

ESSAYS ON ENERGY PORTFOLIO MANAGEMENT

Von der Wirtschaftswissenschaftlichen Fakultät
der Universität Leipzig
genehmigte

DISSERTATION

zur Erlangung des akademischen Grades

Doktor rerum politicarum
(Dr. rer. pol.)

vorgelegt von

Hendrik Kohrs

geboren am 15. Juli 1991 in Bremervörde

Gutachter:

Prof. Dr. Frank Schuhmacher (Universität Leipzig)
Prof. Dr. Benjamin R. Auer (BTU Cottbus-Senftenberg)

Tag der Verleihung:
20. April 2022

Bibliographische Beschreibung:

Kohrs, Hendrik
Essays on energy portfolio management
Universität Leipzig, Dissertation
132 S., 322 Lit., 9 Tab., 29 Abb.

Referat:

Diese englischsprachige Dissertation behandelt ausgewählte Fragen zum Thema Portfoliomanagement in Energiemärkten. Im Kontext der modernen Portfoliotheorie werden theoretische Verteilungsannahmen untersucht, die einen optimalen Mittelwert-Varianz-Ansatz implizieren. Der Bereich zu Energiemärkten befasst sich einerseits mit Kurzfristprognosen von Day-Ahead-Preisen auf dem Strommarkt. Andererseits werden auf dem Erdgasmarkt die von komplexen Energiederivaten impliziten Volatilitäten analysiert. Einige interessante Beiträge, die diese Dissertation liefert, sind beispielsweise (i) die Erkenntnis, dass sich der Mittelwert-Varianz-Ansatz zur Bestimmung eines optimalen Portfolios von Vermögensgegenständen auch im Falle einer schiefen Renditeverteilung theoretisch rechtfertigen lässt, (ii) eine umfangreiche Vergleichsstudie mit verschiedenen Ansätzen zur Reduktion der Komplexität von multivariaten Strompreisprognosen und (iii) die Entwicklung eines theoretischen Rahmens und effizienten Algorithmus zur Übersetzung von Preisen für Swing-Optionen in implizite Volatilitäten.

Danksagung

Diese Dissertation ist in den Jahren 2017 bis 2021 während und nach meiner Zeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Finanzierung und Investition der Universität Leipzig entstanden. Ohne die Hilfe, Unterstützung und Inspiration verschiedener Personen und Institutionen wäre diese Arbeit niemals geschrieben worden.

Meine aufrichtige Dankbarkeit gilt meinen Doktorvätern Professor Frank Schuhmacher und Professor Benjamin R. Auer für ihre wissenschaftliche Betreuung und die großen Freiräume während meiner Forschungszeit. Ihre stetige Unterstützung und wertvollen Denkanstöße haben die Qualität dieser Dissertation ungemein verbessert.

Ein besonderer Dank geht an meinen Arbeitgeber VNG Handel & Vertrieb GmbH, der mir die Forschung in einem hochinteressanten Teilgebiet der Energiewirtschaft überhaupt ermöglicht hat. Dabei danke ich insbesondere meinen Teamkollegen in der Quantitativen Analyse für die vielen anregenden Diskussionen und konstruktiven Ideen sowie Hermann Mühlichen für das Vertrauen in meine Arbeit und seine stetige Unterstützung.

Zu guter Letzt möchte ich mich bei all meinen Freundinnen und Freunden für die unzähligen Gespräche und schönen gemeinsamen Momente bedanken, die mein Leben so lebenswert machen und mich stets weiter vorantreiben. Ich danke insbesondere meinem Freund Aleks für die Harmonie, die Abenteuer und den Quatsch sowie alles weitere, mit dem du mein Leben bereicherst. Meine tiefste Dankbarkeit gilt meinen Eltern für ihre bedingungslose Unterstützung während meines gesamten Lebens. Ihr habt mir eine so wundervolle und sorgenfreie Kindheit geschenkt, die mich zu dem gemacht haben, der ich heute bin.

Hendrik Kohrs

Contents

1	Introduction	1
2	Mean-variance portfolio theory	5
2.1	Introduction	5
2.2	Preliminary result	8
2.3	Main result	11
2.4	Additional results	14
2.4.1	Alternative sufficiency proof	14
2.4.2	Irrelevance of performance measure choice	15
2.4.3	Two-fund separation	16
2.4.4	Identifying the parameters determining portfolio return distributions	18
2.5	Empirical perspective	19
2.6	Conclusion	23
3	Multivariate electricity price forecasting	25
3.1	Introduction	25
3.2	Data and methodology	28
3.2.1	Sample selection	28
3.2.2	General technical issues	28
3.2.3	Random forests	30
3.2.4	Variable importance scores	32
3.2.5	Competing forecasting models	35
3.3	Empirical results	38
3.3.1	Forecasting error values	38
3.3.2	Diebold-Mariano test	42
3.3.3	Hansen test	45
3.4	Conclusion	47
4	Swing option-implied volatility	50
4.1	Introduction	50
4.2	Swing option theory	53
4.2.1	Contract definition	53
4.2.2	Valuation theory	54
4.2.3	Sensitivity analysis	57
4.3	Price model	59

4.3.1	General setup	59
4.3.2	Volatility function	60
4.3.3	Return specification	61
4.4	Implied volatility algorithm	64
4.4.1	Option valuation algorithm	64
4.4.2	Root-finding algorithm	67
4.5	Numerical study	71
4.5.1	Preliminaries	71
4.5.2	Value drivers	72
4.5.3	Mean-reversion rate	75
4.5.4	Seasonality	77
4.5.5	Moneyness	79
4.6	Conclusion	82
5	Summary	84
	References	86
	Appendices	101
A	Appendix of Chapter 2	101
A.1	Proofs	101
A.2	Robustness checks	105
B	Appendix of Chapter 3	107
B.1	Additional empirical results	107
B.2	Benchmark models	114
B.3	Feature extraction	120
B.4	Regularization	122
B.5	Feature selection	127
B.6	Forecast combinations	128
C	Appendix of Chapter 4	130
C.1	Proofs	130

List of Figures

3.1	Correlations between current and lagged hourly prices	30
3.2	Aggregated lag importance	32
3.3	Variable importance profiles for selected hours	34
3.4	Importance score evaluation	35
3.5	Diebold-Mariano test with alternative ‘less accurate’	46
3.6	Diebold-Mariano test with alternative ‘more accurate’	46
3.7	Hansen test	48
4.1	Gas indexation	54
4.2	Digital consumption state grid	58
4.3	Delta in additive and multiplicative models	65
4.4	Continuation value regression	68
4.5	Convex value function	71
4.6	Volatility calibration errors	72
4.7	Implied volatilities of swing options	73
4.8	Exemplary implied volatility functions	74
4.9	Quote fitting procedure	77
4.10	Estimates of the mean-reversion rate	78
4.11	Full sample seasonal classification	78
4.12	Rolling window seasonal classification	79
4.13	Intrinsic execution	81
4.14	Moneyness approximation error	81
4.15	Moneyness classification	82
B.1	Importance profiles ‘night’ I	107
B.2	Importance profiles ‘night’ II	108
B.3	Importance profiles ‘morning’	109
B.4	Importance profiles ‘high noon’	110
B.5	Importance profiles ‘afternoon’	111
B.6	Importance profiles ‘evening’ I	112
B.7	Importance profiles ‘evening’ II	113

List of Tables

2.1	NASDAQ stock returns	22
2.2	Fama-French portfolios and additional asset classes	23
3.1	Most important input variables	33
3.2	Forecasting errors	40
3.3	Mean absolute errors per hour	43
3.4	Root mean square errors per hour	44
4.1	Implied volatility function fits	74
A.1	Robustness checks	105
B.1	Optimal elastic net mixture parameters	124

Chapter 1

Introduction

Over the past decades, energy markets worldwide have undergone rapid deregulation, creating a more competitive market environment and exposing participants to potentially much greater risks. Not surprisingly, this development has made portfolio and risk management a primary concern for energy companies (see Bunn, 2004; Eydeland and Wolyniec, 2003). In fact, there is a bunch of sources driving the transformation of energy markets. With investment banks being drawn into the area as they look for new markets to operate, traditional approaches from the banking industry like mean-variance portfolio selection (see Markowitz, 2000) find introduction to the energy sector. There is also an increasing number of power marketers entering the field making accurate electricity price forecasts a valuable input to optimally self-schedule and engage in profitable bilateral contracts (see Nowotarski and Weron, 2018; Weron, 2014). In order to control increasing exposure to energy prices, the use of complex derivatives moves into the focus of many energy companies, as they often provide the simplest and most flexible solution of precise risk management (see Clewlow and Strickland, 2000).

All of these topics can easily grow into tricky problems embedded in complex theoretical frameworks. There are still approaches to reduce complexity and manage these problems. These concepts include (i) searching for theoretical frameworks which justify the restriction on most relevant indicators like mean and variance in portfolio optimization, (ii) the usage of feature selection or factor models in multivariate electricity price forecasting and (iii) the translation of quotes for complex energy option in implied volatility. In this doctoral thesis, each of these approaches is covered in one of three articles. The first article answers the question which distributional requirements of asset returns are necessary and sufficient for mean-variance analysis in portfolio optimization. The second one deals with dimensionality problems in multivariate electricity price forecasting. Finally, the third article develops an implied volatility framework for energy swing options.

In the first paper, we focus on *portfolio theory* and investigate the theoretical and empirical justification of mean-variance analysis in optimal portfolio selection which is still subject to controversial discussions in literature (see Markowitz, 2014). In an important contribution, Chamberlain (1983) shows that, in presence of a risk-free asset, the distribution of all scaled portfolios are determined by mean and variance if and only if asset returns are jointly elliptically distributed. This result has been cited quite frequently because it is usually the focus of academic interest and is often used by practitioners to reject

mean-variance analysis in portfolio optimization. However, Meyer and Rasche (1992) already point out that Chamberlain’s findings build on the premise of an unlimited portfolio universe, which even includes portfolios with negative wealth, and that restricting this universe extends the class of distributions implying mean-variance utility functions. While Chamberlain (1983) looks at a rather unusual portfolio universe, we consider the returns of portfolios and search for a larger family of distributions which allow mean-variance analysis. For this, we modify the generalized location-scale (GLS) distribution used in Meyer and Rasche (1992) by introducing elliptical noise. We then show that, if the returns of risky assets follow what we call skew-elliptical GLS distribution, all investments are ranked equivalently under expected utility theory and the mean-variance approach. This holds regardless whether there is a risk-free asset in the choice set or not. Besides this main result, we illustrate that our approach allows a simple and more catchy representation of the distributions described in Chamberlain (1983, Theorem 2). Furthermore, as we work out important model links to the mean-variance-skewness framework of Simaan (1993a), we derive the famous two-fund separation property. Finally, we extend the framework of Schuhmacher and Auer (2014) to portfolios including more than one risky asset and conclude that in our model setting investment decisions are independent of the performance measure choice. To round off our theoretical results, we finish our study by presenting some evidence on the empirical relevance of our skew-elliptical GLS model. To this end, we modify the testing approach proposed by Meyer and Rasche (1992) to be applicable to our specific GLS case. Roughly speaking, our skew-elliptical GLS model can be tested by checking normalized out-of-sample CAPM residuals for sphericity (a special case of ellipticality) via the uniform techniques of Liang et al. (2008). Using a sample of common stocks, well-known factor portfolios (capturing size, value, momentum, reversal and industry effects) and alternative investment vehicles (i. e., indices reflecting the performance of commodity futures and hedge funds), we find that the skew-elliptical GLS distribution cannot be rejected in the majority of considered settings.

The second paper deals with *electricity price forecasting* (EPF). In short-term forecasting of day-ahead electricity prices, many authors emphasize the importance of incorporating the autoregressive intraday dependency structure (see Maciejowska and Weron, 2013; Ziel, 2016). That is, we have to take into account that the price for tomorrow’s hour h may depend not only on previous prices for this hour of the day but also on previous prices for other hours. However, considering more and more autoregressive lags as explanatory variables quickly leads to models with too many parameters. There are several approaches to tackle the curse of dimensionality which also have been analysed in several studies in the context of electricity price forecasting. Among these are factor models (see Maciejowska and Weron, 2013), elastic net regressions (see Ludwig et al., 2015; Uniejewski et al., 2016; Ziel and Weron, 2018), bayesian shrinkage (see Joutz et al., 1995; Panagiotelis and Smith, 2008) as well as feature selection and splitting (see Gonz  les et al., 2016; Ludwig et al., 2015; P  rtoles et al., 2018). Unfortunately, these studies are based on data sets for different markets and, in addition, they tend to compare the employed methods to a few selected benchmarks only. With a focus on the German/Austrian market, we provide a comprehensive comparative study featuring all of these approaches instead. Our study provides several interesting results and insights. First, we systematically analyse the role of lagged hourly prices and weekday seasonalities in EPF with importance scores drawn

from random forests. These importance scores confirm previous results that prices lagged by one week have considerably more explanatory power than directly preceding smaller lags and that prices of specific daytimes can be forecasted with greater accuracy when lagged prices of different daytimes are used as predictors (see Ziel, 2016). Second, we get a profound picture of which forecasting approaches perform well in a multidimensional EPF setting. Using the most important predictors in a support vector machine is an approach which shines in comparison to other established forecasting methods. Third, we find that some forecasting methods are more accurate for certain hours of the day than others which can be explained by the presence of distinct importance profiles for different hourly prices. We see that each daytime appears to have its ideal forecasting model. For example, elastic nets are superior models when it comes to forecasting evening prices. Finally, to round off our comparison of predictive accuracy, a last contribution of our study is motivated by a growing body of literature emphasizing that combinations of forecasts derived from different models are often more reliable than forecasts from a single model alone (see Baumeister and Kilian, 2015; Clemen, 1989; Timmermann, 2006). These turn out to improve forecast accuracy even further which supports the line of research indicating that simple combination methods often work reasonably well.

In the last paper, we turn to *energy swing options* and focus on an implied volatility concept for these options. Implied volatility is one of the most important metrics for the analysis and presentation of financial options. Even though Ball et al. (1985) demonstrate that implied volatility may be calculated for exotic options as well, the literature tends to focus on standard options. In particular, there exists no concept to quote commodity swing options in terms of implied volatility despite the increasing interest of academics and practitioners in swing options (see Carmona and Ludkovski, 2010; Jaillet et al., 2004). This is where we step into the picture. We are the first to develop an approach to quote commodity swing options in terms of implied normal volatility by combining a Monte Carlo valuation procedure with the Newton-Raphson root-finding method. For this combination to work, two sets of results from the literature have to be taken into account. First, Barrera-Esteve et al. (2006) and Bardou et al. (2010) show that the optimal purchase is of digital type for swing options with different volume constraints and penalties. This simplification allows us to extend the methodology of Longstaff and Schwartz (2001), originally designed for American-style options, to the case of Swing options. Second, Bonnans et al. (2012) justify a pathwise derivative approach for swing options to calculate first-order sensitivities. It enables us to derive a formula for the vega of swing options, which is used as an input for the Newton-Raphson iteration. The convergence of this iteration can be ensured via monotonicity and convexity arguments depending on the specific price model choice. There are additional requirements on the price model to support our main objective. First, the traditional notion of implied volatility demands the measure to be unique which cannot be achieved in multi-factor or regime-switching environments because they assume distinct volatilities for each factor and state (see Wahab and Lee, 2011). Second, we require a model which captures all dimensions of flexibility offered to the option holder, notably outright and time-spread optionality (see Berger et al., 2018; Keppo, 2004). Finally, the price model has to exhibit necessary theoretical properties like the Markov property or convexity, as discussed above, to ensure an efficient valuation and root-finding procedure (see Kohrs et al., 2019). Taking these requirements into account, we formulate an additive one-factor

forward curve model in the spirit of Clewlow and Strickland (1999) which is equipped with a deterministic volatility function of negative exponential form. We complete our theoretical discussion with an empirical study. Working with our framework, we analyse the shape of the value function as well as seasonal and moneyness patterns in implied volatilities of swing option quotes. Similar to observations for standard electricity options (see Fanelli and Schmeck, 2019), we find the implied volatilities of swing option quotes to exhibit seasonality with respect to the delivery period. As far as the the moneyness of swing options is concerned, volatility smirks are not particularly pronounced in contrast to typical plain vanilla commodity market analyses (see Jia et al., 2021).

The remainder of thesis is organised as follows. We present its three articles in Chapters 2 - 4, where each chapter starts with an abstract containing the main contribution and the results of the corresponding paper. Chapter 5 then concludes with a summary of the most important findings of the thesis.

Chapter 2

Mean-variance portfolio theory

We show that, in the presence of a risk-free asset, the return distribution of every portfolio is determined by its mean and variance if and only if asset returns follow a specific skew-elliptical distribution. Thus, contrary to common belief among academics and practitioners, skewed returns do not allow a rejection of mean-variance analysis. Our work differs from Chamberlain (1983) by focusing on the returns of portfolios, where the weights over the risk-free asset and the risky assets sum to unity. Furthermore, it extends Meyer and Rasche (1992) by introducing elliptical noise into their generalized location-scale framework. To emphasize the relevance of our skew-elliptical model, we additionally provide empirical evidence that it cannot be rejected for the returns of typical portfolios of common stocks or popular alternative investments.

2.1 Introduction

In the 1950s, Markowitz (1952, 1959) laid the foundations for efficient portfolio optimization frameworks by linking expected utility maximization to mean-variance analysis. His ideas have been extended in various ways and are at the core of many approaches in modern finance. The capital asset pricing model (CAPM) of Sharpe (1964), Lintner (1965) and Mossin (1966) is probably one of the most important follow-up results. Despite the increasing complexity and uncertainty in global financial markets, the framework introduced by Markowitz continues to find wide application (see Bernard and Vanduffel, 2014; Gao and Nardari, 2018; Kolm et al., 2014; Levy and Levy, 2004; Markowitz, 2000; Ray and Jenamani, 2016). However, the theoretical and empirical *justification of mean-variance analysis remains controversial*. There are even *widespread misunderstandings concerning the necessary and sufficient conditions for its application* (see Markowitz, 2014). For example, if asset returns follow a multivariate normal distribution, or if the utility function is quadratic, it is sufficient to consider mean and variance of returns in portfolio selection because these conditions ensure consistency between the mean-variance and the expected utility model (see Baron, 1977). However, many researchers argue that these conditions either have theoretical defects or lack empirical support. Quadratic utility implies increasing absolute risk aversion and that marginal utility eventually becomes negative for risk-averse agents (see Brockett and Golden, 1987; Pratt, 1964). Furthermore, normality of asset returns is typically rejected based on empirical evidence of skewness or heavy tails (see Mandelbrot, 1963; Peiró, 1999). While these observations are correct, researchers

are often unaware that there are more general conditions which imply mean-variance utility functions such that we cannot reject Markowitz's principles based on such simplistic arguments.

Two strands of literature provide more general justifications for mean-variance analysis. Empirical studies examine whether functions of mean and variance at least serve as good approximations for expected utility in practice. Markowitz (2014) reviews this issue and stresses that many authors have ignored the favorable results of this body of research. Using different methodologies, preference settings and asset return data, Markowitz (1959), Levy and Markowitz (1979), Dexter et al. (1980), Pulley (1983), Kroll et al. (1984), Hlawitschka (1994) and Simaan (1993b, 2014) show that mean-variance-efficient portfolios can provide good approximations for expected-utility-maximizing portfolios. Hence, for the practical use of mean-variance analysis, it is necessary and sufficient to ensure that a portfolio choice derived from the mean-variance-efficient frontier approximately maximizes expected utility for a wide variety of utility functions. While this pragmatic justification motivates intensive empirical research using different data sets, it does not explain which general circumstances justify mean-variance analysis. Identifying theoretical models which imply mean-variance portfolio utility makes it much easier to justify Markowitz's approach. Moreover, as we observe good approximations, there has to be a (at least approximate) theoretical foundation for this phenomenon.

This brings us to the theoretical literature. General consistency conditions should hold for any risk-averse agent. If we allow the utility function to be an arbitrary increasing function, we can equivalently search for conditions under which the distributions of portfolio *returns* are determined by mean and variance. This refers to the *standard* (see Black, 1972; Markowitz, 2000) or *canonical* (see Ingersoll, 1987) portfolio optimization problem, which is characterized by a *full investment condition*, i. e., the constraint that the portfolio weights of all assets (including the risk-free asset, if it exists) must sum to unity. In an important contribution, Chamberlain (1983) instead considers the distributions of all *scaled* portfolios. He argues that expected utility is a function of mean and variance for every increasing utility function if and only if every portfolio with equal mean and variance has the same distribution. Then he uses this relationship to provide some insights into which distributional assumption is *necessary and sufficient* for portfolio distributions to be determined by mean and variance. Specifically, he shows that, in portfolio selection with a risk-free asset, portfolio distributions are determined by mean and variance if and only if asset returns are jointly elliptically distributed (see Chamberlain, 1983, Theorem 1). Without a risk-free asset, elliptical symmetry is still important; however, there is one additional degree of freedom with respect to the asset return distributions (see Chamberlain, 1983, Theorem 2). His result for the portfolio problem with a risk-free asset has been cited quite frequently because it is usually the focus of academic interest — as it is in our study — and is often used by practitioners to reject the Sharpe ratio as a measure of investment performance. Another important contribution has been made by Meyer and Rasche (1992). They point out that Chamberlain's findings build on the premise of an unlimited portfolio universe, which even includes portfolios with negative wealth, and that restricting this universe extends the class of distributions implying mean-variance utility functions. Specifically, they illustrate that a generalized location-scale (GLS) distribution, which allows asymmetries in asset returns, is *sufficient* to imply mean-variance utility for

the primary assets in the choice set (in other words, for very simple portfolios with only one risky asset), but not necessarily for portfolios thereof.

While Chamberlain (1983) and Meyer and Rasche (1992) look at rather unusual portfolio universes, we consider asset combinations, which are the basis of a wide variety of portfolio applications. Our research question is, which distributions imply mean-variance utility functions in the context of standard portfolio optimization with *full investment constraint*. With this in mind, we automatically search for the distributions relevant for portfolio *returns* as opposed to those relevant for *scaled* portfolio values, which do not necessarily represent returns.¹ Our answer can be divided between a preliminary result (sufficient condition) and a main result (necessary condition, full characterization). To derive the former, we modify the GLS distribution used in Meyer and Rasche (1992) by introducing *elliptical noise*. We show that, if the returns of risky assets follow what we call skew-elliptical GLS distribution, all investments are ranked equivalently under expected utility theory and the mean-variance approach. This holds regardless whether there is a risk-free asset in the choice set or not. In comparison to Meyer and Rasche (1992), we require a larger set of portfolios to be ranked equivalently, namely all portfolio returns. In this respect, our result is stronger but requires the specification of elliptical noise, which is not needed in the original framework of Meyer and Rasche (1992). Our preliminary result is not in conflict with the condition of elliptical symmetry contained in Chamberlain (1983, Theorem 1) because the full investment restriction eliminates, for example, portfolios with negative wealth. The sufficient condition of elliptical symmetry is still sufficient. However, now there is a more general sufficient condition which directly calls for an analysis of the necessary condition for mean-variance utility. Our main result highlights that the skew-elliptical GLS distribution is both a sufficient and a necessary condition. We arrive at this core finding by two important insights. First, we show that switching the perspective from returns to *excess returns*, which is possible under the full investment constraint, allows us to connect the situation with risk-free asset and full investment constraint to the situation without risk-free asset and no full investment constraint. Consequently, when what we call skew-elliptical mean-standard deviation (MS) distribution used in Chamberlain (1983, Theorem 2) holds for excess returns, it implies mean-variance utility for portfolio returns. Second, we show that this distribution is equivalent to a version of our skew-elliptical GLS distribution.

