



University of Zurich  
Zurich Open Repository and Archive

Winterthurerstr. 190  
CH-8057 Zurich  
<http://www.zora.uzh.ch>

---

*Year: 2008*

---

## An econometric analysis of emission allowance prices

Paoletta, M S; Taschini, L

Paoletta, M S; Taschini, L (2008). An econometric analysis of emission allowance prices. *Journal of Banking and Finance*, 32(10):2022-2032.

Postprint available at:  
<http://www.zora.uzh.ch>

Posted at the Zurich Open Repository and Archive, University of Zurich.  
<http://www.zora.uzh.ch>

Originally published at:  
*Journal of Banking and Finance* 2008, 32(10):2022-2032.

# An econometric analysis of emission allowance prices

## Abstract

Knowledge of the statistical distribution of the prices of emission allowances, and their forecastability, are crucial in constructing, among other things, purchasing and risk management strategies in the emissions-constrained markets. This paper analyzes the two emission permits markets, CO<sub>2</sub> in Europe, and SO<sub>2</sub> in the US, and investigates a model for dealing with the unique stylized facts of this type of data. Its effectiveness in terms of model fit and out-of-sample value-at-risk-forecasting, as compared to models commonly used in risk-forecasting contexts, is demonstrated.

# An Econometric Analysis of Emission Trading Allowances\*

Marc S. Paolella<sup>a</sup>      Luca Taschini<sup>a†</sup>

<sup>a</sup>*Swiss Banking Institute, University of Zurich, Switzerland*

November 2006

## Abstract

World power and gas markets have a natural relationship with global tradable carbon permits markets, including the U.S. Clean Air Act Amendments and the EU Emissions Trading Scheme, the latter officially launched in January 2005. Electric utilities operate their power plants based in part on the price of the power and the relative cost of coal and natural gas. As both carbon dioxide and sulphur dioxide are by-products of the coal burning process, the new factors of SO<sub>2</sub> and CO<sub>2</sub> emissions allowances come into play in a carbon constrained economy. Now that a price has been put on such allowances, the differences in carbon intensity for coal and gas could potentially change the way companies run their power plants. Moreover, knowledge of the statistical distribution of emission trading allowances, and its forecastability, becomes crucial in constructing optimal hedging and purchasing strategies in the carbon market. This paper provides an in-depth analysis of available data addressing the unconditional tail behavior and the inherent heteroskedastic dynamics in the returns on the emissions allowances.

**Keywords:** Environmental Finance, GARCH, Greenhouse Gases, Mixture Models, Tail Estimation.

*JEL Codes:* C16; C32; C51; C52; C53

---

\*Part of the research of the authors has been carried out within the National Centre of Competence in Research “Financial Valuation and Risk Management” (NCCR FINRISK), which is a research program supported by the Swiss National Science Foundation. Corresponding author: [taschini@isb.unizh.ch](mailto:taschini@isb.unizh.ch).

†The help from Enviro Climate Change Group as well as the data base on sulfur dioxide spot prices from the Chicago Climate Exchange are gratefully acknowledged. The authors would like to thank Pauline Barrieu, Marc Chesney, George Daskalakis, Helyette Geman, Daniel Giamouridis, Rajna Gibson, Wolfgang Härdle, Gikas Hardouvelis, Raphael Markellos, Nikitas Pittis, Dimitris Psychoyios, Costas Vorlow, and Alexander Wagner for their helpful discussions and comments. We wish to also thank Sven-C. Steude for the programming assistance on the mixed normal models used in the paper.

# 1 Introduction

Since the climate change and environmental pressure problems were acknowledged at the Earth Summit in Rio de Janeiro in 1992, governments realized that something must be done, even if there is considerable disagreement as to exactly what. At the time of writing, the media is starting with the discussion of the gravity and consequences of global warming due to carbon dioxide output. In the financial world, innovative products have been developed to deal specifically with such environmental problems; see Tietenberg (2006) for an historical account.

Perhaps the best known of the new family of environmental financial products are emissions allowances. The most celebrated example was created by the United States (U.S.) Acid Rain Program, which in 1995 imposed emissions limits on major sources of sulfur dioxide (SO<sub>2</sub>)—the main cause of acid rain. Each of these sources (mainly power plants) was allocated a limited number of emission allowances, each representing one ton of SO<sub>2</sub>. Companies that emit more than the number of allowances they hold face financial penalties. These “assets” may be freely bought and sold: Those that can reduce their SO<sub>2</sub> output relatively cheaply will sell their excess allowances to those that find it more costly to curb their emissions.

As a consequence, the same approach is being applied to other environmental problems, such as nitrogen oxide in the U.S. in order to combat local smog problems, or to carbon dioxide (CO<sub>2</sub>) in the European Union (EU) and the U.S. as part of their efforts to tackle global warming.

According to the World Energy Outlook (WEO) 2002, CO<sub>2</sub> emissions will increase by 1.8% per year between 2000 and 2030. From 12,369 million tons (Mt) of CO<sub>2</sub> equivalent<sup>1</sup> in 2000, emissions are forecasted to reach about 16,400 Mt in 2030 for the Organization for Economic Cooperation and Development (OECD) countries, representing an average increase of 0.9% per year. The power generating sector has actually contributed more than 50% of the increase in global emissions between 2000 and 2006, and will remain the largest source of CO<sub>2</sub> emissions in 2030. During the years 2000 to 2003, power generation in OECD countries was responsible for 40% of emissions of greenhouse gases; in 2006, it was 65%.

The CO<sub>2</sub> emission allowance has grown in importance since the Kyoto Protocol entered into force on February 16, 2005, obliging all affected utilities in the countries that ratified the Protocol to curb their greenhouse gas emissions according to the program. The EU has decided to introduce a CO<sub>2</sub> emission trading scheme, in short, the EU ETS, in two phases (2005-2007 and 2008-2012) to coincide with the first Kyoto Protocol commitment period. Even in the U.S., which has not signed the Protocol, New York Attorney General Eliot Spitzer demanded in July 2004 that the nation’s

---

<sup>1</sup>The CO<sub>2</sub> equivalent is a metric measure used to compare the emissions from various greenhouse gases based upon their global warming potential (GWP). They are commonly expressed as *million metric tons of CO<sub>2</sub> equivalents* (MMTCDE). The CO<sub>2</sub> equivalent for a gas is derived by multiplying the amount of gas, in tons, by the associated GWP. For example, the GWP for methane is 21 and for nitrous oxide 310. This means that emissions of one million metric tons of methane and nitrous oxide are equivalent to emissions of 21 and 310 million metric tons of CO<sub>2</sub>, respectively.

largest utilities significantly reduce greenhouse gas emissions. The states of Massachusetts and California unveiled plans to cut emissions by 10% and 20%, respectively, by 2020. At the same time, U.S. energy companies are also planning for the future: American Electric Power, the number one coal-burning utility, is trying to accumulate credits for cutting CO<sub>2</sub>, as are Du Pont, Alcoa Inc., and General Electric.

The introduction of emission allowances is obviously modifying operating costs in the power generation sector. Electric utilities operate their power plants based in part on the price of the power and the relative cost of coal and natural gas. In a carbon constrained economy, the right to pollute is a new factor which comes into play. Now that a price has been put on such emission allowances, the differences in carbon intensity for coal and gas could potentially change the way companies run their power plants and lead to an increasing amount of trading of emission rights.

Along with the working papers of Daskalakis et al. (2006) and Benz and Trück (2006), this paper provides one of the first econometric investigations of the behavior of the new emission allowance commodities. Our approach is completely different than that used in both the aforementioned papers.<sup>2</sup> We discuss forecast methods based on the analysis of a variety of factors, including analysis of supply and demand fundamentals, and also based on the spot-future parity. We show that these two approaches lead to implausible conclusions due, respectively, to the complexity of the market and to the particular behavior of the emission allowance commodity. More effective are statistical models relying exclusively on historical price information. In particular, we examine (i) the riskiness of the asset via estimation of the tail thickness of the unconditional distribution, and (ii) a particularly well-suited GARCH-type model for the conditional distribution of the returns on the emission allowance spot prices.

The remainder of the paper proceeds as follows: Section 2 explains the emission trading framework describing the existing SO<sub>2</sub> and CO<sub>2</sub> markets. Section 3 introduces the data used in the study and describes their main features. Section 4 discusses the misleading conclusions resulting from forecasting approaches based on either analysis of fundamentals or the future-spot parity. Sections 5 and 6 provide the empirical analysis of the SO<sub>2</sub> and CO<sub>2</sub> returns, respectively. Section 7 provides concluding remarks and ideas for future research.

---

<sup>2</sup>Daskalakis et al. (2006) use a jump-diffusion model to approximate the random behavior of the carbon dioxide emission spot price. Benz and Trück (2006) analyze the short-term spot price behavior of carbon dioxide emission allowances, and employ a Markov-switching model to capture the heteroskedastic behavior of the return series. In contrast, we advocate use of a new GARCH-type structure. It is noteworthy that the two approaches (Markov switching and GARCH) could be combined (see, e.g., Haas et al., 2004b, and Bauwens et al., 2006, and the references therein), though we do not pursue this because, firstly, the data sets which were available at the time of writing are somewhat short for such elaborate models, and secondly, tests for structure in the occurrence of zeros in the data set we examine do not reveal any Markov structure; see Section 5.3 below.

## 2 Political and Market Framework

Title IV of the Clean Air Act Amendment (CAAA) in the U.S., and the Emission Trading Scheme (ETS) in Europe, created *de facto* property rights for emissions, referred to as *allowances* in the programs, that can be freely traded. The right gives affected subjects<sup>3</sup> complete flexibility in determining how they will comply with their obligations under the programs. To allow utilities to take advantage of inter-temporal cost savings, Title IV in the U.S. allows *inter-phase banking*, i.e., unused allowances in Phase I can be used in Phase II. The EU ETS allows only *within-phase* banking and prevents inter-phase banking. Therefore, within each phase, allowances can be banked from one year to the next, but unused Phase I allowances are not valid during Phase II. However, this restriction is under discussion because it can be perceived as a potential source of market distortion; see Cason and Gangadharan (2005).<sup>4</sup>

Allowances can be traded nationally in the case of the U.S., and internationally under the EU ETS, with no necessity of prior approval. The purchase and holding of allowances is not restricted to the companies affected by the programs—which means that all sources, as well as third parties such as brokers, are free to buy and sell allowances with any other party. Also, small private investors can participate in this market, thanks to the Emission Certificates sold by private banks.

### 2.1 How the programs work

Title IV of the 1990 CAAA established the first large scale, long-term environmental program to rely on tradable emissions permits to control pollution. This program was designed to cut acid rain by reducing SO<sub>2</sub> emissions from electric generating plants to about half their 1980 level, beginning in 1995. Title IV specifies the initial allocation of SO<sub>2</sub> allowances according to a fairly complicated set of rules discussed by Joskow and Schmalensee (1998). The restrictions on SO<sub>2</sub> emissions are applied in two phases. Phase I covers the 263 “dirtiest” (this being the terminology of the CAAA, referring to the largest SO<sub>2</sub> producers) generating units in the U.S. and required them, in the aggregate, to reduce their emissions to about 5.7 million tons per year, during the period 1995-1999. Phase II, which began in 2000, tightened the emissions cap further and extended the program to all the electric generation units in the U.S.

The EU ETS is the largest multi-country, multi sector greenhouse gas (GHG) emissions trading scheme in the world.<sup>5</sup> It covers almost 12,000 installations in 25 countries, responsible for nearly

---

<sup>3</sup>Title IV in the U.S. affects mainly electric utilities and EU ETS affects different sectors like iron, steel, cement, glass and ceramics, pulp and paper producers and the energy sector as well.

<sup>4</sup>France and Poland received permission from the European Commission (EC) to bank from Phase I to Phase II. In light of this, the EC should rapidly clarify all inter-phase and intra-phase banking rules so as to reduce regulatory uncertainty.

<sup>5</sup>Greenhouse gases refer to CO<sub>2</sub>, methane CH<sub>4</sub>, nitrous oxide N<sub>2</sub>O, hydrofluorocarbons HFC<sub>s</sub>, perfluorocarbons PFC<sub>s</sub> and sulphur hexafluoride SF<sub>6</sub>. These promote warming of the earth’s atmosphere and are released primarily through the combustion of fossil fuels, industrial processes and land use changes.

half of the EU's CO<sub>2</sub> emissions. They have been allocated allowances giving them the right, over the next three years, to emit 6.6 billion tons of CO<sub>2</sub>. The basic approach to emissions control embedded in the U.S. CAAA and EU ETS programs is straightforward: an aggregate annual cap on SO<sub>2</sub> and GHG emissions, respectively, defines the number of emission allowances available for allocation to the different sources each year. An emission allowance is the right to emit one ton of the traded gas into the atmosphere. To legally emit the offending gas during a given year, an affected unit must have enough allowances that are good for use in that year to cover all its emissions. They are also obligated to have a continuous emissions monitoring system to measure actual emissions and report the data to the authorities. In the beginning of March, each source must have deposited enough allowances into an account maintained for it by the authority to cover all of its recorded emissions for the previous year.

Crucial for both programs is the fact that a penalty is levied if a firm does not deliver a sufficient amount of allowances on time. The level of fines is set at 2,500 USD per ton SO<sub>2</sub> according to the U.S. CAAA and €40 per ton CO<sub>2</sub> under the EU ETS.<sup>6</sup> Because payment of the fine does not remove the obligation to achieve compliance, it does not cap the price of allowances. Each allowance specifies a particular year, referred to as its *vintage*, in which it is first available to be used to cover emissions. An allowance can also be totally or partly *banked* and used in the future. That means a 2006 vintage allowance can be used to cover emissions in 2006 or in any later year. An affected source can buy allowances to cover its present emissions or its future emissions from any type of trading partner.

## 2.2 The new markets

SO<sub>2</sub> allowance and Carbon Financial Instruments (CFI) are the terms used to describe the spot and futures contracts listed and traded on different electronic trading platforms. The underlying commodity of a CFI is the CO<sub>2</sub> emission allowance issued by designated European Union Member States and other units representing CO<sub>2</sub> emissions that are specified as eligible for delivery.

Futures markets are derivative markets—they exist in relation to cash markets, which are the underlying primary markets in which actual physical commodities are bought and sold. In the case of U.S. CAAA, the underlying cash market is traded mainly through private transactions executed in the over-the-counter markets as well as in the Chicago Climate Exchange spot market. In the case of the EU ETS, the future contracts and the underlying cash market are traded in the following markets:

- Nord Pool in Norway started in February 2005 trading future contracts with a initial daily

---

<sup>6</sup>The EU ETS takes place, like U.S. Title IV, in two phases, with Phase I running from 2005 to 2007, and Phase II from 2008 to 2012. The penalty during EU ETS Phase I is €40 per ton CO<sub>2</sub>, and it increases to €100 per ton in Phase II. Such an increase is directly related to the coinciding of the EU ETS Phase II and the Kyoto Protocol's first commitment period—which requires a deeper GHG reduction.

volume of 150,000 CO<sub>2</sub> tons.

- European Energy Exchange (EEX) in Leipzig began in March 2005 to trade spot contracts with a starting daily volume of 50,000 CO<sub>2</sub> tons.
- European Climate Exchange (ECX) in Amsterdam started in April 2005 trading future contracts with a initial daily volume of 300,000 CO<sub>2</sub> tons, i.e., 40% of the daily volume traded in all Europe.
- Powernext in France started June 24th, 2005, trading spot CO<sub>2</sub> contracts. The market began trading 20,000 CO<sub>2</sub> tons per day and is now the most liquid spot market among the European exchanges.
- SendeCO<sub>2</sub> in Spain started at the end of 2005.

