}\,R\,T\,\xi_0\right)^2} - G_{C_0}\,R\,T\,\xi_0}{2\left(C_0 - k\right)^2 Q_\tau}. \tag{1.46}$$

Equation (1.46) defines the maximum amount of mineral refined to the given concentration k that can be extracted in a period of time τ given the energy input E. Under our assumptions a fixed proportion of raw material and energy is required, for each level of production, which only depends on the scale of production and time available for the production process.

Solving relation (1.46) w.r.t. E we find that:

$$E\left(N\right) \quad = \quad \frac{\left(C_0 - k\right)^2 Q_\tau\, N^2 + G_{C_0}\,R\,T\,\xi_0\,N}{G_{C_0}}. \tag{1.47}$$

The energy $E\left(N\right)$ in (1.47) can be interpreted as a conditional energy demand (input) when the other factor is at the optimal level. From expression (1.47) we can compute the marginal and average costs of production in terms of energy, $\frac{\partial E}{\partial N}$ and $\frac{E}{N}$:

$$\frac{\partial E}{\partial N} = \frac{2\left(C_0 - k\right)^2 Q_\tau\, N + G_{C_0}\,R\,T\,\xi_0}{G_{C_0}}, \tag{1.48}$$

$$\frac{E}{N} = \frac{\left(C_0 - k\right)^2 Q_\tau\, N + G_{C_0}\,R\,T\,\xi_0}{G_{C_0}}. \tag{1.49}$$

Within the limit $\tau \to +\infty$ we have that $Q_\tau \to 0$ and then:

$$\lim_{\tau \to +\infty} \frac{\partial E}{\partial N} = \lim_{\tau \to +\infty} \frac{E}{N} = R\,T\,\xi_0. \tag{1.50}$$

This simple result could be also obtained taking the limit $\tau \to +\infty$ of equation (1.34). Equation (1.50) means that if the process is reversible then the technology is linear, that is marginal and average costs are the same technological constant $R\,T\,\xi_0$.

1.4.2 Two-Period Production Technology

We assume that the production process by which the mineral is extracted occurs over two periods of time. First we allow our extraction plant to work in a basin, for a period τ_1 , with a number of initial moles N_0 of a mixture $(C_0, 1-C_0)$, where the first component of the mixture is the useful mineral and the second component is waste rock. We suppose that the concentration of the useful material is lower than that of the waste rock, that is $C_0 < 1-C_0$. The final consumption good is defined by the mixture $(k, 1-k)$, with $k > 1-C_0$. If the mining operator produces N_1 moles of the consumption good in a time τ_1 he needs the energy given by equation (1.47).

In the first time period the producer does not completely deplete the basin, allowing for additional production to be carried out in a second time period, τ_2, which we assume to be equal to τ_1.

We suppose that the depletion of the resource is neglectable, i.e. the concentration of the valuable substance is kept unchanged for the two periods of extraction. A more realistic model should take into account for the ore degradation. Resource depletion is discussed in Roma and Pirino (2008). We omit this refinement of the model because all results are not influenced by the presence of degradation.

Therefore at the end of the second period the mining operator produces a number of moles:

$$N_2(E_2) = \frac{\sqrt{4\,E_2\,(C_0-k)^2\,Q_\tau\,G_{C_0} + (G_{C_0}\,R\,T\,\xi_0)^2} - G_{C_0}\,R\,T\,\xi_0}{2\,(C_0-k)^2\,Q_\tau}, \tag{1.51}$$

where E_2 indicates the energy spent for production in the second period. Note that accounting for resource depletion would introduce a non-linear dependence between N_2 and E_1. If the mining depletes the resource then after having extracted a quantity $N_1(E_1)$ of refined mineral the producer will face, at the second period of extraction, a new initial concentration $C_1[N_1(E_1)] < C_0$ which depends on E_1 through N_1. Therefore depletion is negligible when $\frac{\partial N_2}{\partial E_1} \approx 0$.

1.5 Effects of the Absence of Substitutability in Equilibrium

In what follows we sketch a simple general equilibrium framework in which the qualitative effect of the lack of substitutability can be easily analysed. To keep the structure of the economy as simple as possible, we assume that the production described in our model is the only industry sector that exists in the economy. This may sound at odds with the basic resource nature of our production, which may itself be an input in other productive sectors of the economy. However, as we only need to highlight the role of the consumer's preferences in the valuation, we avoid any specific connotation of the good in this section, without loss of generality. This will avoid a more complex description of the productive structure of the economy which would obscure appreciation of the effects we will be highlighting.

The limit to the substitutability of useful energy by other factors of production has theoretical implications for the optimal allocation of resources in the economy. General equilibrium theory combines the optimal choice of individual consumers regarding the allocation of their budget, in order to achieve the highest possible utility with the production decision of profit-maximizing firms that produce and supply the consumption good to the economy. By matching of supply and demand a market clearing mechanism

determines values (i.e. market prices) according to which resources are allocated. A basic requirement of the general equilibrium analysis is that all values should be comparable, that is expressed in the same unit of account. This is achieved by normalizing prices in terms of one of the goods available in the economy, the numeraire. In a Neoclassical equilibrium, in which every good can be bartered against any other good, the choice of the numeraire is irrelevant as it will not alter the quantities finally produced and consumed. However, a key feature of Neoclassical general equilibrium models is also the perfect substitutability, on the production side, between factors of production. The lack of substitutability of energy implied by the thermodynamically consistent production function derived in the previous sections affects the equivalence between different factors of production and determines extra rigidity in the firm's decisions. We will provide some insight into the effect that this will have on the equilibrium allocation. We will find that the invariance of the equilibrium with respect to the unit of account may no longer hold. Our discussion will focus on the issue of the metric to be adopted for judging the optimality of resource allocation and production decisions.

An initial discussion of the role of thermodynamic efficiency in general equilibrium decisions can be found in Berry et al. (1978), who model the existence of a minimum energy input in production and identify it (as later done by Ruth (1995)) as that required by a limiting reversible process. In the energy dimension, isoquants of the production function do not approach a zero level within the limit but instead they reach positive constants which depend on the level of production. In their view, a thermodynamic efficiency criterion of resource allocation would select a combination of inputs (a point on the achievable isoquant) as close as possible to the minimum energy asymptote. They acknowledge that the optimal combination of inputs adopted by a firm in general equilibrium will be driven by overall cost minimization and may deviate from thermodynamic efficiency. If other inputs are scarce enough more energy will be used than the minimum thermodynamic limit. The neoclassical Criteria of Pareto Optimality in the allocation of resources (whereby no agent can improve their own welfare without damaging someone else's) would not usually coincide with thermodynamic efficiency in the use of resources. Roma (2006), however, argues that in the presence of negative externalities associated with entropic waste (e.g. global warming) the maximization of thermodynamic efficiency in production may improve global welfare.

We analyse here the problem of a value-maximising producer who operates over two periods under the irreversible technology we have described in the previous sections. The producer is only constrained by scarcity of energy, which does not allow full exploitation of the natural resources available in the time frame provided. The problem is that of allocating the use of available energy, given the prices at which the commodity produced can be sold in the two periods.

Given that the value of the goods produced in the different periods must be expressed in a common numeraire, we take it to be **energy** rather than the good produced in one of the two periods. That is, we assume that when the producer sells the final product he receives energy in exchange. It is assumed (following Ruth (1995)) that no capital enters the production process and that the raw material (mixture of mineral and rock) is freely available in an unlimited quantity[12].

When energy is chosen to be the unit of account for the producer, the profitability of the firm will differ

[12]Intertemporal production functions that do not explicitly involve the use of capital are common in the financial literature, see for example Cox et al. (1985), Fama (1972) and Hartwick (1970).

from that achievable when the unit of account is a finished good, from which utility can be derived directly. When energy is used as the unit of account, the producer's decision regarding its use in investments will essentially be based on thermodynamic considerations. If these are disregarded, production will be driven entirely by the non-satiated consumer who derives utility from the finished good but not from the energy needed to produce it.

The idea that accounting in terms of energy rather than in standard monetary terms may modify the attractiveness of an investment strategy may seem counter intuitive. However, this concept is firmly established within the literature concerning the concept of EROEI (Energy Return On Energy Investment) of power plants. EROEI is defined as the ratio of gross energy produced by an energy supply process to the total (direct plus indirect) energy cost of its production (Cleveland, 1992, 2005; Gately, 2007). Thus the EROEI approach provides the net energy analysis of an energy-based production process and can lead to completely different results from a pure financial assessment. Financial evaluation of investments deals only with market prices for the construction of the power plant plus maintenance costs and consequent financial revenues (Kaldellis et al., 2005; Tian and Wang, 2001; T.Tsoutsos et al., 2003)[13].

In our model we recognize that, if energy is taken to be the numeraire and means of exchange, energy itself represents "value", rather than the product. This implies that unused energy balances will enter the producer's value function. These specific features of the model force the producer to account for the profit of his investment in terms of energy. If the additional energy spent on production has a negative return it will not be used but rather retained at its face value as an inventory of the firm. The limit to the substitutability of the energy input will play a central role in this result. The total energy available for production is E_T, which is paid to the firm by the consumer/shareholder in exchange for future dividends before the production process begins. Given that P_e, the price of the good in terms of energy, and B, the real discount factor defined in expression (1.13), are both outside the firm's control in a competitive market, the producer's problem can be written as (see the two commodities model of section 1.1.4):

$$\begin{aligned} \max_{E_1,E_2} \quad & N_1\,P_e + B\,N_2\,P_e - E_1 - E_2\,, \\ sub \quad & N_1 = f^1\left(E_1\right)\,, \\ & N_2 = f^1\left(E_2\right)\,, \\ & E_1 + E_2 \leq E_T\,, \\ & E_1 > 0\,, \\ & E_2 > 0, \end{aligned} \tag{1.52}$$

where E_1 denotes the energy used to produce N_1 and E_2 the energy used to produce N_2.

[13] A comparative analysis between financial evaluation of power plants and EROEI can be found in Hall et al. (1979). The authors deal with a problem that arose in the NYSEG (New York State Electric and Gas Corporation) service area in the late 70's, when there was a proposal for a new 870 MW coal-fired generating station. The authors compare the energy (with the EROEI approach) and the dollar cost of building the power plant with the cost of a comprehensive regional program of insulation. They results show that the insulation program is more efficient in conserving energy than the power plant is in providing it, by a factor of at least 4 in economic terms and a factor of 15 in EROEI terms. This enforces the idea that the decision-making strategy can be quite different if an energy approach is considered.

Function $f^1(E_1)$ (or $f^1(E_2)$) is defined by expressions (1.46) (or 1.51).

In problem (1.52) the producer, who may use up to the amount of energy E_T, may choose not to use part of the available energy for production and retain it as an asset of the firm. The unused energy will be $E_T - E_1 - E_2$ and will add to the value of the firm.

On the consumer's side of the economy we assume, for simplicity, that a representative consumer exists who only derives utility from the finished good. In order to define the consumer's budget constraint, we consider two alternative exchange mechanisms. We will initially assume that the amount of scarce energy E_T is in the hands of the consumer, who exchanges it with the firm in order to obtain rights to the consumption good produced. In this case we give energy a special role in the exchange mechanism, by requiring that every good can only be purchased by exchanging it for a pre-specified good, rather than by bartering. This is the essence of a cash in advance constraint (Clower, 1967), which is used to provide a role for fiat money (paper money without any intrinsic value forced onto the economy by a Central Bank) that would otherwise have no economic role in the economy.

Alternatively, we can assume that the same amount of energy E_T is already an endowment of the firm and that the consumer has the right to receive the product of the firm by way of dividends. We will analyze this case below.

However, in a barter economy in which only physical goods with a direct economic use are present the imposition that a specific good should be exchanged against all the others should have no effect. Nevertheless if energy is exchanged for the availability of the good produced then a price of the good in terms of energy will be defined. This will allow the use of energy as a numeraire, and will provide information for its efficient allocation in the economy.

The consumer's problem is:

$$\max_{C_1,C_2} \quad u(C_1, C_2)\,, \tag{1.53}$$

$$sub \quad P_e\,(C_1 + B\,C_2) \leq E_T\,, \tag{1.54}$$

where C_1 and C_2 denote consumption of the good at the end of the time intervals 1 and 2 respectively. The consumer trades his endowment of energy E_T with the firm before production begins, in exchange for future production. As energy does not enter the utility function in the case in which the endowment of the consumer is energy, there will be no incentive for the consumer to avoid spending some of the available energy in exchange for the good, thus pushing the energy price of the good to the highest level allowed by the budget constraint. This will in turn provide the producer with an incentive to produce more. Hence (1.54) will hold as an equality.

The consumer's first order conditions imply that:

$$B = \frac{\frac{\partial u}{\partial C_2}}{\frac{\partial u}{\partial C_1}}, \tag{1.55}$$

and:

$$P_e = \frac{E_T}{C_1 + B\,C_2}, \tag{1.56}$$

where in equilibrium $C_1 = N_1$ and $C_2 = N_2$. Hence:

$$P_e = \frac{E_T}{N_1 + B\,N_2}. \tag{1.57}$$

1.5.1 Use of Energy by the Producer.

In the setting described above, where energy is the unit of account, it will not generally be optimal for the producer to use all the available energy, unless specific restrictions are placed on technology and the consumer's preferences. Production will not exceed the level where the marginal energy cost is equal to the price set by the market-clearing condition (1.57). This result derives from the inherent non linearity of the irreversible production function, which displays increasing marginal cost and from the limit to the substitutability of energy by other factors of production. Hence if energy is used as the numeraire, given that there is no an alternative technology in the economy under which unused energy balances may efficiently enter production, and therefore the producer will not be able to exchange unused energy with a competitor, equilibrium in the economy will leave some energy resources unused. This would not be the case under perfect substitutability of the two production inputs (e.g. a Cobb-Douglas production function). In such a case the producer would instead use all the available energy, as it would be possible to supplement it with a high enough amount of the other input to achieve an increase in production that justifies the energy spent. Perfect substitutability would be inconsistent with the unlimited availability of the second input, and a price would have to be introduced for the second input which must be assumed to be scarce. In our case we have already substituted in the production frontier (1.46) the optimal amount of the free input, which minimizes the energy cost of production. This may, on the other hand, have positive environmental consequences as less entropy is generated in the economy (Kummel, 1989; Roma, 2006).

The general solution of problem (1.52) is obtained by maximizing the Lagrange function:

$$\mathcal{L}\left(E_1, E_2, \lambda\right) = N_1\left(E_1\right)\, P_e + B\, N_2\left(E_2\right)\, P_e - E_1 - E_2 + \lambda\, \left(E_T - E_1 - E_2\right) \tag{1.58}$$

w.r.t. its arguments. Maximization w.r.t. E_1 and E_2 gives the following pair of equations if the energy bound holds as a strict inequality in an internal solution (namely $E_1 + E_2 < E_T$) and $\lambda = 0$:

$$\begin{aligned} P_e\, \frac{\partial N_1}{\partial E_1} - 1 &= 0, \\ B\, P_e\, \frac{\partial N_2}{\partial E_2} - 1 &= 0. \end{aligned} \tag{1.59}$$

The second condition of (1.59) implies that $\frac{\partial E_2}{\partial N_2} = B\, P_e$, that is the marginal cost of production in the second period is equal to the discounted price $B\, P_e$.

In Section 1.6 we report an analytical solution for the problem (1.52) and show that it corresponds to a solution with a non-zero energy slack.

Slack arises because additional use of energy for production does not increase the value of the firm[14].

[14] An extension of the production side of the economy to include additional production processes (i.e. other firms producing additional goods) is unlikely to eliminate the potential slack in the use of energy if each additional process is subject to a similar minimum energy threshold. An average minimum energy bound would arise for the production side of the economy,

We now turn to the case in which energy is not the unit of account. In a typical Neoclassical model E_T would be treated as an endowment of the firm which has no value if not used, hence its use would not be a cost and the produced good itself would be the numeraire. If scarce energy did not have to be valued on the market, its full rational use would follow over the time span of the model. Then energy cost does not matter and the producer's problem becomes:

$$\begin{aligned} \max_{E_1,E_2} \quad & N_1 + B\,N_2\,, \\ sub \quad & N_1 = f^1\left(E_1\right)\,, \\ & N_2 = f^1\left(E_2\right)\,, \\ & E_1 + E_2 \leq E_T\,, \\ & E_1 > 0\,, \\ & E_2 > 0. \end{aligned} \tag{1.60}$$

The objective function is strictly increasing in both N_1 and N_2. In turn N_1 and N_2 are strictly increasing in E_1 and E_2, until $E_1 + E_2 = E_T$. So the energy inputs E_1 and E_2 will be increased by the producer until the constraint holds as an equality and $E_2 = E_T - E_1$. This can be seen analytically from equations (1.59). If the maximization problem is (1.60), the first order conditions (1.59) for a solution with an energy slack become:

$$\begin{aligned} \frac{\partial N_1}{\partial E_1} &= 0, \\ B\,\frac{\partial N_2}{\partial E_2} &= 0. \end{aligned} \tag{1.61}$$

Conditions (1.61) are never satisfied because $\forall E_1 > 0$ and $\forall E_2 > 0$ we have that $\frac{\partial N_1}{\partial E_1} > 0$ and $\frac{\partial N_2}{\partial E_2} > 0$. Disregarding the energy cost leads to the full use of energy, E_T.

If all the available energy is used (namely $E_1 + E_2 = E_T$) , the problem (1.52), as well as (1.60), reduces to a maximization in one single variable E_1, that is it can be equivalently rewritten as:

$$\begin{aligned} \max_{E_1} \quad & N_1\,P_e + B\,N_2\,P_e - E_T\,, \\ sub \quad & N_1 = f^1\left(E_1\right)\,, \\ & N_2 = f^1\left(E_T - E_1\right)\,, \\ & E_T \geq E_1 > 0\,, \end{aligned} \tag{1.62}$$

The first order condition is:

$$\frac{\partial N_1}{\partial E_1} + B\,\frac{\partial N_2}{\partial E_1} = 0. \tag{1.63}$$

which would prevent substitution of energy by other factors. Similarly, the possibility of including energy as an argument of the consumer's utility function would not rule out slack. If E_c is the portion of the consumer's initial endowment which is consumed directly and therefore not exchanged against the goods produced the resulting energy price $P_e = \frac{E_T - E_c}{N_1 + B\,N_2}$ will in general be lower, decreasing the incentive to produce even further.

That is the equilibrium where all the finite amount of available energy E_T is used for production is independent from the price P_e. The consumer's budget constraint can be rewritten as $C_1 + B\,C_2 = N_1 + B\,N_2$. The price P_e plays no role.

Note that if the consumer is forced to exchange energy for the consumption good, no additional constraint is imposed compared to the case in which the budget constraint is $C_1 + B\,C_2 = N_1 + B\,N_2$, as the amount of energy available for production is the same. What is modifying the equilibrium is the definition of the energy price P_e, that is the introduction of a market for energy which imposes a thermodynamic efficiency criterion on the firm's decision, whereby energy cost is compared to the energy revenue of the irreversible technology.

It is clear from the previous discussion that the producer's energy efficiency differs between the two cases in which there is an energy slack $E_1 + E_2 < E_T$ or in which all the energy available is spent for production and $E_1 + E_2 = E_T$. Accounting in terms of energy is a technical criterion that leads to different decisions with respect to standard monetary accounting. A key role in this sense is played by the EROEI index, which can be interpreted as the ratio between the energy collected by the producer and the energy spent for production. In the case of our firm, the producer collects an amount of energy equal to $P_e\,N_1 + B\,P_e\,N_2 = E_T$, facing an energy expenditure of $E_1 + E_2$, so that the EROEI for our production is:

$$EROEI = \frac{P_e\,N_1 + B\,P_e\,N_2}{E_1 + E_2} = \frac{E_T}{E_1 + E_2}. \tag{1.64}$$

This will be higher than one if the energy is used efficiently in problem (1.52) and will be equal to one if the energy cost of production is disregarded in the decision.

1.6 Analytical and Numerical Solutions

1.6.1 Positive Energy Slack

We found analytical solutions for the producer-consumer's problem (1.52)-(1.53) when depletion is neglected and when $u\,(C_1, C_2) = C_1^{\beta}\,C_2^{1-\beta}$, with $\beta \in (0,1)$. In this situation $B = \gamma\,\frac{N_1}{N_2}$ where $\gamma = \frac{1-\beta}{\beta}$ and the solutions are:

$$\begin{aligned} E_1^* \;=\; & \frac{1}{8\,(C_0-k)^2\,Q_\tau\;(1+\gamma)}\Bigg\{4\,E_T\,k^2\,Q_\tau + 2\,{C_0}^2\,\left(2\,E_T\,Q_\tau - (1+\gamma)\;(R\,T\,\xi_0)^2\right) - \\ & - 2\,C_0\,\left(4\,E_T\,k\,Q_\tau - (1+\gamma)\;(R\,T\,\xi_0)^2\right) - \\ & - R\,T\,\xi_0\,\left(R\,T\;(1+\gamma)\;\xi_0 - \sqrt{G_{C_0}}\,\sqrt{1+\gamma}\,\sqrt{8\,E_T\,(C_0-k)^2\,Q_\tau + G_{C_0}\;(1+\gamma)\;(R\,T\,\xi_0)^2}\right)\Bigg\}, \end{aligned} \tag{1.65}$$

$$\begin{aligned} E_2^* \;=\; & \frac{1}{8\,(C_0-k)^2\,Q_\tau\;(1+\gamma)}\Bigg\{4\,E_T\,k^2\,Q_\tau\,\gamma + 2\,{C_0}^2\,\left(2\,E_T\,Q_\tau\,\gamma - (1+\gamma)\;(R\,T\,\xi_0)^2\right) - \\ & - 2\,C_0\,\left(4\,E_T\,k\,Q_\tau\,\gamma - (1+\gamma)\;(R\,T\,\xi_0)^2\right) - \\ & - R\,T\,\xi_0\,\left(R\,T\;(1+\gamma)\;\xi_0 - \sqrt{G_{C_0}}\,\sqrt{1+\gamma}\,\sqrt{8\,E_T\,(C_0-k)^2\,Q_\tau\,\gamma + G_{C_0}\;(1+\gamma)\;(R\,T\,\xi_0)^2}\right)\Bigg\}. \end{aligned} \tag{1.66}$$

It is possible to verify that $E_1 + E_2 < E_T$, that is (1.65)-(1.66) actually correspond to a solution with an energy slack regardless of the value of E_T. In order to see this, consider the fraction of the total energy spent in the production:

$$\begin{aligned}\frac{E_1^* + E_2^*}{E_T} &= \frac{1}{8\,E_T\,(C_0-k)^2\,Q_\tau\,(1+\gamma)}\Bigg\{2\,(1+\gamma)\,\left(2\,E_T\,(C_0-k)^2\,Q_\tau - G_{C_0}\,(R\,T\,\xi_0)^2\right) + \\ &+ \sqrt{G_{C_0}\,(1+\gamma)}\,R\,T\,\xi_0\left[\sqrt{8\,E_T\,(C_0-k)^2\,Q_\tau + G_{C_0}\,(1+\gamma)\,(R\,T\,\xi_0)^2} + \right. \\ &+ \left.\sqrt{8\,E_T\,(C_0-k)^2\,Q_\tau\,\gamma + G_{C_0}\,(1+\gamma)\,(R\,T\,\xi_0)^2}\right]\Bigg\}. \end{aligned} \tag{1.67}$$

We will show that:

$$\frac{E_1^* + E_2^*}{E_T} < 1, \tag{1.68}$$

whatever is E_T. Straightforward computations show that inequality (1.68) holds if and only if:

$$\begin{aligned}\Psi\,(E_T) &\equiv -2\,(1+\gamma)\,\left(2\,E_T\,(C_0-k)^2\,Q_\tau + G_{C_0}\,(R\,T\,\xi_0)^2\right) + \\ &+ \sqrt{G_{C_0}\,(1+\gamma)}\,R\,T\,\xi_0\left[\sqrt{8\,E_T\,(C_0-k)^2\,Q_\tau + G_{C_0}\,(1+\gamma)\,(R\,T\,\xi_0)^2} + \right. \\ &+ \left.\sqrt{8\,E_T\,(C_0-k)^2\,Q_\tau\,\gamma + G_{C_0}\,(1+\gamma)\,(R\,T\,\xi_0)^2}\right] < 0. \end{aligned} \tag{1.69}$$

It is easy to see that $\Psi\,(E_T) \to -\infty$ when $E_T \to +\infty$. Moreover, the equation $\Psi\,(E_T) = 0$ has only complex solutions, given by:

$$E_T^{\pm} = \frac{-G_{C_0}\,(C_0-k)^2\,Q_\tau\,(1+\gamma)^2\,(R\,T\,\xi_0)^2 \pm \sqrt{-\left[{G_{C_0}}^2\,(C_0-k)^4\,{Q_\tau}^2\,(\gamma^2-1)^2\,(R\,T\,\xi_0)^4\right]}}{2\,(C_0-k)^4\,{Q_\tau}^2\,(1+\gamma)^2}, \tag{1.70}$$

i.e. the function $\Psi\,(E_T)$ does not intercept the real axis and then $\Psi\,(E_T) < 0, \forall E_T$. This shows inequality (1.68).

It is interesting to analyze the producer's choice within the limit of infinite energy available (i.e. $E_T \to +\infty$). From expressions (1.65)-(1.66) it is easy to see that:

$$\lim_{E_T \to +\infty} \frac{E_1^*}{E_T} = \frac{1}{2\,(1+\gamma)}, \tag{1.71}$$

$$\lim_{E_T \to +\infty} \frac{E_2^*}{E_T} = \frac{\gamma}{2\,(1+\gamma)}. \tag{1.72}$$

This means that the fractions of the energy spent in the two production periods depend only on the consumer's preference γ. Moreover:

$$\lim_{E_T \to +\infty} \frac{E_1^* + E_2^*}{E_T} = \frac{1}{2}. \tag{1.73}$$

That is, only half of the available energy is used for production.

1.6.2 Null Energy Slack

Analytical solutions can be also found for the slackless case (1.62) if depletion is neglected. Straightforward computations show that, in this case, the producer's choice is:

$$E_1^{no-slack} = \frac{E_T}{(1+B^2)} - \frac{\left(B^2-1\right)\, G_{C_0}\, \left(R\,T\,\xi_0\right)^2}{4\,\left(1+B^2\right)\,\left(C_0-k\right)^2\,Q_\tau}, \tag{1.74}$$

$$E_2^{no-slack} = \frac{B^2\,E_T}{(1+B^2)} + \frac{\left(B^2-1\right)\, G_{C_0}\, \left(R\,T\,\xi_0\right)^2}{4\,\left(1+B^2\right)\,\left(C_0-k\right)^2\,Q_\tau}. \tag{1.75}$$

Obviously, $E_1^{no-slack} + E_2^{no-slack} = E_T$. Similarly to the case of Section (1.6.1), if the producer is provided with an infinite amount of energy (i.e. $E_T \to +\infty$), the fractions of energy spent in first and second period depend only on the consumer's preferences (contained in parameter B). In formula:

$$\lim_{E_T \to +\infty} \frac{E_1^{no-slack}}{E_T} = \frac{1}{1+B^2} \tag{1.76}$$

$$\lim_{E_T \to +\infty} \frac{E_2^{no-slack}}{E_T} = \frac{B^2}{1+B^2} \tag{1.77}$$

1.7 Arbitrage Analysis

The energy price of the good produced will be subjected to arbitrage bounds. Suppose that we observe its market price in terms of energy. We have at our disposal a quantity of moles N_Q of the final good characterized by the highly refined mixture $(k, 1-k)$. We wonder whether it is possible, in some way, to extract the energy E_Q that has been used to produce the moles N_Q and compare it to the market price. This will be possible under certain conditions. The mining operator has at his disposal an engine which is able to carry out the thermodynamic transformation $(C_0, 1-C_0) \to (k, 1-k)$. We call this transformation the forward process. The operator could plan to extract the energy E_Q spent to produce the moles N_Q by running the engine in a backward process $(k, 1-k) \to (C_0, 1-C_0)$. This is only fully possible, however, if the engine works reversibly, that is within the limit $\tau \to +\infty$. As discussed above, this thermodynamic limit has no economic sense because production processes must be completed in a finite time and thus the production time τ has a fixed finite value. However, if the transformation (of raw material into finished good), which occurs in a finite time, is so small that the thermodynamic system remains in equilibrium, it can be assumed that the transformation occurs in a reversible way. Suppose that an infinitesimal quantity dN of the final good can be purchased on the market. It would be possible to perform a reversible transformation in a finite time τ, extracting the infinitesimal quantity of energy dE held in dN. This is possible because the transformation now involves only infinitesimal variation of the thermodynamic variables of the system.

Assume for now that the technology is totally reversible, that is we are within the limit $\tau \to +\infty$. We suppose that our engine is able to work backward, namely we can make the transformation $(k, 1-k) \to (C_0, 1-C_0)$. We buy N_Q moles of the final good, characterized by a concentration vector $(k, 1-k)$, and we compute the energy necessary to re-obtain the initial mixture $(C_0, 1-C_0)$:

$$E_R^{back} = RT\,N_Q\,\left[C_0\,\log C_0 + (1-C_0)\,\log(1-C_0) - k\,\log k - (1-k)\,\log(1-k)\right]. \tag{1.78}$$

It can be shown, in our setting where $k > 1-C_0 > C_0$, that the energy variation expressed by formula (1.78) is negative, that is we gain an amount of energy $|E_R^{back}|$. The energy required to obtain N_Q moles of the final good $(k, 1-k)$ is independent (in the reversible case) from the number of the initial moles of the mixture $(C_0, 1-C_0)$ and equal to:

$$E_R^{for} = RT\,N_Q\,\left[k\,\log k + (1-k)\,\log(1-k) - C_0\,\log C_0 - (1-C_0)\,\log(1-C_0)\right] = -E_R^{back} > 0. \tag{1.79}$$

Thus under a fully reversible technology we require, for the forward process, an amount of energy equal, in absolute value, to the energy released by the backward process.

As a consequence, under the reversible technology the energy market price of the good produced must be equal to the marginal cost of production (which is, in this case, a constant independent from the production level). Suppose that $P_e > \frac{\partial E_1}{\partial N_1} = RT\,\xi_0$; then we can produce and sell N_s moles of the final good with an amount of energy equal to $RT\,\xi_0\,N_s$ and we receive as payment an amount of energy $N_s\,P_e > RT\,\xi_0\,N_s$ with a net energy gain $N_s\,(P_e - RT\,\xi_0) > 0$. If $P_e < \frac{\partial E_1}{\partial N_1} = RT\,\xi_0$, then we buy N_b moles of the good, paying an amount of energy $P_e\,N_b$, and then we run our engine backward, extracting an amount of energy $RT\,\xi_0\,N_b$, with a net energy gain of $N_b\,(RT\,\xi_0 - P_e) > 0$. It follows that if the technology is reversible $P_e = RT\,\xi_0$. This is a no-arbitrage condition which is obtained irrespective of the preferences in the economy.

The situation is different if the technology is not reversible. In the irreversible case the energy required to transform N_b moles of concentration $(k, 1-k)$ into N_b moles of concentration $(C_0, 1-C_0)$ is:

$$E_{IR}^{back} = E_R^{back} + \frac{2\,N_b^2}{\tau\,\alpha}\,(k-C_0)^2\,. \tag{1.80}$$

Since the second term on the right-hand side of expression (1.80) is positive if we want the backward process to release energy, that is $E_{IR}^{back} < 0$, we have to operate our engine for a sufficiently long time τ. More precisely:

$$\tau \geq \frac{2\,N_b^2}{RT\,\xi_0\,\alpha}\,(k-C_0)^2\,. \tag{1.81}$$

Even if inequality (1.81) holds, part of the energy used to produce N_b cannot be recovered. Moreover the forward process $(C_0, 1-C_0) \to (k, 1-k)$ requires an amount of energy given by:

$$E_{IR}^{for} = E_R^{for} + \frac{2\,N_b^2}{\tau\,\alpha}\,(k-C_0)^2\,, \tag{1.82}$$

and then $E_{IR}^{for} > E_{IR}^{back}$ because $E_R^{for} > 0$ and $E_R^{back} < 0$. If the technology is not reversible it is not possible to replicate the argument used in the reversible case to determine P_e for a finite quantity of material and energy. However, the no-arbitrage argument is still partially valid for an infinitesimal variation in production at economic equilibrium. Suppose that we have irreversible technology and we achieve economic equilibrium at some production level N_* (without loss of generality in the first production

period). If $P_e > \frac{\partial E_1}{\partial N_1}|_{N_*}$, we can produce and sell an extra infinitesimal quantity of the final good dN, spending an energy $dE = \frac{\partial E_1}{\partial N_1}|_{N_*}\, dN$, with a net energy gain $dN\left(P_e - \frac{\partial E_1}{\partial N_1}|_{N_*}\right) > 0$[15]. If $P_e < \frac{\partial E_1}{\partial N_1}|_{N_*}$, the producer can buy an extra quantity of refined material dN at a price $dN\,P_e$ and reversibly extract the energy $dN\,R\,T\,\xi_0$. By inspecting expression (1.48), it can be seen that $\frac{\partial E_1}{\partial N_1}|_{N_*} \geq R\,T\,\xi_0, \forall N_*$. Thus the condition $P_e < \frac{\partial E_1}{\partial N_1}|_{N_*}$ is not sufficient to establish that the quantity of energy $dN\,R\,T\,\xi_0 - dN\,P_e$ delivered by this strategy is positive and therefore there is an arbitrage opportunity. To find an arbitrage opportunity in the irreversible case, the price P_e must be lower than the reversible marginal cost of production. In the range $P_e \in \left[R\,T\,\xi_0, \frac{\partial E_1}{\partial N_1}|_{N_*}\right]$ we have no arbitrage restrictions on the price P_e.

The same type of arbitrage arguments are used in Roma (2006) to determine the no-arbitrage price in terms of energy for a more elementary product (hot fluid). This case is quite different from ours and leads to a complete identification of the energy price P_e of the good through no-arbitrage considerations because the product itself is thermal energy (stored in a fluid) and there is no need to operate any thermodynamic engine backward to extract it.

1.8 The Producer's Problem under the Reversible Technology

We turn to the analysis of production decisions within the limit of reversible technology. As discussed, this case is unrealistic. We show that the slack (under-utilization) in energy, which characterizes the production decision of a profit-maximising firm that accounts in terms of energy under the irreversible extraction technology, will no longer exist. Moreover, as the price P_e will be set by no-arbitrage considerations in this case, the unconstrained solution will not necessarily be consistent with the investor's preferences. To see this, recall the form that the production function takes under full reversibility. From (1.34) we have a linear technology in this case:

$$N_1 = \frac{E_1}{R\,T\,\xi_0}. \tag{1.83}$$

Neglecting the depletion of the mine, in the second period we have:

$$N_2 = \frac{E_2}{R\,T\,\xi_0}, \tag{1.84}$$

Under this technology the firm's profit maximisation problem, with the inequality constraint $E_1 + E_2 < E_T$, will deliver inconsistent results. The first order conditions (1.59) are:

$$P_e\,\frac{1}{R\,T\,\xi_0} - 1 = 0, \tag{1.85}$$

$$B\,P_e\,\frac{1}{R\,T\,\xi_0} - 1 = 0. \tag{1.86}$$

Equation (1.85) gives the no-arbitrage condition $P_e = R\,T\,\xi_0$ while equation (1.86) imposes that $B = 1$. In this situation the maximized value of the firm equals zero:

$$P_e\,\frac{E_1}{R\,T\,\xi_0} + P_e\,\frac{E_2}{R\,T\,\xi_0} - E_1 - E_2 \equiv 0, \tag{1.87}$$

[15] Note that this argument is valid without the introduction of the concept of reversibility.

for whatever choice of E_1 and E_2. Moreover the price clearing condition (1.57) requires that:

$$P_e = \frac{E_T}{N_1 + B\,N_2} = R\,T\,\xi_0\,\frac{E_T}{E_1 + E_2}, \tag{1.88}$$

which, together with $P_e = R\,T\,\xi_0$, implies that $E_1 + E_2 = E_T$. No slack is found even in the case $\lambda = 0$. If $E_1 + E_2 = E_T$ the firm's value becomes:

$$E_1 + B\,(E_T - E_1) - E_T, \tag{1.89}$$

and the first order condition implies $B = 1$.

Again, the maximised value of the firm less the energy cost will be identically equal to zero for whatever choice of E_1 and E_2 and the resource allocation between the two periods will not depend on the consumer's preferences. The producer's profit will not deviate from zero whatever the production choice.

As a concluding remark we can state that reversibility leads to the complete use of energy and an equilibrium independent of the choice of the numeraire. In this sense the reversible limit could be interpreted as the limit to a neoclassical equilibrium, discussed in Section 1.1.4.

1.9 Conclusions.

In this chapter we have addressed the issue of substitutability between inputs in a production function and, in particular, the existence of a lower bound on the amount of useful energy necessary for production to take place. The hypothesis of perfect substitutability has convenient implications for general equilibrium modelling, but does not plausibly represent many actual production processes.

Our methodological approach provides a clear cut answer to this much debated question. In this light we revisit the production problem of a mining operator who refines a mineral from its natural concentration C_0 to a strike concentration k, defining the consumption good. The mining operator faces an initial mixture of useful mineral and waste rock. The production process is modelled analytically, following the thermodynamic theory of an irreversible separation process. The model provides a clear answer to the question of substitutability between natural resources and energy for the specific production considered. It confirms the existence of an energy bound below which production is impossible. This prevents full substitutability between factors of production and contradicts the standard Cobb-Douglas representation of production. If we assume, following the view in Ayres and Miller (1980), that the original mixture from which the commodity has to be extracted is a freely available natural resource, while energy is in aggregate scarce, the inputs for every level of production will be combined with fixed coefficients which depend only on the scale of production and the time available for the production process. The key role of energy as a scarce non-substitutable resource that underlies thermodynamic optimization in the engineering literature can hence be reconciled with economic optimization in a microeconomic model. The derived production function is non-linear in the inputs as long as the process occurs in a finite time, and displays increasing marginal cost in terms of energy.

This technological description is used within the optimisation problem of a profit-maximising mining operator who produces over two periods and sells the product in a competitive market. As generally

highlighted by Berry et al. (1978) Pareto optimality will not necessarily coincide with the most efficient use of energy in a general equilibrium framework. In our economy the only choice available to the producer concerns the use of energy, which provides a framework for the analysis of the relationship between economic and thermodynamic optima. In a neoclassical framework, the optimal choice would imply the full use of energy with an inter-temporal allocation that maximizes the consumer's utility. However, if we impose energy as the means of payment and the unit of account in the economy the equilibrium will be twisted towards thermodynamic efficiency. The lack of substitutability between energy and other inputs implies that the price, in terms of energy, of the finished product established in the market may not be high enough to induce the producer to employ all the energy available in production. Since no other factors of production can be increased in order to improve the marginal productivity of energy, the under-utilization of available energy will follow. In this way the producer will take the irreversibility in the use of energy into account. The choice of numeraire and means of payment will not be irrelevant, in the sense that the creation of a market where goods are exchanged against energy will determine an energy price and provide guidance to the producer on how to use it efficiently. When energy is the unit of account it will represent "value" and will not be wasted. If the producer's objective function is consistent with this criterion, energetically inefficient investments will be ruled out. However, the under-utilization of the energy resources available resulting from energy accounting will not necessarily lead to Pareto suboptimality if the negative externalities of energy degradation during production are taken into account. As highlighted by Kummel (1989) and Roma (2006), the efficient use of energy may have positive environmental effects. On the other hand, if no value is attributed to the energy resources per se (but only the finished product is valued) full utilization of the available energy will follow. This result may be interpreted in the light of the literature on EROEI (Energy Return on Energy Investment).

In the specific case, our assumptions lead to a one-to-one relationship between energy employed and production of the basic commodity considered. Whether this may be considered a reasonable approximation for what will be obtained in the case of more complex production processes, or possibly an abstraction that may apply to the description of aggregate production in a macroeconomic sense, should be the subject of further research. We note here that a vast body of literature devoted to the successful estimation of aggregate production, using electric energy as the key explanatory variable, assumes the existence of such a one-to-one relationship (Bodo and Signorini, 1987; G. and Signorini, 1991).

Another matter for further research is the empirical application of thermodynamics based production functions.

Only when the production process becomes fully reversible will the producer's analysis in energy or financial (produced good) terms be the same. However, reversibility does not describe any real world production process.

Chapter 2

Measure Theory and Time Series Analysis

2.1 Probability Spaces

Measure theory is a fundamental tool for financial econometrics. This section provides an overview of the theory with important results useful for an exhaustive comprehension of the thesis. Most part of the results are taken from Bjork (1998), Cont and Tankov (2004) and Shreve (2004).

Probability theory is based on the introduction of a probability space Ω: a set which describes all possible events or scenarios that may occur. A generic element of Ω will be denoted by ω. For example if we are looking at the toss of a coin infinitely many times we need to work with the set of all possible infinite sequences of heads and tails: $\Omega = \{\omega = \omega_1 \omega_2 ... | \omega_n \text{ is result of the n-th coin toss}\}$.

Usually one observes a random process X which is the real source of randomness. It is natural to identify the space of all possible scenarios with all possible outcomes of X. For example, if X is the result of a dice rolling experiment, one would have $\Omega = \{1, 2, 3, 4, 5, 6\}$. With this choice the random process X is the identity map:

$$X : \Omega \to \Omega, \quad X(\omega) = \omega. \tag{2.1}$$

Representation (2.1) means that if we observe the outcome n we are in the scenario n. This identification is called the *canonical* representation. It is the simplest representation we can construct. However, despite its simplicity, this choice is not useful when dealing with complex financial model. With the choice (2.1) we cannot introduce a second source of randomness which is independent from X. Being X the unique source of randomness all processes will be correlated to it. To be more precise we give an example: the pricing of a derivative depend on the value of an asset S_T at some future date T. This value can be affected by many factors such demand and supply, economic indicators or even external news which do not have a numeric representation. All these factors are summed up in a variable ω. Thus the variable S_T will be a function of the economic scenario described by ω: $S_T \equiv S_T(\omega)$.

It is evident that we must separate the space of possible scenarios Ω from the space of possible values

for a given asset S_T at some date T. The first one will remain an abstract space, needed to formulate the theory. The second one will instead be described by the real line $\mathbb{R}$.

The second tool we need is a probability measure on the abstract probability space Ω. With *probability measure* we mean some function which assigns to the events of Ω a real number to be interpreted as the probability of those events. The domain of definition of the probability measure must satisfies certain condition in order to obtain the usual properties of probability. We thus introduce the idea of σ-algebra:

Definition 2.1.1 *Let Ω be a nonempty set and let $\mathcal{F}$ be a collection of subset of Ω. We say that $\mathcal{F}$ is a σ-algera if:*

1. $\emptyset \in \mathcal{F}$.
2. *If $A \in \mathcal{F}$ then, denoted by $A^c = \Omega/A$ its complement, then $A^c \in \mathcal{F}$.*
3. *If $(A)_{n\in\mathbb{N}} \in \mathcal{F}$ is a sequence of set of $\mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.*

In particular if $\mathcal{F}$ is a σ-algebra of Ω then $\Omega \in \mathcal{F}$. A characterization of the σ-algebra is given by the following proposition:

Proposition 2.1.2 *Given a collection $\mathcal{A}$ of subsets of Ω, there exists a unique σ-algebra denoted $\sigma(\mathcal{A})$ with the following property: if any σ-algebra $\mathcal{F}'$ contains $\mathcal{A}$ then $\sigma(\mathcal{A}) \subset \mathcal{F}'$. That is $\sigma(\mathcal{A})$ is the smallest σ-algebra containing $\mathcal{A}$ or the minimal completion of $\mathcal{A}$ to a σ-algebra. $\sigma(\mathcal{A})$ is called the σ-algebra generated by $\mathcal{A}$.*

Suppose that we are working in $\mathbb{R}^d$. We can consider the collection $\mathcal{A}$ of all open subsets of $\mathbb{R}^d$. In this case the σ-algebra generated by $\mathcal{A}$ is called the Borel σ-algebra and denoted by $\mathcal{B}$. An element $B \in \mathcal{B}$ is called a Borel set. Each open and closed set is a Borel set.

We can now give the generic definition of measure on a set Ω.

Definition 2.1.3 *Let Ω be a nonempty space and let $\mathcal{F}$ be a σ-algebra of subset of Ω. A positive measure on $(\Omega, \mathcal{F})$ is a function:*

$$\begin{aligned} \mu: \quad \mathcal{F} &\rightarrow [0,\infty] \\ \forall A \in \mathcal{F}: \quad A &\rightarrow \mu(A) \end{aligned} \tag{2.2}$$

such that:

1. $\mu(\emptyset) = 0$.
2. *If $(A)_{n\in\mathbb{N}} \in \mathcal{F}$ is a sequence of disjoint set of $\mathcal{F}$ (i.e. $A_i \bigcap A_j = \emptyset \forall i \neq j$) then:*

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n). \tag{2.3}$$

An element $A \in \mathcal{F}$ is called a measurable set and $\mu(A)$ its measure. The triplet $(\Omega, \mathcal{F}, \mu)$ is called a measurable space.

If $\mu(\Omega) = 1$ then μ is called a probability measure and the triplet $(\Omega, \mathcal{F}, \mu)$ is called a probability space. For probability measure we will adopt the symbol $\mathbb{P}$ in order to distinguish from a generic measure μ.

The notion of a finite measure is generalized with the introduction of Radon measure:

Definition 2.1.4 *Let $E \subset \mathbb{R}^d$. A Radon measure on $(E, \mathcal{B})$ is a measure such that for each compact measurable set $B \in \mathcal{B}$, $\mu(\mathcal{B}) < \infty$.*

An example of Radon measure is the Lebesgue measure λ on $\left(\mathbb{R}^d, \mathcal{B}\right)$:

$$\lambda(A) = \int_A dx \quad \forall A \in \mathcal{B}. \tag{2.4}$$

An important example of measure is the Dirac measure:

Definition 2.1.5 *Given a point $\omega_0 \in \Omega$, the Dirac measure associated to ω_0 is defined, $\forall A \in \mathcal{B}$ as:*

$$\delta_{\omega_0}(A) \equiv \begin{cases} 1 & \text{if } \omega_0 \in A \\ 0 & \text{if } \omega_0 \notin A \end{cases}$$

The Dirac measure can be extended to any countable set $\omega = \{\omega_i, i = 1, 2, 3, ...\} \subset \Omega$ introducing the counting measure:

$$\mu_\omega(A) = card\{i, \omega_i \in A\} = \sum_i 1_{\omega_i \in A}, \tag{2.5}$$

where $card\{E\}$ is the cardinality of a set E. Conversely a measure which gives zero measure to any point is said to be *diffuse*:

$$\forall \omega_0 \in \Omega, \quad \mu(\{w_0\}) = 0. \tag{2.6}$$

On $(E, \mathcal{B})$ with $E \subset \mathbb{R}^d$ each Radon measure admits a particular decomposition, as stated by the following proposition:

Proposition 2.1.6 *Any Radon measure μ can be decomposed in the following way:*

$$\mu = \mu_0 + \sum_{j \geq 1} b_j \, \delta_{x_j} \quad x_j \in E, \quad b_j > 0, \tag{2.7}$$

where μ_0 is diffuse.

2.2 Random Variables

Real random variables are fundamental for financial applications. They describe the observed values of a set of generic economic variables, like assets, interest rate etc. In order to define what a random variable is we need to introduce the concept of measurable function.

Definition 2.2.1 *Let $(E_1, \mathcal{F}_1, \mu_1)$ and $(E_2, \mathcal{F}_2, \mu_2)$ be two measurable space. A function $f : E_1 \to E_2$ is said to be measurable if for any measurable set $A \in \mathcal{F}_2$, the set:*

$$f^{-1}(A) = \{x \in E_1 | f(x) \in A\}, \tag{2.8}$$

is a measurable subset of E_1.

For each state of the economy $\omega \in \Omega$ we would like to associate a vector of real numbers, thus we introduce the following definition:

Definition 2.2.2 *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a set $E \subset \mathbb{R}^d$ a real random variable X is any measurable function:*

$$X : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{B}, \mu) . \tag{2.9}$$

where μ is a measure on $(E, \mathcal{B})$.

Given two random variables X and Y we say that $X = Y$ almost surely if and only if:

$$\mathbb{P}\{\omega \in \Omega | X(\omega) = Y(\omega)\} = 1. \tag{2.10}$$

Given a real random variable X we can define a new measure μ_X on the Borel σ-algebra $\mathcal{B}$ in the following way:

$$\forall A \in \mathcal{B}, \quad \mu_X(A) = \mathbb{P}\{\omega \in \Omega | X(\omega) \in A\} . \tag{2.11}$$

The measure of μ_X is called the law (or distribution) of X.

2.2.1 Expectations

Let X be a random variable of $(\Omega, \mathcal{F}, \mathbb{P})$. One would like to compute an average value of X. If Ω is finite the average value of X is simply defined as:

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\ \mathbb{P}(\omega) . \tag{2.12}$$

If, as in most cases, Ω is infinite the sum appearing in (2.12) must be modified in an integral. For this purpose we assume for the moment that $0 < X(\omega) < \infty$ for all $\omega \in \Omega$. Let $\Pi = \{y_0, y_1, ..., y_k, ...\}$ with $0 = y_0 < y_1 < ... < y_k < ...$ be a partition of $[0, \infty]$. We set:

$$A_k = \{\omega \in \Omega | y_k \leq X(\omega) < y_{k+1}\} \tag{2.13}$$

and we define the expected value of X as:

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) \equiv \lim_{||\Pi|| \to 0} \sum_{k=1}^{\infty} y_k \mathbb{P}(A_k) , \tag{2.14}$$

where $||\Pi||$ is the maximal distance between the y_k partition points. If the random variable X is not a positive function we use the decomposition $X(\omega) = X^+(\omega) - X^-(\omega)$, where:

$$X^+(\omega) = \max[X(\omega), 0] , \quad X^-(\omega) = \max[-X(\omega), 0] . \tag{2.15}$$

and we define:

$$\int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \int_\Omega X^+(\omega)\, d\mathbb{P}(\omega) - \int_\Omega X^-(\omega)\, d\mathbb{P}(\omega) \tag{2.16}$$

If $\int_\Omega X^+(\omega)\, d\mathbb{P}(\omega)$ and $\int_\Omega X^-(\omega)\, d\mathbb{P}(\omega)$ are both finite we say that X is integrable.

The integral defined in equation (2.16) is called Lebesgue integral. The most notably properties of the Lebesgue integral is that it is a linear operator and the condition of integrability of X is equivalent to the condition of integrability of $|X|$, that is X is integrable if and only if $|X|$ is integrable.

2.3 Convergence of Random Variables

Econometric estimates are usually improved increasing the data density, that is increasing the frequency of observations of the random variable under study[1]. As a consequence a huge part of the results concerning econometric theory are asymptotic and a notion of convergence for random process is needed. The most simple notion of convergence is the *pointwise* convergence: given a sequence $(X_n)_{n\in\mathbb{N}}$ valued in a normed space like $\mathbb{R}^d$, one would require that for each $\omega \in \Omega$ the sequence $X_n(\omega)$ converges to a limit variable $X(\omega)$. This notion does not take into account that many events could have probability zero and then are negligible. The following definition of convergence is a valuable one that avoids this problem:

Definition 2.3.1 *A sequence of real random variables $(X_n)_{n\in\mathbb{N}}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and valued in $(E, \mathcal{B}, \mu)$ is said to converge **almost surely** to a random variable X if:*

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1. \tag{2.17}$$

A weaker definition of convergence is the convergence in probability:

Definition 2.3.2 *A sequence of real random variables $(X_n)_{n\in\mathbb{N}}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and valued in $(E, \mathcal{B}, \mu)$ is said to converge **in probability** to a random variable X if:*

$$\forall \epsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0. \tag{2.18}$$

The convergence in probability is usually denoted by $X_n \xrightarrow{\mathbb{P}} X$. Almost sure convergence implies convergence in probability. On the contrary convergence in probability does not imply almost sure convergence.