This set of results highlights that introducing a full investment constraint into Chamberlain's framework with risk-free asset — in other words, focusing on portfolio returns — extends the class of distributions implying mean-variance utility by, in particular, asymmetric distributions. It allows us to emphasize that, *in the presence of a risk-free asset, conducting mean-variance analysis is justified even when asset and portfolio returns are skewed*. Many seem unaware of this issue because they tend to dismiss mean-variance analysis based on the lack of empirical evidence for elliptical return symmetry (or even

¹In Section 2.2, we discuss in more detail why this difference is important and why ignoring it has led to crucial misunderstandings in the literature.

normality) in decision problems where all wealth is invested.² Our skew-elliptical GLS representation captures the relevant family of distributions by a simple two-factor linear equation and allows risk to be broken down into two components: an elliptical variance component and a non-elliptical component which covers skewness (and higher odd cumulants) of asset returns. This simplicity (in contrast to the rather abstract definition of Chamberlain's skew-elliptical MS distribution) can help to make the conditions for mean-variance utility more accessible to academics and practitioners and thus prevent misunderstandings in future research. Furthermore, our GLS model is a valuable tool for the derivation of additional results, including more intuitive ways of obtaining results of the early mean-variance literature and the generalization of results provided in more recent research. We give some examples of such applications after the derivation and discussion of our main findings.

To round off our theoretical results, we finish our study by presenting some evidence on the empirical relevance of our skew-elliptical GLS model. To this end, we modify the testing approach proposed by Meyer and Rasche (1992) to be applicable to our specific GLS case. Roughly speaking, our skew-elliptical GLS model can be tested by checking normalized out-of-sample CAPM residuals for sphericity (a special case of ellipticality) via the uniform techniques of Liang et al. (2008). Using a sample of common stocks, well-known factor portfolios (capturing size, value, momentum, reversal and industry effects) and alternative investment vehicles (i. e., indices reflecting the performance of commodity futures and hedge funds), we find that the skew-elliptical GLS distribution cannot be rejected in the majority of considered settings.

2.2 Preliminary result

Let $X = (X_1, \dots, X_n)'$ be a $n \times 1$ vector of stochastic (primary) asset returns with finite variances and $r > 0$ the non-stochastic return on the risk-free asset, if there is one. Suppose that we have an initial budget $B > 0$ to construct a portfolio consisting of these assets. In this context, we have to decide on the amount of capital $w_i B$ to be assigned to each asset i ($i = 0, 1, \dots, n$), where $w_i \in \mathbb{R}$ is a weighting factor for asset i . The term w_0 refers to the risk-free asset, whereas the terms for the risky assets are collected in the vector $w = (w_1, \dots, w_n)'$. We allow the weights to take negative values, i. e., short sales are permitted. To preserve the budget constraint, however, the weights have to sum to unity. That is, we introduce the full investment condition $w_0 + w'e = 1$, where e is a vector of ones, such that the sum of invested capital equals B . If the full investment condition is imposed, the linear combination $P := w_0 r + w'X$ can be interpreted as the return of a portfolio of assets weighted by $(w_0, w) \in \mathbb{R}^{n+1}$. In standard portfolio optimization, we

²Grootveld and Hallerbach (1999), Kapsos et al. (2014) and Lwin et al. (2017) are examples motivating their otherwise high-quality work in such a way. Studies generally discarding mean-variance analysis in the light of non-normality are, for example, Zakamouline and Koekebakker (2009), Homm and Pigorsch (2012) and Levy and Kaplanski (2015).

choose portfolio weights by maximizing the expected utility of the portfolio return (see Ingersoll, 1987, Section 3.1). That is, we solve the optimization problem

$$\begin{aligned} & \max_{(w_0, w) \in \mathbb{R}^{n+1}} \mathbb{E}[u(w_0 r + w' X)] \\ & \text{subject to } w_0 + w' e = 1, \end{aligned} \tag{P}$$

where the utility function u represents our preferences. In the portfolio selection process, the full investment constraint ensures a well-posed optimization problem, which delivers optimal asset ratios, i. e., relative amounts of investment capital, for each asset. Consequently, all investors can utilize the results of the optimization regardless of the sum they wish to invest because they can proportionally adapt an obtained optimal portfolio to their needs.

With this focus, it is clear that theoretical justification of mean-variance analysis has to build upon portfolio *returns*. This is important and often overlooked in articles using Chamberlain (1983) to reject mean-variance analysis. By pushing the limits of an initially introduced budget constraint to infinity in the ‘early’ optimization steps, he describes the distributions of an unrestricted universe of scaled portfolios $B(w_0 r + w' X)$ for an arbitrary budget $B \in \mathbb{R}$. From a scientific perspective, there is nothing wrong with this approach. However, readers tend to misinterpret the results. Chamberlain’s findings cannot be used to reject mean-variance analysis based on skewness in empirical returns because he is not looking at portfolio returns.³

By imposing the full investment restriction, we limit the portfolio universe to practically relevant constellations and simultaneously deal with this interpretation problem because the restriction enforces a portfolio return focus in our distributional considerations. As emphasized by Meyer and Rasche (1992), when additional constraints are imposed, larger families of distributions can be sufficient to justify mean-variance analysis. In the following, we show that our constraint leads to a significant and highly relevant expansion.

Within our standard portfolio optimization setting, the following distribution will be essential for our results:

Definition (Skew-elliptical GLS distribution). A random vector $X \in \mathbb{R}^n$ is said to have a skew-elliptical generalized location-scale (GLS) distribution with constant $r \in \mathbb{R}$, if its components X_i ($i = 1, \dots, n$) can be written as

$$X_i = r + \beta_i \cdot Y + \gamma_i \cdot Z_i, \tag{2.2.1}$$

where, conditional on Y , the vector $Z = (Z_1, \dots, Z_n)'$ is spherically distributed, and Y is a real-valued random variable with $\mathbb{E}[Y] \neq 0$ and $\text{Var}[Y] = 1$. The coefficients β_i, γ_i are real numbers with $\beta_i \neq 0$ for at least one $i = 1, \dots, n$.

Meyer and Rasche (1992) introduced the GLS representation in a more general form, i. e., they did not restrict the specific distributional form of Y and Z , and showed that their

³In other words, when critics argue based on returns they are assuming that the full investment condition holds. However, in this case, they cannot use Chamberlain (1983) to back up their arguments.

GLS model leads to mean-variance utility functions for the primary assets in the choice set, but not necessarily for portfolios thereof. In order to analyze not only primary assets but also non-trivial portfolios of the primary assets, we differ from their specification by defining Z to be spherical while leaving Y untouched. Note that spherical distributions are invariant under orthogonal transformations and, per definition, all Z_i ($i = 1, \dots, n$) are identically distributed with zero mean and unit variance. We do not require $\gamma_i \geq 0$ because the distribution of the Z_i is symmetric around the origin such that Z_i and $-Z_i$ have identical distributions.⁴ Our denomination 'skew-elliptical' arises from two facts. First, the noise vector $\xi := (\gamma_1 Z_1, \dots, \gamma_n Z_n)'$ is elliptically distributed about the origin because elliptical random variables are generated by affine transformations of spherically distributed ones.⁵ Second, while the spherical component Z is symmetric, the distribution of the non-spherical component Y is not restricted. This means that a skewed Y can introduce skewness into X (and higher odd moments are influenced analogously). Fat tails of X can originate from excess kurtosis in Z or Y .

Formulation (2.2.1) has the appealing property that it can be interpreted as a market model (see Elton et al., 2007, Chapter 7), where Y and Z_i are systematic and unsystematic components driving the returns X_i of asset i . Thus, from this perspective, our distributional specification of Y and Z implies that non-zero skewness in asset returns comes primarily from market-wide shocks. A risk-free asset can be easily introduced, if $r > 0$ is interpreted as the risk-free rate of return. That is, if $\beta_i = 0$ and $\gamma_i = 0$ for one $i = 1, \dots, n$, we have the risk-free return $X_i = r$. This is important because, at first glance, it may appear that some of the following theoretical results only hold for portfolios of risky assets, when, in fact, they also hold for portfolios including the risk-free asset.

In general, any linear combination of an elliptically distributed random vector is also elliptical (see McNeil et al., 2005, Section 3.3.3). With our modification, the GLS distribution receives a similar *stability* property. That is, with elliptical noise, a linear combination of GLS random variables inherits the GLS distribution from its components. Because this important feature drives our preliminary result, we state it in a lemma.

Lemma 1. *Let $X = (X_1, \dots, X_n)'$ be a random vector satisfying the skew-elliptical GLS property (2.2.1). Then, every linear combination $P = w'X$ with $w \in \mathbb{R}^n$ subject to $w'e = 1$ belongs to the same skew-elliptical GLS family (2.2.1).*

Proof. See Appendix A. □

We are now in a position to formulate our preliminary result (sufficient condition). If we interpret the X_i as asset returns and if they satisfy (2.2.1), the portfolio returns P have a representation of the form (2.2.1) as well. In other words, portfolio building does not

⁴Furthermore, although Y and Z_i might be stochastically dependent, we still have $\text{Cov}[Z_i, Y] = 0$ because $\mathbb{E}[Z_i|Y] = 0$ (see Hunter, 1972). In other words, Y and Z_i might affect each other, but there is no linear relationship between them.

⁵In our definition, we could alternatively assume $Z = (Z_1, \dots, Z_n)'$ to be elliptically distributed about the origin and appropriately scale the coefficients γ_i to obtain the same conclusions in our following discussions. In subsequent proofs, we often use this perspective (with 'dropped' scaling argument) when verifying that a random vector follows a skew-elliptical GLS distribution.

enlarge the family of return distributions. The stability (or transmission) property has the consequence that not only are the distributions of the primary asset returns determined by mean and variance, so are those of portfolio returns. The following proposition captures this result.

Proposition 1. *If a vector X of asset returns satisfies the skew-elliptical GLS property (2.2.1), then the distribution of a portfolio return $P = w'X$ is determined by its mean and variance for every $w \in \mathbb{R}^n$ with $w'e = 1$.*

Proof. See Appendix A. □

Researchers and practitioners often use the ellipticity statement of Chamberlain (1983, Theorem 1) and the observation that empirical returns tend to be asymmetric to argue that mean-variance analysis should be abandoned. In addition to our previous discussion, Proposition 1 shows that Chamberlain (1983, Theorem 1) cannot be used in such a way. It highlights that, in portfolio selection with a risk-free asset, mean-variance analysis is justified even when asset returns and portfolio returns are skewed. We can nicely see that the set of relevant distributions significantly extends from distributions with no skewness to distributions with skewness, when the full investment condition is imposed, i. e., when we are interested in a theory for portfolio returns. Intuitively, the reason for this is that fewer portfolios have to be ranked such that the conditions for equal rankings under expected utility and mean-variance framework relax.

2.3 Main result

Chamberlain (1983, Theorem 1) provides a full characterization of the distributions that imply mean-variance utility functions for all scaled portfolios, including those with negative wealth. Since our preliminary result shows that the full investment condition crucially influences his result, i. e., enlarges the set of distributions that imply mean-variance utility functions, we now aim at the question of whether we can provide a full characterization of the distributions that imply mean-variance utility for portfolio returns.

Interestingly, it turns out that the mathematical tools for deriving an answer can be found in Chamberlain (1983, Theorem 2) because the setting of this theorem and our portfolio selection problem exhibit, in a specific way, the same degrees of freedom of distributional assumptions for portfolio returns to be determined by mean and variance. The family of distribution relevant in this theorem is defined as follows.

Definition (Skew-elliptical MS distribution). A random vector $X \in \mathbb{R}^n$ is said to have a skew-elliptical mean-standard deviation (MS) distribution, if it is a linear transformation of a random vector in which the last $n - 1$ components are spherically distributed conditional on the first component, which has an arbitrary distribution. Formally, we require

$$TX = \begin{pmatrix} m \\ S \end{pmatrix}, \quad (2.3.1)$$

where T is a non-singular matrix, m is a real-valued random variable with $\mathbb{E}[m] \neq 0$ and $S = (S_2, \dots, S_n)'$ is a random vector with $S|m$ being spherically distributed.

According to Chamberlain (1983, Theorem 2), in the absence of a risk-free asset, the distribution of every scaled portfolio is determined by its mean and variance if and only if the primary assets have a skew-elliptical MS distribution (2.3.1). Proposition 1 shows that the skew-elliptical GLS distribution (2.2.1) applied to asset returns is sufficient for mean-variance-determined portfolio returns. Our first main result is that these two distributions can be considered identical.⁶

Theorem 1. *A random vector $X \in \mathbb{R}^n$ follows the skew-elliptical MS distribution (2.3.1) if and only if it follows the skew-elliptical GLS distribution (2.2.1) with no constant.*

Proof. See Appendix A. □

Because the *skew-elliptical MS distribution* and the *skew-elliptical GLS distribution* are different representations of the same family of distributions, our GLS framework provides a concise summary of skewed distributions implying mean-variance utility functions.

With this insight, we can now turn to our core objective: the full characterization of which distributional assumptions are necessary and sufficient for portfolio returns to be determined by mean and variance. To our surprise, its derivation is quite simple. The introduction of the full investment condition in the situation with a risk-free asset and n risky assets has the consequence that the portfolio excess return is equal to the weighted sum of asset excess returns:⁷

$$P - r = w_1(X_1 - r) + \dots + w_n(X_n - r). \quad (2.3.2)$$

Otherwise, this equality would not hold. If we denote the risky asset excess returns with $\bar{X}_i := X_i - r$ and the portfolio excess return with $\bar{P} := P - r$, we can rewrite the portfolio excess return as $\bar{P} = w_1\bar{X}_1 + \dots + w_n\bar{X}_n$. Because, formally, the risk-free asset and the full investment condition ‘disappear’, this is exactly the situation analyzed in Chamberlain (1983, Theorem 2). Hence, the distribution in Chamberlain (1983, Theorem 2) applied to excess returns characterizes the distributions that imply mean-variance utility functions for portfolio returns.⁸ With the help of Theorem 1, we are now in a position to formulate our second main result.

Theorem 2. *Assume there exists at least one $i = 1, \dots, n$ such that $\mathbb{E}[X_i] \neq r$, where X_i is the i th element of the risky asset vector X . In the presence of a risk-free asset, the distribution of portfolio returns $P = w_0r + w'X$ is determined by its mean and variance*

⁶Showing the equivalence of the distributional representations is not trivial. An elliptically distributed random vector $Z \in \mathbb{R}^n$ is, in general, generated by an affine transformation of a spherically distributed random vector of the same dimension n . The vector S , however, is of dimension $n - 1$. We can still choose the arbitrary random variable m appropriately such that it covers the missing last degree of freedom and the random variable Y .

⁷For $w_0 + w_1 + \dots + w_n = 1$, we have $P - r = (w_0r + w_1X_1 + \dots + w_nX_n) - r = (w_0r + w_1X_1 + \dots + w_nX_n) - (w_0 + w_1 + \dots + w_n)r = w_1(X_1 - r) + \dots + w_n(X_n - r)$.

⁸It is not important to distinguish whether portfolio returns or portfolio excess returns are determined by mean and variance because they contain the same information: $\mathbb{E}[\bar{P}] = \mathbb{E}[P - r] = \mathbb{E}[P] - r$ and $\text{Var}[\bar{P}] = \text{Var}[P - r] = \text{Var}[P]$.

for every $(w_0, w) \in \mathbb{R}^{n+1}$ with $w_0 + w'e = 1$ if and only if the asset returns X have a skew-elliptical GLS distribution (2.2.1) with r being the risk-free rate.

Proof. See Appendix A. □

Theorem 2 contains the core message of our study: *elliptical symmetry is not necessary for mean-variance utility of portfolio returns*. In standard portfolio optimization, mean-variance analysis is justified even when returns are skewed.

To put an end to the longstanding debate on the distributions that imply mean-variance utility for portfolio returns, we require one more theoretical step. Theorem 2 presents a necessary and sufficient condition for the situation with a risk-free asset. For the situation *without a risk-free asset*, Proposition 1 shows that the same condition is sufficient. Recall that Chamberlain (1983, Theorem 2) combined with our Theorem 1 states that, in the latter situation, the skew-elliptical GLS distribution (2.2.1) is a necessary condition if we consider all scaled portfolios. Thus, the question arises whether introducing the full investment constraint also relaxes Chamberlain (1983, Theorem 2) such that a larger family of distributions justifies mean-variance analysis. As illustrated in the following, it does indeed. We assume that there is a risky asset, say $X_0 = \mathcal{R}$, such that $X_i - \mathcal{R}$ and \mathcal{R} are stochastically independent for each $i = 1, \dots, n$. If \mathcal{R} is just a constant, this condition holds vacuously true and \mathcal{R} can again be interpreted as the risk-free rate of return. The premise of this setup, which might look artificial at first glance, is that with a variable \mathcal{R} , which can be treated just as the constant risk-free rate, our proof can be based on the same chain of reasoning that we used before.⁹

Corollary 1. *Let $X_0 = \mathcal{R}$ be a risky asset such that $X_i - \mathcal{R}$ and \mathcal{R} are stochastically independent for each $i = 1, \dots, n$ in the risky asset vector X . Further, assume $\mathbb{E}[X_i] \neq \mathbb{E}[\mathcal{R}]$ for at least one $i = 1, \dots, n$. Then, in the absence of a risk-free asset, the distribution of a portfolio return $P = w_0 X_0 + w'X$ is determined by its mean and variance for every $(w_0, w) \in \mathbb{R}^{n+1}$ with $w_0 + w'e = 1$ if and only if the asset returns X_i have a skew-elliptical GLS distribution of the form*

$$X_i = \mathcal{R} + \beta_i \cdot Y + \gamma_i \cdot Z_i, \quad (2.3.3)$$

where, conditional on Y , the vector $Z = (Z_1, \dots, Z_n)'$ is spherically distributed, Y is a real-valued random variable with $\mathbb{E}[Y] \neq 0$ and $\text{Var}[Y] = 1$, and \mathcal{R} is independent of Y and Z .

Proof. See Appendix A. □

Instead of having a constant risk-free rate, the family of distributions (2.3.3) offers space for another risky return component \mathcal{R} . However, since there is no additional coefficient to choose, this enlargement is rather unimportant. We did not expect it to be significant because, if we wish to remain in a mean-variance world, introducing a real third degree

⁹Note that, because the following Corollary 1 requires \mathcal{R} to satisfy specific conditions, the necessary condition for the very general case of n risky assets and no risk-free asset remains open.

of freedom is not possible. The variance measures symmetric uncertainty such that one component of the distribution also has to be symmetric. The other component, which corresponds to the mean, allows an arbitrary distribution. Thus, the skew-elliptical GLS distribution (2.2.1) already offers the maximum degrees of freedom.

2.4 Additional results

Our analysis has produced many additional theoretical results because, based on our framework, findings of the portfolio literature can be derived and/or generalized quite easily.

First, we can alternatively derive the sufficiency part of Chamberlain (1983, Theorem 1) by realizing that our GLS model contains an elliptical version of the Sinn (1983) and Meyer (1987) LS model. If this special case holds, the distributions of all scaled portfolios and portfolio returns are determined by mean and variance. This result also helps to illustrate the distributional expansion we observe with the introduction of the full investment constraint because we are jumping from a LS model to a GLS model.

Second, we generalize the findings of Schuhmacher and Auer (2014). Most importantly, we show that under our skew-elliptical GLS model and some mild axiomatic requirements (with respect to the quantification of risk), the investment performance rankings of multi-asset portfolios are identical regardless of the choice of performance measure. This explains the so far puzzling empirical observation that the Sharpe ratio and many other popular measures tend to produce the same rank ordering of investment funds with highly non-normal return distributions (see Eling and Schuhmacher, 2007). It also highlights that switching from the Sharpe ratio to conceptionally different (typically asymmetric) alternatives – an approach practitioners often promote for skewed return environments (see Bacon, 2008) – does not necessarily improve the investment process. In other words, using a symmetric risk measure in an asymmetric (skew-elliptical GLS) world still leads to the optimal portfolio decision.

Third, by linking our GLS model to the framework of Simaan (1993a), we can adopt his efficient set calculations to show that the famous two-fund separation property of Tobin (1958) holds in our setting and to derive the corresponding tangency portfolio formula.

Finally, we illustrate that the parameters determining the distribution of portfolio returns can be identified in a very elegant fashion and derive a *generally* applicable two-step guide for use in future research. According to this guide, in the first step, we have to show that a given family of distributions is stable with respect to portfolio formation. If this is true, we can, in the second step, identify the parameters that determine the distribution of a primary asset belonging to this family by following the implicit function approach used in our proofs and in Meyer and Rasche (1992). The resulting parameters consequently determine the return distribution of the primary asset and the return distribution of portfolios.

The full discussion is located in the following subsections.

2.4.1 Alternative sufficiency proof

The sufficiency part of Chamberlain (1983, Theorem 1) can be derived by realizing that it resembles a special case of our GLS framework. Similar to the proof of Lemma 1, it can

be shown that our family of asset returns also exhibits portfolio stability for $r > 0$ and all $w \in \mathbb{R}^n$, if Y is a non-stochastic constant. In this case, (2.2.1) reduces to an elliptical version of the LS condition of Sinn (1983) and Meyer (1987), which is fulfilled by a vector $X = (X_1, \dots, X_n)'$ if its components X_i ($i = 1, \dots, n$) can be written as

$$X_i = r + \mu_i + \gamma_i \cdot Z_i, \quad (2.4.1)$$

where $\mu_i \in \mathbb{R}$ and $Z = (Z_1, \dots, Z_n)'$ is spherically distributed.¹⁰ A vector X satisfying (2.4.1) is simply an elliptical random vector such that the following statement is in line with Chamberlain (1983, Theorem 1).

Corollary 2. *Assume that the vector X of asset returns satisfies the elliptical LS property (2.4.1). Then, the distribution of a portfolio $P = w'X$ is determined by its mean and variance for every $w \in \mathbb{R}^n$.*

2.4.2 Irrelevance of performance measure choice

Schuhmacher and Auer (2014) investigate investment decisions concerning portfolios of the risk-free asset and *only one* risky asset (e. g. an investment fund) belonging to the original GLS family of Meyer and Rasche (1992). With our Lemma 1, we can expand their setting to more general portfolios consisting of the risk-free asset and n risky assets.¹¹

Schuhmacher and Auer (2014, Section 3.3) analyze the conditions under which expected utility and mean-variance efficient sets are identical. The following corollary extends their result to the n -asset case.

