Bjarne Sæther

Temporal variations in the tail indices of German natural gas prices

Masteroppgave i Finansiell Økonomi
Veileder: Anne Neumann
Medveileder: Sjur Westgaard
Desember 2022

Masteroppgave

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi

NTNU
Kunnskap for en bedre verden

Bjarne Sæther

Temporal variations in the tail indices of German natural gas prices

Masteroppgave i Finansiell Økonomi
Veileder: Anne Neumann
Medveileder: Sjur Westgaard
Desember 2022

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi

NTNU
Kunnskap for en bedre verden

Preface

This master thesis is part of my Master of Science degree in financial economics at the Norwegian University of Science and Technology, Department of economics.

I would like to thank my supervisor Anne Neumann for excellent and patient guidance going well beyond what can be expected from a supervisor.

I would also like to thank my co-supervisor Sjur Westgaard for interesting discussions and, in particular, his efforts to get this project off the ground.

In addition, big thanks Evan Kyritsis for spending his spare time discussing tail index estimation and energy commodities with me.

Finally I would like to thank my family. Both my spouse Bjørg and my father Bernt-Erik deserve many thanks for reading more about natural gas prices than they probably ever envisioned in connection with this project.

Trondheim, 29.11.2022

Bjarne Sæther

Abstract

Natural gas is an important energy input for heating and electricity production in Germany. In recent years natural gas prices have soared and volatility has increased. The combination of lower domestic natural gas production in Europe and competition for liquefied natural gas in the global market has made Europe more vulnerable to a shortfall in natural gas deliveries.

The tail index estimator is a measure for tail fatness in statistical distributions and gives an indication for the probability of extreme prices. We use the tail index estimator to study the development of the risk of extreme prices in the German natural gas market over time.

We find that between 2012 and 2021 the probability of extremely high prices in the German natural gas market increases significantly. Similarly we find that the probability of extremely low prices increase, while to a lower degree. We test these effects for similar temperature conditions in the German winter months and we test if the introduction of renewable energy supply into the German electricity market has an effect in the price distribution for natural gas prices in Germany. Our results suggest that for similar temperature conditions the probability for extreme prices increase over time. We also find that the level of renewable energy supply in the German electricity market adds to explaining more extreme prices in the German natural gas market.

The increased likelihood for extreme prices in the German natural gas prices implies that the natural gas system in Germany to an increasing degree struggles to match supply and demand for natural gas. Using daily prices we provide empirical evidence suggesting that the German natural gas system has become more fragile over time.

Sammendrag

Naturgass er en viktig innsatsfaktor for oppvarming og elektrisitetsproduksjon i Tyskland. I de senere årene har prisene på naturgass steget mye og volatiliteten har økt. Kombinasjonen av lavere hjemlig naturgassproduksjon i Europa og konkurranse om flytende naturgass i det globale markedet har gjort Europa sårbart for bortfall av naturgassleveranser.

Haleindeksestimatoren er et mål for hale-fethet i statistiske fordelinger og gir en indikasjon for sannsynligheten for ekstreme priser. Vi bruker haleindeksestimatoren til å studere utviklingen av risikoen for ekstreme priser i det tyske naturgassmarkedet over tid.

Vi viser at mellom 2012 og 2022 så har sannsynligheten for ekstremt høye priser i det tyske naturgassmarkedet økt betydelig. Tilsvarende viser vi at sannsynligheten for ekstremt lave priser øker, men i mindre grad. Vi tester disse effektene for tilsvarende temperaturbetingelser i de tyske vintermånedene og vi tester om introduksjonen av fornybar energiproduksjon i det tyske elektrisitetsmarkedet gir en effekt på prisfordelingen til naturgassprisene i Tyskland. Vi viser at for tilsvarende temperaturbetingelser så øker sannsynligheten for ekstreme priser over tid. Vi viser også at nivået av fornybar energi i det tyske elektrisitetsmarkedet bidrar til mer ekstreme priser i det tyske naturgassmarkedet.

Den økte sannsynligheten for ekstreme priser i det tyske naturgassmarkedet impliserer at naturgass-systemet i Tyskland i økende grad sliter med balansen mellom tilbud og etterspørsel av naturgass. Våre funn viser at det tyske naturgass-systemet har blitt mer sårbart over tid.

Table of contents

1. Introduction

After years of cheap, easily accessible energy in Europe, energy policy and security of supply have again arrived at the forefront of public discourse. In recent years natural gas prices and price volatility have soared, as reported by IEA (2020). Kyritsis et al. (2017) shows that there has been increasing price fluctuations in the German electricity market. Prices started dropping below zero in hours with low consumption and high intermittent renewable electricity supply (RES). In France electricity prices have soared in periods of low availability of nuclear electricity supply and high electricity consumption with the consequence of France introducing a price cap on electricity prices in the retail market in the summer of 2022[1]. In recent years the phenomenon of increased volatility has also been observed in European coal markets as reported by IEA, (2022).

Natural gas is an important energy input for heating and electricity production in Germany. The German natural gas market has changed structurally since 2008 when a process of reducing the number of market areas from then 19 market areas started. Heather (2021) writes about the recent merger of the two remaining German market areas and the results of previous mergers. A market area or a balancing zone is a non-physical hub for trading natural gas in a certain area of the gas market. Within the balancing zone the price for natural gas is the same wherever you buy or sell it. The balancing point then represents the price between all entry and exit points within the market area. Those balancing zones were then over time gradually merged until only two remained. In the Northern part of Germany the main hub was Gaspool (GPL) and in the Southern part of Germany Netconnect (NCG). In October 2021 these two natural gas hubs were merged into one named Trading Hub Europe (THE) as discussed by Heather (2021). One of the main purposes for merging the natural gas hubs since 2008 was to increase liquidity for market participants. This failed after the creation of NCG and GPL and it remains to be seen if THE succeeds in this mission. The German natural gas hub still lags far behind both the Dutch Title Transfer Facility (TTF) and the British National Balancing Point (NBP) with regards to liquidity. In

[1] https://www.bloomberg.com/news/articles/2022-08-27/france-pledges-to-shield-households-from-soaring-power-prices?leadSource=uverify%20wall

2020 the aggregated volume at the TTF was 46,690 TWh of natural gas while at the NBP the number was 10,060 TWh of natural gas. In comparison, the numbers for 2020 at GPL and NCG were 1,350 TWh and 1,965 TWh natural gas traded respectively, analyzed by Heather (2021).

The German natural gas market is the largest natural gas market, in terms of consumption, in Europe. Natural gas is used primarily for heating, chemical industries and for electricity production. Germany has also functioned as a major entry point for Russian gas into Northern Europe due to inflows in pipelines on land via Ukraine, Belarus and Poland. In addition, Nord Stream 1 has operated as a seaborne pipeline directly from Russia to Germany with a yearly capacity of 59 billion cubic meters (bcm) natural gas. The intention for Nord Stream 2 was to increase the seaborne capacity for natural gas flows directly from Russia to Germany by 55 bcm.

Natural gas is an important source of energy for heating and electricity production in Northern Europe, but the domestic extraction of natural gas has declined in the region. Comparing numbers from 2012 to 2020 from *Eurostat*[2] we see that the domestic production in Northern Europe has declined dramatically. In 2012 Germany produced 13.1 bcm of natural gas, while in 2020 the number was 5.8 bcm. Denmark produced 5.8 bcm in 2012 compared with 1.4 bcm in 2020. The Netherlands produced 82 bcm in 2012, while merely 24.1 bcm in 2020. The numbers for the EU in total are 134 bcm in 2012 and 57 bcm in 2020. In the United Kingdom, connected to Belgium through the Zeebrugge pipeline, the local production has been stable at 41 bcm in 2012 and at 40 bcm in 2020. To compensate for the regional supply shortfall, Northern Europe has increasingly become dependent on liquefied natural gas (LNG) supply. Of the three operational terminals in Northern Europe, one is located in Swinoujscie on the border of Germany and Poland (6.3 bcm annual capacity), in Rotterdam (12 bcm annual capacity) and in Zeebrugge (11.4 bcm annual capacity). As opposed to regional production, and the supply from Norway and Russia, LNG is traded in the global market with competition in supply, especially from Eastern Asia. Xiaoyi & Haichun (2018) concludes that there is strong evidence for spot LNG price convergence, indicating price competition in a global LNG-market.

[2] https://ec.europa.eu/eurostat/databrowser/view/NRG_CB_GAS__custom_3820478/default/table?lang=en

Over the last decade European countries have progressively moved away from fixed long term contracts of natural gas delivery linked to the prices of oil products such as fuel-oil and gas-oil in favor of purchasing natural gas in spot delivery and long term contracts of volume delivery linked to the spot price of gas. Here the volume is fixed, but the price is referenced to a natural gas spot price at a relevant balancing point of delivery. The long term spot linked supply contracts are, generally, of shorter time frames than the old oil linked contracts. For a more thorough introduction to the development of long-term contracts of natural gas we recommend Franza (2014). Typically, the long term contracts were attractive for both buyer and seller as they shared the risk between buyer and seller. The buyer hedged the volume risk and received security of supply. The seller hedged the price risk long term. The original idea of risk sharing in natural gas supply contracts has been abolished and market prices are to reflect these risks now. The motivations for exchanging long term contracts of delivery with spot delivery were twofold. The main reason was the expectation of lower prices for natural gas deliveries. The second reason was a change in policy to remove fossil fuels from the energy supply altogether and exchange it with renewable energy sources emitting less greenhouse gases. This strategy has been reinforced by the recent program proposed by the European Commission called *"Fit for 55"*[3].

The move towards spot delivery of natural gas makes both the consumer and the producer of natural gas more vulnerable to short term price changes. Since 2015 we have seen examples of both gluts and shortages of energy commodities such as natural gas, coal and electricity (IEA, 2020)).

By 2030 the EU target is to reduce emission of GHG by at least 55% compared to the emission of GHG in 1990. By 2050 the target is that the EU should be climate neutral[4]. This has great ramifications for the energy industry in a region where especially coal fired power plants have been an integral part of the electricity systems in member countries of the EU. This also helps explaining the declining capacity of regional natural gas production. Investors are unwilling to invest in new projects in a region where their investment is unwanted by the European Union.

[3] https://www.consilium.europa.eu/en/policies/green-deal/fit-for-55-the-eu-plan-for-a-green-transition/
[4] https://climate.ec.europa.eu/eu-action/climate-strategies-targets/2050-long-term-strategy_en

Petroleum projects take a long time to realize and a long time to generate return on capital. In the time schedule for a fossil free European Union this makes new projects financially unattractive for investors.

The relation of natural gas prices and wholesale electricity prices in Germany is well documented by Hirth (2018) and Frydenberg et al. (2014). The electricity production capacity from natural gas in Germany is 32 GW, accounting for roughly 15% of total installed capacity of electricity production (211 GW) in 2022. Installed capacity of onshore/offshore wind power production and solar power production are 66 GW and 64 GW respectively[5].

While natural gas fired power plants constitute a modest part of the total German electricity production capacity, its importance is higher than its production capacity reveals. Natural gas fired power plants are, as opposed to coal fired power plants and nuclear electricity production, flexible. These power plants are able to ramp up (and down) production within hours without damaging the power plants. This property makes the natural gas fired power plants attractive parts of the German electricity system where RES is increasing its market share.