Many theorems proposed in this thesis concern the asymptotic distribution of a random variable in the limit of infinite frequency of observations. For this reason it is worth to introduce the definition of convergence in distribution for a real random variable:

Definition 2.3.3 *A sequence of real random variables $(X_n)_{n\in\mathbb{N}}$ valued in $(E, \mathcal{B}, \mu)$ is said to converge **in distribution** to a random variable X if, for any bounded function $f : E \to \mathbb{R}$, it happens that:*

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] \tag{2.19}$$

The convergence in distribution is usually denoted by $X_n \xrightarrow{d} X$. Note that while the almost sure convergence and the convergence in probability are well defined only for those sequence defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, this is not the case for the convergence in distribution.

[1] In most cases the frequency of observations cannot be too high because of the presence of different sources of noise.

2.4 Stochastic Processes and Information

Assets, indexes, interest rates and other financial variables are usually modelled as stochastic processes. A stochastic process is a family $(X_t)_{t\in[0,T]}$ of random variables indexed by time.

Suppose we are observing a stochastic process X_t. As observation takes place we acquire more and more information about the system under study. The precise mathematical formulation of this feature is obtained introducing the filtration:

Definition 2.4.1 *A filtration or information flow on $(\Omega, \mathcal{F}, \mathbb{P})$ is an increasing family of $\sigma-$algebras $(\mathcal{F}_t)_{t\in[0,T]}$ such that:*

$$\forall\ 0 \leq s \leq t, \quad \mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F} \tag{2.20}$$

The idea underlying the definition (2.20) is that, as the time flows, the information at our disposal increase and then the probability of an event will change. This change is modelled taking as given and invariable the probability measure $\mathbb{P}$ and conditioning all the expectations values on the information $\mathcal{F}_t$ available at time t.

If the information held in a σ-algebra are enough to determine the value of a random variable X we will say that the variable is measurable with respect to that σ-algebra. This idea can be formalized introducing the σ-algebra generated by a random variable X:

Definition 2.4.2 *Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The σ-algebra generated by X, denoted by $\sigma(X)$, is the collection of all subset of ω of the form $\{\omega \in \Omega | X(\omega) \in B\}$ where B ranges over the Borel subsets of $\mathbb{R}$.*

We can now introduce the notion of independence between σ-algebras and random variables:

Definition 2.4.3 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{H}$ and $\mathcal{G}$ be σ-algebras such that $\mathcal{H} \subseteq \mathcal{F}$ and $\mathcal{G} \subseteq \mathcal{F}$. We say that $\mathcal{H}$ and $\mathcal{G}$ are independent if:*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\ \mathbb{P}(B) \quad \text{for all } A \in \mathcal{G}, B \in \mathcal{H}. \tag{2.21}$$

Let X and Y be two random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, we say that X and Y are independent if $\sigma(X)$ and $\sigma(Y)$ are independent. We say that the random variable X is independent of the σ-algebra $\mathcal{G}$ if $\sigma(X)$ and $\mathcal{G}$ are independent.

The second step is to introduce the definition of a random variable measurable with respect to a σ-algebra:

Definition 2.4.4 *Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}$ a σ-algebra contained in $\mathcal{F}$. We say that X is $\mathcal{G}$-measurable if every set in $\sigma(X)$ is also in $\mathcal{G}$.*

All events ω belonging to a filtration $\mathcal{F}_t$ are such that, given the information $\mathcal{F}_t$ at time t, the observer can establish if the event has occurred or not. This fact leads to the natural definition of adapted stochastic process:

Definition 2.4.5 *A stochastic process X_t is said to be adapted to the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ if X_t is $\mathcal{F}_t$-measurable $\forall \quad t \in [0,T]$.*

In the same way it is possible to introduce the information flow generated by the observation of X_t:

Definition 2.4.6 *The filtration $\left(\mathcal{F}_t^X\right)_{t\in[0,T]}$ associated to the process X_t is the $\sigma-$algebra generated by the past values of the process:*

$$\mathcal{F}_t^X = \sigma\left(X_s, s \in [0,t]\right). \tag{2.22}$$

We have now the correct mathematical tools to introduce the conditional expectation of a random variable with respect to a $\sigma-$algebra $\mathcal{G}$:

Definition 2.4.7 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G}$ a σ-algebra of $\mathcal{F}$. Let X be a random variable that is either nonnegative or integrable (see Section 2.2.1). We define the conditional expectation of X given the information $\mathcal{G}$, denoted by $\mathbb{E}[X|\mathcal{G}]$, any* ***random variable*** *such that:*

1. *$\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable,*
2. *$\forall A \in \mathcal{G}$ it happens that:*

$$\int_A \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P}(\omega) = \int_A X(\omega)\, d\mathbb{P}(\omega). \tag{2.23}$$

If $\mathcal{G}$ is the σ-algebra generated by some other random variable W (i.e. $\mathcal{G} = \sigma(W)$) we write $\mathbb{E}[X|W]$ instead of $\mathbb{E}[X|\sigma(W)]$.

The first property guarantees that, although the estimate of X based on $\mathcal{G}$ is itself a random variable, the value of such estimate can be determined from the information held in $\mathcal{G}$. The second property ensures that $\mathbb{E}[X|\mathcal{G}]$ is indeed an estimate of X. As the number of sets in $\mathcal{G}$ increase the σ-algebra becomes finer and the estimate $\mathbb{E}[X|G]$ improves, approaching X. It can be shown that $\mathbb{E}[X|\mathcal{G}]$ always exists and it is unique except for a set of zero measure (almost surely unique). It can be shown also that conditional expectation has the following properties:

- (**Linearity**)

$$\mathbb{E}[c_1 X_1 + c_2 X_2|\mathcal{G}] = c_1\,\mathbb{E}[X_1|\mathcal{G}] + c_2\,\mathbb{E}[X_2|\mathcal{G}]. \tag{2.24}$$

- (**Taking out what is known**) If X, Y and XY are integrable random variables and X is $\mathcal{G}$-measurable then:

$$\mathbb{E}[XY|\mathcal{G}] = X\,\mathbb{E}[Y|\mathcal{G}]. \tag{2.25}$$

- (**Iterated conditioning**) If $\mathcal{H}$ is σ-algebra contained in $\mathcal{G}$ then:

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]. \tag{2.26}$$

- (**Independence**) If X is $\mathcal{G}$-independent then:

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]. \tag{2.27}$$

2.5 Martingales and Predictable Process

In order to include model with abrupt changes it will be useful to deal with non-continuous stochastic processes. For this purpose we introduce a particular class of functions: the *cadlag* functions. A functions $f : [0,T] \to \mathbb{R}$ is said to be cadlag (from the French *continue à droite limit a gauche*) if it is right-continuous with left limit:

$$\exists f(t_-) = \lim_{s \uparrow t} f(s), \exists f(t_+) = \lim_{s \downarrow t} f(s), \tag{2.28}$$

and $f(t) = f(t_+)$. Discontinuity points of a cadlag function will be interpreted as "jumps". In particular if t is a discontinuity point the jump amplitude at instant t is:

$$\Delta f(t) = f(t) - f(t_-). \tag{2.29}$$

The choice of right-continuous functions has a deep motivation. We want to model jump as unpredictable sudden variations of state variables. Therefore left-continuous processes have to be excluded because the value of the variables at the jump instants could be approximated by the previous one.

Martingales are particular stochastic processes widely used in finance. For the sake of simplicity we give first the formal mathematical definition:

Definition 2.5.1 *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ a cadlag process X_t is said to be a martingale if it is $(\mathcal{F}_t)_{t\in[0,T]}$-adapted, $\mathbb{E}[|X_t|]$ is finite for any $t \in [0,T]$ and:*

$$\forall s > t, \quad \mathbb{E}[X_s|\mathcal{F}_t] = X_t. \tag{2.30}$$

The property (2.30), which identifies the martingales, can be rephrased as follows: the best guess for the future expected value of X_t conditioned on the information available today is its actual value. As a consequence of equation (2.30) a martingale has constant expectation: $\mathbb{E}[X_t] = \mathbb{E}[X_0]$. This does not mean that any driftless process is a martingale.

We can interpret a stochastic process $X_t(\omega)$ as a function defined on $[0,T] \times \Omega$. This leads to the definition of predictable processes:

Definition 2.5.2 *The predictable σ-algebra is the σ-algebra $\mathcal{P}$ generated on $[0,T] \times \Omega$ by all adapted left-continuous processes. A process $[0,T] \times \Omega \to \mathbb{R}^d$ which is measurable with respect to $\mathcal{P}$ is called a predictable process.*

As anticipated the name "predictable" is quite intuitive. Any left-continuous process is predictable (by definition) and the value X_t is "announced" by its previous values: $\lim_{s\uparrow t} X_s = X_t$.

2.6 Jump Process

We focus on a particular class of jump processes. We want the discontinuous variations of the state variable to be memoryless. As will be widely discussed further jumps in financial assets can be interpreted as the market response to unpredictable, uncorrelated events. For this purpose we start our discussion with the introduction of time random variables with exponential distribution.

A random time is nothing else a positive random variable $\tau > 0$ which represents the time at which some event is going to take place. When we are provided with an information flow $(\mathcal{F}_t)_{t\in[0,T]}$ it is natural to investigate if the knowledge of $(\mathcal{F}_t)_{t\in[0,T]}$ it is enough to determine whether the event has happened or not. If the answer is positive the random time T is called a non-anticipating random time or *stopping time*. To be more precise a random time τ is called a nonanticipating random time or a $\mathcal{F}_t$-stopping time if:

$$\forall t \geq 0, \ \ \{\omega \in \Omega | \tau(\omega) \leq t\} \in \mathcal{F}_t. \tag{2.31}$$

Given an adapted process X_t and a $\mathcal{F}_t$-stopping time τ one can define a new process $X_{\tau\wedge t}$, the process X stopped at τ, by[2]:

$$X_{\tau\wedge t} = X_t \text{ if } t < \tau, \ X_{\tau\wedge t} = X_\tau \text{ if } t \geq \tau. \tag{2.32}$$

We say that the time random variable τ has an exponential distribution if its density is given by:

$$f(\tau) = \begin{cases} \lambda e^{-\lambda\tau} & \text{if } \tau \geq 0 \\ 0 & \text{if } \tau < 0, \end{cases}$$

where $\lambda > 0$. The variable τ is said to be an exponential random variable. The parameter λ is related to the half-life h of the variable τ by the relationship $h = \frac{\log 2}{\lambda}$. The expected value of an exponential random variable can be easily computed as:

$$\mathbb{E}[\tau] = \int_0^{+\infty} \tau f(\tau)\, d\tau = \frac{1}{\lambda}, \tag{2.33}$$

while the cumulative distribution function is given by:

$$\mathbb{P}\{\tau \leq t\} = \int_0^t f(\tau)\, d\tau = 1 - e^{-\lambda t} \quad t \geq 0, \tag{2.34}$$

that is:

$$\mathbb{P}\{\tau > t\} = e^{-\lambda t} \quad t \geq 0. \tag{2.35}$$

A process (such as a radioactive decay) which has random exponential time-occurrences is *memoryless*. This means that, after s time units have passed, the probability to wait an additional t time units is the same as the probability of having to wait t time units starting at time 0. This property in formula reads as:

$$\mathbb{P}\{\tau > t+s | \tau > s\} = \frac{\mathbb{P}\{\tau > t+s \text{ and } \tau > s\}}{\mathbb{P}\{\tau > s\}} = \frac{\mathbb{P}\{\tau > t+s\}}{\mathbb{P}\{\tau > s\}} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}\{\tau > t\}. \tag{2.36}$$

To model jump arrivals in a econometric context we start with a sequence $(\tau_n)_{n\in\mathbb{N}}$ of exponential random variables (*inter arrival times*) with $\mathbb{E}[\tau_n] = \frac{1}{\lambda} \quad \forall n \in \mathbb{N}$. The first jump occurs at time τ_1, the second jump occurs τ_2 time units after the first jump and so on. In this context the n-th jump occurs at time:

[2]We adopt the notation $a \wedge b = \inf(a, b)$.

$$S_n = \sum_{k=1}^{n} \tau_k. \tag{2.37}$$

The random variables $(S_n)_{n\in\mathbb{N}}$ are called the *arrival times*. The counting *Poisson process* $N(t)$ is defined as:

$$N(t) = \begin{cases} 0 & \text{if } 0 \le t < S_1 \\ 1 & \text{if } S_1 \le t < S_2 \\ \vdots & \\ n & \text{if } S_n \le t < S_{n+1} \\ \vdots & \end{cases}$$

Defined as in (2.6) the counting process $N(t)$ is right-continuous i.e. $N(t) = \lim_{s\downarrow t} N(s)$. Because the expected time between arrivals is $\frac{1}{\lambda}$ the jumps arrive at an average rate of λ per unit of time. The parameter λ is called the *intensity* of the Poisson process $N(t)$. The distribution density of the Poisson process (2.6) can be derived from the distribution density of the random variables $(S_n)_{n\in\mathbb{N}}$, which is derived in the following:

Proposition 2.6.1 *For all $n \ge 1$ the distribution density of the random variable S_n is:*

$$D_{S_n}(s) = \frac{(\lambda s)^{n-1}}{(n-1)!}\,\lambda\, e^{-\lambda s}, \quad s \ge 0 \tag{2.38}$$

Proof. For $n = 1$ we have that $S_1 = \tau_1$ has an exponential distribution which can obtained from (2.38) with $n = 1$. Thus the proposition is true for $n = 1$. Now suppose that (2.38) is true for n and prove it for $n + 1$. This means that we know that the distribution of S_n is given by (2.38) and we want to compute the distribution of $S_{n+1} = S_n + \tau_{n+1}$. Given that S_n and τ_{n+1} are independent random variables the distribution of their sum is given by the convolution:

$$\int_0^s D_{S_n}(v)\, f(s-v)\, dv = \int_0^s \frac{(\lambda v)^{n-1}}{(n-1)!}\,\lambda\, e^{-\lambda v}\,\lambda\, e^{-\lambda(s-v)}\, dv = \tag{2.39}$$

$$= \frac{\lambda^{n+1} e^{-\lambda s}}{(n-1)!}\int_0^s v^{n-1}\, dv = \frac{(\lambda s)^n}{n!}\,\lambda\, e^{-\lambda s} = D_{S_{n+1}}(s) \tag{2.40}$$

□

From Proposition 2.6.1 it is possible to derive the distribution density for the counting process $N(t)$.

Proposition 2.6.2 *The Poisson process $N(t)$ with intensity λ has the distribution:*

$$\mathbb{P}\{N(t) = k\} = \frac{(\lambda t)^k}{k!}\, e^{-\lambda t} \tag{2.41}$$

Proof. For $k \ge 1$ we have $N(t) \ge k \Leftrightarrow S_k \le t$ that is $\mathbb{P}\{N(t) \ge k\} = \mathbb{P}\{S_k \le t\} = \int_0^t D_{S_k}(s)\, ds$. Similarly $\mathbb{P}\{N(t) \ge k+1\} = \mathbb{P}\{S_{k+1} \le t\} = \int_0^t D_{S_{k+1}}(s)\, ds$. Integration by part of (2.38) for $n = k + 1$ yields:

$$\mathbb{P}\{S_{k+1} \le t\} = \int_0^t \frac{(\lambda s)^k}{(k)!}\,\lambda\, e^{-\lambda s}\, ds = -\frac{(\lambda t)^k}{k!}\, e^{-\lambda t} + \int_0^t \frac{(\lambda s)^{k-1}}{(k-1)!}\,\lambda\, e^{-\lambda s}\, ds = \tag{2.42}$$

$$= -\frac{(\lambda t)^k}{k!}\, e^{-\lambda t} + \mathbb{P}\{S_k \le t\} = -\frac{(\lambda t)^k}{k!}\, e^{-\lambda t} + \mathbb{P}\{N(t) \ge k\}. \tag{2.43}$$

That is:

$$\mathbb{P}\{N(t) \geq k+1\} = -\frac{(\lambda t)^k}{k!} e^{-\lambda t} + \mathbb{P}\{N(t) \geq k\}. \tag{2.44}$$

This implies that for $k \geq 1$:

$$\mathbb{P}\{N(t) = k\} = \mathbb{P}\{N(t) \geq k\} - \mathbb{P}\{N(t) \geq k+1\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \tag{2.45}$$

For $k = 0$ we have that $\mathbb{P}\{N(t) = 0\} = \mathbb{P}\{S_1 > t\} = \mathbb{P}\{\tau_1 > t\} = e^{-\lambda t}$ which is the (2.41) with $k = 0$. □

The Poisson process $N(t)$ has the following properties, if $0 \leq s \leq t$ then:

$$\mathbb{P}\{N(t) - N(s) = k\} = \frac{\lambda^k (t-s)^k}{k!} e^{-\lambda (t-s)}, \tag{2.46}$$

$$\mathbb{E}[N(t) - N(s)] = \lambda (t-s), \tag{2.47}$$

$$\Sigma[N(t) - N(s)] = \lambda (t-s), \tag{2.48}$$

where $\Sigma[X] \equiv \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$ denotes the variance of X.

In what follows will be useful the theorem that states the martingale property of the so called *compensated Poisson process*:

Theorem 2.6.3 *If $N(t)$ is a Poisson process with intensity λ then the compensated Poisson process defined by $M(t) \equiv N(t) - \lambda t$ is a martingale w.r.t. the filtration $\mathcal{F}_t$ generated by $N(t)$.*

Proof. Let $0 \leq s < t$. The increment $N(t) - N(s)$ is independent from the filtration $\mathcal{F}_s$ thus $\mathbb{E}[N(t) - N(s) | \mathcal{F}_s] = \mathbb{E}[N(t) - N(s)]$. Moreover we know that has expected value $\mathbb{E}[N(t) - N(s)] = \lambda (t-s)$. Thus:

$$\mathbb{E}[M(t) | \mathcal{F}_s] = \mathbb{E}[M(t) - M(s) + M(s) | \mathcal{F}_s] = \mathbb{E}[M(t) - M(s) | \mathcal{F}_s] + \mathbb{E}[M(s) | \mathcal{F}_s] = \tag{2.49}$$

$$= \mathbb{E}[N(t) - N(s) - \lambda (t-s) | \mathcal{F}_s] + M(s) = \tag{2.50}$$

$$= \mathbb{E}[N(t) - N(s)] - \lambda (t-s) + M(s) = M(s). \quad \square \tag{2.51}$$

The definition of the Poisson process presented above is not enough to define a jump process useful for financial applications. We need an extra process to model the jump sizes at each jump time. Therefore given a Poisson process $N(t)$ let $(Y_n)_{n \in \mathbb{N}}$ be a sequence of i.i.d. random variables with expected value $\beta = \mathbb{E}[Y_i]$. It is usually assumed that the sequence $(Y_n)_{n \in \mathbb{N}}$ is independent of the Poisson process. The *compound Poisson process* is defined as:

$$J(t) = \sum_{n=1}^{N(t)} Y_i, \quad t \geq 0. \tag{2.52}$$

With this definition jumps in $J(t)$ and $N(t)$ occur at the same times but the jump sizes of the Poisson process is always 1 whereas they are random variables for the process $J(t)$. As for the Poisson process the increment $J(t) - J(s)$ for $0 \leq s < t$ are independent of $J(s)$. Moreover it is easy to see that:

$$\mathbb{E}[J(t)] = \beta \lambda t. \tag{2.53}$$

The analogue of Theorem (2.6.3) is valid for the compensated compound process $J(t) - \lambda t$:

Theorem 2.6.4 *If $J(t)$ is a compound Poisson process with intensity λ then the compensated compound Poisson process defined by $L(t) \equiv J(t) - \lambda t$ is a martingale w.r.t. the filtration $\mathcal{F}_t^Q$ generated by $J(t)$.*

Proof. The proof replicates the steps used for Theorem (2.6.3). □

2.7 The Poisson Random Measure

The Poisson process $N(t)$ defined in the previous section is a counting process, it simply tell us the number of jumps between 0 and t. If $(\tau_n(\omega))_{n\in\mathbb{N}}$ are the jumps instants, then:

$$N(t,\omega) = \#\{i|\tau_i(\omega) \in [0,t]\}. \tag{2.54}$$

This counting procedure defines a measure $M(\omega,.)$ on $[0,\infty[$ in the following way:

Definition 2.7.1 *For any measurable set $A \in \mathcal{B}$ the jump random measure of A is defined as:*

$$M(\omega,A) \equiv \#\{i|\tau_i(\omega) \in A\}. \tag{2.55}$$

The measure $M(\omega,.)$ is a random measure, in the sense that it depends on the uncertainty parameter ω. With this definition we have $N(t,\omega) = M(\omega,[0,t])$. The intensity λ of the poisson process determines the average value of the associated random measure: $\mathbb{E}[M(\omega,A)] = \lambda|A|$, where $|A|$ is the Lebesgue measure of A. We will use the notation $M(\omega,A) = \int_A M(\omega,ds)$. A random measure can be also associated to the compensated Poisson process defining:

$$\tilde{M}(\omega,A) = M(\omega,A) - \lambda|A|. \tag{2.56}$$

The measure $\tilde{M}(\omega,A)$ is the centered version of $M(\omega,A)$ and it is neither positive nor integer. Moreover it verifies: $\mathbb{E}\left[\tilde{M}(\omega,A)\right] = 0$ and $\Sigma\left[\tilde{M}(\omega,A)\right] = \lambda|A|$. Following the definition (2.1.5) of the Dirac measure, the jump measure associated to a Poisson counting process with jump instants $(\tau_n(\omega))_{n\in\mathbb{N}}$ can be written in a more compact form:

$$M = \sum_i \delta_{\tau_i(\omega)}. \tag{2.57}$$

More generally it is possible to define a measure on $[0,\infty[\times \mathbb{R}^d$ associated to a compound Poisson process X_t with jump instants $(\tau_n(\omega))_{n\in\mathbb{N}} \in [0,\infty[$ and jump amplitude $(Y_n(\omega))_{n\in\mathbb{N}} \in \mathbb{R}^d$. The natural extension of (2.57) is:

$$J_X = \sum_i \delta_{(\tau_i,Y_i)}, \tag{2.58}$$

where, for simplicity, it has been suppressed the dependence on ω. It is possible to define an integral with respect to the measure (2.58) for any regular enough function $f : [0,\infty[\times \mathbb{R}^d \to \mathbb{R}^d$:

$$\int_0^t \int_{\mathbb{R}^d} f(s,y)\ J_X(ds \times dy) = \sum_{\{i|\tau_i \le t\}} f(\tau_i, Y_i). \tag{2.59}$$

In this way, the sum of the squared increments of X in the time window $[0,T]$ is given by:

$$\int_0^T \int_{\mathbb{R}} y^2 \, J_X \left(dtdy\right) = \sum_{t \in [0,T]} \left(\Delta X_t\right)^2, \tag{2.60}$$

where:

$$\Delta X_t = X_t - \lim_{s \uparrow t} X_s. \tag{2.61}$$

The sum on the right hand side of equation (2.60) is a very important quantity whose meaning will be explained later.

2.8 Brownian Motion and Stochastic Integrals.

Brownian motion is the fundamental brick for modelling financial asset variations. It has been widely used among researchers and practitioners due to its capability to capture all the essential features of financial assets, especially when considered on large time scales[3]. In what follows we give the definition of Brownian motion:

Definition 2.8.1 *A stochastic process W is called a Brownian Motion if the following conditions hold:*

- $W\left(0\right) = 0$.
- *W has independent increments, that is if $r \leq s \leq t \leq u$ then $W_u - W_t$ and $W_s - W_r$ are independent stochastic processes.*
- *For $s \leq t$ the stochastic process $W_t - W_s$ has the Gaussian distribution $N\left(0, \sqrt{t-s}\right)$.*
- *W has continuous trajectories.*

It can be shown that the Brownian motion is nowhere derivable with probability 1 and then it is not possible to define a derivative function or a differential in the usual sense. Despite this fact it is possible to define the integral of a stochastic process g with respect to the Brownian motion. For this purpose it is required to identify a class of integrable functions: the $\mathcal{L}^2\left[a,b\right]$ class.

Definition 2.8.2 *We say that the stochastic process g belongs to the $\mathcal{L}^2\left[a,b\right]$ class if the following conditions holds:*

- $\mathbb{E}\left[\int_a^b g_t^2 dt\right] < \infty$.
- *The process g is adapted to the $\mathcal{F}_t^W$ filtration.*

We say that g belongs to the class $\mathcal{L}^2$ if $g \in \mathcal{L}^2\left[0,t\right]$ for all $t > 0$.

[3] On short time scales, like days or hours, discontinuous variations are not negligible. For more details see the introduction of Cont and Tankov (2004)

Firstly, the integral is defined on simple processes. Suppose that $g \in \mathcal{L}^2[a,b]$ is constant in the sub-intervals of a deterministic partition $a = t_0 < t_1 < ... < t_n = b$ of $[a,b]$. This means that $g_t = g_{t_k}$ for $t_k \leq t < t_{k+1}$. Such a process is said to be *simple*. The integral of a simple process with respect to the Brownian motion is naturally defined as:

$$\int_a^b g_t \, dW_t \equiv \sum_{k=0}^{n-1} g_{t_k} \left[W_{t_{k+1}} - W_{t_k} \right]. \tag{2.62}$$

Note that the increments of the Brownian motion entering the sum on the right hand side of expression (2.62) are forward increments. This has a deep economic interpretation: g_{t_k} is $\mathcal{F}_{t_k}^W$-adapted while the increment $W_{t_{k+1}} - W_{t_k}$ is not revealed at t_k. Moreover $W_{t_{k+1}} - W_{t_k}$ is independent of g_{t_k} because the Brownian motion has independent increments. All these facts imply that:

$$\mathbb{E}\left[\int_a^b g_t \, dW_t \right] = \sum_{k=0}^{n-1} \mathbb{E}\left[g_{t_k} \right] \mathbb{E}\left[W_{t_{k+1}} - W_{t_k} \right] = 0. \tag{2.63}$$

The integral with respect to the Brownian motion defined with forward increments has always zero expectation. It can be shown that for all $g \in \mathcal{L}^2[a,b]$ there is a sequence of simple processes g^n such that:

$$\int_a^b \mathbb{E}\left[\{ g_t^n - g_t \}^2 \right] dt \to 0. \tag{2.64}$$

The series g^n approximate g in the $\mathcal{L}^2$ sense and for each n the integral $Z_n = \int_a^b g_t^n \, dW_t$ is well defined. Therefore we define:

$$\int_a^b g_t \, dW_t \equiv \lim_{n \to \infty} \int_a^b g_t^n \, dW_t. \tag{2.65}$$

It is possible to show that such a limit exists in the $\mathcal{L}^2$ sense. The integral defined in (2.65) has the following properties:

Proposition 2.8.3 *If $g \in \mathcal{L}^2$ then:*

- $\mathbb{E}\left[\int_a^b g_t \, dW_t \right] = 0.$
- $\mathbb{E}\left[\left(\int_a^b g_t \, dW_t \right)^2 \right] = \int_a^b \mathbb{E}\left[g_t^2 \right] \, dt.$
- $\int_a^b g_t \, dW_t$ *is $\mathcal{F}_b^W$-measurable.*

The first property is inherited from relation (2.63) which holds for simple processes. As anticipated it is a consequence of the forward increments defining the stochastic integral (2.65).

Usually it is assumed that the temporal dynamics of the log-price X_t of a financial assets is driven by a drift term $\mu(s, X_s) \, ds$ plus a Brownian diffusion $\sigma(s, X_s) \, dW_s$:

$$X_t = X_{t=0} + \int_0^t \mu(s, X_s) \, ds + \int_0^t \sigma(s, X_s) \, dW_s, \tag{2.66}$$

where the diffusion amplitude $\sigma(s, X_s)$ is allowed to be a stochastic process. The amplitude $\sigma(s, X_s)$ is commonly referred as "instantaneous volatility" and we will adopt this nomenclature. A less cumbersome notation for expression (2.66) is:

$$\begin{cases} dX_t = \mu\left(t, X_t\right)\, dt + \sigma\left(t, X_t\right)\, dW_t, \\ X_{t=0} = X_0. \end{cases} \tag{2.67}$$

As anticipated it is not possible to define a differential form dW_t for a Brownian motion, therefore notation (2.67) is just intuitive and must always be interpreted as a simplified version of (2.66).

Martingales and stochastic integrals are related by the following result:

Lemma 2.8.4 *Assuming enough integrability a stochastic process X is a martingale if and only if:*

$$dX_t = \sigma\left(t, X_t\right)\, dW_t, \tag{2.68}$$

i.e. X has no drift term.

This means that the stochastic process defined in (2.67) is not a martingale. For this reason we give the definition of a wider class of stochastic processes: semimartingales. For this purpose we first introduce the simple predictable processes:

Definition 2.8.5 *A stochastic process $(\phi_t)_{t\in[0,T]}$ is called a simple predictable process if it can be represented as:*

$$\phi_t = \phi_0\, 1_{t=0} + \sum_{i=0}^{n-1} \phi_i\, 1_{]\tau_i,\tau_{i+1}]}\left(t\right), \tag{2.69}$$

where $0 = \tau_0 < \tau_1 < ... < \tau_n = T$ are nonanticipating random times and each ϕ_i is a $\mathcal{F}_{\tau_i}$-measurable random variable.

The set of simple predictable processes on $[0,T]$ will be denoted by $\mathbb{S}\left[0,T\right]$. Semimartingales are thus introduced in the following:

Definition 2.8.6 *Let S_t be a cadlag adapted stochastic process. S_t is a semimartingale if the stochastic integral of simple predictable processes with respect to S_t:*

$$\phi_t = \phi_0\, 1_{t=0} + \sum_{i=0}^{n-1} \phi_i\, 1_{]\tau_i,\tau_{i+1}]}\left(t\right) \rightarrow \int_0^T \phi\, dS = \phi_0\, S_0 + \sum_{i=0}^{n-1} \phi_i\, \left(S_{\tau_{i+1}} - S_{\tau_i}\right) \tag{2.70}$$

verifies the following continuity property: for every ϕ^n, $\phi \in \mathbb{S}\left[0,T\right]$ if

$$\lim_{n\rightarrow\infty} \sup_{(t,\omega)\in[0,T]\times\Omega} |\phi_t^n\left(\omega\right) - \phi_t\left(\omega\right)| = 0 \text{ then } \mathbb{P} - \lim_{n\rightarrow\infty} \int_0^T \phi_t^n dS_t = \int_0^T \phi_t dS_t, \tag{2.71}$$

where $\mathbb{P} - \lim$ indicates the limit in probability as in definition (2.3.2).

If an investor trades at random times $0 = \tau_0 < \tau_1 < ... < \tau_n = T$ detaining a quantity ϕ_i of the asset during the period $]\tau_i, \tau_{i+1}]$ then the capital gain resulting from market fluctuations is given by the integral (2.70) of the simple predictable process ϕ_t with respect to the asset S_t. Modelling assets as semimartingales guarantees that the capital gain of trading strategy is stable with respect to small error in the composition of the portfolio. Definition (2.8.6) could appear rather abstract. Finite variation processes help to better classify the class of semimartingales. The total variation of a function $f_t : [0,T] \rightarrow \mathbb{R}^d$ is defined as:

$$\mathcal{TV}\left[f_t\right] = \sup \sum_{i=0}^{n} \left| f\left(t_{i+1}\right) - f\left(t_i\right) \right|, \tag{2.72}$$

where the supremum is taken over all finite partitions $0 = t_0 < t_1 < ... < t_n = T$ of $[0, T]$. A stochastic process S_t is said to be finite variation if:

$$\mathbb{P}\left\{\omega \in \Omega | \mathcal{TV}\left[S_t\left(\omega\right)\right] < \infty\right\} = 1, \tag{2.73}$$

that is S_t has almost-surely finite variation trajectories.

The following lemma links the finite variation processes with the semimartingales:

Lemma 2.8.7 *Every finite variation process is a semimartingale.*

All assets modelled in this work are assumed to be finite variations process and then semimartingales.

Note that while the integral with respect to a Brownian motion has been defined for cadlag processes, the integral with respect to a semimartingale has been defined for simple predictable processes which are caglad. This distinction derives from the fact that instantaneous volatility is itself a stochastic process allowed to have unpredictable jumps, and then must be chosen among the cadlag functions. However integrals with respect to an asset (semimartingale) are interpreted as portfolio capital gains, and then are defined for simple predictable processes, the trading strategy being trivially predictable. A continuity property of the integral with respect to a semimartingale allow us to extend such integral to all caglad processes:

Proposition 2.8.8 *If S_t is a semimartingale then for every $\phi^n,\ \phi \in \mathbb{S}\left[0, T\right]$ such that:*

$$\lim_{n \to \infty} \sup_{(t,\omega) \in [0,T] \times \Omega} \left| \phi_t^n\left(\omega\right) - \phi_t\left(\omega\right) \right| = 0, \tag{2.74}$$

it happens that:

$$\mathbb{P} - \lim_{n \to \infty} \sup_{t \in [0,T]} \left| \int_0^t \phi_t^n dS_t - \int_0^t \phi_t dS_t \right| = 0. \tag{2.75}$$

It can be shown that every caglad process ϕ can be approximated by a sequence of simple predictable processes $\phi^n \in \mathbb{S}\left[0, T\right]$ in the sense of equation (2.74) (uniform convergence). Proposition (2.8.8) allows us to define the integral $\int_0^t \phi_t dS_t$ as the limit of $\int_0^t \phi_t^n dS_t$ in the sense of equation (2.75).

2.9 Quadratic Variation and Realized Volatility

The continuous model $dX_t = \mu\left(t, X_t\right) dt + \sigma\left(t, X_t\right) dW_t$ requires the introduction of an unobservable quantity: the instantaneous volatility $\sigma\left(t, X_t\right)$. This quantity gives a measure of the uncertainty held in X_t and economic theory links it with the information flow and the size of the economy. For these reasons a task of a great importance is to obtain a good estimation of the volatility dynamics. The typical approach consists in estimate the daily integrated volatility, that is $\int_0^T \sigma_s\, ds$ with $T = 1$ day, starting from the observation of X_t on a grid of equally spaced instants $\Pi^n = \left\{t_k = k\,\delta | k = 0, 1, ..., n \text{ and } n = \lfloor \frac{T}{\delta} \rfloor\right\}$. The *realized volatility* $V_X\left[\Pi^n\right]$ of X_t on the grid Π^n is defined as the sum of squared increments:

$$V_X(\Pi^n) = \sum_{t_k \in \Pi^n} \left(X_{t_{k+1}} - X_{t_k}\right)^2. \tag{2.76}$$

Each terms of the sum (2.76) can be re-written as:

$$\left(X_{t_{k+1}} - X_{t_k}\right)^2 = X_{t_{k+1}}^2 - X_{t_k}^2 - 2\, X_{t_k}\, \left(X_{t_{k+1}} - X_{t_k}\right)^2. \tag{2.77}$$

Therefore the realized volatility (2.76) becomes:

$$V_X(\Pi^n) = X_T^2 - X_0^2 - 2 \sum_{t_k \in \Pi^n} X_{t_k}\, \left(X_{t_{k+1}} - X_{t_k}\right)^2. \tag{2.78}$$

When X is a semimartingale and $n \to \infty$ the realized variance approaches a quantity that can be computed using the following proposition:

Proposition 2.9.1 *Let S_t be a martingale and let ϕ be a caglad process. Let:*

$$\Pi^n = \{0 = \tau_0^n < \tau_1^n < ... < \tau_n^n = T\} \tag{2.79}$$

be a sequence of random partitions of $[0,T]$ such that $|\Pi^n| = \sup_k |\tau_{k+1}^n - \tau_k^n| \to 0$ almost surely when $n \to \infty$. Then:

$$\mathbb{P} - \lim_{n\to\infty} \sup_{t\in[0,T]} \left| \phi_0\, S_0 + \sum_{k=0}^{n-1} \phi_{\tau_k}\, \left(S_{\tau_{k+1}\wedge t} - S_{\tau_k \wedge t}\right) - \int_0^t \phi_{u_-}\, dS_u \right| = 0, \tag{2.80}$$

i.e. the process $\phi_0\, S_0 + \sum_{k=0}^{n-1} \phi_{\tau_k}\, \left(S_{\tau_{k+1}\wedge t} - S_{\tau_k \wedge t}\right)$ converges uniformly on $[0,T]$ in probability to $\int_0^t \phi_{u_-}\, dS_u$.

If X_t is a semimartingale then it is cadlag and the process $X_{u_-} = \lim_{t\uparrow u} X_t$ is caglad. Therefore applying proposition (2.9.1) to the sum (2.78) we obtain that:

$$\mathbb{P} - \lim_{n\to\infty} \sum_{t_k \in \Pi^n}^{0\le t_k < t} \left(X_{t_{k+1}} - X_{t_k}\right)^2 = [X,X]_t\,, \tag{2.81}$$

where the convergence is uniform in $t \in [0,T]$ and the process $[X,X]_t$ is defined as:

$$[X,X]_t \equiv X_t^2 - X_0^2 - 2 \int_0^t X_{u_-}\, dX_u. \tag{2.82}$$

In the case of a Brownian motion W_t we have $[W,W]_t = t$. Heuristically the sum (2.78) can be interpreted as a Riemann sum and the relation $[W,W]_t = t$ is formally written as: $dW^2 = dt$. Furthermore it can be shown that if $X_t = t$ then $[X,X]_t = 0$. This relation is formally written as: $dt^2 = 0$.

The cross-quadratic variation between two semimartingales X_t and Y_t is defined as:

$$[X,Y]_t \equiv X_t\, Y_t - X_0\, Y_0 - \int_0^t X_{u_-}\, dX_u - \int_0^t Y_{u_-}\, dY_u. \tag{2.83}$$

It can be shown that:

$$\mathbb{P} - \lim_{n\to\infty} \sum_{t_k \in \Pi^n}^{0\le t_k < t} \left(X_{t_{k+1}} - X_{t_k}\right)\, \left(Y_{t_{k+1}} - Y_{t_k}\right) = [X,Y]_t\,. \tag{2.84}$$

where the convergence is uniform in $t \in [0,T]$. The cross-quadratic variation between the Brownian motion W_t and the process $X_t = t$ is identically zero: $[W,t] \equiv 0$. This leads to the formal expression $dW\,dt = 0$. Summarizing we have the following useful **formal** identities:

$$dW^2 = dt, \tag{2.85}$$

$$dt^2 = 0, \tag{2.86}$$

$$dW\,dt = 0. \tag{2.87}$$

2.10 Lévy Process

Lévy processes are a broad class of stochastic processes which includes Poisson and Wiener processes. Moreover it is possible to show that Brownian motion and Poisson processes are the building blocks of any Lévy process which can be always decomposed as a superposition of a Brownian motion and a (possibly infinite) number of Poisson processes. We will not go deep in the theory of Lévy processes because it is outside our purposes. We are mainly interested in the definition of activity of a Lévy process because it will be used in different contexts.

Definition 2.10.1 *A cadlag stochastic process $(X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with value in $\mathbb{R}^d$ such that $X_0 = 0$ is called a Lévy process if it possesses the following properties:*

- *For every increasing sequence of times t_0,...,t_n, the random variables X_{t_0},$X_{t_1} - X_{t_0}$, ...,$X_{t_n} - X_{t_{n-1}}$, are independent.*
- *The law of $X_{t+h} - X_t$ does not depend on t.*
- *$\forall \epsilon > 0$, $\lim_{h\to 0} \mathbb{P}\left(|X_{t+h} - X_t| \geq \epsilon\right) = 0$.*

The third property does not imply in any way that the Lévy process is continuous. It is needed to exclude jumps at fixed deterministic time t such as calendar effects. Jumps may occur only at random times. The measure associated to a Lévy process is given in what follow:

Definition 2.10.2 *Let $(X_t)_{t\geq 0}$ a Lévy process on $\mathbb{R}^d$. The measure ν on $\mathbb{R}^d$ is defined by:*

$$\forall \;\; \mathcal{A} \in \mathbb{R}^d, \quad \nu(\mathcal{A}) = \mathbb{E}\left[\#\left\{t \in [0,1] : \Delta X_t \neq 0, \Delta X_t \in \mathcal{A}\right\}\right], \tag{2.88}$$

i.e. $\nu(\mathcal{A})$ is the expected number of jumps, per unit of time, whose size belongs to $\mathcal{A}$. We report here the decomposition theorem for Lévy process, which is fundamental for modelling purposes:

Theorem 2.10.3 *Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$ and ν its Lévy measure. Then:*

- *ν is a Radon measure (see Definition 2.1.4) on $\mathbb{R}^d \setminus \{0\}$ and verifies:*

$$\int_{|x|\leq 1} |x|^2 \nu(dx) < \infty \qquad \int_{|x|\geq 1} \nu(dx) < \infty \tag{2.89}$$

- *The jump measure of X, denoted by J_X, is a Poisson random measure on $[0,\infty[\times \mathbb{R}^d$ with intensity measure $\nu(dx)\,dt$.*

- *There exists a vector γ and a d-dimensional Brownian motion $(W_t)_{t\geq 0}$ with covariance matrix A such that:*

$$\begin{aligned} X_t &= \gamma\, t + W_t + X_t^l + \lim_{\epsilon \downarrow 0} \hat{X}_t^\epsilon, \ \textit{where} \\ X_t^l &= \int_{|x|\geq 1, s\in[0,t]} x\, J_x\left(ds \times dx\right) \ \textit{and} \\ \hat{X}_t^\epsilon &= \int_{\epsilon\leq|x|\leq 1, s\in[0,t]} x\left\{J_x\left(ds\times dx\right) - \nu\left(dx\right) ds\right\}. \end{aligned} \tag{2.90}$$

  The terms in the equation (2.90) are independent and the convergence in the last term is almost sure and uniform in t on $[0,T]$.

Theorem 2.10.3 states that every Lévy process can be decomposed using a Brownian motion and some jump process. Once one assumes that log-prices of financial assets are described by a Lévy process it is possible to use this useful decomposition. Moreover Definition (2.10.2) allows the introduction of the *Blumenthal-Getoor index* of a Lévy process:

$$\alpha \overset{def}{=} \inf\left\{\delta \geq 0 : \int_{|x|\leq 1} |x|^\delta \nu\left(dx\right) < +\infty\right\}. \tag{2.91}$$

By definition, $\alpha \in [0,2]$. The index defined by equation (2.91) is also referred as activity of the process, it measures how frenetic the jump activity is: the smaller the α the milder the activity. A finite activity Lévy pure jump process is defined as a a jump process with only a finite number of jumps in any finite time interval. The Blumenthal-Getoor index of such a process is equal to zero. Conversely an infinite activity jump process has an infinite number of jumps in any finite time interval and its Blumenthal-Getoor index is above zero.

2.11 Time Series Analysis

The theory presented till now concerns the continuous time behaviour of stochastic process. It is the fundamental brick of our analysis, however all empirical results are based on finite series sampled at discrete times. For this reason in this section we describe some fundamental ideas of time series analysis. The interested reader is referred to Brockwell and Davis (2001) for a more exhaustive approach to the argument.

2.11.1 Observed Data and Random Variables

Given an observed time series $\{x_t\}_{t\in\mathbb{N}}$ it is natural to suppose that each observation x_t is a possible realization of an abstract random variable $X\left(t,\omega\right)$:

Definition 2.11.1 *A time series model for the observed data $\{x_t\}_{t\in\mathbb{N}}$ is a specification of the joint distribution of a sequence of random variables X_t of which x_t is postulated to be a realization.*

Therefore a time series model is specified once is defined the set of probabilities:

$$\mathbb{P}\left\{X_1 \leq x_1, ..., X_n \leq x_n\right\}, \quad x_i \in (-\infty, \infty), \quad n = 1, 2, ... \tag{2.92}$$

The simplest model one could think is made of a sequence of independent and identically distributed random variables. We will refer to this model as *i.i.d. noise*. Therefore if X_1, X_2,..., are i.i.d. random variable then we have:

$$X_i \quad i.i.d. \Rightarrow \mathbb{P}\{X_1 \leq x_1 ... X_n \leq x_n\} = F(x_1) \cdots F(x_n), \tag{2.93}$$

where $F(\cdot)$ is the cumulative distribution function of the identically distributed random variables X_1,X_2,... .

The cumulative process obtained from an i.i.d. noise is called *random walk*:

Definition 2.11.2 *If X_i is an i.i.d. noise then a random walk S_n is defined as:*

$$S_0 \equiv 0, \tag{2.94}$$

$$S_n \equiv \sum_{i=1}^{n} X_i, \quad \forall n \geq 1. \tag{2.95}$$

2.11.2 Stationarity, Autocovariance and Autocorrelation

The idea of absence of trends or seasonality in a time series is expressed by the definition of stationarity. Roughly speaking a time series is said to be stationary if it has the same statistical properties of its time-shifted version:

Definition 2.11.3 *Let $\{X_t\}$ be a time series with $\mathbb{E}\left[X_t^2\right] < \infty$. We say that X_t is stationary if $\mathbb{E}[X_t]$ is independent of t and if the covariance:*

$$cov(X_{t+h}, X_t) \equiv \mathbb{E}[(X_{t+h} - \mu)(X_t - \mu)], \tag{2.96}$$

where $\mu = \mathbb{E}[X_t]$, is independent of t for each h.

When X_t is stationary we can define the *autocovariance function* $\gamma_X(h)$ as:

$$\gamma_X(h) \equiv \mathbb{E}[(X_{t+h} - \mu)(X_t - \mu)], \tag{2.97}$$

and the *autocorrelation function* $\rho_X(h)$ as:

$$\rho_X(h) \equiv \frac{\gamma_X(h)}{\gamma_X(0)}. \tag{2.98}$$

Obviously if X_i is an i.i.d. noise we have:

$$X_i \quad i.i.d. \Rightarrow \rho_X(h) = \begin{cases} 1 & \text{if } h = 0 \\ 0 & \text{otherwise} \end{cases}$$

When dealing with sampled data $\{x_1, ..., x_n\}$ we are not able to compute the autocorrelation (2.98), however we can define the analogous versions of (2.97) and (2.98) for the sample under analysis:

Definition 2.11.4 *Let $\{x_1, ..., x_n\}$ be observations of a time series. Its sample mean is defined as:*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{2.99}$$

The sample autocovariance function is defined as:

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{i=1}^{n-|h|} \left(x_{i+|h|} - \bar{x}\right) (x_i - \bar{x}), \tag{2.100}$$

while the sample autocorrelation function is defined as:

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}. \tag{2.101}$$

When x_t is an i.i.d. noise with finite variance it can be shown that the sample autocorrelation coefficient $\hat{\rho}(h)$ are approximately $N\left(0, \frac{1}{n}\right)$:

$$(x_i)_{i=1,\dots,n} \text{ i.i.d.} \Rightarrow \hat{\rho}(h) \sim N\left(0, \frac{1}{n}\right). \tag{2.102}$$

This results allows the caluclation of confidence bands for testing the null of uncorrelated time series. For example, given a data sample x_i and the corresponding sample autocorrelation coefficients $\hat{\rho}(h)$, we **reject** at 95% confidence level the hypotesis that x_i are uncorrelated at lag h if $|\hat{\rho}(h)| \geq \frac{1.96}{\sqrt{n}}$, where n is the sample length.

2.11.3 Long and Short Range Dependence

From now on we suppose that the autocorrelation function (2.98) is integrable. In this case we can have two cases:

$$\int_0^\infty \rho(h)\, dh = \begin{cases} < \infty \\ = \infty \end{cases}$$

In the first case we can identify a characteristic time $h_c = \int_0^\infty \rho(h)\, dh$ and we can model the autocorrelation function as: $\rho(h) \sim e^{-h/h_c}$, i.e. the process drastically loses memory of itself after a time h_c. This types of processes are called *weakly autocorrelated*.

In the second case we cannot identify a characteristic time. This mean that the memory of the process is persistent. We will refer this type of processes as *strongly autocorrelated* or *long-memory* processes. Examples of long-memory processes are those displaying a power-law autocorrelation function: $\rho(h) \sim \frac{1}{h^\alpha}$, with $\alpha \in (0, 1)$.

Chapter 3

Electricity Market: A Non-parametric Approach.

This chapter is inspired from the paper of Pirino and Renò (2008).

3.1 Introduction

In this chapter, we analyze the price dynamics of electricity in four European markets: European Energy Exchange (Germany), Power Next (France), Amsterdam Power Exchange (The Netherlands) and Operador del Mercado de Electricidad (Spain), as well as in four American markets: California (SP 15), Texas (ERCOT), Ohio (Cinergy) and Pennsylvania (PJM). Each of these markets follows peculiar regulatory dynamics which regulate the trading process. They differ for total capacity and mechanism of price determination, even if they are pretty similar in spirit.

Electricity is a special commodity: it cannot be stored, thus its price dynamics is quite different from those of the other commodities. Arbitrage is limited (Eyeland and Geman, 1998) and more than one factor is needed to price forward contracts of different maturities (Lucia and Schwartz, 2002; Longstaff and Wang, 2004). Electricity prices are quite volatile (Duffie et al., 1998; Simonsen, 2005): the demand and the supply of energy can be influenced by shortages in the distribution network (Mount et al., 2006) and electricity price exhibits very pronounced daily changes: returns of more than +100% are not rare. Moreover, the price of energy is under the direct influence of weather conditions, thus electricity prices display pronounced seasonality (Wilkinson and Winsen, 2002; Knittel and Roberts, 2005).

Many efforts have been spent to model the electricity price dynamics. Lucheroni (2007) simulates typical stylized facts of electricity prices using FitzHugh-Nagumo excitable dynamics (see also De Sanctis and Mari (2007)). Guercia et al. (2005) adopt an agent-based description. Exploiting real demand data, the main features of the electricity markets arise from the interaction among agents.

In this chapter, we propose an alternative approach with respect to the existing models in the literature. All the models employed in the literature are parametric, in the sense that their dynamics depends on real parameters which have to be estimated. The set of possible parametric model is clearly very large; however,

once the model has been specified, it imposes a constraint on the dynamics. We provide an estimate of the dynamics of electricity prices which is independent, as far as possible, from modelling assumptions. The advantage of this approach is clearly its flexibility, since we do not need any parametric structure. To this purpose, we adopt nonparametric estimation techniques. By means of our approach, we show very clearly that it is convenient to separate the normal operational status from spike dynamics, and to this purpose we borrow from very recent advancements in financial econometrics.

We proceed as follows. Firstly, we separate the spike component from the normal operational process, since these two are driven by an arguably different dynamics (Huisman and Mahieu, 2003; Guirguis and Felder, 2004). Therefore, we describe the dynamics of logarithmic returns[1] of the electricity prices as a sum of a continuous part, which is driven by a drift term and a diffusion term with stochastic volatility, and a discontinuous part to model its abrupt changes (spikes) driven by Poisson jumps[2] Our filtering technique is based on non-parametric threshold estimators developed by Mancini (2007). It identifies as spike those returns which are larger than an estimated threshold. We propose an iterative estimate of the threshold. We then estimate the drift, the diffusion and the jump intensity function of logarithmic returns using Nadaraya-Watson estimators (Hardle, 1990), which are close, in spirit, to those proposed by Florens-Zmirou (1993); Bandi and Phillips (2003); Mancini and Renò (2008), and which have been typically employed in financial applications (Stanton, 1997; Jiang and Knight, 1997; Jiang, 1998).

This chapter is organized as follows: in Section 3.2 we discuss the filtering procedure and the technique used for the estimation of the threshold. In Section 3.3 we describe the Monte Carlo simulations used to compute confidence bands of the proposed estimators. Finally, in Section 3.4 we discuss our results and we report our conclusions.