The SO<sub>2</sub> and CO<sub>2</sub> cash and future markets have grown enormously since their inception; see Figures 1 and 2. The value of ECX CFI futures contracts traded on the International Petroleum Exchange, for example, has climbed above the €1 billion mark in less than six months of trading and the open interest in the contracts now exceeds €200 million.

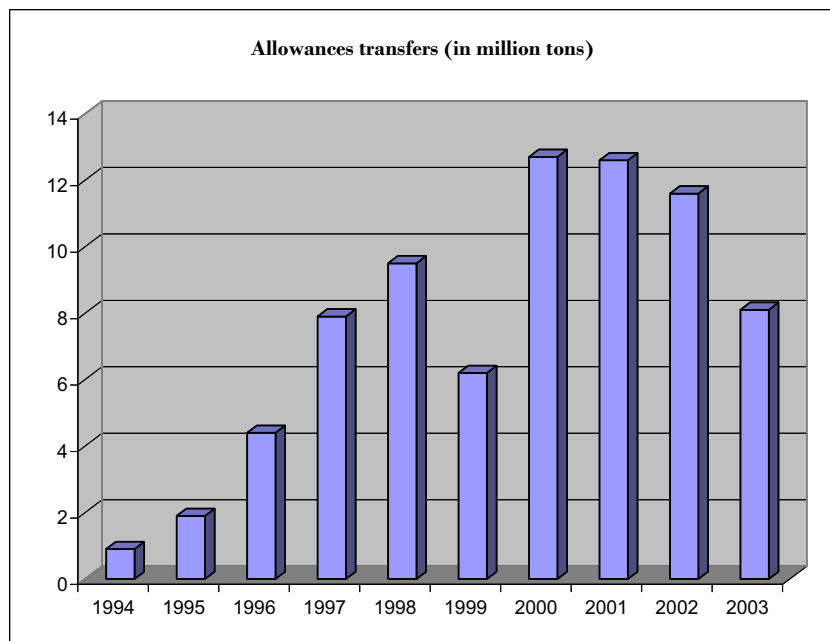


Figure 1: Annual volumes of SO<sub>2</sub> emission allowance transfers across economically distinct entities.

Source: U.S. Environmental Protection Agency USEPA.

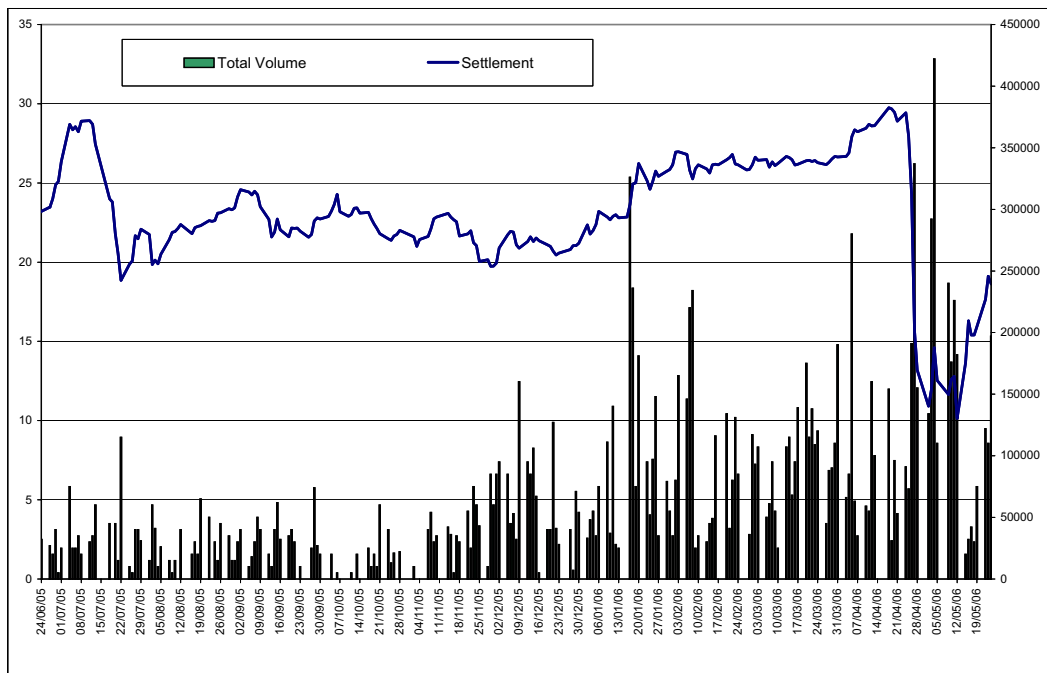


Figure 2: Spot CO<sub>2</sub> volumes (vertical bars) and prices from April 2005 to May 2006. Source: European Climate Exchange.

### 3 Spot Price Dynamics of Emission Allowances

In contrast to stock prices, which grow on average because the investor is rewarded for the time value of money augmented by a risk premium, commodity prices do not generally exhibit trends over long periods. Even if sharp rises are observed during short periods for specific events, such as the weather or political conditions, commodity prices tend to revert to normal levels in the long run. The resulting properties of commodity prices are a consequence of the general behavior of mean-reversion combined with spikes in prices caused by shocks in the supply/demand balance.

Under the U.S. CAAA-Title IV, the EU ETS, and the Kyoto Protocol frameworks, the emission allowances can be considered as a new kind of commodity. The aim to gradually create a scarcity of allowances must ultimately generate an upward price trend. Thus, one expects to observe a mean-reversion trend around an upward slope; such supposition could be strengthened by a first analysis of the daily SO<sub>2</sub> price, as shown in Figure 3. Indeed such a path, at least from June 2003 until November 2005, could be consistent with a stochastic mean-reverting process with a constant positive drift. Clearly, the enormous price drop after November 2005 indicates that such an assumption would have been detrimental. As such, we opt to consider a time-series model applied to the return series,  $r_t = 100(\ln p_t - \ln p_{t-1})$ , generated by the price sequence  $p_t$ .

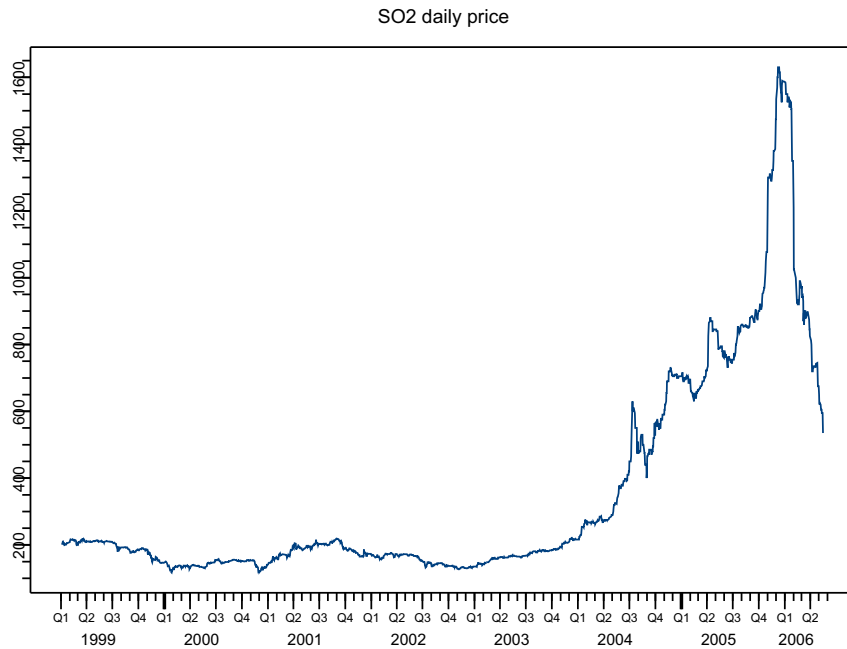


Figure 3: Daily SO<sub>2</sub> allowance prices over the period January 1999 - May 2006. Source: Chicago Climate Exchange.

### 3.1 SO<sub>2</sub> spot price dynamics

The spot closing prices SO<sub>2</sub> have been collected by the Chicago Climate Exchange on the OTC market from January 4th, 1999 up to May 26th, 2006. Table 1 shows some basic statistics for the returns on the SO<sub>2</sub> price series. As is common with returns on more traditional financial securities, there is high kurtosis, although virtually no skewness.

SO <sub>2</sub> Returns	Mean	Std. Dev.	Skewness	Kurtosis
1999 - 2006	0.00	0.018	-0.52	15.83
2003 - 2006	0.71	0.307	0.31	13.79

Table 1: First four moments for SO<sub>2</sub> allowances price return over period January 1999 to May 2006, and from January 1, 2003 to May 23, 2006.

From Figure 3, it might appear that the SO<sub>2</sub> daily price is mean-stationary until the beginning of 2003, but this is an artifact of the scaling of the graph, and magnifying shows that the data resemble a random walk. Nevertheless, based on the blatant difference occurring around 2003, it is reasonable to assume that the data-generating process has changed. Moreover, there is at least one plausible explanation for why the spot price increased so dramatically in 2003: In 2000, the more stringent Phase II in Title IV emissions limitations became effective. In addition, because Phase II extends to all electric generating units in the continental U.S., significantly more utilities

are affected by Title IV starting in the year 2000 and they must plan accordingly for compliance purposes, see Bailey (1998). As discussed in Ellerman (2003), there also exists strong evidence of a rational banking behavior of affected sources during Phase I provided by the change in total emissions from 1999 and 2000—when the allocation for the affected units was reduced by 50%. Inspecting Figure 1 and considering that, in 1999, affected sources continued to bank allowances which, in 2000, they started to draw down, it is plausible that the price started rising when the allowance–buffer began suffering. In fact, the contraction of the supply side of emission allowances translated into a slight decrease of transferred volumes and an increase of the price. Moreover, after a short running period, the presence of a wide variety of market participants and the scarcity of emission allowances boosted the price.

The spot prices of SO<sub>2</sub> broke through the \$1,000 per ton barrier for the first time on October 25, 2005, and approached \$1,200 per ton on November 1, 2005, as utilities hoarded allowances in expectation of further price rises. That represents a steep climb from the \$130 per ton in early 2003. Although, according to the EPA, while five to six million excess allowances were currently banked by utilities, generators were drawing those down at a rate of 1 million tons a year, thus driving the speculative trend. Anticipation in late 2005 about a warm 2006 winter season reduced the speculative enthusiasm. Sure enough, the winter season turned out to be the fifth warmest, and the month of January was *the warmest* on record. This obviously affected the demand for heating, energy production, and, thus, the pollution level itself. While the spot price decreased by \$600 per ton, the overall liquidity of the market increased compared to the beginning of 2005.

### 3.2 CO<sub>2</sub> Spot Price Dynamic

Daily data on spot prices for CO<sub>2</sub> are available from Nord Pool, European Energy Exchange (EEX) and Powernext Exchange in Paris for time series starting from, respectively, February, beginning of March, and end of June. In our study we examine the closing spot price recorded by Powernext, the exchange with the largest traded volumes of spot-contracts. This yields 337 returns, from June 25, 2005 to November 3, 2006, with basic statistics reported in Table 2. Figure 4 plots the returns.

CO <sub>2</sub> Returns	Mean	Std. Dev.	Skewness	Kurtosis
June 05 - October 05	−0.10	0.52	−1.06	3.73
June 05 - November 06	−0.23	4.79	−1.98	30.12

Table 2: First four moments for CO<sub>2</sub> allowances returns.

Although a futures market for allowances in the EU ETS started as early as 2003-2004, a real CO<sub>2</sub> trading market commenced just in February 2005 with the first quoting of the Nord Pool Exchange. According to Evolution Markets data, one-month historical price volatility has ranged from 20% to 90% since the start of 2005. Over the same period, the price has ranged from slightly

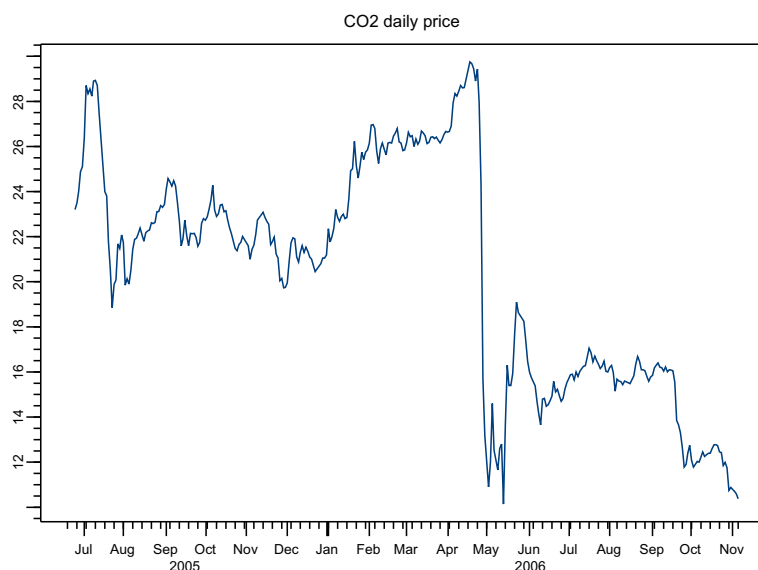


Figure 4: Daily CO<sub>2</sub> allowance prices over the period June 25, 2005 to November 3, 2006. Source: Pownert Exchange.

less than €7 at the beginning of the year to nearly €30 in July. At that point in time, already four exchanges have been trading CO<sub>2</sub> spot and futures allowances for some hundreds million, and by December 2005, after a few months from the first trade, the total traded value reached the one billion level.

On May 2nd, 2006, CO<sub>2</sub> prices in the EU ETS nearly halved after several countries reported lower-than-expected CO<sub>2</sub> emissions in 2005. The spot price closed at € 10.90 per ton on Pownert on that day, compared with a price of just under €30 per ton at the close of trading on April 24, 2006. The price collapse followed reports from Belgium, the Czech Republic, Estonia, France, the Netherlands and Spain on how much their affected companies had emitted in 2005. In all cases, they were significantly less than expected, signaling an excess supply of allowances. Despite the controversies and the uncertainty surrounding the aforementioned data release, the market had remained liquid throughout: Trading volumes had been high and parties looking to buy or sell allowances had been able to do so (a counterparty was virtually always available) during the period of the drop.

Essentially, if Europe has a large allowance surplus, then the first-phase price will go to zero due to the inter-phase banking prevention. But before March 2008, no one will be able to irrefutably demonstrate if Phase I is in a shortage or surplus position. The market itself illustrates such discord: If everyone were to accept the theory of the market being fundamentally in surplus, then the CO<sub>2</sub> allowance would be trading at nearly zero. On the contrary, the CO<sub>2</sub> allowance

spot price was just under €20 on May 23, 2006.

Moreover, by the first half of 2006, the overall market liquidity increased, with Powernext, for example, exchanging 100,000 CO<sub>2</sub> tons per day on average. And around 200 million tons of CO<sub>2</sub> have been traded up to the end of 2005. This represents a market *churn* (number of allowances traded annually divided by the total number issued each year) of around 8%. By April 2006, this value reached 15%. It is interesting to compare this with the SO<sub>2</sub> market, which has a market churn in excess of 200%, according to the USEPA. Given that the U.S. market has been running for ten years, this potentially points the way to a more liquid future for the EU ETS. Moreover, the EU directive allows for considerable changes in EU ETS Phase II. The main options are the inclusion of new gases; the extension of the scheme to include other sources of CO<sub>2</sub> in the sectors covered by the directive; and the inclusion of new sectors like aluminum production, aviation, road transport and chemicals. Such enlargements, combined with the gradual decrease of the cap in EU ETS Phase II, will surely raise the likelihood of an upward trend in trading values.