RES has the attractive properties of producing electricity without emitting greenhouse gases and its marginal cost of production is close to zero. However, a drawback is that the production is intermittent. Therefore, the system as a whole needs flexible backup generation capacities to step into the electricity system when wind or sun is absent (until cost efficient storage solutions are added to the electricity system). In periods with high RES and low demand, electricity producers will pay consumers to take electricity there is no demand for. Biber et al. (2022) addresses the problem and resolution of negative electricity prices. In periods with low RES the share of natural gas fired electricity production will be higher than in periods with high RES. In the natural gas market this will in turn increase demand for natural gas.

This can lead to substantial differences in wholesale electricity prices from day to day and from hour to hour. The German wholesale electricity price is (at the time of writing) determined by a merit order. The merit order is a way of ranking available electricity sources with regard to the

[5] https://www.energy-charts.info/charts/installed_power/chart.htm?l=en&c=DE&stacking=single&chartColumnSorting=default

increasing order of cost. (A more detailed introduction to the merit order pricing mechanism and the effect of RES is Erdmann (2017)). For every hour day ahead the generation bids are aggregated into one supply curve. Then the demand curve is matched with the generation bid that has the lowest marginal cost that meets demand. The resulting market price in this hour is then the price all market participants pay (consumers) or get paid (generators). Figure 1.1 illustrates this concept with an example from one hour in the German electricity market where there is low RES and also the merit order with high RES available to the electricity system.

An illustration of the merit order in the German electricity market

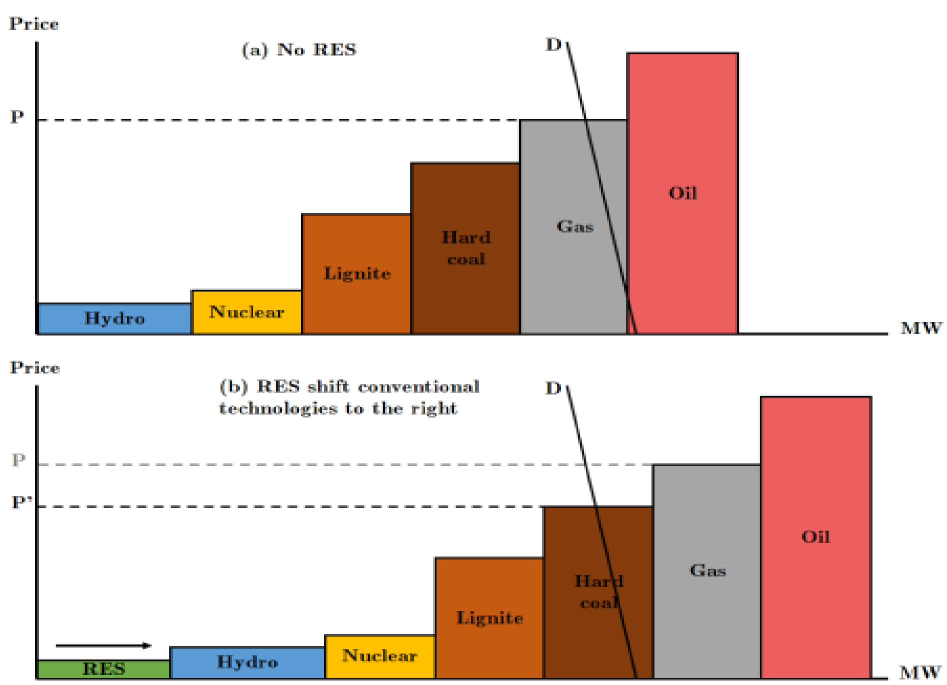

Figure 1.1: The graph represents an illustrative example for the relationship between the merit order pricing mechanism in the German electricity market, gas demand in Germany and RES. The demand curve is denoted by *"D"* while the supply curve is aggregated with different technologies and different prices. Without RES in the system the gas fired power plants will determine the German electricity price (top graph). With RES in the system, the hard coal fired power plants will determine the price, as the supply curve shifts right. Since no natural gas fired power plants run that day, then the demand for gas will be lower as well. Source: Liebensteiner&Wrienz (2020).

With varying RES over time demand for natural gas in Germany will vary in relation to available RES. The introduction of RES may in effect introduce a feedback mechanism into the natural gas market where not only the electricity price is affected by the price of natural gas, but where the natural gas price is affected by the availability of RES. This may be most prevalent in hours with a tight supply and demand balance in the natural gas market. The massive introduction of RES in Germany has given natural gas fired electricity production a more prominent place in the German electricity system. Intermittent energy production increases the demand for flexible electricity production in the electricity grid as analyzed by Biber et al. (2022).

The introduction of RES into the electricity system has made electricity prices more volatile and increased uncertainty for future price developments for market participants. For electricity producers the uncertainty of future prices increases the uncertainty about future production of electricity. For consumers the same applies as the risk of future cost increases (or reductions) are hard to predict.

On the financial level, changes in the price distributions of energy markets have consequences for the energy industry and in the related financial industry. With higher volatility and fatter tails in the price distributions of energy commodities, the cost of doing business increases as the demand for collateral capital surges. Both energy exchanges and clearing banks will become more reluctant to accepting business from companies with weak balance sheets. This is the beginning of a vicious circle. The threshold and cost of entering the energy markets become higher. Liquidity disappears and less liquidity increases the risk of participating in the market for the remaining market actors. This will in turn negatively impact the pricing mechanism.

The main question we aim to address in this paper is whether or not there has been a trend toward fatter tails in the price distributions of German natural gas prices since 2012. This is interesting and relevant because fatter tails in the German natural gas prices reveal structural changes in both the supply side and the demand side of the natural gas industry. Fatter right tails in natural gas prices are a signal of a tighter supply and demand balance in the physical natural gas market. If that is the case, then it has implications for fundamental actors in the natural gas market such as retail consumers, electricity producers, natural gas producers, industrial companies and natural gas retail companies. If the price distributions of natural gas change to fatter tails they will have to adapt in the long term. For retail consumers and chemical industry this means a risk of higher costs. For natural gas retail companies it means higher capital costs as counterparties and exchanges will require more collateral capital. For natural gas producers it means a chance for higher income.

Fatter left tails will indicate bigger probability for gluts in the natural gas market. While the domestic gas production has fallen in the EU since 2012, there might still be periods where the market is over supplied as domestic production have been replaced by Russian natural gas and LNG.

We use daily natural gas prices in Germany from 2012 to 2021 to analyze tail fatness. Our assumption is that the right tail of natural gas prices will be fatter in the second period than in the first period. This is because the domestic production cuts of natural gas in Europe during the time period we are testing makes the European natural gas supply more vulnerable to supply squeezes. Our assumption for the left tails is that they have also become fatter over the time period we are looking at. This is because we assume that while domestic natural gas production has declined in the EU, the pipelines from Russia and LNG should be able to cover the production gap. In periods with mild weather and high RES we expect the natural gas system to experience very low prices.

In order to test our hypothesis of increased tail fatness over time we split the data sample. If we find a difference in tail fatness between the two subsamples we aim to explore contributing factors for this. One candidate is the level of RES in the electricity system. We therefore control for that. As elaborated previously, we argue that a fatter tail (high demand for natural gas) could be explained by low temperature through heating demand. Similarly, low natural gas demand is also a product of high temperatures as heating demand drops. We compare the two time periods with respect to temperatures to test if the tail fatness is different, in general, between the two time periods for similar temperature conditions.

Temperature is an important determinant of natural gas demand since roughly 50% of German households are heated with natural gas[6]. Comparing the tails of German natural gas prices for

[6] https://www.statista.com/statistics/1189752/household-heating-sources-germany/

similar temperatures over time allow us to test if the tail fatness has increased, corrected for temperature.

We also want to control for the level of RES and investigate if this has indeed created a feedback mechanism to the natural gas market. We will do this by comparing tail fatness over time with similar RES-levels in the electricity system. We will test both for a fixed RES-level between the time periods and for a relative RES level between the time periods.

To quantify the tail fatness we will use a technique of estimating tail indices for the price distributions of German natural gas prices introduced by Huisman et al. (2001). In contrast to the widely used Hill-index, introduced by Hill (1975), this method produces estimates that are robust and unbiased for small samples.

The remainder of the thesis is structured as follows. In chapter 2 we provide the theoretical background for our calculations of the tail fatness. Then, in chapter 3, we present and analyze the data used in the thesis. In chapter 4 we present the results of our analysis. In chapter 5 we offer a discussion about the results presented in chapter 4. In chapter 6 we submit the main conclusions about the tail fatness in German natural gas prices.

2. Methodology

We aim to analyze the tail fatness of German natural gas prices. Hill (1975) introduced an estimator for calculating the tail index of a distribution. The proposed Hill tail index is a measure for tail fatness in data. Huisman et al. (2001) introduces a method which is more robust and produces less biased estimates in small samples.

In this section we will first introduce the theoretical foundation of either technique, before we show how the method suggested by Huisman et al. (2001) works in practice.

Further we will also explain the Welch's t-test, used to test the statistical significance of differences between the estimated tail indices.

2.1 Measuring tail fatness

Extreme value theory (EVT) is a branch of statistics concerned with the distribution of tail observations in probability distributions. Bryson (1974) defines fat-tailed probability distributions as distributions which, unlike the normal distribution, do not express exponential decay in the tails. Furthermore, asymptotically the tail distribution follows a Pareto distribution. This is an important assumption that has been shown to hold in financial time series of energy commodities in general and natural gas prices in particular by Benth et al. (2008) and Lv and Shan (2013).

The Pareto distribution was first introduced by Pareto (1897). In a practical example he showed that the number of people, N_x, with incomes higher than *x* could be modeled with a *power law*

(2.1) $$N_x = Ax^{-\alpha}.$$

A is a constant, and the distribution is defined by α which is the exponent of the power law. The probability distribution function that follows a power law is defined by the general distribution function

(2.2) $$p(x) = Cx^{-\alpha}.$$

Once α is fixed, the constant C is calculated such that the requirement that the distribution function sums to 1 is fulfilled.

In connection with the power law the tail index γ is defined as the inverse of α

(2.3) $$\gamma = \frac{1}{\alpha}.$$

The tail fatness increases with γ, so with higher γ (lower α), the slower the probability density function decays to zero.

2.1.1 The Hill index

The Hill tail index is a measure for calculating tail fatness introduced by Hill (1975). It is a non-parametric approach that does not make assumptions about the distribution of the global sample from which the tail data is drawn from. The Hill estimator is popular due to its simplicity and ease of use. It is an asymptotically unbiased estimator and performs well for very large samples as explained in Hill (1975). However, it exhibits bias in relatively small and more realistic sample sizes.

Suppose that a sample of *n* positive independent observations is drawn from some unknown distribution that is assumed to follow a power law. We then sort the sample in increasing order such that *x(i)* represents the *i*th-order statistic such that $x(i) \geq x(i-1)$ for $i = 2, \dots, n$ and we choose to include κ observations of the right tail to represent the right tail of our sample. (Hill, 1975) then proposes an estimator for the right tail index given by

(2.4) $$\gamma(\kappa)_{right} = \frac{1}{\kappa}\sum_{j=1}^{\kappa} \ln(x(n-j+1) - \ln(x(n-\kappa)).$$

This is the maximum likelihood estimator for a conditional Pareto distribution. The threshold which is chosen to be the tail of the sample is given by κ.