3.2 Modelling and Estimation

In this Section, we describe the proposed model and our estimation technique, and we estimate the model on electricity spot prices. We start with a description of the data.

3.2.1 Data Description

Here we focus on the empirical description of the data at our disposal. Table 3.1 reports the time span used (we exclude Saturdays, Sundays and zero prices), and sample kurtosis and standard deviation of logarithmic returns for each analyzed market. Each market is characterized by a level of kurtosis much higher than that of a Normal diffusion, suggesting the presence of heteroskedasticity and spikes. The four European markets display a standard deviation which is very similar, with the Dutch market displaying the highest level of volatility. American markets are less volatile. The time series of the spot price of European and American markets are shown in Figure 3.1.

[1] In modelling electricity spot returns instead of prices we borrow from finance, where this approach is customary. Such an approach has risen some criticism recently Guthrie and Videbeck (2007), we manage to obtain an accurate description of the *daily* data.

[2] This separation is common in the electricity dynamics literature. For example, Geman and Roncoroni (2006), explicitly disentangle continuous from discontinuous variation (using a constant threshold) to estimate a fully para- metric model on the normal status.

	Sample			return statistics	
	start date	end date	data	st. dev.	kurtosis
Germany	01/01/01	25/07/05	1191	0.329	29.9
France	27/11/01	30/07/05	960	0.300	12.6
Spain	01/01/01	26/05/05	1205	0.201	10.5
The Netherlands	01/01/01	10/08/05	1203	0.406	11.6
California	02/01/02	12/02/07	1273	0.095	52.9
Texas	02/01/02	19/03/07	1244	0.096	27.4
Ohio	03/01/01	11/05/07	1542	0.168	4.9
Pennsylvania	02/01/01	09/05/07	1581	0.150	12.4

Table 3.1: Reports data sample periods expressed as gg/mm/yy together with kurtosis and standard deviations of the sample logarithmic returns. Saturdays and Sundays are excluded.

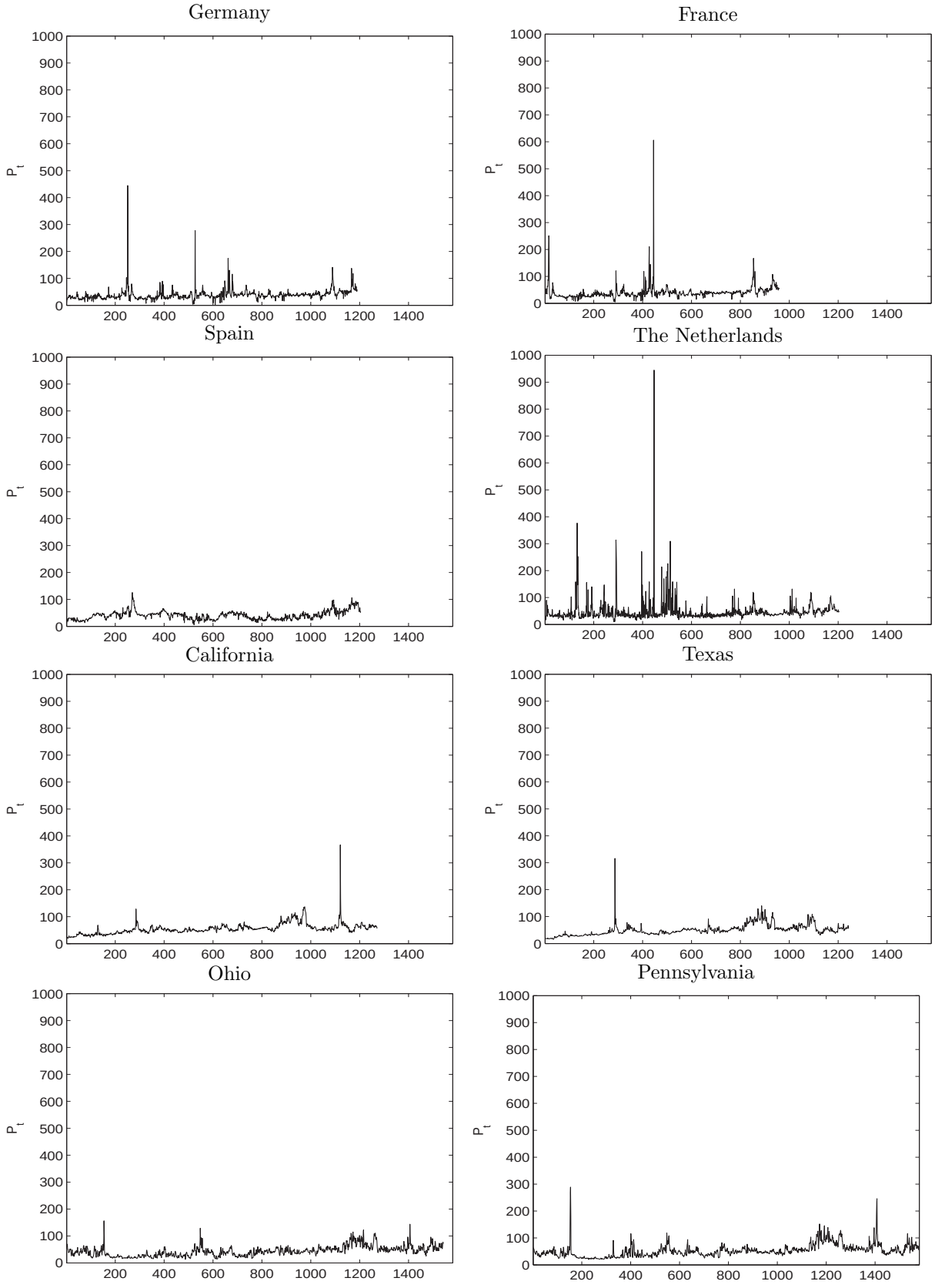

Figure 3.1: The time series of spot electricity prices of the European and American markets.

3.2.2 The Model

Denote by P_t the price of electricity on day t and by $r_t = \log P_t - \log P_{t-1}$ the logarithmic return. Our model for electricity prices is specified by:

$$P_t = P_t^c + J_t, \tag{3.1}$$

that is we separate the *normal status price*, which we denote by P_t^c, from *spikes* which we denote by J_t. We assume that the normal status is driven by the following equation:

$$r_{t+1}^c = \mu\left(r_t^c\right) + \sigma\left(r_t^c\right)\varepsilon_t \tag{3.2}$$

where $r_t^c = \log P_t^c - \log P_{t-1}^c$ and ε_t are IID standard Normal shocks. The dependence of today's continuous returns from yesterday's ones is then modelled via the functions $\mu(\cdot)$ and $\sigma(\cdot)$.

The spike component J_t is associated with a compound Poisson process with intensity λ_t, which we allow to depend on time in a periodic fashion, to allow for spike seasonality[3]. When the associated Poisson process jumps at time t, J_t is a spike drawn by a parametric distribution with positive support; $J_t = 0$ otherwise.

We do not specify the functions $\mu(\cdot), \sigma(\cdot)$ and λ_t, as it is common in the parametric literature. Instead, we estimate these function on data themselves, providing more flexibility to the model. The only parametric assumption we make is on the distribution of the size of spikes.

Thus, our aim is to get an estimate of the functions $\mu(\cdot)$ and $\sigma(\cdot)$ in (3.2), as well as an estimate of the intensity function λ_t and the statistical features of the spike component J_t, by using N discrete observations of the variable r_t in a the time interval $[0, T]$. In order to obtain a reliable estimate, we first disentangle the spike component of the data, which is represented by J_t, by the normal status variations.

3.2.3 The Threshold Theory

How can discontinuous variations be separated from continuous ones? Mancini (2004) proposes an elegant theory to answer this question. The main idea is based on Paul Lévy's law of modulus of continuity of Brownian motion. When W_t is a Brownian motion it can be shown that:

$$\text{a.s.} \quad \lim_{\delta \to 0} \sup_{\substack{0 \le s_1 \le s_2 \le T \\ |s_2 - s_1| = \delta}} \frac{|W_{s_1} - W_{s_2}|}{\sqrt{2\,\delta \log \frac{1}{\delta}}} = 1.$$

Equation (3.2.3) states that $\sqrt{2\,\delta \log \frac{1}{\delta}}$ is a measure of the speed at which the increment of the Brownian motion goes to zero while the time interval δ shrinks to zero. Given a stochastic process X_t we wonder whether the variation $\Delta X_{t,\delta} = X_{t+\delta} - X_t$ in the time window $[t, t+\delta]$ is purely Brownian or not. When a jump occurs between t and $t+\delta$ the quantity $\Delta X_{t,\delta}$ is discontinuous and thus not Brownian.

Intuitively if we choose a function h_δ of the time interval δ which goes to zero slower than the modulus of continuity of the Brownian motion, we can *asymptotically* label as "discontinuous" all variations $\Delta X_{t,\delta}$ which are greater than the chosen threshold h_δ.

[3]Also the functions $\mu(\cdot)$ and $\sigma(\cdot)$ may easily incorporate a seasonal component. However, we find seasonality to be much more important for the jump intensity than for drift and volatility. Also λ_t may depend on the state variable r_t^c, but this would be more difficult to estimate given the scarcity of spikes. Finally, different seasonalities (monthly, weekly) can be accounted for in the same fashion.

With *asymptotic* identification we mean that the smaller the value of δ the better the estimate of the discontinuous component.

In order to put the presented ideas in a mathematical form we first introduce the formal context which will be the basis of all our results.

We work in a filtered probability space $(\Omega, (\mathcal{F}_t)_{t\in[0,T]}, \mathcal{F}, \mathcal{P})$, satisfying the usual conditions (Protter, 1990). Our assumptions about the jump-diffusion model are very mild.

Assumption 3.2.1 *$(X_t)_{t\in[0,T]}$ is a real-valued process such that $X_0 \in \mathbb{R}$ and*

$$dX_t = a_t dt + \sigma_t dW_t + dJ_t \tag{3.3}$$

where a_t is predictable, σ_t is cádlág and J_t is a finite activity jump process, that is we can write

$$J_t = \sum_{j=1}^{N_t} c_j \tag{3.4}$$

where N_t is a Poisson process whose intensity is an adapted stochastic process λ_t, the times of the corresponding jumps are $(\tau_j)_{j=1,\dots,N_T}$ and c_j are adapted random variables measuring the size of the jump at time τ_j.

Here we will not deal with the infinite activity case, see Sato (2001) for reference. Consider now an interval $[0, T]$ with T fixed. Consider a real number $\delta = T/n$. We define the evenly sampled returns as:

$$\Delta_j X = X_{j\delta} - X_{(j-1)\delta}, \qquad j = 1, \dots, n \tag{3.5}$$

This definition is used a little bit "liberally" when δ is generic; in this case $n = T/\delta$ is not guaranteed to be integer; when this happens, T is thought to be re-defined as equal to δn. During all the dissertation we use previous-tick interpolation, that is if the interpolated time $\tau = n\,\delta$ (for same integer n) falls between the observed times t_{k-1} and t_k we take $X(t_{k-1})$ as the interpolated log-price. In formula:

$$\tau \in [t_{k-1}, t_k) \to X(\tau) \equiv X(t_{k-1}) \tag{3.6}$$

Linear interpolation is not suitable for the construction of intraday returns at high frequency as discussed in Dacorogona et al. (2001). However this is not the case for electricity prices which are available only on a daily basis. However we put the discussion in a more general form which will be useful in the next chapter where high-frequency data are exploited.

The following definition introduces the concept of threshold functions.

Definition 3.2.2 *A threshold is a deterministic function h_δ such that:*

$$\lim_{\delta\to 0} h_\delta = 0, \qquad \lim_{\delta\to 0} \frac{\sqrt{\delta \log \frac{1}{\delta}}}{h_\delta} = 0. \tag{3.7}$$

We will use the notation $\vartheta_\delta = h_\delta^2$.

The limit (3.7) formalizes the idea that, when $\delta \to 0$, threshold functions shrink to zero slower than the Brownian motions.

The ideas presented heuristically above can be now stated in the following theorem due to Mancini (2004):

Remark 3.2.3 (Mancini, 2004) *Suppose that:*

1. $\exists C(\omega)$ *such that:*

$$a.s. \quad \limsup_{\delta \to 0} \frac{\sup_i \left| \int_{t_{i-1}}^{t_i} a_s \, ds \right|}{\sqrt{\delta \, \log \frac{1}{\delta}}} \leq C(\omega) < \infty. \tag{3.8}$$

2. σ *is such that* $\int_0^T \sigma_s^2 \, ds < \infty$ *and* $\exists M(\omega)$ *such that:*

$$a.s. \quad \limsup_{\delta \to 0} \frac{\sup_i \left| \int_{t_{i-1}}^{t_i} \sigma_s^2 \, ds \right|}{\delta} \leq M(\omega) < \infty. \tag{3.9}$$

3. h_δ *is a threshold function and* $\vartheta_\delta = h_\delta^2$.

then for all P-almost $\omega \in \Omega$ $\exists \bar{\delta}(\omega)$ *such that* $\forall \delta \leq \bar{\delta}(\omega)$ *we have*

$$\forall i = 1, ..., n, \qquad I_{\{\Delta_i N = 0\}}(\omega) = I_{\left\{(\Delta_i X)^2 \leq \vartheta_\delta\right\}}(\omega). \tag{3.10}$$

Theorem (3.2.3) is a fundamental brick of the presented work. For this reason we report the full proof in what follows.

Proof. We first clarify the notation. Having chosen $\omega \in \Omega$ and $i \in \{1, ..., n\}$ we have that:

$$I_{\{\Delta_i N = 0\}}(\omega) \equiv \begin{cases} 1 & \text{if } \Delta_i N(\omega) = 0, \\ 0 & \text{otherwise,} \end{cases}$$

and similarly for $I_{\left\{(\Delta_i X)^2 \leq \vartheta_\delta\right\}}$. Equation (3.10) states that, asymptotically, the set of interval without jumps ($\Delta_i N(\omega) = 0$) will approach the set of intervals of under-threshold increments ($\Delta_i X^2(\omega) \leq \vartheta_\delta$).

As a consequence of (3.2.3) we have that:

$$\text{a.s.} \quad \lim_{\delta \to 0} \sup_{\substack{0 \leq s_1 \leq T \\ 0 < s_2 - s_1 < \delta}} \frac{|W_{s_1} - W_{s_2}|}{\sqrt{2 \, (s_2 - s_1) \, \log \frac{1}{s_2 - s_1}}} = 1, \tag{3.11}$$

and thus:

$$\text{a.s.} \quad \lim_{\delta \to 0} \sup_{i \in \{1, ..., n\}} \frac{|\Delta_i W|}{\sqrt{2 \, \delta \, \log \frac{1}{\delta}}} \leq 1, \tag{3.12}$$

the supremum (3.12) being computed on a smaller set than supremum (3.11). It can be shown that (see Ikeda and Watanabe, 1981), after a proper extension of σ_t on $]T, \infty]$, the stochastic integral $\sigma \cdot W = \int_0^t \sigma_s dW_s$ is in turn a time changed Brownian motion, i.e. defining $Y_t \equiv \int_0^t \sigma_s^2 ds$ and $\xi_t \equiv \inf \{v : Y_v > t\}$, we have that the stochastic process $B_t = \int_0^{\xi_t} \sigma_s dW_s$ is a Brownian motion and $\Delta_i (\sigma \cdot W) = B_{Y_{t_i}} - B_{Y_{t_{i-1}}}$. Therefore:

$$\text{a.s.} \quad \sup_{i \in \{1, ..., n\}} \frac{|\sigma \cdot W|}{\sqrt{2 \, \delta \, \log \frac{1}{\delta}}} = \sup_i \frac{\left| B_{Y_{t_i}} - B_{Y_{t_{i-1}}} \right|}{\sqrt{2 \, \delta \, \log \frac{1}{\delta}}} \quad = \tag{3.13}$$

$$\sup_i \left(\frac{\left| B_{Y_{t_i}} - B_{Y_{t_{i-1}}} \right|}{\sqrt{2 \, \Delta_i Y \, \log \frac{1}{\Delta_i Y}}} \, \frac{\sqrt{2 \, \Delta_i Y \, \log \frac{1}{\Delta_i Y}}}{\sqrt{2 \, M \, \delta \, \log \frac{1}{M \, \delta}}} \, \frac{\sqrt{2 \, M \, \delta \, \log \frac{1}{M \, \delta}}}{\sqrt{2 \, \delta \, \log \frac{1}{\delta}}} \right) \leq \tag{3.14}$$

$$\sup_i \frac{\left| B_{Y_{t_i}} - B_{Y_{t_{i-1}}} \right|}{\sqrt{2 \, \Delta_i Y \, \log \frac{1}{\Delta_i Y}}} \, \sup_i \frac{\sqrt{2 \, \Delta_i Y \, \log \frac{1}{\Delta_i Y}}}{\sqrt{2 \, M \, \delta \, \log \frac{1}{M \, \delta}}} \, \sup_i \frac{\sqrt{2 \, M \, \delta \, \log \frac{1}{M \, \delta}}}{\sqrt{2 \, \delta \, \log \frac{1}{\delta}}}. \tag{3.15}$$

We now analyze each terms appearing in expression (3.15):

- The first term is bounded by 1 because second assumption (see expression 3.9) guarantees that $\Delta_i Y = \int_{t_{i-1}}^{t_i} \sigma_s^2 ds = O_p(\delta)$ and thus using property (3.12) we have that:

$$\text{a.s.} \quad \lim_{\delta \to 0} \sup_i \frac{\left| B_{Y_{t_i}} - B_{Y_{t_{i-1}}} \right|}{\sqrt{2\,\Delta_i Y\,\log \frac{1}{\Delta_i Y}}} \leq 1. \tag{3.16}$$

- As $\lim_{x \downarrow 0} x \log \frac{1}{x} = 0$ and $\Delta_i Y = O_p(\delta)$ we have that:

$$\text{a.s.} \quad \lim_{\delta \to 0} \sup_i \frac{\sqrt{2\,\Delta_i Y\,\log \frac{1}{\Delta_i Y}}}{\sqrt{2\,M\,\delta\,\log \frac{1}{M\,\delta}}} \leq 1. \tag{3.17}$$

- The third term converges to $\sqrt{M(\omega)}$.

Therefore the quantity $\sup_{i \in \{1,...,n\}} \frac{|\sigma \cdot W|}{\sqrt{2\,\delta\,\log \frac{1}{\delta}}}$ is a.s. bounded. In formula, for small δ, we can write:

$$\text{a.s.} \quad \sup_{i \in \{1,...,n\}} \frac{|\sigma \cdot W|}{\sqrt{2\,\delta\,\log \frac{1}{\delta}}} \leq Z(\omega) < \infty. \tag{3.18}$$

As a consequence, for small δ, we have:

$$\sup_i \frac{\left| \int_{t_{i-1}}^{t_i} a_s ds + \int_{t_{i-1}}^{t_i} \sigma_s dW_s \right|}{\sqrt{2\,\delta\,\log \frac{1}{\delta}}} \leq \sup_i \frac{\left| \int_{t_{i-1}}^{t_i} a_s ds \right|}{\sqrt{2\,\delta\,\log \frac{1}{\delta}}} + \sup_i \frac{\left| \int_{t_{i-1}}^{t_i} \sigma_s dW_s \right|}{\sqrt{2\,\delta\,\log \frac{1}{\delta}}} \leq C(\omega) + Z(\omega). \tag{3.19}$$

Where expression (3.8) of assumption 1) has been used.

In order to proof the theorem we need to demonstrate that, for small δ, we have:

$$\forall i \quad I_{\{\Delta_i N = 0\}}(\omega) \leq I_{\left\{(\Delta_i X)^2 \leq \vartheta_\delta\right\}}(\omega), \tag{3.20}$$

and

$$\forall i \quad I_{\{\Delta_i N = 0\}}(\omega) \geq I_{\left\{(\Delta_i X)^2 \leq \vartheta_\delta\right\}}(\omega). \tag{3.21}$$

Proof of inequality (3.20). For all $\omega \in \Omega$ let $J_{0,\delta}$ be the set of indexes $i \in \{1, ..., n\}$ of the intervals **without** jumps: $J_{0,\delta} \equiv \{i \in \{1, ..., n\} : \Delta_i N = 0\}$. If we demonstrate that:

$$\sup_{i \in J_{0,\delta}} (\Delta_i X)^2 \leq \vartheta_\delta \tag{3.22}$$

then for all i such that $I_{\{\Delta_i N = 0\}} = 1$ we would have: $I_{\left\{(\Delta_i X)^2 \leq \vartheta_\delta\right\}} = 1$, which demonstrates inequality (3.20). In order to verify inequality (3.22) we compute:

$$\sup_{i \in J_{0,\delta}} \frac{(\Delta_i X)^2}{\vartheta_\delta} = \sup_{i \in J_{0,\delta}} \left(\frac{\left| \int_{t_{i-1}}^{t_i} a_s ds + \int_{t_{i-1}}^{t_i} \sigma_s dW_s \right|}{\sqrt{2\,\delta\,\log \frac{1}{\delta}}} \right)^2 \frac{2\,\delta\,\log \frac{1}{\delta}}{\vartheta_\delta}. \tag{3.23}$$

The first factor is a.s. bounded as demonstrated in (3.19). The second factor approaches zero as stated by definition (3.7). Then, for small δ, $\sup_{i \in J_{0,\delta}} \frac{(\Delta_i X)^2}{\vartheta_\delta}$ is dominated by 1:

$$\sup_{i \in J_{0,\delta}} (\Delta_i X)^2 \leq \vartheta_\delta. \tag{3.24}$$

*Proof of inequality (3.21).*We now define the analogous of $J_{0,\delta}$ for the interval **with** jumps: $J_{1,\delta} \equiv \{i \in \{1, ..., n\} : \Delta_i N \neq 0\}$. If we demonstrate that:

$$\inf_{i \in J_{1,\delta}} (\Delta_i X)^2 > \vartheta_\delta, \tag{3.25}$$

then for all i such that $I_{\{\Delta_i N \neq 0\}} = 1$ we would have: $I_{\left\{(\Delta_i X)^2 > \vartheta_\delta\right\}} = 1$, which means that:

$$I_{\{\Delta_i N \neq 0\}} \leq I_{\left\{(\Delta_i X)^2 > \vartheta_\delta\right\}}, \tag{3.26}$$

as well as:

$$I_{\{\Delta_i N=0\}} = 1 - I_{\{\Delta_i N\neq 0\}} \geq 1 - I_{\left\{(\Delta_i X)^2 \geq \vartheta_\delta\right\}} = I_{\left\{(\Delta_i X)^2 \leq \vartheta_\delta\right\}}, \tag{3.27}$$

which demonstrates inequality (3.21). In order to evaluate $\inf_{J_{1,\delta}} (\Delta_i X)^2$ we compute:

$$\forall i \in J_{1,\delta} \quad \frac{(\Delta_i X)^2}{\vartheta_\delta} = \frac{\left(\int_{t_{i-1}}^{t_i} a_s ds + \Delta_i \sigma \cdot W\right)^2}{2\,\delta\,\log\frac{1}{\delta}} \frac{2\,\delta\,\log\frac{1}{\delta}}{\vartheta_\delta} \quad + \tag{3.28}$$

$$2\,\frac{\int_{t_{i-1}}^{t_i} a_s ds + \Delta_i \sigma \cdot W}{\sqrt{\vartheta_\delta}} \frac{\sum_{l=1}^{\Delta_i N} c_l}{\sqrt{\vartheta_\delta}} + \frac{\left(\sum_{l=1}^{\Delta_i N} c_l\right)^2}{\vartheta_\delta}. \tag{3.29}$$

The first term converges a.s. to zero uniformly w.r.t. i. For small δ we have that $\Delta_i N \leq 1$, therefore the other terms become:

$$\frac{c_{\tau_i}}{\sqrt{\vartheta_\delta}} \left[\frac{\int_{t_{i-1}}^{t_i} a_s ds + \Delta_i \sigma \cdot W}{\sqrt{\vartheta_\delta}} + \frac{c_{\tau_i}}{\sqrt{\vartheta_\delta}}\right]. \tag{3.30}$$

Again the first term in brackets converges a.s. to zero uniformly w.r.t. i. Therefore:

$$\text{a.s.} \lim_{\delta\to 0} \inf_{i\in J_{1,\delta}} \frac{(\Delta_i X)^2}{\vartheta_\delta} = \lim_{\delta\to 0} \inf_{i\in J_{1,\delta}} \frac{c_{\tau_i}^2}{\vartheta_\delta} \geq \lim_{\delta\to 0} \frac{(\min_i c_i)^2}{\vartheta_\delta} = +\infty, \tag{3.31}$$

having used $\lim_{\delta\to 0} \vartheta_\delta = 0$. As a consequence, for small δ, we obtain inequality (3.25). □

In Theorem (3.2.3) the threshold ϑ_δ is assumed to be a constant function. As remarked in Mancini and Renò (2008) Theorem (3.2.3) holds even if ϑ_δ is replaced by $\tilde{\vartheta}_{t,\delta} \equiv c_t\,\vartheta_\delta$, where c_t is a bounded deterministic function which is also bounded away from zero.

In finite sample it is very useful to have a time varying threshold. This gives the possibility to adapt the threshold function to the local volatility of the time series.

When volatility is high we can observe large price variations even in the normal status. A constant threshold could not distinguish between a "real" spike and a continuous variation generated when volatility is very high. In what follows we will show that it is possible to choice a suitable threshold which is indeed an estimate of the local volatility.

3.2.4 Non-parametric Estimators

The model of electricity price returns (3.1)-(3.2) is the discretized version of (3.3) where $a_s = \mu(r_s^c)$ and $\sigma_s = \sigma(r_s^c)$. Florens-Zmirou (1993) proposes non-parametric estimation of $\mu(x)$, $\sigma(x)$ starting from the notion of local time. Let:

$$L_t(x) = \lim_{b\to 0} \frac{1}{2\,b} \int_0^T I_{\{|X_s - x|<b\}}\, ds \tag{3.32}$$

be the local time in x during $[0,T]$ of a process X_s. $L_t(x)$ estimates how much X_s visits x during $[0,T]$. Following Florens-Zmirou (1993) the local time (3.32) can be estimated by:

$$L_t^n(x) = \frac{1}{2\,n\,b_n} \sum_i^{\lfloor n\,t\rfloor} I_{\left\{|X_{(i/n)} - x|<b_n\right\}}, \tag{3.33}$$

when $n\to\infty$, $b_n\to 0$ and $n\,b_n\to\infty$.

The idea of non-parametric estimation is based on the fact that, if:

$$dX_s = \mu(X_s)\,dt + \sigma(X_s)\,dW_s, \tag{3.34}$$

then using multiplication rules (2.85)-(2.86)-(2.87) we get:

$$\mathbb{E}\left[dX_s|X_s=x\right]=\mu\left(x\right)\,dt\Rightarrow\mu\left(x\right)=\frac{1}{dt}\mathbb{E}\left[dX_s|X_s=x\right], \tag{3.35}$$

and:

$$\mathbb{E}\left[\left(dX_s\right)^2|X_s=x\right]=\sigma^2\left(x\right)\,dt\Rightarrow\sigma^2\left(x\right)=\frac{1}{dt}\mathbb{E}\left[\left(dX_s\right)^2|X_s=x\right]. \tag{3.36}$$

The time interval dt is given by $dt=\frac{T}{n}$, where T is the time horizon and n the number of observation:

$$\mu\left(x\right)=\frac{n}{T}\mathbb{E}\left[dX_s|X_s=x\right],\quad\sigma^2\left(x\right)=\frac{n}{T}\mathbb{E}\left[\left(dX_s\right)^2|X_s=x\right]. \tag{3.37}$$

Following the result (3.33) for the local time, as an heuristic argument, the expected values $\mathbb{E}\left[dX_s|X_s=x\right]$ and $\mathbb{E}\left[\left(dX_s\right)^2|X_s=x\right]$ can be approximated by the quantities:

$$\mathbb{E}\left[dX_s|X_s=x\right]\approx\frac{\sum_{i=1}^{n-1}I_{\left\{\left|X_{(i/n)}-x\right|<b_n\right\}}\left(X_{(i+1)/n}-X_{i/n}\right)}{\sum_{i=1}^{n}I_{\left\{\left|X_{(i/n)}-x\right|<b_n\right\}}}, \tag{3.38}$$

$$\mathbb{E}\left[\left(dX_s\right)^2|X_s=x\right]\approx\frac{\sum_{i=1}^{n-1}I_{\left\{\left|X_{(i/n)}-x\right|<b_n\right\}}\left(X_{(i+1)/n}-X_{i/n}\right)^2}{\sum_{i=1}^{n}I_{\left\{\left|X_{(i/n)}-x\right|<b_n\right\}}}. \tag{3.39}$$

It is intuitive that we can replace the indicator function $I_{\left\{\left|X_{(i/n)}-x\right|<b_n\right\}}$ with a normalized symmetric kernel function $K\left(\frac{X_{i/n}-x}{b_n}\right)$ without changing the content of (3.38)-(3.39). Therefore our candidate estimators are:

$$\hat{\mu}\left(x\right)=\frac{n\sum_{i=1}^{n-1}K\left(\frac{X_{i/n}-x}{b_n}\right)\left(X_{(i+1)/n}-X_{i/n}\right)}{T\sum_{i=1}^{n}K\left(\frac{X_{i/n}-x}{b_n}\right)}, \tag{3.40}$$

$$\hat{\sigma}^2\left(x\right)=\frac{n\sum_{i=1}^{n-1}K\left(\frac{X_{i/n}-x}{b_n}\right)\left(X_{(i+1)/n}-X_{i/n}\right)^2}{T\sum_{i=1}^{n}K\left(\frac{X_{i/n}-x}{b_n}\right)}. \tag{3.41}$$

In fact it can be shown that (Florens-Zmirou, 1993):

$$\hat{\mu}\left(x\right)\xrightarrow{P}\mu\left(x\right),\ \hat{\sigma}^2\left(x\right)\xrightarrow{P}\sigma^2\left(x\right)\ \forall x\text{ visited by }\left(X_s\right)_{s\in[0,T]}. \tag{3.42}$$

when $\left(X_s\right)_{s\in[0,1]}$ is driven by a continuous diffusion like (3.34) and under very mild assumption on $K\left(\cdot\right)$.

However estimators (3.40)-(3.41) are not suitable for our purposes. Model (3.3) is discontinuous and expressions (3.40)-(3.41) must be modified in order to suppress the jump component. Using the result of theorem (3.2.3) we guess as candidates for the unbiased estimators of $\mu\left(x\right)$ and $\sigma\left(x\right)$the quantity:

$$\hat{\mu}\left(x\right)=\frac{n\sum_{i=1}^{n-1}K\left(\frac{X_{i/n}-x}{b_n}\right)\left(X_{(i+1)/n}-X_{i/n}\right)I_{\left\{\left(X_{(i+1)/n}-X_{i/n}\right)^2\leq\vartheta_{i,\delta}\right\}}}{T\sum_{i=1}^{n}K\left(\frac{X_{i/n}-x}{b_n}\right)}, \tag{3.43}$$

$$\hat{\sigma}^2\left(x\right)=\frac{n\sum_{i=1}^{n-1}K\left(\frac{X_{i/n}-x}{b_n}\right)\left(X_{(i+1)/n}-X_{i/n}\right)^2I_{\left\{\left(X_{(i+1)/n}-X_{i/n}\right)^2\leq\vartheta_{i,\delta}\right\}}}{T\sum_{i=1}^{n}K\left(\frac{X_{i/n}-x}{b_n}\right)}, \tag{3.44}$$

having used the substitution $dt = \delta$ to put relations in an more homogeneous form. The function $\vartheta_{i,\delta}$ is supposed to be a suitable time-varying threshold function. As expected it happens that estimators (3.43)-(3.43) are unbiased in presence of jumps, as stated by the following theorem due to Mancini and Renò (2008):

Theorem 3.2.4 (Mancini and Renò, 2008) *Suppose that:*

1. $dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t + d\left(\sum_{k=1}^{N_t} c_k\right)$, *with* $P\{c_k = 0\} = 0$ *for each* k *and* $J_t = \sum_{k=1}^{N_t} c_k$ *is a finite activity doubly stochastic compound Poisson process with level-dependent intensity* $\lambda(X_{t-})$*;*

2. $\vartheta(\delta)$ *is a threshold function.;*

3. $n\,h^4 \to 0$ *as* $n \to \infty$ *and* $\exists \beta > 1$ *such that* $n\,h^\beta \to \infty$*;*

4. $\sigma(x)$ *is bounded on its domain and* $\mathbb{E}\left[\int_0^\infty \sigma(X_s)\,ds\right] < \infty$*;*

5. $\lambda(x)$ *is bounded on its domain.*

Define:

$$\hat{\mu}_n(x) = \frac{n\sum_{i=1}^{n} K\left(\frac{X_{t_{i-1}} - J_{t_{i-1}} - x}{h}\right) \Delta_i X \cdot I_{\left\{(\Delta_i X)^2 \le \vartheta\left(\frac{T}{n}\right)\right\}}}{T\sum_{i=1}^{n} K\left(\frac{X_{t_{i-1}} - J_{t_{i-1}} - x}{h}\right)}, \tag{3.45}$$

and

$$\hat{\sigma}_n(x) = \frac{n\sum_{i=1}^{n} K\left(\frac{X_{t_{i-1}} - J_{t_{i-1}} - x}{h}\right) (\Delta_i X)^2 \cdot I_{\left\{\Delta_i X^2 \le \vartheta\left(\frac{T}{n}\right)\right\}}}{T\sum_{i=1}^{n} K\left(\frac{X_{t_{i-1}} - J_{t_{i-1}} - x}{h}\right)}, \tag{3.46}$$

where $K(\cdot)$ *is a bounded kernel function differentiable in a neighborhood of zero with bounded derivative and such that* $\int_{\mathbb{R}} |K(u)|\,du < \infty$*. Then for all* x *visited by* X*:*

$$\hat{\mu}_n(x) \to_P \mu(x) \tag{3.47}$$

$$\hat{\sigma}_n(x) \to_P \sigma(x) \tag{3.48}$$

Even if our model is in discrete time, we can follow this technique closely.

We adopt the following threshold, which iterates on the integer Z:

$$\vartheta^Z(t) = 9\,\frac{\sum_{i=-L, i\neq -1,0,1}^{L} K\left(\frac{i}{L}\right) r_{t+i}^2\, I_{\{r_{t+i}^2 \le \vartheta^{Z-1}(t+i)\}}}{\sum_{i=-L, i\neq -1,0,1}^{L} K\left(\frac{i}{L}\right) I_{\{r_{t+i}^2 \le \vartheta^{Z-1}(t+i)\}}}, \tag{3.49}$$

The parameter c_V sets the number of standard deviations after which an observation is considered above threshold. Its value potentially determines the number of detected jumps; however, for electricity prices spikes are typically so large that the choice of this value is not influential. We set $c_V = 3$, corresponding to three standard deviations, as it is customary in these kinds of applications.

When r_t^c is conditionally Gaussian with mean μ_t and standard deviation σ_t we have that 99.7% of observations are within $\mu_t \pm 3\,\sigma_t$. Thus the probability value 99.7% can be interpreted as a confidence level for the identification of jumps with the threshold function (3.49).

Equation (3.49) is a nonparametric recursive smoothing filter for the variance (Fan and Yao, 2003), with the starting value $\theta^0(t) = +\infty$, which corresponds to no threshold in the first step. The condition that identifies a jump is the following:

$$r_t^2 \geq \theta_t^Z \tag{3.50}$$

Thus, in equation (3.49) the indicator function excludes returns which are detected as jumps in the previous iteration.

The iterations on Z are stopped as soon as $\vartheta^Z(t) \equiv \vartheta^{Z-1}(t)$. For simplicity we indicate with $\vartheta(t)$ the threshold obtained in the last iteration.

The bandwidth parameter L determines the time scale of returns included in the estimation of the threshold around point t. Its choice can be crucial: increasing the value of L yields a smoother threshold.

The choice of the kernel function $K(\cdot)$ is not influential in this kind of applications (Silverman, 1986; Wand and Jones, 1995). We use a Gaussian kernel.

$$K(y) = \left(1/\sqrt{2\,\pi}\right) \exp\left(-y^2/2\right). \tag{3.51}$$

When estimating the threshold at time t, we exclude not only the observation at time t, but also observations at times $t+1, t-1$. In electricity markets jumps are always consecutive and the occurrence of a jump at time t is always accompanied with another jump at time $t-1$ or $t+1$ and then excluding the nearest observations allows a better estimate of the local volatility .

As a whole, our filtering technique can be viewed as a refinement of the iterative technique proposed by (Clewlow and Strickland, 2000), where the proposed threshold is a constant. On the contrary we allow it to vary with time and we estimate it locally, using observations around the point of interest.

We report the results of the filtering procedure on the considered markets in Tables 3.2-3.3, which show, for each market, the number of jumps detected at each iteration for different values of the bandwidth L. Our results show that the procedure converges quickly. After few iterations the number of jumps detected becomes constant for approximately all values of L.

Figure 3.2 shows an example of threshold on the time series of the logarithmic returns of the German market. We compute the thresholds for seven values of the bandwidth parameter, i.e. $L = 15, 20, 25, ..., 45$, see Figure 3.2.

Figures 3.3-3.4 report the percentage efficiency (defined as the number of detected jumps among actually generated ones) and percentage noise (defined as the number of detected jumps among generated normal status returns) of the spike detection procedure. Efficiency and noise are computed via Monte Carlo simulations of estimated models on European and American data, see Section 3.3. We can introduce a signal to noise ratio, defined as the ratio between jump (interpreted here as signal) standard deviation and mean continuous volatility (which play the role of noise). The higher the ratio the higher the percentage detection efficiency and the lower the percentage noise. Inspection of figures 3.3-3.4 reveals that Germany and California have the largest signal to noise among European and American markets respectively. Besides, comparing the results for the American and European markets we observe that the procedure is less noisy and more efficient for the American markets rather than for their European counterparts.

Market	Z	L						
		15	20	25	30	35	40	45
Germany	1	38	36	34	32	30	25	26
	2	53	49	50	48	46	41	42
	3	60	55	56	55	52	49	49
	4	62	57	56	57	55	53	52
	5	62	57	56	57	57	56	53
	6	62	57	56	57	57	56	53
	7	62	57	56	57	57	56	53
	8	62	57	56	57	57	56	53
France	1	31	31	25	22	22	21	22
	2	39	41	34	28	29	30	34
	3	41	41	34	30	33	38	37
	4	41	41	34	33	35	40	38
	5	41	41	34	34	36	41	38
	6	41	41	34	34	37	43	38
	7	41	41	34	34	39	43	38
	8	41	41	34	34	39	43	38
Spain	1	28	28	25	23	23	24	24
	2	38	38	36	36	37	36	35
	3	41	42	41	38	40	39	40
	4	43	43	44	39	42	40	40
	5	43	43	45	39	42	40	40
	6	43	43	45	39	42	40	40
	7	43	43	45	39	42	40	40
	8	43	43	45	39	42	40	40
The Netherlands	1	28	21	19	20	22	22	23
	2	37	32	28	31	35	35	33
	3	39	36	32	33	40	42	41
	4	39	38	34	34	43	46	46
	5	39	38	34	35	45	49	48
	6	39	38	34	35	47	50	49
	7	39	38	34	35	48	50	49
	8	39	38	34	35	48	50	49

Table 3.2: Number of jumps detected at each iteration of the filtering procedure for each European market and for each value of L.

Market	Z	L						
		15	20	25	30	35	40	45
California	1	23	19	19	17	16	16	17
	2	31	28	32	26	26	26	29
	3	35	32	34	33	31	31	32
	4	35	34	34	34	32	32	32
	5	35	34	34	34	32	32	32
	6	35	34	34	34	32	32	32
	7	35	34	34	34	32	32	32
	8	35	34	34	34	32	32	32
Texas	1	26	21	20	19	21	22	22
	2	36	32	28	29	33	32	33
	3	39	37	33	34	36	36	35
	4	39	38	36	36	37	36	35
	5	39	38	36	36	37	36	35
	6	39	38	36	36	37	36	35
	7	39	38	36	36	37	36	35
	8	39	38	36	36	37	36	35
Ohio	1	16	20	16	16	17	16	13
	2	19	21	22	21	19	21	19
	3	20	21	23	21	20	21	21
	4	21	21	23	21	20	21	22
	5	21	21	23	21	20	21	22
	6	21	21	23	21	20	21	22
	7	21	21	23	21	20	21	22
	8	21	21	23	21	20	21	22
Pennsylvania	1	19	21	22	21	21	23	20
	2	23	26	27	26	30	30	29
	3	25	28	29	27	31	33	32
	4	25	28	29	27	31	33	32
	5	25	28	29	27	31	33	32
	6	25	28	29	27	31	33	32
	7	25	28	29	27	31	33	32
	8	25	28	29	27	31	33	32

Table 3.3: Number of jumps detected at each iteration of the filtering procedure for each American market and for each value of L.

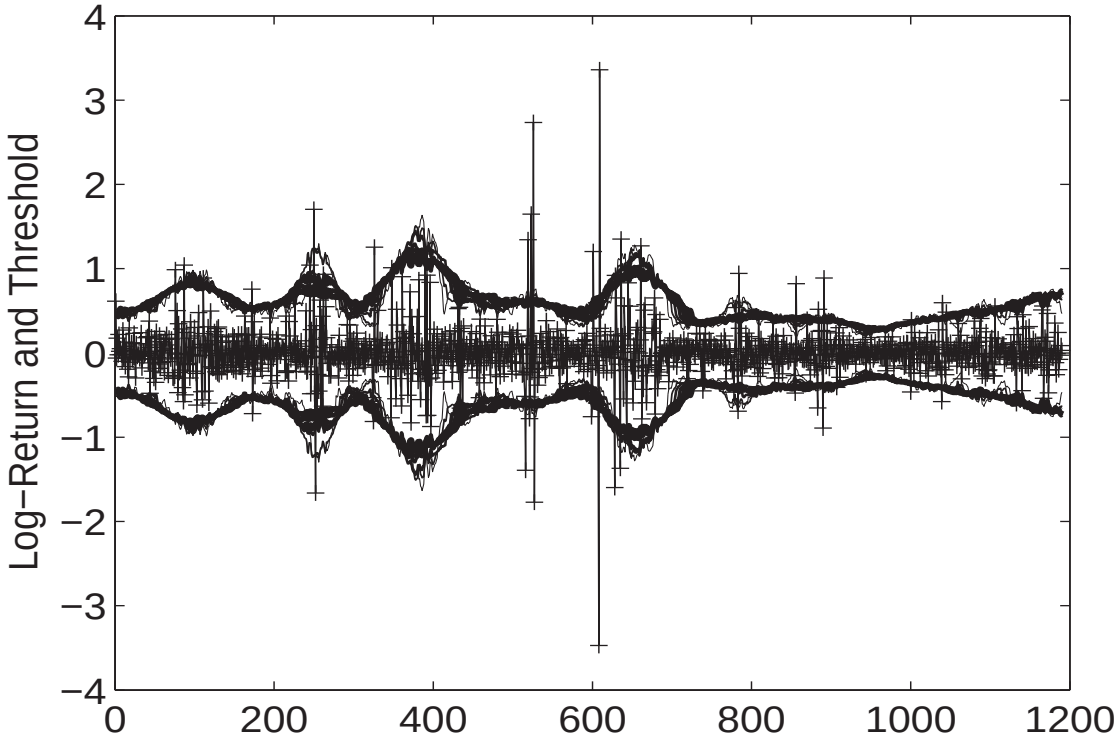

Figure 3.2: Logarithmic returns and estimated nonparametric threshold for different values of L for the German market. We compute thresholds from $L = 15$ (line with lower thickness) to $L = 45$ (line with higher thickness).

3.2.5 Estimation of the Spike Dynamics

We use the time series of spikes to compute a nonparametric estimator of the intensity function λ_t, borrowing from Mancini and Renò (2008). Our aim is to model yearly seasonality, thus we denote by $\tau \in [0,1]$ the relative position in the year, with $\tau = 0$ denoting January, 1^{st} and $\tau = 1$ denoting December, 31^{th} and set:

$$\widehat{\lambda}(\tau) = 0.5 \cdot \frac{N \sum_{i=1}^{N} K\left(\frac{\tau_i - \tau}{h_\tau}\right) c_i \, I_{\{r^2_{iT/N} \geq \theta_{iT/N}\}}}{T \sum_{i=1}^{N} K\left(\frac{\tau_i - \tau}{h_\tau}\right)}, \tag{3.52}$$

where τ_i is the relative position in the year of the $i-$th observation, and $h_\tau = 10/252$, corresponding to a bandwidth of roughly two weeks. The factor 0.5 means that we count only upward jumps, and we do not count the subsequent downward jump induced by a spike. The threshold θ_i that appears in equation (3.52) is the threshold obtained by successive iterations of equation (3.49).

The multiplying factors c_i can be set to compensate the loss of jumps which are below threshold. In fact we identify as jumps those variations which are larger than the threshold and so all jumps which are less than the threshold are not taken into account in the estimator (3.52). If we choose:

$$c_i = 1 + R(\vartheta_i), \tag{3.53}$$

where:

$$R(\vartheta_i) = \frac{\mathbb{P}(|J| \leq \vartheta_i)}{\mathbb{P}(|J| > \vartheta_i)} \tag{3.54}$$

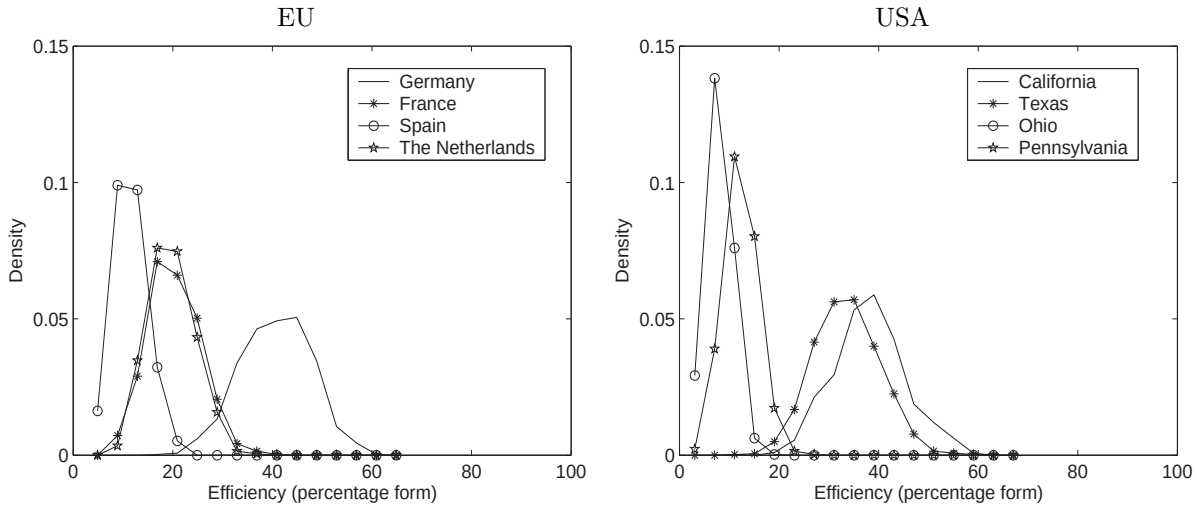

Figure 3.3: On $1,000$ replications of the estimated models reports the distribution of efficiency in detecting jumps for Europe (left) and USA (right) in percentage form.

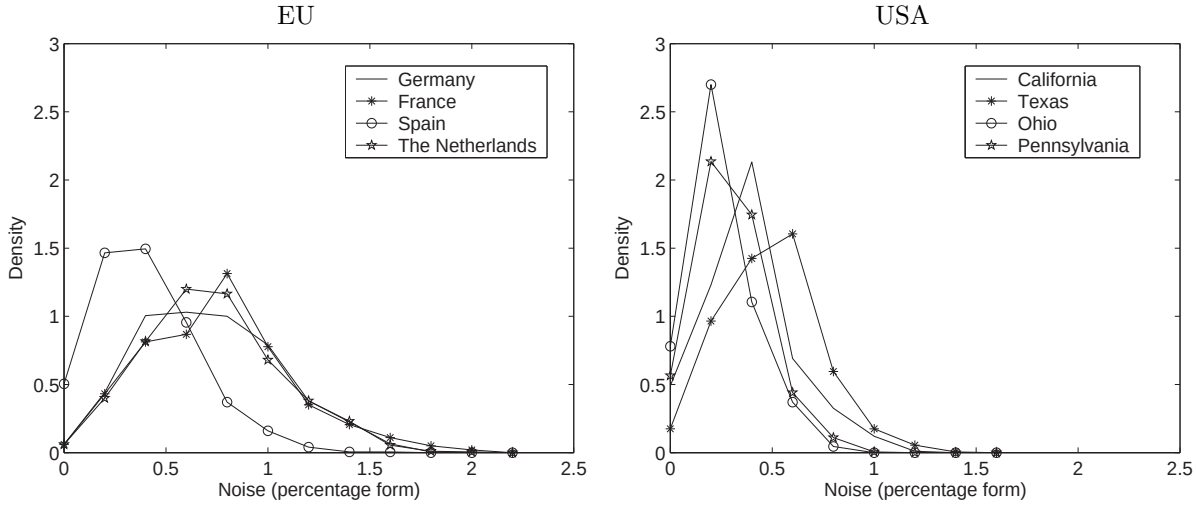

Figure 3.4: On $1,000$ replications of the estimated models reports the distribution of noise in detecting jumps for Europe (left) and USA (right) in percentage form.

represents the ratio between lost and detected jumps, we can multiply each term of the sum in the numerator of expression (3.52) accounting for the under-threshold jumps.

If the jump size is Normally distributed[4] with mean 0 and variance σ_J^2, expression (3.53) becomes:

$$c_i = \frac{1}{2\mathcal{N}\left(-\frac{\sqrt{\vartheta_i}}{\sigma_J}\right)}, \tag{3.55}$$

where $\mathcal{N}(x)$ is the cumulative of a Normal distribution function with zero mean and standard deviation equal to one.

Under the assumption of Normally distributed spike sizes, we can estimate σ_J by a simple method of moment estimator. If X is a Normal random variable such that $X \sim N\left(0, \sigma_J^2\right)$ straightforward calculations show that:

$$\mathbb{E}\left[X^2 \middle| |X| \geq c\right] = \left[\sigma_J^2 + \frac{\sigma_J\, c\, \exp\left(-\frac{c^2}{2\,\sigma_J^2}\right)}{\mathcal{N}\left(-\frac{c}{\sigma_J}\right)\sqrt{2\,\pi}}\right]. \tag{3.56}$$

Therefore the jump standard deviation can be estimated by solving the following minimization problem:

$$\min_{\sigma_J} \sum_j \left(Y_j^2 - \left(\sigma_J^2 + \frac{\sigma_J\sqrt{\theta_j}\,\exp\left(-\frac{\theta_j}{2\,\sigma_J^2}\right)}{\mathcal{N}\left(-\frac{\sqrt{\theta_j}}{\sigma_J}\right)\sqrt{2\,\pi}}\right)\right)^2$$

where Y_j is the estimated jump series detected by the threshold θ_j by means of equation (3.50).

The estimated jump standard deviation is reported in Table 3.4 for all the considered markets. They are fairly robust to the choice of the parameter L (in the subsequent analysis, we set $L = 25$). American markets present a lower σ_J than European ones. Standard errors on these numbers can be evaluated by means of Monte Carlo simulation, see Section 3.3. Figure 3.5 plots the distribution of the relative percentage error in estimating σ_J, as obtained on replications of the estimated model for European and American markets. The estimator has a slight negative bias due to normal status returns above threshold, which are spuriously detected as low spikes.

The results of the estimation of annualized jump intensity functions are reported in Figure 3.6, reporting relevant seasonal variation in the intensity of spikes, in line with the results, among others, in Geman and Roncoroni (2006). Standard errors are computed via replicating the estimated model, see Section 3.3.