Corollary 3. *Assume the vector X of asset returns satisfies the skew-elliptical GLS condition (2.2.1) and let $w, \tilde{w} \in \mathbb{R}^n$. Then, we have $\mathbb{E}[u(w'X)] \geq \mathbb{E}[u(\tilde{w}'X)]$ for all non-decreasing and concave utility functions u , with at least one strict inequality, if and only if we have $\mathbb{E}[w'X] \geq \mathbb{E}[\tilde{w}'X]$ and $\text{Var}[w'X] \leq \text{Var}[\tilde{w}'X]$, with at least one strict inequality.*

Schuhmacher and Auer (2014, Section 3.4) also look at conditions under which different performance measures can be written as an increasing function of the Sharpe ratio. While Chen et al. (2011) show that many measures are a monotonic function of the Sharpe ratio when asset returns are elliptical, Schuhmacher and Auer (2014) introduce skewness via the original GLS family and prove monotonicity under such distributions. A performance measure R of a portfolio P is defined as the ratio of expected excess return to investment risk captured by a risk measure ρ . Thus, we have $R(\bar{P}) := \mathbb{E}[\bar{P}]/\rho(\bar{P})$. The Sharpe ratio S , as the most famous performance measure, sets ρ to the standard deviation of (excess) portfolio returns. Other popular measures use, for example, lower partial moments, the value-at-risk or the expected shortfall. Considering two properties of meaningful risk measures (positive homogeneity and risk consistency; as in Schuhmacher and Auer, 2014), we

¹⁰Note that the parameter $r \in \mathbb{R}$ is actually redundant, because, in contrast to our GLS framework, it does not enlarge the family of distributions in the LS framework. We merely use it as a reference to the risk-free rate.

¹¹In other words, the following corollaries are a consequence of our Lemma 1 and the results of Schuhmacher and Auer (2014). Put differently, we could also say that they follow from our Proposition 1.

can extend their result in the following corollary. Positive homogeneity means that the risk of a portfolio is proportional to its size, i. e., we have $\rho(\varphi\bar{P}) = \varphi\rho(\bar{P})$ for every $\varphi \geq 0$ (see Artzner et al., 1999). Further, ρ must be consistent with a specific order relation. This risk consistency implies $\rho(\bar{P}_1) > \rho(\bar{P}_2)$ for two portfolio excess returns \bar{P}_1 and \bar{P}_2 , if \bar{P}_1 is obtained from a random variable by a higher mean-preserving spread than \bar{P}_2 . Formally, the portfolios can be written as $\bar{P}_1 = Y + b_1 Z_1$ and $\bar{P}_2 = Y + b_2 Z_2$ for some random variables Y, Z_1, Z_2 and real numbers $b_1 > b_2$, where Z_1 and Z_2 are identically distributed with $\mathbb{E}[Z_i|Y] = 0$ (see also Bigelow, 1993; Hadar and Russell, 1971; Hanoch and Levy, 1969; Ortobelli et al., 2005; Rothschild and Stiglitz, 1970).

Corollary 4. *Assume the vector X of asset returns satisfies the skew-elliptical GLS condition (2.2.1). Further, let ρ be any risk measure satisfying positive homogeneity and denote by R the corresponding performance measure. Then, we have:*

- (i) *If $S(w'\bar{X}) = S(\tilde{w}'\bar{X})$ for some $w, \tilde{w} \in \mathbb{R}^n$, this implies $R(w'\bar{X}) = R(\tilde{w}'\bar{X})$.*
- (ii) *Assume ρ additionally fulfills risk consistency. Then, if we have $S(w'\bar{X}) > S(\tilde{w}'\bar{X})$ for some $w, \tilde{w} \in \mathbb{R}^n$, this implies $R(w'\bar{X}) > R(\tilde{w}'\bar{X})$.*

This shows that for all risk measures satisfying positive homogeneity and risk consistency, the corresponding performance measure is a strictly increasing function in the Sharpe ratio, if the skew-elliptical GLS condition holds. A portfolio has the highest Sharpe ratio if and only if it has the highest value of all other admissible performance measures. Simply put, the efficient sets corresponding to different performance measures are identical.

Corollary 4 is important because practitioners often reject the Sharpe ratio, arguing that empirical return skewness makes other performance measures more suitable. Our results show that, even when returns are skewed, the Sharpe ratio and many alternative performance measures may produce identical rankings of portfolio alternatives.¹²

2.4.3 Two-fund separation

An important implication of mean-variance preferences is a separation property. That is, in the mean-variance framework with risk-free asset, Tobin (1958) found that an investor can find his/her optimal portfolio by first identifying the ideal portfolio of risky assets and then combining this portfolio with the risk-free asset. Ross (1978) completely characterized the class of distributions that imply such two-fund separability for risk-averse investors.

We can easily derive the efficient set and the two-fund separation property for our setting by using the fact that our skew-elliptical GLS distribution is linked to the class of distributions analyzed in Simaan (1993a). In our notation, his class can be stated as

$$X_i = \mu_i + \beta_i \cdot Y + \gamma_i \cdot Z_i, \quad (2.4.2)$$

where, conditional on Y , $Z = (Z_1, \dots, Z_n)'$ is spherically distributed, and Y is a real-valued random variable with $\mathbb{E}[Y] = 0$ and $\text{Var}[Y] = 1$. Again, $\xi := (\gamma_1 Z_1, \dots, \gamma_n Z_n)'$ is elliptically

¹²This describes the population. Empirical rankings are occasionally different because of sampling error.

distributed about the origin, i. e., $\xi \sim E(0, \Sigma, \psi)$ where Σ is the covariance matrix and ψ the characteristic function.¹³

In comparison to (2.2.1), (2.4.2) has an additional degree of freedom because it allows $\mathbb{E}[X_i] = \mu_i$ to be independent of β_i . Consequently, in light of our Theorem 2, under (2.4.2) portfolio returns $P = w_0 r + w'X$ with $(w_0, w) \in \mathbb{R}^{n+1}$ and $w_0 + w'e = 1$ lose the property of being determined by mean and variance only. However, in line with intuition, these changes lead to a setting where portfolio return distributions depend on three parameters. As stated in Simaan (1993a, Theorem 3), the distribution of any portfolio return is determined by its mean, variance and skewness, if (2.4.2) holds.¹⁴ More precisely, the distribution of P is a function of the mean parameter $w'\mu$, the non-sphericity and skewness parameter $w'\beta$ and the elliptical variance parameter $w'\Sigma w$. In empirical portfolio selection, such a setting has the effect that an additional parameter type has to be estimated, which adds estimation error to the one already documented for mean and (co)variance (see, for example, Best and Grauer, 1991; Britten-Jones, 1999).¹⁵

Condition (2.4.2) reduces to (2.2.1) with $r = 0$ if $\mu_i/\beta_i =: c$ is constant for $i = 1, \dots, n$. We then have $X_i = \beta_i \cdot (c + Y) + \gamma_i \cdot Z_i$. In this case, the mean and skewness parameters of a portfolio P provide the same information: $w'\mu = \sum_{i=1}^n w_i \mu_i = c \sum_{i=1}^n w_i \beta_i = c \cdot w'\beta$. Hence, the mean/skewness parameter $w'\beta$ and the elliptical variance parameter $w'\Sigma w$ are enough to determine the distribution of portfolio returns. By reintroducing a risk-free asset (i. e. simply adding it to the class again) we obtain a special case of the Simaan (1993a) distribution which allows us to model *excess returns* instead of returns.¹⁶ With this focus and in the presence of the full investment condition, we can adopt the efficient set calculations of Simaan (1993a) under (2.4.2) to derive the efficient set of our portfolio selection problem. Adding a constant does not render his general line of reasoning invalid. The proof of Corollary 5, which derives our efficient set and tangency properties, outlines this issue in detail.

¹³The multivariate extended skew-normal and closed extended skew-normal distributions can be represented via (2.4.2). Adcock (2014) shows that, with adequately chosen parameters, this also holds for the Student- t extensions of these distributions. Furthermore, McNeil et al. (2005, pp. 77 et sqq.) illustrate that random variables following the generalized hyperbolic distribution can be written as $\mu_i + \beta_i W + \sqrt{W} \tilde{Z}_i$, where $\tilde{Z} \sim N(0, \Sigma)$, $W \geq 0$ has a generalized inverse Gaussian distribution independent of \tilde{Z} , and $\sqrt{W} \tilde{Z}|W$ is elliptically distributed.

¹⁴Simaan (1993a, Theorem 3) considers no risk-free asset and no full investment constraint. Thus, similar to our use of Chamberlain (1983, Theorem 2) in our setting, we can rely on Simaan's result even after adding the risk-free rate and the full investment condition. We only have to slightly modify his arguments.

¹⁵For a further discussion of Simaan (1993a) type models (e. g. variants considering right-hand variables with non-zero mean), see Adcock (2014, Section 6).

¹⁶This is why the tangency portfolio in the following corollary slightly differs from the one stated in Simaan (1993a, Theorem 5), i. e., we have $f = \frac{V^{-1}\beta}{e'V^{-1}\beta}$ instead of $f = \frac{V^{-1}(\beta-re)}{e'V^{-1}(\beta-re)}$.

Corollary 5. *Consider the standard portfolio problem for a risk-averse agent, i. e., any concave non-decreasing utility function u . In the presence of a risk-free asset, solve*

$$\begin{aligned} \max_{(w_0, w) \in \mathbb{R}^{n+1}} \quad & \mathbb{E}[u(w_0 r + w'X)] \\ \text{subject to} \quad & w_0 + w'e = 1. \end{aligned} \tag{P}$$

Assume that the vector X of asset returns satisfies the skew-elliptical GLS condition (2.2.1) with risk-free rate $r > 0$. Then, the efficient set is spanned by the risk-free asset and the tangency portfolio

$$f = \frac{V^{-1}\beta}{e'V^{-1}\beta},$$

where $V = \Sigma + \beta\beta'$ is the covariance matrix of X .

Proof. See Appendix A. □

We have two-fund separation such that any investor's optimal portfolio can be constructed by holding each of two funds (risk-free and tangency) in appropriate ratios.

2.4.4 Identifying the parameters determining portfolio return distributions

Looking at the link between (2.4.2) and (2.2.1) from a different angle, we can obtain the result of Simaan (1993a, Theorem 3) more easily than the author by following the same reasoning we used to derive Proposition 1. For $w \in \mathbb{R}^n$, we have $P = w'X = \sum_{i=1}^n w_i\mu_i + \sum_{i=1}^n w_i\beta_i Y + \sum_{i=1}^n w_i\gamma_i Z_i = \tilde{\mu} + \tilde{\beta}Y + \tilde{Z}$, where $\tilde{\mu} := \sum_{i=1}^n w_i\mu_i$, $\tilde{\beta} := \sum_{i=1}^n w_i\beta_i$ and $\tilde{Z} := \sum_{i=1}^n w_i\gamma_i Z_i$. The random variable \tilde{Z} is elliptically distributed about the origin with the same characteristic function (see proof of Lemma 1), such that portfolio P belongs to the same family of distributions as the asset returns, namely (2.4.2). Similar to our setting, the distribution family of asset returns is stable under linear combinations. Consequently, it is sufficient to show that the distribution of an asset X_i fulfilling (2.4.2) is determined by its mean, variance and skewness. That is, we only have to verify whether the coefficients μ_i , β_i and γ_i are determined uniquely by $\mathbb{E}[X_i]$, $\text{Var}[X_i]$ and skewness $\nu[X_i]$ because, without loss of generality, all Z_i are identically distributed with unit variance, and Y has unit variance. We have $\mathbb{E}[X_i] = \mu_i$, $\text{Var}[X_i] = \beta_i^2 + \gamma_i^2$ and $\nu[X_i] = \mathbb{E}[X_i^3] = \beta_i^3 \mathbb{E}[Y^3]$, which we can solve for μ_i , β_i and $|\gamma_i|$.¹⁷ Thus, the distribution of any portfolio which combines assets satisfying (2.4.2) is determined by its mean, variance and skewness.

It is not a coincidence that we can derive results of studies that do and do not cover higher moments based on the same reasoning. As soon as a multivariate family of distributions is stable in portfolio formation, i. e., it has a transmission property ensuring that a linear combination of random variables belonging to the family is part of the same family, we simply have to determine the parameters determining the distributions of the

¹⁷Again, we do not need to know the sign of γ_i because the distribution of Z_i is symmetric about the origin.

original variables and automatically obtain the parameters determining the distribution of the combination. This important observation leads to a simple concept or guideline to identify parameters determining the distribution of portfolio returns (and to construct stable families).

Remark. We propose two steps to derive the parameters that determine portfolio returns. In a first step, we have to show that the considered family of distributions is stable with respect to portfolio building. If this is not the case, the portfolio returns cannot be determined by a specific set of parameters. The second step consists of identifying the parameters that determine the distribution of an asset belonging to the considered family. For this, we can follow Meyer and Rasche (1992) by utilizing the implicit function theorem.

2.5 Empirical perspective

Even though the focus of our study is theoretical in nature, we shed some light on whether our skew-elliptical GLS distribution has empirical support. Meyer and Rasche (1992) propose an intuitive testing procedure for their original GLS model which uses the CAPM to fill it with life, i. e., to establish a link to empirically observable variables. We adopt their general idea but, at a certain point, have to deviate from it because our GLS model is more specific.

According to the CAPM the random return of an asset or portfolio i is given by $X_i = r + \beta_i(X_m - r) + \xi_i$, where X_m is the random return of the market portfolio, β_i governs systematic risk and ξ_i represents the random idiosyncratic component of the return. In this model, X_m and ξ_i are independent, and the ξ_i are independent of one another. If the ξ_i follow the form $\xi_i = d + \gamma_i Z_i$, then the return of i becomes $X_i = (r + d) + \beta_i(X_m - r) + \gamma_i Z_i$, which is of the original GLS form. Consequently, to obtain empirical evidence in favor of the original GLS condition in a sample of N ($i = 1, \dots, N$) assets with T ($t = 1, \dots, T$) returns each, we should focus on the properties of $Z_{it} = [X_{it} - (r + d) - \beta_i(X_{mt} - r)]/\gamma_i$.

To obtain empirical observations of Z_{it} , Meyer and Rasche (1992) use a recent 30-year sample of stocks with continuous NASDAQ trading history. In our study, such a selection yields monthly total return data from January 1989 to December 2018 for 503 stocks. The market portfolio is proxied by the CRSP value-weighted index and the risk-free rate is captured by the one-month Treasury bill rate (from Ibbotson Associates).¹⁸ The stock returns are obtained from Thomson Reuters Datastream whereas the two remaining variables are taken from Kenneth R. French's data library.¹⁹ The data is split into two subsamples, A and B , of the same number of return observations for each stock. Subsample A serves to construct values for the parameters determining the Z_{it} . That is, the excess returns for each stock are regressed against the contemporary excess return of the market. No intercept is included in the regression equation. The regression slope coefficient for each stock is used as its β_i . Next, the β_i are used to obtain the $\xi_{it} = (X_{it} - r) - \beta_i(X_{mt} - r)$.

¹⁸Following Eling and Schuhmacher (2007), we have used several alternative ways of modeling the risk-free rate (e. g., initial fixes or diverse averages). None of them affected our main conclusion.

¹⁹See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

The average of ξ_{it} over all i and t gives d , whereas the standard deviation of ξ_{it} for each i delivers γ_i . In subsample B , these parameter values then allow the calculation of the Z_{it} .

Because this procedure only looks at individual stocks, Meyer and Rasche (1992) employ a vector sampling procedure to construct portfolios. In contrast to pure random sampling, this method selects blocks of stocks from a beta-sorted stock vector such that the resulting portfolios do not overlap (allowing independent idiosyncratic risk terms) and provide diversity with respect to β_i . Within each portfolio, the stocks are weighted equally (as often in practice; see DeMiguel et al., 2009) and the portfolio Z_{it} are determined similar to individual stocks. To illustrate typical empirical test results, we use $K = 3, 5, 7$ portfolios of $S = 2, 4, 6, 8$ stocks in each, which are common sizes for individual investors (see Barasinska et al., 2012; Kumar and Lim, 2008; Statman, 1987).²⁰

To test whether there is empirical support for their GLS specification, Meyer and Rasche (1992) analyze whether empirical realizations of Z_{it} are likely to have resulted as independent samples of the same population and whether their mean is significantly different from zero or not. Specifically, they use Kolmogorov-Smirnov (KS) multisample tests with simulated probability distribution (see Gardner et al., 1980) and classic t-tests for this purpose.²¹ Implicitly testing a *joint hypothesis* that the GLS condition holds across portfolios and that its variables and parameters are adequately chosen, they cannot reject their model in almost all settings they consider. Because we closely follow Meyer and Rasche (1992), our testing procedure will have a similar joint nature. In other words, we build a *conservative* setup which is not tilted in favor of the GLS model because the additional components of the joint hypothesis (such as our potentially suboptimal market portfolio choice for the common factor) make our tests more likely to reject the GLS distribution.²²

Because our GLS requirements are more restrictive with respect to the model constant (risk-free rate) and the noise component (elliptical), we drop d and cannot apply the KS test. At first glance, we might jump to established tests for elliptical symmetry, like the ones of Manzotti et al. (2002), Schott (2002) or Huffer and Park (2007), and apply them to the ξ_{it} of subsample B .²³ However, testing for spherical symmetry of the Z_{it} is more suitable because, apart from several technical advantages (related to scaling parameter measurement), it takes into account that the mean of the noise has to be zero. In general, there are many statistics for this purpose. However, most of them either converge slowly

²⁰For full transparency, the results for other K and S are available from the authors upon request. Also note that individual investors can, of course, also form portfolios based on characteristics other than betas (see Green et al., 2017). However, considering them all is beyond the scope of our study.

²¹For alternative testing procedures (like the Cramer-von Mises test), which can also be used in the context of the original GLS model, see Vassalos et al. (2012).

²²Also note that, by relying on the CAPM framework, we assume that Y and Z are independent, which is not required in our GLS world. While this simplifies the testing procedure (because conditional and unconditional distributions become identical), it introduces an additional hurdle for empirical data to pass our test.

²³More recently proposed ellipticity tests have not found wide acceptance because their implementation is computationally too expensive for large numbers of dimensions and data points (see Paoletta, 2019, Section C.2.4).

to their limiting distributions or cannot be easily evaluated in numerical computation. We use some of the tests recently outlined and extended in Liang et al. (2008, 2019) because they do not suffer from such shortcomings.

According to Liang et al. (2008), the null hypothesis that the cumulative distribution function underlying an independently identically distributed K -dimensional sample $\{z_{tj}\}_{t=1,\dots,T,j=1,\dots,K}$ is spherical can be tested via univariate or multivariate uniform statistics calculated with observations $\{v_{tj}\}_{t=1,\dots,T,j=1,\dots,K-1}$ originating from well-defined transformations of the sample data.²⁴ This is because a test for spherical symmetry can be substituted by a test of the null hypothesis H_0^1 : *the v_{tj} are uniformly distributed in $(0, 1)$* or a test of the null hypothesis H_0^2 : *the $v_t = (v_{t1}, \dots, v_{t,K-1})'$ are uniformly distributed in the hypercube $[0, 1]^{K-1}$* . Based on extensive simulations, Quesenberry and Miller Jr. (1977) and Miller Jr. and Quesenberry (1979) recommend the (modified) Watson (1961, 1962) statistic or the Neyman (1937) smooth statistic with fourth-degree polynomials for testing univariate uniformity. Critical values for these tests have been provided by Stephens (1970) and Miller Jr. and Quesenberry (1979), respectively. As far as multivariate uniformity is concerned, Liang et al. (2001) propose two statistics: one asymptotically normal and one asymptotically chi-square with two degrees of freedom. Each of them can be computed using three different measures of discrepancy: symmetric, centered and star (see Hickernell, 1998). For all tests, high statistics (in absolute terms) indicate evidence of non-uniformity and thus non-sphericity. In the following, we implement the (modified) Watson (1961, 1962) test and the normal form of the Liang et al. (2001) test because, in comparison to the two alternatives, they are characterized by better size and power under many known spherical distributions.

Table 2.1 presents our test results. We find that the null hypothesis of sphericity – and thus our GLS model – cannot be rejected in the majority of cases. Given that our approach is conservative and yet the number of rejections is low, this is strong evidence that our distribution model is relevant. As expected, the few observable rejections mainly occur for less diversified portfolios with $S = 2$.²⁵ There are also some rejections for more diversified portfolios with $S = 8$. Similar to the empirical findings of Meyer and Rasche (1992) there is no clear-cut relationship between rejection and portfolio size. That is, test statistics do not monotonically fall with a rising number of assets within the portfolios. However, they tend to be somewhat higher when more portfolios enter the test procedure. All this also holds in other sampling exercises.²⁶

To analyze larger portfolios (with different weighting schemes) and to go beyond our initial (limited) stock selection, we also test whether well-known portfolios of French's

²⁴For technical details on these transformations and the formal reasons for the reduction from K to $K - 1$, see Liang et al. (2008, pp. 684-685).

²⁵However, in line with earlier evidence (see McNeil et al., 2005, Section 3.3.5), very small portfolios and even single stocks are not generally in conflict with spherical symmetry.

²⁶In addition to vector sampling, we implemented random sampling similar to Paoletta (2019, Section C.2.4). That is, we randomly drew $S \cdot K$ of our NASDAQ stocks, formed equal-weighted portfolios, calculated our test statistics and repeated this 1,000 times. With few exceptions, density plots of the statistics nicely matched the null distributions of our tests. Performing this procedure for industry subsets (given by the Worldscope general industry classification) delivered similar results.

	Watson	Liang		
		Symmetric	Centered	Star
3 portfolios				
2 stocks	0.35	0.03	4.30	2.17
4 stocks	0.08	0.16	0.42	0.67
6 stocks	0.03	0.38	0.47	-0.07
8 stocks	0.05	-1.05	0.31	0.27
5 portfolios				
2 stocks	0.07	1.50	0.62	1.12
4 stocks	0.05	-0.91	0.68	0.25
6 stocks	0.04	0.34	-0.60	0.29
8 stocks	0.31	-2.65	2.37	-3.41
7 portfolios				
2 stocks	0.74	1.83	0.89	5.86
4 stocks	0.10	1.65	0.96	2.91
6 stocks	0.13	1.34	-0.05	-0.89
8 stocks	0.19	0.15	1.73	-1.53

Following the procedure outlined in the main text, this table tests the null hypothesis of sphericity against the alternative of non-sphericity within a universe of NASDAQ stocks. For several sets of portfolios with specific sizes, we obtain CAPM-filtered normalized noise, transform the noise as suggested by Liang et al. (2008) and report the uniform statistics of Watson (1961, 1962) and Liang et al. (2001), where the latter is calculated in its normal version. A significant statistic implies that the normalized noise may be considered non-spherical and the return data is in conflict with our GLS model specification. Significance at the 1 % level is highlighted in bold print.

TABLE 2.1: NASDAQ stock returns

library, which are typically used as benchmarks in asset pricing studies and have high relevance in institutional trading (see Ang, 2014, Chapter 10), are in line with our GLS model. That is, we perform additional tests for 6 size/book-to-market, 6 size/momentum, 6 size/reversal as well as 5 and 10 industry portfolios (defined and used in, for example, Abhakorn et al., 2013; Fama and French, 2012). Furthermore, we look at two additional asset classes by considering the 5 futures-based subindices of the Goldman Sachs Commodity Index (GSCI) and the 13 strategy subindices of the Credit Suisse Hedge Fund Index (CSHFI).²⁷ While the stock portfolios and the commodity indices have the same sample size as our previous analysis, the availability of the hedge fund data is limited to the period from January 1994 to December 2018.