For the left tail the formula is similar, but with opposite signs

(2.5) $$\gamma(\kappa)_{left} = \frac{1}{\kappa}\sum_{j=1}^{\kappa} \ln(x(n-\kappa) - \ln(x(n-j+1)).$$

The estimator in itself is standard to calculate, but the choice of κ is crucial Beirlant et al. (1996) since choosing different values for κ may give widely different values for γ. Selecting too many values in κ increases bias since it includes data that are not part of the tail. Selecting too few data in κ, results in the estimator being dominated by its variance.

Dacorogna et al. (1995) deduces an asymptotic approximation to the bias of the Hill tail index for the class of distributions with the characteristic function

(2.6) $$F(x) = 1 - ax^{-\alpha}(1 + bx^{-\beta}) .$$

Here *α* and *β* are larger than zero, *a* and *b* are real numbers. Equation (2.6) provides the second order expansion of the cumulative distribution function (cdf.) for almost all fat tailed distributions. The cdf. for the Pareto distribution is given by

(2.7) $$F(x) = 1 - x^{-\alpha}.$$

This is equation (2.6) with *a = 1* and *b = 0.*

(Hall, 1990) shows that the expected value of the Hill-estimator for a given value of *κ* in the family of distributions defined in eq. (2.6) is given by

(2.8) $$E(\gamma(\kappa)) \approx \frac{1}{\alpha} - \frac{b\beta}{\alpha(\alpha+\beta)} a^{\frac{-\beta}{\alpha}} \left(\frac{\kappa}{n}\right)^{\frac{\beta}{\alpha}}.$$

Equation 2.8 shows that the bias increases in *κ* (since γ = $\frac{1}{\alpha}$ by definition, the unbiased estimator for γ would have expected value of $\frac{1}{\alpha}$). Selecting too many data from the center of the distribution leads to an increase in the bias.

(Hall, 1990) also shows that the asymptotic variance for the Hill estimator for the family of distribution defined in 2.6 to be

(2.9) $$var(\gamma(\kappa)) \approx \frac{1}{\kappa\alpha^2}.$$

As opposed to the bias inherent in the expected value of the tail index, the variance of the estimator decreases with *κ*. Equation 2.8 implies that there is always a bias for κ > 0. When applying the proposed method one is faced with a trade-off between bias and precision when calculating the tail index using the estimator for the Hill tail index.

There is extensive literature dedicated to making the optimal choice of κ with respect to bias and variance. Jansen&de Vries (1991) uses a method which demands to make an assumption of the underlying global probability distribution. Resnick&Starica (1997) suggests smoothing the Hill-estimates for different values of κ with a rolling average and plotting the averaged estimators against κ for subjective assessment of the true value of the estimator. Bhattacharya&Kallitsis (2019) proposes an adaptive, data driven method for trimming the parameters, following the tradition of choosing an optimal asymptotically unbiased estimator for κ.

2.1.2 Tail index estimates in small samples

Huisman et al. (2001) proposes an alternative method for estimating tail indices in small samples. They observe that for values $k < \kappa$ when calculating the Hill tail index, the values for $\gamma(k)$ increase in an approximately linear fashion with k. This means that for $k = 2,\dots,\kappa-1$ then $\gamma(\kappa-1) \geq \gamma(\kappa-2) \geq \dots \gamma(2)$.

Imposing the condition that $\alpha=\beta$ in eq. (2.8) makes the asymptotic bias linear in κ[7]. Using this condition allows to transform equation (2.8) to

$$(2.10) \qquad \gamma(k) = \beta_0 + \beta_1 k + \varepsilon(k), k = 1, \dots, \kappa.$$

Huisman et al. exploit the linearity in κ of the bias function from imposing the condition that $\alpha=\beta$. Rather than selecting an optimal value for κ they propose to calculate all values for γ(k) where $k = 1,\dots,\kappa$ and then using the estimates for the tail indices $\gamma(k)$ for $k = 1,\dots,\kappa$ to estimate the parameters in equation (2.10). For small values of κ then the intercept, β_0, will provide an unbiased estimate of the tail index. This solves the bias-variance trade-off as long as κ is chosen such that the function $\gamma(k)$ for $k = 1,\dots,\kappa$ is approximately linear, according to Huisman et al. (2001).

[7] For further discussion around the appropriateness of imposing $\alpha=\beta$ we refer to Hall, P. (1990). Using the Bootstrap to Estimate Mean Square Error and Select Smoothing Parameters in Non-parametric Problems. *Journal of Multivariate Analysis*, *32*, 177-203. and Dacorogna, M., Müller, U., Pictet, O., de Vries, C. (1995). The Distribution of Extremal Foreign Exchange Rate Returns in Extremely Large Datasets. In T. Institute (Ed.), *Discussion paper*..

It is possible to use ordinary least squares (OLS) to calculate the parameters. However, equation (2.9) indicates that the variance differs for different values of *k.* This means that the error term in eq. (2.10), $\varepsilon(k)$, is heteroskedastic. The variance will be different for different values of *k.* To solve this problem we use weighted least squares (WLS) to calculate the parameters in eq. (2.10).

However, the variables used in the WLS calculation are correlated, since they use overlapping data to calculate *γ(k)* for the different values of *k.* This excludes using the standard formulas for OLS and WLS to calculate the standard errors for our estimate of β_0.

To calculate the WLS estimates for we use a weighting matrix of dimension (κ x κ) with weights $\{\sqrt{1}, \sqrt{2}, \dots, \sqrt{\kappa}\}$ on the diagonal and 0 elsewhere. The *Z*-matrix is a (κ x 2)-matrix with 1's in the first column and numbers 1,…,κ in the second column. The OLS-estimator is then given by

(2.11) $$\gamma^* = Z\beta + \varepsilon,$$

in matrix notation. Transforming the OLS-estimator in eq. (2.11) gives the WLS-estimator for β by

(2.12) $$b_{wls} = (Z^T W^T W Z)^{-1} Z^T W^T W \gamma^*.$$

The intercept and unbiased estimator for the tail index, β_0, is then the first element of the (2 x 1)-matrix b_{wls}. Huisman et al. (2001) gives an alternative way to calculate the tail index. To calculate the weighted average of the Hill-estimators from k = 1,…,κ one can use the formula

(2.13) $$\gamma^m(k) = \sum_{k=1}^{\kappa} w(k)\gamma(k).$$

where *w(k)* are the diagonal elements of the weight matrix *W* and *γ(k)* are the respective Hill-tail indices calculated for *k = 1,…,κ.*

For calculating the standard errors and adjust for the aforementioned heteroscedasticity we must adjust the regular calculations of standard errors from WLS.

Let *y(i) = ln(x(i)).* Then the Hill estimator is a linear combination of *y(i)* (eq. (2.4) and eq. (2.5)) and the tail index estimation can be expressed as

(2.14) $$\gamma^* = Ay$$

in matrix notation. Here γ^* is the estimated Hill indices consisting of $\gamma(k), k = 1, \dots, \kappa$, while A is the transformation matrix. Then $\Omega = \mathrm{A}\Sigma\mathrm{A}^{\mathrm{T}}$ is the covariance matrix for the Hill estimates in γ^*. For the WLS based estimator in equation (2.13) we can calculate the standard errors of the estimates using

$$cov(\beta)_{wls} = (Z^T W^T W Z)^{-1} Z^T W^T W \Omega W^T W Z (Z^T W^T W Z)^{-1}. \tag{2.15}$$

Here W is the aforementioned weight matrix with $\sqrt{1}, \sqrt{2}, \dots, \sqrt{\kappa}$ on the diagonal and 0 elsewhere, while Z is a *(κ x 2)*-matrix with 1's in the first column and the vector $\{1,2,3,\dots,\kappa\}^{\mathrm{T}}$ in the second column. Further we use the fact that increasing order statistics *z(i), (i =1,2,…,κ+1)* from a sample of size *n* are multivariate normally distributed, as proven by Cox&Hinkley (1974) to calculate Σ.

The A-matrix can be derived from the Hill-estimator calculated by WLS as

$$A = \begin{bmatrix} 0 & 0 \cdots 0 & -1 & & 1 \\ 0 & \dots -1 & \frac{1}{2} & & \frac{1}{2} \\ 0 \;\cdots & -1 \;\; \frac{1}{3} & \frac{1}{3} & & \frac{1}{3} \\ \vdots & \ddots & & & \vdots \\ -1 & \frac{1}{\kappa} \cdots \frac{1}{\kappa} & \frac{1}{\kappa} & \frac{1}{\kappa} & \frac{1}{\kappa} \end{bmatrix}. \tag{2.16}$$

To obtain Σ we use the multivariate normal distribution of increasing order statistics with mean *μ(i)* and covariances between order statistics *y(i)* and *y(j)* given by *v(i,j)* where

$$\mu(i) = F_y^{-1}(p(i)) \tag{2.17}$$

and

$$v(i,j) = \frac{p(i)(i - p(j))}{n f_y(\mu(i)) f_y(\mu_j))} \qquad for\ i \le j. \tag{2.18}$$

Further we approximate $p(i) \approx \frac{1}{n}$ and use the Pareto distribution, that we assume is the distribution of *x,* such that

$$F_x(x) = 1 - x^{-\alpha}. \tag{2.19}$$

Since the Hill estimator is a linear combination of the natural logarithm of ordered data based on *x,* we can approximate the mean with

$$\mu(i) = \ln\left(\left(1 - p(i)\right)^{(1-\alpha)}\right). \tag{2.20}$$

Then the standard errors are fully defined. For further details regarding the calculation of standard errors we recommend the appendix of Huisman et al. (2001).

2.2 Welch's t-test

In order to compare the statistical significance of our tail index we will face a problem to compare our estimates from samples of different sample sizes and different variances. Welch's t-test was originally introduced by Welch (1947) to handle the problem of different sample sizes and different variances for testing whether two populations have different means. Welch's t-test assumes that the means are normally distributed, but that they may have different variances.

The *t*-statistic is defined by:

$$(2.21) \qquad t = \frac{\overline{X_1 - \overline{X_2}}}{\sqrt{s_{\overline{X_1}}{}^2 + s_{\overline{X_2}}{}^2}}.$$

Further $s_{\overline{X_i}}$ is defined by

$$(2.22) \qquad s_{\bar{X}_i} = \frac{s_i}{\sqrt{N_i}}.$$

The respective sample means are denoted by $\overline{X_i}$, while $s_{\overline{X_i}}$ is its standard error. N_i is the sample size of sample *i* and s_i is the corrected sample standard deviation. Finally, the degrees of freedom for the variance estimate in (2.17), ν, is approximated by the Welch-Satterthwaite equation, developed by Satterthwaite, (1941):

$$(2.23) \qquad \nu = \frac{\left(\frac{s_1^2}{N1} + \frac{s_2^2}{N2}\right)^2}{\frac{s_1^4}{N_1^2 \nu_1} + \frac{s_2^4}{N_2^2 \nu_2}}$$

Using this test rather than a standard t-test will allow us to compare our estimates for the tail indices.

3. Data

In this chapter we present the data we used and the results on the properties with respect to statistical distribution.

3.1 Data sources

3.1.1 German natural gas prices

The dataset consists of daily day-ahead natural gas prices from the two German natural gas hubs Gaspool (GPL) and Netconnect Germany (NCG) from 2/1/2012 to 31/3/2021. The data in the period from 1/10/2021 to 31/12/2021[8] were collected from Trading Hub Europe (THE), which is, from 1/10/2021, the only gas hub in Germany and a result of merging GPL and NCG into one common hub. We include only weekdays from Tuesday to (and including) Friday due to data availability. We analyze natural gas prices for the winter seasons, from (and including) October to (and including) March only thus avoiding the seasonality effect with much lower heating demand in summer than in winter. We assume consistent natural gas storage operation by owners. The data was downloaded via Eikon Reuters and is available on request. We computed the average of the day-ahead contract prices in the two natural gas hubs and used the average price as a collective natural gas price for Germany as a whole.