[4]Again, note that while spikes in price are always positive, jumps on returns are both positive and negative. Thus, when simulating the model we assume that the spike is always positive by taking the absolute value of the Normal variable. Such a spike is added directly to the price, see Section 3.3.

	L						
	15	20	25	30	35	40	45
Germany	0.800	0.835	0.850	0.850	0.855	0.870	0.890
France	0.600	0.570	0.610	0.595	0.560	0.545	0.570
Spain	0.415	0.395	0.385	0.405	0.395	0.395	0.405
The Netherlands	0.780	0.795	0.835	0.835	0.770	0.770	0.775
California	0.250	0.245	0.240	0.235	0.245	0.245	0.245
Texas	0.240	0.235	0.230	0.240	0.245	0.240	0.235
Ohio	0.260	0.265	0.275	0.275	0.270	0.275	0.280
Pennsylvania	0.355	0.335	0.330	0.340	0.325	0.320	0.325

Table 3.4: Estimated standard deviation of the jump size for the four European markets and for the four American markets and for each value of L.

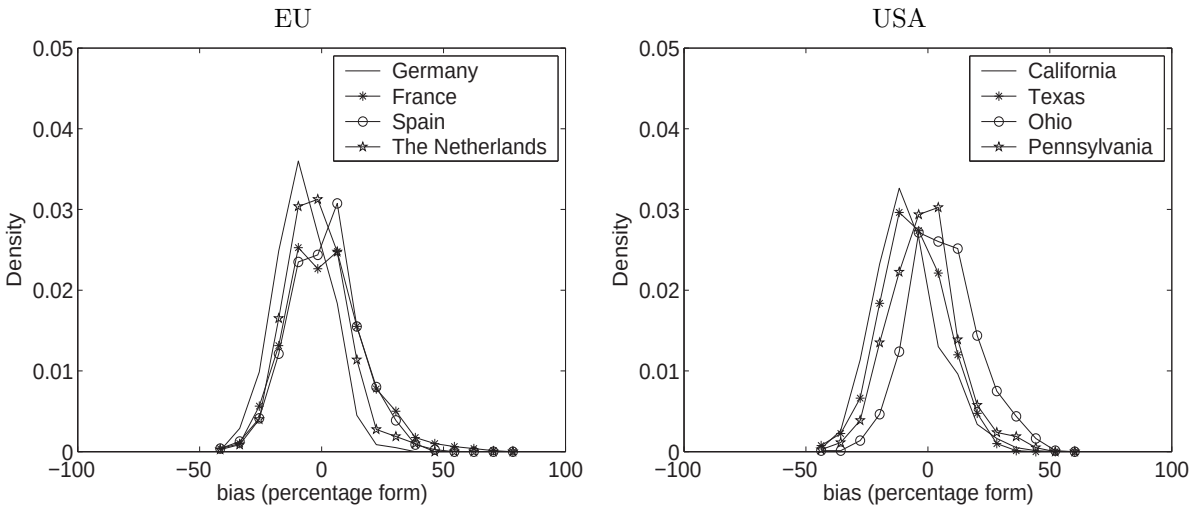

Figure 3.5: Reports the distribution of the relative bias in estimating σ_J on $1,000$ replications of the estimated model for Europe (left) and USA (right).

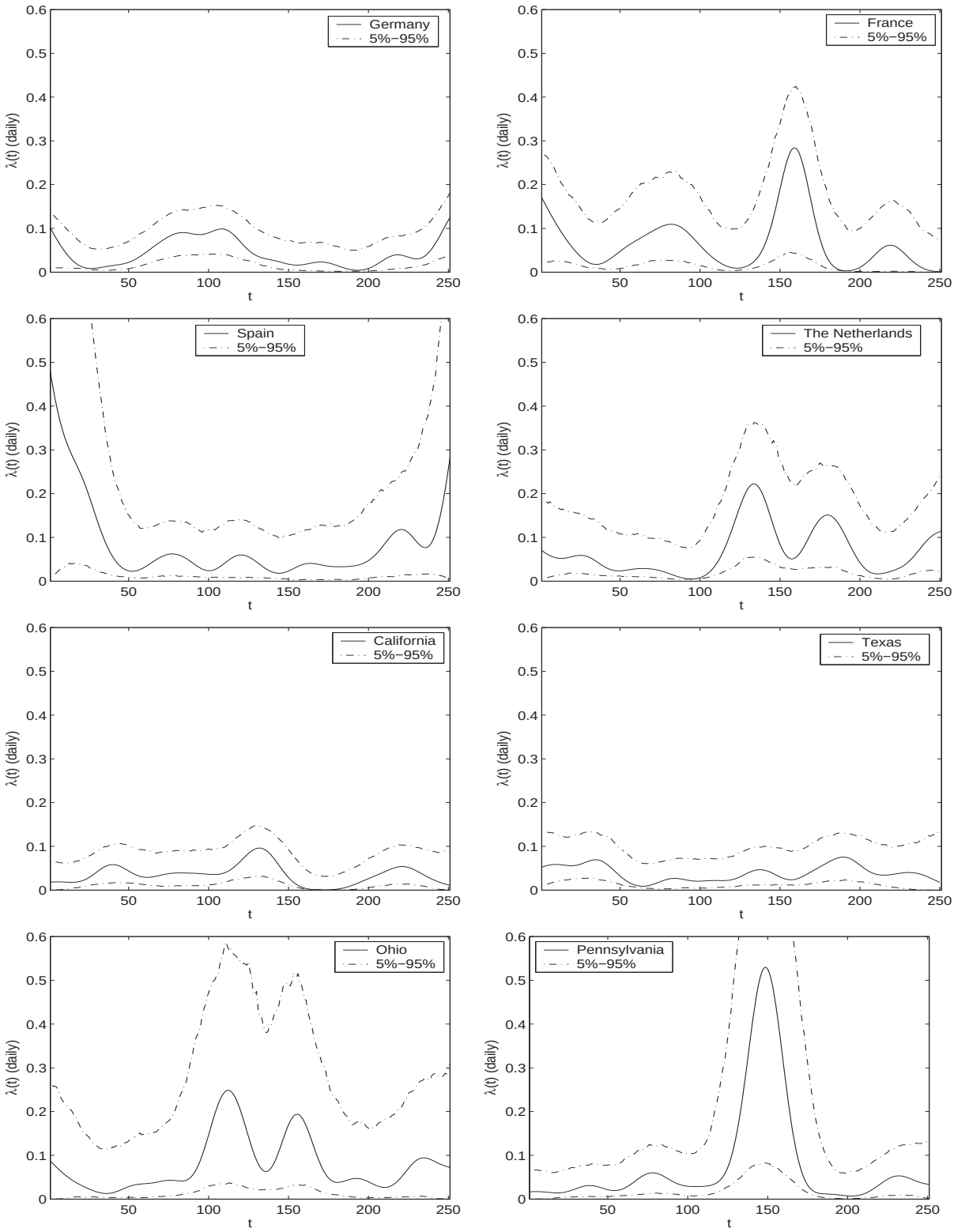

Figure 3.6: Reports the estimate of the seasonal spike intensity function λ_t over one year for Europe and USA.

3.2.6 Estimation of the Normal Status Dynamics

After removing the estimated spikes, we are left with a time series of normal status variations which we denote by r_t^c. When a spike occurs, we cannot observe the normal status return, thus we do not use that observation in the estimation procedure. To test for absence of jumps on the filtered series, that is to test whether the procedure of spike removal has been successful, we compute the $C - Tz$ test proposed by Corsi et al. (2008) on the filtered series. The test is reported in Chapter 4 and it is a fundamental part of this work. In Chapter 4 it is shown that $C - Tz$ test is distributed as a standard Normal under the null of no jumps in the time series. Results, shown in Table 3.5, show that while jumps are present in spot electricity prices, as witnessed by the high values of the $C - Tz$ statistics (with the exception of Ohio), the filtered series present values of such statistics which are consistent with the null.

Model (3.2) is then estimated on filtered returns. To estimate the functions $\mu(\cdot)$ and $\sigma(\cdot)$ we borrow again, as discussed, from the results of Mancini and Renò (2008) using Nadaraya-Watson estimators:

$$\widehat{\mu}(x) = \frac{N \sum_{i=1}^{N-1} K\left(\frac{r^c_{iT/n} - x}{h}\right) r^c_{(i+1)T/N}}{T \sum_{i=1}^{N} K\left(\frac{r^c_{iT/N} - x}{h}\right)}, \tag{3.57}$$

$$\widehat{\sigma}^2(x) = \frac{N \sum_{i=1}^{N-1} K\left(\frac{r^c_{iT/N} - x}{h}\right) \left(r^c_{(i+1)T/N}\right)^2}{T \sum_{i=1}^{N} K\left(\frac{r^c_{iT/N} - x}{h}\right)}, \tag{3.58}$$

where N is the number of observations of the filtered data and $T = N\,\Delta t$ is the width of the time window. When using daily time units, $\Delta t = 1$ and T and N disappear from equations (3.57)-(3.58). We use again the Gaussian kernel (4.20). The bandwidth parameter h is set according to the thumb rule:

$$h = \frac{h_s\,\overline{\sigma}^c}{N^{\frac{1}{5}}} \tag{3.59}$$

where $\overline{\sigma}^c$ is the sample standard deviation of the filtered data and h_s is a smoothing parameter which we set equal to 1.06 (Scott, 1992).

The results of the estimation are reported for European and American markets in Figures 3.7-3.8, together with standard errors obtained via Monte Carlo. The estimate of $\mu(\cdot)$ indicates significant mean-reversion in Europe, and mild or null mean-reversion in USA. Volatility functions are U-shaped for both European and American markets. In general European markets are more volatile and more mean-reverting than American ones. These results are in line with those obtained with parametric methods, see e.g. Gianfreda (2008) for the European markets and Hadsell et al. (2004) for American markets. Such differences are likely due to the different institutional frameworks of electricity markets in Europe and in the US, as well as to regulatory policies and different marginal costs of generation sources. Li and Flynn (2004) analyze the sources of volatility in deregulated markets.

Since standard errors are obtained re-estimating the model on simulated data which are generated by the estimates, they provide a straightforward specification test. Indeed, they show that estimates are stable, since the original estimate is inside the confidence bands. The estimate of volatility is slightly upward biased, and we find similar biases, with different magnitudes, in all the considered markets. We also plot the residuals of the estimated models on normal status dynamics in Figure 3.9. The good comparison with a standard Normal distribution indicates a satisfactory fit of the model on real data. Residuals are

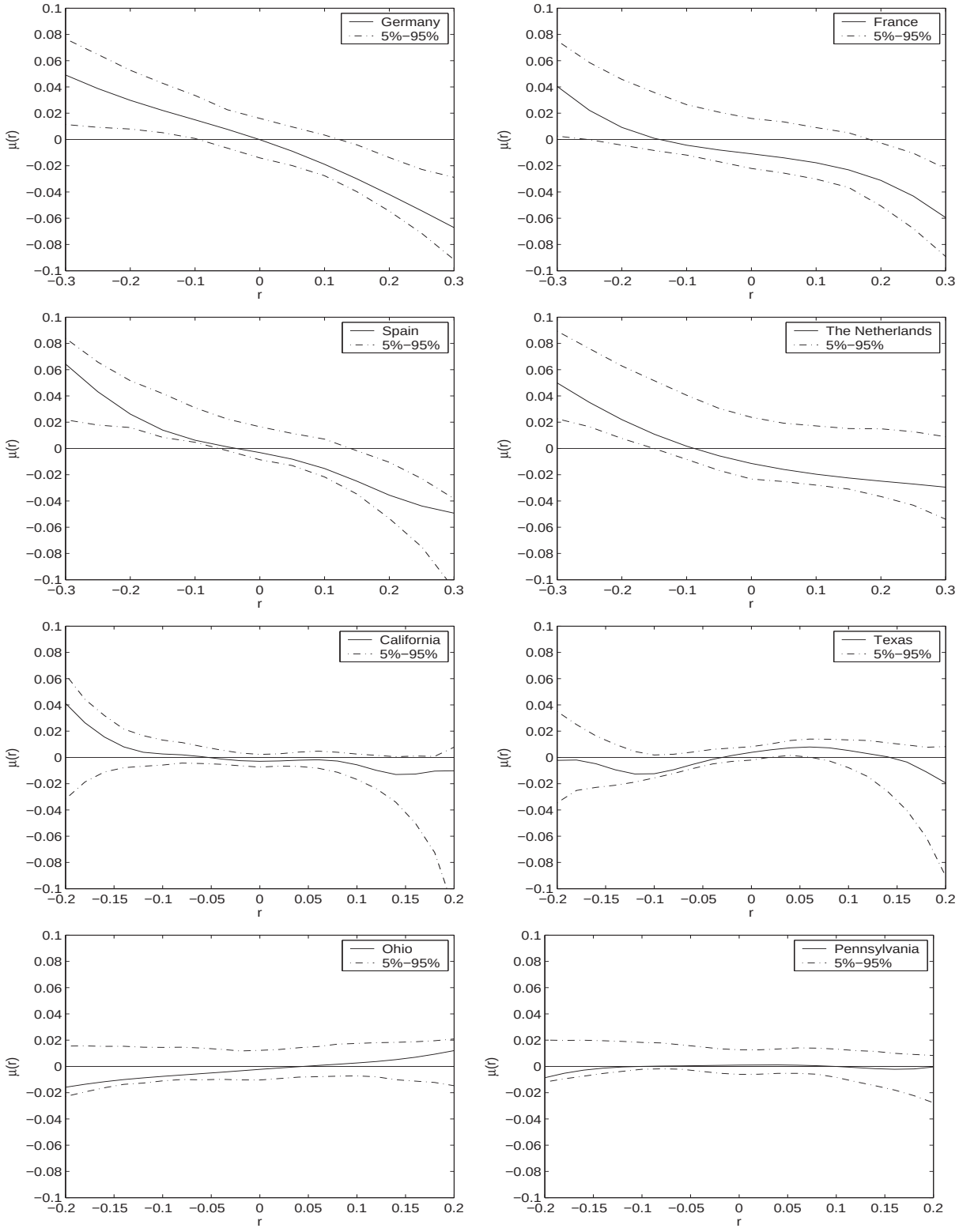

Figure 3.7: Reports the estimate of the drift function $\widehat{\mu}(r)$ for Europe and USA.

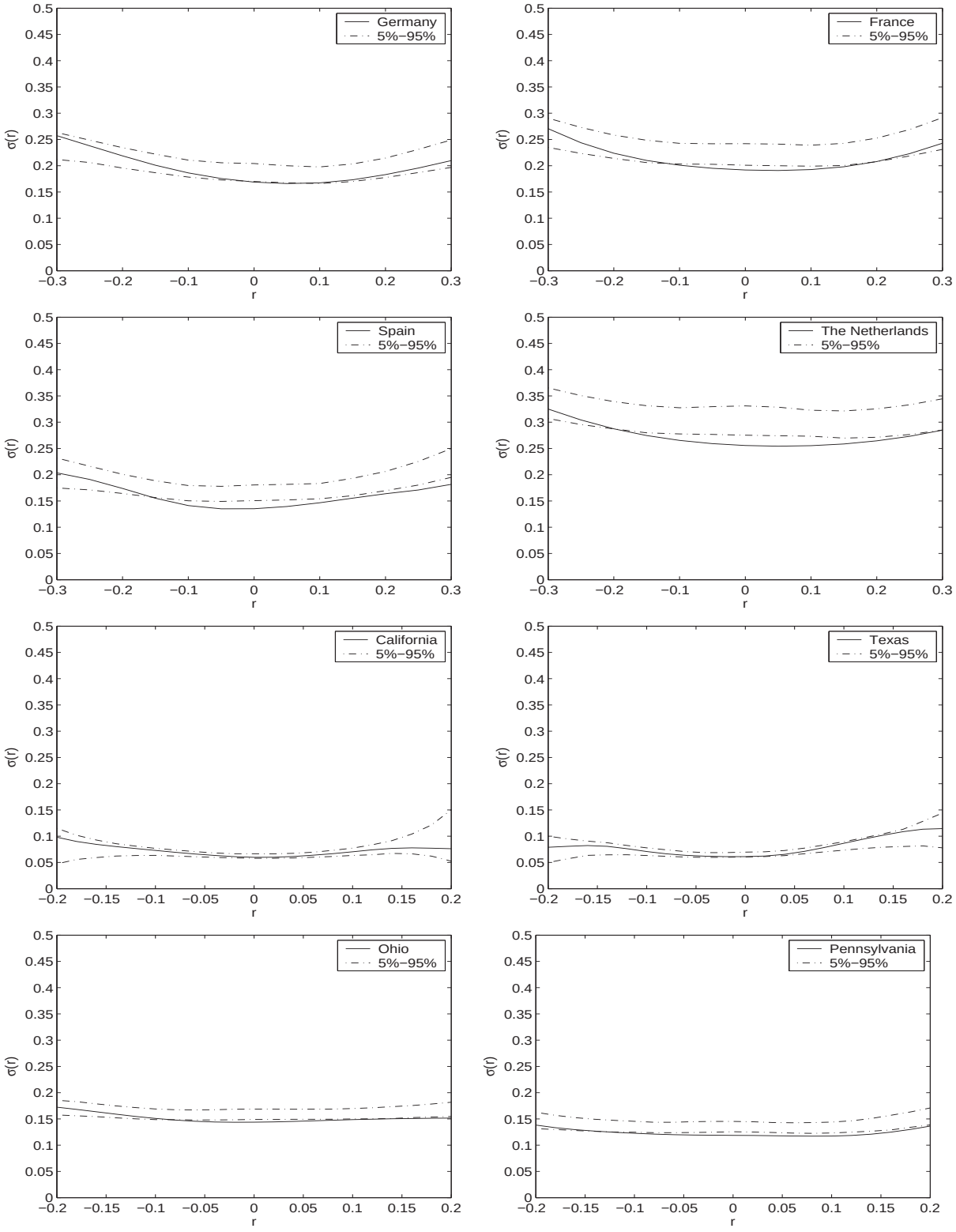

Figure 3.8: Reports the estimate of the volatility function $\widehat{\sigma}(r)$ for Europe and USA.

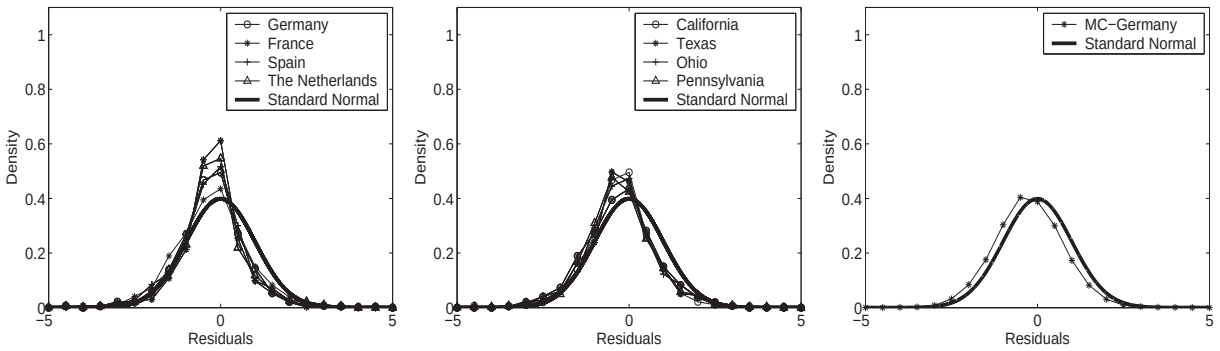

Figure 3.9: Reports the residuals of the model for the normal status dynamics, together with a standard Normal distribution, for European markets (left), American markets (center) and a Monte Carlo simulation for the German market (right).

		CT-z	
	$\widehat{\rho}(1)$	Before filtering	After filtering
Germany	-0.09	10.68	-0.29
France	-0.07	4.37	0.71
Spain	-0.08	4.99	0.28
The Netherlands	-0.09	3.20	-0.23
California	-0.02	7.37	-0.42
Texas	0.03	8.18	1.59
Ohio	0.03	1.42	0.65
Pennsylvania	0.02	5.79	1.31

Table 3.5: In the first column, reports the residuals first order autocorrelation coefficients. In the second and third column, reports the CT-z test for jumps before filtering spikes and after filtering spikes.

slightly more negatively skewed than the Normal distribution. This is (at least in part) a by-product of the estimation procedure: spikes that are not detected (when they are below threshold) contribute spuriously to continuous returns, and in an asymmetric way since they are always positive. This can be seen from the right panel of Figure 3.9, where we show the residual distribution for a Monte Carlo simulation, which shows a similar skewness. Finally, the residuals are not serially correlated: Table 3.5 reports the estimated first autocorrelation coefficient of the residuals time series, showing that they are not significantly different from zero.

Thus, in spite of its simplicity, the nonparametric model, associated with the separation of the spike dynamics from the normal status dynamics, is able to reproduce significant statistical features of the data.

3.3 Monte Carlo Simulation

To evaluate the confidence bands reported in this chapter, as well as to test the efficiency of the procedure, we use simulated replications of the estimated models.

The Nadaraya-Watson estimation procedure provides the functions $\mu(\cdot)$, $\sigma(\cdot)$ and $\lambda(\tau)$, as well as the

parameter σ_J^2, where $\tau \in [0,1]$ denotes the relative position in the year. To replicate this model, we simulate separately the normal status process r_t^c, using equation (3.2), and the spikes using a Poisson process, then we add the two components according to (3.1). More precisely, for each instant $t_k = k$ corresponding to a position τ_k in the year, we generate a random number ψ uniformly distributed in $[0,1]$; if $\psi \leq \lambda(\tau_k)$ then we add a positive spike to the normal status process. The spike is added to price P_t^c and it is set equal to $P_t^c\left(e^{|\varphi|}-1\right)$, with φ Normally distributed with mean 0 and variance σ_J^2. We use a linear interpolation on a fixed grid to get the values of the function $\mu(\cdot)$ and $\sigma(\cdot)$ for every observation.

When computing standard errors, we generate $1,000$ replications of the time series with the same number of data and, on each replication, we estimate again the functions μ, σ and λ. We use $L = 25$ for the smoothing parameters of the nonparametric threshold of equation (3.49). Moreover we choose the optimal $h_s = 1.06$ for the bandwidth parameter of the estimators (3.57)-(3.58) and a smoother $h_\tau = 10/252$ for the estimator (3.52). Confidence bands are obtained accordingly. This is also a test of specification of the model, since in the case of good specification re-estimated functions should be close to the originally estimated ones and they can be statistically evaluated using confidence bands.

3.4 Discussion of Results and Conclusions

We propose a simple nonparametric model for electricity prices in which we separate two regimes: the normal status dynamics and the spike dynamics. The normal status dynamics is flexibly driven by the observed logarithmic returns, which is the only state variable of the system. Spikes are driven by a seasonal compound Poisson process. With our model, not only we avoid all the computational burden of parametric models, but also introduce flexibility without further assumptions.

The fit of the proposed model on eight electricity spot price time series shows that the model, in spite of its simplicity of estimation, is able to capture the main statistical features of the data. We account for mean reversion, nonlinear volatility and seasonal spikes. The model residuals are not correlated and very close to a standard Normal distribution. Moreover, our spike separation technique, based on threshold methodology, is able to produce a filtered series in which jumps are absent. This result suggests that recursive threshold estimation is a suitable technique for detecting spikes in electricity time series. This is due to the fact that jumps in electricity prices are very pronounced.

We also find that the four European markets do not show distinct dynamic properties, as well as the four American markets, indicating that the pricing formation mechanism is pretty similar, which is a crucial indication for theoretical underpinning of electricity pricing. Both American and European markets show heteroskedasticity with a U-shaped volatility function. However, European and American markets are quite different. European markets display an higher level of mean reversion with respect to their American counterparts. Moreover, European markets are more volatile, both for the pure diffusion and the jump component.

Clearly, the proposed model may not reproduce statistical features which may emerge in particular markets or with longer time series, such as *inverse leverage effects* or more complicated seasonalities. In the proposed framework, however, all these refinements can be accomplished in a straightforward manner: they were not needed though to fit the considered time series.

Chapter 4

Threshold Multipower Variation

This chapter is inspired from the paper of Corsi, Pirino, and Renò (2008).

4.1 Introduction

Volatility estimation is a topic widely studied in the recent literature as it plays a key role in asset price modelling and option pricing. The typical approach consists in modelling the log-price of an asset as a continuous semimartingale (e.g. Zygadlo, 2003) plus a discontinuous process to catch its abrupt changes (e.g. Daly and Porporato, 2006). Continuous time processes with jump transitions are very common in different fields, like neuron-spikes activity in biophysics (Greinstein and Mandelbrot, 1964), modelling of earthquakes (Shcherbakov et al., 2003), hydrological models (Porporato and D'Odorico, 2004), *inter alia*. The Poisson process is the most used stochastic process in literature to induce jump occurrences (Daly and Porporato, 2007).

The continuous part of the log-price is usually modelled as a drift term and a Brownian diffusion, with the continuous volatility being the stochastic diffusion coefficient of the Brownian motion. A natural question may arise at this point: why we have to use Brownian motion for modelling asset price behaviour? We recall that two main empirical facts are addressed to the log-price of financial asset:

- No persistence is found in the autocorrelation of log-returns (note that the absolute value and the square of log-returns usually show persistence).
- The log-returns are stationary.

These stylized facts enforce the use of Lévy process (see Definition 2.10.1 of Chapter 2) for modelling the behaviour of financial assets. Therefore using Theorem 2.10.3 of Chapter 2 we can always decompose the process as a sum of a Brownian motion, with appropriate diffusion coefficient, plus a jump process.

Many stylized facts regarding volatility have been established in the past years. The literature refers to volatility clustering to that phenomenon analyzed by Mandelbrot (1963) who observed that "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes".

Return intervals τ between price volatilities that are above a certain threshold q are investigated by Wang et al. (2006) and Eisler et al. (2007). A similar approach is taken by Lillo and Mantegna (2003) who investigate the number of times the absolute value of an index return is exceeding a given threshold value after a financial crash (like Black Monday). They found a power-law evolution (known as Omori law) for the aftercrash distribution of the number of events exceeding the selected threshold. They further observe that this type of evolution lasts for approximately 60 trading day. Such a relaxation behaviour is observed in many complex systems like the earth crust: in an earthquake the number of the aftershock following the main shock obeys the Omori law. Finally Lillo and Mantegna (2003) suggest that "one possibility for this common occurrence is that the Omori law is a phenomenological manifestation of underlying common microscopic mechanisms governing the dynamics of complex systems after an extreme event". The same approach is taken by Weber et al. (2006) for volatility return intervals. They found that Omori law holds also for volatility in the aftercrash period. Moreover they found that the aftercrash behaviour of the volatility is related to the long-term correlation found for volatility sequences of financial markets. This is a key issue linked to the results proposed in this chapter, where an important connection between jumps and volatility forecasting is found.

Physicists have developed fascinating techniques to investigate economic information held in financial stocks, as those suggested in the paper of Mantegna (1999). He proposes a topological space in order to successfully identify a hierarchical structure in the stocks entering the Dow Jones Industrial Average (DJIA) and the Standard and Poor's 500 (S&P500). This kind of analysis is useful in detecting common economic factors affecting different stock. A similar approach (clustering algorithm) is used by McDonald et al. (2008) to investigate the impact of shocking news (like the 9/11 terrorist attack) on the foreign exchange (FX) market dynamics.

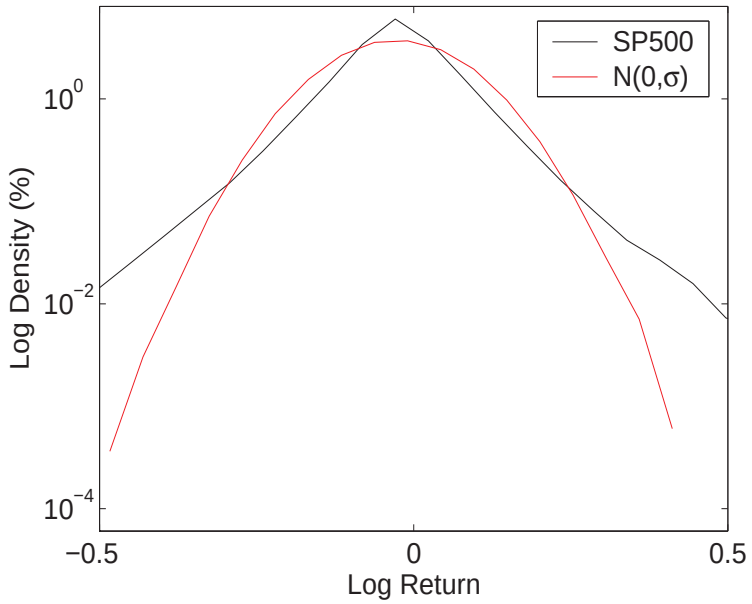

Figure 4.1: Shows a comparison between empirical distribution of S&P500 (for the period $1990 - 2004$), reported in black, and a Normal distribution with the same sample mean and standard deviation, reported in red. The plot has a logarithmic scale on the vertical axis in order to show heavy tails.

Shocking news, unpredictable events and governmental announcements affect financial markets as a trigger for large (positive or negative) variations. The presence of such variations results in heavy tails in the empirical distribution of returns. The tails of the distribution decay slowly and very large moves have a non-zero probability to occur. This can be seen by inspection of Figure (4.1). We report a comparison between the empirical log-distributions of S&P500 five minutes log-returns (15 years of observations) and gaussian log-returns, both series having the same sample standard deviation. While S&P500 experiences variations of even $3\% - 4\%$ the simulated gaussian process is confined within absolute variations smaller than 0.5%.

A local volatility diffusion process like:

$$dX_t = \mu_t \, dt + \sigma \left(t, X_t \right) \, dW_t, \tag{4.1}$$

even though driven by a gaussian noise dW_t is not a gaussian process. An appropriate choice of the nonlinear diffusion coefficient $\sigma \left(t, X_t \right)$ can generate diffusion processes with arbitrary heavy tails. However modelling log-return variations with a continuous model like (4.1) or with stochastic volatility like:

$$\begin{aligned} dX_t &= \mu_t^X \, dt + \sqrt{v_t} \, dW_t^a \\ d \log v_t &= \left(\alpha - \beta \, \log v_t \right) \, dt + \eta \, dW_t^b \end{aligned} \tag{4.2}$$

forces to have an high value for the volatility of volatility η or an highly varying diffusion coefficient $\sigma \left(t, X_t \right)$, in order to reproduce heavy tails. Moreover models (4.1)-(4.2) cannot generate sudden, discontinuous price moves: tail events are the result of the accumulation of many small moves.

These issues can be easily circumvented modelling jumps with a random counting process, like the compound poisson process defined in Section (2.6). Besides the introduction of jumps in financial modelling is enforced analyzing the option market.

Stochastic modelling in finance has been mainly developed for option pricing. Options are contracts written on underlying variables, which could be assets, indexes, interest rates and similar assets. As an example a *call* option with strike price K, maturity date T, which pays no dividends and written on S_t, gives to the holder the final payoff: $\phi \left(T, K \right) = \max \left(S_T - K, 0 \right)$. The price of such an option at a generic instant t depends on the model assumed for S_t. The specification of the model is then adapted fitting the theoretical curve for option prices with the empirical data. A misspecification of the model often results in a bad fit, which does not satisfactorily reproduce the empirical features of option prices. As firstly observed by Merton (1976) jump-diffusion models provide a better comprehension of option price curves providing exhaustive explanations of the main stylized facts.

A list of recent studies on jumps in financial assets includes test specification of Aït-Sahalia (2004), Jiang and Oomen (2006), Barndorff-Nielsen and Shephard (2006) and Y.Aït-Sahalia and Jacod (2008), as well as the empirical studies of Bollerslev, Law and Tauchen (2007) and Maheu and McCurdy (2004); Bollerslev et al. (2007); Andersen et al. (2006); nonparametric estimation in the presence of jumps, as in Bandi and Nguyen (2003); Johannes (2004); Mancini and Renò (2008); option pricing as in Duffie et al. (2002); Eraker et al. (2003); Eraker (2004). Interesting references for a review are Cont and Tankov (2004) and Barndorff-Nielsen and Shephard (2007).

However, while jumps have been shown to be relevant in economic and financial applications, they still have no direct role for volatility forecasting.

We start from an apparent puzzle contained in the study of Andersen et al. (2007) (henceforth ABD) and Forsberg and Ghysels (2007); Giot and Laurent (2007); Busch et al. (2006). In these works, jumps have been found to possess a negative or null impact in determining future volatility. We find this result puzzling in at least two respects. First, visual inspection of realized volatility time series reveals that bursts in volatility are usually initiated by a large and unexpected movement of asset prices; this suggests that jumps should have a forecasting power for volatility. Figure (4.2) provides a clarification of this point. It shows the series of S&P500 five-minutes daily realized volatility[1] for the same sample used for Figure (4.1). The top of the figure reports the whole sample, formed by 3736 days, while in the bottom an enlargement of an afterjump period is shown. The visualized jump is a variation of 13.06% which takes place at day number 3135, but greater variations are still present. In both plots, for a better identification of the phenomenon, we have depicted in red the period of 300 days after the jump. It is evident that, after the jump, realized volatility begins a burst which lasts for about the whole marked period[2].

Second, it is well known that volatility is associated with dispersion of beliefs and heterogeneous information, see e.g. Shalen (1993); Wang (1994). The occurrence of a jump is likely to increase the uncertainty on fundamental values, and thus it is likely to have a positive impact on future volatility.

In this and subsequent chapters three main contributions to the literature are proposed. The first contribution is to show that this puzzle is due to the fact that preceding studies in the literature use a biased estimator of the continuous integrated volatility: the bipower variation (defined in Section 4.2). It is possible to show that bipower variation, even in presence of jumps, asymptotically converges to continuous integrated volatility when the time distance between two consecutive log-price observations shrinks to zero.

However at very high sampling frequency financial assets display a particular behaviour known as "market microstructure". The microstructure noise is generated by the mechanism of price formation, when bid and ask quotes try to match. At high frequency the transaction price bounces between the bid and ask quotes. For this reason the time interval between two consecutive observations must be chosen high enough to reduce the bias generated by the microstructure [3]. As a consequence the bias generated by jumps is not completely removed.

Our simulation studies show that, in finite samples, bipower variation is largely upper biased in presence of jumps, and this implies a large underestimation of the jump component.

An alternative estimator of integrated volatility has been introduced by Mancini (2007) and studied by Aït-Sahalia and Jacod (2007) in a broader context. This estimator is potentially less biased than bipower

[1] I.e. the daily sum of the squared five-minutes log-returns: equation (2.76), with $T = 1$ day, $\delta = 5$ min, $X_t = \log(\text{S\&P500})_t$.

[2] This is a purely heuristic argument.

[3] Attempts to study and correct bipower variation under microstructure noise can be found in Podolskij and Vetter (2006), and Christensen et al. (2008). ABD and Huang and Tauchen (2005) propose a staggered version of bipower variation. Fan and Wang (2007) study jumps and microstructure noise with wavelet methods. The impact of microstructure noise on threshold estimation of Mancini (2007) is instead unknown. Our results can potentially be extended to account for microstructure noise. Directly incorporating microstructure noise can improve volatility forecasting, see e.g. Aït-Sahalia and Mancini (2008).

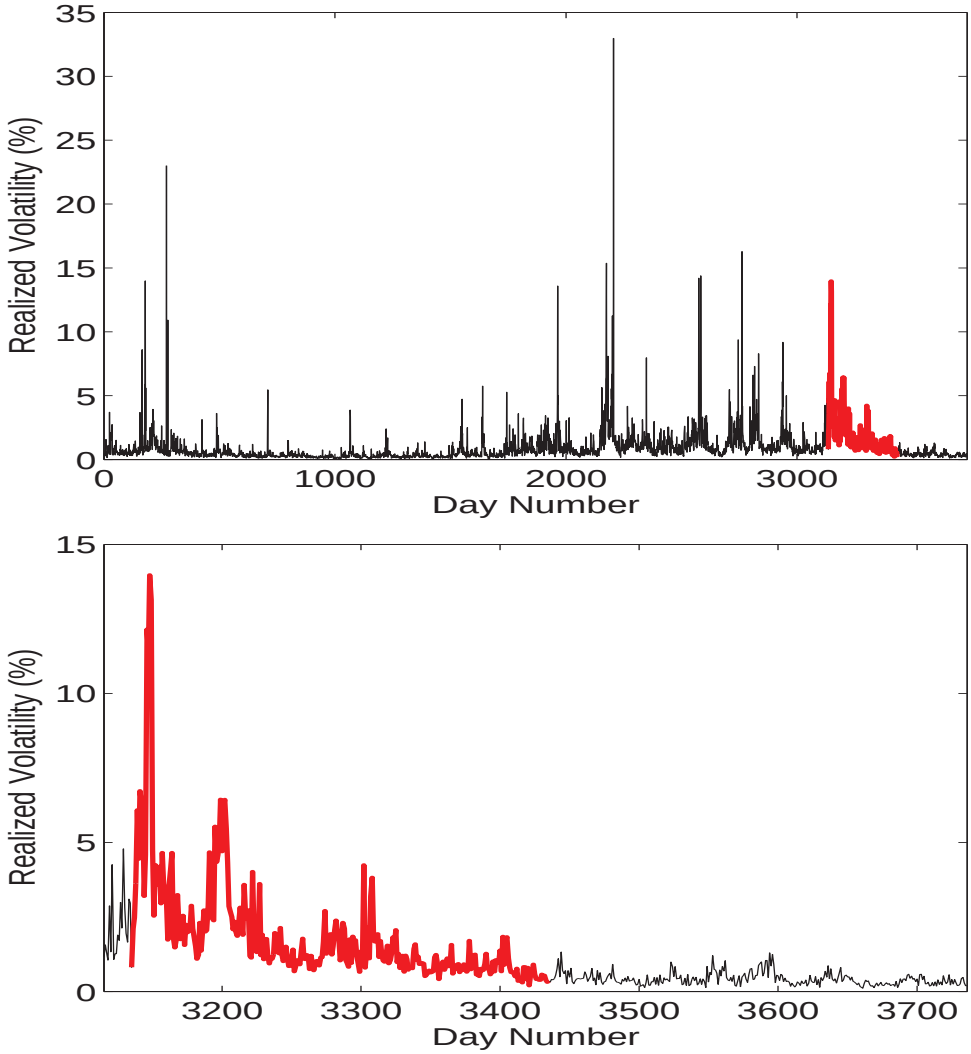

Figure 4.2: (Top) Realized volatility (in percentage form) computed for 15 years of S&P500 starting from December 7, 1989 and ending at December 22, 2004. A period of 300 days after a variation of 13.06% taking place at day number 3135 is marked in red. (Bottom) Enlargment of the top panel of the highlighted period. The high variation of volatility generates a burst which lasts for about the overall 300 days.

variation, but it requires the auxiliary estimation of a threshold function; thus, when it is used in its original form, it provides estimates of continuous volatility which are very sensitive to the specification of the threshold.

The second contribution of this work is thus to introduce an alternative estimator of integrated powers of volatility in presence of jumps. We introduce the concept of *threshold multipower variation*, which is a combination of the above mentioned techniques. Using realistic simulation of asset prices (Section 4.4), we show that threshold bipower variation is nearly unbiased on continuous trajectories and, importantly, also in the presence of jumps. Moreover, it is robust to the choice of the threshold function, in the sense that the impact of the threshold on estimation is marginal. Thus, it is an ideal candidate to estimate dynamic models of volatility in which we use separately the continuous and discontinuous volatility as explanatory variables.

Our third contribution is the introduction of a novel test for jump detection in time series. Our test is basically a correction of the z statistics of Barndorff-Nielsen and Shephard (2006) based on threshold multipower variation. This correction removes the bias in estimating the integral of the second and fourth power of continuous volatility in presence of jumps. We show that our $\mathsf{C}-\mathsf{Tz}$ test is sized as the traditional z-test under the null. In presence of jumps instead, the corrected test has significantly more power than the z test, especially when jumps are consecutive, a situation which is quite frequent in high-frequency data. We show this on simulated data (Section 4.4) and on time series which are well known to display very large jumps (Section 4.5), namely electricity prices and interest rate data, for which the z-test has a disappointingly low power. The forecasting model inspired by the ideas proposed in this chapter is reported in Chapter 6.

4.2 Disentangling Diffusion from Jumps

4.2.1 Introductory Concepts

We work in a filtered probability space $(\Omega, (\mathcal{F}_t)_{t\in[0,T]}, \mathcal{F}, \mathcal{P})$, satisfying the usual conditions (Protter, 1990). We assume that an economic variable X_t, for example the logarithmic price of a stock or an interest rate, satisfies the following assumption:

Assumption 4.2.1 *$(X_t)_{t\in[0,T]}$ is a real-valued process such that $X_0 \in \mathbb{R}$ and*

$$dX_t = \mu_t dt + \sigma_t dW_t + c_t dN_t \tag{4.3}$$

where μ_t is predictable, σ_t is cádlág and N_t is a Poisson process whose intensity is an adapted stochastic process λ_t, the times of the corresponding jumps are $(\tau_j)_{j=1,\dots,N_T}$ and c_j are i.i.d. adapted random variables measuring the size of the jump at time τ_j.

Typically, in financial econometrics a time window T is fixed, e.g. one day, and we define the quantities of interest on a time span of length T. Quadratic variation of such a process over a time window is defined as (see Section 2.9):

$$[X]_t^{t+T} := X_{t+T}^2 - X_t^2 - 2\int_t^{t+T} X_{s-} dX_s, \tag{4.4}$$

where t indexes the interval, typically a day, and it can be decomposed into its continuous and discontinuous component, as:

$$[X]_t^{t+T} = [X^c]_t^{t+T} + [X^d]_t^{t+T} \tag{4.5}$$

where $[X^c]_t^{t+T} = \int_t^{t+T} \sigma_s^2 ds$ and $[X^d]_t^{t+T} = \sum_{j=N_t}^{N_{t+T}} c_{\tau_j}^2$, where c_{τ_j} is the size of the j-th jump at time τ_j. To estimate these quantities, we divide the time interval $[t, t+T]$ into n subintervals of length δ, thus $\delta = T/n$. On this grid, we define the evenly sampled returns as:

$$\Delta_{j,t}X = X_{j\delta+t} - X_{(j-1)\delta+t}, \qquad j = 1, \dots, n \tag{4.6}$$

For simplicity, in what follows we omit the subscript t and we simply write $\Delta_j X$. An estimator of $[X]_t^{t+T}$ is given by realized volatility, defined as:

$$\mathsf{RV}_\delta(X)_t = \sum_{j=1}^{n} (\Delta_j X)^2 \xrightarrow{\delta\to 0} [X]_t^{t+T} \tag{4.7}$$

where the above convergence is in probability.

To disentangle the continuous quadratic variation from the discontinuous one, multipower variation has been introduced by Barndorff-Nielsen and Shephard (2006) and it is defined as:

$$\mathsf{MPV}_\delta(X)_t^{[\gamma_1,\ldots,\gamma_M]} = \delta^{1-\frac{1}{2}(\gamma_1+\ldots+\gamma_M)} \sum_{j=M}^{[T/\delta]} \prod_{k=1}^{M} |\Delta_{j-k+1}X|^{\gamma_k} \tag{4.8}$$

It is a natural extension of the concept of bipower variation, studied in Barndorff-Nielsen and Shephard (2004); Barndorff-Nielsen et al. (2006). Asymptotic properties of multipower variation have been studied by Barndorff-Nielsen et al. (2006) in absence of jumps, and by Barndorff-Nielsen, Shephard and Winkel (2006) and Woerner (2006) in presence of jumps. With very mild assumptions, and in some cases for infinite activity Levy jump processes, Barndorff-Nielsen, Shephard and Winkel (2006) show that:

$$p - \lim_{\delta\to 0} \mathsf{MPV}_\delta(X)_t^{[\gamma_1,\ldots,\gamma_M]} = \left(\prod_{k=1}^{M} \mu_{\gamma_k}\right) \int_t^{t+T} \sigma_s^{\gamma_1+\ldots+\gamma_M} ds \tag{4.9}$$

where $\mu_\gamma = E(|u|^\gamma) = 2^{\gamma/2}\frac{\Gamma(\frac{\gamma+1}{2})}{\Gamma(1/2)}$ and the above convergence is in probability. For practical applications, multipower variation is used for the estimation of $\int_t^{t+T} \sigma_s^2 ds$ and $\int_t^{t+T} \sigma_s^4 ds$. For this motivation, we define some relevant special cases of multipower variation, according to suitable choices of the vector $\gamma_1, \ldots, \gamma_M$. The most important example is bipower variation:

$$\mathsf{BPV}_\delta(X)_t = \mu_1^{-2}\mathsf{MPV}_\delta(X)_t^{[1,1]} = \mu_1^{-2} \sum_{j=2}^{[T/\delta]} |\Delta_{j-1}X| \cdot |\Delta_j X| \xrightarrow{\delta\to 0} \int_t^{t+T} \sigma_s^2 ds \tag{4.10}$$

with $\mu_1 \simeq 0.7979$. Estimators of $\int_t^{t+T} \sigma_s^4 ds$ are widely described in Section 4.3.

Mancini (2007) provides an alternative estimator, with respect to multipower variation, of squared and fourth power volatility. These estimators are defined as follows:

$$\mathsf{TRV}_\delta(X)_t = \sum_{j=1}^{[T/\delta]} |\Delta_j X|^2 I_{\{|\Delta_j X|^2 \le \Theta_j(\delta)\}} \xrightarrow{\delta\to 0} \int_t^{t+T} \sigma_s^2 ds \tag{4.11}$$

$$\mathsf{TQV}_\delta(X)_t = \frac{1}{3\delta} \sum_{j=1}^{[T/\delta]} |\Delta_j X|^4 I_{\{|\Delta_j X|^2 \le \Theta_j(\delta)\}} \xrightarrow{\delta\to 0} \int_t^{t+T} \sigma_s^4 ds \tag{4.12}$$

where the above convergence is in probability, and where the threshold functions has to satisfy

$$\lim_{\delta\to 0} \Theta_j(\delta) = 0, \qquad \lim_{\delta\to 0} \frac{\delta \log \frac{1}{\delta}}{\Theta_j(\delta)} = 0, \tag{4.13}$$

that is they have to vanish slower than the modulus of continuity of the Brownian motion, as previously discussed in Section 3.2.3 of Chapter 3. Mancini (2007) also establishes a central limit theorem for TRV.

4.2.2 Threshold Multipower Variation

We now introduce the following extension of multipower variation. In what follows, we use a strictly positive threshold function $\vartheta_s : [t, t+T] \to \mathbb{R}^+$, which, contrary on the threshold function used in (4.11)-(4.12), does not depend on δ, so it is fixed and does not need to vanish as $\delta \to 0$. For brevity, we denote by $\vartheta_j = \vartheta_{j\delta}$. We define the threshold multipower variation estimator as follows:

Definition 4.2.2 *We define the (realized) threshold multipower variation as:*

$$\mathsf{TMPV}_\delta(X)_t^{[\gamma_1,\ldots,\gamma_M]} = \delta^{1-\frac{1}{2}(\gamma_1+\ldots+\gamma_M)} \sum_{j=M}^{[T/\delta]} \prod_{k=1}^{M} |\Delta_{j-k+1}X|^{\gamma_k} I_{\{|\Delta_{j-k+1}X|^2 \leq \vartheta_{j-k+1}\}} \tag{4.14}$$

The intuition behind the concept of threshold multipower variation is the following. Suppose $\Delta_j X$ contains a jump. In the case of bipower variation, it will multiply two adjacent returns, $\Delta_{j-1}X$ and $\Delta_{j+1}X$. Asymptotically, both these returns will vanish and bipower variation will converge to integrated continuous volatility. But for finite δ these returns will not vanish, causing a positive bias which will be larger as $\Delta_j X$ increases. This consideration suggests that the bias of multipower variation will be extremely large in case of consecutive jumps, as will be shown in Section 4.5 . For estimators (4.14) instead, if $\Delta_j X$ contains a jump larger than ϑ_j, the corresponding indicator function vanishes both for finite δ and when $\delta \to 0$, thus correcting for the bias. This intuition is supported by the analysis in the subsequent sections.

Formally, we can state the following Theorem regarding asymptotic behavior of threshold multipower variation:

Theorem 4.2.3 *Assume 4.2.1 holds, and that θ_t is a real positive mapping defined on $[0,T]$. Then, as $\delta \to 0$,*

1. *if $\max(\gamma_1,\ldots,\gamma_M) < 2$ we have:*

$$\mathsf{TMPV}_\delta(X)_t^{[\gamma_1,\ldots,\gamma_M]} \longrightarrow \left(\prod_{k=1}^{M} \mu_{\gamma_k}\right) \int_t^{t+T} \sigma_s^{\gamma_1+\ldots+\gamma_M} ds \tag{4.15}$$

   where the above convergence is in probability.

2. *if $\max(\gamma_1,\ldots,\gamma_M) < 1$ we have:*

$$\delta^{-\frac{1}{2}} \left(\mathsf{TMPV}_\delta(X)_t - \int_t^{t+T} \sigma_s^{\gamma_1+\ldots+\gamma_M} ds\right) \longrightarrow c_\gamma \int_t^{t+T} \sigma_s^{\gamma_1+\ldots+\gamma_M} dW'_s \tag{4.16}$$

   where the above convergence is in law and:

$$c_\gamma^2 = \prod_{k=1}^{M} \mu_{2\gamma_k} - 2(M-1)\prod_{k=1}^{M} \mu_{\gamma_k}^2 + 2\sum_{j=1}^{M-1} \left(\prod_{k=1}^{j} \mu_{\gamma_k} \prod_{k=M-j+1}^{M} \mu_{\gamma_k} \prod_{k=1}^{M-j} \mu_{\gamma_k+\gamma_{k+j}}\right) \tag{4.17}$$

Proof. Under Assumption 4.2.1, we write:

$$X = Y + Z$$

where $Y_t = \int_t^{t+T} \mu_s ds + \int_t^{t+T} \sigma_s dW_s$. If $Z = 0$, the Theorem has been proved by Barndorff-Nielsen et al. (2006). Since Z is a finite activity jump process and Y is continuous, for every trajectory there exists δ' such that the number of returns such that $|\Delta X_j|^\gamma > \vartheta$ is finite. For $\delta < \delta'$, the difference $\mathsf{TMPV}_\delta(X)_t^{[\gamma_1,\ldots,\gamma_M]} - \mathsf{TMPV}_\delta(Y)_t^{[\gamma_1,\ldots,\gamma_M]}$ is zero for those terms who do not contain jumps; in the remainder every term contains at most one jump, which is summed up over M finite terms, with only one term containing a jump. For the others, we have:

$$\delta^{-\frac{1}{2}} |\Delta_j X| = O_p(|\log\delta|^{\frac{1}{2}}).$$

Thus the difference is $O_p\left(\left(\delta \log\frac{1}{\delta}\right)^{\frac{1}{2}(\sum_{i=1}^{M}\gamma_i - \max(\gamma_1,\ldots,\gamma_M))}\right)$. Thus

$$\delta^{1-\sum_{i=1}^{M}\gamma_i} \left(\mathsf{TMPV}_\delta(X)_t^{[\gamma_1,\ldots,\gamma_M]} - \mathsf{TMPV}_\delta(Y)_t^{[\gamma_1,\ldots,\gamma_M]}\right) =$$

$$= O_p\left(\delta^{1-\frac{1}{2}\max(\gamma_1,\ldots,\gamma_M)}\left(\log\frac{1}{\delta}\right)^{\frac{1}{2}(\sum_{i=1}^{M}\gamma_i-\max(\gamma_1,\ldots,\gamma_M))}\right)$$

The latter term is at least $O_p(1)$ if $\max(\gamma_1,\ldots,\gamma_M)<2$, which implies that $\mathsf{TMPV}_\delta(X)_t^{[\gamma_1,\ldots,\gamma_M]}$ has the same limit in probability of $\mathsf{TMPV}_\delta(Y)_t^{[\gamma_1,\ldots,\gamma_M]}$; and it has the same limit in law if $\max(\gamma_1,\ldots,\gamma_M)<1$, since in this case the difference is $O_p(\delta^{\frac{1}{2}})$. □

Then, not surprisingly, threshold multipower variation has the same distribution of multipower variation and threshold power variation when $\delta\to 0$. However, our interest is in estimating $\int\sigma_s^2 ds$ when δ is finite and large enough to avoid microstructure effects (typically, $\delta=5$ minutes), and in this case we will show that the three estimators are different.