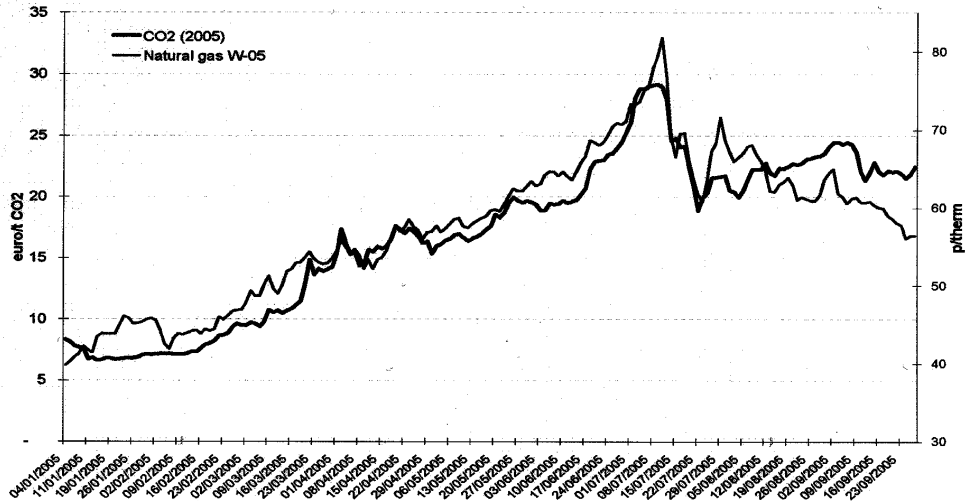
## 4 CO<sub>2</sub> Price Scenarios According to Current Approaches

One aim of this paper is to provide the appropriate methodology to get insights into future emission market price scenarios. Because CO<sub>2</sub> is the underlying of the largest multi-country, multi sector GHG emissions trading scheme in the world, we begin considering the trading strategies currently applied to it. The following sections briefly consider the CO<sub>2</sub> market from a qualitative and quantitative point of view, as well as the potential existence of a future-spot parity. We illustrate that both approaches fail to be reliable tools for price scenario analysis, motivating our atheoretical, statistical study of the historical price series in Section 5.

### 4.1 Approach based on Fundamentals of CO<sub>2</sub>

The theoretical optimal market price during EU ETS Phase I should be determined by the marginal cost of achieving compliance over the first three years. Pricing movements experienced in the last nine months have reflected traders' perceptions of this marginal activity. Determining this price setting activity is complex, requiring consideration of a number of demand-side and supply-side factors. Because the largest affected sources by the EU ETS are power utilities, we can assume, as elementary proxies of the input market drivers, the *fuel prices* and *economic growth*. The former, treated in this section, is related to the cost of fossil fuel itself and the fuel switching in the power sector; the latter is related to the future electricity demand.

Backed by a high correlation level since the EU ETS inception (on January 4, 2005) between EU allowance prices and average winter 2005-2006 gas prices (see Figure 5), the risk management offices of the companies covered by the ETS were mainly trading according to the oil price trend. Indeed, the wholesale gas price, i.e., one of the cost proxies, is linked to the international oil



Source: Barclays Capital & Enviro

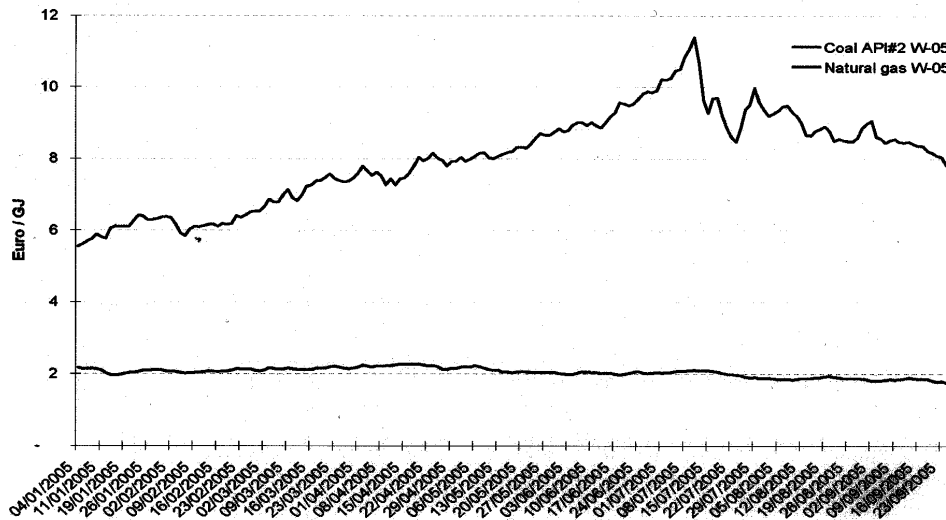
Figure 5: CO<sub>2</sub> and forward natural gas price from January to September 2005. Source: Enviro Climate Change Group.

price, which had increased significantly since 2004. However, such a relationship lasted until August, when trends started to diverge. After this point, winter gas prices started falling but CO<sub>2</sub> prices showed a slight upward trend from a low on July 27, 2005. A possible explanation could be a higher expectation of GHG emissions by the end of 2005 due to larger coal use, or traders' realization that the initial tradable allocation entitled by the European governments to individual firms was not too generous.

As discussed in Bailey (1998) and Montero and Ellerman (1998), the theoretical CO<sub>2</sub> price should be close to the marginal cost for reducing emissions, e.g., the marginal cost of fuel switching.<sup>7</sup> In addition, major trading desks like Fortis and Evolution Markets claimed that the rising price differential between gas and coal has been the major factor determining the increase in CO<sub>2</sub> prices on the market, notably during the first half of 2005; see Figure 6. The fuel switching price level, however, did not totally reflect such differences and, contradicting the theory, the CO<sub>2</sub> price remained far away from such levels; see Figure 7.

Therefore, *fuel prices* and, ultimately, *fuel switching*, are the primary approximate fundamentals that determine the CO<sub>2</sub> price level. There exist several possible explanations for such features that potentially pave the way for future research. For example, the opportunity costs to switch from coal to natural gas or the fact that the market does not expect the current high natural gas price to continue. It turns out that any price scenario delineation based on the trend analysis of a few variables considered as proxies for the fundamentals that affect the supply-side

<sup>7</sup>Fossil fuel units can reduce their emissions by switching fuel from higher to lower sulfur oil or coal, or from oil and coal to natural gas.



Source: Barclays Capital & Enviro

Figure 6: Forward winter prices for coal and gas from January to September 2005. The nearly straight line hovering around €2/GJ is the forward gas price. Source: Enviro Climate Change Group.

cannot suffice.<sup>8</sup>

## 4.2 Approach based on Future-Spot Parity of CO<sub>2</sub>

In a world of certainty without transaction costs, with banking across all relevant periods, arbitrage between current and expected future compliance costs and between allowances of differing vintages will cause the immediate settlement prices, i.e., the spot price for allowances of different vintages, to be equal. In all periods, an individual holder of allowances chooses a level of GHG abatement such that the current marginal cost of abating in that period equals the current spot price. According to the theoretical frameworks of Bailey (1998) and Montero and Ellerman (1998), across any two periods with banking, the individual abates so that the discounted marginal costs of abatement are equal among all companies covered by the program. Because the discounted marginal cost of abatement should be equal across periods and current marginal cost of abatement is equal to the current allowance price in every period, the spot prices for allowances of differing vintages should be equal. That is to say, allowances of differing vintages should sell for the same price today. For example, the immediate settlement price for a vintage 2006 allowance in 2006 should equal the immediate settlement price of a vintage 2007 allowance in 2006. The allowance market, though, is marked by uncertainty and positive transaction costs. One source of uncertainty is the cost of compliance. Costs of abatement can be different than

<sup>8</sup>One might consider implementing the fundamental analysis into the mean equation of the return process and use GARCH-type structures, as discussed below, for the variance equation, though we leave such ideas for future possible research.



Figure 7: CO<sub>2</sub> prices and fuel switching costs from January to September 2005.  
Source: Enviros Climate Change Group.

expected. Second, actual emissions of GHG are stochastic even when abatement costs are certain, so that a firm's need for GHG emission allowances is also stochastic. For these reasons, GHG emissions in any given compliance year cannot be known with absolute certainty.

Because GHG emissions in any given year are uncertain and it is costly to purchase and sell allowances due to the much higher transaction costs (compared to those on the major stock exchanges), a firm benefits from holding a stock of allowances on hand to buffer itself against unexpectedly high prices.<sup>9</sup> The benefit that accrues from holding a stock of allowances on hand, called *convenience yield*, is the transaction cost saved from not having to make additional transactions and/or undo the transaction just done. To the extent that there is uncertainty and positive transaction costs, allowances of different vintages will not sell for the same price today. Therefore, a premium is placed on holding the commodity in the spot market rather than having a future position. The convenience yield tends to keep spot prices high relative to future prices. Such a relation between spot and future prices leads to a pricing structure called *backward*,

$$F_t(S, T) = Se^{(r-\delta)(T-t)},$$

where  $F$  is the futures price,  $t$  is the time at which the futures is evaluated,  $T$  is the maturity of the contract,  $S$  the spot price,  $r$  is the risk-free rate and  $\delta$  is the convenience yield. The first feature of a commodity market with backwardation is that the convenience yield is so large as to make the future price less than the spot price. Second, the future price decreases as time to

<sup>9</sup>A second reason could be the risk involved in buying back the permits due to the peculiarity of these traded assets as described in Section 5.1.

maturity increases.

Empirical evidence shows exactly the opposite of our expectations. Comparing prices of futures contracts on CO<sub>2</sub> emission allowances<sup>10</sup> provided by the European Climate Exchange with the spot prices, we obtained a *contango* term structure, i.e., the futures prices are greater than the spot prices. In the literature, a commodity with a *contango* structure leads to a *cost of carry* model that works best in pricing futures on commodities with large readily available inventories, that can be easily stored and accessed, and that have a stable supply and demand flow. This is in contradiction with what we stated above. Even if emission allowances have no cost for storing and are easily accessed, we illustrated how challenging the analysis is of supply fundamentals due to the complexity of several factors that play different roles. On the other hand, any inventory is deeply dependent on the actual level of emissions of GHG that are stochastic and thus hardly predictable. Therefore, the CO<sub>2</sub> emission allowance convenience yield does not behave according to our theory and a price scenario analysis based on the existence of a well-behaved future-spot parity is useless as long as the political uncertainty will largely affect long futures maturities. As such, a deeper understanding of the convenience yield behavior backed by longer time series, namely the determination of a stochastic rather than a deterministic process, must be taken into account for a more exhaustive and complete price dynamic analysis. Further consideration of the stochastic behavior of the convenience yield for CO<sub>2</sub> can be found in Trück et al. (2006).

## 5 An Econometric Approach for Analyzing the SO<sub>2</sub> Data

In the previous section, we explained some flaws in the approaches currently applied by the risk management offices of the companies covered by the ETS. An alternative approach explicitly designed to meet the statistical characteristics of the underlying data is to use a pure time series model. This section discusses in detail the analysis on the SO<sub>2</sub> spot price data set, using the 1,780 returns from January 4, 1999 to May 16, 2006. The CO<sub>2</sub> series is analyzed in Section 6.

### 5.1 Illiquidity

The emission allowances market is nonstandard due to the fact that the traded asset is itself a so-called *nonstandard commodity*. Electric power companies, for example, do not physically need the emission-right to produce and, therefore, to pollute; in most cases it is feasible to delay action and wait for new information before purchasing permits. This also applies the other way around: A firm that holds more permits than it expects to need may still hold onto the surplus because they have some option value, given that it may be costly to get them back once they are sold.

---

<sup>10</sup>We considered futures contract prices by October 20, 2005 with vintage December 2005, March 2006, June 2006, September 2006, December 2006 and December 2007-2008-2010 and 2012.

So, *illiquidity* arises endogenously from the fact that firms can emit without having permits and thus fear that they may face a market squeeze at the end of the year.

Historically, markets for permits have never been completely liquid. However, liquidity on the emission allowance markets has increased over time (based on traded volumes, see footnote 11), particularly from the end of 2005. Nevertheless, the data set exhibits a relatively large (29%) number of zero returns, due in part because of the relatively small number of agents interacting on the market. This is a common finding of exchange-traded assets which, on a daily scale, possesses a low floating capital and an even lower traded volume. For example, SO<sub>2</sub> is a regional problem in the U.S., where covered utilities include only a few hundred large energy producers (a few thousand facilities). Another plausible explanation (but which remains to be confirmed) is that within-firm trading at the same price could be taking place. For example, AES, the largest electricity producer in the U.S., has numerous facilities covered by the CAAA Title IV, and it is not unreasonable to assume that emission allowances are financially transferred from one balance sheet to another, at market price.<sup>11</sup>

## 5.2 Stable Paretian Distribution

Below we will present a statistical analysis of the returns, emphasizing the interplay between the standard features of the data (fat tails and volatility clustering) and the less-standard fact that the data exhibit a much greater percentage of zero-returns than the more commonly analyzed financial markets. This analysis will involve use of the stable Paretian distribution, which is of significant theoretical interest in probability theory; see, e.g., Gut (2005, Chapter 9). From a more practical point of view, it is a fat-tailed, possibly skewed distribution, and thus highly suited for the stylized facts of asset returns. It was introduced for the unconditional analysis of asset returns in the pioneering papers of Mandelbrot (1963) and Fama (1965), and has since gained considerable popularity.<sup>12</sup> Very briefly, the characteristic function of a symmetric stable random variable  $X$  is given by

$$\varphi_X(t; \alpha) = E[\exp\{itX\}] = \exp\{-|t|^\alpha\}, \quad 0 < \alpha \leq 2.$$

For  $0 < \alpha < 2$ , the fractional absolute moments of  $X$  of order  $\alpha$  and higher do not exist, i.e.,  $E[|X|^r]$  is finite for  $0 \leq r < \alpha$ , and infinite otherwise. For  $\alpha = 2$ ,  $\varphi_X(t; \alpha) = \exp\{-t^2\sigma^2/2\}$  with  $\sigma^2 = 2$ , i.e., for  $\alpha = 2$ ,  $X \sim N(0, 2)$ , while for  $\alpha = 1$ ,  $X$  is a Cauchy random variable. A standard calculation involving the inversion formula linking the characteristic function to the density of a

---

<sup>11</sup>At the time of writing, data to confirm this are not available, although from the USEPA website, the intra-company transaction volume is available (graphic available upon request), thus confirming at least that within-firm trading does take place on a large scale.

<sup>12</sup>See the monograph of Rachev and Mittnik (2000) and the survey paper of McCulloch (1997a) for an introduction, overview, and extensive bibliography of the use of the stable Paretian distribution in quantitative financial applications.

random variable yields

$$f_X(x) = \frac{1}{\pi} \int_0^\infty \cos(tx) e^{-t^\alpha} dt, \quad (1)$$

which can be used for evaluating the likelihood function of a model which uses independent stable variates as the residual term. Except for  $\alpha = 1$  and  $\alpha = 2$ , (1) is not expressible in closed-form and needs to be numerically evaluated. Particularly in financial applications, when the number of data points is large, it is far faster to use an algorithm based on the fast Fourier transform and interpolation for computing the likelihood; see Paoletta (2007) for a detailed presentation and computer code.