3.1.2 German intermittent renewable energy supply

The data for electricity generation from solar and wind in Germany is extracted from the respective German transmission system operators' websites (Amprion, 50Hertz, ENBW and Tennet). The data were originally presented in 15 minutes intervals and needed to be aggregated to a daily average to match the data granularity of the day-ahead NCG/GPL/THE-data. We add solar and wind power production to approximate an aggregated daily estimate for renewable energy supply (RES). The data used is made available on request.

[8] We have excluded data from 2022 since the situation in the natural gas market moved beyond what could be expected from market mechanism effects after that.

3.1.3 German temperature data

The temperature data was collected from Point Carbon and is weighted with respect to electricity consumption in the cities included in the weighted temperature index. This is the industry standard. The temperature data are also daily aggregates, averaged from hourly measurements. From each selected measurement station actual temperature data are used with hourly (where available) or three-hourly granularity. The data is then aggregated to a temperature for Germany as a whole where the aggregated temperature is weighted with regards to power consumption for the respective city. The cities used in the aggregated, consumption weighted temperature are Düsseldorf, Berlin, Hamburg, Stuttgart, Frankfurt, München and Bremen.

The advantage of using a consumption weighted temperature for electricity is that although only around 5% of the population in Germany heats with electricity, electricity demand is still responsive to temperature. Germany shares interconnectors with several countries that rely on electric heating and on days with cold temperatures electricity exports from Germany will increase.

By using data with weighted temperature for electricity demand- we implicitly impose the same for natural gas demand. Ideally we would be able to use temperature data weighted with gas demand, but we have been unable to locate such data. The data used is made available on request.

The data provider of the consumption weighed temperature data only reveals which German cities were used to calculate their temperature index, but not how these cities were weighted in the index. Therefore we construct a temperature index by collecting the temperature data for the cities used in the temperature index that the data provider company from provided from *The European climate and dataset*[9].

To calculate our temperature index we used ordinary least squares and the model in eq. (3.1).

[9] https://www.ecad.eu

(3.1) $y_T = \beta_0 + \beta_1 T_{\text{Düsseldorf}} + \beta_2 T_{Berlin} + \beta_3 T_{Hamburg} + \beta_4 T_{Stuttgart} + \beta_5 T_{Frankfurt} + \beta_6 T_{\text{München}} + \beta_7 T_{Bremen} + \varepsilon.$

In eq. (3.1) y_T represents the temperature index from Point Carbon which we are reverse-engineering and is daily averaged. *T* is the daily average temperature for the respective cities in Germany. Estimating the model yields the following parameter values and t-values, presented in table 3.1:

Estimated weights for the cities used in the temperature index

	$\boldsymbol{\beta_0}$	$\boldsymbol{\beta_1}$ Düsseldorf	$\boldsymbol{\beta_2}$ Berlin	$\boldsymbol{\beta_3}$ Hamburg	$\boldsymbol{\beta_4}$ Stuttgart	$\boldsymbol{\beta_5}$ Frankfurt	$\boldsymbol{\beta_6}$ München	$\boldsymbol{\beta_7}$ Bremen
Values	-0.02	0.209	0.149	0.100	0.198	0.188	0.078	0.067
t-values	-2.8	21.56	52.57	63.25	29.24	76.64	13.53	66.53

Table 3.1: The estimated parameter values of the OLS-estimation of the temperature index. The parameters, from β_1 to β_7 , are the different measuring point's weights into the aggregated temperature index for Germany. The t-values are displayed in the third row.

All parameters are statistically significant. The residual standard error is 0.1909 on 3646 degrees of freedom. Both the R-squared statistics and the adjusted R-squared statistics give the value 0.9993 and indicate a very good model fit. The advantage to this way of estimating the temperature index is that we know the weights of the cities used to calculate the index.

Figure 3.1 shows the estimated index in comparison to the original index from Point Carbon. From visual inspection and in combination with the estimated model diagnostics, we conclude that our estimated index compares well to the data provided by Point Carbon.

3.1.4 Descriptive Statistics

Temperature data, price data and RES data are split into two periods, from 2/1/2012-31/12/2016, and from 2/1/2017-31/12/2021. This is done because first, we aim to test whether or not the tail distributions of German natural gas prices have changed over time. Second, we are interested in the effect of temperature (as a proxy of consumption) and RES on the tail distributions. The dataset for the first period consists of 514 data points, while the dataset for the second period consists of 527 observations.

A comparison of temperature indices for Germany

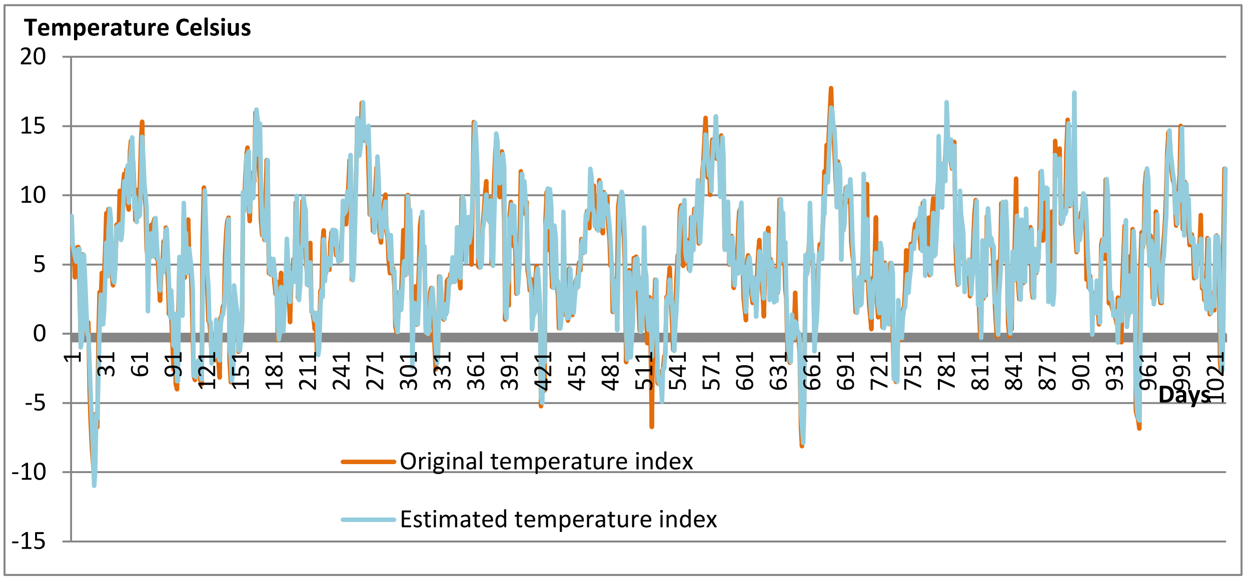

Figure 3.1: Visualization of our temperature data. The original temperature index is orange. Our estimated temperature index is colored light blue.

Next we split the two data sets with regards to temperature, defining the temperatures below 5.42 degrees Celsius (consumption weighted) as “low” and above or equal to 5.42 degrees Celsius[10] as “high”. Approximately 50% of German homes are heated with natural gas[11] and therefore we expect natural gas consumption to be sensitive to temperature.

Finally we have also split the data with respect to renewable energy production. This is done because we want to test the effect on daily gas prices from RES. We have aimed to classify the data between “high” and “low” renewable energy production using the 13000 MW as a fixed threshold. This was a heuristic choice to ensure fairly equal sample sizes of data between the datasets. As the capacity for renewable energy production increases dramatically during the time period 2012-2021 we have also defined “high” and “low” renewable energy supply with

[10] 5.42 degrees Celsius refers to the median temperature in the consumption weighted temperature data from 2012 to 2021.

[11] https://www.bmwi-energiewende.de/EWD/Redaktion/EN/Newsletter/2015/09/Meldung/infografik-heizsysteme.html

different thresholds in each period. We chose a threshold of 8900 MW in the first period and almost twice that, 17900 MW in the second period as the difference between "high" and "low" RES in any given day. 8900 MW was the median value of RES in time period 1. In time period 2 we chose a value slightly higher than the median (17600 MW) as the threshold since the mean value of the RES in the second period was 18900 MW. Figures 3.2 a), 3.2 b) and 3.2 c) illustrate a schematic overview of the organization of the data we used for our analysis.

Illustration of the partitioning of the dataset with respect to temperature

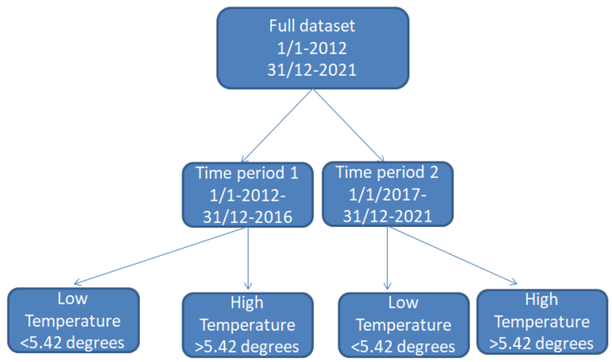

Figure 3.2 a): Illustration of the initial split of the full dataset with respect to temperature. All the data (temperature, RES and German natural gas prices are split into two time periods; time period 1 and time period 2. Then the data in both time periods are split with respect to temperature. This means that for dates where temperatures are below the threshold, all the RES data and German natural gas price data associated with these dates are put into this sub-dataset where the corresponding temperature is below the threshold. We do the same for data above the threshold, for both time periods.

Illustration of the partitioning of the dataset with respect to RES.

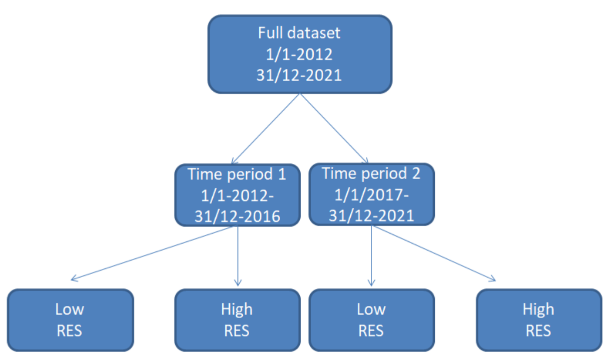

Figure 3.2 b): Illustration of the initial split of the full dataset with respect to RES. All the data (temperature, RES and German natural gas prices are split into two time periods; time period 1 and time period 2. Then the data in both time periods are split with respect to RES. This means that for dates where RES are below the threshold, all the temperature data and German natural gas price data associated with these dates are put into this sub-dataset where the corresponding RES is below the threshold. We do the same for data above the threshold, for both time periods.