Table 2.2 presents our additional test results. Based on the Watson statistic, our stock, commodity and hedge fund data is not in conflict with a skew-elliptical GLS world. The test statistics are insignificant for all data sets. Turning to the different specifications of the normal Liang test, we see a slightly different picture. While there are still many instances of non-rejection, the hedge fund results may appear less fortunate. However, this is only a slight bump in the otherwise supportive nature of our empirical results because investors typically trade individual hedge funds which are more GLS-compatible (see Schuhmacher, 2012). Furthermore, we must keep in mind that we are testing a joint hypothesis and that rejection may simply stem from the fact that, instead of the market portfolio proxy, a more suitable variable capturing common variation of returns is needed.

Because our conclusions may be sensitive to some of the settings in our research design, we have conducted several robustness checks with respect to a potential survivorship bias,

²⁷The GSCI data is available in Thomson Reuters Datastream. The CSHFI data can be obtained via <https://lab.credit-suisse.com>.

	Watson	Liang		
		Symmetric	Centered	Star
Additional stock portfolios				
FF 6 size/book-to-market	0.21	5.55	2.98	-0.18
FF 6 size/momentum	0.11	4.53	1.72	0.70
FF 6 size/reversal	0.21	1.39	1.72	-1.99
FF 5 industries	0.04	1.39	1.22	-1.32
FF 10 industries	0.05	1.11	2.60	-1.60
Additional asset classes				
GS 5 commodity indices	0.03	1.40	1.90	0.11
CS 13 hedge fund indices	0.23	8.78	8.66	4.16

This table repeats the tests of Table 2.1 for additional data sets. Specifically, we use the well-known Fama-French (FF) size, book-to-market, momentum, reversal and industry portfolios. Furthermore, we look at the subindices of the Goldman Sachs (GS) commodity futures index and the Credit Suisse (CS) hedge fund index. Again, significance at the 1 % level is marked bold and implies rejection of our GLS specification.

TABLE 2.2: Fama-French portfolios and additional asset classes

the choice of sample split in the filter procedure and the selected common factor in the case of commodity and hedge fund data. The results, presented in Appendix A, provide additional support for the GLS model and the practical relevance of our work.

2.6 Conclusion

Almost 70 years after Markowitz’s groundbreaking work on mean-variance portfolio theory and decades of ambiguity over whether its application is justified when returns are skewed, we build on the important contributions of Chamberlain (1983) and Meyer and Rasche (1992) to offer a resolution: if there is a risk-free asset and we introduce a full investment condition, i. e., focus on the distributions of portfolio returns, then mean-variance analysis is justified even when asset and portfolio returns are skewed. This cuts across the widespread belief that mean-variance analysis should be abandoned in the case of asymmetric return distributions. Using the symmetric risk measure variance in an asymmetric world leads to the same optimal portfolio decisions as popular asymmetric risk measures.

The key to this result is the discovery that, if the skew-elliptical MS distribution used in Chamberlain (1983, Theorem 2) holds for excess asset returns, it completely characterizes the distributions that imply mean-variance utility for portfolio returns. With the additional insight that this class of distributions is identical to an extension of the GLS family of Meyer and Rasche (1992), which we obtain by specifying elliptical noise, we can provide a concise representation of the distributions that justify mean-variance analysis. The resulting factor-model-like framework is not only easy to interpret but also allows the findings of previous studies to be reproduced and generalized in a quick and elegant fashion. In some empirical tests, we additionally show that its validity cannot be rejected in several data sets containing, for example, portfolios typically held by private investors or portfolios popular in asset pricing tests.

While we have imposed a wealth restriction, another common assumption in portfolio selection bans short sales (see Agarwal and Naik, 2004; Board and Sutcliffe, 1994; Grundy et al., 2012; Jagannathan and Ma, 2003), i. e., requires all portfolio weights to be non-negative. Meyer and Rasche (1992) emphasize that the necessity of elliptical symmetry in

Chamberlain (1983) is related to the fact that, for each portfolio, the mirror counterpart likewise has to be appropriately evaluated. Hence, future theoretical research might examine whether banning short sales also enlarges the family of distributions which are sufficient for portfolio return distributions to be determined by mean and variance. To derive specific results, one might think of modifying the assumption that the noise in our GLS model is elliptically distributed. A closer look at the stability discussion in Cass and Stiglitz (1970) and Owen and Rabinovitch (1983) may be a suitable starting point for such an endeavor.²⁸ A separate analysis of self-financing portfolios, where the sum of the investment weights is zero (see Korkie and Turtle, 2002), is similarly interesting. As far as additional empirical work is concerned, Schuhmacher and Auer (2014) provide some thoughts on how we might find or construct a common GLS factor which is not a market portfolio. Consequently, our empirical setup should be considered as the starting point of a full-scale empirical analysis of the skew-elliptical GLS model. Given the considerable flexibility of the model, in that we can set any distribution for the common factor, Monte Carlo studies may also, in a wide variety of distributional settings, compare the size and power characteristics of the many alternative methods (identifiable in the references of Liang et al., 2008) we have to test our GLS model.²⁹

²⁸One may also study Genton and Loperfido (2005) and Shushi (2016, 2018).

²⁹Test extensions in the spirit of Bai (2003) and Li and Tkacz (2011) may also have their merits.

Chapter 3

Multivariate electricity price forecasting

In short-term forecasting of day-ahead electricity prices, incorporating intraday dependencies is vital for accurate predictions. However, it quickly leads to dimensionality problems, i.e., ill-defined models with too many parameters. In an application for the German/Austrian market, we address this issue by deriving variable importance scores from a random forest algorithm and arrive at several interesting insights. First, we can develop full profiles stating which hours of which past days have the highest predictive power for specific hours in the future. Second, in a comparison to other approaches (such as dynamic factor models, penalized regressions or Bayesian shrinkage) that are commonly used to resolve dimensionality problems, feeding the identified variables into a support vector machine leads to a promising forecasting technique. Finally, our variable importance profiles provide a possible explanation why some forecasting methods are more accurate for certain hours of the day than others. The profiles can therefore serve as a guide for model selection. In addition, they help to explain why simple forecast combination schemes tend to outperform the full battery of models considered in our comprehensive comparative study.

3.1 Introduction

Over the past decades, electricity markets worldwide have undergone rapid deregulation, creating a more competitive market environment characterized by more volatile prices in auction-based day-ahead electricity markets. Not surprisingly, this development has (i) made energy risk management a primary concern for electricity producers, traders and consumers, and (ii) elevated price forecasts to a key input in decision making processes (see Bunn, 2004; Eydeland and Wolyniec, 2003; Weron, 2006). Electricity producers and retailers, for example, have to predict market-clearing prices to optimally self-schedule, derive bidding strategies in auctions and engage in profitable bilateral contracts (see Conejo et al., 2005). Motivated by the practical need for accurate price forecasts, scientific research has brought forth a wide variety of electricity price forecasting (EPF) models. Aggarwal et al. (2009), Weron (2014) and Nowotarski and Weron (2018) provide excellent reviews and evaluations of the methods which have been proposed so far.

In many day-ahead markets, agents submit their bids and offers for electricity, delivered during specific hours of the next day, before closing time at noon. Thus, when modeling and forecasting hourly electricity prices, we have to consider that the prices for all contracts of the next day are determined at the same time using the same available

Several recent studies have analyzed the potential of selected remedies. While, for example, Maciejowska and Weron (2013) use principal component analysis to condense input variables, Ludwig et al. (2015), Uniejewski et al. (2016) and Ziel and Weron (2018) opt for special cases of elastic net regressions. Joutz et al. (1995) and Panagiotelis and Smith (2008) focus on Bayesian techniques. Lago et al. (2018) suggest variable dropping via a functional analysis of variance, whereas Ludwig et al. (2015), Gonz  les et al. (2016) and P  rtoles et al. (2018) have investigated the performance of different types of regression trees. Unfortunately, these studies are based on data sets for different markets and, in addition, they tend to compare the employed methods to a few selected benchmarks only. Consequently, it is difficult to identify which one should be preferred by decision makers in specific electricity markets (see G  rtler and Paulsen, 2018). This is where we step into the picture.

With a focus on the German/Austrian market, which appears to be neglected in many studies tackling the dimensionality issue, we start by systematically analyzing the role of lagged hourly prices and weekday seasonalities in EPF. More precisely, we derive *importance rankings* which can reduce the number of available predictors by identifying the ones that do not contain substantial additional information. Because, in very high-dimensional input variable spaces, unreasonable trust in heuristics and simple filter scores often lead to variable subsets of poor quality (see Sorjamaa et al., 2007; Yoon et al., 2005), we implement a supervised machine learning algorithm called random forest. This method, proposed by Breiman (2001) and quite popular in bioinformatics (see Díaz-Uriarte and Alvarez de Andrés, 2006; Heidema et al., 2006; Strobl et al., 2007), relies on simple, yet powerful data sampling and feature splitting techniques to provide reliable importance scores within a reasonably short computation time. In our study, random forests confirm previous results that prices lagged by one week have considerably more explanatory power than directly preceding smaller lags and that prices of specific daytimes can be forecasted with greater

accuracy when lagged prices of different daytimes are used as predictors (see Ziel, 2016).¹ Extending previous research, we can present detailed variable importance profiles for hours of different daytimes which allow several interesting insights. Using the block bid definitions of the day-ahead auction, we find that, for example, evening prices of the previous day are highly relevant predictors for after-midnight prices, whereas weekday dummy variables and yesterday’s afternoon prices are very important for the determination of afternoon prices.

After the derivation of the most important predictors, we use them in a standard regression setting (statistical model) and a support vector machine (machine learning model), which has attracted much interest in the EPF literature (see Chaâbane, 2014; Naumzik and Feuerriegel, 2020; Niu et al., 2010; Yan and Chowdhury, 2010) and tends to shine in comparison to other established computational intelligence methods (such as some forms of neural networks; see Che and Wang, 2010; Sansom et al., 2003), to see how the variable selection performs in out-of-sample forecasts. At the same time, we provide a *full-scale comparison* with popular alternative forecasting approaches. That is, besides simple benchmark models (including naive predictions, unrestricted VARs and expert models), we have a closer look at other techniques commonly applied to deal with dimensionality issues. Among the alternatives we have, for example, dynamic factor models (principal component analysis followed by assigning a time series model to the identified factors) for feature extraction, as well as penalized regressions (such as the elastic net) and Bayesian shrinkage of VAR models for regularization. As a first result, we find that expert models relying on a rough variable set derived from our random forest profiles tend to outperform large unrestricted models. Second, when combining the detailed random forest feature selection with a support vector machine, we obtain lower prediction errors compared to a feed of the variables into a classic statistical model and also find that the former approach performs well relative to other remedies for dimensionality. Third, in an analysis of the prediction errors on a hourly level, we see that each daytime appears to have its ideal forecasting model. For example, elastic nets are superior models when it comes to forecasting evening prices. Our importance profiles can help us to explain such a result. They show the most relevant variables such that, based on the design of a forecasting approach, we can determine whether or not the approach includes these variables.

Finally, to round off our comparison of predictive accuracy, a last contribution of our study is motivated by a growing body of literature emphasizing that *combinations of forecasts* derived from different models are often more reliable than forecasts from a single model alone (see Baumeister and Kilian, 2015; Clemen, 1989; Timmermann, 2006). Because electricity studies have mainly focused on simple time series models so far (see Bordinon et al., 2013; Nowotarski et al., 2014), we look at combination schemes incorporating more advanced forecasting techniques. When applied to our entire model selection, they outperform each single model. This is partially because the combinations strongly benefit

¹In contrast, for many multivariate economic series, importance usually declines with higher lags and cross-dependencies are substantially less pronounced such that univariate models with lags covering the recent own past are often sufficient.

from our observation of daytime-dependent model performance. Consequently, if we are interested in accurate predictions for all hours of the day, we do not necessarily have to search for one model with this capacity. Instead, we can develop models tailored to specific daytimes and include them in combined forecasts to achieve high overall predictive power.

The remainder of our article is organized as follows: In Section 3.2, we describe our data, implement our variable importance calculations using the random forest algorithm and present an overview of competing forecasting models, including our two random forest models. In Section 3.3, we evaluate the accuracy of our price models based on their forecasting errors and selected statistical tests. Section 3.4 concludes, highlights the implications of our study and outlines directions for future research.

3.2 Data and methodology

3.2.1 Sample selection

We use European Power Exchange (EPEX) hourly day-ahead electricity prices (in EUR/MWh) covering the market area Germany/Austria from January 2013 to December 2017.² Our sample size is motivated by previous studies which favor short periods covering only the most recent (typically six, three or even fewer) years (see Lago et al., 2018; Ludwig et al., 2015; Ziel and Weron, 2018) in order to not overweight less relevant price characteristics of the very distant past. This is important because, especially in the recent years, the regulatory changes in electricity markets have caused significant structural breaks in price and volatility dynamics (see Auer, 2016). Our initial model calibrations rely on data from 2013 to 2015; the remaining years 2016 and 2017 are used to evaluate out-of-sample forecasts. For some methods, the calibration period is split up into the two-year training period 2013-2014 and the one-year validation period 2015.

3.2.2 General technical issues

For each hour h , our goal is to identify the most relevant inputs to explain the price variable $y = (P_{d,h})_{d=1,\dots,N}$. We focus on the explanatory power of past spot prices and weekday effects. Thus, we work within an almost pure multivariate time series context. As for past prices, we consider lags up to a maximum order of $p = 21$ days. Thus, we have to choose from $24 \cdot 21 = 504$ lagged price variables $(P_{d-l,k})_{d=1,\dots,N}$ resulting for the lags $l = 1, \dots, p$ and hours $k = 0, \dots, 23$. As noted by Weron (2006), electricity prices depend on the day of the week, meaning that weekends tend to exhibit price patterns different from

²The data is available from the European Energy Exchange (EEX) via <https://www.eex.com/en/market-data>. Because the loss of an hour due to daylight saving time causes some data problems, we interpolate the missing hour in March with the two hours around the missing hour. Furthermore, we use the average of the double hours in October such that we have 24 observable prices each day.

weekdays. This also holds for holidays, where prices are lower than on business days. To incorporate the weekday effect, we introduce weekday dummies

$$W_i(d) = \begin{cases} 1, & \mathcal{W}(d) = i, \\ 0, & \mathcal{W}(d) \neq i, \end{cases} \quad (3.2.1)$$

where $\mathcal{W}(d)$ is a function that gives a number which represents each day of the week ($i = 0$ for Sunday, $i = 1$ for Monday, ..., $i = 6$ for Saturday). To capture holidays, we follow Ziel et al. (2014). We treat each public holiday as a Sunday. In addition, each regional holiday that is not a Sunday is treated as a Saturday. Moreover, we take the 12 hours after a regional or public holiday as Monday-am and the 12 hours before each holiday as Friday-pm.

All possible input variables are gathered in the $(N \times J)$ regressor matrix $X = (x^j)_{j=1, \dots, J}$. Each $(N \times 1)$ column x^j represents either a weekday dummy variable $(W_i(d))_{d=1, \dots, N}$ for a day $i = 0, 1, 5, 6$ or a lagged price variable $(P_{d-l,k})_{d=1, \dots, N}$ for some lag $l = 1, \dots, p$ and hour $k = 0, \dots, 23$. The standard linear regression model is then given by

$$y = X\psi + \varepsilon \quad (3.2.2)$$

with $(J \times 1)$ coefficient vector ψ and $(N \times 1)$ error term ε .³

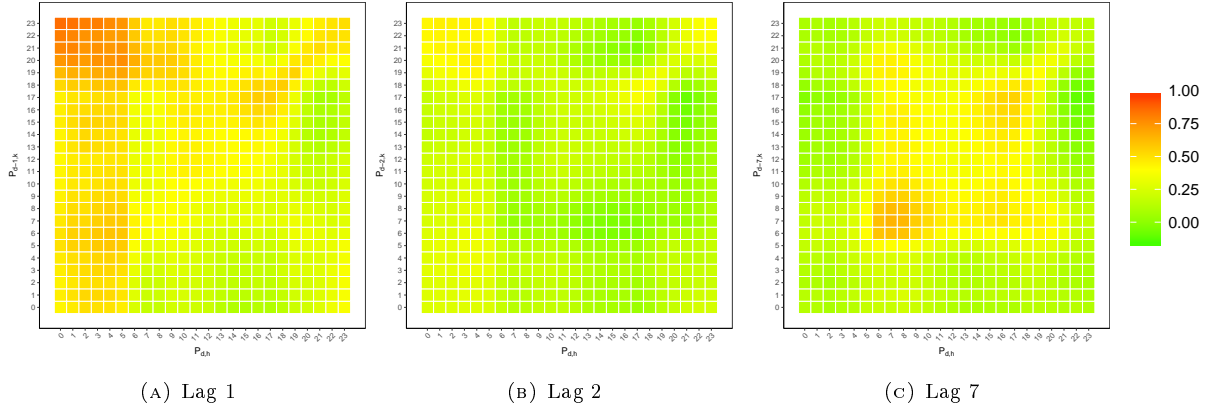
In this setup, we derive the importance profile of hourly prices with reference to year 2015. In the case of an univariate time series, relevant lag variables are traditionally selected by inspecting the empirical autocorrelation function and choosing all variables with statistically significant correlation (see Box et al., 2015). We also implement this approach as a first simple benchmark for our own random forest proposal. That is, we consider the sample correlation coefficient between variables x^j and y as an importance score I for variable x^j :

$$I(x^j) = \frac{\sum_{d=1}^N (x_d^j - \bar{x}^j)(y_d - \bar{y})}{\sqrt{\sum_{d=1}^N (x_d^j - \bar{x}^j)^2} \sqrt{\sum_{d=1}^N (y_d - \bar{y})^2}}. \quad (3.2.3)$$

An obvious deficiency of this score is that it looks at each feature separately and consequently cannot reveal mutual information. A more suitable score should assign higher importance to a few extremely relevant input features, and in return rank features lower which mainly cover information already supplied by more relevant features.

Using a heat map design, Figure 3.1 presents the correlation matrices between current and lagged hourly prices. For a one-day lag, we find an overall positive correlation between prices of adjacent days. In addition, we can observe a correlation ‘break’ between hours 05 and 06. While today’s after-midnight prices are highly correlated with yesterday’s prices (in particular with prices in the late evening hours), the remaining correlations are notably

³Note that, for notational convenience, our matrix X does not contain the typical initial vector of ones. Of course, in our empirical application, models are estimated with constant terms (see Allen and Stone, 2005).



Based on electricity market data for Germany/Austria in 2015, this figure plots heat maps for sample correlations between current spot prices $P_{d,h}$ and lagged spot prices $P_{d-l,k}$ for different hours $h, k = 0, \dots, 23$ and lags $l = 1, 2, 7$.

FIGURE 3.1: Correlations between current and lagged hourly prices

lower. While, at first glance, this appears unspectacular, recall that today's price at hour 00 is formed 24 hours after yesterday's price at hour 23. What might also come as a surprise is that high correlations on the matrix diagonal (i.e., strong relationships of specific hours to their own past) appear only at evening hours. Turning to the matrix for a seven-day lag, we can see that the highest correlations occur at daylight from hours 06 to 18. In a comparison of the seven-day with the two-day lag, we find higher correlations for the weekly lag.

The simple correlation profile suggests that hours of different daytimes have distinct importance, i.e., the explanatory power of lagged input variables appears to vary depending on daytime. In the following, we provide a more sophisticated analysis of this claim by deriving variable importance rankings based on random forests. For better visualization and to identify systematic patterns within these rankings, we use the daytime (or block) definitions of the energy exchange. In the auction forming the day-ahead spot price, agents can not only place single hour bids but also block bids for combinations of hours. Particularly popular blocks are night (00-05), morning (06-09), high noon (10-13), afternoon (14-17) and evening (18-23).⁴

3.2.3 Random forests

A random forest is an ensemble of N_{tree} decision trees, where each tree is grown using a subsample (or bootstrap sample) from the training set and choosing randomly at each node a subset of explanatory variables. In contrast to a classification and regression tree (CART) model building strategy (see Breiman et al., 1984), the random forest algorithm builds unpruned regression trees, where at each node v , features of given number m are randomly sampled and the best split is calculated only based on these variables (see Genuer

⁴There are many more block definitions including, for example, business hours (08-15).

et al., 2008). A split divides the remaining observations of a node v over two child nodes by choosing one of the m features and a threshold value. Split evaluation in regression trees is performed via a residual sum of squares criterion which defines the impurity at node v as

$$\iota(v) = \frac{1}{N_v} \sum_{n=1}^{N_v} (y_{d_n} - \bar{y})^2, \quad (3.2.4)$$

where N_v is the number of observations in node v , y_{d_n} is the response value of observation n (i.e., of day d_n), and \bar{y} is the average response of the observations in v . The reduction in impurity at node v achieved by splitting using the variable x^j is given by

$$\Delta\iota(x^j, v) = \iota(v) - \sum_c w(v_c|v) \iota(v_c) \quad (3.2.5)$$

where v_c denotes a child node and $w(v_c|v)$ is the proportion of observations in v that are assigned to v_c . At each node, a random set of m features is evaluated, and the feature x^j with the maximum $\Delta\iota(x^j, v)$ is used for splitting the node v . Binary splitting continues until the subsample (or bootstrap sample) is used up.

Random forests use their so-called out-of-bag (OOB) samples to construct a variable importance measure: the permutation importance of Breiman (2001). The OOB sample of a tree includes all observations (of the training sample) not used to grow the tree. For each feature, the OOB error captures the mean decrease in accuracy by removing the association between the feature and the target. This is achieved by randomly permuting the values of the feature and measuring the resulting increase in error. Formally, let $\mathcal{B}^{(t)}$ be the OOB sample for a tree t , with $t \in \{1, \dots, N_{tree}\}$. Then the variable importance of x^j in tree t is

$$I^{(t)}(x^j) = \frac{1}{|\mathcal{B}^{(t)}|} \sum_{d \in \mathcal{B}^{(t)}} \left(y_d - \hat{y}_d^{(t)} \right)^2 - \frac{1}{|\mathcal{B}^{(t)}|} \sum_{d \in \mathcal{B}^{(t)}} \left(y_d - \pi_j \hat{y}_d^{(t)} \right)^2, \quad (3.2.6)$$

where $\hat{y}_d^{(t)}$ and $\pi_j \hat{y}_d^{(t)}$ are the predicted values for observation d before and after permuting the values of variable x^j . We predict by simply moving up the tree. The raw importance score for each variable is computed as the mean importance over all trees:

$$I(x^j) = \frac{1}{N_{tree}} \sum_{t=1}^{N_{tree}} I^{(t)}(x^j). \quad (3.2.7)$$

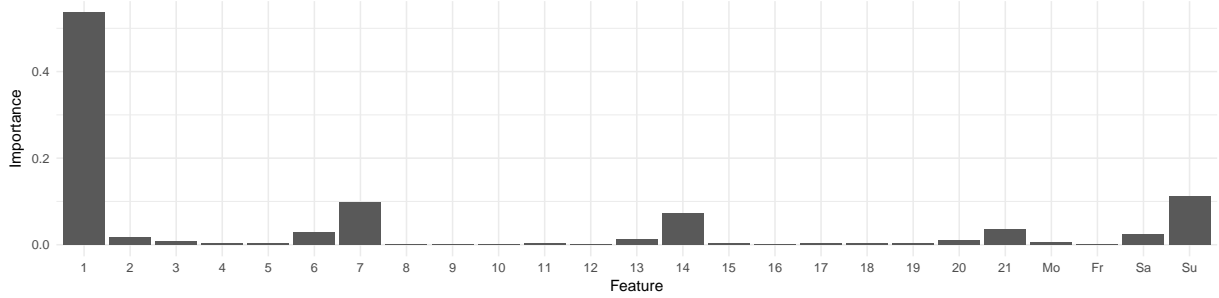
Standard implementations of random forests often provide a scaled version of the permutation importance which is obtained by dividing the raw importance by its standard error. However, because this alternative measure has statistical defects (see Strobl and Zeileis, 2008), we focus on the unscaled version.⁵ Furthermore, it is important to note that

⁵The ability of the scaled version to identify irrelevant variables does not improve with increased sample size.

the permutation importance for variables of different types (in our case, real-valued vs. dummies) is unbiased only when subsampling (instead of bootstrapping) is used to build the trees (see Strobl et al., 2007). Therefore, we use the R package **party** for implementation, set $N_{tree} = 500$ and $m = \lfloor J/3 \rfloor$, and opt for subsampling without replacement.