If  $X$  is a symmetric stable random variable with location parameter  $\mu$  and scale parameter  $c > 0$ , then we write  $X \sim S_\alpha(\mu, c)$ . That is, if  $Z \sim S_\alpha(0, 1)$  and  $X = \mu + cZ$ , then  $X \sim S_\alpha(\mu, c)$  with  $\varphi_X(t) = \exp(i\mu t - c^\alpha |t|^\alpha)$  and density at point  $x$  denoted by  $f_X(x; \alpha, \mu, c)$ .

While there are many distributions which exhibit fat-tails, the stable distribution has the summability (or stability) property (hence the name). In particular, with  $X_i \stackrel{\text{indep}}{\sim} S_\alpha(\mu_i, c_i)$  and  $S = \sum_{i=1}^n X_i$ ,

$$\varphi_S(t) = \prod_{i=1}^n \varphi_{X_i}(t) = \exp(i\mu_1 t - c_1^\alpha |t|^\alpha) \cdots \exp(i\mu_n t - c_n^\alpha |t|^\alpha) = \exp(i\mu t - c^\alpha |t|^\alpha),$$

i.e.,

$$S \sim S_\alpha(\mu, c), \quad \text{where} \quad \mu = \sum_{i=1}^n \mu_i \quad \text{and} \quad c = (c_1^\alpha + \cdots + c_n^\alpha)^{1/\alpha}.$$

Another appealing property of  $S_\alpha$  is its relation to the generalized central limit theorem, or GCLT, which states that the only non-trivial possible limit of normalized sums of independent and identically distributed (i.i.d.) random variables is stable Paretian (with special case the normal distribution). This is appealing because the innovation sequence, or error terms, in econometric models can often be interpreted as being the sum of a large number of (small) factors.

The other part of the name, Paretian, reflects the fact that the asymptotic tail behavior of the  $S_\alpha$  distribution is the same as that of the Pareto, i.e.,  $S_\alpha$  has power tails for  $0 < \alpha < 2$ . In particular, it can be shown that, for  $X \sim S_\alpha(0, 1)$ ,  $0 < \alpha < 2$ , as  $x \rightarrow \infty$ ,

$$\bar{F}_X(x) = \Pr(X > x) \approx k(\alpha) x^{-\alpha},$$

where  $k(\alpha) = \pi^{-1} \sin(\pi\alpha/2) \Gamma(\alpha)$  and  $a \approx b$  means that  $a/b$  converges to one in the limit.

The asymmetric stable distribution, denoted  $S_{\alpha,\beta}$ , involves the additional parameter  $\beta$ ,  $-1 \leq \beta \leq 1$ , and is asymmetric for  $\beta \neq 0$  and  $\alpha \neq 2$ . It also shares the properties of summability and asymptotic Pareto-tails, and preserves the relation to the GCLT. The SO<sub>2</sub> return series appears symmetric<sup>13</sup> so use of the  $S_{\alpha,\beta}$  distribution is not required, and we omit its details.

<sup>13</sup>Fitting the  $S_{\alpha,\beta}$  distribution via maximum likelihood yields the practically insignificant value of  $\hat{\beta} = 0.002$ ,

### 5.3 Basic Analysis

The top panel in Figure 8 plots the return data, from which we clearly see the presence of volatility clustering. The sample autocorrelation function (SACF) for the returns is typical in the sense that very little correlation structure is present in the data, and is not shown. Unsurprisingly, there is much stronger correlation involving the absolute and squared returns, with the SACF of the former shown in the middle panel of Figure 8. Given the large percentage of zero-returns in the data, one might expect that the correlation structure of the absolute returns would differ if the zeros were removed. This is indeed the case, and is shown in the bottom panel. Thus, other than the larger-than-usual number of zeros, the returns exhibit the usual stylized facts of asset returns, including a very low predictive component for the mean, strong volatility clustering, and tails which are far fatter than the normal.

To emphasize the tail behavior, Figure 9 shows a kernel density estimate of the returns data, but having removed the zeros (explaining the small dip in the curve near zero), which does not affect the tail of the distribution, but would otherwise jeopardize the quality of a fitted distribution.<sup>14</sup> We see that the nonzero returns are virtually symmetric, obviating the need for distributions which support asymmetry. The graph also shows an overlaid normal with matching mean and variance, and a location-scale  $S_\alpha(\mu, c)$  distribution (fit via maximum likelihood). While the normal fit is disastrous, the stable distribution fits the data rather well, with estimated tail index (and estimated standard error)  $\alpha = 1.453$  (0.045), location term  $\mu = 0.092$  (0.047) and scale term  $c = 0.980$  (0.037).

In light of the excess number of zeros, a conditional time series model for the returns would have to account for any dependency structure in the occurrence of zeros. To test this, we use the standard runs-test, reducing the data to a sequence of zeros (zero return) and ones (nonzero return), and using the asymptotic normality of the test statistic. Inspection of the return series shows that relatively more zeros occur in the first third of the data set (from January, 1999 up to March, 2002) than later, so that the runs-test applied to the whole series leads to blatant rejection of the null hypothesis. To account for this, we could test just, say, the last two thirds of

---

suggesting we can safely assume it is zero. One should keep in mind that a statistical test for symmetry based on  $\hat{\beta}$  is, in general, problematic because (i) it would be based on the asymptotic normality of the maximum likelihood estimator (DuMouchel, 1973) and the approximate standard errors obtained by computing the Hessian matrix (or use of a bootstrap), and so is not exact, and (ii) its behavior when the data are not i.i.d.  $S_{\alpha,\beta}$  is not known. Nevertheless, a conventional test for skewness involves computing the sample third moment of the data, but its theoretical counterpart, and the higher moments required to compute the variance of the sample skewness, may not exist in general for fat-tailed finance data, and do not exist if the data are stable Paretian with  $\alpha < 2$ . As such, a test based on  $\hat{\beta}$  might be preferred, or a test based on the significance of the asymmetry parameter of another (flexible, fat-tailed, asymmetric) distribution. This is generally referred to as partially adaptive estimation, and is useful because it avoids some of the disadvantages of nonparametric inference; see, for example, McDonald (1991, 1997).

<sup>14</sup>The kernel density of the data with the zeros looks similar, but with a higher peak near zero. Fitting a normal or stable distribution to the returns with the zeros leads to a very misleading fit in the tails. A mixture of normals could be fit to such data, and this is done in the more general context of the conditionally heteroskedastic model below.

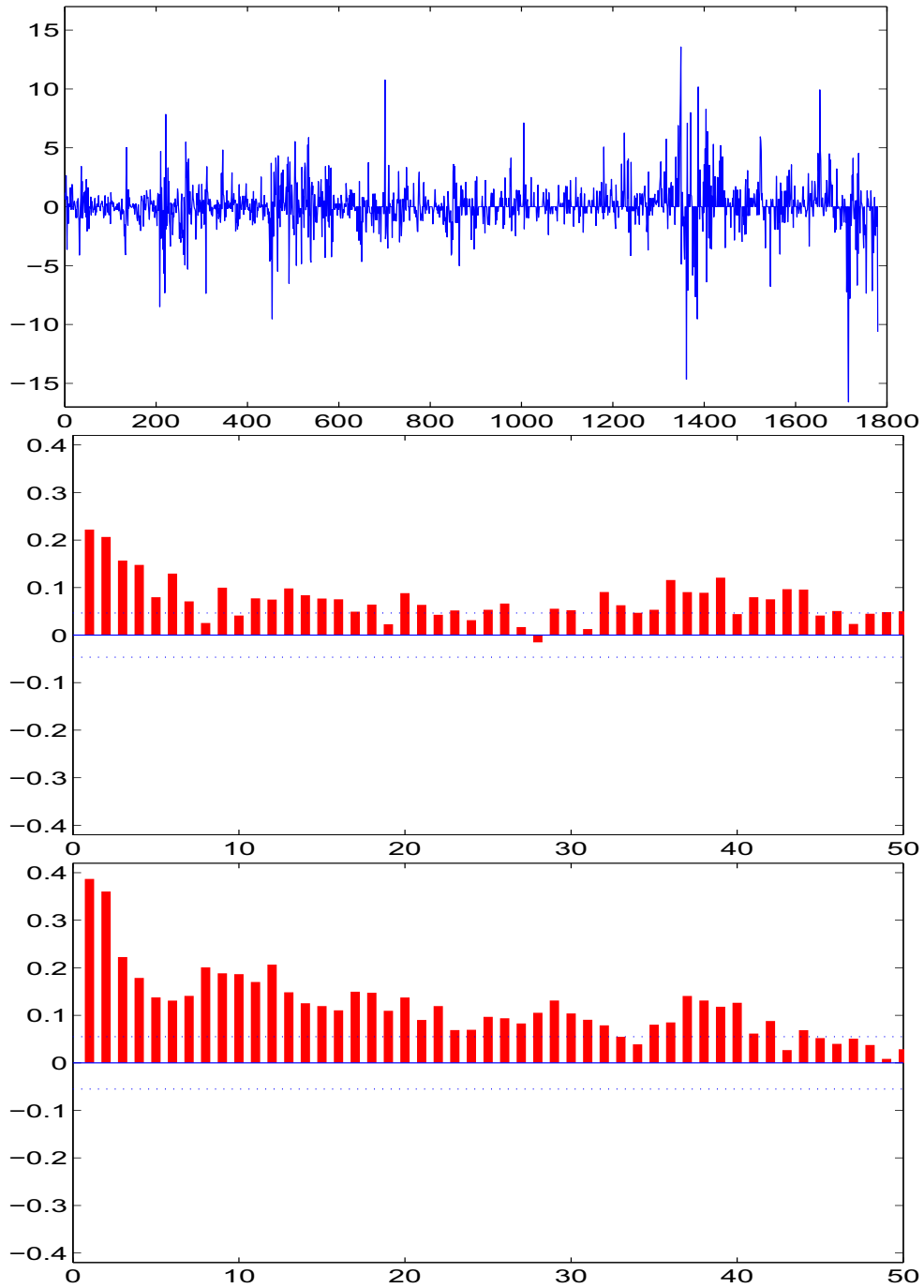


Figure 8: Daily SO<sub>2</sub> returns (top), the SACF of the absolute returns (middle) and the SACF of the zeros-removed absolute returns (bottom).

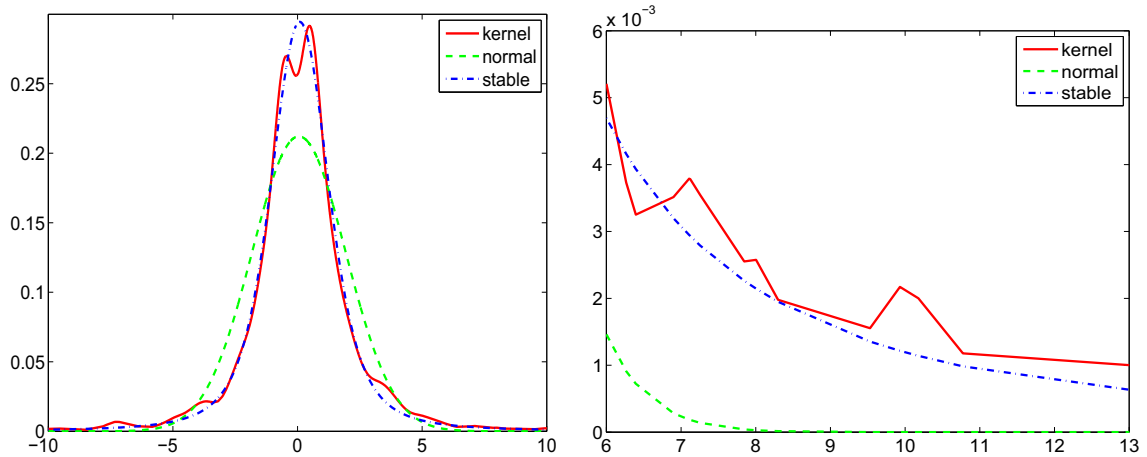


Figure 9: Kernel density (solid) of the SO<sub>2</sub> return series, with the best-fitting normal density (dashed) and best-fitting symmetric stable density (dash-dot). Right panel is just the magnified view of the right tail.

the data, but more informative is to perform the test on the return series starting from the  $s$ th return to the end of the series,  $s = 1, \dots, T - 50$ , where  $T = 1,780$  is the total number of returns. Figure 10 plots the resulting  $p$ -values, and shows that, for the latter half of the data set, the null hypothesis of no correlation structure in the occurrence of zeros cannot be rejected. This fact is important for the conditional model employed below; if the assumption of randomness of the zeros were not tenable, then a more complicated model involving Markov-switching structures would have been necessary.

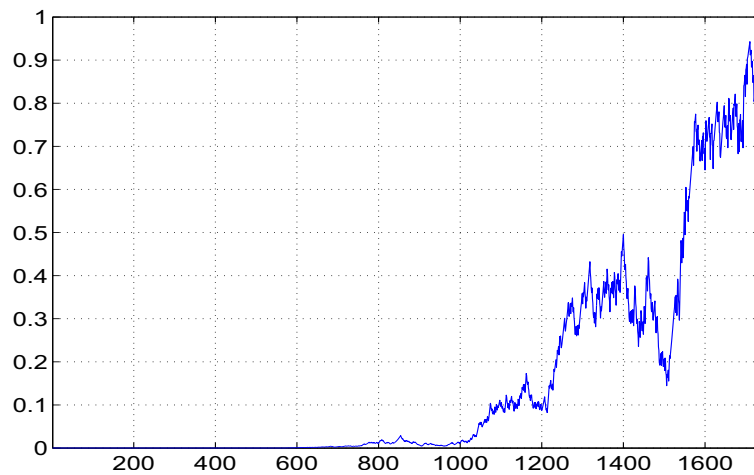


Figure 10: The  $p$ -values from the runs-test performed on segments of the SO<sub>2</sub> return series. The first segment is the returns in the whole series, the second is from the second return to the end, etc., up to the  $(T - 50)$ th observation to the end.

## 5.4 Stable GARCH Model

Perhaps the most common effective conditional model used in both academic and financial institutions contexts for the analysis of asset returns data is GARCH(1,1), as given in (6) below with  $r = s = 1$ , with the distributional assumption for the innovations having fatter-than-normal tails and, possibly, allowing for asymmetry (see, e.g., Kuester et al., 2005, and the references therein, for surveys). The arguably most popular among this class of distributions is the Student's  $t$ , yielding the  $t$ -GARCH model, as first studied by Bollerslev (1987). A less popular choice, due in part to the higher complexity of its numeric implementation, is the use of the (asymmetric) stable Paretian distribution.

The stable-Paretian-GARCH model, denoted  $S_{\alpha,\beta}$ -GARCH and given in (6) below, goes back to McCulloch (1985) and has subsequently been extended and used by several authors (see Mittnik et al., 2002, for references and the definition and stationarity conditions of the general case), and possesses at least three advantages over use of the  $t$ -GARCH. Firstly, the  $S_{\alpha,\beta}$ -GARCH in-sample fit and out-of-sample Value-at-Risk (VaR) forecasting ability are superior to  $t$ -GARCH (Mittnik and Paolella, 2003). Secondly, the use of stable Paretian innovations is theoretically more appealing because of its relation to the GCLT and which, via the stability property of the distribution, can be tested and confirmed to be applicable in some financial asset return series (Mittnik, Paolella and Rachev, 2000; Paolella, 2001). Lastly, because the tail index,  $\alpha$ , is estimated jointly in a conditional volatility model, it is directly related to the risk measures obtained by methods in extreme value theory and nonparametric tail estimation, which are critically useful tools in quantitative financial risk management; for detailed discussion of this important relation and how it is used in practice, see, e.g., the textbook presentations in Tsay (2002), Dowd (2005), Christoffersen (2003), Ruppert (2004) and McNeil et al. (2005).