Illustration of the partitioning of data with regards to both RES and temperature

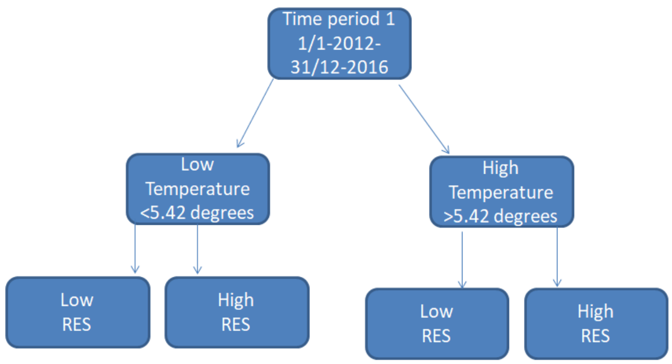

Figure 3.2 c): Illustration of the split of the dataset with respect to both temperature and RES illustrated with the first time period. The data in both time periods are split with respect to temperature. This means that for dates where temperature are below the threshold, all the temperature data and German natural gas price data associated with these dates are put into this sub-dataset where the corresponding temperature is below the threshold. We do the same for data above the threshold, for both time periods. Then we repeat the exercise for the data already separated with respect to temperature, but now with respect to RES.

The aggregate production capacity of solar and wind power in Germany has increased from 65,000 MW in 2012 to 130,000 MW by the end of 2021. This was the reasoning for introducing relative thresholds. We noticed we got too many sub-datasets with sparse data when we separated the data with a fixed threshold of 13,000 MW RES in both time periods. The fixed threshold is useful for a direct comparison between the time periods. However, adding the relative threshold provides information about how the natural gas system reacts to the structurally different electricity system in period 2.

3.2 Data analysis

German natural gas prices are generally stable, but frequently experience price jumps and price drops (cf. Figure 3.3). The statistical properties of the time series are presented in table 3.1.

In the Normal distribution skewness is 0 (symmetric around the mean). We observe that the skewness for the German natural gas prices is skewed to the right in the distribution with skewness of 4.23. Kurtosis is an indicator of tail fatness and based on the very high levels of kurtosis (kurtosis in the normal distribution is 3) in these data we argue that our price data for German natural gas prices does show indications of fat tail behavior.

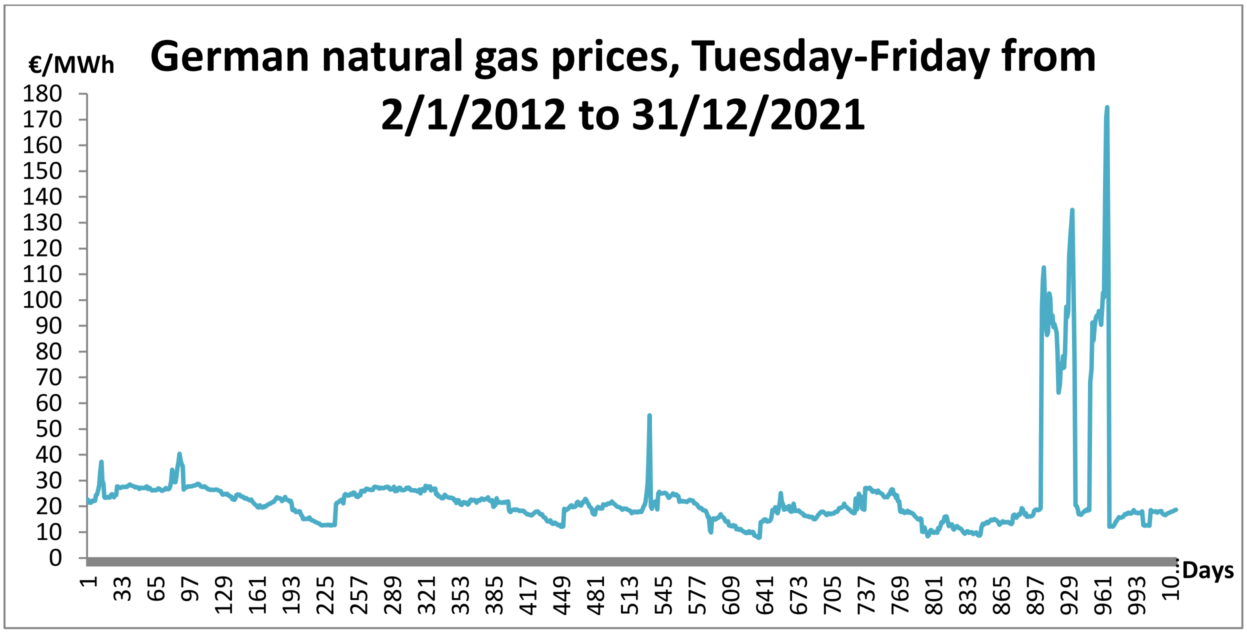

Figure 3.3: The time series for the German natural gas data used in our analysis. The data consists of weekday data from Tuesday to, and including, Friday every week in the winter quarters (Q4 and Q1) from Q1 2012 to, and including, Q4 2021.

Thus we conclude that the German natural gas prices are not normally distributed.

Statistical properties of German gas prices

	Mean	Variance	Standard deviation	Skewness	Kurtosis
German natural gas price	23.80	328	18.11	4.23	23.68

Table 3.1: Display of statistical properties (mean, variance, standard deviation, skewness and kurtosis) of the German natural gas prices.

The statistical properties of the temperature data and the RES data are presented in table 3.2. The second time period is slightly milder than the first time period. We expect some more data in the sub-dataset for the high temperature threshold than the low temperature threshold in the second period. For the RES we clearly see that the numbers are generally higher in the second period than in the first period. This is a consequence of much higher capacity for RES in Germany in time period 1 compared with time period 2.

The daily averaged RES-production in Germany over both time periods are shown in figure 3.4.

Statistical properties of temperature and RES

	Mean	Median	Variance	Standard deviation	Skewness	Kurtosis
Temperature period 1	5.37	5.25	20.32	4.5	-0.16	3.12
Temperature period 2	5.74	5.59	18.93	4.35	-0.028	3.05
RES period 1	10,465	8893	40,894,564	6,394	0.95	3.44
RES period 2	18,936	17676	93,375,975	9,663	0.49	2.6

Table 3.2: Display of statistical properties (mean, median, variance, standard deviation, skewness and kurtosis) of the temperature data and the RES data, divided by their respective time periods.

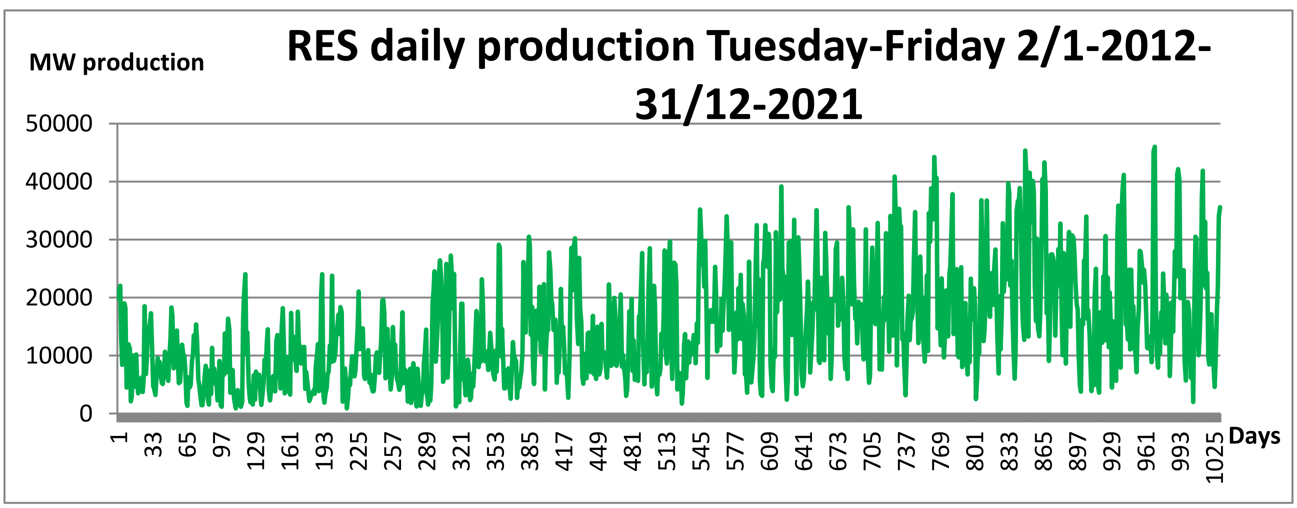

Figure 3.4: Daily-averaged RES production in Germany from 2/1/2012-31/12-2021. The weekdays Tuesday-Friday are included in the data.

The visual representation of data confirms what we see in table 3.2. The mean of RES clearly increases with time. The variability is also higher at the end of the time series than it is at the beginning.

4. Results

We will first present estimates of the tail indices for the full time period. The full sample from 2012-2021 is used and we use roughly 30 % of the available data for calculation (15% of the data with the lowest values, 15% of the data with the highest values) with κ equal to 150. This is consistent with Huisman et al. (2001), who recommends using 10-20 % of the data on either side for calculation of the tail index.

Tail index estimation for the full dataset

	Left tail	**Right tail**
Full data	0.3615 (0.00019) N=1031 κ=150	0.578 (0.00014) N=1031 κ=150
Difference Right tail – Left tail:	*0.217* *t = 917*	

Table 4.1: The table shows tail indices for the entire dataset. The standard errors are shown in parenthesis, the number of data in total is denoted by N and the number of data used as the tail of the data is denoted by κ. In the third row of the table we see the difference between the left tail and the right tail (with the corresponding t-value).

For the entire time period table 4.1 shows that the right tail is fatter than the left tail.

Next, we want to compare the two time periods of the data. We have split the data between 2/1/2012 - 31/12/2016 (which we refer to as the *early* time period or time period one) and 2/1/2017 – 31/12/2021 (which we refer to as the *late* time period or time period two). The left tail index is smaller in the first time period than in the second time period (table 4.2). The right tail index is very small in the first time period, indeed lower than the left tail index in the first period, but increases significantly in the second period. We should note that we used a different value of *κ* in this calculation compared to the results shown in table 4.1 as the number of data used in each calculation is roughly halved compared to the number of data used to compute the results in table 4.1. The differences between the tails in both periods are large at a high level of confidence.

Tail estimation for the two time periods

	Left tail	Right tail
Early period	0.2342 (0.0003) N = 515 κ=70	0.0858 (0.0002) N=515 κ=70
Right tail – Left tail:	*-0.1484* *t = -411*	
Late period	0.3013 (0.0003) N=516 κ=70	0.5587 (0.00006) N=516 κ=70
Right tail – Left tail:	*0.2574* *t = 841*	

Table 4.2: The table shows the tail indices for the two time periods. The standard errors are shown in parenthesis, the number of data in total is denoted by N and the number of data used as the tail of the data is denoted by κ. In the third row of the table we see the difference between the left tail and the right tail (with the corresponding t-value).

As the data suggest that the tail indices increase over time, we introduce temperature and intermittent renewable electricity supply (RES) as possible explanations for why this happens. We have already separated the data into two time periods (early and late) and now we split the data in both time periods with regards to RES. We do two calculations in that case as we want to investigate both what happens if we use a threshold for *high* and *low* RES that is fixed between the time periods and a relative threshold that is different in time period one and time period two. The fixed threshold is 13000 MW RES for both periods. The thresholds for relative RES is set to 8900 MW in period one and 17900 MW in period two.

The second variable we wish to investigate is temperature. Since natural gas is used for heating in Germany, we expect increased demand for natural gas when temperatures are low. The results are presented in table 4.3 where the values for κ are adjusted for the number of data used to calculate the tail indices in each individual case.