3.2.4 Variable importance scores

We now investigate the random forest importance scores resulting for the market data in the year 2015. To evaluate the overall importance of specific lags, we start with an aggregate perspective. That is, for each lag, we sum the corresponding scores across hours and plot the totals in Figure 3.2. Afterward we look at the detailed hourly results. Table 3.1 reports the seven most important input variables (and their scores) for each hour of the day. Figure 3.3 then presents the full importance profiles for selected hours, where, for each daytime, we pick a representative hour and restrict our attention to the most important lags $l = 1, 7$.⁶



This figure plots the relative permutation importance scores (computed via the random forest algorithm of Section 3.2.3) of lagged price variables $P_{d-l,k}$ (lags $l = 1, \dots, 21$ days) and weekday dummies W_i (days $i = 0, 1, 5, 6$) for the explanation of prices $P_{d,h}$. For better visualization, the scores have been aggregated across all hours h and k .

FIGURE 3.2: Aggregated lag importance

Our entire set of relative percentage scores supports the idea that the variables which are most important for price forecasting change over daytime. However, we can also see that, across hours, the most relevant variables belong to the same lag order. That is, the first lag exhibits the greatest explanatory power, followed by the weekly lags of 7, 14 and 21 days. Interestingly, weekly lags minus 1 (i.e., the lags 6, 13 and 20) are not negligible. Weekday effects are dominant during daylight hours, while they vanish during night and evening hours. Overall we can sketch the following profile for our block-oriented daytime definitions:

Night (00-05): Evening prices of the previous day are not only the most relevant variables; they also explain night prices almost exclusively. With importance scores between 15.1% (for hour 05) and 41.5% (for hour 00), hour 23 of the previous day stands out. Other

⁶The importance profiles for all 24 hours of the day can be found in Appendix B.

		Top seven explanatory variables								
	h	#1	#2	#3	#4	#5	#6	#7		
Night	00	$P_{d-1,23}$ (41.5)	$P_{d-1,22}$ (26.7)	$P_{d-1,21}$ (15.6)	$P_{d-1,20}$ (08.3)	$P_{d-1,19}$ (03.6)	$P_{d-1,18}$ (01.2)	$P_{d-1,17}$ (00.5)		
	01	$P_{d-1,23}$ (40.2)	$P_{d-1,22}$ (20.6)	$P_{d-1,21}$ (19.6)	$P_{d-1,20}$ (07.9)	$P_{d-1,19}$ (04.1)	$P_{d-1,18}$ (01.8)	$P_{d-1,15}$ (00.7)		
	02	$P_{d-1,23}$ (40.2)	$P_{d-1,22}$ (20.7)	$P_{d-1,21}$ (14.7)	$P_{d-1,20}$ (09.8)	$P_{d-1,19}$ (04.4)	$P_{d-20,19}$ (02.2)	$P_{d-1,18}$ (01.1)		
	03	$P_{d-1,23}$ (39.0)	$P_{d-1,22}$ (16.5)	$P_{d-1,21}$ (14.9)	$P_{d-1,20}$ (13.0)	$P_{d-1,19}$ (05.3)	$P_{d-1,18}$ (02.1)	$P_{d-1,5}$ (01.8)		
	04	$P_{d-1,23}$ (28.4)	$P_{d-1,20}$ (20.0)	$P_{d-1,21}$ (15.7)	$P_{d-1,22}$ (13.9)	$P_{d-1,19}$ (08.0)	$P_{d-1,18}$ (03.0)	$P_{d-1,5}$ (02.8)		
	05	$P_{d-1,20}$ (26.4)	$P_{d-1,21}$ (17.0)	$P_{d-1,23}$ (15.1)	$P_{d-1,22}$ (11.9)	$P_{d-1,19}$ (08.1)	W_0 (04.0)	$P_{d-1,18}$ (02.5)		
Morning	06	W_0 (23.6)	$P_{d-1,20}$ (16.9)	$P_{d-7,7}$ (08.9)	$P_{d-1,21}$ (07.6)	$P_{d-1,19}$ (04.7)	$P_{d-1,23}$ (04.2)	W_6 (03.9)		
	07	W_0 (26.2)	$P_{d-7,7}$ (12.3)	$P_{d-1,20}$ (08.4)	W_6 (06.5)	$P_{d-14,7}$ (05.5)	$P_{d-1,19}$ (04.3)	$P_{d-14,8}$ (04.0)		
	08	W_0 (30.9)	$P_{d-7,7}$ (12.6)	$P_{d-1,20}$ (09.1)	W_6 (06.6)	$P_{d-1,19}$ (04.4)	$P_{d-1,23}$ (03.9)	$P_{d-7,8}$ (03.9)		
	09	W_0 (30.4)	$P_{d-1,20}$ (09.1)	$P_{d-1,23}$ (07.6)	$P_{d-7,7}$ (07.4)	W_6 (04.3)	$P_{d-7,8}$ (04.1)	$P_{d-1,19}$ (03.7)		
High noon	10	W_0 (27.7)	$P_{d-1,23}$ (09.0)	$P_{d-7,7}$ (08.9)	$P_{d-7,8}$ (05.3)	$P_{d-1,19}$ (04.8)	$P_{d-1,20}$ (04.2)	W_6 (03.5)		
	11	W_0 (18.1)	$P_{d-1,23}$ (08.8)	$P_{d-7,7}$ (08.6)	$P_{d-1,13}$ (05.1)	$P_{d-1,16}$ (04.9)	$P_{d-7,8}$ (04.8)	$P_{d-1,15}$ (04.6)		
	12	W_0 (21.3)	$P_{d-1,23}$ (07.7)	$P_{d-7,7}$ (07.0)	$P_{d-1,16}$ (06.3)	$P_{d-1,15}$ (05.3)	$P_{d-1,13}$ (04.8)	$P_{d-7,8}$ (04.5)		
	13	W_0 (26.5)	$P_{d-7,7}$ (09.0)	$P_{d-1,16}$ (07.0)	$P_{d-1,15}$ (05.3)	$P_{d-1,17}$ (05.2)	$P_{d-1,23}$ (04.3)	$P_{d-1,13}$ (03.5)		
Afternoon	14	W_0 (25.3)	$P_{d-1,17}$ (08.4)	$P_{d-1,16}$ (08.3)	$P_{d-7,7}$ (06.8)	$P_{d-1,15}$ (05.4)	$P_{d-14,7}$ (04.3)	$P_{d-14,8}$ (03.0)		
	15	W_0 (18.3)	$P_{d-1,17}$ (13.1)	$P_{d-1,16}$ (08.2)	$P_{d-7,7}$ (05.1)	$P_{d-14,7}$ (04.9)	$P_{d-7,17}$ (04.2)	$P_{d-1,18}$ (03.9)		
	16	$P_{d-1,17}$ (16.5)	W_0 (08.7)	$P_{d-1,16}$ (08.0)	$P_{d-7,17}$ (06.7)	$P_{d-1,18}$ (05.5)	$P_{d-14,17}$ (04.2)	$P_{d-7,16}$ (04.0)		
	17	$P_{d-1,17}$ (18.9)	$P_{d-1,18}$ (08.6)	$P_{d-7,17}$ (08.4)	$P_{d-14,17}$ (06.8)	$P_{d-1,16}$ (05.1)	$P_{d-7,16}$ (03.6)	$P_{d-14,15}$ (03.1)		
Evening	18	$P_{d-1,18}$ (23.5)	$P_{d-1,19}$ (08.1)	$P_{d-1,17}$ (06.9)	$P_{d-7,17}$ (03.8)	$P_{d-7,18}$ (03.6)	$P_{d-14,17}$ (03.4)	$P_{d-14,18}$ (02.7)		
	19	$P_{d-1,19}$ (27.0)	$P_{d-1,20}$ (08.0)	$P_{d-1,18}$ (05.4)	$P_{d-6,19}$ (05.1)	$P_{d-7,7}$ (04.9)	$P_{d-14,19}$ (03.7)	$P_{d-14,9}$ (03.3)		
	20	$P_{d-1,20}$ (16.7)	$P_{d-1,21}$ (11.0)	$P_{d-21,7}$ (06.7)	$P_{d-1,19}$ (06.2)	$P_{d-7,7}$ (05.2)	$P_{d-14,20}$ (03.8)	$P_{d-14,7}$ (03.5)		
	21	$P_{d-1,21}$ (24.2)	$P_{d-1,22}$ (11.2)	$P_{d-1,20}$ (04.9)	W_6 (04.9)	$P_{d-14,21}$ (04.7)	$P_{d-21,7}$ (03.4)	$P_{d-14,20}$ (03.0)		
	22	$P_{d-1,22}$ (22.5)	$P_{d-1,23}$ (16.8)	$P_{d-1,21}$ (09.9)	W_6 (05.6)	$P_{d-2,22}$ (03.7)	$P_{d-2,23}$ (03.1)	$P_{d-14,21}$ (02.0)		
	23	$P_{d-1,23}$ (25.3)	$P_{d-1,22}$ (15.6)	$P_{d-1,21}$ (06.0)	W_6 (04.9)	$P_{d-2,22}$ (04.4)	$P_{d-1,0}$ (03.7)	$P_{d-2,23}$ (03.5)		

This table reports the most important input variables for explaining the prices $P_{d,h}$ ($h = 0, \dots, 23$). The relative permutation importance scores (obtained via the random forest algorithm of Section 3.2.3) are given in parentheses.

TABLE 3.1: Most important input variables

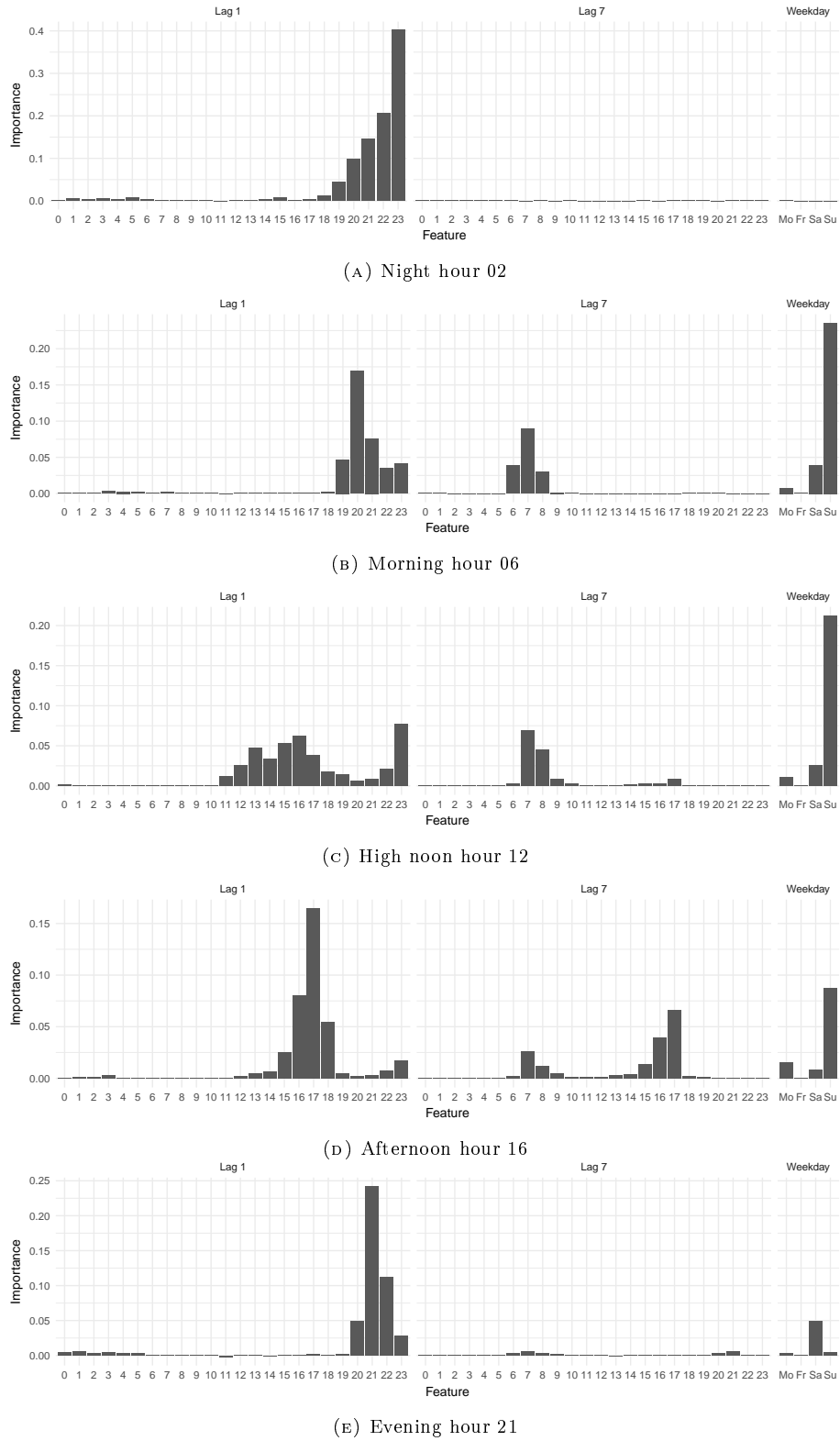
variables are barely important. For example, we have almost no weekday effect at night hours.

Morning (06-09): Last evening's prices are still important. For example, hour 20 has high scores of around 10%. Additionally, morning prices lagged by one week gain in significance, with hour 07 exhibiting scores between 7.4% and 12.6%. The weekday effect grows in strength with sunrise, where the Sunday dummy variable is the most important.

High noon (10-13): For high noon prices, many variables have significant importance scores. The Sunday dummy is the most relevant one with scores between 18.1% and 27.7%. The prices of morning hours 07 and 08 lagged by one week also have high scores. As far as one-day lags are concerned, high noon, afternoon and evening prices are relevant: hour 23 provides the highest scores between 7.7% and 9.0% for the hours 10 to 12.

Afternoon (14-17): In contrast to high noon hours, the most important one-day lags are concentrated around a few hours, with hour 17 scoring highest. Morning hour 07 and afternoon hour 17 of the previous week are significant. The Sunday effect declines with sunset.

Evening (17-23): Evening hours are the only hours for which the same hour the previous day always has the highest importance score. In comparison to afternoon hours, the explanatory power of weekly lagged morning hours wanes. Weekday effects, among which Saturday reaches higher scores than Sunday, are of little relevance.

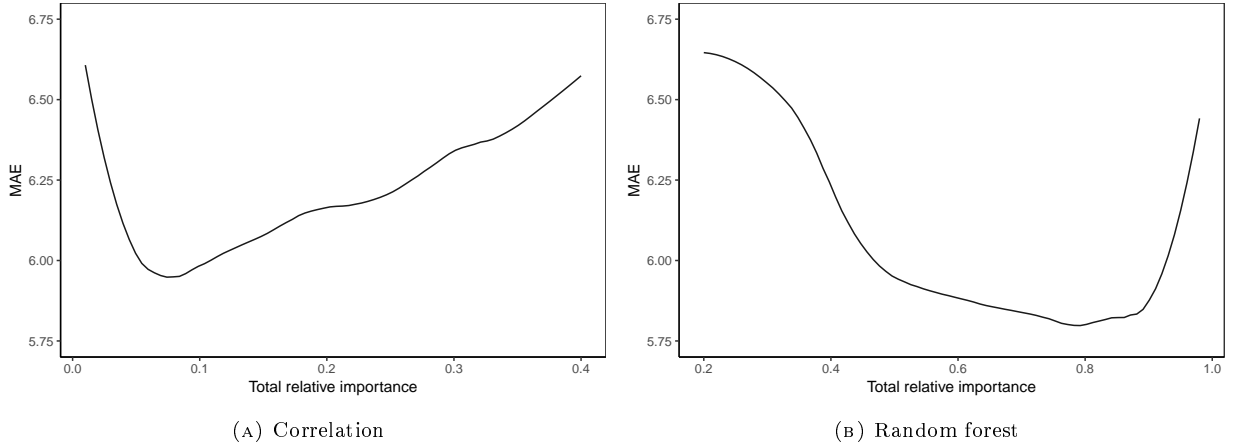


This figure illustrates the relative permutation importance scores (computed via the random forest algorithm of Section 3.2.3) of lagged price variables $P_{d-l,k}$ (lags $l = 1, 7$ days) and weekday dummies W_i (days $i = 0, 1, 5, 6$) for the explanation of some selected hourly price variables.

FIGURE 3.3: Variable importance profiles for selected hours

Before turning to our comprehensive forecasting study, it is instructive to pre-evaluate the quality of our random forest scores in comparison to the simple correlation scores of Section 3.2.2. Specifically, we use the variable rankings delivered by the two approaches for a simple feature selection procedure and investigate the resulting out-of-sample forecasting errors for the years 2016 and 2017. In both cases, we choose the most important features whose relative importance in the validation period scores sum up to a pre-set total relative importance score I_{total} . We then calculate the out-of-sample mean absolute error (MAE, as defined in Section 3.3.1) of forecasts produced by linear models which use the identified feature subsets.

In Figure 3.4, we illustrate the relationship between the choice of total importance and the corresponding forecasting errors. For the correlation benchmark, the minimum MAE of 5.95 is reached for $I_{total} = 7\%$. For the random forest algorithm, we have a lower minimum MAE of 5.77 for $I_{total} = 79\%$. For a large range of I_{total} values, random forest errors are below the best result of the correlation approach and indicate a better selection of feature subsets by the random forest method.



Using (a) correlation coefficients (see Section 3.2.2) and (b) permutation scores of random forests (see Section 3.2.3), we select the most important features whose relative importance scores sum up to a pre-set total relative score and calculate the out-of-sample mean absolute errors (MAE) of electricity price forecasts produced by linear models equipped with the identified feature subsets. The out-of-sample period is given by the years 2016 and 2017. This figure plots the pre-set total scores against the resulting MAE values.

FIGURE 3.4: Importance score evaluation

In our data, we have many variables with similarly high correlation to the dependent variable and thus similar correlation-based scores. In contrast, the random forest algorithm assigns high scores only to a few input variables, whereas many other features receive scores around 0. This makes feature selection easy because we could simply sort out features with importance scores close to 0.

3.2.5 Competing forecasting models

Our comparative analysis builds on a wide variety of popular forecasting models. We focus on the explanatory power of past spot prices, such that only pure price models

without exogenous variables (like weather or load) are considered. One exception is made for weekday dummies W_i , which are included in all price models. Because many of the models are well-known, the following discussion focuses on an overview and the specification issues relevant for our electricity data.⁷

Benchmark models. We set up three simple and three advanced benchmark models. In the first group, we start with the naive or *persistent* model which assumes that price changes are completely random such that the best forecast is an appropriate past hourly price (see Nogales et al., 2002). Furthermore, we use *24 univariate AR(p)* models (henceforth abbreviated with 24AR) which are based on centered hourly prices and estimated via Yule-Walker solving (see Ziel, 2016). The model orders p_h are determined by minimizing the Akaike information criterion (AIC) with an upper bound of $p_{\max} = 21$. Finally, we consider an *unrestricted VAR(p)* model (henceforth just VAR) which is estimated via equation-by-equation ordinary least squares (OLS) (see Raviv et al., 2015; Ziel and Weron, 2018). Again, we choose the lag order p via AIC with $p_{\max} = 21$. We obtain $p = 2$.

The second group of benchmark models consists of expert models which use a fixed parsimonious autoregressive structure based on prior knowledge of experts. Screening previous research (see Weron and Misiorek, 2008) or our random forest importance scores, a straightforward expert choice is to include all lagged price variables $P_{d-l,k}$ for the lags $l = 1, 7$ and weekday dummy variables W_i for Monday, Friday, Saturday and Sunday, i.e., $i = 0, 1, 5, 6$. Using these variables in a VAR setup leads to an *expert VAR* implementable via equation-by-equation OLS (see Maciejowska et al., 2016; Misiorek et al., 2006). Besides a standard time series model, we also feed the expert variable choice into two forms of machine learning models. First, because energy market research (including Catalão et al., 2007; Kuo and Huang, 2018; Lago et al., 2018; Szkuta et al., 1999) has paid much attention to neural networks (NNs), we construct an *expert NN*. Specifically, for each hour h , we train a (feed-forward) multi-layer perceptron with one hidden layer and tangent-sigmoid transfer function via the Levenberg-Marquardt algorithm (see Cruz et al., 2011; García-Ascanio and Maté, 2010). To derive the optimal number of hidden nodes, we consult several rules of thumb (see Blum, 1992; Boger and Guterman, 1997; Hagan et al., 1996) and train networks with a number of hidden nodes ranging from 1 to 12. As far as the network generalization error (the mean square error in the validation sample) is concerned, we find 8 hidden nodes to be a reasonable layer size. Second, because support vector machines (SVMs), depending on the market, can be more accurate than NNs (see Che and Wang, 2010; Sansom et al., 2003), we also set up an *expert SVM* in Schölkopf et al. (2000) form and with Gaussian radial basis kernel function. The optimal parameters of the SVM regression are obtained by setting reasonable parameter ranges and then selecting values via a real-value genetic algorithm (see Scrucca, 2013). In this context, the fitness function is set similar to the expert NN.

Feature extraction. While random forests concentrate on selecting the most important features, another approach to dimension reduction is to extract (or generate) new features from the set of input variables and to use these new variables instead of the original ones

⁷Full technical model descriptions can be found in Appendix B.

(see Maciejowska and Weron, 2013). To follow this idea, we assume that each hourly price can be decomposed into a small set of K common factors and an idiosyncratic component. These quantities are estimated by principal component analysis (PCA) with an optimal number of factors $K = 7$ derived from the tests of Bai and Ng (2002), Alessi et al. (2010), Onatski (2010) and Ahn and Horenstein (2013). These factors explain 97.3% of the intraday variability in our price data. To forecast future prices, we model the idiosyncratic component via an AR(q) model and the truncated factors via an unrestricted VAR. In the latter, AIC lag determination yields $p = 7$. We call the overall price forecasting approach $PC(K)$ -VAR. In addition to this VAR setting, we also apply the least absolute shrinkage operator (LASSO, to be discussed below) and a SVM (discussed above) to the factors, which leads to the $PC(K)$ -LASSO and $PC(K)$ -SVM models. In both cases, we set $p = 7$. We use $K = 7$ and $K = 11$ because the latter number of factors yields the best forecasts for PC-LASSO and PC-SVM.