For the SO<sub>2</sub> data, because of the aforementioned issue with a preponderance of zeros in the return series, the data generating process (DGP) is not consistent with any typical distributional assumption (such as Student's  $t$  or stable Paretian), in either the unconditional or conditional (GARCH) case (though a mixture model suggests itself, and is the method used below in the conditional modeling section). Because fatter-than-normal tails of a unimodal distribution imply a more peaked center, the excess amount of zeros will have the effect of causing the tail index (the thickness of the tail) to be biased downwards (thicker), thus overestimating the risk of extreme tail events.

To illustrate, we estimate<sup>15</sup> the  $S_{\alpha,\beta}$ -GARCH model for the SO<sub>2</sub> return series. As with the unconditional stable Paretian model for the returns, the estimated asymmetry term is  $\hat{\beta} = -0.003$ , which is practically and statistically insignificant. Note that the resulting estimate of the tail index  $\alpha$  pertains to the stable Paretian innovations of the GARCH process describing the returns, i.e.,

---

<sup>15</sup>See Mittnik, Paolella and Rachev (1998b, 2002) for details on the joint maximum likelihood estimation of the parameters of the model.

the GARCH effects (which also give rise to the fat-tails of the returns) are taken into account, so that the resulting index will be greater (i.e., correspond to thinner tails) than the unconditional counterpart. The resulting estimate of the tail index (and approximate standard error) are  $\hat{\alpha} = 1.0278$  (0.015), suggesting the plausibility of Cauchy ( $\alpha = 1$ ) innovations, which do not even possess a finite mean and is therefore not a tenable assumption. If we restrict  $\hat{\alpha}$  to be above 1.3, we obtain  $\hat{\alpha} = 1.498$  (0.018), showing that a local maximum of the likelihood does exist in a “plausible” region of the sample space. This occurs because the innovations in the conditional model are not stable Paretian (there are too many zeros, yielding the near-Cauchy fit), but the tails are thinner than Cauchy, which resulted in the trade-off value for  $\hat{\alpha}$  of about 1.5.

If, for illustrative purposes, we strip all the zeros from the return series and then fit the  $S_{\alpha,0}$ -GARCH model, we obtain  $\hat{\alpha} = 1.640$  (0.022). This value is in agreement with the range of estimated tail indexes of numerous other financial data sets (see Lau et al., 1990; Mittnik and Paoletta, 2003, and the references therein) and is a far better reflection of the true thickness of the (conditional) tail. However, given the ad-hoc nature of the removal of zeros, this is still an unsatisfying approach for building a realistic model of the return series. Similar results are obtained if we use the  $t$ -GARCH model.

In fact, the abundance of zeros also precludes effective use of other GARCH-type models which otherwise tend to perform excellent in terms of VaR forecasting. In particular, one might think that the GARCH-EVT of Diebold et al. (1998) and McNeil and Frey (2000), which focuses on tail estimation of the residuals of GARCH-filtered returns via methods of extreme-value-theory, would be particularly suited for VaR prediction of the SO<sub>2</sub> data, given its unique behavior in the center (but not the tails) of the distribution. The problem, however, is that the choice of innovations assumption used with the GARCH filter in the first step of the GARCH-EVT model is decisive for its forecasting performance, as detailed in Kuester et al. (2005). Thus, the same problem arises as with the use of conventional fat-tailed-GARCH models. Similar findings, as detailed in Kuester et al. (2005), apply to the use of the otherwise highly effective method of filtered historical simulation (FHS), as proposed and studied by Barone-Adesi et al. (1999, 2002). More encouragingly, Kuester et al. (2005) show in their study that the mixed-normal GARCH model (which is our proposed solution to the zeros problem) delivers highly competitive VaR forecasts on par with the quality of GARCH-EVT and FHS.

Because of the zeros-problem and the resulting non-applicability of the aforementioned conditional models, we explore two options which have more validity. The first is the unconditional analysis of the tails of the data, i.e., ignoring the volatility clustering and, as the zeros are in the center, directly avoiding the zeros-problem. Such an approach is both common and advantageous in situations when interest focuses on risk measures and the tail of the distribution; see the discussions in DuMouchel (1983), Hols and de Vries (1991) and Danielsson, Jansen and deVries (1992) for details. The second is a conditional analysis, whereby a model is considered which adequately

accounts for the GARCH-type effects and also the zeros-problem. The following sections detail these two approaches.

## 5.5 Unconditional tail estimation

The Hill estimator of the tail thickness of an i.i.d. data set  $X_1, \dots, X_n$ , coined after the seminal paper by Bruce Hill (1975) in which it was introduced, is by far the most commonly used tail estimator for the tail index of a distribution (Kratz and Resnick, 1996; Resnick, 1997). With  $X_{j:n}$  denoting the  $j$ th order statistic of the sample, it is given by

$$\hat{\alpha}_{\text{Hill}}(k) = \frac{1}{(1/k) \sum_{j=1}^k \ln(X_{n+1-j:n}) - \ln X_{n-k:n}}, \quad (2)$$

with approximate standard error

$$\widehat{\text{SE}}(\hat{\alpha}_{\text{Hill}}; k) = \frac{k \hat{\alpha}_{\text{Hill}}(k)}{(k-1)(k-2)^{1/2}}, \quad k > 2. \quad (3)$$

If the right tail of the distribution is asymptotically Pareto (as  $x \rightarrow \infty$ ,  $\bar{F}(x) \approx cx^{-\alpha}$  for some constant  $c > 0$ ), then, given an appropriate choice of  $k$ ,  $\hat{\alpha}_{\text{Hill}}(k)$  provides an estimate of Pareto tail index  $\alpha$ . Recall that the Pareto distribution possesses absolute moments of order less than  $\alpha$ . Because of the relationship between the Pareto distribution (for  $0 < \alpha < 2$ ) and the tails of the stable Paretian distribution, the Hill estimator is also commonly used to estimate the stable index  $\alpha$ .

Moreover, it was shown by Mason (1982) that the Hill estimator is consistent for distributions in the (max-) domain of attraction of the (max-) type extreme value distributions  $\exp(-x^{-\alpha})$ . These include, among others, the Student's  $t$ , Pareto and stable Paretian distributions. Goldie and Smith (1987) proved its asymptotic normality,  $(\hat{\alpha}_{\text{Hill}}^{-1} - \alpha^{-1}) k^{1/2} \sim N(0, \hat{\alpha}^{-2})$ , (as would be suggested from the consistency and (3)), so that standard inference procedures can be used (see, for example, Koedijk, Schafgans and de Vries, 1990; and Lux, 1995).<sup>16</sup>

In practice, with finite samples, the choice of  $k$  involves a tradeoff: it must be sufficiently small so that  $X_{n-k:n}$  is in the tail of the distribution; but if too small, the estimator will lack precision. Its optimal value will depend not only on the sample size, but also on the unknown tail index itself, resulting in inevitable and potentially “outrageous” bias (Kratz and Resnick, 1996). This problem has traditionally thwarted effective use of the Hill estimator, and has prompted consideration of improved tail estimators, such as those proposed by Danielsson, Jansen and deVries (1996), Pictet et al. (1996) and Resnick and Stărică (1997). The performance of all these

---

<sup>16</sup>The rate at which  $k(n)$  increases to infinity is specified in the main result of Goldie and Smith (1987); see also Hall (1982) and Hall and Walsh (1984). Further details and references regarding the consistency of the Hill estimator can be found in Dekkers, Einmahl and de Haan (1989) and Dekkers and de Haan (1993).

estimators also depends on unspecified parameters like  $k$ , leaving this issue still unresolved.

Figure 11 plots the Hill estimates of the the SO<sub>2</sub> spot price data set, as a function of  $k$ , based on the 1,780 sorted absolute returns. (Taking absolute returns effectively doubles the sample size, assuming the left and right tails behave the same, which appears to be the case from the analysis of the asymmetry term given above.) The graph is typical of Hill estimator plots applied to financial returns data, and a sizeable region for which  $\hat{\alpha}_{\text{Hill}}$  is “flat”, or roughly constant in  $k$ , cannot be found. What can be said is that, for  $k > 200$ , or a choice of  $k$  corresponding to at least 11% of the sample size,  $\hat{\alpha}_{\text{Hill}}$  is less than two, emphasizing the extremely highly fat-tailed nature of the data and also the possible justification of use of the stable Paretian assumption (recall  $\alpha$  needs to be less than two). In fact, as emphasized in McCulloch (1997b) and Mittnik, Paoletta and Rachev (1998b), under the assumption of stable Paretian data, the choice of  $k$  should be much larger than the usually recommended values of 5% or 10%, and the unexpectedly large choice for  $k$  of 40% of the sample size is recommended. Using this (giving  $k = 712$ ), yields  $\hat{\alpha}_{\text{Hill}} = 1.1$ , which, interestingly, is not far from the value obtained via maximum likelihood estimation given above when the zeros were not removed from the data set.

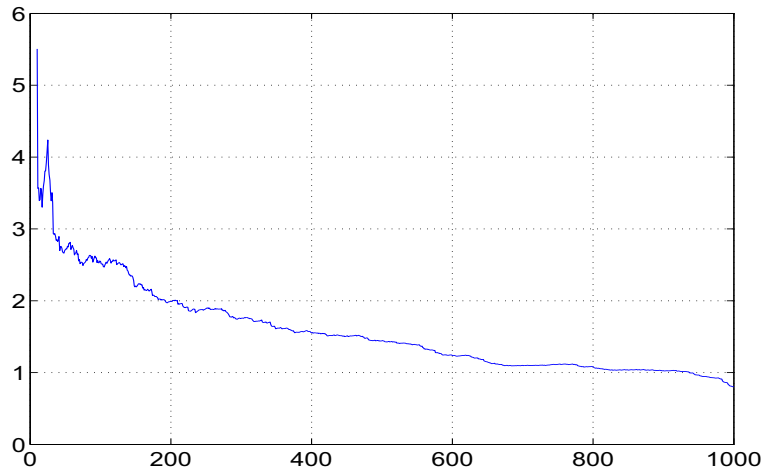


Figure 11: The Hill (1975) tail estimates as a function of cutoff value  $k$  for the SO<sub>2</sub> returns.

The Hill estimator for any value of  $k$  cannot be considered particularly reliable. A tail estimator designed explicitly for stable Paretian data and which exhibits excellent small sample properties which can exceed those of maximum likelihood estimators was developed in Mittnik and Paoletta (1999) and referred to as the Hill–intercept estimator, or  $\hat{\alpha}_{\text{Hint}}$ . It is based on a set of Hill estimators for a range of  $k$ -values, and computed as

$$\hat{\alpha}_{\text{Hint}} = -0.8110 - 0.3079 \hat{b} + 2.0278 \hat{b}^{0.5}, \quad (4)$$

where  $\hat{b}$  is the intercept in the simple linear regression of  $\hat{\alpha}_{\text{Hill}}(\mathbf{k})$  on  $\mathbf{k}/1000$ , where the elements of  $\mathbf{k}$  are such that  $0.2T \leq k \leq 0.8T$  in steps of  $\max\{\lfloor T/100 \rfloor, 1\}$ . Besides being trivially computed,

even in samples as small as  $T = 50$ ,  $\hat{\alpha}_{\text{Hint}}$  is unbiased for  $\alpha \in [1, 2]$  and virtually exactly normally distributed. Furthermore, for given sample size, the variance of  $\hat{\alpha}_{\text{Hint}}$  is practically constant across  $\alpha$ , reaching its maximum for  $\alpha = 1.5$ . For sample sizes  $50 < T < 10,000$ , this is approximated by

$$\widehat{\text{SE}}(\hat{\alpha}_{\text{Hint}}) \approx 0.0322 - 0.00205T_* + 0.02273T_*^{-1} - 0.0008352T_*^{-2}, \quad (5)$$

where  $T_* = T/1000$  and  $T$  is the sample size. For the  $\text{SO}_2$  returns, we obtain  $\hat{\alpha}_{\text{Hint}} = 1.46$  with standard error 0.043. This value is much more in accordance with tail-index values observed for other financial series, and is practically the same as the value of the maximum likelihood estimator of the tail index using the zero-stripped return series, as discussed above. As such, we recommend use of the tail index 1.46 for computing VaR estimates, as discussed in the textbooks mentioned above in Section 5.4, as well as the methodology discussed in Novak and Beirlant (2006), which relies on extreme value theory and nonparametric tail estimation for evaluating the magnitude of possible market crashes, and computing VaR and expected shortfall measures.

## 5.6 Mixed Normal GARCH: Model and Numerical Issues

The traditional power-GARCH( $r, s$ ) model is given by

$$r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t, \quad \sigma_t^d = \theta_0 + \sum_{i=1}^r \theta_i |\varepsilon_{t-i}|^d + \sum_{j=1}^s \phi_j \sigma_{t-j}^d, \quad (6)$$

for  $d > 0$  and (sufficient for ensuring  $\sigma_t > 0$  and, for  $r = s = 1$ , also necessary),  $\theta_0 > 0$ ,  $\theta_i \geq 0$ ,  $i = 1, \dots, r$ , and  $\phi_j \geq 0$ ,  $j = 1, \dots, s$ , and where  $z_t \stackrel{\text{iid}}{\sim} f_Z(\cdot)$  with  $f_Z$  a zero-location, unit-scale continuous probability density function (pdf). In Bollerslev's (1986) model,  $f_Z$  is the Gaussian density and  $d = 2$ , while in the  $S_{\alpha, \beta}$ -GARCH model,  $f_Z$  is the  $S_{\alpha, \beta}$  density and  $d = 1$ .<sup>17</sup> The problem with this formulation for the data under study is its excess number of zeros, which precludes use of the usual array of distributions used in this context, such as Gaussian, Student's  $t$ , skewed  $t$  extensions, and  $S_{\alpha, \beta}$ .

One candidate distribution which is perfectly suited for capturing this phenomenon is to use a mixture-model, i.e., taking  $f_Z$  to be a weighted sum of two or more pdfs. It might seem natural to have one component be degenerate at zero, and the other(s) continuous, but it suffices, and is operationally easier to implement, to choose all components of  $f_Z$  to be continuous pdfs, in this case, each Gaussian, with the first one possessing a very small (but positive) variance and a mean at zero (more on this below).

An advantage of the mixture model not shared by other distributional assumptions is that it lends itself to economic interpretation in several ways. A mixture of two or more normals could

<sup>17</sup>In general,  $d$  can be estimated, with  $0 < d < \alpha$ , but in practice,  $\alpha > 1$  and, for numerous financial return series, the out-of-sample forecasting ability is barely affected by the choice of  $d \in [1, \alpha)$ .

arise from different groups of actors, with one group acting, for example, more volatile than the others or, possibly, processing market information differently. In this last case, as mentioned in Section 3.1, with the approaching of the more stringent Phase II in Title IV and with the SO<sub>2</sub> emission level taking shape for the different utilities, companies obtained a better indication of their Phase I net positions and some appeared to refrain from speculation covering their short positions on a forward basis and saving the remaining allowances. This uncertainty, combined with the nonstandard feature of this commodity resulted in no trading, also explains part of the zero return value.