The right tail indices are uniformly higher in the second period than in the first period for corresponding RES and temperature regimes. The right tail indices in the second period are also responding to RES, such that the tail indices decrease when there is high RES. This is apparent both when we use fixed and relative thresholds for RES. The right tail indices are also lower for high temperatures than for low temperatures in both periods, which is in line with our expectations.

Tail estimation with respect to RES and temperature

	Early		Late	
	Left tail	**Right tail**	**Left tail**	**Right tail**
Low RES (relative)	0.291 (0.0003) N=259 κ=25	0.1011 (0.0004) N=259 κ=25	0.244 (0.0003) N=256 κ=25	0.3651 (0.0002) N=256 κ=25
Right tail – Left tail:	*-0.1899* *t = -379.8* *v = 478*		*0.1211* *t = 335.9* *v = 444*	
High RES (relative)	0.081 (0.0005) N=256 κ=25	0.0985 (0.0004) N=256 κ=25	0.1514 (0.0004) N=260 κ=25	0.1214 (0.0004) N=260 κ=25
Right tail – Left tail:	*0.018* *t = 27* *v = 486*		*-0.0044* *t = -53* *v = 518*	
Low RES (fixed)	0.295 (0.0002) N=373 κ=40	0.100 (0.0004) N=373 κ=40	0.310 (0.0006) N=154 κ=20	0.623 (0.0003) N=154 κ=20
Right tail – Left tail:	*-0.1915* *t = -436* *v = 547*		*0.313* *t = 466.6* *v = 225*	
High RES (fixed)	0.111 (0.001) N=142 κ=20	0.071 (0.001) N=142 κ=20	0.169 (0.0003) N=362 κ=40	0.364 (0.0002) N=362 κ=40
Right tail – Left tail:	*-0.04* *t = -28.3* *v = 282*		*0.198* *t = 540.8* *v = 266*	
Low temp.	0.09 (0.0004) N=267 κ=25	0.1251 (0.0004) N=267 κ=25	0.320 (0.0003) N=250 κ=25	0.6031 (0.00015) N=250 κ=25
Right tail – Left tail:	*0.0315* *t = 55.6* *v = 532*		*0.2831* *t = 844* *v = 366*	
High temp.	0.188	0.0108	0.125	0.1570

	(0.0004) N=248 κ=25	(0.0007) N=248 κ=25	(0.0004) N=266 κ=25	(0.0003) N=266 κ=25
Right tail – Left tail:	*-0.1772* *t = -219.8* *v = 392*		*0.032* *t = 64* *v = 491*	

Table 4.3: The table shows the tail indices for the two time periods, separated, in turn, with regards to different explanatory variables. The standard errors are shown in parenthesis, the number of data in total is denoted by N and the number of data used as the tail of the data is denoted by κ. We see the difference between the left tail and the right tail (with the corresponding t-value) below the estimates for the tail indices, with corresponding *t*-value and with *v* denoting the degrees of freedom in the T-distribution.

The results for the left tail indices are more heterogeneous. They do not increase uniformly from period one to period two as the right tail indices do and they are lower for high RES than for low RES both in the first and the second time period. In addition they increase for lower temperatures in the first period, while they decrease with lower temperatures in the second period. The difference between the right tail and left tail is small when temperature is high in the second period and at its biggest when temperature is low. The right tail also responds by falling in the second time period when we move from a regime of low RES to high RES and in particular in comparison with the left tail index.

The most important thing to note about the tail indices in the first period is that the right tail index, in general, is small for all variables and most of the time at level or below the left tail index in the same period. This is in contrast to the second period when the right tail index is significantly higher than the left tail index, except for when there is high RES.

We also present the differences in the tail indices between the time periods (table 4.4). We compare the estimates for the late time period with the corresponding tail indices in the early time period. While the right tail index increases in all cases, we see that this change is biggest for low RES. For the left tail indices the changes are generally lower and, for low RES, the left tail index actually decreases from time period one to time period two.

Finally we want to split the data on both explanatory variables; RES and temperature. This introduces a new problem as the number of data dwindles for each separation of data. As we can see (table 4.5) the number of data available after splitting the data with regards to both temperature and RES can become less than as low as N = 47 as in the case of low fixed RES and high temperature in the second period. This is also the motivation for choosing a *relative* threshold with regards to RES and especially in this calculation the results from the calculation with a fixed threshold for RES should be read with some skepticism. With relative thresholds the data becomes more fairly partitioned between the datasets. As in the earlier cases, the level of κ is adjusted with an eye to the number of data available for the calculation.

Tail index difference between the early and late time period in the sample

	Left tails		Right tails	
	Early	**Late**	**Early**	**Late**
Low RES (relative)	0.291 (0.0003) N=259 κ=25	0.244 (0.0003) N = 256 κ = 25	0.1011 (0.0004) N=259 κ=25	0.3651 (0.0002) N=256 κ=25
Tail (late) – tail(early):	*-0.047* *t = -110.8* *v = 522*		*0.264* *t = 590* *v = 375*	
High RES (relative)	0.081 (0.0005) N=256 κ=25	0.154 (0.0004) N=260 κ=25	0.0985 (0.0004) N=256 κ=25	0.1214 (0.0004) N=260 κ=25
Tail (late) – tail(early):	*0.0703* *t = 7.13* *v = 488*		*0.0229* *t = 40.5* *v = 513*	
Low RES (fixed)	0.295 (0.0002) N=373 κ=40	0.310 (0.0006) N=154 κ=20	0.0931 (0.0004) N=356 κ=40	0.670 (0.0005) N=130 κ=20
Tail (late) – tail(early):	*0.015* *t = 17.7* *v = 433*		*0.5769* *t = 900.97* *v = 302*	
High RES (fixed)	0.111 (0.001) N=142 κ=20	0.169 (0.0003) N=362 κ=40	0.0960 (0.0009) N=159 κ=20	0.334 (0.0002) N=386 κ=40
Tail (late) – tail(early):	*0.058* *t = 55.5* *v = 166*		*0.238* *t = 258* *v = 173*	
Low temp.	0.09 (0.0004) N=267 κ=25	0.32 (0.0003) N=250 κ=25	0.1251 (0.0004) N=267 κ=25	0.6031 (0.00015) N=260 κ=25
Tail (late) – tail(early):	*0.23* *t = 460* *v = 485*		*0.478* *t = 1119* *v = 339*	
High temp.	0.188	0.125	0.0108	0.1570

	(0.0004) N=248 κ=25	(0.0004) N=266 κ=25	(0.0005) N=248 κ=25	(0.0003) N=266 κ=25
Tail (late) – tail(early):	*-0.063* *t = -111* *v = 511*		*0.1462* *t = 250* *v = 407*	

Table 4.4: The table shows the tail indices for the two time periods, separated, in turn, with regards to different explanatory variables. The standard errors are shown in parenthesis, the number of data in total is denoted by N and the number of data used as the tail of the data is denoted by κ. We see the difference between the tail indices of the same part of the distribution compared between the different time periods below the estimates for the tail indices, with corresponding *t*-value and with *v* denoting the degrees of freedom in the T-distribution.

For the second time period a notable effect is that both the tail indices drop significantly for high level of RES when we go from low to high temperatures. This happens for both the left tail and the right tail and for both relative and fixed thresholds. Both temperature and RES affects the tail indices for German gas prices. For the right tail indices these effects are as expected. When there is abundant RES, the tails indices are lower, and when temperatures are higher, tail indices fall. The difference between the tail indices between time periods early and late is presented in table 4.6.

Tail indices calculated with respect to both temperature and RES

	Early				**Late**			
	Left tail		**Right tail**		**Left tail**		**Right tail**	
	Low temp.	**High temp.**	**Low temp.**	**High temp.**	**Low temp.**	**High temp.**	**Low temp.**	**High temp.**
Low RES (relative)	0.292 (0.0005) N=159 κ=20	0.2790 (0.0016) N=100 κ=20	0.1054 (0.0009) N=159 κ=20	0.0136 (0.0032) N=100 κ=20	0.4022 (0.0005) N=147 κ=20	0.3332 (0.0012) N=109 κ=20	0.7917 (0.0003) N=147 κ=20	0.6327 (0.0009) N=108 κ=20
High RES (relative)	0.1695 (0.0017) N=108 κ=20	0.1573 (0.0009) N=148 κ=20	0.1257 (0.0019) N=108 κ=20	0.0126 (0.0014) N=148 κ=20	0.4285 (0.0012) N=103 κ=20	0.1302 (0.0008) N=157 κ=20	0.6382 (0.001) N=103 κ=20	0.3288 (0.0005) N=157 κ=20
Low RES (fixed)	0.1242 (0.0004) N=211 κ=20	0.2638 (0.0005) N=162 κ=20	0.113 (0.0005) N=207 κ=20	0.0112 (0.0014) N=149 κ=20	0.3747 (0.0016) N=95 κ=15	0.3468 (0.003) N=59 κ=15	0.8061 (0.0009) N=83 κ=15	0.7290 (0.005) N=47 κ=15
High RES (fixed)	0.1470 (0.0029) N=56 κ=10	0.1903 (0.003) N=86 κ=20	0.170 (0.0024) N=60 κ=10	0.0170 (0.0032) N=99 κ=20	0.3321 (0.0005) N=155 κ=20	0.2077 (0.0007) N=207 κ=30	0.6432 (0.0003) N=167 κ=20	0.4750 (0.0004) N=219 κ=30

Table 4.5: The table shows the tail indices for the two time periods, separated, in turn, with regards to different explanatory variables. In this case we split the data in both time periods with regards to *both* temperature and RES. We calculate the tail indices both using relative RES and fixed RES thresholds. For the relative RES we use 8900 MW as threshold for the early time period, 17900 MW as threshold for the late time

period. The fixed threshold used is 13000 MW. The number of data in total is denoted by N and the number of data used as the tail of the data is denoted by κ.

The right tail index increases significantly from time period one to time period two. The difference increases with regard to RES and temperature and an absence of RES combined with low temperatures make the biggest impact for the right tail indices. In general we can conclude that the changes in the left tail indices over time are lower than the changes in the right tail indices.

Tail index differences for tail indices calculated with respect to temperature and RES

	Late tail – early tail			
	Left tails		Right tails	
	Low temp.	**High temp.**	**Low temp.**	**High temp.**
Low RES (relative)	0.1102 t = 189 v = 256	0.0542 t = 27.1 v = 187	0.6863 t = 723 v = 192	0.6191 t = 186.2 v = 114
High RES (relative)	0.259 t = 124.5 v = 190	-0.0271 t = -22.5 v = 296	0.5125 t = 1005 v = 161	0.3162 t = 212.7 v = 184
Low RES (fixed)	0.2505 t = 151.9 v = 105	0.083 t = 27.3 v = 61	0.6931 t = 673.2 v = 135	0.7178 t = 138.2 v = 53
High RES (fixed)	0.1851 t = 62.9 v = 58	0.0174 t = 5.65 v = 94	0.4732 t = 195.6 v = 60	0.458 t = 142.0 v = 101

Table 4.6: The table shows the differences in tail indices between the two time periods. In this case we split the data in both time periods with regards to *both* temperature and RES. We calculate the differences between the right tail indices and the left tail indices between the two time periods. The number of data in total is denoted by N and the number of data used as the tail of the data is denoted by κ, with corresponding *t*-value and with *v* denoting the degrees of freedom in the T-distribution.