As a special case for estimating integrated variance, we introduce threshold bipower variation as follows:

$$\mathsf{TBPV}_\delta(X)_t=\mu_1^{-2}\mathsf{TMPV}_\delta(X)_t^{[1,1]}=\mu_1^{-2}\sum_{j=2}^{[T/\delta]}|\Delta_{j-1}X|\cdot|\Delta_jX|I_{\{|\Delta_{j-1}X|^2\le\vartheta_{j-1}\}}I_{\{|\Delta_jX|^2\le\vartheta_j\}} \tag{4.18}$$

For estimators of the integrated fourth-power of volatility, see Section 4.3.

Now, we describe our technique to select a suitable threshold, and show how to implement threshold multipower variation when δ is finite. In what follow we adopt the same non-parametric threshold described in Section 3.2.4 of Chapter 3.

4.2.3 Estimating the Threshold Function

The threshold we use is based on the estimate of the continuous local variance by means of an iterating smoothing filter which excludes jumps using the threshold function itself.

To estimate the local variance, we use a local linear volatility estimator (Fan and Yao, 2003) adapted for the presence of jumps:

$$\hat{V}_t^Z=\frac{\sum_{i=-L,\ i\neq-1,0,1}^{L}K\left(\frac{i}{L}\right)(\Delta_{t+i}X)^2\,I_{\{(\Delta_{t+i}X)^2\le c_V\cdot\hat{V}_{t+i}^{Z-1}\}}}{\sum_{i=-L,\ i\neq-1,0,1}^{L}K\left(\frac{i}{L}\right)I_{\{(\Delta_{t+i}X)^2\le c_V\cdot\hat{V}_{t+i}^{Z-1}\}}},\qquad Z=1,2,\ldots \tag{4.19}$$

with the starting value set to $\hat{V}^0=+\infty$, which corresponds to using all observations in the first step, and $c_V=9$. Asymptotic properties for this kind of estimator in absence of jumps can be found in Kristensen (2007).

The estimation of local variance via equation (4.19) is iterating in Z. At each iteration, jumps are detected by the condition $(\Delta_tX)^2>c_V\cdot\hat{V}_t^{Z-1}$ and removed from the time series by means of the indicator function; each estimate of the variance is multiplied by c_V to get the threshold for the following step. For estimation of the local variance at time t, we do not use the adjacent observations ($i\neq-1,0,1$). The iterations stop when the removed jumps are the same. On high frequency data, this always happens with $Z=2,3$ iterations.

The bandwidth parameter L determines the time scale of returns included in the estimation of the threshold around point t. In our application, its choice results not to be crucial. We set $L=25$. We recall

(Section 3.2.4, Chapter 3) that the choice of the kernel function $K(\cdot)$ is not influential in this kind of applications (Silverman, 1986; Wand and Jones, 1995). We use a Gaussian kernel:

$$K(y) = \left(1/\sqrt{2\,\pi}\right) \exp\left(-y^2/2\right). \tag{4.20}$$

Then we set the threshold function proportional to the estimated local variance

$$\vartheta_t = c_\vartheta^2 \cdot \hat{V}_t^Z \tag{4.21}$$

A typical value of c_ϑ is $c_\vartheta = 3$. However, the dimensionless parameter c_ϑ can be used to scale the threshold with respect to local variance, and by varying it we can test the robustness of proposed estimators with respect to the choice of the threshold.

4.2.4 A Corrected Test for Jump Detection

While threshold multipower variation has the same asymptotic law of multipower variation, we expect it to provide better estimates in small samples. However, for finite δ , when $|\Delta X|^2 > \vartheta$ the indicator function in (4.14) zeros its relative addend. This can be an issue when testing for the presence of jumps, for example with the z statistics introduced by Barndorff-Nielsen and Shephard (2006), since under the null of no jumps there are still variations larger than the threshold. These variations are also the larger ones, thus TMPV is expected to be negatively biased under the null. However, this issue can be solved by correcting the estimator according to the following rule: when $|\Delta X|^2 > \vartheta$ we replace $|\Delta X|^\gamma$ with its expected value under the null (instead of 0). To this purpose, we need the following computation; when X is Normal distributed with zero mean and variance σ^2, we have:

$$\mathbb{E}\left[|X|^\gamma \left| X^2 > \vartheta\right.\right] = \frac{\Gamma\left(\frac{\gamma+1}{2}, \frac{c_\vartheta^2}{2}\right)}{2N(-\frac{\sqrt{\vartheta}}{\sigma})\sqrt{\pi}} \left(2\sigma^2\right)^{\frac{1}{2}\gamma}, \tag{4.22}$$

where $N(x)$ is the standard normal cumulative function and $\Gamma(\alpha, x)$ is the upper incomplete gamma function.[4] Now, we can exploit the estimate of the continuous local volatility given by the threshold itself, which is given by (4.21), replacing σ^2 by ϑ/c_ϑ^2 in (4.22). Then, we define the corrected realized threshold multipower estimator as:

$$\mathsf{C-TMPV}_\delta(X)_t^{[\gamma_1,\ldots,\gamma_M]} = \delta^{1-\frac{1}{2}(\gamma_1+\ldots+\gamma_M)} \sum_{j=M}^{[T/\delta]} \prod_{k=1}^{M} Z_{\gamma_k}(\Delta_{j-k+1}X, \vartheta_{j-k+1}) \tag{4.23}$$

where the function $Z_\gamma(x, y)$ is defined as:

$$Z_\gamma(x,y) = \begin{cases} |x|^\gamma & if\ x^2 \le y \\ \dfrac{1}{2N(-c_\vartheta)\sqrt{\pi}} \left(\dfrac{2}{c_\vartheta^2} y\right)^{\frac{\gamma}{2}} \Gamma\left(\dfrac{\gamma+1}{2}, \dfrac{c_\vartheta^2}{2}\right) & if\ x^2 > y \end{cases} \tag{4.24}$$

[4] Precisely,

$$N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$$

and

$$\Gamma(\alpha, x) = \int_x^{+\infty} s^{\alpha-1} e^{-s} ds.$$

When $\alpha = 1$, $\Gamma(1, x) = e^{-x}$.

Relevant cases which will be examined in what follows are $\gamma = 1, 2, 4/3$. In this special cases we have, with $c_\vartheta = 3$ and $x^2 > y$, $Z_1(x,y) \simeq 1.094 \cdot y^{\frac{1}{2}}$, $Z_{4/3}(x,y) \simeq 1.129 \cdot y^{\frac{2}{3}}$, and $Z_2(x,y) \simeq 1.207 \cdot y$ respectively.

For example, the corrected version of (4.18) is the corrected threshold bipower variation defined as:

$$\mathsf{C-TBPV}_\delta(X)_t = \mu_1^{-2}\mathsf{C-TMPV}_\delta(X)_t^{[1,1]} = \mu_1^{-2}\sum_{j=2}^{[T/\delta]} Z_1(\Delta X_j, \vartheta_j) Z_1(\Delta X_{j-1}, \vartheta_{j-1}) \tag{4.25}$$

While the corrected version (4.23) of threshold bipower variation is expected to be unbiased in absence of jumps, it will introduce a positive bias if jumps are present in the trajectory of X. Indeed, suppose that there is a jump ΔJ in a given interval, such that $\Delta X = \Delta X^c + \Delta J$. When detecting the jump, we approximate $(\Delta X)^\gamma$ with $\mathbb{E}\left[|X^c|^\gamma \left| (X^c)^2 > \vartheta \right.\right]$, which is much larger than $\mathbb{E}\left[|X^c|^\gamma\right]$, which is our estimation target. Thus, the correction should not be implemented when the test statistics detects a jump in the trajectory. However, as will be shown in the next section, the correction is essential for building test statistics; indeed, the correction provides unbiased estimates under the null. The test statistics which is used in our empirical analysis is based on this correction and it is defined by:

$$\mathsf{C-Tz} = \delta^{-\frac{1}{2}} \frac{\left(\mathsf{RV}_\delta\left(X\right)_T - \mathsf{C-TBPV}_\delta\left(X\right)_T\right) \cdot \mathsf{RV}_\delta\left(X\right)_T^{-1}}{\sqrt{\vartheta \, \max\left\{1, \frac{\mathsf{C-TTriPV}_\delta(X)_T}{\left(\mathsf{C-TBPV}_\delta(X)_T\right)^2}\right\}}}, \tag{4.26}$$

A correction similar to (4.23) can also be introduced for the threshold estimators of Mancini (2007). We define:

$$\mathsf{C-TPV}_\delta(X)_t^{[\gamma]} = \delta^{1-\frac{1}{2}\gamma}\sum_{j=1}^{[T/\delta]} Z_\gamma(\Delta_j X, \vartheta_j) \tag{4.27}$$

where, for large γ, we can exploit the integration by parts formula $\Gamma(\alpha+1, x) = \alpha\Gamma(\alpha, x) + x^\alpha\, e^{-x}$.

4.3 Quarticity Estimators and Jump Detection Statistics

For estimation of the integrated quarticity, the literature focused on the following quantities:

$$\mathsf{QPV}_\delta(X)_t = \mu_1^{-4} \cdot \mathsf{MPV}_\delta(X)_t^{[1,1,1,1]} = \mu_1^{-4}\frac{1}{\delta}\sum_{j=4}^{[T/\delta]} |\Delta_{j-3}X| \cdot |\Delta_{j-2}X| \cdot |\Delta_{j-1}X| \cdot |\Delta_j X| \xrightarrow{\delta\to 0} \int_t^{t+T} \sigma_s^4 ds \tag{4.28}$$

and

$$\mathsf{TriPV}_\delta(X)_t = \mu_{\frac{4}{3}}^{-3} \cdot \mathsf{MPV}_\delta(X)_t^{[\frac{4}{3},\frac{4}{3},\frac{4}{3}]} = \mu_{\frac{4}{3}}^{-3}\frac{1}{\delta}\sum_{j=3}^{[T/\delta]} |\Delta_{j-2}X|^{\frac{4}{3}} \cdot |\Delta_{j-1}X|^{\frac{4}{3}} \cdot |\Delta_j X|^{\frac{4}{3}} \xrightarrow{\delta\to 0} \int_t^{t+T} \sigma_s^4 ds \tag{4.29}$$

where $\mu_{\frac{4}{3}} \simeq 0.8309$.

Barndorff-Nielsen and Shephard (2006) develop a theory, based on multipower variation, that allows to test for the presence of jumps in a time window $[0, T]$ with the desired level of significance.

The following proposition defines the proper statistic to test for jumps.

Proposition 4.3.1 (Barndorff-Nielsen and Shephard, 2006) *Assume that:*

1. $c_t dN_t = 0$ *in model (4.3).*

2. *The volatility process is bounded away from zero.*

3. The joint process (μ, σ) is independent of the Brownian motion W driving the variable X.

then, as $\delta \to 0$, we have:

$$G = \frac{\delta^{-\frac{1}{2}} \left(\mu_1^{-2} \, \mathsf{MPV}_\delta \left(X \right)_T^{[1,1]} - \mathsf{RV}_\delta \left(X \right)_T \right)}{\sqrt{\vartheta \int_0^T \sigma_s^4 \, ds}} \to^L N(0,1) \tag{4.30}$$

and

$$H = \frac{\delta^{-\frac{1}{2}} \left(\frac{\mu_1^{-2} \, \mathsf{MPV}_\delta(X)_T^{[1,1]}}{\mathsf{RV}_\delta(X)_T} - 1 \right)}{\sqrt{\vartheta \, \frac{\int_0^T \sigma_s^4 \, ds}{\left(\int_0^T \sigma_s^2 \, ds \right)^2}}} \to^L N(0,1), \tag{4.31}$$

with $\vartheta = \frac{\pi^2}{4} + \pi - 5$.

Proposition (4.3.1) is based on the fact that $p\text{-}\lim_{\delta \to 0} \mathsf{MPV}_\delta \left(X \right)_T^{[1,1]} = \mu_1^2 \int_0^T \sigma_s^2 ds$ and thus combining this result with equation (4.7) we have that:

$$\mu_1^{-2} \, \mathsf{MPV}_\delta \left(X \right)_T^{[1,1]} - \mathsf{RV}_\delta \left(X \right)_T \to^p - \sum_{j=1}^{N_T} c_j^2 \leq 0, \tag{4.32}$$

$$\frac{\mu_1^{-2} \, \mathsf{MPV}_\delta \left(X \right)_T^{[1,1]}}{\mathsf{RV}_\delta \left(X \right)_T} - 1 \to^p - \frac{\sum_{j=1}^{N_T} c_j^2}{\int_0^T \sigma_s^2 ds + \sum_{j=1}^{N_T} c_j^2} \leq 0. \tag{4.33}$$

A feasible jump test can be now constructed using multipower estimators of the integrated quarticity given by equations (4.28)-(4.29). In their study, ABD use the test:

$$z = \delta^{-\frac{1}{2}} \frac{\left(\mathsf{RV}_\delta \left(X \right)_T - \mathsf{BPV}_\delta \left(X \right)_T \right) \cdot \mathsf{RV}_\delta \left(X \right)_T^{-1}}{\sqrt{\vartheta \, \max \left\{ 1, \frac{\mathrm{TriPV}_\delta(X)_T}{\left(\mathsf{BPV}_\delta(X)_T \right)^2} \right\}}}, \tag{4.34}$$

Monte Carlo studies of Huang and Tauchen (2005) showed that the statistics G and H in Proposition 4.3.1 have a better power if implemented in their logarithmic forms. We then define the logarithmic jump test statistics as follows:

$$z_1 = \delta^{-\frac{1}{2}} \frac{\log \mathsf{RV}_\delta \left(X \right)_T - \log \mathsf{BPV}_\delta \left(X \right)_T}{\sqrt{\vartheta \, \frac{\mathrm{QPV}_\delta(X)_T}{\left(\mathsf{BPV}_\delta(X)_T \right)^2}}}, \tag{4.35}$$

$$z_2 = \delta^{-\frac{1}{2}} \frac{\log \mathsf{RV}_\delta \left(X \right)_T - \log \mathsf{BPV}_\delta \left(X \right)_T}{\sqrt{\vartheta \, \frac{\mathrm{TriPV}_\delta(X)_T}{\left(\mathsf{BPV}_\delta(X)_T \right)^2}}}. \tag{4.36}$$

In what follows, we will use the following special cases of threshold multipower variation, which are the counterparts of (4.28) and (4.29):

$$\mathsf{TQPV}_\delta(X)_t = \mu_1^{-4} \cdot \mathsf{TMPV}_\delta(X)_t^{[1,1,1,1]} \xrightarrow{\delta \to 0} \int_t^{t+T} \sigma_s^4 ds, \tag{4.37}$$

$$\mathsf{TTriPV}_\delta(X)_t = \mu_{\frac{4}{3}}^{-3} \cdot \mathsf{TMPV}_\delta(X)_t^{[\frac{4}{3},\frac{4}{3},\frac{4}{3}]} \xrightarrow{\delta \to 0} \int_t^{t+T} \sigma_s^4 ds. \tag{4.38}$$

The theory of threshold multipower variation allows for additional tests, which are the natural counterparts of the statistics (4.35),(4.36). We then introduce the corresponding threshold z statistics, defined

as:

$$\mathsf{Tz} = \delta^{-\frac{1}{2}} \frac{\left(\mathsf{RV}_\delta\left(X\right)_T - \mathsf{TBPV}_\delta\left(X\right)_T\right) \cdot \mathsf{RV}_\delta\left(X\right)_T^{-1}}{\sqrt{\vartheta \max\left\{1, \frac{\mathrm{TTriPV}_\delta(X)_T}{\left(\mathsf{TBPV}_\delta(X)_T\right)^2}\right\}}}, \tag{4.39}$$

$$\mathsf{Tz}_1 = \delta^{-\frac{1}{2}} \frac{\log \mathsf{RV}_\delta\left(X\right)_T - \log \mathsf{TBPV}_\delta\left(X\right)_T}{\sqrt{\vartheta \frac{\mathrm{TQPV}_\delta(X)_T}{\left(\mathsf{TBPV}_\delta(X)_T\right)^2}}}. \tag{4.40}$$

$$\mathsf{Tz}_2 = \delta^{-\frac{1}{2}} \frac{\log \mathsf{RV}_\delta\left(X\right)_T - \log \mathsf{TBPV}_\delta\left(X\right)_T}{\sqrt{\vartheta \frac{\mathrm{TTriPV}_\delta(X)_T}{\left(\mathsf{TBPV}_\delta(X)_T\right)^2}}}. \tag{4.41}$$

and their corrected versions, which performs well under the null of no jumps, as discussed in Section 4.2.4:

$$\mathsf{C}-\mathsf{Tz}_1 = \delta^{-\frac{1}{2}} \frac{\log \mathsf{RV}_\delta\left(X\right)_T - \log \mathsf{C} - \mathsf{TBPV}_\delta\left(X\right)_T}{\sqrt{\vartheta \frac{\mathsf{C}-\mathrm{TQPV}_\delta(X)_T}{\left(\mathsf{C}-\mathsf{TBPV}_\delta(X)_T\right)^2}}}, \tag{4.42}$$

$$\mathsf{C}-\mathsf{Tz}_2 = \delta^{-\frac{1}{2}} \frac{\log \mathsf{RV}_\delta\left(X\right)_T - \log \mathsf{C} - \mathsf{TBPV}_\delta\left(X\right)_T}{\sqrt{\vartheta \frac{\mathsf{C}-\mathrm{TTriPV}_\delta(X)_T}{\left(\mathsf{C}-\mathsf{TBPV}_\delta(X)_T\right)^2}}}. \tag{4.43}$$

4.4 Simulation Study

To assess the small sample properties of the concurrent estimators we use Monte Carlo simulations of realistic stochastic processes which have been extensively used to model stock index prices. The purpose of this section is to show that bipower variation is a biased estimator of integrated volatility in the presence of jumps, while threshold-based estimator, both power and multipower, are much less sensitive to jumps and accordingly less biased. Moreover, we show that while threshold power variation is particularly sensitive to the choice of the threshold, threshold bipower variation is instead largely robust to this choice. This latter feature is particularly important, since it suggests that results obtained in empirical applications using threshold multipower variation are not too sensitive to the threshold employed.

We simulate two models. The first model, labelled as SV1FJ, is a one-factor jump-diffusion model with stochastic volatility, described by the coupled stochastic differential equations:

$$\textbf{SV1FJ} \qquad \begin{aligned} dX_t &= \mu\, dt + \sqrt{v_t} dW_{x,t} + dJ_t, \\ d\log v_t &= (\alpha - \beta \log v_t)\, dt + \eta dW_{v,t}, \end{aligned} \tag{4.44}$$

where W_x and W_v are standard Brownian motions with $\mathrm{corr}\left(dW_y, dW_v\right) = \rho$, v_t is a stochastic volatility factor, dJ is a compound Poisson process with constant activity λ and random jump size which is normally distributed with zero mean and standard deviation σ_J. We use the model parameters estimated by Andersen et al. (2002) on S&P500 prices and reported in Table 4.1. Similar estimates have been obtained by Bates (2000); Pan (2002); Chernov et al. (2003). We label as SV1F the continuous model obtained from (4.44) imposing $dJ_t = 0$, for all t. In the top of Figure 4.3 a SV1F continuous trajectory for the normalized price $P_t/P_0 = \exp(X_t)/\exp(X_0)$ is shown. The price is simulated with a time step of 1 sec and sampled every 60 sec. The middle graph depicts the corresponding one-minute percentage returns (as black line) and the computed threshold (as red line). All observations that exceed the threshold in absolute value are detected as jumps. As discussed in the previous sections, this procedure unavoidably produces a fake

jump component J_t^{fake} which can be interpreted as noise introduced with the threshold. The autocorrelation function[5] of the fake jump component (in absolute value) appears in the bottom of the mentioned figure together with the 95% confidence levels for an i.i.d. process. As expected the estimated fake jump component does not show persistence and the detected jump component is Markovian.

The second model is a continuous model introduced by Chernov et al. (2003). We label it SV2F, as it uses two stochastic volatility factors. The second volatility factor has a feedback in the volatility function that may produce abrupt changes in the price variable.

The model can be expressed as:

$$\textbf{SV2F} \quad \begin{aligned} dX_t &= \mu\, dt + \exp\left[\beta_0 + \beta_1\, v_{1,t} + \beta_2\, v_{2,t}\right] dW_{x,t}, \\ dv_{1,t} &= \alpha_{v_1}\, v_{1,t}\, dt + dW_{v_1,t}, \\ dv_{2,t} &= \alpha_{v_2}\, v_{2,t}\, dt + \left[1 + \beta_{v_2}\, v_{2,t}\right]\, dW_{v_2,t}. \end{aligned} \tag{4.45}$$

where the correlation of the Brownian motions are $corr(dW_x, dW_{v_1}) = \rho_1$ and $corr(dW_x, dW_{v_2}) = \rho_2$. Table 4.1 reports model parameters of equation (4.45), which are those used by Huang and Tauchen (2005). The first volatility factor $v_{1,t}$ is a slow mean-reverting factor, while v_2 is a fast mean-reverting variable, since $|\alpha_{v_1}| \ll |\alpha_{v_2}|$. We use this model since this is a challenging one for the z-test; indeed, very large price movements can be misleadingly detected as jumps, even if SV2F is a continuous model. Top panel of Figure (4.4) shows a realization of a SV2F price path. It is evident that SV2F price displays sudden variations. Middle panel of Figure (4.4) reports in black the corresponding one-minute percentage returns together with the estimated threshold as red line. As for SV1F we obtain a fake jump component, the SV2F being a continuous model. However the autocorrelation of the detected SV2F jump component (in absolute value), reported in bottom panel, shows that SV2F jumps are not Markovian. A persistence lasting for about 1.5 days is detected. This is the reason why SV2F model is incorrect for modelling asset prices with uncorrelated jumps.

SV1FJ	
Parameter	Value
μ	0.0304
α	-0.0120
β	0.0145
η	0.1153
ρ	-0.6127
λ	0.0137
σ_J	1.51

SV2F	
Parameter	Value
μ	0.030
β_0	-1.2
β_1	0.04
β_2	1.5
ρ_1	-0.3
ρ_2	-0.3
α_{v_1}	-0.00137
α_{v_2}	-1.386
β_{v_2}	0.250

Table 4.1: Parameters of models (4.44)-(4.45), expressed in percentage form and on daily basis.

[5]We report only lags greater or equal than one, the zero-lag autocorrelation being trivially one.

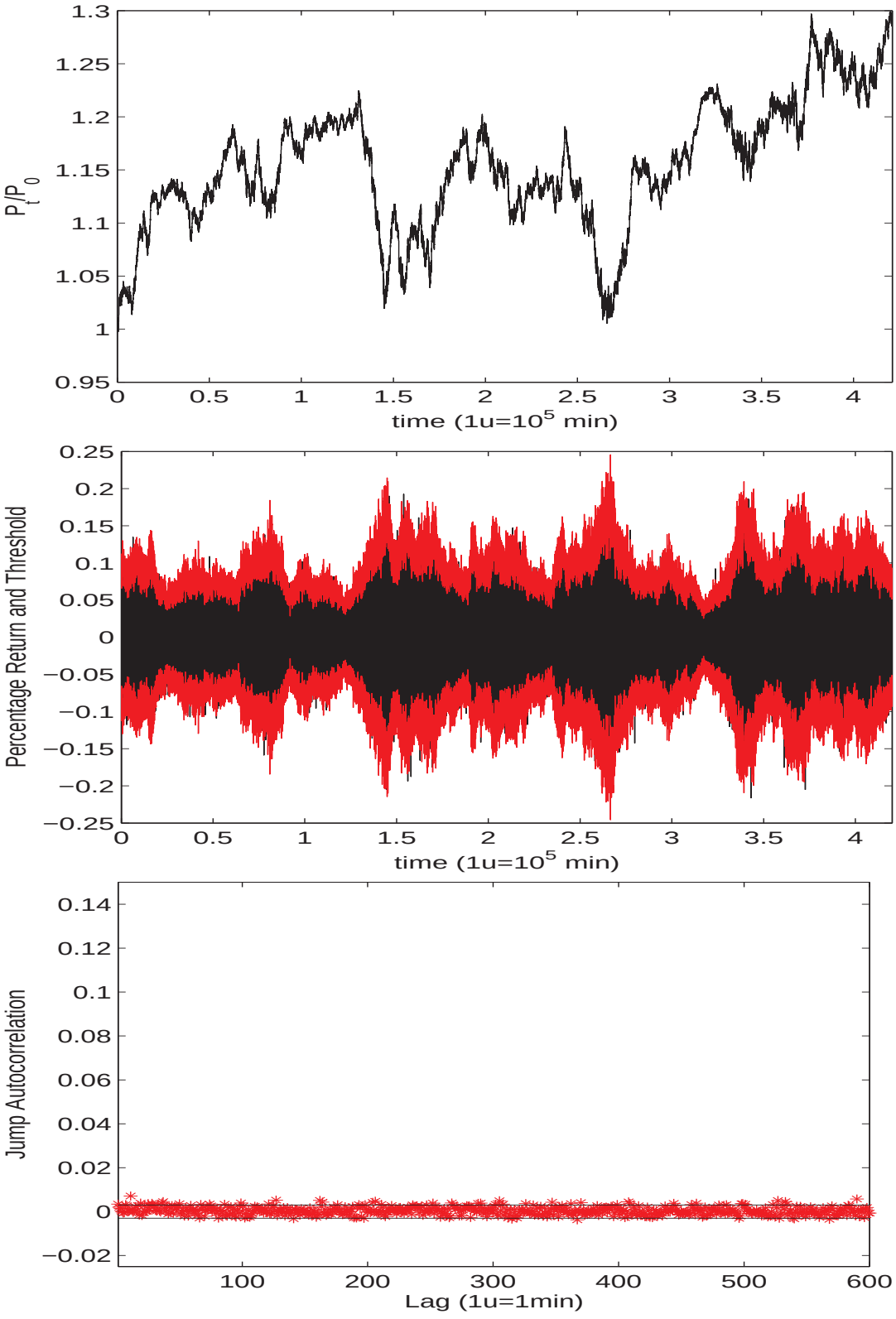

Figure 4.3: From the top to the bottom. (Top) Simulated price path for continuous model (4.44) where the jump component has been suppressed. The price is simulated with a time step of 1 sec and sampled every minute. (Middle) One-minute log returns (black line) of the simulated price path and estimated threshold (red line) for $c_\vartheta = 3$. (Bottom) Autocorrelation coefficients (red stars) for lags greater or equal than one of the absolute value of estimated (fake) jump component for SV1F and 95% i.i.d. process confidence levels (black lines).

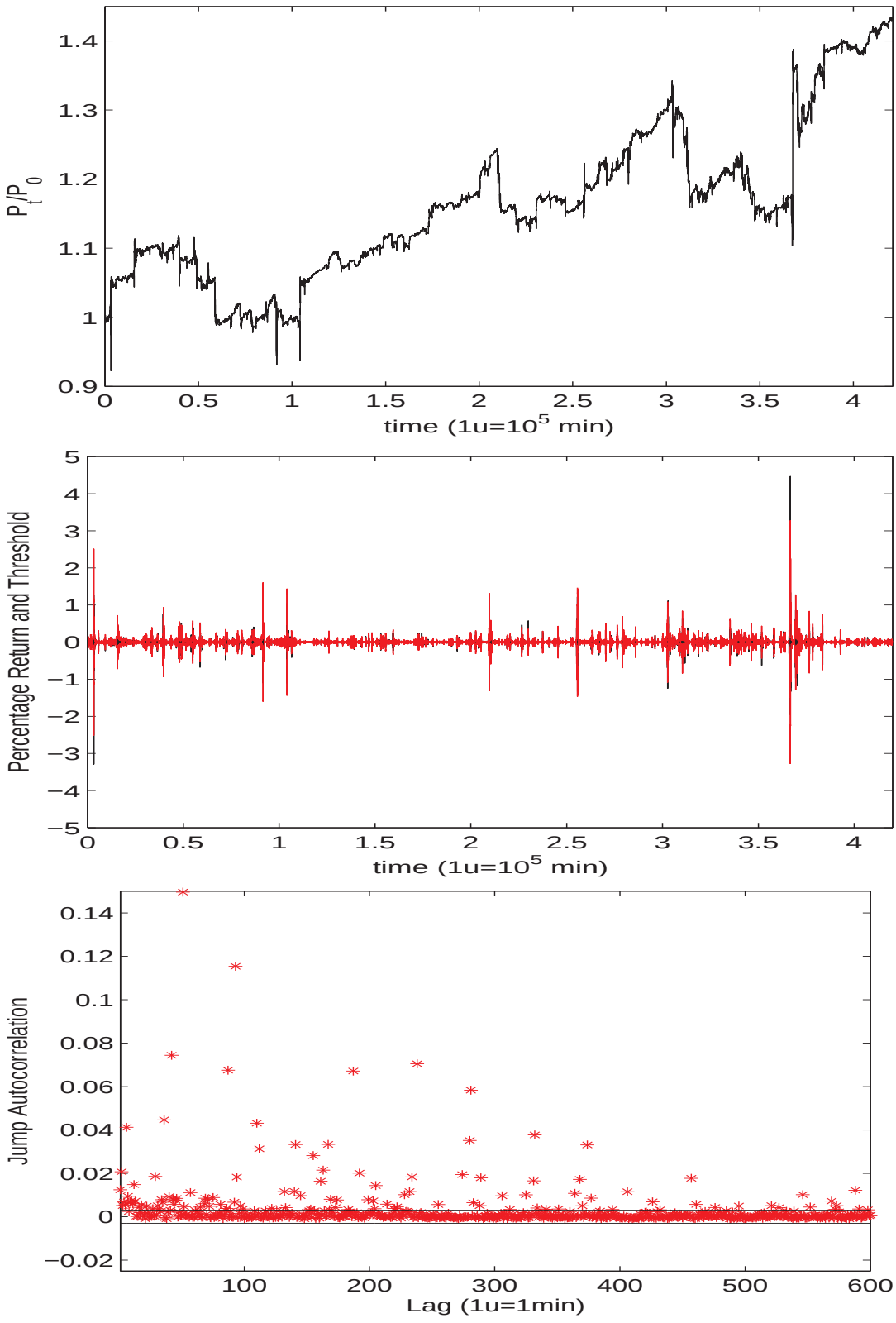

Figure 4.4: From the top to the bottom. (Top) Simulated price path for continuous model (4.45). The price is simulated with a time step of 1 sec and sampled every minute. (Middle) One-minute log returns (black line) of the simulated price path and estimated threshold (red line) for $c_\vartheta = 3$. (Bottom) Autocorrelation coefficients (red stars) for lags greater or equal than one of the absolute value of the estimated (fake) jump component for SV1F and 95% i.i.d. process confidence levels (black lines).

Each day, we simulate $7 \cdot 60 \cdot 60$ steps corresponding to seven hours. When the sampling is performed with $\delta = 5\,\text{min}$ we obtain 84 returns per day, while we get 420 returns per day when $\delta = 1\,\text{min}$.

The numerical integration of models (4.44) and (4.45) is performed with the Euler scheme, using a discretization step of $\Delta = 1$ second. Each day, we simulate $7 \cdot 60 \cdot 60$ steps corresponding to seven hours. We then use $\delta = 5$ minutes, that is 84 returns per day.

The Monte Carlo experiments are devised to compare the efficiency of the proposed estimators in estimating the integral of σ^2 and σ^4. Since we are keenly interested in studying the performance of competing estimators in presence of jumps, we generate different samples (with $1,000$ "daily" replications each) in the following way. In the first sample, we do not generate jumps at all. In the second sample, we generate a single jump for each day. In the third sample, we generate exactly two jumps per day. In the fourth sample, we generate two jumps per day and we force them to be consecutive (i.e., the second jump is forced to occur 300 seconds after the first). This allows us to compute the expected value conditioned to the presence of zero, one, two jumps, consecutive or not. For every simulated daily trajectory, we compute the estimates of BPV and their fourth-power counterparts QPV,TriPV as well as threshold estimators TRV, TQV, TTriPV and threshold multipower estimators TBPV, TQPV,TTriPV. We compute daily percentage estimation error and compute averages and standard deviations across the sample. On each day, we compute the percentage error in estimating the quantity of interest, and we average this error and compute its standard deviation across replications. All results are reported in Table 4.2.

Results are compelling. Bipower variation (as realized volatility) does a good job in estimating integrated squared volatility in case of no jumps. However, it is significantly biased if there is a jump in the trajectory (+45.71%) and largely biased (+96.30%) if there are two jumps in the trajectory. If the two jumps are consecutive, the bias is huge (+567.48%). The bias of multipower variation in estimating integrated quarticity is even larger. The QPV and $TriPV$ estimators show an average bias of +80.01% and +185.84%, respectively, in presence of one jump per day.

Threshold-based estimators, instead, are much more robust to the presence of jumps. The bias of threshold power variation of Mancini (2007) in estimating integrated squared volatility is around −6% in absence of jumps and around −7% in presence of one and two jumps, consecutive or not. The same happens when estimating quarticity, the bias being around −15%. The presence of a negative bias is due to the fact that, by their proper definition, threshold estimators remove completely observations larger than the threshold. When we correct for this as indicated in Section 4.2.4, the bias turns out to be positive since, when an observation is above the threshold, we replace its power with its expected value under the assumption that the observation was actually above threshold; which is true under the null of no jumps, but needs not to be true in presence of an actual jump, thus inducing a positive bias.

The estimators based on threshold multipower variation, introduced in this study, yield equally good results. Threshold bipower variation has a bias of −5.75% in the case of no jumps, of −7.29% with a single jump, of −8.84% in the case of two jumps, and of −8.51% in the case of two consecutive jumps. When estimating quarticity, these biases range between −15% and −16% according to the number of jumps and the estimator used. Again, the corrected versions largely correct the bias under the null of no jumps, but

turn the negative bias in a positive one in the case of jumps. However, from our simulated experiment we can conclude that threshold-based estimators perform much better than multipower variation in presence of jumps.

The strength of the threshold estimators is shown by the histograms of the relative errors $\frac{BPV-\int_0^T \sigma_t^2 dt}{\int_0^T \sigma_t^2 dt}$, $\frac{QPV-\int_0^T \sigma_t^4 dt}{\int_0^T \sigma_t^4 dt}$ and $\frac{TriPV-\int_0^T \sigma_t^4 dt}{\int_0^T \sigma_t^4 dt}$, $T = 1\ day$, reported in Figure 4.5 for the ODOJ case. In the threshold and corrected-threshold case the histograms profile are very thin w.r.t. the classical estimators. Moreover no particular differences are revealed between multipower estimators and their threshold counterpart for the SV1F as shown by the histograms in Figures 4.6, confirming that our proposed estimators are well-behaved under the null.

Quantity	Estimator	Relative bias (%)			
		no jumps	one jump	two jumps	two consecutive jumps
$\int \sigma_s^2 ds$	BPV	-2.18 (0.53)	45.71 (1.68)	96.30 (3.10)	567.48 (21.44)
	TRV	-5.58 (0.50)	-5.91 (0.53)	-6.88 (0.52)	-6.96 (0.53)
	C − TRV	-1.39 (0.46)	9.38 (0.55)	18.89 (0.61)	18.89 (0.62)
	TBPV	-5.75 (0.56)	-7.29 (0.59)	-8.84 (0.57)	-8.51 (0.57)
	C − TBPV	-2.26 (0.53)	4.52 (0.60)	10.94 (0.64)	20.11 (0.71)
$\int \sigma_s^4 ds$	QPV	-3.05 (1.04)	80.01 (4.54)	198.41 (10.50)	1553.70 (85.49)
	TQV	-11.08 (1.24)	-15.54 (0.98)	-16.42 (0.98)	-16.38 (1.01)
	C − TQV	-5.18 (1.28)	38.34 (1.54)	76.90 (2.01)	76.17 (2.10)
	TQPV	-16.37 (0.91)	-13.34 (1.28)	-16.48 (1.22)	-15.00 (1.25)
	C − TQPV	-4.08 (1.01)	7.56 (1.46)	22.05 (1.81)	47.63 (2.49)
	TriPV	-3.99 (1.21)	185.84 (11.23)	493.81 (43.49)	7875.79 (461.73)
	TTriPV	-10.51 (1.18)	-11.99 (1.22)	-14.34 (1.19)	-13.73 (1.21)
	C − TTriPV	-4.23 (1.21)	11.01 (1.46)	27.76 (1.87)	71.32 (2.86)

Table 4.2: Reports the mean error in percentage form in estimating $\int \sigma_s^\gamma ds$, with $\gamma = 2, 4$, using the corresponding estimator, in the case of no jumps, one jump, two jumps and two consecutive jumps when simulating model (4.44). In parenthesis, the standard error of the mean is reported.

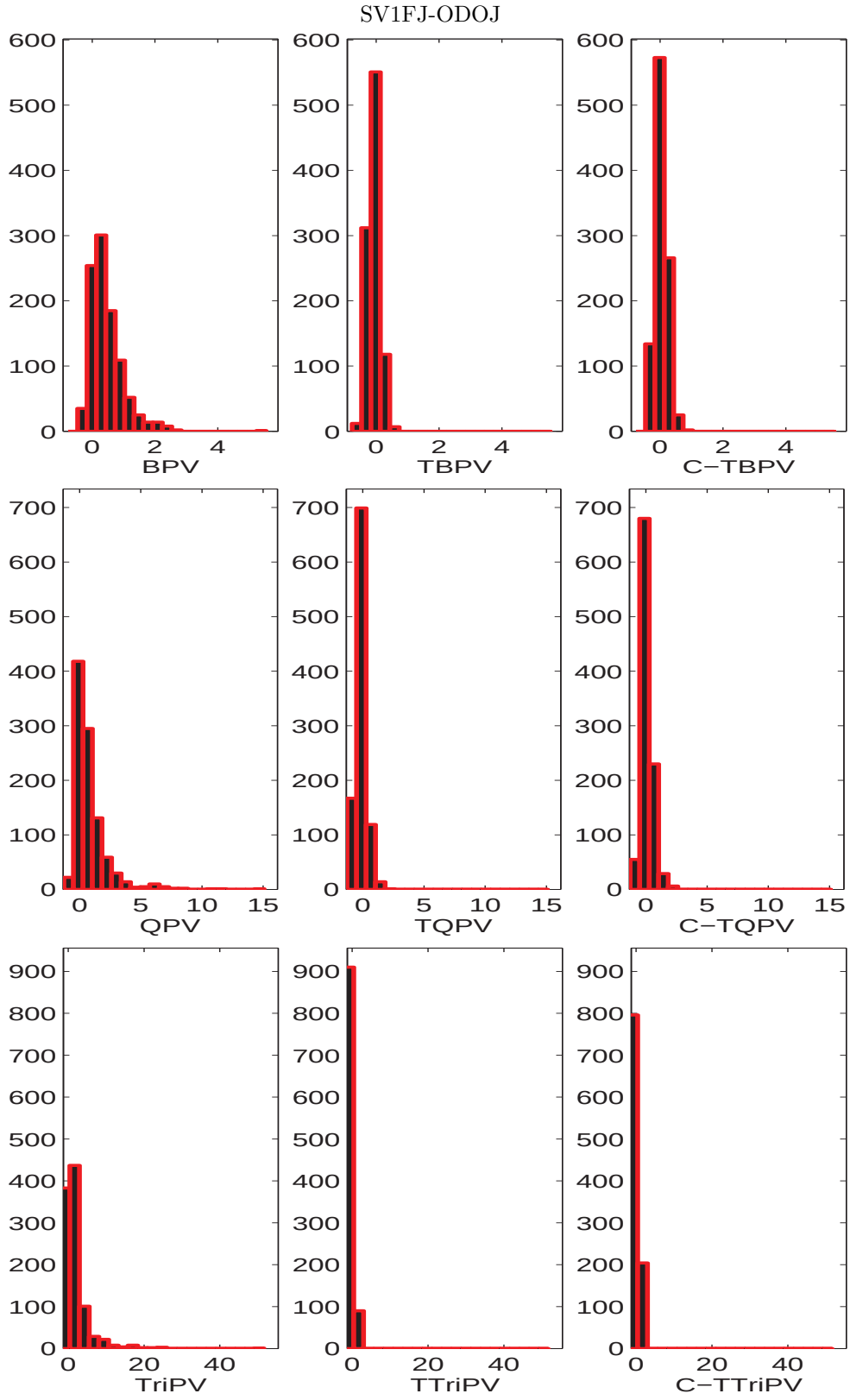

Figure 4.5: (Up) Histograms for the SV1FJ-ODOJ model of the relative error $\frac{BPV-\int_0^T \sigma_t^2 dt}{\int_0^T \sigma_t^2 dt}$, $T = 1\ day$, and its threshold versions. (Middle) Histograms for the SV1FJ-ODOJ model of the relative errors $\frac{QPV-\int_0^T \sigma_t^4 dt}{\int_0^T \sigma_t^4 dt}$ and (down) $\frac{TriPV-\int_0^T \sigma_t^4 dt}{\int_0^T \sigma_t^4 dt}$, as well as their threshold versions.

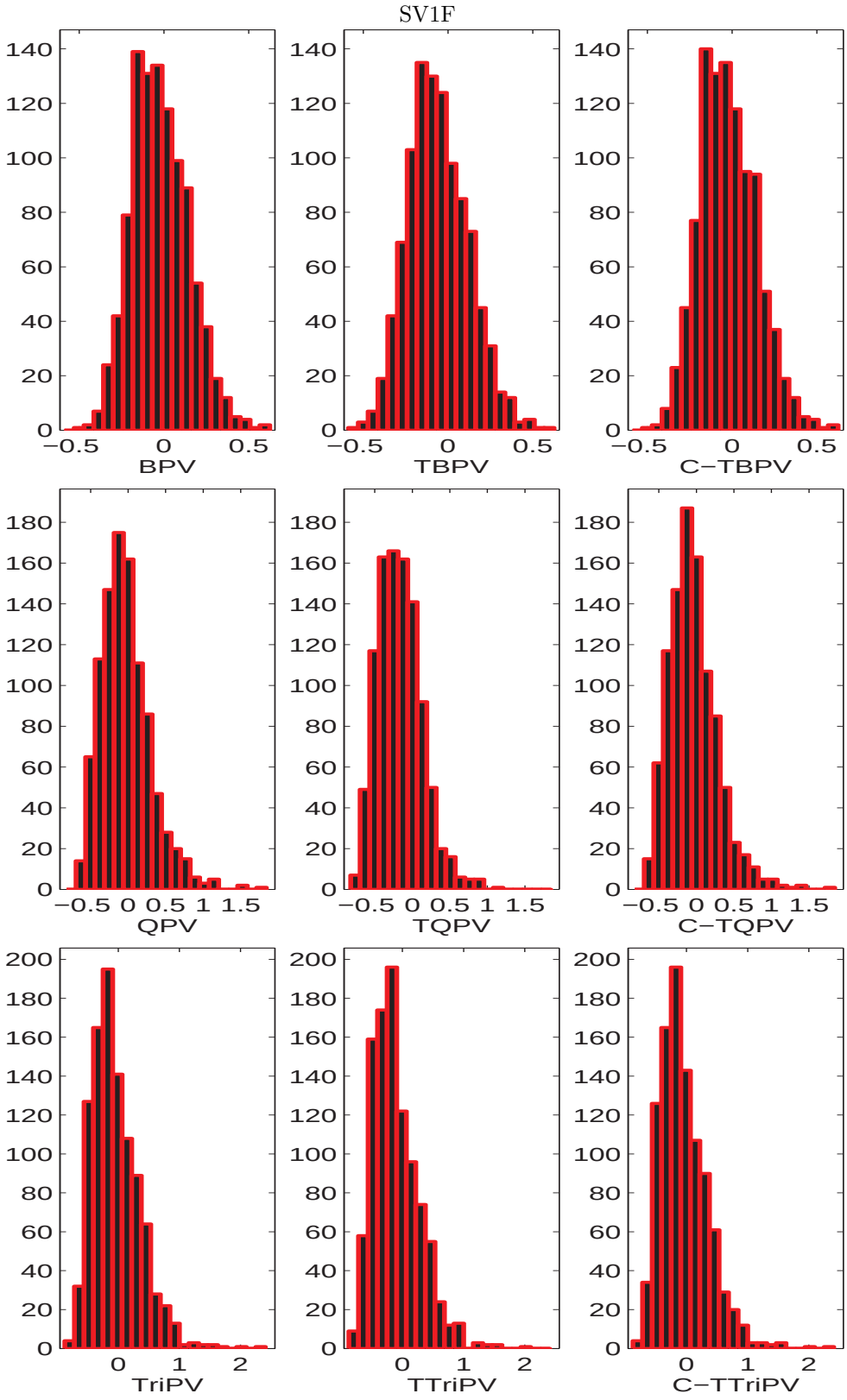

Figure 4.6: (Up) Histograms for the SV1F model of the relative error $\frac{BPV-\int_0^T \sigma_t^2 dt}{\int_0^T \sigma_t^2 dt}$, $T = 1$ *day*, and its threshold versions. (Middle) Histograms for the SV1FJ model of the relative errors $\frac{QPV-\int_0^T \sigma_t^4 dt}{\int_0^T \sigma_t^4 dt}$ and (down) $\frac{TriPV-\int_0^T \sigma_t^4 dt}{\int_0^T \sigma_t^4 dt}$, as well as their threshold versions.

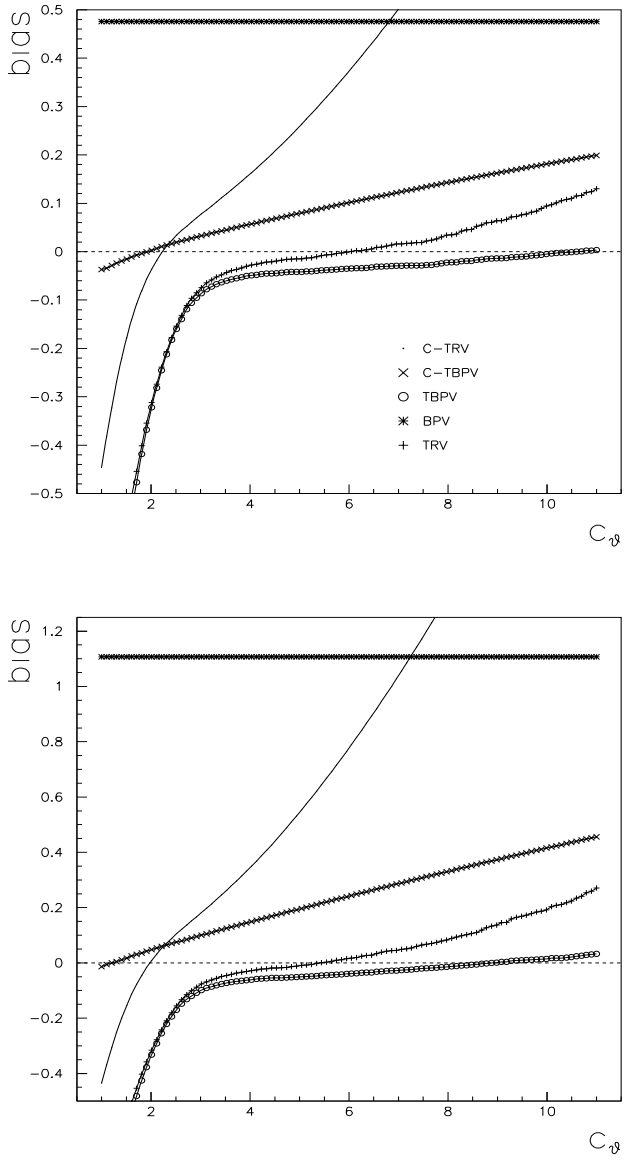

Figure 4.7: Bias of the different estimators of $\int \sigma_s^2 ds$ in the presence of a single jump (top) and two jumps (bottom), as a function of the threshold parameter c_ϑ.

Threshold estimators deliver more accurate estimates in the presence of jumps at the cost of introducing this extra-parameter (the results presented above are based on the threshold (4.21) with $c_\vartheta = 3$). Between the two competing threshold estimators (power and multipower), our simulation experiments highlight a substantial advantage in using threshold multipower variation instead of threshold power variation. This conclusion stems from considering the robustness of jump detection and volatility measuring with respect to the parameter c_ϑ. We use simulated experiments to check the robustness of threshold estimators with respect to this parameter, in the case of a single jump and in the case of two jumps. Results are shown in Figure 4.7.

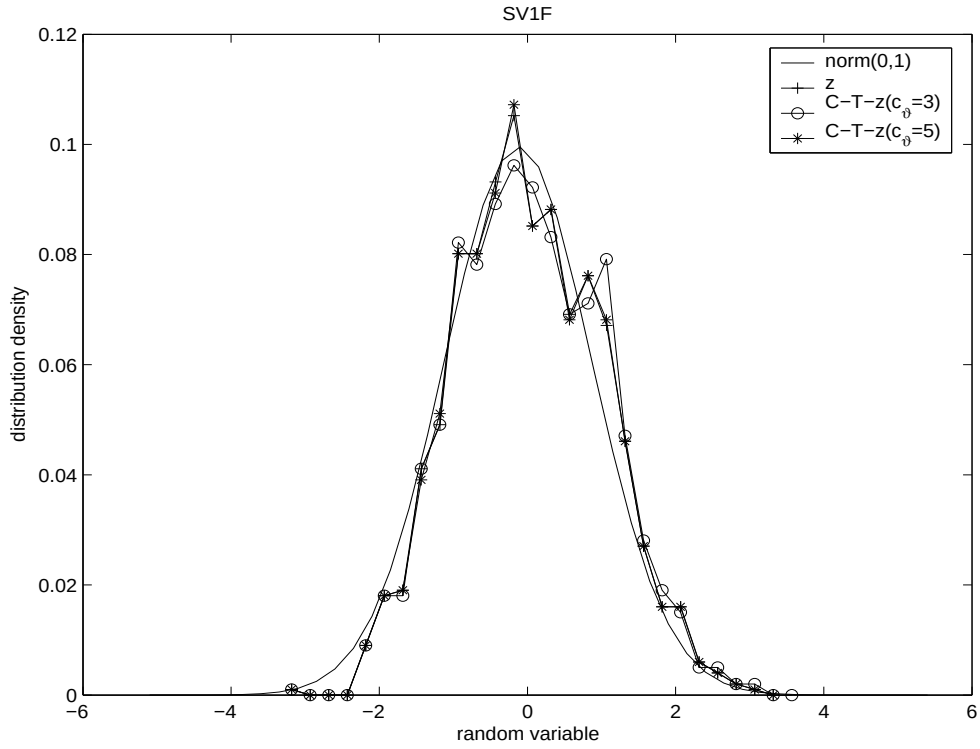

Figure 4.8: Distribution of the jump statistics z and $\mathsf{C}-\mathsf{Tz}$ (with $c_\vartheta = 3, 5$) on 1,000 replications of model (4.44) with $cdN = 0$.