A GARCH-type model with mixed normal innovations, denoted MixN-GARCH, has already been proposed and studied independently by Haas et al. (2004a) and Alexander and Lazar (2006).<sup>18</sup> It was not designed with the zeros-problem in mind but rather motivated by (i) the aforementioned economic interpretation of different groups of market participants, as was pursued in the unconditional (non-GARCH) case by Fama (1965), Kon (1984), Tucker and Pond (1988), and Aparicio and Estrada (2001); and (ii) the fact that the mixture of normals distribution is extremely flexible, fat-tailed and asymmetric, thus easily able to capture the distributional regularities of financial returns data. A third benefit of the model is that it automatically induces time-varying skewness and kurtosis, which have been advocated in this context by Hansen (1994), Harvey and Siddique (1999), Rockinger and Jondeau (2002), and Brännäs and Nordman (2003). Finally, and of great practical importance, Haas et al. (2004a) and Kuester et al. (2005) have demonstrated that the model offers a plausible decomposition of the contributions to market volatility, and also delivers highly competitive out-of-sample VaR forecasts.

The time series  $\{\epsilon_t\}$  is generated by an  $n$ -component MixN-GARCH( $r, s$ ) process if the conditional distribution of  $\epsilon_t$  is an  $n$ -component mixed normal with zero mean, i.e.,

$$\epsilon_t \mid \mathcal{F}_{t-1} \sim \text{MN}(\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}_t^2), \quad (7)$$

where the mixed normal pdf is given by

$$f_{\text{MN}}(z; \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = \sum_{j=1}^n \omega_j \phi(z; \mu_j, \sigma_j^2),$$

where  $\phi$  is the normal pdf,  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)'$  is the set of component weights such that  $\omega_j \in (0, 1)$  and  $\sum_{j=1}^n \omega_j = 1$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$  is the set of component means, such that, to ensure  $\text{E}[\epsilon_t] = 0$ ,  $\mu_n = -\sum_{j=1}^{n-1} (\omega_j/\omega_n) \mu_j$ , and  $\boldsymbol{\sigma}_t^{(2)} = (\sigma_{1t}^2, \dots, \sigma_{nt}^2)' \in \mathbb{R}_+^n$  are the positive component variances

---

<sup>18</sup>The appealing theoretical and empirical properties of the model have led to a number of extensions, including a Bayesian analysis by Bauwens and Rombouts (2005), methods of explicitly capturing the leverage effect (Alexander and Lazar, 2005; Haas et al., 2006a), generalization to Markov switching structures (Haas et al., 2004b; Bauwens et al., 2006) and different multivariate extensions (Bauwens et al., 2006; Haas et al., 2006b).

at time  $t$ . It is straightforward to show that, if  $Y \sim \text{MN}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\sigma}^{(2)})$ , then

$$\text{E}[Y] = \boldsymbol{\omega}'\boldsymbol{\mu}, \quad \text{and} \quad \text{Var}(Y) = \boldsymbol{\omega}'(\boldsymbol{\sigma}^{(2)} + \boldsymbol{\mu}^{(2)}) - (\boldsymbol{\omega}'\boldsymbol{\mu})^2, \quad (8)$$

where  $\boldsymbol{\mu}^{(p)} = (\mu_1^p, \dots, \mu_n^p)'$ . In order to capture the dynamics in the second (and higher) moments of the returns, the  $n \times 1$  component variances  $\boldsymbol{\sigma}_t^{(2)}$  are allowed to evolve according to the GARCH-like structure

$$\boldsymbol{\sigma}_t^{(2)} = \boldsymbol{\gamma}_0 + \sum_{i=1}^r \boldsymbol{\gamma}_i \epsilon_{t-i}^2 + \sum_{j=1}^s \boldsymbol{\Psi}_j \boldsymbol{\sigma}_{t-j}^{(2)}, \quad (9)$$

where  $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, \dots, \gamma_{in})'$ ,  $i = 0, \dots, r$ , are  $n \times 1$  vectors, and  $\boldsymbol{\Psi}_j$ ,  $j = 1, \dots, s$ , are  $n \times n$  matrices with typical entry  $\psi_{jhk} = [\boldsymbol{\Psi}_j]_{hk}$  (which we write as  $\psi_{hk}$  when, as in most applications,  $s = 1$ ). We restrict  $\boldsymbol{\Psi}_j$  to be diagonal, which, as discussed in Haas et al. (2004a), yields a much more parsimonious model with little loss in goodness-of-fit. In this case, and with  $r = s = 1$ , the parameter constraints  $\gamma_{0i} > 0$ ,  $\gamma_{1i} \geq 0$ , and  $\psi_{ii} \geq 0$ ,  $i = 1, \dots, n$ , are necessary and sufficient to ensure the nonnegativity of the variance terms. With one component ( $n = 1$ ), the model reduces to the Bollerslev (1986) GARCH model. With two or more components, the model is able to capture the asymmetry and most of the excess kurtosis common in normal-GARCH residuals. Moreover, with  $n \geq 2$ , the structure of (9) also gives rise to rich conditional dynamics in the 2nd, 3rd and 4th moments which cannot be modeled by the classic GARCH model of the form (6) with *any* distributional assumption, but do appear in real financial returns data (see Haas, et al., 2004a).

It has been found that the component of the mixture assigned to the most volatile observations often consists of randomly and infrequently occurring jumps in the volatility, so that a GARCH structure is not required. We denote by  $\text{MixN}(n, g)$  the model given by (7) and (9), with  $n$  component densities, but such that only  $g$ ,  $g \leq n$ , follow a GARCH process (and  $n - g$  components restricted to be constant). In the context of the  $\text{SO}_2$  returns, the component which picks up the zeros will also not require a GARCH component, so that we only entertain models of the form  $\text{MixN}(n, g)$ ,  $1 \leq g < n$ , for  $n \geq 2$ , and, for comparison purposes, the  $\text{MixN}(1, 1)$ , which is just the Bollerslev (1986) GARCH model.

In line with the vast majority of studies involving the Bollerslev (1986) GARCH model and those involving the  $\text{MixN-GARCH}(r, s)$ , the choice  $r = s = 1$  has been found to be adequate for the  $\text{SO}_2$  returns, and we subsequently suppress reference to  $r$  and  $s$ . As is common in GARCH applications, the  $\text{AR}(1)$  structure  $r_t = a_0 + a_1 r_{t-1} + \epsilon_t$  is included in the model to capture the extremely mild autocorrelation structure in the mean. Thus, all future reference to a  $\text{MixN}(n, g)$  model implies an  $\text{AR}(1)$ - $\text{MixN-GARCH}(1, 1)$  structure with diagonal  $\boldsymbol{\Psi}_1$  matrix. So, for example, in the  $\text{MixN}(3, 2)$  case (which, anticipating our results below, is the preferred model), (9) takes

the form

$$\begin{bmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \\ \sigma_{3t}^2 \end{bmatrix} = \begin{bmatrix} \gamma_{01} \\ \gamma_{02} \\ \gamma_{03} \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \gamma_{12} \\ 0 \end{bmatrix} \epsilon_{t-1}^2 + \begin{bmatrix} \psi_{11} & 0 & 0 \\ 0 & \psi_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_{1,t-1}^2 \\ \sigma_{2,t-1}^2 \\ \sigma_{3,t-1}^2 \end{bmatrix}. \quad (10)$$

In this case, there are 13 parameters to estimate. These are listed in Table 4, noting that  $\mu_3$  and  $\omega_3$  are constrained, as discussed above, and  $\gamma_{13}$  and  $\psi_{33}$  are held at zero.

In general, the number of components,  $n$ , needs to be determined empirically. As discussed in Haas et al. (2004), standard model likelihood-based selection criteria can be successfully employed to compare models with different numbers of components. For a model with  $K$  parameters,  $T$  observations and log-likelihood  $L$ , evaluated at the maximum likelihood estimator, we report the  $AIC = -2L + 2K$  and  $BIC = -2L + K \log T$ ; see Burnham and Anderson (2003) for original references and a detailed textbook presentation of the use of these measures.

The likelihood of the general  $AR(p)$ -MixN-GARCH( $r, s$ ) model is straightforward to program and evaluate, and its numeric maximization has proven to be unproblematic using standard quasi-Newton-type optimization routines (as implemented in Matlab). One nonstandard issue which arises in the context of the data in our study involves the point masses at zero, which are picked up as one of the  $n$  components in the MixN( $n, g$ ) model. Because these form a degenerate distribution, one variance component, namely  $\gamma_{03}$  in (10), is zero, and the likelihood is not defined. One way around this is to set  $\gamma_{03}$  and  $\mu_3$  to zero and  $\omega_3$  (the weight of this component) to the percentage of zeros in the data set. This turns out to be problematic because the other normal components (which are close to centered around zero) have a “gap” at zero, which (given the discrete nature of the returns data) renders the normal distribution inappropriate. Instead, we propose to replace the zero returns with realizations of i.i.d. normally distributed random variables with mean zero and standard deviation  $\sigma_k$ , where  $\sigma_k$  is chosen to be a small number relative to the unconditional variance of the returns.

At first blush, this appears to be an uncomfortable solution, because adding random noise to the data implies both a loss of “objectivity” as well as non-reproducibility of our estimation results. With respect to the first issue of objectivity, we note the relationship between this approach and that of Hamilton (1991), which is a quasi-Bayesian approach to estimating the (unconditional) mixed normal distribution involving (algebraically) adding observations to the data which reflect prior information and then maximizing a quasi-likelihood. His method not only results in greater numeric stability of estimation, but also (with a nonzero and very small amount of prior information) leads to better small-sample estimation performance than pure maximum likelihood. In our model context, we observe the same results in terms of improved reliability of the numerical maximization of the likelihood and, as verified by simulations, more accurate parameter estimates. For the second issue of non-reproducibility, we remark that the parameter estimates are not sensitive (with respect to their approximate standard errors) to a

range of  $\sigma_k$  values from 0.01 to 0.2. In what follows, we use this method with  $\sigma_k = 0.1$  and estimate model (10) without any constraints on  $\gamma_{03}$ ,  $\mu_3$  and  $\omega_3$ .

### 5.7 Mixed Normal GARCH: Estimation Results and Diagnostics

The top segment of Table 3 reports the likelihood-based goodness-of-fit measures for the various MixN( $n, g$ ) fitted models. As expected, the worst performer is MN(1,1), the standard (one-component) normal-GARCH model. The best model, by far, is the MN(3,2). To help confirm that there is no structure in the pattern of zeros throughout the return series, we also estimated the MN(2,2) model. Observe that its likelihood is virtually the same as the MN(2,1), showing that there is no GARCH dynamic in the component which picks up the zero-returns.

Model	K	L	AIC	BIC
MN(1,1)	5	-4072.0	8134.0	8181.42
MN(2,1)	8	-2919.6	5855.2	5899.07
MN(2,2)	10	-2919.3	5858.6	5913.44
MN(3,1)	11	-2873.8	5769.6	5829.93
MN(3,2)	13	-2835.7	<b>5697.4</b>	<b>5768.70</b>
MN(4,2)	16	-2834.5	5701.0	5781.75
MN(4,3)	18	-2831.6	5699.2	5797.92

Table 3: Likelihood-based goodness-of-fit for SO<sub>2</sub>. The best values for each criteria are marked in boldface. MN(1,1) coincides with the usual, AR(1)-normal-GARCH model, MN( $n, g$ ) is the  $n$ -component AR(1)-mixed-normal-GARCH(1,1) model (7) and (9) with diagonal  $\Psi_1$  matrix and such that only  $g$  of the  $n$  components are endowed with a GARCH structure.

The parameter estimates of the MN(3,2) model are shown in Table 4. Observe that the weight of the component associated with the zero-returns (the third component) is 23.9%, which is somewhat less than the unconditional (model free) estimate of the percentage of zeros (29%) in the data set, because the other two components are normal distributions which have their modes near zero, and thus account for some of the zero-returns.

The responsiveness to shocks of the individual components is illustrated in Figure 12. It shows the SO<sub>2</sub> return time series and (the square roots of) the fitted variances of the  $g = 2$  GARCH components in the MN(3,2) model. The relatively large value of the constant  $\hat{\gamma}_{01} = 0.621$  in the first component is reflected in the high floor of  $\sigma_{1t}$  at roughly one.

From the general formula in Haas et al. (2004a), the unconditional variance of  $\epsilon_t$  is, for the diagonal model with  $r = s = 1$  and  $n$  components,

$$E[\epsilon_t^2] = \frac{\sum_{j=1}^n \omega_j \mu_j^2 + \sum_{j=1}^n \omega_j \gamma_{0j} / (1 - \psi_{jj})}{\sum_{j=1}^n \omega_j (1 - \gamma_{1j} - \psi_{jj}) / (1 - \psi_{jj})}, \quad (11)$$

which indeed reduces for  $n = 1$  component to  $\theta_0 / (1 - \theta_1 - \phi_1)$ , as given in Bollerslev (1986) for

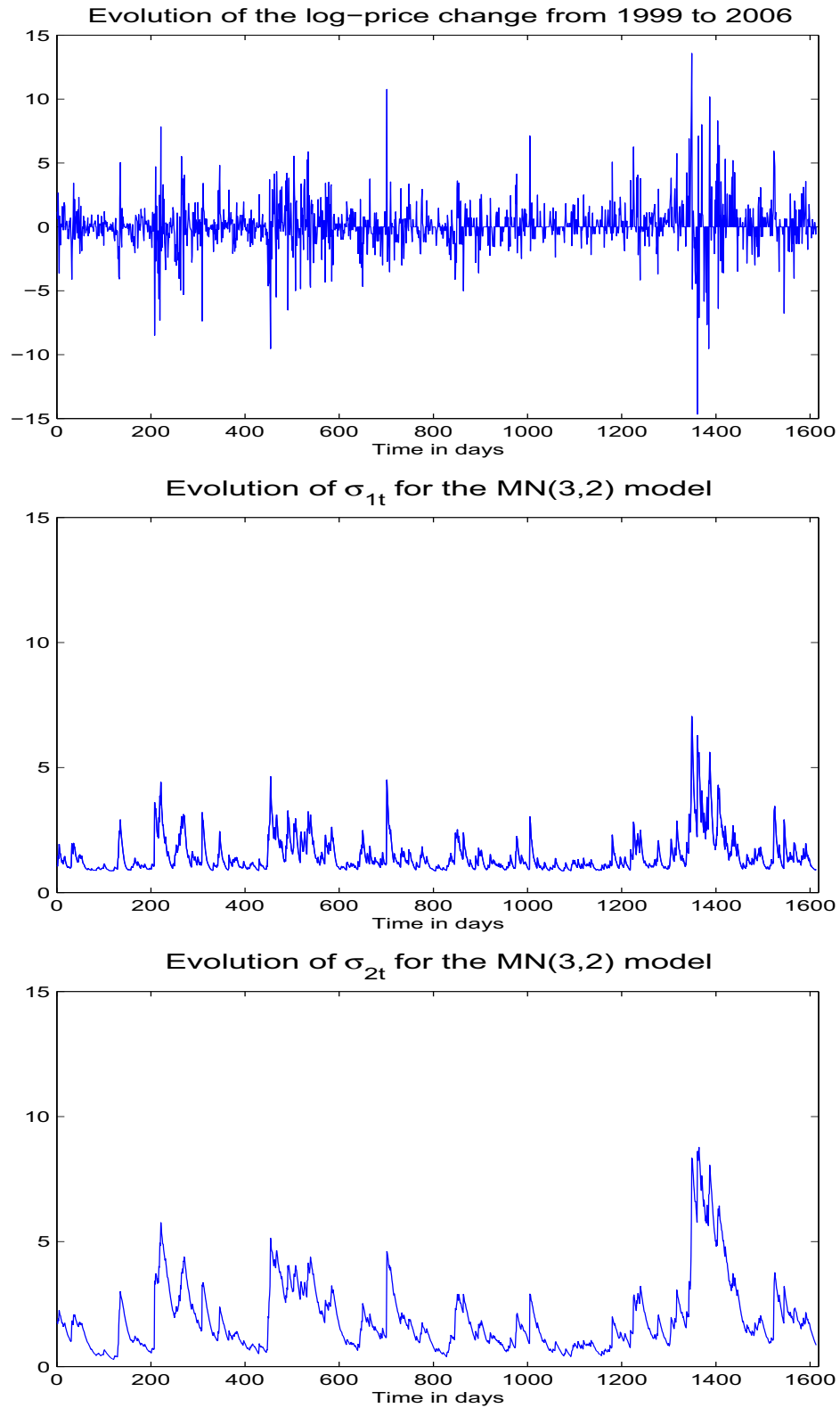


Figure 12: Daily return time series for the SO<sub>2</sub> and volatility evolution for the MN(3,2)-GARCH model.