The amount of data is unevenly distributed between the datasets where we use fixed thresholds for RES and we believe that is a limitation of the analysis. We believe the results with relative RES is more accurate with regards to the effect of RES on the German gas prices.

5. Discussion

In this section we will discuss our findings in more detail and provide some economic interpretation. However, we will first briefly discuss and verify the method we used for our analysis.

Existing literature on analyzing tail fatness follows two different strands: The more traditional and well-established approach from Hill (1975) and the more recent suggested by Huisman et al. (2001). In order to calculate robust estimates we are using Huisman et al. (2001). The main advantage is that our estimates are less sensitive to the choice of κ.

We still need to make a qualitative assessment of the choice of κ and follow Huisman et al. (2001) who recommends using 10%-15% of the sample for the left tail, and similar for the right tail. If this is the correct choice needs to be judged upon qualitatively from the available data. In addition, we note that the assumption of linearity in the estimates is a necessity for unbiased estimates of the regression parameters. Once the linearity condition is fulfilled the choice of κ is non-consequential as the omission of a small number of data (or addition of more than necessary data) into a regression does not affect the estimate to the same degree as it would in a single point estimation done with the estimator proposed by Hill (1975).

Turning to the discussion of the results we note that the most apparent effect we observe is that the right tail indices of German natural gas prices increase over time. While they rise with a varying degree depending on which variable (RES or temperature) we analyze, the rise is uniform in all tests we perform. The left tails do not increase uniformly over time. The changes in the left tail indices are also of a smaller magnitude than for the right tail indices from time period 1 to time period 2.

The tail index can be considered as a measure for tightness (or lack of tightness) in the pricing mechanism. When the tail index is low the natural gas system is matching demand with supply most days without problems. When the tail index is high, such as we see in the right tail index in the second time period, it may be a symptom of a natural gas system which is stressed where the system struggles to match supply with demand. Similarly, when the left tail index rises it may be a symptom of gluts of supply that cannot be matched with low demand. We observe both these effects in the results as both the right tail index and the left tail index increases in the second time period.

An extreme price is a relative phenomenon, where the *extremeness* of the price distribution is determined by the distance of the tails from the majority of the prices in the distribution. That makes the estimation of tail indices a good measure for tightness in the natural gas system. If the frequency of extreme prices increase and the displacement from the center of the distribution is large, then it is an indication of a market where either gluts or shortages are more frequent and more expensive to remedy.

The increase in the right tail fatness over time (table 4.2) is in line with our assumptions, outlined in the introduction. The domestic production of natural gas in the European Union has declined since 2012[12]. While the volumes in domestic production have been replaced by Russian gas and liquefied natural gas (LNG) the reduced domestic production introduces vulnerability to the supply system for European consumers of natural gas. As domestic production of natural gas has dropped the vulnerability with respect to Russian gas pipelines and the global LNG-market has increased. In periods of high demand for LNG from Asia, Europe has been forced to compete in bidding wars for LNG-supply. When the war in Ukraine started the natural gas flow from Russia dwindled suddenly. These are examples of the increased vulnerability that lower domestic natural gas supply entails.

The increase in the left tail fatness over time (table 4.2) is more modest than for the right tail fatness. That it increases at all is in fact an interesting phenomenon at least in the light of our explanation for why the right tail fatness increases. If the explanation for increased right tail fatness is correct (that the natural gas system is suffering from a tighter supply and demand balance in the second time period than in the first), then we would expect the left tail fatness to decline. A tighter supply and demand balance make gluts of natural gas less likely to occur.

[12] https://ec.europa.eu/eurostat/databrowser/view/NRG_CB_GAS__custom_3820478/default/table?lang=en

One explanation for why this does not happen might be that the natural gas supply, also in the second time period, has been healthy most of the time, thus making very low prices possible. In 2020 the demand for natural gas plummeted as a consequence of the Covid-19 pandemic, as discussed in more detail by Norouzi (2021). Since the prices were low over a long time period, the left tail of the price distribution was more truncated. This in turn may have reduced the distance between the prices at the left tail of the German natural gas prices, resulting in a lower tail index than we see for the right tails of the natural gas price distribution.

To compare the tail indices over time for similar demand-conditions, we use temperature as proxy for natural gas demand. We observe that particularly the right tail index increases a lot in time period 2 compared to time period 1 for low temperatures (table 4.3). When the temperature is low the natural gas demand, through heating demand, in Germany increases. This confirms our assumption that the balance between supply and demand is tighter in the second time period compared to the first time period. In the second time period it takes less of an increase in demand to generate more extreme prices in the German natural gas market. While the most extreme prices occur in the fall 2021 it is still an open question what came first. Did the cold autumn of 2021, with low RES cause the high prices or were the high prices a result of years of weakening supply and demand balance in the European gas markets?

When we calculate the tail indices with respect to RES we see an increase in tail indices from period 1 to period 2 for the right tail when RES is low (table 4.3). This increase confirms what we see in the results when we test the tail indices with respect to temperature. We argue that the supply and demand balance in the natural gas system in Germany is tighter when demand increases through heating demand. Similarly, the supply and demand balance tightens in the German natural gas system when RES is low. This effect occurs because demand for natural gas from electricity production increases when there is less RES in the electricity system. Then gas fired electricity production is needed to replace the lacking RES to balance the electricity system. This is explained through the merit order (chapter 1 and Erdmann (2017)) and explains why RES may have an impact in German natural gas prices. RES might have contributed to a stronger feedback mechanism between the pricing in the electricity market and the pricing in the natural gas market.

When we look at the impact of high RES we see a different picture. In days with high RES we observe that the tail indices are generally lower in the second period compared to the tail indices when RES is low. For days with high RES (relative threshold) the right tail index only exhibit minor changes. When we test the changes with a fixed threshold the change from the first time period to the second time period is more muted for the right tail than the case was in the days with low RES. This confirms our hypothesis that the introduction of RES into the German electricity market affects the price distribution in the German natural gas market.

The same effect is seen in the left tail indices with respect to RES. On days with low RES the left tail index is higher. We do observe a slight decrease in the left tail indices on days with high RES. This confirms that the introduction of RES into the German electricity market has an impact also in the German natural gas market. The impact is twofold. The left tail indices decreases with RES, such that the left tails of the German natural gas prices get less fat with less RES in the German electricity market. The right tail indices of the German natural gas prices increase when there is low RES in the German electricity market. The increase in RES decreases the probability for extremely high natural gas.

The increased tail fatness in the German natural gas market is in line with the results Huisman et al. (2022) found for the German electricity market. They showed that the tail indices of German electricity prices are higher when both the RES and non-intermittent supply of electricity is low. In our research we show a similar effect in the German natural gas market. This indicates that the effects of RES in the German electricity market is contagious and spreads to the natural gas market as well. The mechanism for this spread is explained through the merit order pricing mechanism in the German electricity market , explained by Erdmann (2017). Depending on (inelastic) demand in both the German natural gas market and in the German electricity market we show that the introduction of RES into the German electricity market increases the probability of extreme prices in the German natural gas market. This is inconsistent with the findings of Koch (2014). He argues that extreme prices in the energy futures markets are a product of financial demand for energy futures rather than fundamental demand and supply shifts. Our study shows that the extreme prices in the spot price (in the German electricity market) and in the day ahead market for German natural gas are connected with fundamental factors such as supply and demand. Although we have not analyzed this specifically in this study it is not unreasonable to infer that fundamental changes in the short term markets affects the futures markets for energy products.

The effect of RES on German natural gas prices is confirmed by the next step of our analysis. Here we look at the tail indices of the German natural gas market and split the data both with respect to temperature and with respect to RES. When we look at the combination of low RES and temperature with the relative threshold for RES, we see that the effect of low RES is clearly a major contributor in the tail fatness for both the left tail and the right tail (table 4.5). The right tail index for low temperature and low RES (relative threshold) is higher than for high temperature and low RES. For high RES and low/high temperatures the right tail indices are much lower. This effect shows that the tail index is lower when more RES is in the German electricity for the same temperature conditions. It also shows that once the high RES and the high temperatures coincide, then the right tail index drops off significantly and the probability of extreme prices drops considerably. The same effect is observed when we use the fixed threshold for RES, although the numbers differ.

Looking at the left tail estimates when we combine temperature and RES (table 4.5) we generally see an increase in the tail indices from the first period to the second period. However, as already noted, the increase is more muted than for the right tail indices. For the combination high RES/high temperature the tail index decreases slightly from period one to period two when using the relative threshold for RES. One effect of the left tails that should be noted is that in contrast to the right tails, the left tails are limited. Natural gas prices never (in our sample) go below zero, while the upside is unlimited (in principle) for the right tails. This might contribute to lower left tail indices as well.

Looking at the results and the discussion thus far we confirm our hypothesis that we introduced in the introduction chapter. Tail indices in the German natural gas markets do increase over time. And, as we expected in our hypothesis, the tail indices for the right tails increase the most.

The direct consequence of increased tail indices in the German natural gas market is a higher probability for extreme prices. We find support for our claim that the right tail becomes more sensitive to extreme prices than the left tail over time. This indicates a tightened supply and demand balance in the German natural gas market which is exacerbated by the introduction of RES into the German electricity market. Germany produced less than 50% of the natural gas it produced in 2012 in 2020. In the EU as a whole the same scenario has been unfolding over the same time period, where the EU produced less than 50% of the natural gas it produced in 2012 in 2020. Since the natural gas market is a market with local constraints of supply due to infrastructure, the loss of local production increases vulnerability of supply of natural gas.

Although it was already known that German natural gas prices were not normally distributed, what is new in this study is that the non-normality (tail fatness) in German natural gas prices can be forecasted with demand and volume of RES. For risk managers this implies that models should be made conditional on these variables. One should also use models where the tail structures can be easily adjusted to shifting supply and demand conditions in the natural gas market. This will also be useful for hedging decisions when hedging for extreme losses.

The German electricity market and natural gas market are connected through gas fired power plants and heating demand. The Northern European electricity market is tightly interconnected, with Germany at its center. The Northern-European natural gas market is connected through an extensive network of natural gas pipelines. The policy of one country affects many countries. Policy makers should know the risks and beware the balance of supply and demand in interconnected energy markets before implementing new policies.

6. Conclusion

We estimate the tail indices of German natural gas prices for two time periods from 2012-2021. The domestic supply of natural gas in Germany in particular, and the European Union as a whole, has decreased during this time period. We test the hypothesis that tail fatness in German day-ahead natural gas prices has changed over time.

For the analysis of the tail indices for the German day-ahead contracts we split our data into two time periods: 2012-2016 and 2017-2021. Using the approach suggested by Huisman et al. (2001) we conclude that both the left and the right tail index have increased over time. This implicates a higher probability of extreme prices in the second period.

In addition, we have analyzed the tail indices for the German day-ahead contracts for natural gas with respect to temperature, renewable energy supply (RES) and temperature and RES in combination. Here we find that the left tail fatness increases for low temperatures and decreases for high temperatures. For the right tail fatness the increase is bigger for low temperatures than the increase for high temperatures. This confirms our assumption that the probability for extremely high prices has increased in the second time period. We argue that this is a result of a more vulnerable supply system of natural gas in Germany in the second time period.

With low RES in the electricity system the right tail indices increase in the second period, both for a fixed threshold and for a relative threshold of RES. There is an increase in the right tail indices also with high RES in the electricity system, but the increase is much larger for low RES.