Regularization. To solve ill-posed optimization problems or to prevent overfitting, regularization incorporates additional information into the estimation process. In *ridge* regression, this is an l_2 -penalty (see Uniejewski et al., 2016). In contrast, the *LASSO* method adds an l_1 -penalty (see Ziel, 2016). In both cases, a tuning parameter λ controls the strength of the penalty. While the ridge penalty shrinks the coefficients of correlated predictors towards each other, the LASSO tends to pick one of them and discard the others. The *elastic net* combines LASSO and ridge with weights of α and $(1 - \alpha)$, respectively (see Zou and Hastie, 2005). For $p_{max} = 21$, we calibrate the three methods with the pathwise coordinate descent algorithm of Friedman et al. (2010). That is, we define a grid of λ (and α) values and afterward determine the optimal ones for each hour by 10-fold cross-validation minimizing the mean square error in the calibration set. In majority, we obtain α values between 0.8 and 1.0, that is, elastic nets strongly tilted towards the LASSO approach.

With the intention of reducing parameter uncertainty, another form of regularization – Bayesian estimation – introduces prior information on the distribution of model parameters (see Karlsson, 2013). In our application, we implement the reduced form Bayesian VAR with Minnesota prior of Litterman (1986). We assume that a random walk model is a reasonable ‘center’ of beliefs about parameter behavior. Consequently, the prior mean is 1 for the first own lag of each equation and 0 for all other variables. The hyperparameters determining the prior variances are specified to ensure a low harmonic decay of variances with rising lags, to forbid a shrinkage of the coefficients of exogenous variables and to minimize the mean absolute error in 10-fold cross-validation (see Sims and Zha, 1998). For the lag order, we set $p = 7$ such that computation time remains within reasonable limits.

Finally, we consider a regularized random forest (RRF). In comparison to a standard RF, it introduces a penalty λ such that, depending on the level of λ , a new feature needs more or less predictive power to be selected for splitting a node. If the penalty is defined individually for each feature and as a function of the normalized importance score of a preliminary standard RF and a control γ , we arrive at the guided RRF (GRRF) (see Deng and Runger, 2013). We implement the GRRF using the RF specification of Section 3.2.3 and $\gamma = 0.75$, for which it performs best.

Feature selection. We use three distinct feature selection techniques. First, we implement a *stepwise* model building procedure (henceforth abbreviated with Step-Linear) which

starts with the intercept and then iteratively adds (removes) input variables to (from) a linear model (see Uniejewski et al., 2016). At each step and up to order $p_{max} = 21$, this method searches for the greatest improvement of the model fit (measured by the AIC), and terminates if no further improvement can be made. Second, besides using LASSO variables in a standard linear model (see regularization), we also combine LASSO feature selection with SVM regression (see Becker et al., 2009). This *LASSO-SVM* model is calibrated similar to our LASSO and SVM discussion above. Finally, we turn to our random forest scores. They are a natural basis for feature selection because they measure how much we would regret not including a feature in a model and, even though they are based on linear calculations, they can detect both linear and non-linear relationships between input and target variables (see Auret and Aldrich, 2012). If relationships were actually non-linear, it would not be surprising to observe that a random forest selection used in a linear prediction model performs poorly whereas a strong non-linear regression tool makes more efficient use of a random forest feature choice. To elaborate on this perspective, we use the K most relevant variables according to their importance score rankings in both a standard linear model and a SVM (as specified above). As for the number K of features to select, we set $K = 50$ for each hour to allow a comparison to our expert models which consider $24 + 24 + 4 = 52$ features. We denote the resulting models *RF(50)-Linear* and *RF(50)-SVM*. Furthermore, in the models *RF-Linear* and *RF-SVM*, we use the K between 10 and 100 that optimizes generalizability.

Forecast combinations. In addition to forecasting via individual models, we look at forecast combinations. Because even rudimental combinations often work reasonably well in comparison to more complex ones (see Genre et al., 2013), we choose two simple weighting schemes. We use the *simple average* of all forecasts produced by our entire model selection. While this approach provides some insurance against crucial forecasting failures (see Hibon and Evgeniou, 2005), it does not consider that the performance of different models can vary with market conditions (see Weron and Misiorek, 2008). Therefore, we also generate combinations where forecasting models are weighted based on their inverse mean square prediction error (MSPE) over the most recent Q periods (see Baumeister and Kilian, 2015). We label these combinations *Inv-MSPE(Q)* and test $Q = 7, 30, 365$ days.

3.3 Empirical results

3.3.1 Forecasting error values

We now present our day-ahead forecasting results for Germany/Austria. Recall that we use a two-year out-of-sample test period from January 2016 to December 2017, which, in contrast to typically used one-year out-of-sample periods, helps to deliver more persuasive results.

Clark and McCracken (2013) and Weron (2014) state that, in forecast evaluation, there is no ‘industry standard’ and the implemented error measures vary considerably. However,

the most often used metrics are based on absolute errors. We use the mean absolute error (MAE) given by

$$\text{MAE} = \frac{1}{24N} \sum_{d=1}^N \sum_{h=0}^{23} |P_{d,h} - \hat{P}_{d,h}|, \quad (3.3.1)$$

where $\hat{P}_{d,h}$ is the forecast of $P_{d,h}$. Because absolute errors are hard to compare across datasets, some studies use measures with absolute percentage errors. However, because electricity can exhibit negative prices, this is no option for us.⁸

Apart from l_1 -type norms, square or l_2 -type norms are also very popular. Therefore, we additionally calculate the root mean square error (RMSE) defined as

$$\text{RMSE} = \sqrt{\frac{1}{24N} \sum_{d=1}^N \sum_{h=0}^{23} |P_{d,h} - \hat{P}_{d,h}|^2}. \quad (3.3.2)$$

While the MAE is a robust measure, the RMSE is a optimal measure for least square problems but very sensitive to outliers.

In Table 3.2, we report both MAE and RMSE for the evaluation of forecasts produced by our model selection. As indicated by the definitions (3.3.1) and (3.3.2), the errors are averaged over all hours $h = 0, \dots, 23$. We also present the estimated standard deviations of MAE and RMSE, which are computed using a residual-based bootstrap with a bootstrap sample size of 10,000, and several MAE- and RMSE-based model rankings.

In line with our expectation, simple benchmarks, which do not take into account intraday dependencies, perform worst. We find the largest errors, i.e., $\text{MAE} = 7.750$ and $\text{RMSE} = 13.244$, for the persistent model. Additionally, the 24AR errors are very close to this benchmark. The expert VAR model is already a great improvement over these approaches because, for example, its MAE of 5.651 and RMSE of 8.981 are 27.1% and 32.2% below the persistent forecasts, respectively. The accuracy of the unrestricted VAR model with AIC-based lag order of $p = 2$ is slightly lower. In both, the group of benchmark models and the entire model selection, the expert SVM shines. That is, for example, its MAE of 5.461 is beaten by only three other individual models: PC(11)-SVM, RF(50)-SVM and RF-SVM. Thus, the expert variable choice (lags $l = 1, 7$ plus weekday dummies) provides solid price forecasts. Within the same set of variables, the choice of regression tool appears to be particularly important. In a direct comparison of the benchmark models using the same expert variable choice, the SVM algorithm outperforms the linear model and the neural network approach. The latter delivers poorer results than the linear model and, in a pure time series context, appears to be unfit for EPF.

All models incorporating PCA feature extraction provide relatively good MAE (RMSE) values below 5.633 (8.716). Feature extraction combined with a VAR model leads to better

⁸We could normalize the error by the average price obtained in a specific evaluation interval (for example, a week). Also note that some EPF applications report week-weighted mean absolute errors.

		Error values				Group ranks		Overall ranks	
		MAE		RMSE		MAE	RMSE	MAE	RMSE
<i>Benchmarks</i>									
1	Persistent	7.750	(.130)	13.244	(.212)	6	6	27	27
2	24AR	6.605	(.117)	10.186	(.175)	4	4	25	25
3	VAR	5.824	(.098)	9.075	(.165)	3	3	23	21
4	Expert VAR	5.651	(.093)	8.981	(.170)	2	2	19	20
5	Expert NN	6.657	(.062)	10.649	(.190)	5	5	26	26
6	Expert SVM	5.461	(.098)	8.709	(.159)	1	1	8	10
<i>Feature extraction</i>									
7	PC(7)-VAR	5.628	(.100)	8.716	(.171)	5	6	16	12
8	PC(7)-LASSO	5.565	(.099)	8.661	(.166)	4	3	14	7
9	PC(7)-SVM	5.461	(.100)	8.684	(.174)	2	4	8	8
10	PC(11)-VAR	5.633	(.101)	8.710	(.173)	6	5	17	11
11	PC(11)-LASSO	5.512	(.098)	8.636	(.164)	3	2	12	6
12	PC(11)-SVM	5.396	(.095)	8.588	(.161)	1	1	7	5
<i>Regularization</i>									
13	Ridge	5.575	(.096)	8.918	(.167)	4	4	15	17
14	LASSO	5.464	(.097)	8.756	(.170)	2	3	11	16
15	Elastic net	5.462	(.097)	8.754	(.171)	1	2	10	15
16	Bayesian VAR	5.516	(.098)	8.685	(.167)	3	1	13	9
17	GRRF	5.761	(.106)	9.514	(.159)	5	5	21	23
<i>Feature selection</i>									
18	Step-Linear	6.301	(.107)	9.611	(.171)	6	6	24	24
19	LASSO-SVM	5.783	(.098)	9.138	(.160)	5	5	22	22
20	RF(50)-Linear	5.655	(.097)	8.922	(.169)	4	3	20	18
21	RF-Linear	5.633	(.095)	8.922	(.167)	3	3	17	18
22	RF(50)-SVM	5.391	(.095)	8.744	(.170)	2	2	6	14
23	RF-SVM	5.385	(.096)	8.726	(.168)	1	1	5	13
<i>Forecast combinations</i>									
24	Simple Average	5.332	(.094)	8.557	(.163)	4	4	4	4
25	Inv-MSPE(7)	5.289	(.093)	8.524	(.165)	1	1	1	1
26	Inv-MSPE(30)	5.307	(.093)	8.533	(.165)	2	2	2	2
27	Inv-MSPE(365)	5.325	(.093)	8.534	(.163)	3	3	3	3

For the forecasting models of Section 3.2.5 and the German/Austrian out-of-sample test period from January 2016 to December 2017, this table reports the mean absolute errors (MAE) and root mean square errors (RMSE) across all hours of the day. The corresponding bootstrap standard deviations are given in parentheses. We also present group and overall model rankings, where the best model has rank 1. To allow a comparison with the hourly Tables 3.3 and 3.4, we additionally implement a heatmap indicating well (\rightarrow green) and poorly (\rightarrow red) performing models.

TABLE 3.2: Forecasting errors

results than the standard VAR approach. In contrast, MAE and RMSE of the PC(7)-VAR and the PC(11)-VAR are only marginally better than the expert VAR benchmark. The best model of the group is the PC(11)-SVM. It also yields the third-lowest MAE of 5.396 across all individual models and barely misses the top rank. In summary, feature extraction (i) slightly stabilizes forecasts, i.e., lowers RMSE values, and (ii) can achieve top MAE performance in combination with SVM regression. However, our results also highlight that the determination of the number of factors K is not trivial. Commonly used test statistics appear to supply a too-small number of factors (in our case $K = 7$), which negatively influences forecasting accuracy (because we could obtain better results for $K = 11$).

In the class of (linear) regularization approaches, the elastic net appears to be a solid, ready-to-use method. It reaches the best-in-class MAE of 5.462 and, among all individual models, is outperformed only by competing models using (non-linear) SVM regressions (except for LASSO-SVM). For reasons discussed in Section 3.2.5, the performance of the LASSO is similar to the elastic net. Consequently, ridge regression performs worse than the LASSO. When turning to the RMSE, the value of 8.756 for the elastic net is outperformed not only by SVM settings but also all PCA approaches. Interestingly, the Bayesian VAR provides a better RMSE value of 8.685, which simultaneously is the best performance

among all individual models excluding PCA approaches. Its MAE of 5.516 is not bad, but worse than many other techniques, including LASSO regression. While providing reliable importance scores, the GRRF is characterized by poor performance and cannot beat the expert VAR benchmark.

High errors of $\text{MAE} = 6.301$ and $\text{RMSE} = 9.611$ suggest that stepwise linear regression can be considered an obsolete approach to feature selection. It is not only outperformed by all alternative feature selection techniques but also almost all other individual models (except for the persistent model, 24AR and the expert NN). While standard linear LASSO regression performs well, using LASSO-selection in a non-linear SVM yields rather unsatisfactory results. That is, we obtain $\text{MAE} = 5.783$ and $\text{RMSE} = 9.138$. Turning to our random forest proposal and starting with the linear model, the detailed variable choice derived from random forest importance scores performs similar to the expert choice. This is not surprising because our importance profiles suggest that lagged prices with $l = 1, 7$ comprise most of the explanatory power. Also, in comparison to the linear regularization and feature extraction settings, random forests cannot shine. However, in combination with a non-linear SVM, the 50 input variables with the highest importance scores result in the second-lowest MAE of 5.391 across all individual models and beat the expert choice. The RMSE of 8.721 is slightly above the expert choice. Optimally tuning the number K of input variables, we obtain the model with the lowest individual MAE of 5.385. While this implies that there appear to be some non-linear tendencies which can be captured best via random forests, the RMSE value of RF-SVM suggests that other non-linear approaches like PC-SVM (and partially even linear ones) have similar qualities.⁹

As far as our forecast combinations are concerned, we find that our selection of combination schemes can beat each individual model in terms of MAE and RMSE. That is, combining models instead of relying on just one has benefits for forecasters. We can isolate two causes of enhanced predictive power. First, because we can already observe lower errors for the simple average, averaging itself increases accuracy. Second, because the errors for $\text{Inv-MSPE}(Q)$ combination schemes (especially for the short evaluation period $Q = 7$) are lower than for the simple average, the former kind of combined forecasts appears to benefit from introducing recent forecasting performance via changing model weights. Another aspect worth noting is related to the consistency of rankings derived from MAE and RMSE. While, for individual models, we have seen that evaluations can differ depending on the error measure (e.g., the elastic net has a lower MAE but a higher RMSE than PC-VAR approaches), both MAE and RMSE agree on the ranks of the combined approaches.

So far our discussions have focused on an aggregate perspective including all hours of the day. Because this may conceal important effects, we complete the picture with a separate look at each hour. Based on Tables 3.3 and 3.4, we gain several interesting insights. First, while prices of night hours 00-05 and evening hours 20-23 can be forecasted with relative ease, it is harder to predict hourly prices during daylight hours 06-19. Second, we can identify different leading models depending on daytime. For example, if we are interested in the night hours 00-05, the RF-SVM approach outperforms the PC(11)-SVM

⁹In Section 3.3.2, we discuss this observation in more detail.

model, whereas, for most daylight hours, the latter is slightly more accurate. In addition, the elastic net performs best for many evening hours. Hence, in our application, another reason for the superiority of forecast combination schemes is related to the fact that they merge very different approaches which are characterized by advantages for specific hours of the day.¹⁰

A close comparison of our variable importance profiles and the observed model performance explains why some models outperform others for certain hours. In other words, it provides a ‘guide’ for model selection which helps us to select the best model depending on the hour we wish to forecast. For night hours, there are only a few important features. These features can be efficiently identified by random forest scores and passed on to a SVM. During day-time, there are more relevant inputs (corresponding to lags of order $l = 1, 7$) which should be transformed to a lower dimensional space utilizing PCA and then transferred to a SVM. For evening hours, large lags of the same hour are relevant input variables. In our setting, the PCA-approaches only consider lag variables up to order $p = 7$ which explains why PC(11)-SVM loses accuracy for these hours. At the same time, there are many other relevant input variables. The elastic net considers lag variables up to order $p = 21$ and therefore performs best for evening hours.

3.3.2 Diebold-Mariano test

To supplement Section 3.3.1, we perform several statistical tests. We start with the Diebold and Mariano (1995) (DM) test of equal predictive ability, which compares two models. Because our predictions for all 24 hours of the next day are made at the same time using the same information set, forecast errors for a particular day typically exhibit high serial correlation. However, the classic version of the test requires their stationarity. Therefore, we follow Bordignon et al. (2013), Nowotarski et al. (2014) and Maciejowska and Nowotarski (2016) by conducting the test for each of the 24 load periods h separately. We define its loss function based on absolute and squared error losses

$$L_i(\varepsilon_d) = |\varepsilon_d|^i = |P_{d,h} - \hat{P}_{d,h}|^i \quad (3.3.3)$$

of the model forecasts, where $i = 1, 2$ and, to simplify notation, we drop the dependence on h on the left hand side. For each model pair (m_1, m_2) and hour h , we compute the loss differential series

$$l_{i,d} = L_i(\varepsilon_d^{(m_1)}) - L_i(\varepsilon_d^{(m_2)}), \quad (3.3.4)$$

which is the main focus of the test. We specify two one-sided DM tests at the 5% significance level. For both tests, the null hypothesis is $H_0 : \mathbb{E}[l_{i,d}] = 0$, i.e., both methods have the same forecast accuracy. However, the tests differ with respect to the alternative hypothesis. For the first test, we have $H_1 : \mathbb{E}[l_{i,d}] < 0$, i.e., model m_2 is less accurate than model m_1 . The second test is complementary with the reverse alternative $\bar{H}_1 : \mathbb{E}[l_{i,d}] > 0$,

¹⁰A further explanation may be that some models work better for certain seasons than others. A detailed analysis of this aspect is left for future research.

ID	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	6.95	6.87	7.19	7.03	6.60	6.15	7.35	8.00	8.20	8.34	8.86	9.11	9.12	9.53	9.35	8.56	8.31	8.24	8.26	7.74	6.97	6.32	6.12	6.74
2	5.39	5.48	5.77	5.66	5.44	5.10	6.38	7.40	7.45	7.28	7.65	7.91	7.94	8.54	8.41	7.67	7.20	7.02	7.13	6.53	5.89	5.16	4.91	5.21
3	2.88	3.09	3.69	3.95	3.93	3.82	5.22	6.35	6.37	6.22	6.56	7.08	7.07	7.52	7.56	7.08	6.91	7.29	7.61	7.18	6.25	5.47	5.25	5.42
4	2.87	3.11	3.70	3.96	3.98	3.81	4.90	5.94	6.03	5.94	6.46	6.97	7.01	7.42	7.45	6.93	6.69	6.89	7.23	6.71	5.98	5.27	4.98	5.35
5	3.33	3.48	4.43	4.80	5.59	4.60	5.74	6.69	8.24	7.29	7.19	7.94	8.70	8.16	8.43	7.60	7.81	8.86	8.90	7.73	6.58	6.31	5.27	6.09
6	2.85	2.96	3.44	3.71	3.57	3.63	4.85	5.57	5.71	5.76	6.28	6.62	6.69	7.13	7.15	6.69	6.54	6.84	7.07	6.60	5.89	5.11	5.32	5.11
7	3.07	3.19	3.68	3.84	3.82	3.73	4.80	5.81	6.06	6.06	6.45	6.94	6.99	7.41	7.53	7.00	6.67	6.91	7.17	6.73	5.98	5.11	4.84	5.21
8	3.09	3.23	3.73	3.90	3.88	3.78	4.76	5.69	5.94	6.00	6.36	6.84	6.88	7.28	7.34	6.64	6.56	6.86	7.04	6.54	5.92	5.08	4.85	5.19
9	3.03	3.11	3.51	3.66	3.66	3.61	4.65	5.58	5.78	5.94	6.31	6.76	6.80	7.13	7.17	6.64	6.45	6.75	7.03	6.52	5.84	5.08	4.85	5.20
10	2.97	3.19	3.72	3.87	3.83	3.70	4.90	5.82	5.98	6.07	6.47	6.96	6.99	7.34	7.43	6.90	6.57	6.93	7.28	6.85	5.93	5.15	4.94	5.39
11	2.98	3.27	3.80	3.99	3.96	3.79	4.83	5.71	5.86	5.89	6.30	6.79	6.80	7.14	7.13	6.62	6.37	6.68	6.91	6.55	5.79	4.99	4.90	5.24
12	2.98	3.08	3.53	3.67	3.67	3.59	4.68	5.54	5.73	5.84	6.17	6.63	6.64	6.96	6.98	6.48	6.29	6.69	6.91	6.47	5.76	5.01	4.92	5.28
13	2.98	3.20	3.63	3.91	3.91	3.85	4.89	5.77	5.94	5.92	6.41	6.93	6.97	7.38	7.43	6.87	6.69	6.95	6.97	6.42	5.60	4.90	4.90	5.35
14	2.71	3.03	3.51	3.75	3.70	3.61	4.86	5.86	5.94	5.83	6.33	6.84	6.90	7.32	7.33	6.83	6.55	6.72	6.83	6.31	5.60	4.93	4.77	5.09
15	2.71	3.03	3.51	3.77	3.76	3.61	4.86	5.84	5.92	5.80	6.33	6.81	6.90	7.30	7.33	6.83	6.56	6.72	6.80	6.32	5.59	4.93	4.76	5.09
16	2.85	3.01	3.48	3.70	3.73	3.61	4.85	5.94	6.01	5.97	6.29	6.83	6.92	7.36	7.33	6.75	6.53	6.73	7.04	6.62	5.77	5.07	4.85	5.14
17	3.17	3.68	4.31	4.53	4.50	4.37	5.15	5.86	6.01	6.07	6.51	7.00	7.02	7.46	7.39	7.01	6.67	6.81	7.01	6.51	5.87	5.18	5.02	5.14
18	3.22	3.57	4.22	4.47	4.43	4.43	5.44	6.44	6.73	6.33	7.42	8.39	8.01	8.27	8.11	7.71	6.83	7.61	8.42	8.10	6.23	5.56	5.26	5.71
19	3.52	3.72	4.18	4.30	4.30	4.40	5.31	6.00	6.11	6.23	6.58	7.11	7.16	7.38	7.30	6.87	6.73	7.21	6.91	6.58	5.89	4.98	4.94	5.05
20	2.85	3.17	3.74	3.89	3.87	3.81	4.91	5.99	6.14	6.13	6.54	7.13	7.14	7.57	7.57	7.02	6.68	7.00	6.97	6.52	5.85	5.13	4.87	5.24
21	2.87	3.15	3.75	3.89	3.87	3.86	4.93	5.95	6.08	6.11	6.51	7.08	7.14	7.52	7.56	6.92	6.64	6.99	7.03	6.45	5.84	4.96	4.84	5.27
22	2.75	2.96	3.34	3.59	3.56	3.81	4.74	5.59	5.77	5.95	6.27	6.76	6.78	7.05	7.15	6.58	6.40	6.66	6.66	6.23	5.78	4.97	4.82	5.20
23	2.74	2.99	3.36	3.62	3.62	3.61	4.86	5.55	5.78	5.82	6.25	6.85	6.75	7.13	7.03	6.58	6.37	6.67	6.69	6.37	5.67	4.94	4.93	5.08
24	2.79	2.99	3.46	3.68	3.70	3.61	4.69	5.53	5.68	5.71	6.19	6.66	6.70	7.10	7.08	6.57	6.28	6.50	6.63	6.24	5.56	4.82	4.71	5.03
25	2.74	2.95	3.40	3.61	3.61	3.59	4.63	5.45	5.62	5.68	6.16	6.62	6.68	7.08	7.06	6.53	6.25	6.46	6.60	6.20	5.53	4.82	4.66	5.01
26	2.74	2.96	3.40	3.63	3.65	3.59	4.65	5.49	5.67	5.70	6.17	6.64	6.69	7.10	7.08	6.55	6.28	6.48	6.62	6.22	5.54	4.84	4.67	5.01
27	2.75	2.96	3.43	3.65	3.67	3.59	4.66	5.53	5.69	5.72	6.19	6.65	6.71	7.10	7.10	6.58	6.28	6.49	6.63	6.24	5.56	4.88	4.72	5.03

TABLE 3.3: Mean absolute errors per hour

For the forecasting models of Section 3.2.5 and our German/Austrian out-of-sample test period from January 2016 to December 2017, this table reports the mean absolute errors per hour. The ID in the first column identifies the forecasting model (see also Table 3.2). A heatmap is used to indicate well (\rightarrow green) and poorly (\rightarrow red) performing models.