Table 4: Maximum likelihood parameter estimates of the mixed normal GARCH models for the SO<sub>2</sub> allowances price return 1999-2006.

Param	MN(2,1)	MN(3,1)	MN(3,2)
$a_0$	0.046	0.040	0.041
$a_1$	0.000	0.000	0.000
$\gamma_{01}$	0.137	0.157	0.621
$\gamma_{11}$	0.243	0.235	0.427
$\psi_{11}$	0.797	0.867	0.846
$\omega_1$	0.709	0.440	0.165
$\mu_1$	0.019	0.012	-0.001
$\gamma_{02}$	0.001	0.505	0.121
$\gamma_{12}$	-	-	0.212
$\psi_{22}$	-	-	0.649
$\omega_2$	0.290	0.334	0.595
$\mu_2$	-0.047	0.019	0.001
$\gamma_{03}$	-	0.012	0.013
$\gamma_{13}$	-	-	-
$\psi_{33}$	-	-	-
$\omega_3$	-	0.226	0.239
$\mu_3$	-	-0.046	-0.045

the usual normal GARCH(1,1) model (6) with  $d = 2$ . For the MN(3,2) model (10) with estimated parameter values given in the last column of Table 4, expression (11) is 4.800, which is close to the sample variance of the nonzero returns (4.97), and larger than the sample variance of all the returns (3.54).

To help confirm that the fitted MixN(3,2) model successfully captures the distributional properties of the SO<sub>2</sub> return series, Figure 13 shows a QQ-plot of the actual data and a (same length) simulated time series generated from the fitted model. The graph is typical of numerous generated ones, and shows that the entire distribution, most notably the tails, is well-captured by the proposed model. Besides demonstrating that the MN(3,2) model is a plausible approximation to the true (and undoubtedly far more complicated) DGP, the excellent fit in the tails has obvious implications for calculations of risk measures such as VaR and expected shortfall.

In addition to an excellent in-sample fit, the mixture model in this context allows for a potential interpretation of the components. Clearly, the third component is used to pick up the zeros and embed them adequately in a stochastic process. The remaining two GARCH mixture components can be viewed as capturing the result of the two major groups of market participants: affected units who buy and sell allowances based primarily on their current and forecasted needs (i.e., the allowances are viewed as a factor of production), and speculative traders or simply non-affected agents (i.e. banks and investment funds). Of course, these two groups could possibly interact. Observe that the general mixed normal GARCH model (7) and (9) allows for a type of

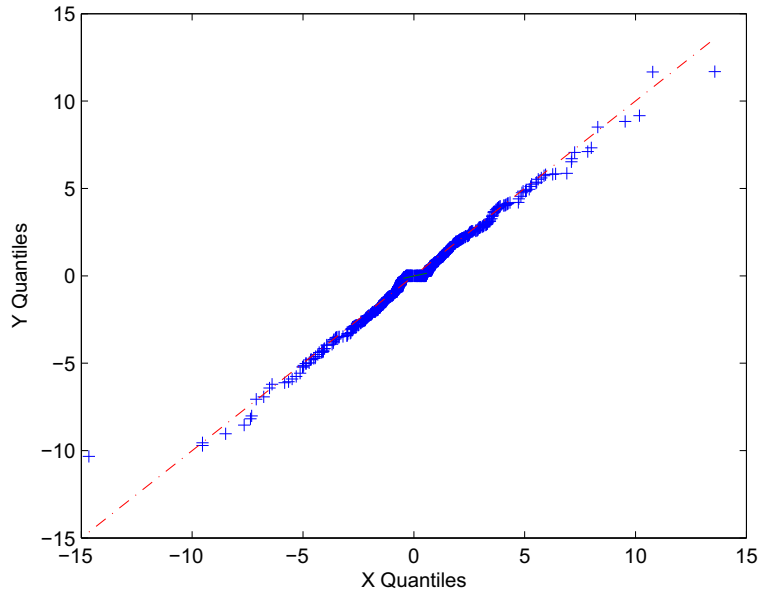


Figure 13: QQ-plot of a simulated time series versus the return series for the SO<sub>2</sub> price index.

dynamic interaction between the components, though for the SO<sub>2</sub> data set, the diagonal model was statistically superior to the full model.

The adequacy of the conditional model also has implications for the *unconditional* estimation of the tail index. In particular,  $\hat{\alpha}_{\text{Hint}}$ , the Hill-intercept tail index point estimator (4) and its confidence interval formed from the standard error (5) were optimally designed for i.i.d. symmetric stable Paretian data, while the data under study are neither i.i.d. nor stable Paretian. As such, it is of value to know the small sample performance of estimator (4) when a better approximation than i.i.d. stable Paretian is used for the true DGP, for which we use the fitted MixN(3,2) model. Figure 14 shows the kernel density estimate of 10,000 values of  $\hat{\alpha}_{\text{Hint}}$  resulting from simulated MN(3,2) series of length 1,780 (the number of SO<sub>2</sub> returns) and using the estimated parameter values given in Table 4. The vertical line is at 1.46, the value of  $\hat{\alpha}_{\text{Hint}}$  for the actual SO<sub>2</sub> returns. The mean, median and mode of the simulated data are 1.48, 1.49 and 1.50, respectively, and the 2.5% and 97.5% quantiles of the simulated  $\hat{\alpha}_{\text{Hint}}$  values are 1.34 and 1.60. These can be compared to the 95% confidence interval from the true data, formed from (5), of (1.38, 1.54). The fact that the point and interval estimators formed from the actual data and simulated data are close is encouraging, and also supports the use of the MN(3,2) model as a valid approximation for the true DGP.

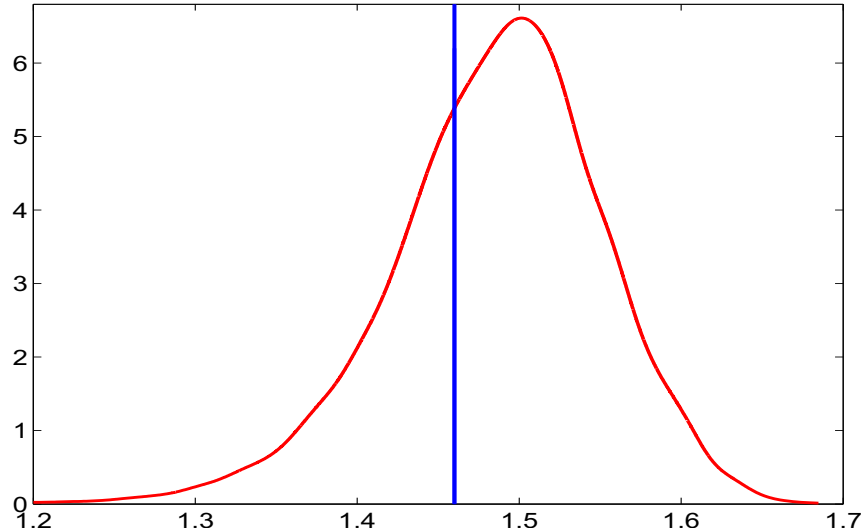


Figure 14: Kernel density estimate of  $\hat{\alpha}_{\text{Hint}}$  based on 10,000 simulated MN(3,2) processes using the MLE and sample size corresponding to the SO<sub>2</sub> returns. The actual  $\hat{\alpha}_{\text{Hint}}$  value is 1.46, shown as the vertical line.

## 5.8 Stable Mixture GARCH Models

The mixed normal GARCH model given by (7) and (9) is clearly capable of capturing the fat-tailed nature of the SO<sub>2</sub> returns data, although it was shown in Kuester et al. (2005) that, for other financial time series, a further improvement in model fit and out-of-sample forecasting ability can be achieved by replacing the normal distribution by a symmetric but fatter-tailed one which nests the normal, such as the generalized exponential (GED). This is admittedly an ad hoc type of improvement which no longer provides a model whose distributional mixture components can be motivated by a central limit theorem. To get around this, Haas et al. (2005) study such a mixture model with symmetric stable Paretian components, so that the components obey the GCLT, and demonstrate that it has a superior out-of-sample VaR forecasting ability and other beneficial properties when compared to both the normal and GED mixture GARCH models.

Analogous to the model given in Equations (7) and (9), time series  $\{\epsilon_t\}$  follows an  $n$ -component mixed stable GARCH( $r, s$ ) process, denoted MixStab-GARCH, if the distribution of  $\epsilon_t \mid \mathcal{F}_{t-1}$  is a weighted mixture of  $n$  symmetric stable distributions, i.e., its density at some real value  $x$  is given by

$$f_{\epsilon_t \mid \mathcal{F}_{t-1}}(x; \boldsymbol{\alpha}, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}_t^{(\delta)}) = \sum_{j=1}^n \omega_j f_S(x; \alpha_j, \mu_j, \sigma_{jt}^\delta), \quad (12)$$

where  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)'$  is the set of tail indices corresponding to the  $n$  symmetric stable distributional components such that  $1 < \alpha_i \leq 2$ ,  $i = 1, \dots, n$ , and, as before,  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)'$  is the set of nonnegative weights which sum to one,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$  is the set of component means,

and now  $\boldsymbol{\sigma}_t^{(\delta)} = (\sigma_{1t}^\delta, \dots, \sigma_{nt}^\delta)'$  is the set of strictly positive scale parameters, and  $f_S(x; \alpha, \mu, c)$  is the location- $\mu$ , scale- $c$ , symmetric stable Paretian density function with tail index  $\alpha$ . As with the MixN model, to ensure zero mean,  $\mu_n = -\sum_{j=1}^{n-1} (\omega_j/\omega_n) \mu_j$ . The component scale terms, analogous to the variance term in the MixN model, evolve according to

$$\boldsymbol{\sigma}_t^{(\delta)} = \gamma_0 + \sum_{i=1}^r \gamma_i |\epsilon_{t-i}|^\delta + \sum_{j=1}^s \boldsymbol{\Psi}_j \boldsymbol{\sigma}_{t-j}^{(\delta)}. \quad (13)$$

Similar to the MixN model, we let  $\text{MS}_\alpha(n, g)$  denote the  $n$ -component stable-mixture-GARCH(1, 1) process with diagonal  $\boldsymbol{\Psi}_1$  matrix, but also with the above constraints on the  $\alpha_i$ ,  $\delta$  restricted to one (estimating it offers little improvement in fit), and such that only  $g$  of the  $n$  components have a GARCH structure.

Maximum likelihood estimation of the model is straightforward given the ability to compute the symmetric stable Paretian density, as discussed in Section 5.2. For the SO<sub>2</sub> data set, we find that virtually no improvement in likelihood is obtained by using the  $\text{MS}_\alpha(n, g)$  model. This is not surprising, given the QQ-plot shown in Figure 13, which shows that the tails are adequately captured by the MN(3,2) model. However, we will see in the next section that, for the CO<sub>2</sub> returns data, a very large improvement of in-sample fit is realized by the  $\text{MS}_\alpha(n, g)$  model.

## 6 Analysis of CO<sub>2</sub> Returns

For the CO<sub>2</sub> price series, we have only 337 consecutive returns, though the larger presence of covered companies under the EU ETS translates into a higher daily traded emission volume and therefore into a much lower presence of zeros in the return time series (only 3 out of the 337 returns were zero). Without the zeros-problem, the stable-GARCH model (or any conventional GARCH model such as  $t$ -GARCH) can be entertained. We will see that the stable-GARCH model is indeed a tenable candidate in this case, along with the mixed-normal-GARCH, although when combining aspects of both (via the mixed-stable-GARCH), the goodness-of-fit is vastly better than either special case.

For the unconditional tail analysis,  $\hat{\alpha}_{\text{Hint}}$  from (4) yields the extremely low value of 1.25 (0.091). This was partly to be expected because of the large downward price jump observed on the day several countries disclosed the lower-than-expected CO<sub>2</sub> emissions in 2005. If we omit this one return from the time series,  $\hat{\alpha}_{\text{Hint}}$  increases only to 1.304 (0.092), indicating that the nature of the data, and not just the single outlier, are indeed extremely fat tailed.

With respect to the conditional GARCH models, Table 5 shows the in-sample fit results for symmetric and asymmetric stable-GARCH and several diagonal MN( $n, g$ ) models, where, as usual in practice, all GARCH models take  $r = s = 1$  and use an intercept and AR(1) term for the mean. First consider the stable models. The asymmetric stable-GARCH model is, according

to all criteria, superior to the symmetric one, with innovations parameters  $\hat{\alpha} = 1.593(0.079)$  and  $\hat{\beta} = -0.273(0.152)$ . One might conjecture that this is not surprising given the massive negative return in April, though when fitting a fat-tailed distribution such as the stable, such “outliers” do not have the impact they would when using a normal distribution. In fact, estimating the model without the value in question yields virtually identical parameter estimates, confirming that the value is not a “high leverage point”, nor improbable when using the estimated stable GARCH model.

For the mixed normal models, the lowest AIC is achieved by the MN(2,2), with a value of 1596.8, which is extremely close to the AIC value of 1597.1 from the MN(3,2), indicating that, even with relatively few observations, the 3-component model might still be preferred to the 2-component. However, the more conservative BIC measure clearly favors the latter; and moreover, the BIC strongly prefers the  $S_{\alpha,\beta}$ -GARCH model, with a value of 1627.1, to any of the MN( $n, g$ ) models, the closest being the MN(2,2), with a BIC value of 1635.0. The superiority of the stable-GARCH model, and the need for the asymmetric distributional assumption, indicates that a very heavy-tailed model which can incorporate skewness is required. As such, one might expect the stable-mixture GARCH model to perform well, and indeed, it completely dominates, by far, all other models used. The  $MS_{\alpha}(3,2)$  achieves the best AIC and BIC measures (with values 1530.6 and 1591.7, respectively), with the latter being extremely impressive, given the conservativeness of the BIC, and the fact that the model has 16 parameters (for 337 data points). For the  $MS_{\alpha}(3,2)$ , the values of  $\hat{\alpha}_i$ ,  $i = 1, 2, 3$ , with standard errors in parentheses, are 1.8021 (0.078), 1.7830 (0.046), and 1.8879 (0.203). The first two are highly different than the Gaussian value of two, while the last one is not.