Bipower variation does not depend on the value of the threshold but it is largely biased, especially with two jumps in the daily trajectory. Threshold estimators are less biased, however we can see that threshold bipower variation is less sensitive to the choice of the threshold than threshold power variation. This is basically due to the fact that, even if both TBPV and TRV converge to $[X^c]$ as $\delta \to 0$, for fixed δ we have $\mathsf{TBPV}_\delta \overset{c_\vartheta \to \infty}{\longrightarrow} \mathsf{BPV}_\delta$ while $\mathsf{TRV}_\delta \overset{c_\vartheta \to \infty}{\longrightarrow} \mathsf{RV}_\delta$.

We also use Monte Carlo experiments to evaluate the efficiency in detecting jumps with the z statistics (4.34), constructed with multipower variation methods, the Tz statistics (4.39), constructed with threshold multipower variation and the $\mathsf{C}-\mathsf{Tz}$ statistics (4.26), corrected as explained in Section 4.2.4, as well as with the corresponding logarithmic test statistics defined in Section 4.3.

Results with different confidence levels are reported in Table 4.3 in the case of no jumps, a single jump, and two consecutive jumps, while Figure 4.8 shows the distribution of selected test statistics under the null. The same analysis could be obtained analyzing Figure 4.9 which plots the specified tests under the null of continuous trajectories. Black lines are 99.99% confidence levels. As we simulate $1,000$ daily replication we expect less than one observation lies outside the confidence bounds. While Tz tests display more outliers than expected the $\mathsf{C}-\mathsf{Tz}$ tests resolve this issue reducing the noise to a level comparable with the classical z test.

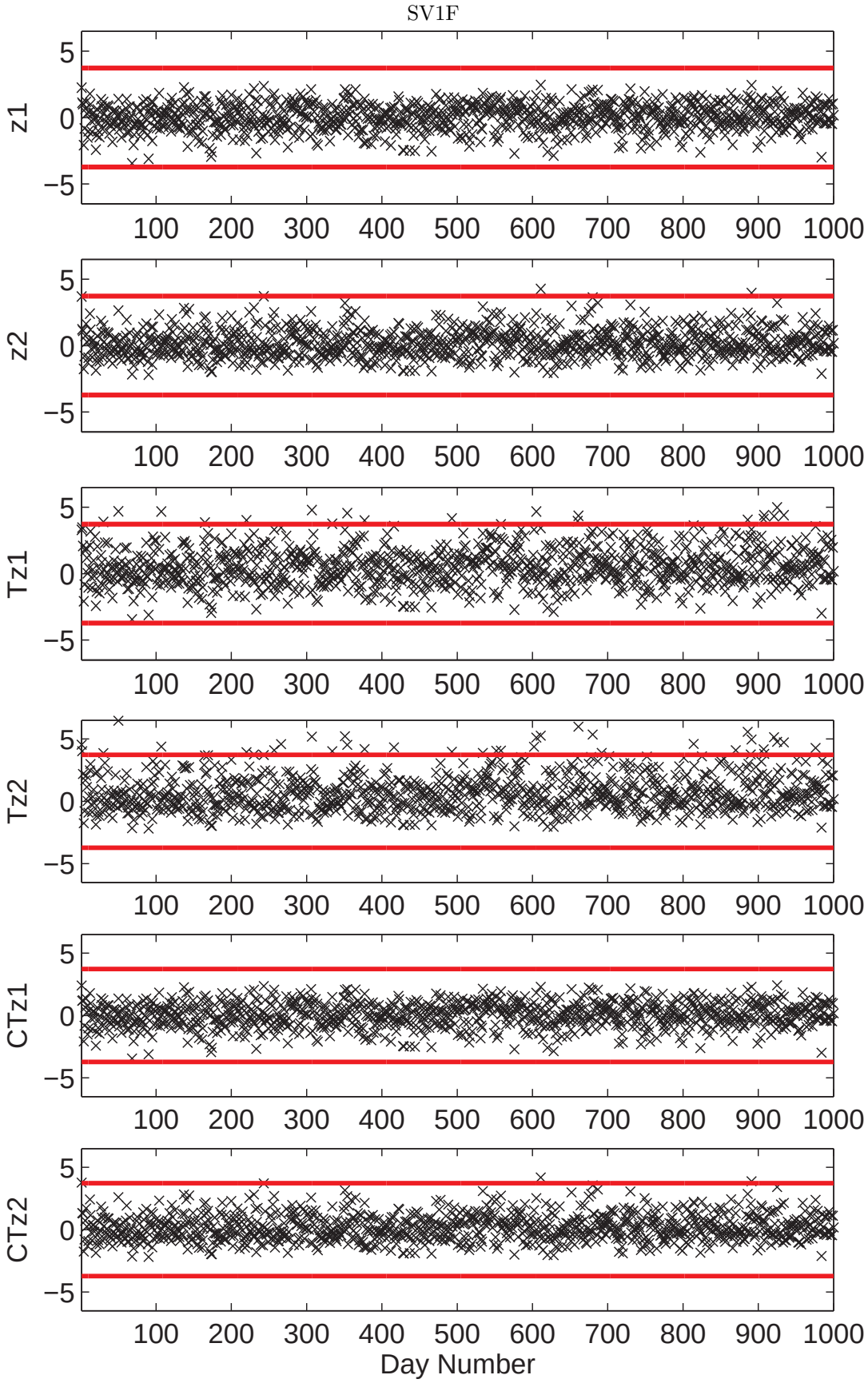

Figure 4.9: z-statistics for the SV1F model. The thick red lines are the ±99.99% confidence levels.

	Panel A				PanelB				PanelC			
	No jumps				Single jump				Two consecutive jumps			
	50%	95%	99%	99.99%	50%	95%	99%	99.99%	50%	95%	99%	99.99%
z	51.90	6.10	1.30	0.00	93.70	83.70	78.90	69.70	96.90	76.60	61.60	39.10
z_1	51.90	3.80	0.30	0.00	93.70	84.50	81.50	74.80	96.90	86.80	82.70	73.40
z_2	51.90	8.70	3.40	0.20	93.70	84.40	80.60	74.00	96.90	78.40	71.00	52.90
Tz	62.60	18.60	9.00	1.00	95.60	90.50	88.00	81.40	99.30	98.70	98.10	95.40
Tz_1	62.60	20.60	11.00	1.90	95.60	91.00	89.50	84.50	99.30	98.70	98.50	96.70
Tz_2	62.60	21.50	12.70	3.60	95.60	91.10	89.30	84.50	99.30	98.70	98.50	96.70
CTz	52.20	6.10	1.50	0.00	94.20	86.20	83.00	75.70	99.10	95.80	94.50	91.20
CTz_1	52.20	4.40	0.30	0.00	94.20	87.30	84.60	79.10	99.10	96.60	95.50	93.40
CTz_2	52.20	9.00	3.50	0.30	94.20	86.90	85.00	78.70	99.10	96.20	95.20	93.10

Table 4.3: Percentage of detected jumps in the case of trajectories with no jumps (Panel A), a single jump per day (Panel B), and two consecutive jumps per day (Panel C), for different significance levels. The $\mathsf{C}-\mathsf{Tz}$ statistics are computed with $c_\vartheta = 3$.

The efficiency of detecting jumps with the Tz statistics is larger than that obtained with z statistics; however also the noise is larger and well above the values expected by the significance level required. This comes from the fact that, under the null of no jumps that we are testing, it is better to use $\mathsf{C}-\mathsf{Tz}$ statistics. Indeed, Table 4.3 shows that the $\mathsf{C}-\mathsf{Tz}$ (with $c_\vartheta = 3$) statistics yield a noise which is almost identical to that of the z statistics still preserving a detection efficiency larger than that of the z statistics. Figure 4.8 shows the distribution of the statistics under the null of no jumps, and we can see that the behavior on that sample of the z and the $\mathsf{C}-\mathsf{Tz}$ statistics, with both $c_\vartheta = 3$ and $c_\vartheta = 5$ is the same. With higher confidence level, the advantage in using $\mathsf{C}-\mathsf{Tz}$ statistics increases both in efficiency and noise. The advantage in using the $\mathsf{C}-\mathsf{Tz}$ statistics is very large if the jumps in the simulated trajectory are consecutive. In this case, the efficiency of the z test at the 99.99% confidence level is just 42.4%, while the corresponding efficiency of the $\mathsf{C}-\mathsf{Tz}$ test is 93.1%. We conclude that, on our simulations, there is a clear advantage in using the $\mathsf{C}-\mathsf{Tz}$ statistics. The z-statistics for the SV1FJ-ODOJ model are plotted in Figure 4.10. The thick red lines are the $\pm 99.99\%$ confidence levels. The same plots are reported in Figures 4.11 for the ODTCJ case. Inspection of Figures 4.10-4.11confirms that the threshold versions of the z-test are more efficient in detecting jumps having a less number of observations in the confidence bounds, especially for the model with two consecutive jumps.

The fact that the z statistics perform very well on continuous trajectories is not surprising, and is in line with the results of Huang and Tauchen (2005); what is surprising is the high power of z-tests in detecting jumps, since the z statistics is based on quantities which are poorly estimated, especially the quarticity. The motivation for this result is that the bias is larger with large jumps; but when the jumps are very large, they are also easier to detect. Thus, even if corrected threshold multipower estimators provide much less biased estimates of integrals of power volatility, the advantage in using $\mathsf{C}-\mathsf{Tz}$ statistics with respect to the corresponding z statistics is not as large as one would expect from Table 4.2. This reasoning is instead completely reversed if the jumps are consecutive; in this case, the quantities used to compute the z statistics are hugely biased, see Table 4.2, and this reduces its efficiency dramatically.

Again, we can check the sensitivity of $\mathsf{C}-\mathsf{Tz}$ tests with respect to the choice of the parameter c_ϑ. Figure 4.12 shows a direct comparison between threshold estimators in the case of a single jump in every trajectory. We observe that the $\mathsf{C}-\mathsf{Tz}$ test has more power than the z test and it is reasonably robust to the choice of c_ϑ.

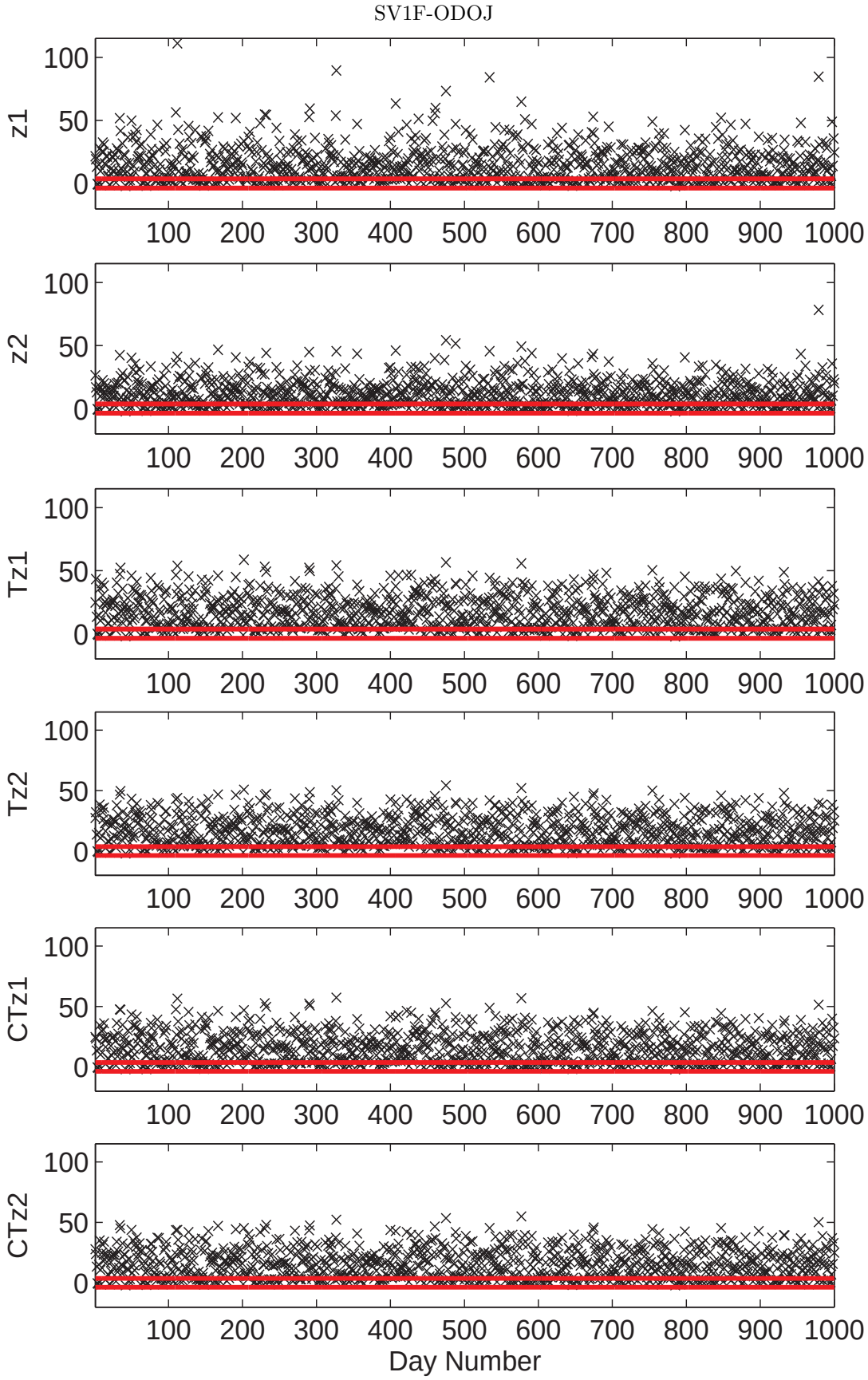

Figure 4.10: z-statistics for the SV1F-ODOJ model. The thick red line are the ±99.99% confidence level.

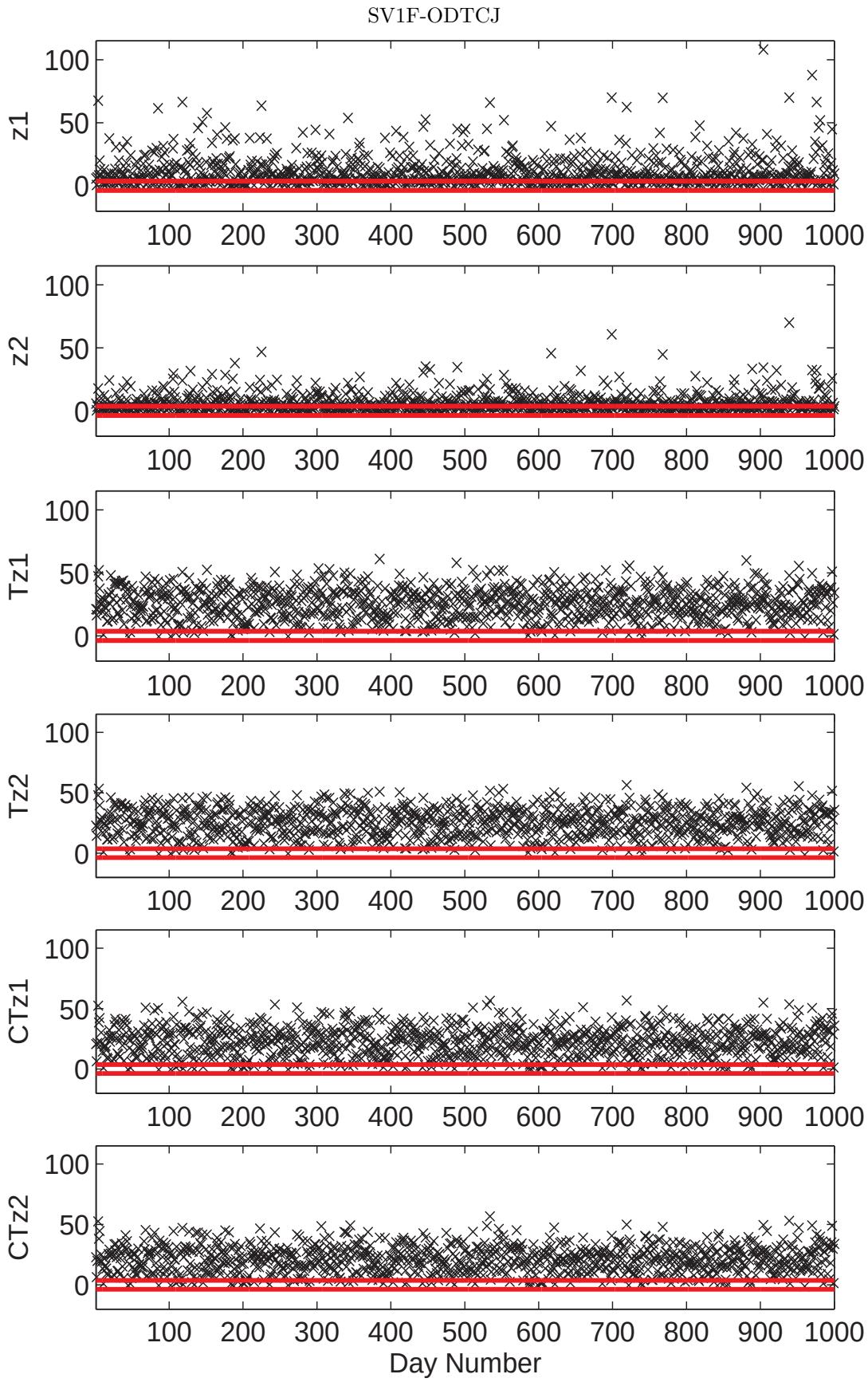

Figure 4.11: z-statistics for the SV1F-ODTCJ model. The thick red line are the ±99.99% confidence level.

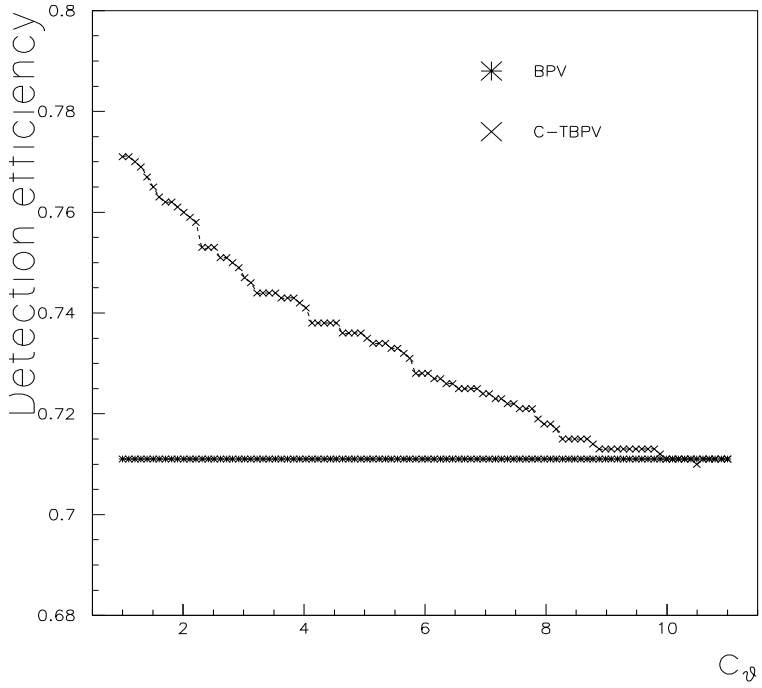

Figure 4.12: Jump detection efficiency for the model (4.44) in the presence of a single jump, as a function of the threshold parameter c_ϑ.

We finally study the behavior of power and multi-power variations when the underlying model is SV2F. As anticipated we expect that, in this case, all the threshold estimators (non corrected) have a strong negative bias. This is confirmed by inspection of Table 4.4 and the motivation is very simple: the threshold identifies as jumps those large price variations that, in fact, are continuous price variations (the SV2F model being a fully continuous model). However this bias is resolved by the corrected version of the threshold multipower.

When the threshold makes a false identification the correction does not delete the correspondent log-return but, substantially, replaces it with its expected value $\mathbb{E}\left[|X|^{\gamma}\left|X^2 > \vartheta\right.\right]$. This provokes a drastic reduction of the bias. For TBPV the bias reduces from -20.59% to -5.65, with similar behaviours in the other estimators.

Table 4.5 and Figure 4.13 show that Tz jump tests are very noisy for the SV2F case (26.90% for Tz_1 and 29.40% for Tz_2 false identifications at 99.99% confidence level) while this noise reduces strongly for the corrected C − Tz tests.

Summarizing our result on simulated experiments, we conclude that:

1. When measuring integrated power volatility in the presence of jumps, multipower variation is largely upward biased, while threshold-based estimator are slightly downward biased.

2. When measuring integrated power volatility, threshold multipower estimators are nearly insensitive to the choice of the threshold for $c_\vartheta \geq 3$ while threshold power estimators are more sensitive to this choice.

3. When testing for the presence of jumps, corrected threshold multipower estimators yield a significant advantage with respect to multipower variation based tests.

4. Corrected multipower variations and jump tests constructed from it well perform even in presence of a "tricky" model like SV2F.

These conclusions will be enforced by results in the next section. Thus, in our empirical analysis we will use threshold bipower variation as an estimator of integrated volatility, and C − Tz statistics as jump detector.

Quantity	Estimator	Relative bias for SV2F (%)
	BPV	-1.23 (0.76)
	TRV	-19.20 (0.72)
$\int \sigma_s^2 ds$	C − TRV	-5.56 (0.63)
	TBPV	-20.39 (0.80)
	C − TBPV	-5.28 (0.71)
	QPV	-19.95 (1.85)
	TQV	-51.45 (2.78)
	C − TQV	-29.38 (1.26)
$\int \sigma_s^4 ds$	TQPV	-48.50 (1.41)
	C − TQPV	-26.95 (1.60)
	TriPV	-13.83 (2.04)
	TTriPV	-47.15 (1.38)
	C − TTriPV	-24.27 (1.60)

Table 4.4: Reports the mean error in percentage form in estimating $\int \sigma_s^\gamma ds$, with $\gamma = 2, 4$, using the corresponding estimator, in the case of no jumps, one jump, two jumps and two consecutive jumps when simulating model (4.44). In parenthesis, the standard error of the mean is reported.

	SV2F			
	50%	95%	99%	99.99%
z	54.40	10.10	4.70	0.80
z_1	54.40	3.60	0.40	0.00
z_2	54.40	11.90	6.30	1.70
Tz	79.50	50.20	40.00	22.70
Tz_1	79.50	53.20	43.80	26.90
Tz_2	79.50	52.90	43.40	29.40
CTz	62.10	17.60	9.80	2.70
CTz_1	62.10	17.10	9.60	3.70
CTz_2	62.10	20.00	12.40	5.00

Table 4.5: Percentage of detected jumps in the case of SV2F trajectories for different significance levels. The Tz and C − Tz statistics are computed with $c_\vartheta = 3$.

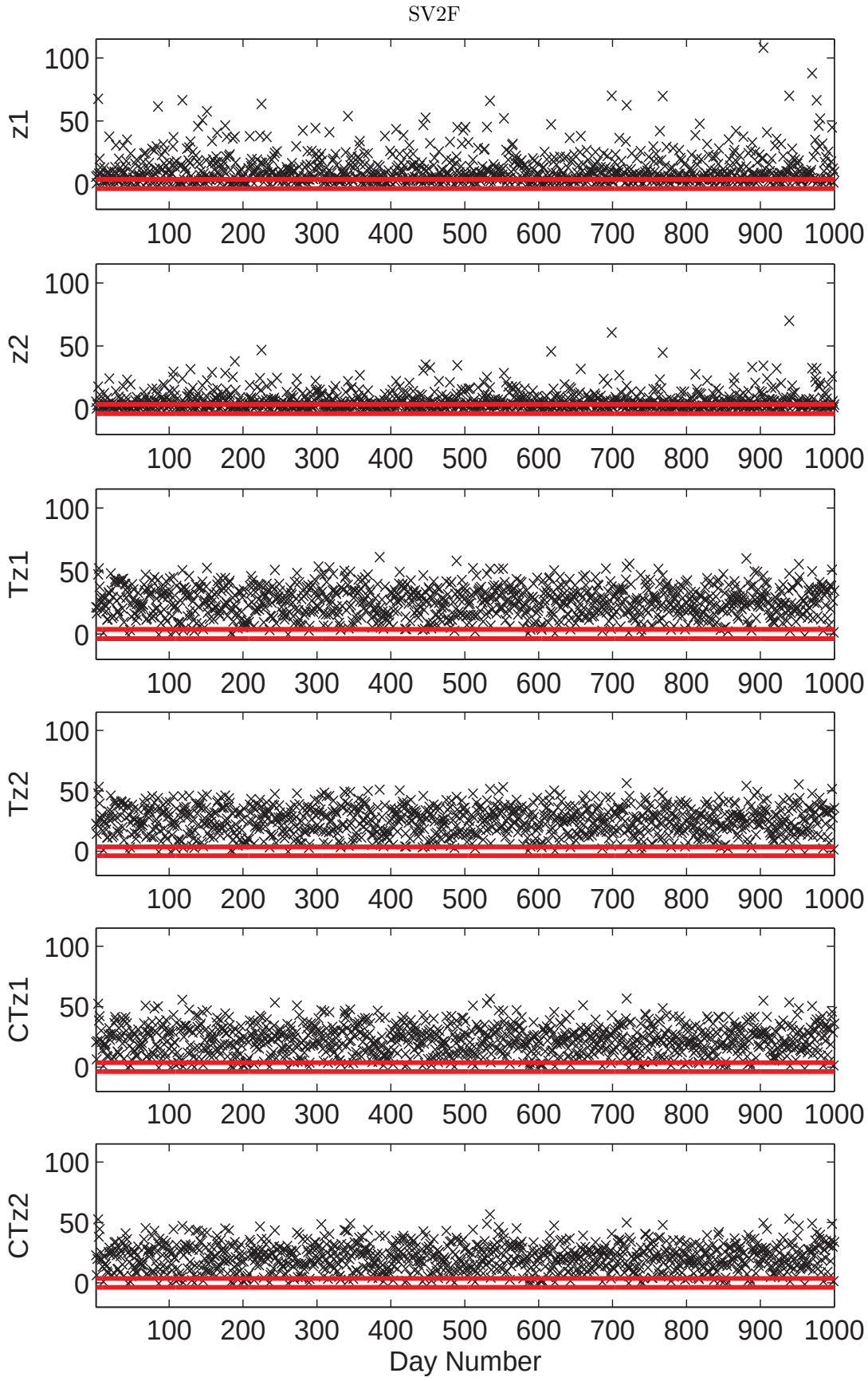

Figure 4.13: z-statistics for the SV2F model. The thick red lines are the ±99.99% confidence levels.

4.5 Detecting Jumps in Data with Jumps

In this section, we test the z statistics and the newly proposed $\mathsf{C}-\mathsf{Tz}$ statistics on data set which are well known to display jumps. We use formulas (4.34) and (4.26) to define the statistics.

The first set is represented by the time series of daily electricity prices. One basic feature of daily electricity prices is to display very pronounced spikes, basically due to power shortages, see e.g Knittel and Roberts (2005); Geman and Roncoroni (2006) and references therein. We compute the jump detecting statistics on the whole time series of daily logarithmic returns of eight electricity markets, both European and American. These data sets display very pronounced jumps (as discussed in Chapter 3). We also consider the test of Lee and Mykland (2007), which is defined as:

$$\mathsf{L}-\mathsf{M}=\delta^{\frac{1}{2}}\frac{X_{t+T}-X_t}{\sqrt{\mathsf{BPV}_t}} \tag{4.46}$$

and which is distributed, in absence of jumps and as $\delta\to 0$, as a standard Normal distribution. Results are shown in Table 4.6.

We find that the z statistics is not able to detect jumps in the electricity time series. Jumps are detected only in Pennsylvania; in the other markets, the null of a continuous trajectory is not rejected, which is completely untenable for electricity prices. This disappointing results can be explained by the fact that most of the jumps in electricity returns are consecutive. After a spike, the price immediately reverts to its original level, generating a second large jump. This completely messes up the measures based on multipower variation, which hinge on the fact that, after a jump, there is a small return vanishing to zero in the limit. Also the $\mathsf{L}-\mathsf{M}$ test fails in detecting jumps, since when there is a spike the sum of two consecutive jumps is zero.

To better point out this effect, we also compute the z test with staggered measures of multipower variation. Staggered bipower variation has been suggested by Huang and Tauchen (2005) and ABD to accommodate microstructure effects in high-frequency time series. The staggered multipower variation is defined as:

$$\mathsf{stag}-\mathsf{MPV}_\delta(X)_t^{[\gamma_1,\ldots,\gamma_M]}=\frac{M}{M-2}\delta^{1-\frac{1}{2}(\gamma_1+\ldots+\gamma_M)}\sum_{j=M+1}^{[T/\delta]}\prod_{k=1}^{M}|\Delta_{j-2(k-1)}X|^{\gamma_k} \tag{4.47}$$

The tests based on staggered multipower variation perform indeed much better, as shown in Table 4.6. However, they would have poor power in detecting jumps in the case of three consecutive jumps, or in the case of "staggered" jumps. All these cases may be relevant in high frequency data.

A much better performance is instead obtained by the corrected test, as suggested by the theoretical properties of threshold multipower variation. When $c_\vartheta=3$ we detect jumps in all the electricity markets, with the exception of Ohio, which has a kurtosis closer to the Normal value[6]; with $c_\vartheta=5$ the test has less power, and difficulties are found in time series with moderately high kurtosis (France, The Netherlands). This is not surprising, since in the limit $c_\vartheta\to\infty$ the $\mathsf{C}-\mathsf{Tz}$ test is identical to the z test. Thus, with an higher c_ϑ, the corrected test gets closer to the z test.

We also analyze the time series of daily interest rate differences. Interest rates are also well known to display jumps, especially in presence of macroeconomic announcements, see e.g. Das (2002); Johannes

[6]It is important to remark that high kurtosis does not imply presence of jumps; the presence of jumps in electricity market is witnessed in the related empirical literature, if not evident by visual inspection of prices.

(2004). Moreover, it is well known that shorter maturity instruments display more jumps, because of liquidity reasons. We then analyze the time series of daily observations for both the 7-day Eurodollar deposit rate and the 3-month Treasury Bill rate, in the period starting in June 1973 and ending in February, 1995, for a total of 5,505 observations. The 7-day time series displays many jumps, and most of them are consecutive; an extensive discussion of the jump properties of these two data sets can be found in Renò et al. (2006).

The results are again shown in Table 4.6. On the 3-months time series, where jumps are rarely consecutive, all the considered z statistics detect the jumps, with a value which is approximately the same. On the 7-days time series jumps are most pronounced; but the z statistics is not able to detect them. The fact that there are many consecutive jumps is signaled by the high value of the staggered z statistics. The $\mathsf{L}-\mathsf{M}$ test is more effective in detecting jumps, even if it has a higher statistics on the less jumpy 3-months time series. The corrected $\mathsf{C}-\mathsf{Tz}$ statistics also detect the jumps; with a value much higher than that of the 3-months time series, in accordance with the evidence that jumps are more frequent and more pronounced in the 7-day time series. The $\mathsf{C}-\mathsf{Tz}$ statistics is lower at $c_\vartheta = 5$, since it becomes closer to the z statistics as c_ϑ increases.

Concluding, consistently with the theory in the previous section, we find that the z statistics, and the $\mathsf{L}-\mathsf{M}$ statistics (4.46), can be very misleading in some situations, e.g. consecutive jumps. On the contrary, the $\mathsf{C}-\mathsf{Tz}$ statistics is more efficient in detecting jumps, especially when consecutive.

4.6 Conclusions

In this chapter we shown that it is possible to disentangle the continuous and discontinuous part of the total quadratic variation. We proposed high-efficient volatility estimator based on an asymptotic theory which only requires the introduction of a threshold. We show that the newly defined estimator is robust to the presence of jumps, well-behaved for continuous trajectories and quite inelastic with respect to the choice of the threshold. We test our estimator on a "tricky" model which, despite it belongs to the class of continuous processes, shows sudden abrupt variations. The proposed method is not significantly deceived, showing small bias even in presence of a fake jump component. Besides, the reliability of the "tricky" model is quite limited: detected discontinuous variations show non-vanishing persistence.

The proposed estimators allow for the construction of accurate jump detection test, like the newly defined $\mathsf{C}-\mathsf{Tz}$. Our empirical results, obtained on US stock index, single stocks and Treasury bond data, also show that jumps can be effectively detected using the proposed statistics. Moreover on Monte Carlo simulations and with electricity price data we find that the jump test well performs even in presence of consecutive jumps.

We start our analysis from the apparent puzzle that jumps have no forecasting power on future volatility. In Chapter 6 we will show that the accurate separation of the continuous and discontinuous part of the quadratic variation proposed in this chapter completely resolves this issue. The sharp identification of a jump via the proposed test will allow the definition of a new high-performing forecasting model.

Market	period	Kurtosis	z	stag-z	$\mathsf{L}-\mathsf{M}$	$\mathsf{C}-\mathsf{Tz}$	
						$c_\vartheta = 3$	$c_\vartheta = 5$
		Electricity markets					
Germany	1 Jan 01 → 25 Jul 05	29.9	1.59	8.63	0.45	10.68	7.82
France	27 Nov 01 → 30 Jul 05	12.6	0.98	8.19	0.16	4.37	1.72
Spain	1 Jan 01 → 26 May 05	10.5	0.83	5.79	1.13	5.00	2.99
The Netherlands	1 Jan 01 → 10 Aug 05	11.6	1.01	6.00	0.60	3.20	1.63
California	2 Jan 02 → 12 Feb 07	52.9	-0.11	6.77	2.45	10.57	6.76
Texas	2 Jan 02 → 19 Mar 07	27.4	-0.91	1.61	2.74	8.18	4.68
Ohio	3 Jan 01 → 11 May 07	4.9	0.51	2.83	0.02	1.41	1.28
Pennsylvania	2 Jan 01 → 9 May 07	12.4	3.45	4.87	0.19	5.79	4.53
		Interest rates					
7-days	1 Jun 1973 - 24 Feb 1995	30.8	2.02	11.51	-16.7	15.39	7.52
3-months	1 Jun 1973 - 24 Feb 1995	18.6	6.90	6.22	-132.6	8.73	7.04

Table 4.6: Jump detection statistics on daily electricity price returns in different markets, and daily interest rate differences for two maturities.

Chapter 5

Jump Detection And Long Range Dependence

This chapter is inspired from the paper of Pirino (2008).

5.1 Introduction and Preliminary Discussion

This chapter is devoted to the investigation of the memory properties of realized volatility time series. Our goal is to highlight that a jump-purified estimate of the realized volatility series allows for a better investigation of its memory. This idea, though from a different perspective, will be widely exploited in the next chapter where a forecasting model is provided.

We analyze autocorrelation properties of realized volatility series with the help of Detrended Fluctuation Analysis (DFA from now on). The basic ideas of this method are explained in what follows.

For a given stationary data sample $(x_i)_{i=1,...,n.}$ we define the integrated series as:

$$y_k = \sum_{i=1}^{k} x_i, \quad k = 1, ..., n. \tag{5.1}$$

Suppose that $\mathbb{E}[x_i] = 0$ and $\mathbb{E}\left[x_i^2\right] = \sigma^2$, then $\mathbb{E}[y_k] = 0$. We investigate the behaviour of the variance $\mathbb{E}\left[y_k^2\right]$ for high values of k:

$$\mathbb{E}\left[y_k^2\right] = \mathbb{E}[\sum_{ij} x_i\, x_j] \quad = \tag{5.2}$$

$$= \sum_{i=j} \mathbb{E}\left[x_i^2\right] + \sum_{i \neq j} \mathbb{E}\left[x_i\, x_j\right] \quad = \tag{5.3}$$

$$= k\,\sigma^2 + 2 \sum_{i=1}^{k-1} (k-i)\ \mathbb{E}\left[x_1\, x_{1+i}\right], \tag{5.4}$$

having used stationarity: $\mathbb{E}[x_i\, x_j] = f(i-j)$. For high values of k we obtain:

$$\mathbb{E}\left[y_k^2\right] \approx k \left\{ \sigma^2 + 2 \sum_{i=1}^{k-1} \mathbb{E}\left[x_1\, x_{1+i}\right] \right\}. \tag{5.5}$$

The dependence of $\mathbb{E}\left[y_k^2\right]$ on k is thus determined by $\sum_{i=1}^{k-1}\mathbb{E}\left[x_1\, x_{1+i}\right]$. We analyze separately the case of weakly and strongly autocorrelated series. If x_i are weakly autocorrelated then:

$$\hat{\rho}\left(h\right) \sim e^{-\frac{h}{h_c}}, \tag{5.6}$$

which gives:

$$\sum_{i=1}^{k-1}\mathbb{E}\left[x_1\, x_{1+i}\right] \approx \int_1^k e^{-\frac{h}{h_c}}dh = h_c\,\left[e^{-\frac{1}{h_c}} - e^{\frac{-k}{h_c}}\right] \overset{k\gg 1}{\sim} h_c\, e^{-\frac{1}{h_c}} = \text{const.} \tag{5.7}$$

Therefore for weakly autocorrelated processes we obtain a linear dependence:

$$x_i \text{ weakly autocorrelated} \Rightarrow \mathbb{E}\left[y_k^2\right] \overset{k\gg 1}{\sim} k. \tag{5.8}$$

While for a long memory process such that:

$$\hat{\rho}\left(h\right) \sim \frac{1}{h^\alpha}, \quad \alpha \in (0,1)\,, \tag{5.9}$$

we obtain:

$$\sum_{i=1}^{k-1}\mathbb{E}\left[x_1\, x_{1+i}\right] \approx \int_1^k \frac{1}{h^\alpha}dh = \frac{1}{1-\alpha}\,\left[k^{1-\alpha} - 1\right] \overset{k\gg 1}{\sim} k^{1-\alpha}, \tag{5.10}$$

Therefore for long memory process we obtain a non-linear dependence:

$$x_i \text{ strongly autocorrelated} \Rightarrow \mathbb{E}\left[y_k^2\right] \overset{k\gg 1}{\sim} k^{2-\alpha}. \tag{5.11}$$

As a consequence memory properties of a time series could be investigated analyzing the dependence of the variance of the integrated series. This is the idea exploited by DFA. More generally we are looking for *self-similarity* in the variance of the integrated series. We say that a function $f\left(t\right)$ is self-similar with parameter β if:

$$f\left(t\right) \overset{d}{=} \lambda^\beta\, f\left(\frac{t}{\lambda}\right), \tag{5.12}$$

where $\overset{d}{=}$ indicates that the two members of the equality have the same statistical distribution and thus the same statistical properties. In the case of long memory process we have shown that, for high k, $\mathbb{E}\left[y_k^2\right] \sim k^{2-\alpha}$, which is self-similar with parameter $2-\alpha$. The DFA approach, after having removed trends which could generate spurious self-similar effects, tries to check whether or not the process under analysis has long memory investigating self-similarity of the integrated series variance.

DFA was proposed, for the first time, by Peng et al. (1994). It has been widely exploited to detect memory in different field of research like economic systems (Liu et al., 1997; Jánosia et al., 1999; Ausloos and Ivanova, 2003), biology and medicine (Ashkenazy et al., 2001; Bhattacharya et al., 2005; Leea et al., 2002; Paraschiv-Ionescu et al., 2008), among others. The method is sketched in what follows.

5.1.1 Construction of DFA Function

The DFA function is built by the following procedure:

1. Given a data sample $(x_i)_{i=1,...,n.}$ compute the detrended integrated time series $y_k = \sum_{i=1}^{k} [x_i - \bar{x}]$, where $\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$.

2. Divide the time interval $i = 1, ..., n$, in $K = \lfloor \frac{n}{L} \rfloor$ sub-sequences of length L without overlapping. We obtain the sub-sampling:

$$y_q^L (p) \quad q = 1, ..., L. \tag{5.13}$$

$$p = 1, ..., K. \tag{5.14}$$

$$\tag{5.15}$$

   For each subsequence $y_q^L,\ q = 1, ..., L$, compute the local trend $\bar{y}_q^L$ as a polynomial fit of r-th degree. If $r = 1$ we are computing first order DFA, if $r = 2$ we get second order DFA and so on.

3. Compute the RMSE with respect to the local trend:

$$g^L (p)^2 = \frac{1}{L} \sum_{q=1}^{L} \left[y_q^L (p) - \bar{y}_q^L (p) \right]^2, \tag{5.16}$$

4. We take the mean value of (5.16) obtaining a function of the sub-sequence length L:

$$F (L) = \frac{1}{K} \sum_{p=1}^{K} g^L (p) = \frac{1}{K} \sum_{p=1}^{K} \sqrt{\frac{1}{L} \sum_{q=1}^{L} \left[y_q^L (p) - \bar{y}_q^L (p) \right]^2}. \tag{5.17}$$

   The procedure returns a function $F (L)$ which is called DFA function.

When the process $(x_i)_{i=1,...,n.}$ is uncorrelated or it has a short-range autocorrelation function the integrated series is a random walk (see definition 2.94) and $F (L) \sim L^{0.5}$ (see Peng et al., 1994). Note that he behaviour $F (L) \sim L^{0.5}$ is the same for both i.i.d. and short-range processes. For a long-memory process we have that $F (L) \sim L^d$ with $d \neq 0.5$.

In this chapter, inspired by the results obtained till now, we propose threshold bipower variation as a proxy for volatility. We then estimate the DFA exponent for both realized volatility and threshold bipower variation. Our Monte Carlo simulations show that realized volatility has worse performance than TBPV which indeed provides estimates of the DFA parameter d unaffected by the presence of jumps.

Finally we perform the simulation of an $AR (1)$ process in order to investigate the significance of the DFA exponent with respect to the sample length. Autoregressive processes of order one have autocorrelation function with exponential decay. Their memory properties are identified by a time parameter α which identifies the characteristic time τ of the autocorrelation function. We simulate different $AR (1)$ processes with increasing values of α and with different sample lengths. For high values of α (which typically characterize the volatility dynamics), the estimates of DFA are shown to be misleading in small samples.

5.2 Monte Carlo Simulations

We simulate 100 replications of two stochastic volatility models: the first one (labeled as SVCJ) has contemporaneous jumps in price and volatility while the second one (labeled as SV) has no jumps. Each replication is composed by a sample of 4000 days $\approx$ 16 years. We follow the approach of Eraker et al. (2003) who estimate the model:

$$dX_t = \mu\, dt + \sqrt{V_{t_-}}\, dW_{1t} + \xi^X\, dN_t \tag{5.18}$$

$$dV_t = \kappa\left(\theta - V_{t_-}\right) dt + \sigma_V \sqrt{V_{t-}}\, dW_{2t} + \xi^V\, dN_t. \tag{5.19}$$

where $corr\left(dW_{1t}, dW_{2t}\right) = \rho\, dt$, $V_{t-} = \lim_{s\uparrow t} V_s$, N_t is Poisson process with constant intensity λ, ξ^X and ξ^V are the jump sizes in returns and volatility respectively. The SVCJ model has correlated jump sizes with $\xi^V \sim \exp\left(\mu_V\right)$ and $\xi^X|\xi^V \sim N\left(\mu_X + \rho_J\, \xi^V, \sigma_X^2\right)$. We use the parameters estimated by Eraker et al. (2003) for the S&P500 (see Table 5.1). The SV model is easily obtained form the SVCJ imposing $dN_t = 0$ for all t. As for the SVCJ model we use the parameters estimated by Eraker et al. (2003) for the same asset and we list them in Table 5.1.

Parameter	SVCJ	SV
μ	0.0554	0.0444
θ	0.5376	0.9502
κ	0.0260	0.0231
σ_V	0.0790	0.1434
μ_X	-1.7533	
ρ_J	-0.6008	
σ_X	2.8864	
μ_V	1.4832	
ρ	-0.4838	-0.3974
λ	0.0066	

Table 5.1: Parameters for models SVCJ and SV expressed in percentage form and on a daily basis. The estimates are provided by Eraker et al. (2003).

On the simulated series we estimate realized volatility and threshold bipower variation with $\delta = 300$ sec. The output of the experiment are 100 replications of the two series RV^i and TBPV^i with $i = 1, ..., 4000$ for the two simulated models SVCJ and SV. As mentioned above the exponent d characterizing the detrended fluctuation function $F\left(l\right) \sim l^d$ provides insights about the memory properties of the series. We choose to adopt first order DFA: a linear trend is removed from the integrated signal for each value of the time window l. Detrended fluctuation of order q (usually indicated as DFA-q) removes trends of order q in the integrated signal and of order $q-1$ in the original signal, as discussed in Hu et al. (2001) and Bunde et al. (2000). We report the results obtained for the DFA-1 because almost no differences are found with higher orders. The chosen windows lengths for the DFA algorithm are: $\left[2^3, 2^4, ..., 2^{\lfloor \log_2 \frac{N}{2} \rfloor}, N\right]$, where N

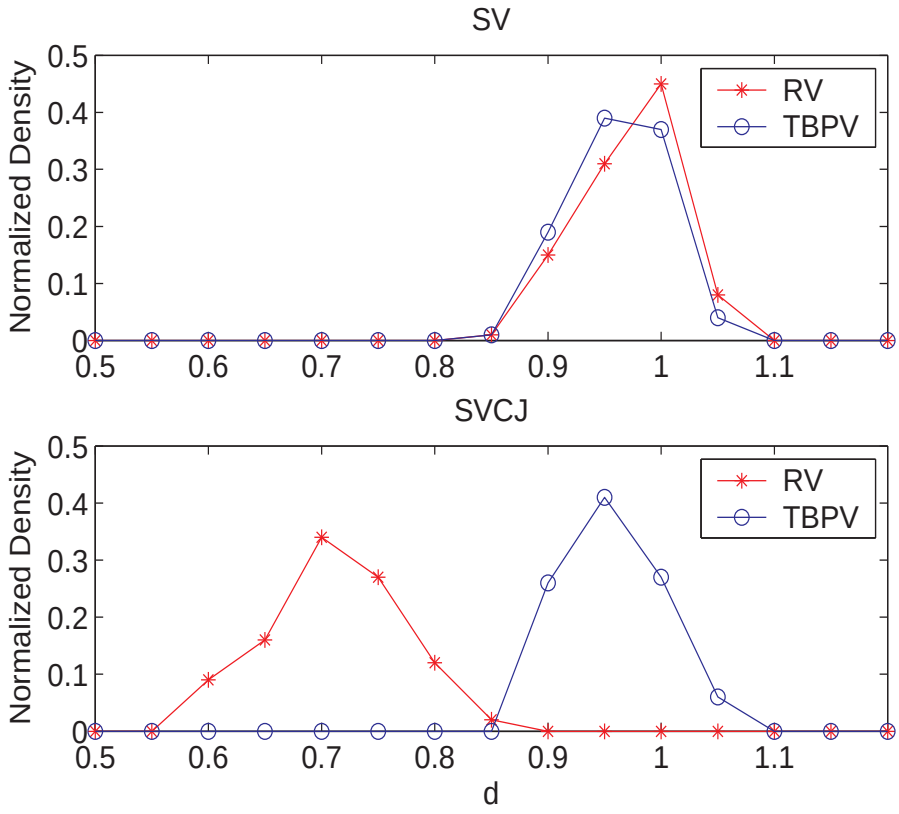

Figure 5.1: Normalized density of the parameter d obtained by DFA on simulated time series for the SV (up) and SVCJ (down) models.

is the sample length. The 100 values of the exponent d obtained from the simulated series form a statistics whose densities are reported in Figure 5.1. It is evident that using TBPV as proxy for the volatility gives the same results for both SV and SVCJ models, while the estimate via RV is affected by the presence of jumps. Table 5.2 reports the mean value and the standard errors (in brackets) for the collected statistics of the DFA exponent. Standard errors are computed as the ratio between standard deviation of the density functions and the square root of the sample length. The value obtained via TBPV is 0.9856 for the SV model and 0.9822 for the SVCJ model. In both cases TBPV reveals the same type of long memory volatility while the estimate by RV is unstable and affected by the jump component. It is important to note that DFA seems to reveal long memory in model (5.18), which is instead a short range data generating process. This issue will be thoroughly discussed in Section 5.4.

Models	$d\left(RV_{300s,1\text{day}}\right)$	$d\left(TBPV_{300s,1\text{day}}\right)$
SVCJ	0.7383(0.0058)	0.9822(0.0038)
SV	0.9975(0.0042)	0.9856(0.0041)

Table 5.2: Mean value of the collected statistics for the DFA exponent with standard errors in brackets. Standard errors are computed as the ratio between standard deviation of the density functions and the square root of the sample length.

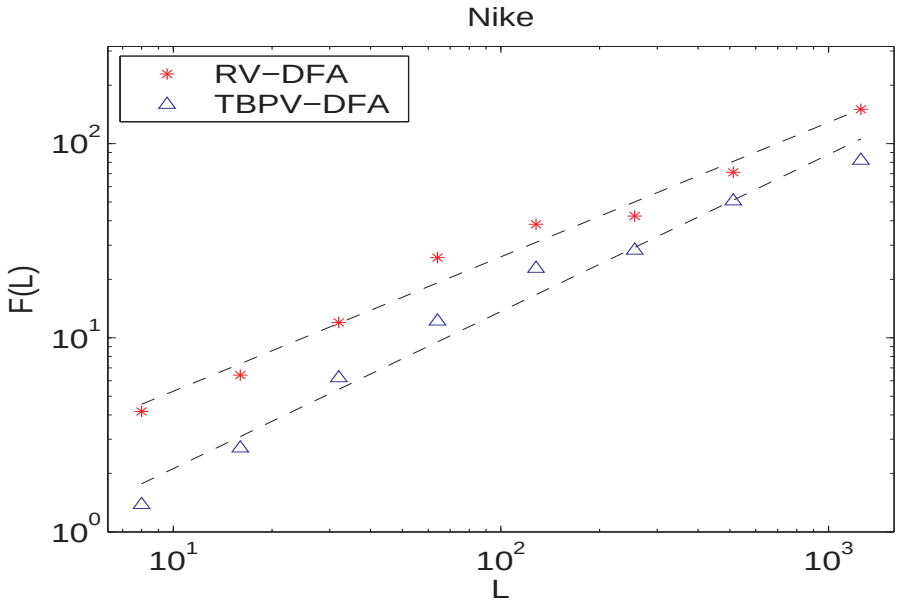

Figure 5.2: Log-log plot of DFA functions for the Nike stock, computed on RV (red starred points) and TBPV (blue triangles) with $\delta = 300\,s$ and on a daily basis ($T = 1$ day). The linear fits of the detrended fluctuation functions are depicted as dashed lines for both RV and TBPV.

5.3 Empirical Data

We now turn to the analysis of empirical data. We have at our disposal 4344 days $\approx$ 17 years of S&P500 and 1252 days $\approx$ 5 years of 10 stocks: Exxon (xom), Pfizer (pfe), Tektronix (tek), Microsoft (msft), Nike (nke), Harley Davidson (hdi), Intel (intc), Citigroup (c), Hasbro (has), Alcoa (aa). The sampling frequency is $\nu = \frac{1}{\delta} = \frac{1}{300\,\text{sec}} = 3.3 \times 10^{-3}$Hz, as for Monte Carlo simulations. Figure 5.2 shows, in a log-log scale, the linear fit of the detrended fluctuation functions computed on the RV and TBPV series (with $\delta = 300\,s$ and on a daily basis), for Nike stock. We choose to represent DFA functions only for the stock with the greatest difference between RV and TBPV estimates of DFA exponent (see Table 5.3). A closer inspection of Figure (5.2) reveals that the DFA points are not extremely collinear while self-similarity would imply a perfect alignment. However the observed curvature is too small to be measured and addressed to a short-range process. This issue is further discussed in Section 5.4.