ID	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	11.59	12.10	12.84	12.29	11.60	11.35	13.70	14.05	14.38	13.59	14.02	14.65	15.24	17.00	16.83	14.68	13.50	12.86	13.52	11.79	10.62	9.85	9.77	12.94
2	8.77	8.83	9.47	9.05	8.59	8.20	9.93	11.00	11.17	10.77	11.13	11.49	11.77	12.87	12.99	11.35	10.72	12.86	13.52	9.41	8.47	7.61	7.51	9.31
3	5.02	5.50	6.58	6.39	6.23	6.24	7.62	9.21	9.29	9.07	9.66	10.18	10.65	11.8	12.15	10.66	10.43	10.44	10.92	9.90	8.72	7.85	7.77	9.33
4	4.88	5.44	6.61	6.39	6.27	7.16	7.55	8.89	9.06	8.89	9.55	10.14	10.73	11.75	12.07	10.46	10.18	10.54	11.16	9.91	8.66	7.83	7.75	9.25
5	5.70	5.99	6.59	6.91	8.54	7.16	9.44	10.28	12.56	10.57	11.52	14.43	14.18	12.85	14.41	12.18	11.54	12.84	12.84	10.95	9.25	8.78	7.91	10.08
6	4.31	4.00	5.52	5.60	5.44	5.82	7.11	8.64	9.04	8.78	9.37	9.74	10.65	11.68	12.00	10.23	9.95	10.31	10.61	9.49	8.41	7.51	7.76	9.05
7	4.70	4.76	5.52	5.61	5.57	5.78	7.39	8.77	9.05	8.90	9.40	9.91	10.46	11.56	11.92	10.33	10.08	10.39	10.79	9.58	8.33	7.42	7.41	9.16
8	4.78	4.84	5.38	5.65	5.60	5.82	7.40	8.70	8.97	8.80	9.43	9.81	10.38	11.47	11.79	10.21	10.05	10.36	10.69	9.36	8.29	7.43	7.42	9.16
9	4.72	4.74	5.38	5.40	5.36	5.64	7.43	8.71	9.02	8.88	9.36	9.94	10.46	11.56	11.84	10.21	10.05	10.36	10.76	9.36	8.30	7.46	7.43	9.16
10	4.67	4.85	5.69	5.76	5.66	5.77	7.29	8.71	8.89	8.80	9.29	9.82	10.41	11.47	11.86	10.25	9.97	10.33	10.88	9.69	8.37	7.50	7.55	9.27
11	4.69	4.84	5.63	5.72	5.66	5.86	7.35	8.89	9.01	8.76	9.29	9.82	10.38	11.44	11.70	10.09	9.89	10.22	10.80	9.30	8.17	7.38	7.46	9.29
12	4.70	4.71	5.50	5.50	5.45	5.70	7.42	8.85	9.01	8.82	9.29	9.82	10.38	11.44	11.70	10.09	9.89	10.22	10.80	9.30	8.17	7.38	7.46	9.29
13	4.67	4.70	5.44	5.40	5.34	5.60	7.38	8.85	9.06	8.82	9.29	9.82	10.38	11.44	11.70	10.09	9.89	10.22	10.80	9.30	8.17	7.38	7.46	9.29
14	4.66	4.66	5.44	5.40	5.34	5.60	7.38	8.85	9.06	8.82	9.29	9.82	10.38	11.44	11.70	10.09	9.89	10.22	10.80	9.30	8.17	7.38	7.46	9.29
15	4.70	4.66	5.44	5.40	5.34	5.60	7.38	8.85	9.06	8.82	9.29	9.82	10.38	11.44	11.70	10.09	9.89	10.22	10.80	9.30	8.17	7.38	7.46	9.29
16	4.70	4.70	5.44	5.40	5.34	5.60	7.38	8.85	9.06	8.82	9.29	9.82	10.38	11.44	11.70	10.09	9.89	10.22	10.80	9.30	8.17	7.38	7.46	9.29
17	5.39	5.77	6.38	6.56	7.13	6.99	8.71	9.44	9.81	10.36	10.63	11.13	11.50	12.48	12.50	11.17	10.73	11.04	11.48	9.74	8.72	7.80	7.66	9.31
18	5.14	5.84	7.10	6.81	7.13	6.92	8.71	9.44	9.81	10.36	10.63	11.13	11.50	12.48	12.50	11.17	10.73	11.04	11.48	9.74	8.72	7.80	7.66	9.31
19	4.89	5.26	6.48	6.38	6.45	6.51	8.62	9.33	9.61	9.67	9.85	10.35	10.88	11.90	12.13	10.60	10.25	10.94	10.98	9.67	8.47	7.56	7.69	9.15
20	4.96	5.26	6.48	6.38	6.45	6.51	8.62	9.33	9.61	9.67	9.85	10.35	10.88	11.90	12.13	10.60	10.25	10.94	10.98	9.67	8.47	7.56	7.69	9.15
21	4.88	5.26	6.48	6.38	6.45	6.51	8.62	9.33	9.61	9.67	9.85	10.35	10.88	11.90	12.13	10.60	10.25	10.94	10.98	9.67	8.47	7.56	7.69	9.15
22	4.34	4.68	5.43	5.53	5.52	5.81	7.57	8.92	8.99	9.08	9.48	10.08	10.47	11.54	12.18	10.21	10.00	10.43	10.89	9.36	8.37	7.59	7.53	9.21
23	4.33	4.61	5.41	5.53	5.59	5.81	7.83	9.03	9.12	9.12	9.43	10.14	10.49	11.70	11.70	10.32	9.98	10.39	10.91	9.44	8.25	7.44	7.46	9.04
24	4.50	4.74	5.52	5.59	5.56	5.86	7.54	8.72	8.95	8.74	9.28	9.78	10.30	11.41	11.65	10.08	9.81	10.09	10.43	9.12	8.09	7.32	7.30	9.04
25	4.41	4.72	5.58	5.57	5.54	5.83	7.39	8.64	8.84	8.60	9.21	9.69	10.28	11.41	11.67	10.08	9.80	10.07	10.42	9.10	8.06	7.29	7.27	9.03
26	4.43	4.67	5.51	5.56	5.53	5.83	7.42	8.66	8.90	8.64	9.25	9.74	10.28	11.41	11.65	10.08	9.81	10.09	10.43	9.12	8.09	7.32	7.31	9.03
27	4.41	4.67	5.47	5.54	5.50	5.81	7.43	8.70	8.91	8.65	9.26	9.74	10.28	11.38	11.65	10.07	9.80	10.09	10.43	9.12	8.09	7.32	7.31	9.03

For the forecasting models of Section 3.2.5 and our German/Austrian out-of-sample test period from January 2016 to December 2017, this table reports the root mean square errors per hour. The ID in the first column identifies the forecasting model (see also Table 3.2). A heatmap is used to indicate well (\rightarrow green) and poorly (\rightarrow red) performing models.

TABLE 3.4: Root mean square errors per hour

i.e., model m_2 is more accurate than model m_1 . We assume that the forecasts of consecutive days (and hence loss differentials) are not serially correlated. Especially for the well-performing models, this is a valid assumption.

The DM test results are summarized in the heatmaps of Figure 3.5 (3.6), where, for each model pair, we report the number of hours for which model m_2 is significantly less (more) accurate than model m_1 . Starting with Figure 3.5, we find that, regardless of the loss measure, some models are significantly less accurate than most other models for almost all hours. The persistent, 24AR, expert NN, GRRF and Step-Linear approaches are typical examples. Only one model is not significantly worse than any other model for all hours: the Inv-MSPE(7) scheme. The additional forecast combination schemes are not inferior when compared to the individual models except for some hours forecasted via the expert SVM, RF(50)-SVM and PC(11)-SVM methods. Figure 3.6 provides stronger evidence in favor of the forecast combinations. It shows that, with the former three exceptions and in terms of MAE, they tend to provide significantly better forecasts than all models for almost all hours. When focusing on the RMSE, outperformance becomes less significant but still favors the combinations.

With respect to our RF-SVM proposal, we can see that, in terms of MAE, it is significantly more accurate than many competing models in almost all hours. It even outperforms its most serious competitors (expert SVM, PC-SVM) for many hours. Turning to the RMSE, it cannot outperform these rivals (and additionally some linear regularization techniques) at all. However, this is no reason to worry because our results also show that, in these cases, the RF-SVM is almost never significantly less accurate than these models. Consequently, it can be considered a promising new technique with the advantage that, in contrast to the similarly performing alternatives, it can deliver variable importance profiles that give valuable insights into the price dynamics in electricity markets and, as we have seen in Section 3.3.1, can guide model selection in future research.

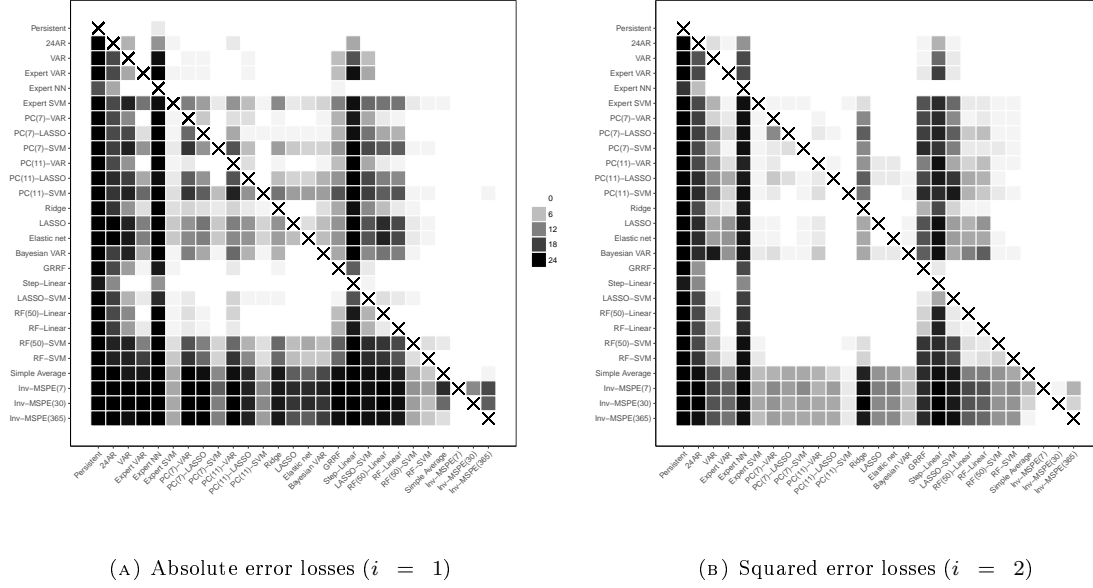
3.3.3 Hansen test

As a second test, we apply the Hansen (2005) test of superior predictive ability. In contrast to the two-model DM test, its idea is to compare a given benchmark model m_b to all alternative models m_s , $s \neq b$. The key variables for the test are relative performance metrics

$$l_{i,d}^{(m_s)} = L_i(\varepsilon_d^{(m_b)}) - L_i(\varepsilon_d^{(m_s)}), \quad (3.3.5)$$

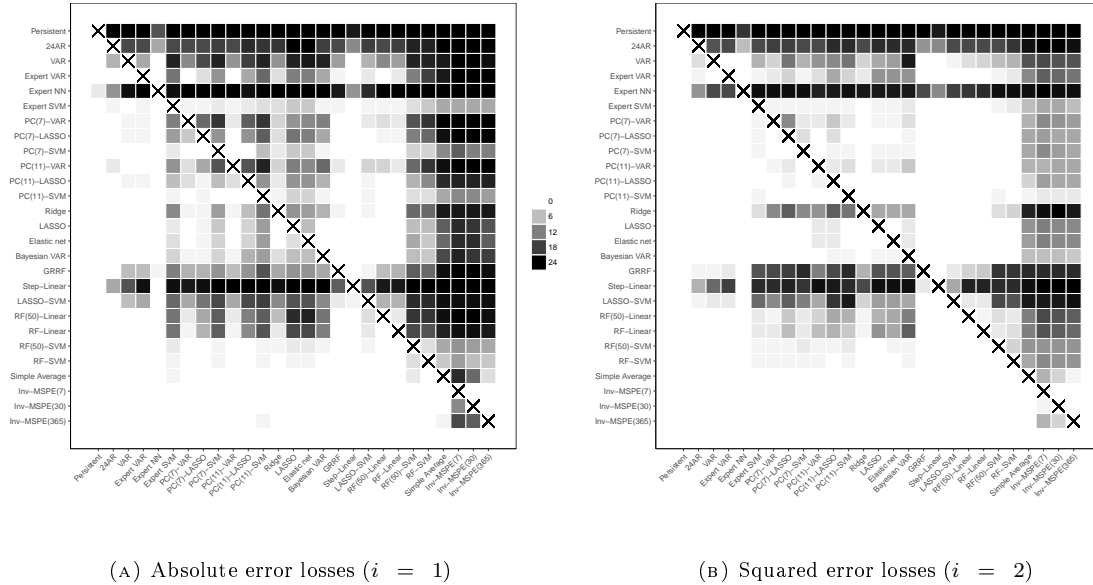
where the loss function L_i is given by (3.3.3) and we again set $i = 1, 2$. Consequently, $l_{i,d}^{(m_s)}$ measures the day- d performance of model m_s relative to the benchmark. We collect the differences of all alternatives to the benchmark in the vector $l_{i,d} = (l_{i,d}^{(m_s)})_{s \neq b}$. Provided that $\mu := \mathbb{E}[l_{i,d}]$ is well-defined, we can then formulate the set of hypotheses $H_0 : \mu \leq 0$ and $H_1 : \exists s, \mu_s > 0$ because the vector element μ_s corresponds to the case where a model m_s is better than m_b .

As in Hansen and Lunde (2005) we use a bootstrap implementation based on the stationary bootstrap of Politis and Romano (1994). It uses pseudo time series $l_{i,\tau_{b,d}}$, $b = 1, \dots, B$, which are resamples of the original data $l_{i,d}$ and where a series $\{\tau_{b,1}, \dots, \tau_{b,N}\}$ is constructed by combining blocks of random length taken from $\{1, \dots, N\}$. More specifically,



For all models of Section 3.2.5 and our German/Austrian out-of-sample period from 2016 to 2017, this figure presents the results of one-sided Diebold and Mariano (1995) tests conducted with the alternative hypothesis ‘less accurate’ and a significance level of 5%. We compare each model with each other model and perform the comparison for each hour separately. We then sum the number of significant differences in forecasting performance across the 24 hours and use a heatmap to indicate the number of hours for which model m_2 (x-axis) is less accurate than model m_1 (y-axis).

FIGURE 3.5: Diebold-Mariano test with alternative ‘less accurate’



In analogy to Figure 3.5, this figure presents the results of one-sided Diebold and Mariano (1995) tests conducted with the alternative hypothesis ‘more accurate’.

FIGURE 3.6: Diebold-Mariano test with alternative ‘more accurate’

we use a number of $B = 1,000$ bootstrap resamples and the block length is assumed to follow a geometric distribution with mean 20.¹¹

We summarize the Hansen test results in Figure 3.7, where heatmaps evaluate the significance of the test statistic at conventional levels of 1%, 5% and 10%. If, in forecasts of the hour- h price, the benchmark model m_b is significantly worse than any alternative model, the significance level is stated and colored in grayscale. In contrast, no level statement and coloring reflect insignificance. In other words, in these cases, the null hypothesis of no inferiority in comparison to the alternatives cannot be rejected at any of the given significance levels. For absolute error losses, we detect only a few models whose test statistics are insignificant for many hours of the day. The expert SVM, PC(7)-SVM and PC(11)-SVM are not inferior to the alternatives for most hours, with the exception of the evening hours 17-23. In contrast, the elastic net is not inferior in most of these evening hours. Only RF(50)-SVM and RF-SVM techniques as well as all three Inv-MSPE forecast combinations show convincing results during almost all hours of day. Switching to squared error losses, we observe more models with insignificant test statistics over many hours and even two models (Inv-MSPE(7) and Inv-MSPE(30)) with insignificance for all hours. Supporting our previous results, the RF-SVM setting often cannot be rejected, but this also holds, for example, for an expert SVM and PC-SVMs.

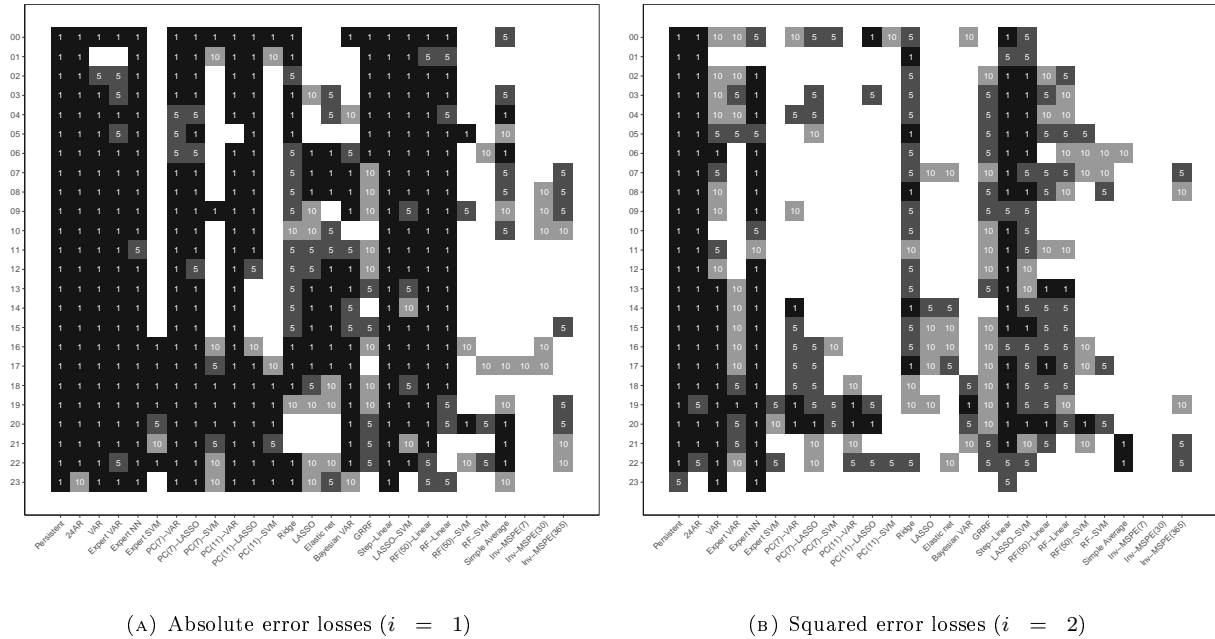
3.4 Conclusion

Recent studies emphasize that taking into account the autoregressive intraday dependency structure of electricity prices is of crucial importance for short-term forecasting. While empirical correlation patterns support this idea, practical implementation is complicated by dimensionality problems. That is, simply including a wide range of lagged prices (for example, in an unrestricted VAR model) can lead to overfitting and ignores that some lags may be more relevant for prediction than others. Luckily the econometrics and machine learning literature offers a battery of methods, including feature selection, feature extraction and regularization techniques, which can help to deal with this issue.

In an analysis of the German/Austrian electricity market, we propose to perform feature selection based on variable importance profiles derived from a random forest algorithm. It is an ideal choice for dimension reduction in high dimensional spaces and, in contrast to other popular methods, allows us to provide a very detailed picture of existing price dependencies. Our analysis shows that previous findings on the dominant position of prices lagged by one day and one week also hold for our data and methodology. Thus, they can be considered robust. Furthermore, we highlight that the predictive power of past prices depends on which hourly contract we would like to forecast. For example, evening prices of the previous day are highly relevant for prices of night hours, but less relevant for prices of morning and high noon hours.

By using the detailed importance profiles in combination with a non-linear support vector machine and comparing its performance to a wide selection of traditional benchmark

¹¹Using a rule of thumb, which suggests $1/(\text{significance level}) \times 100$ resamples, does not influence our conclusions.



shown that our individual models do not perform equally well for all hourly time series. As far as future research is concerned, such a result does not mean that we can stop developing new individual forecasting models. It is the variety of models — each having advantages over the others — that leads to overall improvement. New forms of machine learning (i.e., advanced random forests or support vector machines), which provide better forecasts for afternoon hours, might be added. For example, while we have kept our approach simple, we might alternatively assign permutation scores to subsets of features instead of single features (see Gregorutti et al., 2015). Furthermore, we may wish to leave the boundaries of pure time series analysis by incorporating supply and demand factors, economic variables and electricity market-specific features (like temperature or wind) into the universe of potential predictors (see Avci et al., 2018; Gürtler and Paulsen, 2018; Kristiansen, 2012; Ludwig et al., 2015).

Chapter 4

Swing option-implied volatility

Motivated by the increasing interest of academics and practitioners in swing options, we develop a method for computing swing option-implied volatility. Using recent theoretical advances, we build a dynamic programming option pricing framework supplemented by an additive single-factor forward curve model which, as we show, is a natural candidate for implied volatility estimation. Within this framework, we propose to obtain implied volatility via a combination of Monte Carlo techniques and a root-finding algorithm. In a first empirical study of volatilities implied by natural gas swing options, we illustrate key features of swing option-implied volatilities which serve as a starting point for future research parametrizing the swing option volatility surface and studying its predictive power.

4.1 Introduction

Implied volatility is one of the most important metrics for the analysis and presentation of financial options. Under the Black and Scholes (1973), Cox et al. (1979) and other models, there is a one-to-one link between the option value and the volatility of the underlying asset. Consequently, the option pricing formula can be used to translate the market price of an option into expectations about future asset price fluctuations. This feature has influenced research and practice in various ways. First, in trading software, options are typically quoted in terms of implied volatility rather than price (see Mayhew, 1995).¹ Second, implied volatilities are the key ingredient of major investor sentiment indicators such as the CBOE Volatility Index (VIX) (see Whaley, 1993). Finally, there is a vast literature suggesting that implied volatility is a valuable predictor for future realized volatility (see Bandi and Perron, 2006; Busch et al., 2011; Christensen and Prabhala, 1998) and that profitable investment strategies can be constructed based on option-implied information (see DeMiguel et al., 2013; Xing et al., 2010).²

Calculating implied volatility is not a simple task because, even if models deliver closed-form pricing formulas, they often cannot be inverted analytically. While for the special case

¹This is partially because volatility is the main value driver of a delta-hedged option, i.e., an option hedged against small changes of the underlying's price.