In conclusion our analysis shows that the tail fatness increases over time for German natural gas prices. Further research should address the dynamics of price formation and recent price developments.

References

Beirlant, J., Vynckier, P., Teugels, J. (1996). Tail-index estimation, Pareto Quantile Plots and Regression Diagnostics. *Journal of the American Statistical Association*, *91*, 1659-1667.

Benth, F. E., Benth, J. S., Koekebakker, S. (2008). *Stochastic Modelling of Electricity and Related Markets* (O. E. Barndorff-Nielsen, Ed.). World Scientific.

Bhattacharya, S., Kallitsis, M. (2019). Data-adaptive trimming of the Hill estimator and detection of outliers in the extremes of heavy-tailed data. *Electronic Journal of Statistics*, *13*, 1872-1925.

Biber, A., Felder, M., Wieland, C., Spliethoff, H. (2022). Negative price spiral caused by renewables? Electricity price prediction on the German market for 2030. *The Electricity Journal*, *35*(8).

Bryson, M. C. (1974). Heavy-Tailed Distributions: Properties and Tests. *Technometrics*, *16*(1), 61-68.

Cox, D. R., Hinkley, D. V. . (1974). *Theoretical Statistics*. Chapman Hall.

Dacorogna, M., Müller, U., Pictet, O., de Vries, C. (1995). The Distribution of Extremal Foreign Exchange Rate Returns in Extremely Large Datasets. In T. Institute (Ed.), *Discussion paper*.

Erdmann, G. (2017). *Power system and market integration of renewable electricity* EPJ Web of Conferences, Varenna, Italy.

Franza, L. (2014). *Long-Term Gas Import Contracts in Europe* (Clingendael International Energy Programme, Issue. C. I. E. Programme.

Hall, P. (1990). Using the Bootstrap to Estimate Mean Square Error and Select Smoothing Parameters in Non-parametric Problems. *Journal of Multivariate Analysis*, *32*, 177-203.

Heather, P. (2021). *European Traded Gas Hubs: German hubs about to merge*.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, *3*(5), 1163-1174.

Huisman, R., Koedijk, C.G., Palm, F. C., Kool, C. J., . (2001). Tail Index Estimates in Small Samples [Journal Article]. *American Statistical Association Journal of Business & Economic Statistics*, *19*(1), 208-215.

Huisman, R., Kyritsis, E., Stet, C. (2022). Fat Tails due to Variable Renewables and Insufficient Flexibility: Evidence from Germany. *The Energy Journal*, *43*(5), 231-244. (VATT Institute for Economic Research)

IEA. (2020). *Gas 2020*.

IEA. (2022). *World Energy Outlook*.

Jansen, D., de Vries, C. (1991). On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective. *The Review of Economics and Statistics*, *73*, 18-24.

Koch, N. (2014). Tail events: A new approach to understanding extreme energy commodity prices. *Energy Economics*, *43*, 195-205.

Kyritsis, E., Andersson, J., Serletis, A. (2017). Electricity prices, large-scale renewable integration, and policy implications. *Energy Policy*, *101*, 550-560.

Liebensteiner, M., Wrienz, M. (2020). Do Intermittent Renewables Threaten the Electricity Supply Security? *Energy Economics*, *87*.

Lv, X., & Shan, X. (2013). Modeling natural gas market volatility using GARCH with different distributions. *Physica A: Statistical Mechanics and its Applications*, *392*(22), 5685-5699.

Norouzi, N. (2021). Post-COVID-19 and globalization of oil and natural gas trade: Challenges, opportunities, lessons, regulations, and strategies. *International Journal of Energy Research*, *45*(10).

Pareto, V. (1897). The New Theories of Economics. *Journal of Political Economy*, *5*(4), 485-502.

Resnick, S., Starica, C. (1997). Smoothing the Hill Estimator. *Andvances in Applied Probability*, *29*, 271-293.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika*, *6*(5), 309-316.
Welch, B. L. (1947). The Generalization of 'Student's' Problem when Several Different Population Variances are Involved. *Biometrika*, *34*, 28-35.
Xiaoyi, M., Haichun, Y. (2018). Towards an Integrated Spot LNG Market: An Interim Assessment. *The Energy Journal*, *39*(1), 211-234.

A. Code used for calculation of tail indices

```
library(pracma)

hill_right_tail <- function(vektor, kappa)

{

 lengde = length(vektor)

 indices1 = matrix(data=0, nrow = kappa, ncol = 2)

 gakk = sort(vektor)

 for(i in 1:kappa) #i denotes kappa-value which we use to calculate tail index

 {

  tail_index = 0

  for( j in 1:i) #Calculation of tail index for kappa=i and store it for regression

  {

   tail_index = tail_index + log(gakk[lengde-j+1])-log(gakk[lengde-i])

  }

  tt =  tail_index/i

  indices1[i,1] = tt

  indices1[i,2] = i

 }

 #(indices1)

 return(indices1)

}

hill_left_tail <- function(vektor, Kappa)

{
```

```
 lengde = length(vektor)

 indices1 = matrix(data=0, nrow = Kappa, ncol = 2)

 gakk = sort(vektor)

 for(i in 1:Kappa) #i denotes kappa-value which we use to calculate tail index

 {

  tail_index = 0

  for( j in 1:i) #Calculation of tail index for kappa=i and store it for regression

  {

   tail_index = tail_index + log(gakk[Kappa+1]) - log(gakk[i-j+1])

  }

  tt =  tail_index/i

  indices1[Kappa-i,1] = tt

  indices1[i,2] = i

 }

 return(indices1)

}

calc_H<-function(resultat, Kappa)

{

 #Creating weight matrix

 W = matrix(data=0, nrow = Kappa, ncol = Kappa)

 for(i in 1:Kappa)

 {

  for(j in 1:Kappa)
```

```
 {
  if(i == j)
  {
   W[i,j] = sqrt(i)
  }
 }
}
#Constructing Z-matrix
Z = matrix(data=1, nrow = Kappa, ncol = 2)
for(i in 1:Kappa)
{
 Z[i,2] = i
}
Beta = inv(t(Z)%*%t(W)%*%W%*%Z)%*%t(Z)%*%t(W)%*%W%*%resultat
 return(Beta)
}
calc_H2R<-function(resultat, Kappa)
{
#Calculates the modified Hill estimator by weighted average
 W = matrix(data=0, nrow = Kappa, ncol = Kappa)
 for(i in 1:Kappa)
 {
 for(j in 1:Kappa)
```

```
 {
  if(i == j)
  {
   W[i,j] = sqrt(i)
  }
 }
}
gamma=0
summer = 0
for(j in 1:Kappa)
{
 gamma = gamma+W[j,j]*resultat[j,1]



 summer = summer+W[j,j]


}
 w_gamma = gamma/summer


 return(w_gamma)
}
calc_H2RL<-function(resultat, Kappa)
{
```

```
#Calculates the modified Hill estimator by weighted average
 W = matrix(data=0, nrow = Kappa, ncol = Kappa)


 for(i in 1:Kappa)
 {
  for(j in 1:Kappa)
  {
   if(i == j)
   {
    W[i,j] = sqrt(i)
   }
  }
 }
 gamma=0
 summer = 0
 for(j in 1:Kappa)
 {
  gamma = gamma+W[j,j]*resultat[j,1]


  summer = summer+W[j,j]


 }
  w_gamma = gamma/summer
```

```
  return(w_gamma)

}


calc_SE <- function(xx, estimate, Kappa, n)

{

 A = matrix(data = 0, nrow = Kappa, ncol = Kappa+1)


 for( s in 1:(Kappa))

 {

  for(t in (Kappa):(Kappa-s))

  {


   A[s,t+1] = 1/s

   A[s,Kappa-s+1] = -1

   }

  }


   Sigma = matrix(data=NA, nrow = (Kappa+1), ncol = (Kappa+1))


   for(i in 1:(Kappa+1))

   {

    Intercept = estimate
```

```
 p_i = i/n

 mu_i = log((1-p_i)^(-Intercept))

 f_mu_i = Intercept*mu_i^(-Intercept-1)

for(j in i:(Kappa+1))
{
 p_j = j/n

 j_n = (1-(p_j))

 mu_j = log((1-p_j)^(-Intercept))

 f_mu_j = Intercept*mu_j^(-Intercept-1)
 v_i_j_tell = p_i-p_i*p_j

 v_i_j_nevn = n*f_mu_i*f_mu_j

 v_i_j = v_i_j_tell/v_i_j_nevn
```

```
   Sigma[i,j] = v_i_j
   Sigma[j,i]  = Sigma[i,j]
  }
 }

 Omega = A%*%Sigma%*%t(A)
 return(Omega)
}
xx1 = read.table("Del1LavTysk.txt")
xx2 = read.table("Del1HøyTysk.txt")
xx3 = read.table("Del2LavTysk.txt")
xx4 = read.table("Del2HøyTysk.txt")
xx5 = read.table("THE22.txt")
LowBar1 =  13000
HighFloor1 = 13000#8900
LowBar2 = 13000#17900
HighFloor2 = 13000#17900

xxL1 = xx1[xx1[,3] < LowBar1,]            #LowTemp LowRenew   Early
xxH1  = xx1[xx1[,3] > HighFloor1,]         #LowTemp HighRenew  Early

xxSL1 = xxL1[,1]
```

```
xxSH1 = xxH1[,1]

xxL2 = xx2[xx2[,3] < LowBar1,]             #HighTemp LowRenew   Early
xxH2  = xx2[xx2[,3] > HighFloor1,]          #HighTemp HighRenew  Early

xxSL2 = xxL2[,1]
xxSH2 = xxH2[,1]

xxL3 = xx3[xx3[,3] < LowBar2,]             #LowTemp LowRenew   Late
xxH3  = xx3[xx3[,3] > HighFloor2,]          #LowTemp HighRenew  Late

xxSL3 = xxL3[,1]
xxSH3 = xxH3[,1]

xxL4 = xx4[xx4[,3] < LowBar2,]             #HighTemp LowRenew   Late
xxH4  = xx4[xx4[,3] > HighFloor2,]          #HighTemp HighRenew  Late

xxSL4 = xxL4[,1]
xxSH4 = xxH4[,1]

Bruk = xx5[,1]

kappa1 = 5
```

```
kappa2 = 15

kappa3 = 15

kappa4 = 15


resultat11 = hill_right_tail(Bruk, kappa1)

resultat11L = hill_left_tail(Bruk, kappa1)


huismann11 = calc_H2R(resultat11, kappa1)

huismann11L = calc_H2RL(resultat11L, kappa1)

print(huismann11)

print(length(Bruk))

#print(huismann11)

Omega11 = calc_SE(Bruk, huismann11, kappa1, length(Bruk))


W = matrix(data=0, nrow = kappa1, ncol = kappa1)

 for(i in 1:kappa1)

 {

  for(j in 1:kappa1)

  {

   if(i == j)

   {

    W[i,j] = sqrt(i)

   }
```

```
 }
}

#Lager Z
Z = matrix(data=1, nrow = kappa1, ncol = 2)
for(i in 1:kappa1)
{
 Z[i,2] = i
}

cov11 = inv(t(Z)%*%t(W)%*%W%*%Z)%*%t(Z)%*%t(W)%*%W%*%Omega11%*%t(W)%*%W%*%Z%*%inv(t(Z)%*%t(W)%*%W%*%Z)
#print(huismann11)
print(sqrt(cov11[1,1]))
```

NTNU

Kunnskap for en bedre verden