Model	K	L	AIC	BIC
$S_{\alpha,0}$ -GARCH	6	-799.25	1610.49	1633.42
$S_{\alpha,\beta}$ -GARCH	7	-793.18	1600.36	1627.10
MN(1,1)	5	-983.26	1976.53	1995.60
MN(2,1)	8	-805.30	1626.60	1657.16
MN(2,2)	10	-788.40	1596.80	1635.00
MN(3,1)	11	-800.60	1623.20	1665.22
MN(3,2)	13	-785.57	1597.14	1646.80
MN(3,3)	15	-784.30	1598.60	1655.90
$MS_{\alpha}(2,2)$	12	-785.83	1595.75	1647.36
$MS_{\alpha}(3,2)$	16	-749.29	<b>1530.60</b>	<b>1591.72</b>
$MS_{\alpha}(3,3)$	18	<b>-748.17</b>	1532.38	1601.11

Table 5: Likelihood-based goodness-of-fit for CO<sub>2</sub>.  $S_{\alpha,\beta}$ -GARCH refers to the AR(1)-stable-GARCH(1,1) model (6) with  $Z \sim S_{\alpha,\beta}(0, 1)$ .  $MS_{\alpha}(n, g)$  is similar to the mixed-normal GARCH model, but refers to the  $n$ -component AR(1)-mixed-stable-GARCH(1,1) model (12) and (13). See the footnote in Table 3 for additional comments.

Because the CO<sub>2</sub> data set is not plagued with zeros, we could use quite a variety of conditional

models, such as the aforementioned FHS and GARCH-EVT methods, which, based on the results of Kuester et al. (2005), tend to fair extremely well in terms of out-of-sample VaR forecasting. The reason we forgo this is that these models do not give likelihood-based in-sample measures of goodness-of-fit. Moreover, and more importantly, because the CO<sub>2</sub> price series is so short, an out-of-sample forecasting exercise to assess and compare either the quality of density forecasts, as in Mittnik and Paoletta (2000) and Bao et al. (2003), or VaR forecasts, as in Bao et al. (2004) and Kuester et al. (2005), would be virtually meaningless.

## 7 Conclusions

Title IV of the Clean Air Act Amendment (CAAA) in the U.S. and the EU ETS created *de facto* property rights for emissions. The introduction of such new commodities is modifying operating costs in the power generation sector.

We first showed that current approaches for CO<sub>2</sub> price scenario delineation do not suffice. In particular, (i) the fundamentals-analysis based on few market components overlooks the complexity of the variables that come into play, and (ii) the spot-forward parity approach is, for the moment, inadequate due to the inconsistent behavior of the CO<sub>2</sub> emission allowance convenience yield (which depends on the political uncertainties that largely affect long futures maturities). We rely instead on an empirical approach, which directly addresses the characteristics of the data. Because of the relatively high presence of zeros in the SO<sub>2</sub> return series, we emphasize tail estimation procedures for an unconditional analysis, and, for a conditional analysis, the mixed-normal and mixed-stable GARCH models, which are suitable for dealing with the zeros-problem, and result in an outstanding goodness-of-fit. All methods are suitable for VaR and expected shortfall forecasting.

In addition to forecasting risk measures such as VaR and expected shortfall, knowledge of the unconditional and conditional distribution of emission trading allowance prices is essential for constructing optimal hedging and purchasing strategies in the carbon market, such as the design of new derivatives products. Today, such a derivatives market is fairly nonexistent but it has an enormous growth potential in the future, leading to an increase in the amount of trading of emission rights and also market liquidity (due to a larger participation of speculators and more intensive activity of hedgers). This will be the subject of future research.

## References

Alexander, C. and Lazar, E. (2005). Asymmetries and volatility regimes in the european equity market. ICMA Centre Discussion Papers in Finance 2005–14, The Business School for Financial Markets at the University of Reading.

- Alexander, C. and Lazar, E. (2006). Normal mixture garch(1,1): Applications to exchange rate modelling. *Journal of Applied Econometrics*, 21:307–336.
- Aparicio, F. M. and Estrada, J. (2001). Empirical distributions of stock returns: European securities markets, 1990–95. *European Journal of Finance*, 7:1–21.
- Bailey, E. (1998). Intertemporal pricing of sulfur dioxide allowances. Working paper, MIT Center for Energy and Environmental Policy Research.
- Bao, Y., Lee, T.-H., and Saltoglu, B. (2003). A test for density forecast comparison with applications to risk management. Technical report, University of California, Riverside.
- Bao, Y., Lee, T.-H., and Saltoglu, B. (2004). Evaluating predictive performance of value-at-risk models in emerging markets: A reality check. Technical report, University of California, Riverside.
- Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (1999). Var without correlations for portfolios of derivative securities. *The Journal of Futures Markets*, 19(5):583–602.
- Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (2002). Backtesting derivative portfolios with filtered historical simulation (fhs). *European Financial Management*, 8:31–58.
- Bauwens, L., Hafner, C. M., and Rombouts, J. V. K. (2006a). Multivariate mixed normal conditional heteroskedasticity. CORE Discussion Paper 2006/12, Université Catholique de Louvain.
- Bauwens, L., Preminger, A., and Rombouts, J. V. K. (2006b). Regime switching garch models. CORE Discussion Paper 2006/11, Université Catholique de Louvain.
- Bauwens, L. and Rombouts, J. V. K. (2005). Bayesian inference for the mixed conditional heteroskedasticity model. CORE Discussion Paper 2005/85, Université Catholique de Louvain.
- Benz, E. and Trück, S. (2006). Modeling the price dynamics of CO<sub>2</sub> emission allowances. Working paper, University of Bonn.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.
- Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69:542–547.
- Brännäs, K. and Nordman, N. (2003). Conditional skewness modelling for stock returns. *Applied Economics Letters*, 10:725–728.
- Burnham, K. P. and Anderson, D. (2003). *Model Selection and Multi-Model Inference*. Springer-Verlag, New York, second edition.

- Cason, N. and Gangadharan, L. (2005). Emissions variability in tradable permit markets with imperfect enforcement and banking. Working paper, Purdue University, Department of Economics.
- Christoffersen, P. F. (2003). *Elements of Financial Risk Management*. Academic Press, Amsterdam.
- Danielsson, J., Jansen, D. W., and deVries, C. G. (1996). The method of moments ratio estimator for the tail shape parameter. *Communications in Statistics – Theory and Methods*, 25(4):711–720.
- Daskalakis, G., Psychoyios, D., and Markellos, R. N. (2006). Modeling CO<sub>2</sub> emission allowance prices and derivatives: Evidence from the EEX. Athens University of Economics and Business.
- Dekkers, A. L. M. and de Haan, L. (1993). Optimal choice of sample fraction in extreme-value estimation. *Journal of Multivariate Analysis*, 47:173–195.
- Dekkers, A. L. M., Einmahl, J. H. J., and de Haan, L. (1989). A moment estimator for the index of an extreme value distribution. *The Annals of Statistics*, 17(4):1833–1855.
- Diebold, F. X., Schuermann, T., and Strouhair, J. D. (1998). Pitfalls and opportunities in the use of extreme value theory in risk management. Working paper no. 98-10, The Wharton School, University of Pennsylvania.
- Dowd, K. (2005). *Measuring Market Risk*. Wiley, New York, 2nd edition.
- DuMouchel, W. H. (1973). On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *Annals of Statistics*, 1:948–957.
- DuMouchel, W. H. (1983). Estimating the stable index  $\alpha$  in order to measure tail thickness: A review. *The Annals of Statistics*, 11(4):1019–1031.
- Ellerman, A. (2003). Lessons from phase 2 compliance with the US acid rain program. Technical document, Center for Energy and Environmental Policy Research CEEPR - MIT.
- Fama, E. (1965). The behavior of stock market prices. *Journal of Business*, 38:34–105.
- Goldie, C. M. and Smith, R. L. (1987). Slow variation with remainder: Theory and applications. *Quarterly Journal of Mathematics*, Oxford, 2nd Ser. 38:45–71.
- Gut, A. (2005). *Probability: A Graduate Course*. Springer-Verlag, New York.
- Haas, M., Mittnik, S., Paoletta, M. S., and Steude, S.-C. (2005). Stable mixture GARCH models. NCCR-Finrisk Working Paper Series 257, University of Zurich.

- Haas, M., Mittnik, S., Paoletta, M. S., and Steude, S.-C. (2006a). Analyzing and exploiting asymmetries in the news impact curve. NCCR-Finrisk Working Paper Series 256, University of Zurich.
- Haas, M., Mittnik, S., and Paoletta, M. S. (2004a). Mixed normal conditional heteroskedasticity. *Journal of Financial Econometrics*, 2(2):211–250.
- Haas, M., Mittnik, S., and Paoletta, M. S. (2004b). A new approach to markov switching garch models. *Journal of Financial Econometrics*, 2(4):493–530.
- Haas, M., Mittnik, S., and Paoletta, M. S. (2006b). Multivariate normal mixture garch. Working Paper Series 2006/09, Center for Financial Studies.
- Hall, P. and Welsh, A. H. (1984). Best attainable rates of convergence for estimates of regular variation. *Annals of Statistics*, 12:1079–1083.
- Hall, P. (1982). On simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society, Series B*, 44:37–42.
- Hamilton, J. D. (1991). A quasi-bayesian approach to estimating parameters for mixtures of normal distributions. *Journal of Business and Economic Statistics*, 9(1):21–39.
- Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35:705–730.
- Harvey, C. R. and Siddique, A. (1999). Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis*, 34(4):465–487.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5):1163–1174.
- Hols, M. C. A. B. and de Vries, C. G. (1991). The limiting distribution of extremal exchange rate returns. *Journal of Applied Econometrics*, 6:287–302.
- Joskow, P. and Schmalensee, R. (1998). The political economy of marked-based environmental policy: The u.s. acid rain program. *Journal of Law and Economics*, 41.
- Koedijk, K. G., Schafgans, M. M., and de Vries, C. G. (1990). The tail index of exchange rate returns. *Journal of International Economics*, 29:93–108.
- Koedijk, K. G., Stork, P. A., and de Vries, C. G. (1992). Differences between foreign exchange rate regimes: The view from the tails. *Journal of International Money and Finance*, 11:462–473.
- Kon, S. J. (1984). Models of stock returns: A comparison. *The Journal of Finance*, 39:147–165.

- Kratz, M. and Resnick, S. I. (1996). The qq-estimator and heavy tails. *Communications in Statistics – Stochastic Models*, 12(4):699–724.
- Kuester, K., Mittik, S., and Paoletta, M. S. (2005). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4(1):53–89.
- Lau, A. H., Lau, H., and Wingender, J. R. (1990). The distribution of stock returns: New evidence against the stable model. *Journal of Business and Economic Statistics*, 8(2):217–233.
- Lux, T. (1995). The stable paretian hypothesis and the frequency of large returns: An examination of major german stocks. *Volkswirtschaftliche Diskussionsbeiträge 72*, University of Bamberg, Germany.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 36:394–419.
- Mason, D. M. (1982). Laws of large numbers for sums of extreme values. *The Annals of Probability*, 10:754–764.
- McCulloch, J. H. (1985). Interest-risk sensitive deposit insurance premia: Stable ach estimates. *Journal of Banking and Finance*, 9:137–156.
- McCulloch, J. H. (1997a). Financial applications of stable distributions. In Maddala, G. and Rao, C., editors, *Handbook of Statistics*, volume 14. Elsevier Science, Amsterdam.
- McCulloch, J. H. (1997b). Measuring tail thickness in order to estimate the stable index  $\alpha$ : A critique. *Journal of Business and Economic Statistics*, 15(1):74–81.
- McDonald, J. B. (1991). Parametric models for partially adaptive estimation with skewed and leptokurtic residuals. *Economics Letters*, 37:273–278.
- McDonald, J. B. (1997). Probability distributions for financial models. In Maddala, G. S. and Rao, C. R., editors, *Handbook of Statistics*, volume 14. Elsevier Science.
- McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press, Princeton.
- McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, 7(3/4):271–300.
- Mittnik, S., Paoletta, M. S., and Rachev, S. T. (1998a). A tail estimator for the index of the stable paretian distribution. *Communications in Statistics–Theory and Methods*, 27(5):1239–1262.

- Mittnik, S., Paolella, M. S., and Rachev, S. T. (1998b). Unconditional and conditional distributional models for the nikkei index. *Asia-Pacific Financial Markets*, 5(2):99–128.
- Mittnik, S., Paolella, M. S., and Rachev, S. T. (2000). Diagnosing and treating the fat tails in financial returns data. *Journal of Empirical Finance*, 7:389–416.
- Mittnik, S., Paolella, M. S., and Rachev, S. T. (2002). Stationarity of stable power-garch processes. *Journal of Econometrics*, 106:97–107.
- Mittnik, S. and Paolella, M. S. (1999). A simple estimator for the characteristic exponent of the stable paretian distribution. *Mathematical and Computer Modelling*, 29:161–176.
- Mittnik, S. and Paolella, M. S. (2000). Conditional density and value-at-risk prediction of asian currency exchange rates. *Journal of Forecasting*, 19:313–333.
- Mittnik, S. and Paolella, M. S. (2003). Prediction of financial downside risk with heavy tailed conditional distributions. In Rachev, S. T., editor, *Handbook of Heavy Tailed Distributions in Finance*. Elsevier Science, Amsterdam.
- Montero, J. and Ellerman, A. (1998). Explaining low sulfur dioxide allowance prices: the effect of expectation errors and irreversibility. Technical document, Center for Energy and Environmental Policy Research CEEPR - MIT.
- Novak, S. and Beirlant, J. (2006). The magnitude of a market crash can be predicted. *Journal of Banking and Finance*, 30.
- Paolella, M. S. (2001). Testing the stable paretian assumption. *Mathematical and Computer Modelling*, 34:1095–1112.
- Paolella, M. S. (2007). *Intermediate Probability: A Computational Approach*. John Wiley & Sons, Chichester.
- Pictet, O. V., Dacorogna, M. M., and Müller, U. A. (1996). Behavior of tail estimators under various distributions. Presented at the Satellite Meeting to the 4th World Congress of the Bernoulli Society on “Stable Processes and Other Heavy-tailed Models for Highly Volatile Phenomena”, August 23-25, Wroclaw, Poland.
- Rachev, S. T. and Mittnik, S. (2000). *Stable Paretian Models in Finance*. Wiley & Sons, Chichester.
- Resnick, S. and Stărică, C. (1997). Smoothing the hill estimator. *Journal of Applied Probability*, 29:271–293.

- Rockinger, M. and Jondeau, E. (2002). Entropy densities with an application to autoregressive conditional skewness and kurtosis. *Journal of Econometrics*, 106:119–142.
- Ruppert, D. (2004). *Statistics and Finance: An Introduction*. Springer–Verlag, New York.
- Tietenberg, T. (2006). *Emissions Trading: Principles and Practice*. Resources for the Future Press, Washington, DC, USA.
- Trück, S., Borak, S., Härdle, W., and Weron, R. (2006). Convenience yields for co2 emission allowance future contracts. School of Economics and Finance, Queensland University of Technology, Brisbane Australia.
- Tsay, R. S. (2002). *Analysis of Financial Time Series*. John Wiley & Sons, New York.
- Tucker, A. L. and Pond, L. (1988). The probability distribution of foreign exchange price changes: Tests of candidate processes. *Review of Economics and Statistics*, 70:638–647.

swiss:finance:institute

c/o University of Geneva  
40 bd du Pont d'Arve  
1211 Geneva 4  
Switzerland

T +41 22 379 84 71  
F +41 22 379 82 77  
RPS@sfi.ch  
[www.SwissFinanceInstitute.ch](http://www.SwissFinanceInstitute.ch)