The line slope represents the DFA exponent of the correspondent stock. Table 5.3 reports the estimated DFA exponent for the analyzed series. The estimates on the S&P500 series are nearly the same for both TBPV and RV, while for the ten stocks TBPV reveals to be more stable. This difference may be attributed to the fact that S&P500 is a weighted mean of 500 stocks and hence it is subjected to less discontinuous variations than any other single stock. The effect of spikes on DFA is further investigated by Chen et al. (2001). Both correlated and anticorrelated signals with superimposed jumps are analyzed. Our contribution is different in nature: we provide a technique to drastically reduce the disturbance generated by discontinuous variations on volatility estimates which, in turn, provide cleaner DFA estimates.

Asset	$d\left(RV_{300s,1\text{day}}\right)$	$d\left(TBPV_{300s,1\text{day}}\right)$
xom	0.7672	0.8224
pfe	0.7473	0.8123
tek	0.7496	0.8198
msft	0.7569	0.8021
nke	0.6923	0.8089
hdi	0.7452	0.8251
intc	0.7576	0.8346
c	0.7562	0.8275
has	0.7778	0.8268
aa	0.7367	0.8101
SP500	0.9397	0.9542

Table 5.3: Estimated DFA exponent for the empirical data set.

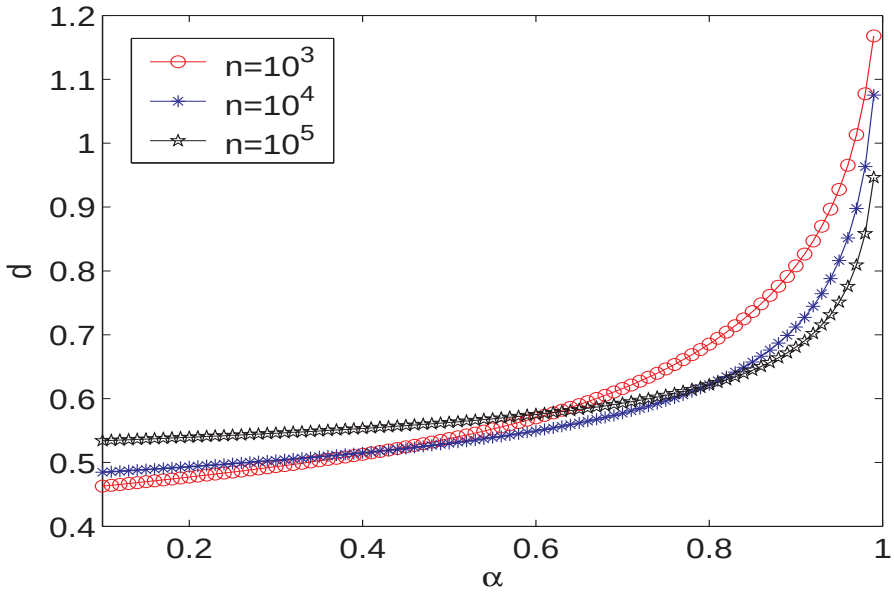

Figure 5.3: Estimated DFA exponent for different values of the parameter α of AR (1) model (5.20) and for different lengths n of the time series.

5.4 Long Memory or Short Range?

This section is dedicated to test the significance of the DFA exponent with respect to the sample length. If the time series has an infinite length then a DFA exponent approximately equal to 1 means that the series is strong autocorrelated with power law decay autocorrelation function. We will show that, in finite samples, the estimate of the DFA exponent is highly sensitive to the length of the series and a DFA exponent near to one does not necessarily correspond to a long memory process[1]. To show this, we simulate a simple AR (1) process:

$$y_0 = 0, \quad y_t = \alpha\, y_{t-1} + \epsilon_{t-1}, \quad t = 1, ..., n, \tag{5.20}$$

where $0 < \alpha < 1$, ϵ_t are normal distributed errors with zero mean and unit variance and n is the length of the time series. The time series y_t defined in (5.20) has an exponential decay autocorrelation function $C(t) \sim \exp(-t/\tau)$ with time constant decay $\tau = \left(\log \frac{1}{\alpha}\right)^{-1}$. This means that after a time τ the process autocorrelation is drastically reduced. Values of α near to 1 correspond to a slow exponential decay (high values of τ) for the autocorrelation function, that is a short range memory with characteristic time comparable to or even grater than the whole duration of y_t. However, regardless of the choice of α, the DFA exponent of y_t should be different from one. Figure 5.3 shows the plot of the estimated DFA exponent for different values of α and for three time series lengths $n = 10^3$, $n = 10^4$ and $n = 10^5$. The DFA algorithm is performed over the same window lengths used in Section 5.2.

If $\alpha \approx 1$ the DFA exponent is ≈ 1 for both $n = 10^3$ and $n = 10^4$. Therefore DFA is potentially misleading: a slow decay autocorrelation function may be confused with a power-law in small samples.

[1]The speed of convergence of DFA estimates is further studied in Bardet and Kammoun (2008), where the asymptotic properties of the DFA functions are established. Empirical DFA confidence intervals have been studied by Weron (2002).

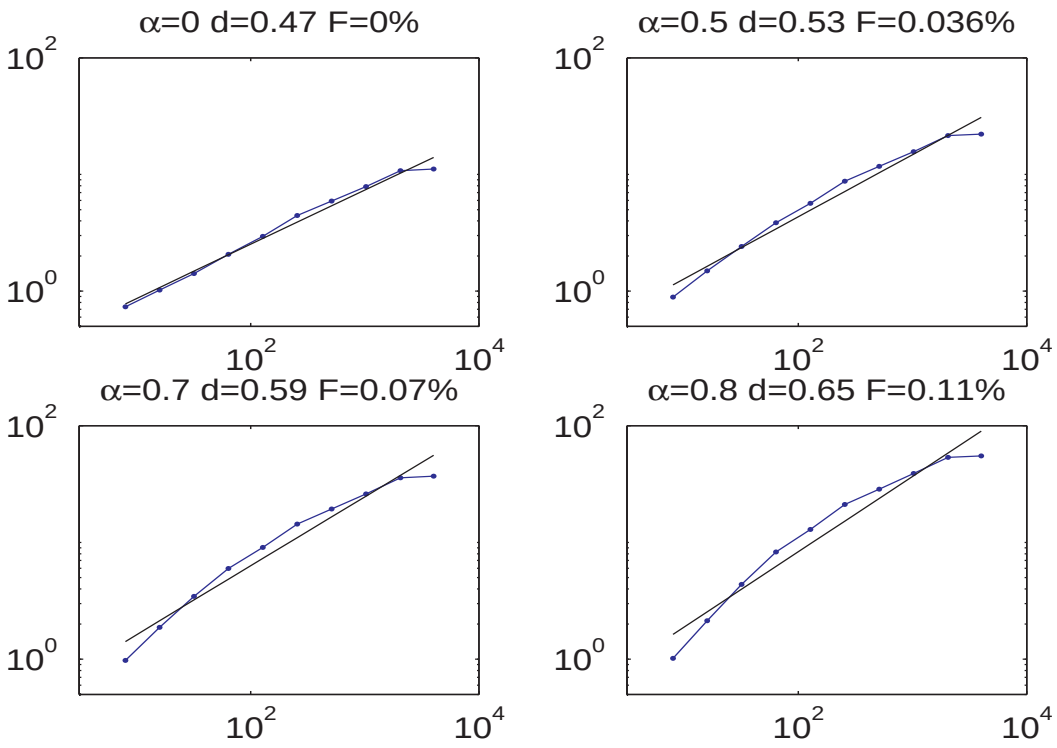

Figure 5.4: Estimated DFA functions of process (5.20) for different values of α and corresponding linear fits, in a log-log scale (the axis labels have been suppressed for clarity). The sample length is $n = 4 \times 10^3$ as for the Monte Carlo simulations. For each plot we report the chosen value of α, the estimated DFA exponent d from the linear regression, and the ratio F between the characteristic time τ of the autocorrelation function and the whole process duration.

Indeed the estimation is drastically improved only if a large number ($n = 10^5$) of observations is provided. In financial econometrics, especially for volatility which is a derived quantity, it is unlikely to deal with a huge number of observations like $n = 10^5$. These results suggest that, in a small sample, DFA is not able to distinguish between a long-memory and a short-range process with a characteristic time comparable to the whole process duration. Nevertheless more efficient semi-parametric estimators of long-memory can be defined. Examples are the local Whittle estimator Robinson (1995), the adaptive global log-periodogram method Moulines and Soulier (2000) and the wavelet based estimator Abry et al. (2002).

Figure 5.4 reports the computed DFA functions for different value of α. The sample length is $n = 4 \times 10^3$ as for the Monte Carlo simulations. For each plot we report, in percentage form, the ration $F = \tau/n$ between the autocorrelation characteristic time and the whole process duration. Besides, the estimate of the DFA exponent d is shown for each plot. When $\alpha = 0$ the process y_t is i.i.d. and the estimate of DFA exponent is, as expected, $d \approx 0.5$. Increasing the value of α, and accordingly the value of F, gives an estimate of d in the range $(0.5, 1)$, i.e. it suggests a long-memory process. However curvatures of DFA points are again too small to be used as a proxy for short-range memory detection.

5.5 Conclusions

In this chapter we compare the estimates of the DFA exponent via TBPV and RV. Monte Carlo simulations of a stochastic model with contemporaneous jumps in price and volatility show that the RV estimates are affected by jumps while TBPV produces the same results regardless of the presence of discontinuous variations. This result leads us to confirm that the persistence of volatility becomes more evident if the continuous and discontinuous part of the price variation are studied separately, as pointed out in the next chapter. We find similar results on empirical data. We analyze two kind of assets: S&P500 index and ten stocks. S&P500 is a weighted sum of 500 stocks, therefore the jump component is smaller than the correspondent one of any other single stock. The estimates of the DFA exponent via RV performed on single stocks and the full index are unstable if compared to those produced using TBPV.

We finally investigate the significance of the DFA exponent when varying the sample length. For this purpose an AR(1) process with different sample lengths has been simulated. Although the simulated process is characterized by an autocorrelation function with exponential decay, the DFA exponent is different from 0.5 unless the series length reaches very high values, unrealistic in financial econometrics. Our findings suggest that DFA is capable to establish if the analyzed time series is a short range process with characteristic time of the autocorrelation function negligible or not with respect to the whole process duration. Still it is not the correct test for detecting long memory in a time series.

Chapter 6

Volatility Forecasting: The Jumps Do Matter.

This chapter is inspired from the paper of Corsi, Pirino, and Renò (2008).

6.1 Introduction

As anticipated in Chapter 4 and Chapter 5, here we propose a forecasting model for volatility which exploits the jump detection techniques previously described. We start with some empirical facts about the memory properties of realized volatility and related literature. Empirical evidence on strong temporal dependence of realized volatility has been already found for instance in Andersen, Bollerslev, Diebold and Labys (2001) and Andersen, Bollerslev, Diebold and Ebens (2001). This evidence, together with our empirical results reported below, suggests that realized volatility series should be described by long-memory type of models, see Andersen et al. (2003); Banerjee and Urga (2005); McAleer and Medeiros (2006). A visual inspection of this empirical fact is provided by Figure 6.1 which shows the autocorrelation function of S&P500 realized volatility from December 7, 1989 to December 22, 2004. Black lines represents 95% i.i.d. process confidence levels. Recently, Corsi (2004) and Corsi et al. (2008) introduced a class of time series models called Heterogeneous Auto-Regressive (HAR) models that successfully achieves the purpose of modeling the long memory behavior of financial variables in a simple and parsimonious way.

Inspired by the HARCH of Müller et al. (1997) where the conditional variance is function of squared returns over many horizons, Corsi (2004) proposes a stochastic additive cascade of three different realized volatility components corresponding to the three main different time horizons operating in the market (daily, weekly and monthly).

The Corsi (2004) model is based on the "Heterogeneous Market Hypothesis" presented by Müller et al. (1993) which assumes that market agents are heterogeneous. It is widely recognized that, empirically, there is a strong positive correlation between volatility and market presence. If the market agents were homogeneous, the addition of new agents would reduce the market volatility. In fact, in an homogeneous market, the more the agents the faster the convergence to price transaction and then the lower the

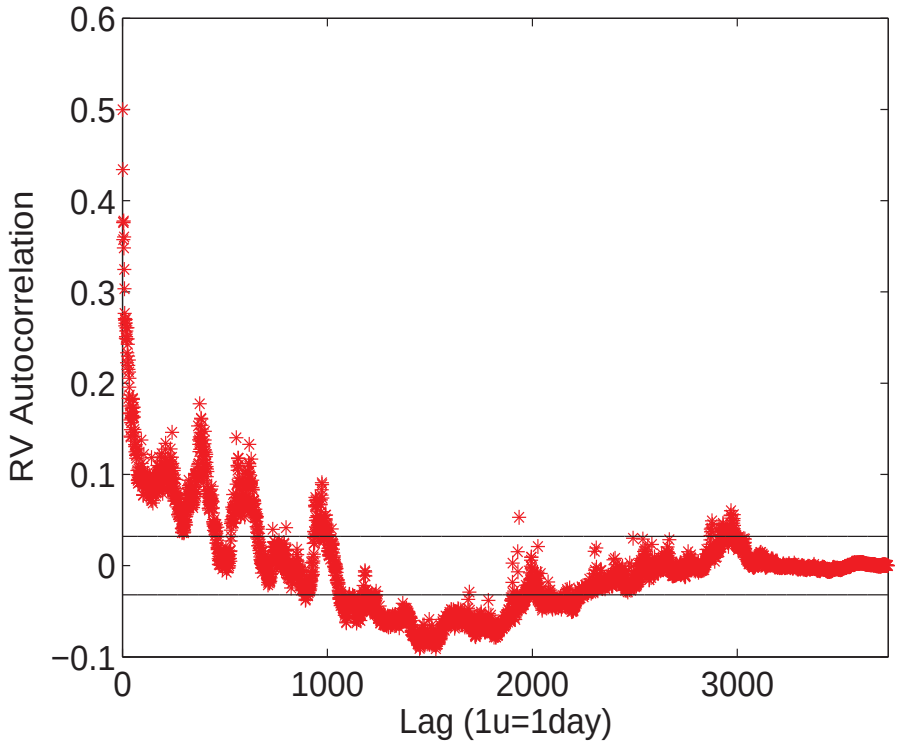

Figure 6.1: Autocorrelation coefficients (red circles) for lags greater or equal than one of the S&P500 Realized Volatility from December 7, 1989 to December 22, 2004. Black lines are 95% i.i.d. process confidence levels.

volatility. On the other hand, if the market agents were heterogeneous, each of them would have different market expectations, budgets, beliefs and thus they would be inclined to buy/sell at different prices. This mechanism creates volatility.

The Corsi (2004) model focuses on the heterogeneity generated by different time horizons. A financial market is composed by participants who operates at different frequencies. At highest frequencies there are intra-day speculators and market dealers (*inter alia*), while at lowest frequency we find central banks and pension fund investors (*inter alia*). Apart from nomenclature and classification of market agents the most relevant fact is that each participant possesses different reaction times to news, events, price changes. The idea is that they generate different volatility components, which are resumed in three main contribution in the Corsi (2004) model: daily, weekly and monthly volatility. The model follows the line sketched by Müller et al. (1990), who found that there is a time scaling behaviour among agents. Long-term participants influence short-term agent choices. This types of hierarchical structures is found in FX markets by Ghashghaie et al. (1996). They observe that the relationship between the probability density of FX price changes and the time delay is very similar to the probability density of the velocity differences of two points in a turbulence flow and their spatial separation. This results generates a strong correspondence between information cascade in FX markets and energy cascade in hydrodynamic turbulence. Moreover a scaling behaviour for the S&P500 is found by Mantegna and Stanley (1995).

The three period stochastic volatility cascade used by Corsi (2004) leads to simple AR-type models

in the realized volatility with the feature of considering realized volatilities defined over different time horizons (the HAR-RV models).

We use the separation of the quadratic variation in its continuous and discontinuous component in a suitable extension of the forecasting model of Corsi (2004), which, in spite of its simplicity, is able to reproduce the main features of volatility dynamics, including long memory (Section 6.2).

Empirical results (Section 6.3) confirm our hypotesis. We find that jumps have a positive and highly significant impact on future volatility, a result which cannot be observed when bipower variation is employed because of its inherent bias. On stock index futures data, our forecasting models provides higher R^2, especially in days following a jump, and lower root mean square error, with respect to the existing models in the literature. We find pretty similar results on single stocks and US bond data, indicating that our findings are not peculiar to the S&P500 index.

6.2 The HAR Model

Although the HAR model does not formally belong to the class of long-memory models, it is able to reproduce a memory decay indistinguishable from that observed in empirical data. As pointed out in Section (5.4) a short range process could be confused for a long memory data type when the characteristic time of the autocorrelation function is comparable to or even greater than the whole observation window.

To explicitly define the HAR models, and to extend them to disentangle the different contributions of continuous and discontinuous quadratic variation, let first define the average multiperiod realized variation i.e. a realized variation measure aggregated over longer horizons and normalized to the daily scale:

$$\mathsf{RV}_{t-h:t} = \frac{1}{h}(\mathsf{RV}_{t-h} + \mathsf{RV}_{t-h+1} + ... + \mathsf{RV}_{t}) \tag{6.1}$$

To keep the HAR model simple and intuitive we will use only three aggregation frequencies: daily ($h = 1$), weekly ($h = 5$) and monthly ($h = 22$). The standard HAR-RV model can then be written as

$$\mathsf{RV}_{t+1:t+h} \quad = \quad \beta_0 + \beta_d \mathsf{RV}_t + \beta_w \mathsf{RV}_{t-5:t} + \beta_m \mathsf{RV}_{t-22:t} + \varepsilon_t \tag{6.2}$$

where ε_t is a standard IID noise and where $h = 1$ forecasts daily volatility, $h = 5$ forecasts weekly volatility and $h = 22$ forecasts monthly volatility. This model has been widely employed in recent applications; see for example Forsberg and Ghysels (2007); Maheu and McCurdy (2006); Clements et al. (2008).

The HAR model can also be applied to the realized volatility, i.e. the square root of the realized variance, or to its log transformation. This two similar nonlinear transformations have the practical advantage of making the residual distribution much closer to the Normal one.

In what follows, we extend the HAR-RV model to directly incorporate the different contribution of jumps. ABD (i.e. Andersen et al., 2007) extended the HAR-RV model by adding a jump component (HAR-RV-J).

$$\mathsf{RV}_{t+1:t+h} \quad = \quad \beta_0 + \beta_d \mathsf{RV}_t + \beta_w \mathsf{RV}_{t-5:t} + \beta_m \mathsf{RV}_{t-22:t} + \beta_j \widehat{J}_t + \varepsilon_t \tag{6.3}$$

where $\widehat{J}_t$ is estimated as

$$\hat{J}_t = I_{t,J} \cdot (\mathsf{RV}_t - \mathsf{BPV}_t)^+ , \tag{6.4}$$

$x^+ = max(x, 0)$, and $I_{t,J}$ is an indicator function which is equal to 1 if jumps are detected on day t, and equal to 0 elsewhere. In their study, ABD use $I_{t,J} = I_{\{z_t > \Phi_\alpha\}}$, with z_t given by (4.34) and Φ_α is the quantile function of the Normal distribution at confidence level α. Clearly, different tests can be adopted, as for example (4.35) and (4.36).

ABD estimated β_j in equation (6.3) to be negative. Now, the total realized variance can be decomposed in its continuous (C_t) and jump (J_t) component, (i.e. $\mathsf{RV}_t = C_t + J_t$) and that RV_t enters both the weekly ($\mathsf{RV}_{t-5:t}$) and monthly ($\mathsf{RV}_{t-22:t}$) measures; as a consequence the jump component J_t is present in all the regressors of the HAR-RV-J model making the interpretation of the impact of J_t non trivial. In order to correctly isolate the impact of the jump component J_t, an estimate of the continuous part should be employed as dependent variable in place of the realized variance. A natural estimate of the continuous component is simply the difference between the realized variance and the estimated significant jump. Estimating the jump component as in equation (6.4) and denoting the corresponding continuous part as

$$\widehat{C_t} = \mathsf{RV}_t - \widehat{J_t} \tag{6.5}$$

we get the following HAR-RV-CJ model, similar to that used by ABD in their forecasting analysis[1]:

$$\mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \widehat{C}_t + \beta_w \widehat{C}_{t-5:t} + \beta_m \widehat{C}_{t-22:t} + \beta_j \widehat{J}_t + \varepsilon_t \tag{6.6}$$

where $\widehat{C}_{t-5:t}$ and $\widehat{C}_{t-22:t}$ are, analogously to the realized volatility measures, the weekly and monthly average multiperiod aggregation of the daily continuous component $\widehat{C_t}$. Using model (6.6), ABD still estimate β_j to be not significant, and the same conclusion has been reached by Forsberg and Ghysels (2007), Giot and Laurent (2007) and Busch et al. (2006); see also the analysis of Ghysels et al. (2006).

In the light of the above sections, it is natural to estimate the jump component using threshold bipower variation instead. Moreover, since we introduced threshold multipower variation, we have more flexibility in choosing the jump detector $I_{t,J}$ and the measures of C_t and J_t in days in which $I_{t,J} = 1$. The motivation for using an alternative bipower variation estimator is that in the measures (6.4) and (6.5) an unbiased estimate of $[X^c]$ in the presence of jumps is needed. Now, Sections 4.4 and 4.5 show that bipower variation is a largely biased measure of $[X^c]$ in days where jumps are present. For this reason, we use TBPV in days in which jumps are detected.

Formally, we define the HAR-RV-TCJ model as:

$$\mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \widehat{TC}_t + \beta_w \widehat{TC}_{t-5:t} + \beta_m \widehat{TC}_{t-22:t} + \beta_j \widehat{TJ}_t + \varepsilon_t \tag{6.7}$$

where we employ the threshold bipower variation measure to estimate the jump component

$$\widehat{TJ}_t = I_{t,J} \cdot (\mathsf{RV}_t - \mathsf{TBPV}_t)^+ \tag{6.8}$$

and the corresponding continuous part $\widehat{TC}_t = \mathsf{RV}_t - \widehat{TJ}_t$, which is equal to RV_t if $I_{t,J} = 0$ and to TBPV_t if $I_{t,J} = 1$.

Its square-root and logarithmic counterparts will read:

$$\mathsf{RV}^{\frac{1}{2}}_{t+1:t+h} = \beta_0 + \beta_d \widehat{TC}_t^{\frac{1}{2}} + \beta_w \widehat{TC}^{\frac{1}{2}}_{t-5:t} + \beta_m \widehat{TC}^{\frac{1}{2}}_{t-22:t} + \beta_j \widehat{TJ}_t^{\frac{1}{2}} + \varepsilon_t \tag{6.9}$$

[1] ABD also consider weekly and monthly aggregation of the jump component.

and

$$\log \mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \log \widehat{TC}_t \beta_w \log \widehat{TC}_{t-5:t} + \beta_m \log \widehat{TC}_{t-22:t} + \beta_j \log\left(\widehat{TJ}_t + 1\right) + \varepsilon_t \quad (6.10)$$

and the same transformations[2] will be estimated for model (6.6).

To evaluate the forecasting performance of the different models, we use the R^2 of Mincer-Zarnowitz forecasting regressions, as well as the (relative) RMSE in predicting the square root of RV, defined as:

$$RMSE = \frac{1}{T}\left[\sum_{t=1}^{T}\left(\frac{\mathsf{RV}_t^{\frac{1}{2}} - \widehat{\mathsf{RV}}_t^{\frac{1}{2}}}{\mathsf{RV}_t^{\frac{1}{2}}}\right)^2\right]^{\frac{1}{2}} \quad (6.11)$$

where $\mathsf{RV}_t^{\frac{1}{2}}$ is the measured value of realized volatility and $\widehat{\mathsf{RV}}_t^{\frac{1}{2}}$ is the predicted value implicit in the model[3]. We use the square root of RV since this is the value mostly employed in risk management applications.

6.3 Empirical Analysis

Our data set cover a long time span of almost 15 years of high frequency data for the S&P 500 futures and US Treasury Bond with maturity 30 years. The purpose of this section is mainly to analyze the impact of jumps on future volatility when threshold bipower variation is employed as a measure of jumps. We will focus not only on the impact of jumps on future realized volatility, but also on the performance of models which explicitly incorporate jumps in forecasting volatility.

Our tables are built at confidence level $\alpha = 99.9\%$ but the most interesting quantities will be computed and plotted for different values of α as well. All the analysis presented in this section is based on measures of threshold multipower variation with a value $c_\vartheta = 3$, and using the $\mathsf{C}-\mathsf{Tz}$ statistics to detect jumps.

6.3.1 Stock index futures S&P500 data

The first data set we analyze is the S&P500 futures time series. We dispose of all high-frequency transactions from January 1990 to December 2004 (3,736 days). In order to mitigate the impact of microstructure effects on our estimates, we choose, as in ABD, a sampling frequency of $\delta = 5$ minutes, corresponding to 84 returns per day.

Figure 6.2 is an example in which using the $\mathsf{C}-\mathsf{Tz}$ statistics is effective. It displays the S&P 500 time series on one specific day, in which there is an evident jump. However, in this day, the z statistics, which is based on multipower variation, is negative and does not reveal a jump at any reasonable significance level; while the $\mathsf{C}-\mathsf{Tz}$ statistics, which is based on threshold multipower variation, does reveal a very significant jump. Our interpretation for this day is that, since the jump appears in the form of two consecutive and very large returns, this creates a huge bias (especially in the quarticity estimates) which makes the z statistics very noisy. This bias is instead completely removed by the corrected threshold estimators.

[2]We remark that in the logarithmic model the term $\log\left(\widehat{TJ}_t + 1\right)$ makes the estimates of the parameters $\beta_d, \beta_w, \beta_m, \beta_j$ not invariant to time scaling. To get approximate invariance, it is advisable to choose a time unit such that $\widehat{TJ}_t >> 1$. For this reason, in what follow we use annualized quantities.

[3]Patton (2006) suggest to use the absolute MSE as a proper loss function, see also the related discussion in Forsberg and Ghysels (2007). However, the ranking of models made in Section 6.3 does not change if we use MSE instead of the relative $RMSE$ (6.11), which we find more intuitive since it gives immediately the relative forecasting error.

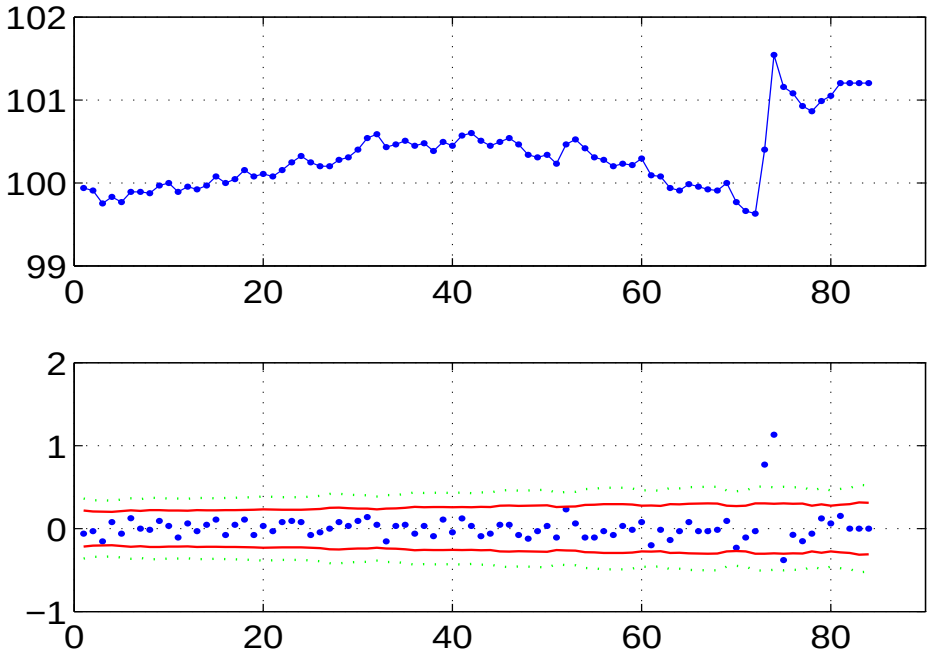

Figure 6.2: Rescaled time series (top) and 5-minutes logarithmic returns (bottom) of the S&P500 on 12 April 1990. The solid and the dotted line are our estimated threshold with $c_\vartheta = 3$ and $c_\vartheta = 5$ respectively. The jump statistics are $z = -0.2545$, $\mathsf{C} - \mathsf{Tz} = 4.5055$ with $c_\vartheta = 3$, $\mathsf{C} - \mathsf{Tz} = 4.4745$ with $c_\vartheta = 5$.

Figure 6.3 shows the number of jumps in this sample, as detected by the condition $\mathsf{C} - \mathsf{Tz} > \Phi_\alpha$ and $z > \Phi_\alpha$ as a function of α. We see that with the statistics based on threshold multipower variation, we get an higher number of jumps.

Here and in what follows, we only discuss the model on RV_t, but we also estimate the corresponding model for $\mathsf{RV}_t^{\frac{1}{2}}$, equation (6.9), and $\log \mathsf{RV}_t$, equation (6.10).

We estimate model (6.7), which we compare with model (6.6). We also estimate the standard HAR-RV model (6.2) for comparison with a model with no separation between continuous and discontinuous component. Results are reported in Table 6.1, where all jumps have been estimated with the $\mathsf{C} - \mathsf{Tz}$ statistics.

Results are unambiguous. When the jump component is measured by means of bipower variation, as in the HAR-RV-CJ model, it is negatively significant for the square model and insignificant for the log and square root model in explaining future volatility. This surprising result is at odd, in our opinion, with the economic intuition which would suggest an increase in volatility after a jump in the price process (especially if large and negative). Moreover, this result is even more puzzling, given that the unconditional correlations between realized variation and jumps lagged by one day is significantly positive and around 20% for the variances, 30% for the volatilities and 25% for the log volatilities.

The explanation for this result could be that the continuous component $\widehat{C_t}$ estimated using bipower variation is still contaminated by the jump component and hence, as for the case with the realized variance in model (6.3), the impact of jumps is also passing through the positive coefficients of the other regressors.

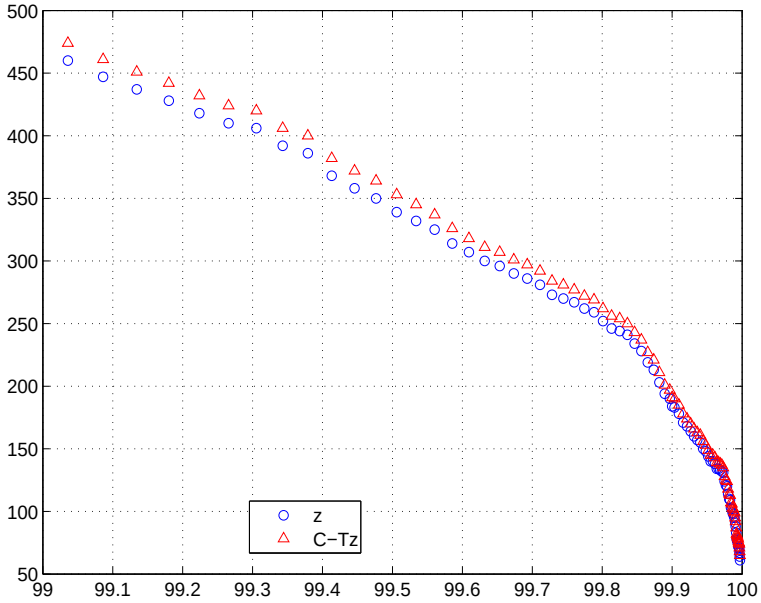

Figure 6.3: Number of days which contain jumps in the S&P500 sample obtained with the $\mathsf{C}-\mathsf{Tz}$ statistics (4.26), as a function of the confidence level α. The total number of days is $3,736$.

When instead the jump component is measured by means of threshold bipower variation, the coefficients β_j are positive, and highly significant in explaining future volatility for the square root and log model. Most importantly, the HAR-RV-TCJ model yields an higher R^2 and a lower $RMSE$, thus showing a better forecasting power. To better understand this point, we divide the sample in days immediately following the occurrence of a jump, and days which follow days with no jumps. On these two samples we compute the R^2 and $RMSE$ statistics separately, denoting them by $J-R^2, J-RMSE$ and $C-R^2, C-RMSE$, respectively. The results in Table 6.1 show that the TCJ model largely improves the forecasting power on realized volatility in days immediately following a jump, and it is still slightly more performing in days which do not follow a jump. Our interpretation of this result is that, since we are better measuring the jump component, we are also removing noise from the continuous component in the explanatory variables; and thus, we also get slightly better results on days in which there were no jumps before.

Our findings are quite robust to the chosen confidence interval. Figure 6.4 displays the most important quantities as a function of the confidence interval α. It shows that the jump component of the HAR-RV-TCJ model, as measured by the $t-$statistics of the coefficient β_j, is always positive for all models and highly significant for square root and log transformations; while the jump component of the HAR-RV-CJ model is mainly significantly negative or not significant. Importantly, it shows that the HAR-RV-TCJ model provides superior forecasts when measured in terms of R^2 and the $RMSE$, irrespective of the confidence level used and model employed.

Figure 6.5 summarizes our findings. It reports the mean square error in forecasting daily realized volatility ($h = 1$) for all the considered models, computed on the full sample (top) and only on days subsequent the occurrence of a jump (bottom). The figure shows that, using the $RMSE$ metrics, in the full sample the ranking depends on the specification: logarithmic models outperform the square root

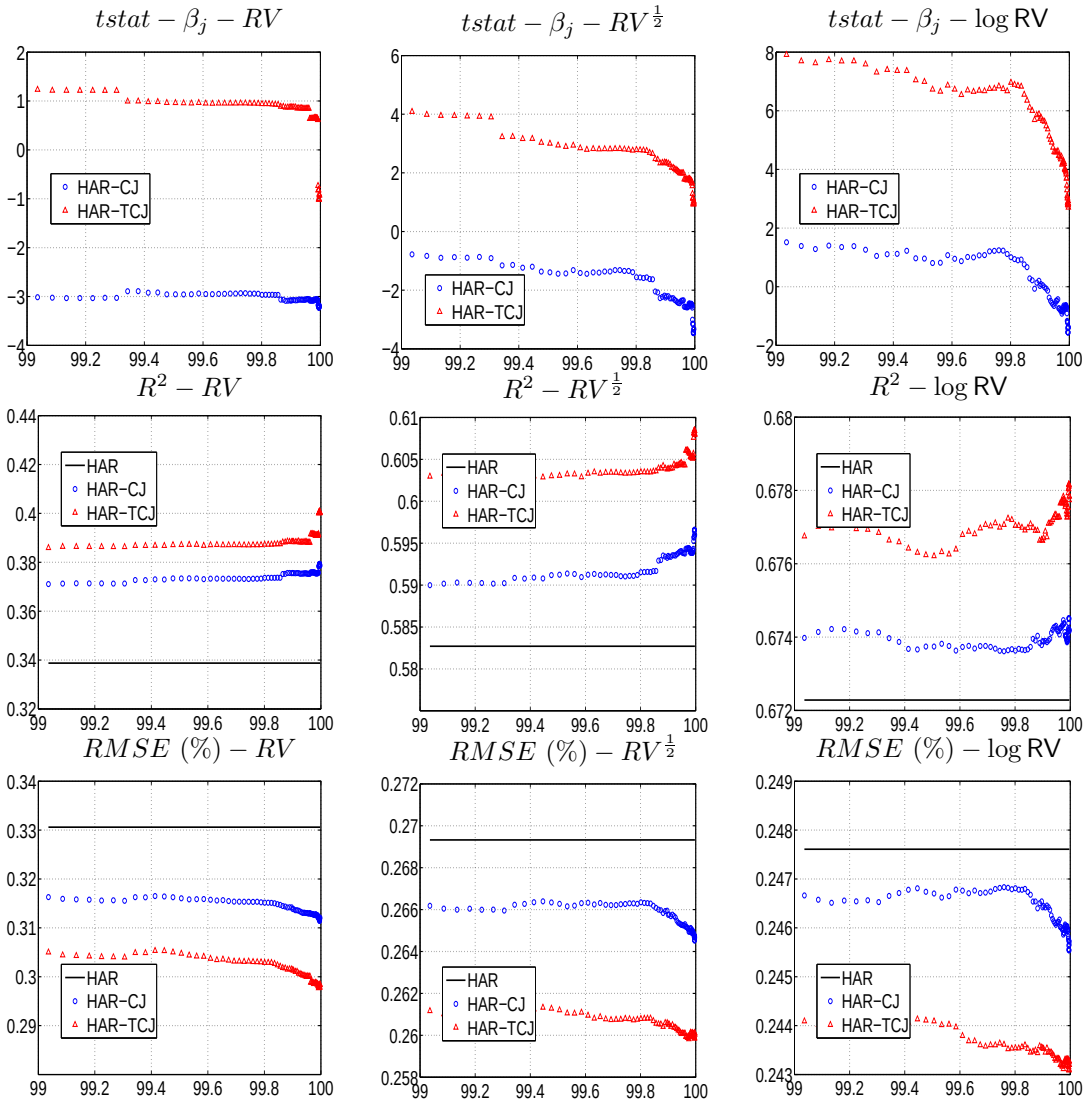

Figure 6.4: Reports the t statistics of the coefficient β_j, the R^2 and the $RMSE$ for the three models estimated on S&P 500 data, for both the HAR-CJ and HAR-TCJ versions, as a function of the confidence level used for detecting jumps with the $\mathsf{C} - \mathsf{Tz}$ statistics.

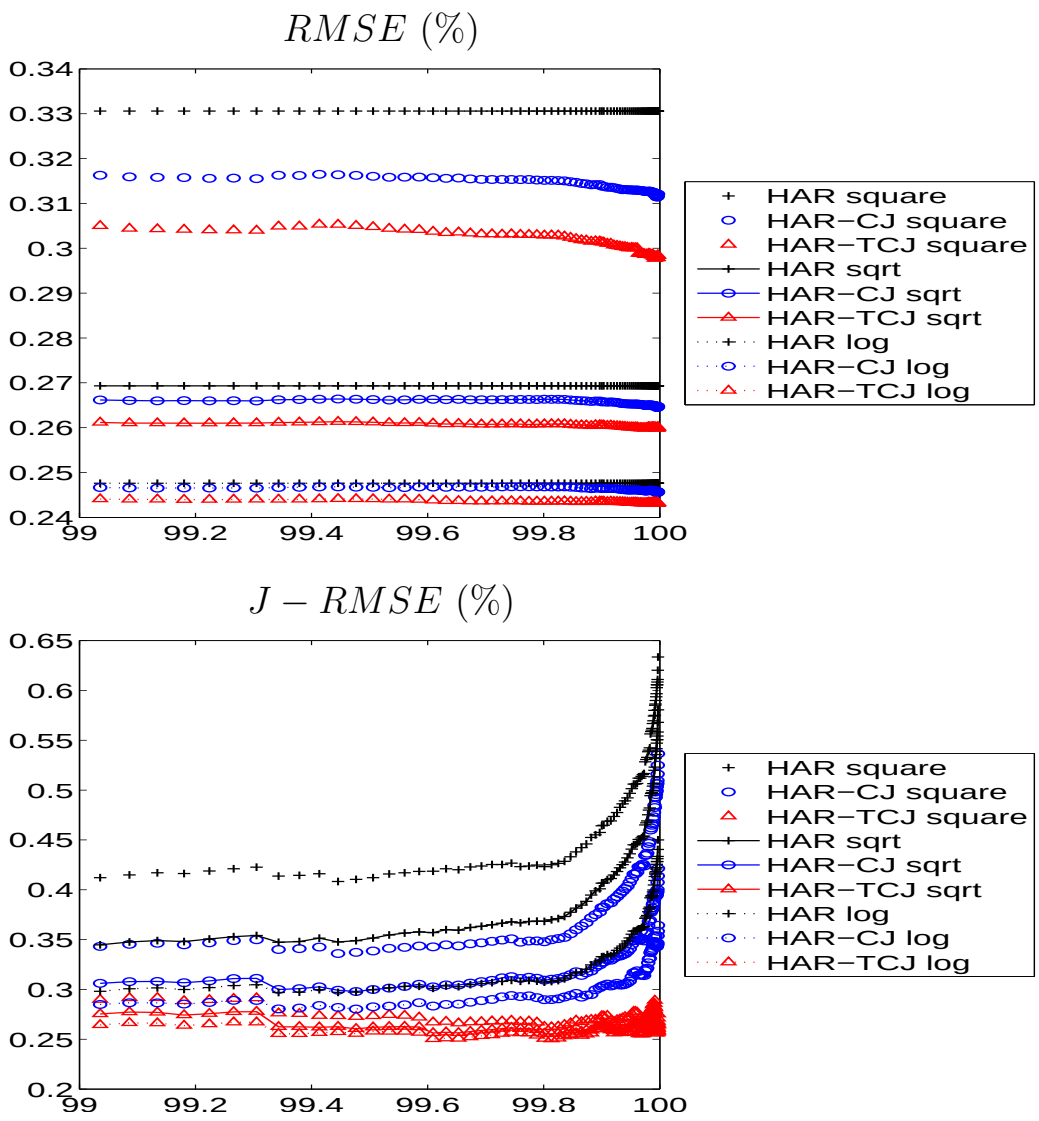

Figure 6.5: Top: $RMSE$ of daily forecasts as a function of the confidence intervals, for all the considered models. Bottom: $J-RMSE$, that is the mean square error for days immediately after a jump, of daily forecasts for all the considered models.

models, which in turn outperform the square models. However, for all the three specifications the best performance is obtained by the $HAR-TCJ$ models which employ threshold bipower variation.

Most importantly, in days following a jump the ranking based on $RMSE$ depends on the measure adopted for continuous variation. Indeed, the $HAR-TCJ$ models provide the smaller $RMSE$, either in the logarithmic, square root or square specification, and their performance does not decline with increasing confidence level.

For the sake of shortness we do not report here the analysis with higher values of c_ϑ. In summary, it is found that the β_j coefficient of the TCJ specification is mildly significant for $c_\vartheta = 4$ and not significant for $c_\vartheta = 5$. This is not surprising, since as we increase c_ϑ we get closer to the results obtained with bipower variation. We also estimate the HAR-RV-CJ model using the jumps detected via the z statistics (4.34), as in ABD, and compare it with the HAR-RV-TCJ model estimated with the jumps detected via the $\mathsf{C}-\mathsf{Tz}$

statistics (4.26). The results, omitted here for brevity, show that in this case the difference between the two models is even higher and that the R^2 and $RMSE$ of the HAR-RV-CJ model are drastically worse for higher α, signaling the difficulty of bipower variation based test to identify larger jumps. When using the same jumps detected with the less performing z statistics, results are qualitatively unchanged, but with milder significance of jumps and with the two models performing more closely for the highest confidence levels.

We also estimate the forecasting models using weekly realized volatility and monthly realized volatility as dependent variables. Results are shown in Tables 6.2 and 6.3, respectively. When forecasting weekly and monthly volatility, the β_j of the HAR-RV-CJ model tends to be negative, sometimes significantly. Instead, for the HAR-RV-TCJ model, the β_j are largely positive and significant in the square root and log model, and insignificant in the square model. Again, the R^2 and the $RMSE$ confirm, in days following a jump, the better forecasting ability of the HAR-RV-TCJ model, which is not worse than HAR-RV-CJ in days not following a jump. Thus, the forecasting ability of the jump component, when it is measured by means of threshold bipower variation, extends to a time span of at least one month.

Thus, our empirical findings further corroborate the theoretical and simulation results in the previous sections on the superior performance of the threshold method in separating and estimating the continuous and jump components of the price process. Moreover, they show that, once the two components are correctly measured and separated, the impact of past jumps on future realized volatility is positive and significant.

6.3.2 Individual stocks

We analyze a sample of six individual stocks, chosen among the most liquid stocks of S&P500. The stocks are Alcoa (aa), Citigroup (c), Intel (intc), Microsoft (msft), Pfizer (pfe) and Exxon-Mobil (xom). Our sample starts on 2 January 2001 and ends on 30 December 2005, containing $1,256$ days per stock. Since these stocks are traded very actively, we still use a sampling frequency of $\delta = 5$ minutes, corresponding to 78 returns per day.

To save space, we focus on the most important quantities (the significance of the jump and the R^2 and $RMSE$ of the forecasting model both on the whole sample and conditioned on days after the occurrence of a jump), which are reported in Table 6.3.2 for the square root model and Table 6.3.2 for the logarithmic model. We report results for $\alpha = 99.9\%$ and $\alpha = 99.99\%$.

The results are still clearly cut, even if the sample size of individual stocks is reduced by more than one third with respect to S&P500 data. Jumps have a substantial impact in determining future volatility, when this effect is measured by means of threshold multipower variation. The t statistics of the β_j coefficient is always larger for the HAR-TCJ model than for the HAR-CJ model, and mostly significant both for the square root and the logarithmic model. On the whole sample, the performance, measured in terms of R^2 and $RMSE$, of the two models is practically the same, but conditioned on the occurrence of a jump, there is a clear advantage in using the HAR-TCJ model, especially in the square root specification.

Thus, the results obtained on the S&P500 portfolio are replicated on its most liquid constituents, indicating that the impact of jumps on future volatility is not peculiar to the S&P500 returns considered

Daily S&P500 Regression ($\mathsf{C}-\mathsf{Tz}$ statistics)

$$\text{HAR:} \quad \mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \mathsf{RV}_t + \beta_w \mathsf{RV}_{t-5:t} + \beta_m \mathsf{RV}_{t-22:t} + \varepsilon_t$$
$$\text{HAR-CJ:} \quad \mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \widehat{C}_t + \beta_w \widehat{C}_{t-5:t} + \beta_m \widehat{C}_{t-22:t} + \beta_j \widehat{J}_t + \varepsilon_t$$
$$\text{HAR-TCJ:} \quad \mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \widehat{TC}_t + \beta_w \widehat{TC}_{t-5:t} + \beta_m \widehat{TC}_{t-22:t} + \beta_j \widehat{TJ}_t + \varepsilon_t$$

	RV_{t+1}			$\mathsf{RV}_{t+1}^{1/2}$			$log\mathsf{RV}_{t+1}$		
	HAR	HAR-CJ	HAR-TCJ	HAR	HAR-CJ	HAR-TCJ	HAR	HAR-CJ	HAR-TCJ
β_0	34.200	26.762	23.259	0.981	0.901	0.767	0.252	0.272	0.284
	(3.771)	(3.363)	(2.753)	(3.951)	(3.795)	(3.298)	(4.457)	(4.866)	(5.168)
β_d	0.220	0.378	0.420	0.323	0.371	0.421	0.334	0.336	0.356
	(2.329)	(5.736)	(6.170)	(6.347)	(8.958)	(11.954)	(13.138)	(13.140)	(14.829)
β_w	0.321	0.263	0.298	0.336	0.317	0.307	0.358	0.356	0.341
	(3.821)	(3.157)	(2.609)	(6.075)	(5.811)	(5.427)	(9.463)	(9.398)	(9.421)
β_m	0.313	0.288	0.253	0.269	0.260	0.238	0.257	0.257	0.253
	(4.817)	(4.786)	(3.852)	(6.842)	(6.725)	(6.103)	(8.873)	(8.959)	(9.067)
β_j		-0.581	0.045		-0.101	0.096		0.007	0.055
		(-2.968)	(0.925)		(-1.626)	(2.653)		(0.683)	(6.384)
R^2	0.339	0.374	0.387	0.583	0.592	0.604	0.679	0.681	0.684
MSE	0.330	0.315	0.302	0.269	0.266	0.260	0.247	0.246	0.243
J-R^2	0.148	0.196	0.306	0.373	0.385	0.480	0.607	0.607	0.643
J-MSE	0.437	0.359	0.268	0.381	0.317	0.261	0.317	0.296	0.256
C-R^2	0.375	0.394	0.399	0.609	0.613	0.616	0.686	0.687	0.688
C-MSE	0.322	0.311	0.304	0.259	0.262	0.260	0.242	0.242	0.242

Table 6.1: OLS estimate for daily ($h = 1$) HAR, HAR-CJ and HAR-TCJ volatility forecast regressions for S&P500 futures from January 1990 to December 2004 (3,736 observations). The significant daily jump are computed using a critical value of $\alpha = 99.9\%$ and the $\mathsf{C}-\mathsf{Tz}$ statistics computed with $c_\vartheta = 3$. Reported in parenthesis are the t-statistics based on Newey-West correction with order 5.

Weekly S&P500 Regression

$$\text{HAR:} \quad \mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \mathsf{RV}_t + \beta_w \mathsf{RV}_{t-5:t} + \beta_m \mathsf{RV}_{t-22:t} + \varepsilon_t$$
$$\text{HAR-CJ:} \quad \mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \widehat{C}_t + \beta_w \widehat{C}_{t-5:t} + \beta_m \widehat{C}_{t-22:t} + \beta_j \widehat{J}_t + \varepsilon_t$$
$$\text{HAR-TCJ:} \quad \mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \widehat{TC}_t + \beta_w \widehat{TC}_{t-5:t} + \beta_m \widehat{TC}_{t-22:t} + \beta_j \widehat{TJ}_t + \varepsilon_t$$

	RV_{t+1}			$\mathsf{RV}_{t+1}^{1/2}$			$log\mathsf{RV}_{t+1}$		
	HAR	HAR-CJ	HAR-TCJ	HAR	HAR-CJ	HAR-TCJ	HAR	HAR-CJ	HAR-TCJ
β_0	47.231	41.199	37.791	1.532	1.457	1.340	0.403	0.422	0.435
	(4.324)	(3.976)	(3.375)	(4.306)	(4.223)	(3.937)	(4.790)	(5.066)	(5.332)
β_d	0.097	0.190	0.210	0.176	0.213	0.244	0.205	0.210	0.229
	(1.892)	(4.858)	(4.402)	(5.360)	(7.925)	(9.392)	(11.742)	(12.129)	(13.815)
β_w	0.367	0.351	0.392	0.368	0.355	0.352	0.358	0.350	0.330
	(4.676)	(4.298)	(3.564)	(6.145)	(5.995)	(5.858)	(8.179)	(8.081)	(8.001)
β_m	0.335	0.320	0.305	0.344	0.339	0.330	0.357	0.359	0.362
	(4.901)	(4.375)	(3.707)	(6.390)	(6.203)	(6.060)	(8.601)	(8.779)	(9.205)
β_j		-0.394	0.007		-0.105	0.040		-0.005	0.031
		(-2.570)	(0.378)		(-2.425)	(2.465)		(-0.689)	(6.190)
R^2	0.499	0.534	0.554	0.690	0.700	0.709	0.768	0.770	0.772
MSE	0.273	0.260	0.252	0.205	0.202	0.199	0.187	0.186	0.185
J-R^2	0.475	0.464	0.445	0.649	0.639	0.632	0.733	0.732	0.729
J-MSE	0.277	0.267	0.265	0.217	0.218	0.219	0.209	0.209	0.210
C-R^2	0.506	0.545	0.568	0.695	0.706	0.716	0.771	0.773	0.776
C-MSE	0.273	0.259	0.251	0.205	0.201	0.198	0.186	0.184	0.183

Table 6.2: OLS estimate for weekly ($h = 5$) HAR, HAR-CJ and HAR-TCJ volatility forecast regressions for S&P500 futures from January 1990 to December 2004 (3,736 observations). The significant daily jump are computed using a critical value of $\alpha = 99.9\%$. Reported in parenthesis are the t-statistics based on Newey-West correction with order 10.

Monthly S&P500 Regression

$$\text{HAR:} \quad \mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \mathsf{RV}_t + \beta_w \mathsf{RV}_{t-5:t} + \beta_m \mathsf{RV}_{t-22:t} + \varepsilon_t$$
$$\text{HAR-CJ:} \quad \mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \widehat{C}_t + \beta_w \widehat{C}_{t-5:t} + \beta_m \widehat{C}_{t-22:t} + \beta_j \widehat{J}_t + \varepsilon_t$$
$$\text{HAR-TCJ:} \quad \mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \widehat{TC}_t + \beta_w \widehat{TC}_{t-5:t} + \beta_m \widehat{TC}_{t-22:t} + \beta_j \widehat{TJ}_t + \varepsilon_t$$

	RV_{t+1}			$\mathsf{RV}_{t+1}^{1/2}$			$log\mathsf{RV}_{t+1}$		
	HAR	HAR-CJ	HAR-TCJ	HAR	HAR-CJ	HAR-TCJ	HAR	HAR-CJ	HAR-TCJ
β_0	78.416	73.455	70.663	2.872	2.797	2.688	0.753	0.767	0.774
	(5.896)	(5.623)	(4.948)	(5.872)	(5.947)	(5.862)	(5.269)	(5.469)	(5.680)
β_d	0.061	0.124	0.129	0.109	0.135	0.149	0.126	0.130	0.138
	(2.555)	(5.292)	(4.875)	(5.914)	(8.403)	(9.179)	(10.053)	(10.516)	(11.717)
β_w	0.219	0.206	0.242	0.279	0.272	0.279	0.266	0.261	0.250
	(4.274)	(3.742)	(4.344)	(5.547)	(5.354)	(5.366)	(6.098)	(6.019)	(5.848)
β_m	0.386	0.386	0.385	0.401	0.400	0.397	0.458	0.460	0.466
	(4.601)	(4.316)	(4.028)	(6.392)	(6.325)	(6.405)	(9.221)	(9.322)	(9.795)
β_j		-0.284	-0.002		-0.091	0.018		-0.010	0.016
		(-2.892)	(-0.114)		(-3.273)	(1.441)		(-1.738)	(3.660)
R^2	0.471	0.500	0.517	0.650	0.659	0.668	0.739	0.743	0.747
MSE	0.300	0.290	0.285	0.209	0.206	0.205	0.187	0.186	0.185
J-R^2	0.533	0.505	0.503	0.662	0.656	0.656	0.713	0.719	0.721
J-MSE	0.308	0.301	0.302	0.221	0.218	0.219	0.197	0.195	0.194
C-R^2	0.467	0.500	0.519	0.649	0.660	0.669	0.741	0.745	0.748
C-MSE	0.299	0.289	0.284	0.209	0.205	0.204	0.187	0.185	0.184

Table 6.3: OLS estimate for monthly ($h = 22$) HAR, HAR-CJ and HAR-TCJ volatility forecast regressions for S&P500 futures from January 1990 to December 2004 (3,736 observations). The significant daily jump are computed using a critical value of $\alpha = 99.9\%$. Reported in parenthesis are the t-statistics based on Newey-West correction with order 44.

in the previous Section, and suggesting that it may simply come from aggregation, at the portfolio level, of the same effect at the individual stock level.

6.3.3 Bond data and the no-trade bias

Finally, we use a sample of 30-years US Treasury Bond futures from January 1990 to October 2003 for a total of $3,231$ daily data points. All the relevant volatility and jump measures are computed again with five-minutes returns, corresponding to 84 returns per day.

The first thing we note on bond data is an unnaturally high number of jumps. At the 99.9% confidence level, the $\mathsf{C}-\mathsf{Tz}$ statistics detects 570 jumps, corresponding to the 17.6 % of our sample. Visual inspection of time series data reveals that in most of these days there are not jumps at all, but many intervals with zero return instead.

The problem hinges from what we could call *the no-trade bias* of bipower variation. This can be explained as follows. Suppose that data are not recorded continuously but, more realistically, that they are recorded discretely. Denote by $\bar{\delta}$ the minimum distance between two subsequent observations. By construction, if $\delta < \bar{\delta}$, then $\mathsf{MPV}_\delta = 0$ identically! Clearly, also $\mathsf{TMPV}_\delta = 0$. This simple reasoning also explains why the presence of null returns caused by absence of trading (stale price) in that interval induces a downward bias in multipower variation measures. Note that realized volatility is immune from this bias instead. Moreover, this bias has a larger impact on the jump detecting statistics, pointing toward rejection of the null. For example, when considering the z statistics, this bias lowers both the BPV and TriPV measure, with the joint effect of increasing the numerator and decreasing the denominator, thus increasing the statistics considerably.

In our case, this problem does not affect the S&P500 index, neither the stocks considered in our empirical analysis. However, it may affect US bond data, which are largely less liquid. Indeed, the percentage of zero 5-minutes return in bond data is very high, nearly 30%.

We accommodate this problem as follows: for bond data, we compute the $\mathsf{C}-\mathsf{Tz}$ statistics using only returns different from zero. Clearly, this biases the test toward the null, meaning that the detected jumps are those which have a larger impact for the statistics. With this correction the number of significant jumps with $\alpha = 99.9\%$ reduced to 112 representing 3.4% of the sample.

Relevant estimation results for bond data when forecasting daily, weekly and monthly volatility are shown in Table 6.6 for $\alpha = 99.9\%$. Corresponding quantities for other values of α are shown in Figure 6.6. We find that the HAR-TCJ model outperforms the HAR-CJ model. This is true even if the impact of the jump on future volatility is generally insignificant, but nevertheless not negatively significant as for the HAR-CJ estimates. An explanation for this finding might be that jumps in the bond market are less "surprising" with respect to those in the stock markets, since most of them are related to scheduled macroeconomic announcements. Indeed, related studies, for example Bollerslev et al. (2000), find two intraday spikes at hours in which announcements are released. However, regarding forecasting, we confirm the results on S&P500 and individual stocks.

Summarizing, our empirical findings further corroborate the theoretical and simulation results in the previous sections on the superior performance of the threshold method in separating and estimating the

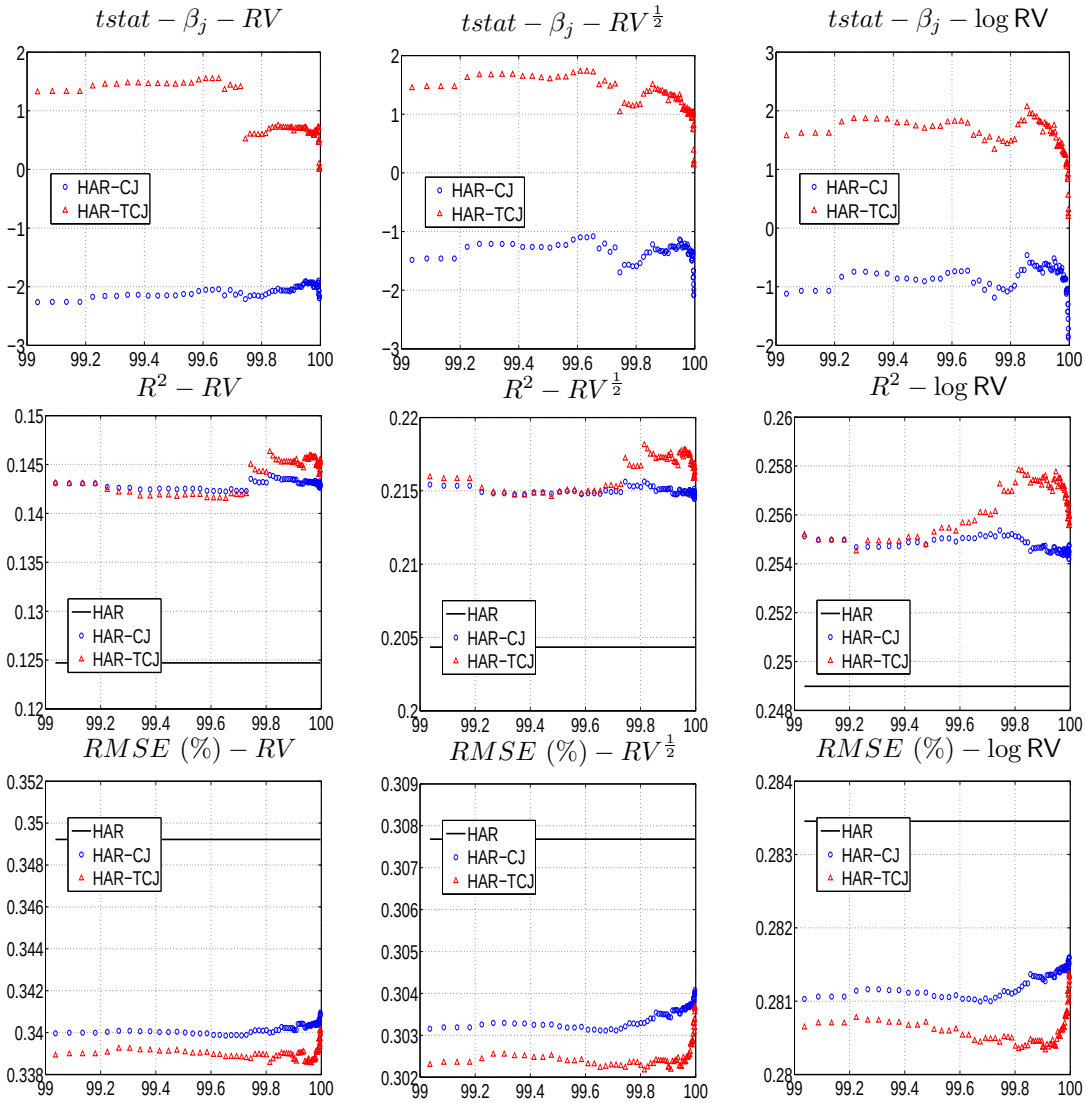

Figure 6.6: Reports the t statistics of the coefficient β_j, the R^2 and the $RMSE$ for the three models estimated on US Bond data, for both the HAR-CJ and HAR-TCJ versions, as a function of the confidence level used for detecting jumps with the $\mathsf{C} - \mathsf{Tz}$ statistics.

continuous and jump components of the price process. Moreover, they show that, once the two components are correctly measured and separated, the impact of past jumps on future realized volatility is positive and significant.

HAR: $\mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \mathsf{RV}_t + \beta_w \mathsf{RV}_{t-5:t} + \beta_m \mathsf{RV}_{t-22:t} + \varepsilon_t$

HAR-CJ: $\mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \hat{C}_t + \beta_w \hat{C}_{t-5:t} + \beta_m \hat{C}_{t-22:t} + \beta_j \hat{J}_t + \varepsilon_t$

HAR-TCJ: $\mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \hat{TC}_t + \beta_w \hat{TC}_{t-5:t} + \beta_m \hat{TC}_{t-22:t} + \beta_j \hat{TJ}_t + \varepsilon_t$

Stock	α (%)	Jumps	β_j t-stat		R^2 (J-R^2)			$RMSE$ (J-$RMSE$)		
			HAR-CJ	HAR-TCJ	HAR	HAR-CJ	HAR-TCJ	HAR	HAR-CJ	HAR-TCJ
aa	99.9	121	1.086	3.684	0.597	0.601	0.596	0.247	0.245	0.245
					(0.534)	(0.559)	(0.569)	(0.314)	(0.292)	(0.289)
	99.99	68	-0.147	2.473	0.597	0.601	0.597	0.247	0.245	0.245
					(0.411)	(0.454)	(0.484)	(0.319)	(0.262)	(0.257)
c	99.9	105	-0.458	3.494	0.723	0.731	0.733	0.248	0.244	0.244
					(0.770)	(0.775)	(0.820)	(0.324)	(0.239)	(0.217)
	99.99	59	-0.937	3.139	0.723	0.732	0.731	0.248	0.244	0.244
					(0.830)	(0.854)	(0.856)	(0.361)	(0.231)	(0.219)
intc	99.9	78	2.706	4.925	0.774	0.774	0.773	0.226	0.226	0.226
					(0.828)	(0.833)	(0.837)	(0.227)	(0.218)	(0.214)
	99.99	43	1.483	3.461	0.774	0.774	0.774	0.226	0.226	0.225
					(0.786)	(0.794)	(0.803)	(0.257)	(0.235)	(0.230)
msft	99.9	92	0.611	2.721	0.748	0.749	0.749	0.242	0.242	0.241
					(0.752)	(0.753)	(0.760)	(0.311)	(0.279)	(0.265)
	99.99	48	0.661	2.567	0.748	0.748	0.748	0.242	0.242	0.241
					(0.652)	(0.643)	(0.674)	(0.363)	(0.332)	(0.311)
pfe	99.9	131	0.500	3.109	0.481	0.488	0.491	0.287	0.281	0.281
					(0.630)	(0.650)	(0.680)	(0.368)	(0.297)	(0.272)
	99.99	83	0.877	2.968	0.481	0.487	0.489	0.287	0.282	0.281
					(0.609)	(0.651)	(0.677)	(0.407)	(0.318)	(0.289)
xom	99.9	98	0.051	2.892	0.674	0.678	0.681	0.237	0.236	0.235
					(0.791)	(0.778)	(0.806)	(0.239)	(0.205)	(0.191)
	99.99	54	0.108	2.411	0.674	0.677	0.680	0.237	0.235	0.235
					(0.749)	(0.730)	(0.776)	(0.289)	(0.246)	(0.224)

Table 6.4: Reports number of jumps, t-stat on β_j, R^2, $RMSE$, $J - R^2$ and $J - RMSE$ for daily ($h = 1$) **square root** version of HAR, HAR-CJ and HAR-TCJ volatility forecast regressions on six single stocks . The significant daily jump are computed using a critical value of $\alpha = 0.999$ and $\alpha = 0.9999$ as reported, with the $\mathsf{C} - \mathsf{Tz}$ statistics.

HAR: $\mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \mathsf{RV}_t + \beta_w \mathsf{RV}_{t-5:t} + \beta_m \mathsf{RV}_{t-22:t} + \varepsilon_t$

HAR-CJ: $\mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \hat{C}_t + \beta_w \hat{C}_{t-5:t} + \beta_m \hat{C}_{t-22:t} + \beta_j \hat{J}_t + \varepsilon_t$

HAR-TCJ: $\mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \hat{TC}_t + \beta_w \hat{TC}_{t-5:t} + \beta_m \hat{TC}_{t-22:t} + \beta_j \hat{TJ}_t + \varepsilon_t$

Stock	α (%)	Jumps	β_j t-stat		R^2 (J-R^2)			$RMSE$ (J-$RMSE$)		
			HAR-CJ	HAR-TCJ	HAR	HAR-CJ	HAR-TCJ	HAR	HAR-CJ	HAR-TCJ
aa	99.9	121	0.684	2.618	0.581	0.585	0.577	0.234	0.232	0.234
					(0.531)	(0.540)	(0.529)	(0.283)	(0.267)	(0.261)
	99.99	68	-0.556	1.818	0.581	0.585	0.578	0.234	0.232	0.234
					(0.436)	(0.463)	(0.460)	(0.279)	(0.246)	(0.244)
c	99.9	105	0.069	4.105	0.815	0.818	0.817	0.233	0.230	0.231
					(0.792)	(0.804)	(0.825)	(0.281)	(0.230)	(0.209)
	99.99	59	-0.202	4.183	0.815	0.819	0.818	0.233	0.230	0.230
					(0.810)	(0.828)	(0.845)	(0.304)	(0.229)	(0.212)
intc	99.9	78	3.101	4.908	0.804	0.804	0.800	0.214	0.215	0.216
					(0.823)	(0.829)	(0.828)	(0.215)	(0.199)	(0.206)
	99.99	43	1.866	3.958	0.804	0.804	0.802	0.214	0.214	0.215
					(0.790)	(0.807)	(0.813)	(0.244)	(0.218)	(0.225)
msft	99.9	92	0.197	2.130	0.796	0.796	0.795	0.227	0.227	0.227
					(0.738)	(0.728)	(0.741)	(0.286)	(0.262)	(0.246)
	99.99	48	0.456	1.859	0.796	0.795	0.794	0.227	0.227	0.227
					(0.605)	(0.593)	(0.620)	(0.335)	(0.313)	(0.294)
pfe	99.9	131	0.422	2.987	0.540	0.545	0.544	0.263	0.261	0.261
					(0.574)	(0.583)	(0.601)	(0.310)	(0.268)	(0.241)
	99.99	83	0.420	2.612	0.540	0.544	0.543	0.263	0.261	0.260
					(0.526)	(0.550)	(0.588)	(0.339)	(0.283)	(0.253)
xom	99.9	98	0.604	2.687	0.681	0.683	0.681	0.234	0.233	0.234
					(0.768)	(0.764)	(0.767)	(0.213)	(0.194)	(0.188)
	99.99	54	0.451	2.147	0.681	0.683	0.682	0.234	0.233	0.233
					(0.720)	(0.708)	(0.716)	(0.256)	(0.229)	(0.216)

Table 6.5: Reports number of jumps, t-stat on β_j, R^2, $RMSE$, $J - R^2$ and $J - RMSE$ for daily ($h = 1$) **logarithmic** version of HAR, HAR-CJ and HAR-TCJ volatility forecast regressions on six single stocks . The significant daily jump are computed using a critical value of $\alpha = 0.999$ and $\alpha = 0.9999$ as reported, with the $\mathsf{C} - \mathsf{Tz}$ statistics.

US Bond Regressions ($\mathsf{C}-\mathsf{Tz}$ statistics)

HAR: $\mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \mathsf{RV}_t + \beta_w \mathsf{RV}_{t-5:t} + \beta_m \mathsf{RV}_{t-22:t} + \varepsilon_t$

HAR-CJ: $\mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \widehat{C}_t + \beta_w \widehat{C}_{t-5:t} + \beta_m \widehat{C}_{t-22:t} + \beta_j \widehat{J}_t + \varepsilon_t$

HAR-TCJ: $\mathsf{RV}_{t+1:t+h} = \beta_0 + \beta_d \widehat{TC}_t + \beta_w \widehat{TC}_{t-5:t} + \beta_m \widehat{TC}_{t-22:t} + \beta_j \widehat{TJ}_t + \varepsilon_t$

	RV_{t+1}			$\mathsf{RV}_{t+1}^{1/2}$			$log\mathsf{RV}_{t+1}$		
	HAR	HAR-CJ	HAR-TCJ	HAR	HAR-CJ	HAR-TCJ	HAR	HAR-CJ	HAR-TCJ
					Daily forecasts				
β_j		-0.095	0.017		-0.038	0.029		-0.010	0.022
		(-2.050)	(0.711)		(-1.339)	(1.355)		(-0.692)	(1.813)
R^2	0.124	0.143	0.145	0.204	0.215	0.217	0.250	0.255	0.258
MSE	0.349	0.340	0.339	0.308	0.303	0.302	0.283	0.281	0.280
J-R^2	0.034	0.056	0.051	0.067	0.077	0.078	0.088	0.086	0.097
J-MSE	0.486	0.375	0.343	0.413	0.337	0.326	0.355	0.316	0.309
C-R^2	0.133	0.146	0.148	0.211	0.219	0.222	0.256	0.261	0.264
C-MSE	0.343	0.339	0.339	0.303	0.302	0.301	0.280	0.280	0.279
					Weekly forecasts				
β_j		-0.071	0.012		-0.050	0.013		-0.026	0.007
		(-3.182)	(1.324)		(-3.363)	(1.301)		(-3.225)	(1.101)
R^2	0.295	0.336	0.343	0.415	0.439	0.445	0.472	0.487	0.492
MSE	0.206	0.197	0.196	0.174	0.170	0.169	0.163	0.161	0.160
J-R^2	0.126	0.146	0.156	0.261	0.277	0.289	0.334	0.340	0.353
J-MSE	0.162	0.159	0.158	0.176	0.175	0.177	0.197	0.197	0.199
C-R^2	0.312	0.356	0.365	0.426	0.451	0.458	0.480	0.494	0.500
C-MSE	0.208	0.199	0.197	0.174	0.170	0.169	0.162	0.160	0.159
					Monthly forecasts				
β_j		-0.059	0.001		-0.049	-0.002		-0.026	-0.002
		(-3.999)	(0.206)		(-4.597)	(-0.279)		(-4.372)	(-0.537)
R^2	0.333	0.383	0.396	0.433	0.466	0.481	0.488	0.509	0.522
MSE	0.168	0.160	0.157	0.135	0.131	0.130	0.128	0.126	0.126
J-R^2	0.334	0.345	0.357	0.430	0.454	0.482	0.467	0.487	0.513
J-MSE	0.145	0.138	0.130	0.122	0.119	0.116	0.124	0.123	0.124
C-R^2	0.332	0.384	0.398	0.433	0.466	0.481	0.488	0.509	0.522
C-MSE	0.169	0.160	0.158	0.136	0.131	0.130	0.128	0.127	0.126

Table 6.6: OLS (partial) estimates for daily ($h = 1$), weekly ($h = 5$), monthly ($h = 22$) HAR, HAR-CJ and HAR-TCJ volatility forecast regressions for US Bond from January 1990 to December 2004 (3,736 observations). The significant daily jump are computed using a critical value of $\alpha = 99.9\%$ and the $\mathsf{C}-\mathsf{Tz}$ statistics. Reported in parenthesis are the t-statistics based on Newey-West correction.

6.4 Conclusions

We show that dividing volatility into jumps and continuous variation yields a substantial improvement in volatility forecasting, because of the positive impact of past jumps on future volatility. This important result has been obscured in the literature since, in finite samples, measures based on multipower variation are largely biased in the presence of jumps.

The models we propose provide a superior forecasting ability, especially in days which follow the occurrence of a jump. The forecasting power on subsequent volatilities extends for a period of at least one month. Clearly, this finding can be of great importance for risk management and other financial applications involving volatility estimation. Recently, the financial econometrics literature focused on the separation of the quadratic variation in its discontinuous and continuous part. While ingenuous, sophisticated and fascinating theories have been contrived to this purpose, these had still little impact on empirical applications. Our study contributes to showing that the above mentioned separation theories (multipower variation, threshold estimation) can indeed be very useful in practical applications, since the two components seem to have different dynamics, with the continuous one being determined endogenously in the market by heterogeneous agents, and the discontinuous one being exogenous and basically unpredictable. Moreover, the correlation between jumps and future volatility can be helpful not only for practical applications, but also in the comprehension of the price formation mechanism. Still, an economic theory explaining our findings may represent an interesting direction for future research.

Chapter 7

Thesis Summary

In the first part of the thesis we have shown, as anticipated in the introduction, that a full thermodynamic description of production technology drastically changes the equilibrium established by the neoclassical approach. Focusing on a the production of a particular commodity we find that perfect substitutability between production inputs is impossible: a minimum amount of energy is required to carry out production. Moreover our approach reveals that irreversibility plays a crucial role in economic equilibrium: in an irreversible technology a producer forced to price goods in terms of energy retains part of available energy as firm inventory. Energy is used in a more parsimonious way. On the contrary, in a reversible technology, all energy is spent for production. Further, energy and finished good are equivalent, they can be converted one into another. As a consequence, reversibility leads to the numeraire-independent neoclassical equilibrium. In this sense our approach shows that neoclassical approach assumes that every transformation is carried out reversibly, which is far away from actual production technologies. Our approach is enforced by similar results obtained by Roma (2006) for the production of a quite simple commodity (hot water).

The second part is very different in nature and focuses on high-frequency data of financial markets. In particular we are interested in the analysis of discontinuous variations of asset prices. In this context, we have developed powerful volatility estimators robust to the presence of jumps. Our technique allows for the construction of highly performing jump detection tests and provides accurate estimates of continuous volatility. Moreover better estimates of detrended fluctuation function exponent are obtained. The separation of realized volatility in its continuous and discontinuous parts is used, in the final part, for the construction of a volatility forecasting model which reveals, for the first time to our knowledge, the positive impact of jumps on future volatility. Markets are switched in a turbulent state after a jump has occurred. For these reasons we believe that the techniques developed in this work could be useful for practitioners and constitute a valid instruments to risk management.

Bibliography

Abel, A. B. (1983). Energy price uncertainty and optimal factor intensity: a mean-variance analysis. *Econometrica 51(6)*, 1839–1845.

Abry, P., R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch (2002). Semiparametric spectral estimation for fractional processes. *IEEE Signal Processing Magazine*.

Aït-Sahalia, Y. (2004). Disentangling diffusion from jumps. *Journal of Financial Economics 74*, 487–528.

Aït-Sahalia, Y. and J. Jacod (2007). Volatility estimators for discretely sampled Lévy processes. *Annals of Statistics 35*, 355–392.

Aït-Sahalia, Y. and L. Mancini (2008). Out of sample forecasts of quadratic variation. *Journal of Econometrics*. Forthcoming.

Andersen, B., P. Salamon, and R. S. Berry (1977). Thermodynamics in finite time: extremals for imperfect heat engines. *Journal of Chemical Physics 66(4)*, 1571–1577.

Andersen, T., L. Benzoni, and J. Lund (2002). An empirical investigation of continuous-time equity return models. *Journal of Finance 57*, 1239–1284.

Andersen, T., T. Bollerslev, and F. Diebold (2007). Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics 89(4)*, 701720.

Andersen, T., T. Bollerslev, F. Diebold, and H. Ebens (2001). The distribution of realized stock volatility. *Journal of Financial Economics 61*, 43–76.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55.

Andersen, T., T. Bollerslev, and F. X. Diebold (2003). Parametric and nonparametric volatility measurement. In L. P. Hansen and Y. Aït-Sahalia (Eds.), *Handbook of Financial Econometrics*. Amsterdam: North-Holland.

Andersen, T., T. Bollerslev, and D. Dobrev (2006). No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effect and jumps: Theory and testable distributional implications. *Journal of Econometrics*. Forthcoming.

Arrow, K. J., C. H. Chenery, B. S., and R. M. Solow (1961). Capital-labor substitution and economic efficiency. *The Review of Economics and Statistics 43(3)*, 225–250.

Ashkenazy, Y., P. C. Ivanov, S. Havlin, and C. K. Peng (2001). Magnitude and sign correlations in heartbeat fluctuations. *Physical Review Letters 86(9)*, 1900(4).

Atkenson, A. and P. J. Kehoe (1980). Models of energy use: Putty-putty versus putty-clay. *The American Economic Review 89(4)*, 1028–1043.

Ausloos, M. and K. Ivanova (2003). Dynamical model and nonextensive statistical mechanics of a market index on large time windows. *Physical Review E 68(4)*, 046122(13).

Ayres, R. U. and A. V. Kneese (1969). Production, consumption, and externalities. *The American Economic Review 59(3)*, 282–297.

Ayres, R. U. and S. M. Miller (1980). The role of technical change. *Journal of Environmental Economics and Management 7*, 353–371.

Bandi, F. and T. Nguyen (2003). On the functional estimation of jump-diffusion models. *Journal of Econometrics 116*, 293–328.

Bandi, F. and P. Phillips (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica 71*(1), 241–283.

Banerjee, A. and G. Urga (2005). Modelling structural breaks, long memory and stock market volatility: an overview. *Journal of Econometrics. 129(1-2)*, 1–34.

Bardet, J. and I. Kammoun (2008). Asymptotic properties of the detrended fluctuation analysis of long range dependence processes. *IEEE Transactions on Information Theory 54(5)*, 2041–2052.

Barndorff-Nielsen, O., E. Graversen, J. Jacod, M. Podoslkij, and N. Shephard (2006). A central limit theorem for realised power and bipower variations of continuous semimartingales. In *From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev*, pp. 33–68.

Barndorff-Nielsen, O. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics 2*, 1–48.

Barndorff-Nielsen, O., N. Shephard, and M. Winkel (2006). Limit theorems for multipower variation in the presence of jumps. *Stochastic Processes and Their Applications 116*, 798–806.

Barndorff-Nielsen, O. E., E. Graversen, J. Jacod, and N. Shephard (2006). Limit theorems for bipower variation in financial econometrics,. *Econometric Theory 22*, 677–719.

Barndorff-Nielsen, O. E. and N. Shephard (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics 4*, 1–30.

Barndorff-Nielsen, O. E. and N. Shephard (2007). Variation, jumps, market frictions and high frequency data in financial econometrics. In *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress*, pp. 328–372. Cambridge University Press.

Bates, D. (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics 94*, 181–238.

Bejan, A. (1996). Entropy generation minimization: the method of thermodynamic optimization of finite-size systems and finite-time processes. *Boca Raton: CRC Press*.

Berndt, E. and D. O. Wood (1975). Technology, prices and the derived demand for energy. *The Review of Economics and Statistics 57(3)*, 259–268.

Berry, R. S., V. Kazakov, A. M. Tsirlin, S. Sieniutycz, and Z. Szwast (2000). Thermodynamic optimization of finite-time processes. *Wiley*.

Berry, R. S., P. Salamon, and G. Heal (1978). On a relation between economic and thermodynamic optima. *Resource and Energy 1(2)*, 125–137.

Bhattacharya, J., J. Edwards, A. N. Mamelak, and E. M. Schuman (2005). Long-range temporal correlations in the spontaneous spiking of neurons in the hippocampal−amygdala complex of humans. *Neuroscience 131*, 547–555.

Bjork, T. (1998). *Arbitrage theory in continuous time*. Wiley.

Bodo, G. and L. F. Signorini (1987). Short-term forecasting of the industrial production index. *International Journal of Forecasting 3(2)*, 245–259.

Bollerslev, T., J. Cai, and F. Song (2000). Intraday periodicity, long memory volatility and macroeconomic announcement effects in the US Treasury bond market. *Journal of Empirical Finance 7*, 37–55.

Bollerslev, T., U. Kretschmer, C. Pigorsch, and G. Tauchen (2007). A discrete-time model for daily S&P500 returns and realized variations: jumps and leverage effects. *Journal of Econometrics*. Forthcoming.

Brennan, M. K. and E. Schwartz (1985). Evaluating natural resource investments. *The Journal of Business 58(2)*, 135–157.

Brockwell, P. and R. Davis (2001). *Introduction to Time Series and Forecasting*. Springer Texts in Statistics.

Bunde, A., S. Havlin, J. W. Kantelhardt, T. Penzel, J. H. Peter, and K. Voigt (2000). Correlated and uncorrelated regions in heart-rate fluctuations during sleep. *Physical Review Letters 85(17)*, 3736(4).

Busch, T., B. Christensen, and M. Nielsen (2006). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Future Markets*. Working Paper, School of Economics and Management, University of Aarhus.

Chen, Z., P. C. Ivanov, K. Hu, and H. E. Stanley (2001). Effect of nonstationarities on detrended fluctuation analysis. *Physical Review E 65(4)*, 041107(15).

Chernov, M., R. Gallant, E. Ghysels, and G. Tauchen (2003). Alternative models for stock price dynamics. *Journal of Econometrics 116*(1), 225–258.

Christensen, K., R. Oomen, and M. Podoslkij (2008). Realised quantile-based estimation of the integrated variance. Working Paper.

Clements, M., A. Galvao, and J. Kim (2008). Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance*. Forthcoming.

Cleveland, C. J. (1992). Energy quality and energy surplus in the extraction of fossil fuel in the U.S. *Ecological Economics 6(2)*, 139–162.

Cleveland, C. J. (2005). Net energy from the extraction of oil and gas in the united states. *Energy 30(5)*, 769–782.

Cleveland, C. J. and M. Ruth (1997). When, where, and by how much do biophysical limits constrain the economic process? a survey of Nicholas Georgescu-Roegen's contribution to ecological economics. *Ecological Economics 22(3)*, 203–223.

Clewlow, L. and C. Strickland (2000). *Energy Derivatives: Pricing and Risk Management*. Lacima Publications.

Clower, R. (1967). A reconsideration of the microfoundations of monetary theory. *Western Economic Journal 6*, 1–9.

Colell, A. M., M. D. Whinston, and J. R. Green (1995). *Microeconomic Theory*. Oxford University Press.

Cont, R. and P. Tankov (2004). *Financial Modelling with Jump Processes*. Chapman & Hall - CRC.

Coratzar, G., E. S. Schwartz, and M. Salinas (1998). Evaluating environmental investments: a real options approach. *Management Science 44(8)*, 1059–1070.

Corsi, F. (2004). A simple long memory model of realized volatility. Working Paper.

Corsi, F., S. Mittnik, C. Pigorsch, and U. Pigorsch (2008). The volatility of realized volatility. *Econometric Reviews. 27(1-3)*, 1–33.

Corsi, F., D. Pirino, and R. Renò (2008). Volatility forecasting: the jumps do matter. *Working Paper*.

Cox, J., J. E. Ingersoll, and S. A. Ross (1985). An intertemporal general equilibrium model of asset prices. *Econometrica 53(2)*, 363–384.

Dacorogona, M. M., R. Genacy, R. B. Muller, U. Olsen, and O. V. Picet (2001). *An Introduction to High-Frequency Finance*. Academic Press, London.

Daly, E. and A. Porporato (2006). Probabilistic dynamics of some jump-diffusion systems. *Physical Review E 73*, 026108–7.

Daly, E. and A. Porporato (2007). Intertime jump statistics of state-dependent Poisson processes. *Physical Review E 75*, 011119–6.

Daly, H. E. (1992). Is the entropy law relevant to the economics of natural resource scarcity? Yes, of course it is! *Journal of Environmental Economics and Management 23(1)*, 91–95.

Daly, H. E. (1997). Georgescu-Roegen versus Solow/Stiglitz. *Ecological Economics 22(3)*, 267–268.

Das, S. (2002). The surprise element: Jumps in interest rates. *Journal of Econometrics 106*, 27–65.

De Sanctis, A. and C. Mari (2007). Modelling spikes in electricity markets using excitable dynamics. *Physica A 384*, 457–467.

Denbigh, K. G. (1956). The second law efficiency of chemical processes. *Chemical Engineering and Processing 6(1)*, 1–9.

Duffie, D., S. Gray, and P. Hoang (1998). Volatility in energy prices. In *Managing Energy Price Risk*. Risk Publications.

Duffie, D., J. Pan, and K. Singleton (2002). Transform analysis and asset pricing for affine junp-diffusions. *Econometrica 68*, 1343–1376.

Eisler, Z., J. Perelló, and J. Masoliver (2007). Volatility: A hidden Markov process in financial time series. *Physical review E 76(1)*, 056105(11).

Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance 59(3)*, 1367–1404.

Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in volatility and returns. *Journal of Finance 58*, 1269–1300.

Ethridge, D. (1973). The inclusion of wastes in the theory of the firm. *Journal of Political Economy 81(6)*, 1430–1441.

Eyeland, A. and H. Geman (1998). Pricing power derivatives. *Risk 11*, 71–73.

Fama, E. F. (1972). *Theory of Finance*. Holt, Rinehart and Winston.

Fan, J. and Y. Wang (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association 102*, 13491362.

Fan, J. and Q. Yao (2003). *Nonlinear time series*. Springer.

Fermi, E. (1956). *Thermodynamics*. Courier Dover Publications.

Field, B. C. and C. Grebenstein (1980). Capital-Energy substitution in U.S. manufacturing. *The Review of Economics and Statistics 62(2)*, 207–212.

Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability 30*, 790–804.

Forland, K. S., T. Forland, and S. K. Ratkje (1988). *Irreversible thermodynamics: theory and applications*. Wiley.

Forsberg, L. and E. Ghysels (2007). Why do absolute returns predict volatility so well? *Journal of Financial Econometrics 5(1)*, 31–67.

Fuss, M. A. (1977). Demand for energy in canadian manufacturing. *Journal of Econometrics 5(1)*, 89–116.

G., B. and A. C. L. F. Signorini (1991). Forecasting the Italian industrial production index in real time. *Journal of Forecasting 10(3)*, 285–299.

Gately, M. (2007). The EROI of U.S. offshore energy extraction: a net energy analysis of the Gulf of Mexico. *Ecological Economics 63(2-3)*, 355–364.

Geman, H. and A. Roncoroni (2006). Understanding the fine structure of electricity prices. *Journal of Business 79*(3), 1225–1261.

Georgescu-Roegen, N. (1971). *The entropy law and the economic process.* Harvard University Press, Cambridge, MA.

Georgescu-Roegen, N. (1979). Energy analysis and economic evaluation. *Southern Economic Journal 45(4)*, 1023–1058.

Ghashghaie, S., W. Breymann, J. Peinke, P. Talkner, and Y. Dodgeparalle (1996). Turbulent cascades in foreign exchange markets. *Letters to Nature 381*, 767 – 770.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics 131(1-2)*, 59–95. Forthcoming.

Gianfreda, A. (2008). Volatility and volume effects in European electricity spot markets. *Working Paper*.

Giot, P. and S. Laurent (2007). The information content of implied volatility in light of the jump/continuous decomposition of realized volatility. *Journal of Future Markets 27(4)*, 337–359.

Greinstein, G. L. and B. Mandelbrot (1964). Random walk models for the spike activity of a single neuron. *Biophysical Journal 4*, 41–68.

Griffin, J. M. and P. R. Gregory (1976). An intercountry translog model of energy substitution responses. *American Economic Review 66(5)*, 845–857.

Guercia, E., S. Ivaldia, S. Pastoreb, and S. Cincottia (2005). Modeling and implementation of an artificial electricity market using agent-based technology. *Physica A 355*, 69–76.

Guirguis, H. and F. Felder (2004). Further advances in forecasting day-ahead electricity prices using time series models. *KIEE International Transactions on PE 4-A*(3), 159–166.

Guthrie, G. and S. Videbeck (2007). Electricity spot price dynamics: beyond financial models. *Energy Policy 35(11)*, 56145621.

Hadsell, L., A. Marathe, and H. Shawky (2004). Estimating the volatility of wholesale electricity spot prices in the US. *The Energy Journal 25*(4), 23–40.

Hall, C., L. Mitchell, and J. Sloane (1979). Efficiency of energy delivery systems: I. An economic and energy analysis. *Environmental Management 3(6)*, 493–504.

Hardle, W. (1990). *Applied nonparametric regression*. Cambridge University Press.

Hartwick, J. (1970). *Investment, Interest, and Capital*. Prentice Hall.

Hartwick, J. (1978). Exploitation of many deposits of an exhaustible resource. *Econometrica 46(1)*, 201–217.

Hotelling, A. (1931). The economics of exhaustible resources. *Journal of Political Economics 39(2)*, 137–175.

Hu, K., P. C. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley (2001). Effect of trends on detrended fluctuation analysis. *Physical Review E 64(1)*, 011114(19).

Huang, X. and G. Tauchen (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics 3*(4), 456–499.

Huisman, R. and R. Mahieu (2003). Regime jumps in electricity prices. *Energy Economics 25*(5), 425–434.

Ikeda, N. and S. Watanabe (1981). *Stochastic differential equations and diffusion processes*. North Holland.

Islam, S. (1985). Effect of an essential input on isoquants and substitution elasticities. *Energy Economics 7(3)*, 194–196.

Jánosia, I. M., B. Janecskó, and I. Kondora (1999). Statistical analysis of 5 s index data of the Budapest Stock Exchange. *Physica A 269*, 111–124.

Jiang, G. (1998). Nonparametric modeling of U.S. interest rate term structure dynamics and implications on the prices of derivative securities. *Journal of Financial and Quantitative Analysis 33*(4), 465–497.

Jiang, G. and J. Knight (1997). A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model. *Econometric Theory 13*, 615–645.

Jiang, G. and R. Oomen (2006). A new test for jumps in asset prices. Working Paper.

Johannes, M. (2004). The statistical and economic role of jumps in continuous-time interest rate models. *Journal of Finance 59*, 227–260.

Kaldellis, J. K., D. S. Vlachou, and G. Korbakis (2005). Techno-economic evaluation of small hydro power plants in Greece: a complete sensitivity analysis. *Energy Policy 33(15)*, 1969–1985.

Khalil, E. L. (1990). Entropy law and exhaustion of natural resources: is Nicholas Georgescu-Roegen's paradigm defensible? *Ecological Economics 2(2)*, 163–178.

Khalil, E. L. (1991). Entropy law and Nicholas Georgescu-Roegen's paradigm: a reply. *Ecological Economics 3(2)*, 161–163.

Kim, I. M. and P. Loungani (1985). *The exergy method of thermal plant analysis*. Butterworths: Bourough Green.

Kim, I. M. and P. Loungani (1992). The role of energy in real business cycle models. *Journal of Monetary Economics 29(2)*, 173–189.

Knittel, C. R. and M. Roberts (2005). An empirical examination of deregulated electricity prices. *Energy Economics 27*(5), 791–817.

Kristensen, D. (2007). Nonparametric filtering of the realised spot volatility: a kernel-based approach. Working Paper.

Krysiak, C. F. (2006). Entropy, limits to growth, and the prospects for weak sustainability. *Ecological Economics 58(1)*, 182–191.

Krysiak, C. F. and D. Krysiak (2003). Production, consumption, and general equilibrium with physical constraints. *Journal of Environmental Economics and Management 46(3)*, 513–538.

Kummel, R. (1989). Energy as a factor of production and entropy as a pollution indicator in macroeconomic modelling. *Ecological Economics 1(2)*, 161–180.

Lee, S. and P. Mykland (2007). Jumps in financial markets: A new nonparametric test and jump clustering. *Review of Financial Studies.*. Forthcoming.

Leea, J.-M., D.-J. Kima, I.-Y. Kima, K.-S. Parkb, and S. I. Kim (2002). Detrended fluctuation analysis of EEG in sleep apnea using MIT/BIH polysomnography data. *Computers in Biology and Medicine 32*, 37–47.

Li, Y. and P. Flynn (2004). Deregulated power prices: comparison of volatility. *Energy Policy 32(14)*, 15911601.

Lillo, F. and R. N. Mantegna (2003). Power-law relaxation in a complex system: Omori law after a financial market crash. *Physical review E 68(1)*, 016119(5).

Liu, Y., P. Cizeau, M. Meyer, C. K. Peng, and H. E. Stanley (1997). Correlations in economic time series. *Physica A 245*, 437–440.

Longstaff, F. and A. Wang (2004). Electricity forward prices: A high-frequency empirical analysis. *Journal of Finance 59*, 1877–1900.

Lozada, G. A. (1991). A defense of Nicholas Georgescu-Roegen's paradigm. *Ecological Economics 3(2)*, 157–160.

Lucheroni, C. (2007). Resonating models for the electric power market. *Physical Review E 76*, 056116–13.

Lucia, J. J. and E. S. Schwartz (2002). Electricity prices and power derivatives: Evidence from the Nordic Power Exchange. *Review of Derivatives Research 5*(1), 5–50.

Magnus, J. R. (1979). Substitution between energy and non-energy inputs in the Netherlands 1950-1976. *International Economic Review 20(2)*, 465–484.

Maheu, J. and T. McCurdy (2004). News arrival, jump dynamics and volatility components for individual stock returns. *Journal of Finance 59*(2), 755–793.

Maheu, J. and T. McCurdy (2006). Do high-frequency measures of volatility improve forecasts of return distributions? Working Paper.

Mancini, C. (2004). Estimation of the parameters of jump of a general Poisson-diffusion model. *Scandinavian Actuarial Journal 1*, 42–52.

Mancini, C. (2007). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*. Forthcoming.

Mancini, C. and R. Renò (2008). Threshold estimation of jump-diffusion models and interest rate modeling. Manuscript, University of Florence and University of Siena.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business 36*, 349–419.

Mantegna, R. and E. Stanley (1995). Scaling behaviour in the dynamics of an economic index. *Nature 376*, 46–49.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B 11(1)*, 193–197.

McAleer, M. and M. Medeiros (2006). Realized volatility: a review. *Econometric Reviews.*. Forthcoming.

McDonald, M., O. Suleman, S. Williams, S. Howison, and N. F. Johnson (2008). Impact of unexpected events, shocking news, and rumors on foreign exchange market dynamics. *Physical review E 77(4)*, 046110(12).

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics 3(1)*, 125–144.

Miller, D. G. (1994). The origins of onsager's key role in the development of linear irreversible thermodynamics. *Journal of Statistical Physics 78(1-2)*, 563–573.

Moulines, E. and P. Soulier (2000). Semiparametric spectral estimation for fractional processes. *Theory and Application of Long-Range Dependence (book cahpter),Birkhäuser*.

Mount, T. D., Y. Ning, and X. Cai (2006). Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters. *Energy Economics 28*, 62–80.

Müller, U., M. Dacorogna, R. Dav, O. Pictect, R. Olsen, and J. Ward (1993). Fractals and intrinsic time - a challenge to econometricians. *XXXIXth International AEA Conference on Real Time Econometrics, 14-15 Oct 1993, Luxembourg*.

Müller, U., M. Dacorogna, R. Davé, R. Olsen, O. Pictet, and J. V. Weizsacker (1997). Volatilities of different time resolutions - analyzing the dynamics of market components. *Journal of Empirical Finance 4*, 213–239.

Müller, U. A., M. M. Dacorogna, R. B. Olsen, O. V. P. M. Schwarz, and C. Morgenegg (1990). Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *Journal of Banking and Finance 14(6)*, 1189–1208.

Murthy, K. V. B. (2002). Arguing a case for the Cobb-Douglas production function. *Review of Commerce Studies 20-21(1)*, 75–91.

Nummedal, L., A. Rosjorde, E. Johannessen, and S. Kjelstrup (2005). Second law optimization of a tubular steam reformer. *Chemical Engineering and Processing 44(4)*, 429–440.

Onsager, L. (1930). Reciprocal relations in irreversible processes I. *Physical Review 37(4)*, 405–426.

Onsager, L. (1931). Reciprocal relations in irreversible processes II. *Physical Review 38(12)*, 2265–2279.

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time series study. *Journal of Financial Economics 63*, 3–50.

Paraschiv-Ionescu, A., E. Buchser, B. Rutschmann, and K. Aminian (2008). Nonlinear analysis of human physical activity patterns in health and disease. *Computers in Biology and Medicine 77(2)*, 021913(10).

Patton, A. (2006). Volatility forecast evaluation and comparison using imperfect volatility proxies. Unpublished paper: Department of Accounting and Finance, LSE.

Peng, C. K., S. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger (1994). Mosaic organization of DNA nucleotides. *Physical Review E 49(2)*, 1685–1689.

Pindyck, R. S. (1979). Interfuel substitution and the industrial demand for energy: an international comparison. *The Review of Economic and Statistics 61(2)*, 169–179.

Pindyck, R. S. (1983). Dynamic factor demands and the effects of energy price shocks. *The American Economic Review 73(5)*, 1066–1079.

Pirino, D. (2008). Jump detection and long range dependence. *Physica A Forthcoming*.

Pirino, D. and R. Renò (2008). Electricity prices: a non-parametric approach. *Working Paper*.

Podolskij, M. and M. Vetter (2006). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. Working Paper.

Porporato, A. and P. D'Odorico (2004). Phase transitions driven by state-dependent Poisson noise. *Physical Review Letters 92*, 110601–4.

Protter, P. (1990). *Stochastic Integration and Differential Equations*. Springer-Verlag.

Renò, R., A. Roma, and S. Schaefer (2006). A comparison of alternative nonparametric estimators of the diffusion coefficient. *Economic Notes 35*(3), 227–252.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics 23(8)*, 1630–1661.

Roma, A. (200). Financial and thermodynamic equilibrium. *Economic Notes 29(3)*, 341–354.

Roma, A. (2006). Energy, money, and pollution. *Ecological Economics 56(4)*, 534–545.

Roma, A. and D. Pirino (2008). A theoretical model for the extraction of natural resources. *Working Paper*.

Ruth, M. (1995). Thermodynamic constraints on optimal depletion of copper and aluminium in the united states: a dynamic model of substitution and technical change. *Ecological Economics 15(3)*, 197–213.

Salamon, P., A. Nitzan, B. Andersen, and R. S. Berry (1980). Minimum entropy production and the optimization of heat engines. *Physical Review A 21(6)*, 2115–2129.

Sato, K. (2001). Basic results on Lévy processes. In O. Barndorff-Nielsen, T. Mikosch, and S. I. Resnick (Eds.), *Lévy Processes, Theory and Applications*, pp. 3–37. Birkhäuser.

Sav, G. T. (1984). Micro engineering foundations of energy-capital complementarity: solar domestic water heaters. *The Review of Economics and Statistics 66(2)*, 334–338.

Scott, D. (1992). *Multivariate Density Estimation*. John Wiley & Sons.

Shalen, C. T. (1993). Volume, volatility and dispersion of beliefs. *Review of Financial Studies 6*, 405–434.

Shcherbakov, R., G. Yakovlev, D. L. Turcotte, and J. B. Rundle (2003). Model for the distribution of aftershock interoccurrence times. *Physical Review Letters 95*, 218501–4.

Shreve, S. (2004). *Stochastic Calculus for Finance II*. Springer.

Sieniutycz, S. (2003). Thermodynamic limits on production or consumption of mechanical energy in practical and industrial system. *Progress in Energy and Combustion Science 29(3)*, 193–246.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC.

Simonsen, I. (2005). Volatility of power markets. *Physica A 355*(1), 10–20.

Solow, R. (1974). The economics of resources or the resources of economics. *The American Economic Review 64(2)*, 1–14.

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance 52*, 1973–2002.

Stiglitz, J. E. (1976). Monopoly and the rate of extraction of exhaustible resources. *The American Economic Review 66(4)*, 665–661.

Thompson, P. and T. G. Taylor (1995). The capital-energy substitutability debate: a new look. *The Review of Economics and Statistics 77(3)*, 565–569.

Tian, L. and Y. Wang (2001). A comparative economic assessment of a 200 MW nuclear heating reactor for seawater desalination by multi-effect distillation (MED). *Desalination 133(3)*, 285–290.

Tsirlin, A. M. and V. Kazakov (2004). Optimal processes in irreversible thermodynamics and microeconomics. *Interdisciplinary Description of Complex System 2(1)*, 29–42.

Tsirlin, A. M., V. A. Mironova, and S. A. Amelkin (1998). Finite-time thermodynamics: Conditions of minimal dissipation for thermodynamic process with given rate. *Physical Review E 58(1)*, 215–223.

Tsirlin, A. M. and I. V. Titova (2004). Estimating the minimum work of separation of an ideal mixture in processes with a specified output. *Theoretical Foundations of Chemical Engineering 38(5)*, 490–498.

T.Tsoutsos, V. Gekas, and K. Marketaki (2003). Technical and economical evaluation of solar thermal power generation. *Renewable Energy 28(6)*, 873–886.

Vallieres, C., E. Favre, X. Arnold, and D. Roizard (2003). Separation of binary mixture by dense membrane processes: influence of inert gas entrance under variable downstream pressure conditions. *Chemical Engineering 58*, 2767–2775.

Wand, M. P. and M. C. Jones (1995). *Kernel smoothing*. Chapman & Hall.

Wang, F., K. Yamasaki, S. Havlin, and H. E. Stanley (2006). Scaling and memory of intraday volatility return intervals in stock markets. *Physical review E 77(1)*, 026117(8).

Wang, J. (1994). A model of competitive stock trading volume. *Journal of Political Economy 102*, 127–168.

Weber, P., F. Wang, I. Vodenska-Chitkushev, S. Havlin, and H. E.Stanley (2006). Relation between volatility correlations in financial markets and omori processes occurring on all scales. *Physical review E 76(1)*, 016109(8).

Weron, R. (2002). Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A 312*, 285–299.

Wijmans, J. G. and R. W. Baker (1995). The solution-diffusion model: a review. *Journal of Membrane Science 107(1)*, 1–21.

Wilkinson, L. and J. Winsen (2002). What we can learn from a statistical analysis of electricity prices in New South Wales. *The Electricity Journal 15*, 60–69.

Woerner, J. (2006). Power and multipower variation: inference for high frequency data. In A. N. Shiryaev, M. do Rosario Grossinho, P. Oliveira, and M. Esquivel (Eds.), *Stochastic Finance*, pp. 343–364. Springer.

Y.Aït-Sahalia and J. Jacod (2008). Volatility estimators for discretely sampled lévy processes. *The Annals of Statistics 35(1)*, 355–392.

Young, J. T. (1991). Is the entropy law relevant to the economics of natural resource scarcity? *Journal of Environmental Economics and Management 21(2)*, 169–179.

Zygadlo, R. (2003). Martingale integrals over Poissonian processes and the Ito-type equations with white shot noise. *Physical Review E 68*, 046117–4.