²Some earlier studies suggested otherwise (see, for example, Canina and Figlewski, 1993; Jorion, 1995).

of European-style at-the-money options there are at least some closed-form approximations (see Brenner and Subrahmanyam, 1988; Curtis Jr and Carriker, 1988), the general approach to obtaining implied volatility is numerical and typically involves feeding the value-price difference into a root-finding algorithm.³ Fast convergence can be achieved if an analytic expression exists for the option's vega, i.e., the derivative of the option value with respect to the volatility parameter. In this case, a Newton-Raphson algorithm can achieve reasonably accurate estimates within just two or three iterations (see Kritzman, 1991; Orlando and Tagliatala, 2017).

Even though Ball et al. (1985) demonstrate that implied volatility may also be calculated for exotic options, the literature tends to focus on standard options.⁴ In particular, there exists no implied volatility concept for commodity swing options which are becoming more important than ever in natural gas markets (see Carmona and Ludkovski, 2010; Jaillet et al., 2004). Subject to periodic volume constraints, such options permit the holder to repeatedly exercise the right to receive large or small quantities of the underlying. While this design provides great flexibility in terms of execution, it complicates option valuation (because subperiod decisions influence future exercise possibilities) and all follow-up work (such as the derivation of option sensitivities and implied volatilities) requiring a solid valuation framework. Fortunately, recent research has achieved significant breakthroughs which now allow us to develop a method for computing swing option-implied volatilities.

Building on the general swing option definition of Jaillet et al. (2004), the core of our proposal combines Monte Carlo option valuation with the Newton-Raphson root-finding method. For this combination to work, two sets of results from the literature have to be taken into account. First, Barrera-Esteve et al. (2006) and Bardou et al. (2010) show that, for swing options with certain types of penalty functions or absolute global constraints, the optimal execution is of digital type. This allows us to simplify the general swing option valuation problem such that the well-known Longstaff and Schwartz (2001) method, originally designed for American-style options, becomes applicable for the valuation of swing options. Second, for the case of an optimal purchasing behavior of the swing option holder, Bonnans et al. (2012) justify a pathwise approach to calculate first-order sensitivities. It enables us to derive a formula for the vega of swing options,⁵ which is required for the root-finding procedure. We ensure the universal convergence of the Newton-Raphson iteration in our setting by showing that the swing option value function is strictly increasing and convex with respect to volatility.

When it comes to deriving implied volatilities, a suitable stochastic model for the price dynamics of the underlying is essential. Keppo (2004) and Berger et al. (2018) show that swing contracts can be viewed as baskets of forwards, calls and time-spread options. Consequently, we cannot simply assume a common geometric Brownian motion for spot prices because it would cause the time-spread optionality of swing options not to be priced.

³Alternatives include the robust but inefficient shotgun method (see Kritzman, 1991) and the bisection method (see Brown, 1990; Chriss, 1996), which requires a starting interval centered on an initial guess.

⁴A notable exception is the coverage of Asian-style options in Yang et al. (2009).

⁵In contrast to previous studies typically focusing on swing options with fixed strike prices, an important novelty of our study is that we also derive vega for swing options with gas-indexed strikes.

We require a model which captures all dimensions of flexibility offered to the option holder and is compatible with our main objective. While Boogert and de Jong (2011), Wahab and Lee (2011) and Chiarella et al. (2016) opt for multi-factor and regime-switching designs to model gas prices, we formulate an additive one-factor forward curve model in the spirit of Clewlow and Strickland (1999) which is equipped with a deterministic volatility function of negative exponential form. This modeling choice is important for three reasons.⁶ First, as many studies breaking new ground, we wish to keep our setup easy to understand and to reproduce (see Jaillet et al., 2004). This is particularly relevant for the practical implementation of our method. Second, the traditional notion of implied volatility (and our root-finding procedure) demands the measure to be unique which cannot be achieved in multi-factor or regime-switching environments because they assume distinct volatilities for each factor and state (see Kohrs et al., 2019; Wahab and Lee, 2011). Finally, our choice of volatility function ensures a Markovian spot price which is a key requirement for efficient dynamic programming in our valuation process (see Bellman, 1957). The implied spot price process is mean-reverting with a mean-reversion parameter that controls the value trade-off between outright and time-spread optionality of swing options. It thus helps us to better understand the flexibility value of swing options.

To provide some insights into typical empirical levels and properties of swing option-implied volatilities, we apply our new method to a dataset of market quotes for European natural gas swing options. This is therefore the first study to translate swing option prices into implied volatilities. Working with a unified implied volatility concept places us in a position to compare swing option quotes,⁷ to identify potential patterns and better understand relevant price drivers. We empirically investigate some generic volatility surface dimensions like time-to-maturity and moneyness. Similar to observations for standard electricity options (see Fanelli and Schmeck, 2019), we find the implied volatilities of swing option quotes to exhibit seasonality with respect to the delivery period. As far as the moneyness of swing options is concerned, we provide a formal definition and a computation method. In contrast to plain vanilla commodity market analyses (see Jia et al., 2021), volatility smirks are not particularly pronounced because swing options are often quoted at-the-money.

The remainder of our article is organized as follows. Section 4.2 introduces the theory on swing option valuation and sensitivity analysis. Section 4.3 discusses the specifics of the price model used for the underlying. Section 4.4 illustrates the core of our implied volatility approach based on Longstaff-Schwartz valuation and Newton-Raphson root-finding iteration. Section 4.5 applies our concept and algorithm to empirical data and searches for noticeable patterns in implied volatilities. Section 4.6 concludes and discusses the implications of our results for energy market practice and future research.

⁶The theoretical motivation for preferring an additive over a multiplicative setup is discussed in Section 4.3.3.

⁷Because natural gas markets are less liquid than, for example, stock markets (see Felix et al., 2013), the identification of inexpensive and expensive deals is particularly important for energy traders.

4.2 Swing option theory

4.2.1 Contract definition

We start with the formal Jaillet et al. (2004) and Barrera-Esteve et al. (2006) swing contract definition. Subject to periodic constraints, a swing option provides flexibility with respect to when and how much natural gas is purchased. The swing option holder can repeatedly exercise the right to receive gas and is allowed to “swing”, i.e., change the quantity each time. The exercise time stamps are some fixed date indices $t = 0, \dots, T - 1$ during the delivery period, where T is referred to as the number of steps.⁸ The quantity q_t purchased at time index t is subject to the “daily” volume constraint

$$q_{\min} \leq q_t \leq q_{\max}, \quad (4.2.1)$$

where q_{\min} and q_{\max} are called the minimum and maximum daily contract quantity (DCQ). The cumulative volume $Q_t = \sum_{s=0}^{t-1} q_s$ purchased up to time $t - 1$ must satisfy the “global” constraint

$$Q_{\min} \leq Q_T \leq Q_{\max}. \quad (4.2.2)$$

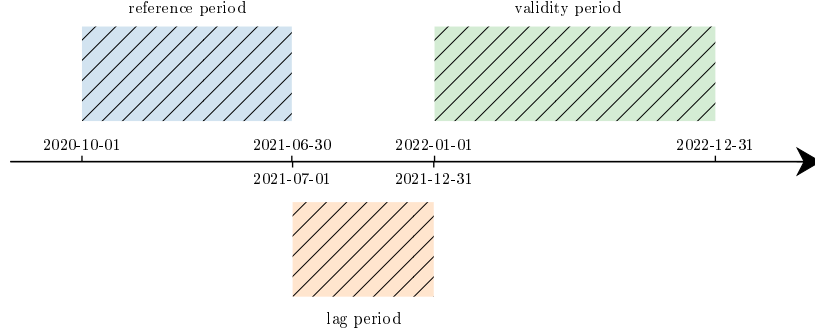
That is, the quantity purchased until the end of the contract is restricted by a minimum and maximum total contract quantity (TCQ).⁹

Swing contracts, whose delivery period spans more than one year, usually restrict the annual contract quantity (ACQ). In our study, we treat each yearly subperiod as an independent swing option such that constraints concerning the ACQ coincide with those concerning the TCQ. Note that this does not necessarily hold for any kind of swing option because, for example, make-up clauses can have a cross-cutting effect on different yearly periods (see Edoli et al., 2013).

The purchase price (or strike price) K_t for natural gas at date t can either be fixed or floating. Gas-indexed contracts rely on realized forward prices averaged over a pre-defined reference period to derive the strike of a related validity period. The simplest approach is the use of a monthly index, where the strike of each month in the delivery period is calculated as the average month-ahead forward price of the previous month. In this case, the length of each reference and validity period is one month, and there is no time-lag between reference and validity period. In a more general approach, we can define the four formula parameters a, b, c, d , where a is the length of the reference period (in months), b the time-lag (in months), c the length of the validity period (in months) and d the starting month of the strike price building formula. The delivery period is composed of M adjacent validity intervals J_1, \dots, J_M with equal lengths of c months. The intervals are sorted in

⁸The indices $0, \dots, T - 1$ correspond to a uniform discretization of the delivery period, i.e., $\tau_0 < \dots < \tau_{T-1}$, which we spare out for the sake of readability.

⁹Violating the constraint (4.2.2) can be allowed but will then be penalized when the option expires. For example, a contract might budget a one-time or a per-unit penalty which could be constant or depend on market prices at expiration (see Jaillet et al., 2004). We treat (4.2.2) as an absolute constraint, i.e., we would define the penalty for violating the terminal constraint as infinite.



For an exemplary time frame, this figure shows the relevant periods of a 9-6-12-1 gas indexation formula. The strike price is obtained by averaging all settlement forward prices observed during the 9-month reference period. These forward prices correspond to the 12-month validity period beginning 6 months after the reference period.

FIGURE 4.1: Gas indexation

ascending order by date and the first interval J_1 starts with month d . The formula strike price K_t for $t \in J_m$ (and $m = 1, \dots, M$) will be determined after the reference period I_m and then be (pathwise) constant in each validity period J_m . That is, we have $K_{J_m} := K_s = K_t$ for each $s, t \in J_m$. It is calculated by averaging the realized forward prices $F(i, J_m)$ for validity period J_m of each trading day i in the reference period I_m ,

$$K_{J_m} := \frac{1}{|I_m|} \sum_{i \in I_m} F(i, J_m) = \frac{1}{|I_m|} \sum_{i \in I_m} \frac{1}{|J_m|} \sum_{j \in J_m} F(i, j) \quad (4.2.3)$$

where $|I_m|$ ($|J_m|$) is the number of days in the reference (validity) period I_m (J_m) of a (c) months. We have a time-lag of b months between the end of the reference period I_m and the start of the validity period J_m . Figure 4.1 illustrates the relevant time periods for a gas indexation with parameters $a = 9$, $b = 6$, $c = 12$ and $d = 1$. The monthly index can be described by $a = 1$, $b = 0$, $c = 1$ and $d = 1$.

Aside from the forward market, natural gas can be purchased at the spot market. The spot price at date index t for delivery at the following day is denoted by S_t . We assume (S_t) to be memoryless or a Markov process in the probability space $\mathcal{L}^2(\Omega, (\mathcal{F}_t), \mathbb{P}; \mathbb{R})$. That is, future price dynamics depend only on the current spot price. The canonical filtration associated with $(S_t) = (S_t)_{t=0, \dots, T-1}$ is denoted by $(\mathcal{F}_t) := \sigma(S_s : 0 \leq s \leq t)$. For gas-indexed strikes, induced forward prices can then be obtained as $F(s, t) = \mathbb{E}[S_t | \mathcal{F}_s]$. In addition, we conveniently assume the interest rate to be constant and zero, although a constant discount factor could be easily introduced. Thus, the payoff of a swing option at index t is $P_t = S_t - K_t$. Again, the strike K_t may be a fixed constant, i.e., $K := K_t$ for every t , or a gas-indexed random variable, i.e., calculated via (4.2.3) and hence dependent on realized forward prices.

4.2.2 Valuation theory

Swing option valuation requires a modification of the dynamic programming techniques known from American-style options with one-time exercise right. In the original framework,

the spot price S_t is the only state variable to be considered. However, for a swing option with gas-indexed strike, the expected future payoff $\mathbb{E}[P_t|\mathcal{F}_s] = \mathbb{E}[S_t - K_t|\mathcal{F}_s]$ for $t > s$ also depends on the “forward” strike prices $K(s, J_m) := \mathbb{E}[K_{J_m}|\mathcal{F}_s]$ if $t \in J_m$.¹⁰ After the initiation of the strike-building period, they are determined by average past forward prices. Thus, S_t is not sufficient to determine future payoff expectations. The vector of price state variables has to be expanded to $\mathfrak{X}_t := (S_t, K(t, J_1), \dots, K(t, J_M))'$. Furthermore, because swing options can be exercised multiple times subject to the constraints (4.2.1) and (4.2.2), even \mathfrak{X}_t is not enough. We also have to add the purchased quantity Q_t such that the variable set (\mathfrak{X}_t, Q_t) forms the dynamics of the stochastic control framework as a discrete-time Markov chain which is controlled by local consumption (q_t) .

That said, the fair value v of a swing contract can be obtained by maximizing its expected payoff over the delivery period. Formally, we have to solve

$$\begin{aligned} v := \sup_{(q_t) \in \mathcal{L}^2} \mathbb{E} \left[\sum_{t=0}^{T-1} P_t(\mathfrak{X}_t) \cdot q_t \right] \\ \text{subject to } q_{\min} \leq q_t \leq q_{\max} \text{ almost surely,} \\ Q_{t+1} = Q_t + q_t, q_0 = 0, \\ Q_{\min} \leq Q_T \leq Q_{\max} \text{ almost surely,} \end{aligned} \quad (4.2.4)$$

where the local consumption q_t is a real-valued, square-integrable and \mathcal{F}_t -measurable random variable for every t . Each consumption control $(q_t) = (q_t)_{t=0, \dots, T-1}$ is hence an element of the functional space $\mathcal{L}^2 := \mathcal{L}^2(\Omega, (\mathcal{F}_t), \mathbb{P}; \mathbb{R})$. A feasible control additionally satisfies the local and global constraints (4.2.1) and (4.2.2) almost surely. The set of feasible controls is denoted by \mathcal{Q} , i.e.,

$$\begin{aligned} \mathcal{Q} := \{ (q_t) \in \mathcal{L}^2 : q_{\min} \leq q_t \leq q_{\max} \text{ for all } t = 0, \dots, T-1 \text{ a.s.} \\ \text{and } Q_{\min} \leq Q_T \leq Q_{\max} \text{ a.s.} \}. \end{aligned} \quad (4.2.5)$$

An optimal control $(q_t^*) \in \mathcal{Q}$ fulfills the equation

$$v = \mathbb{E} \left[\sum_{t=0}^{T-1} P_t \cdot q_t^* \right] \quad (4.2.6)$$

under risk-neutral probability. The set \mathcal{Q}^* of all optimal controls is nonempty.¹¹

¹⁰We add the label “forward” to the strike prices because their definition is close to the fact that forward prices can be seen as conditional expectations under risk-neutral probability, i.e., $F(s, t) = \mathbb{E}[S_t|\mathcal{F}_s]$.

¹¹A priori, we have to ensure the existence of an optimal exercise control. In some settings, the supremum of (4.2.4) may be unbounded or cannot be associated with a feasible control.

However, \mathcal{Q} is a bounded and closed subset of $\mathcal{L}^2 \cap \mathcal{L}^\infty$ as it takes values in $[q_{\min}, q_{\max}]$ and is constrained by (4.2.2). According to the Banach-Alaoglu theorem (see Brezis, 2010, Theorem 3.16), it is hence a weak compact subset of \mathcal{L}^2 . Further, note that the objective function $\mathcal{J} : \mathcal{Q} \rightarrow \mathbb{R}$, defined by $\mathcal{J}(q) = \mathbb{E} \left[\sum_{t=0}^{T-1} P_t \cdot q_t \right]$, is weakly continuous and thus preserves compactness. That is, the image of \mathcal{Q} under the

We now present several steps to simplify the valuation problem (4.2.4). In the first step, we follow Bardou et al. (2010) by splitting up the original swing option into a swap and a normalized swing option with local constraint $\tilde{q}_t \in [0, 1]$.¹² This has the consequence that the value v of the swing contract subject to the original constraints (4.2.1) and (4.2.2) can be obtained via

$$v = (q_{\max} - q_{\min})\tilde{v} + q_{\min} \sum_{t=0}^{T-1} \mathbb{E}[P_t], \quad (4.2.7)$$

where \tilde{v} denotes the value of a swing contract subject to the local constraints $\tilde{q}_t \in [0, 1]$ and the global constraints $\tilde{Q}_t \in [\tilde{Q}_{\min}, \tilde{Q}_{\max}]$ with

$$\tilde{Q}_{\min} = \frac{Q_{\min} - Tq_{\min}}{q_{\max} - q_{\min}} \quad \text{and} \quad \tilde{Q}_{\max} = \frac{Q_{\max} - Tq_{\min}}{q_{\max} - q_{\min}}. \quad (4.2.8)$$

It also allows us to assume in the following that, with no loss of generality, $q_t \in [0, 1]$.

In the second step, we exploit the Markov property of (\mathfrak{X}_t, Q_t) which implies that an optimal exercise decision on day t can be made solely based on the state variables Q_t and \mathfrak{X}_t . This is because the expected future payoff on a day s can be written as a function of these variables:

$$\mathfrak{V}_s(Q_s, \mathfrak{X}_s) := \max_{(q_t) \in \mathcal{Q}} \mathbb{E} \left[\sum_{t=s}^{T-1} P_t(\mathfrak{X}_t) \cdot q_t \right] \quad (4.2.9)$$

subject to $Q_{s+1} = Q_s + q_s$ almost surely.

Bellman (1957) shows that, under such circumstances, the dynamic optimization problem can be broken down into simpler subproblems. For each exercise date, a policy function can be formulated which determines the controls as a function of the states. Furthermore, Bellman (1952) points out that an optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision. This optimality principle allows us to rewrite the original problem as a backward dynamic programming problem:

$$\begin{aligned} \mathfrak{V}_t(Q_t, \mathfrak{X}_t) &= \max_{q_t \in [0, 1]} P_t(\mathfrak{X}_t) \cdot q_t + \mathbb{E}[\mathfrak{V}_{t+1}(Q_t + q_t, \mathfrak{X}_{t+1}) | \mathcal{F}_t] \\ \mathfrak{V}_T(Q_T, \mathfrak{X}_T) &= \begin{cases} 0, & \text{if } Q_{\min} \leq Q_T \leq Q_{\max}, \\ -\infty, & \text{otherwise.} \end{cases} \end{aligned} \quad (4.2.10)$$

mapping \mathcal{J} is also compact, i.e., a closed and bounded subset of \mathbb{R} . Consequently, the function \mathcal{J} attains its maximum.

¹²This can be justified by the fact that the \mathcal{F}_t -measurable random variable q_t is $[q_{\min}, q_{\max}]$ -valued if and only if there exists a $[0, 1]$ -valued \mathcal{F}_t -measurable random variable \tilde{q}_t such that $q_t = q_{\min} + (q_{\max} - q_{\min})\tilde{q}_t$.

The procedure starts at the option's expiration date and steps backward in time while considering two additional dimensions: spot price and consumption level. At each "backward induction" step t , the decision is made by maximizing the future payoff over using or skipping an exercise right. For this, we have to calculate $\mathfrak{V}_t(Q_t, \mathfrak{X}_t)$ for every possible consumption level Q_t and every price state \mathfrak{X}_t , given all the values $\mathfrak{V}_s(Q_s, \mathfrak{X}_s)$ for $s > t$.

To make this manageable, the third step uses the important result that swing contracts can be optimally exercised in digital fashion, i.e., either at the highest or lowest feasible level of q_t . This feature allows us to consider only two possibilities in decision making and only a finite grid of relevant consumption levels for finding the fair option value.

Proposition 2. *Consider the stochastic dynamic decision problem (4.2.4) with $q_{\min} = 0$ and $q_{\max} = 1$. If $Q_{\min}, Q_{\max} \in \mathbb{N}_0$, then there exists an optimal policy $(q_t^*) \in \mathcal{Q}^*$ of digital type, i.e., q_t^* is $\{0, 1\}$ -valued for all $t = 0, \dots, T - 1$.*

Proof. See Bardou et al. (2010). □

Based on this proposition, we can demand $q_t \in \{0, 1\}$ for every t if both Q_{\min} and Q_{\max} are natural numbers. As a consequence, the variable $Q_t \in \mathbb{N}_0$ can be interpreted as the cumulative number of used exercise rights up to day $t - 1$. To refer to the digital case, we often write r_t instead of Q_t . We further call $r := Q_{\max} \in \mathbb{N}_0$ the number of exercise rights and $l := Q_{\min} \in \mathbb{N}_0$ the number of exercise liabilities of a swing option. We refer to the ratio l/r as the take-or-pay (ToP) factor. If $l = r$, we speak of a 100% ToP contract. Figure 4.2 depicts the digital state grid of Q_t for an exemplary option with $l = 10$ and $r = 20$.

With this final simplification, the swing option valuation problem becomes quite similar to that of American-style options. In both cases, an exercise decision is made by comparing the possible instantaneous payoff and the continuation value of the option, i.e., the expected future payoff in the case of retaining the exercise right (see Haugh and Kogan, 2004; Ibáñez and Zapatero, 2004; Longstaff and Schwartz, 2001). However, in contrast to having only one exercise right, the continuation value \mathcal{C}_t^ϱ for a swing option at time t is the expected future payoff with $\varrho = r_{t+1}$ used rights up to and including day t :

$$\mathcal{C}_t^\varrho(\mathfrak{X}_t) := \mathbb{E}[\mathfrak{V}_{t+1}(\varrho, \mathfrak{X}_{t+1}) | \mathcal{F}_t]. \quad (4.2.11)$$

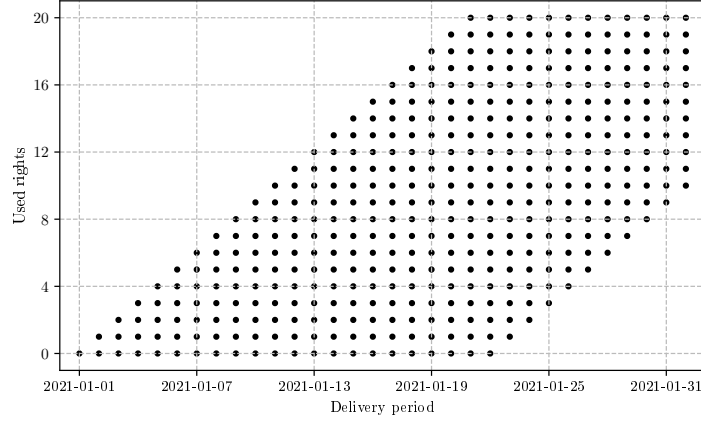
When the backward induction in the simplified framework is completed, one digital optimal policy (q_t^*) can be derived via forward induction

$$q_t^* = \begin{cases} 1, & \text{if } P_t + \mathcal{C}_t^{r_{t+1}} \geq \mathcal{C}_t^{r_t} \\ 0, & \text{otherwise} \end{cases} \quad (4.2.12)$$

starting at $t = 0$ with $r_0 = 0$.

4.2.3 Sensitivity analysis

Sensitivity analysis is the foundation for hedging option portfolios against market risks (i.e., changes of value-determining parameters such as price or volatility of the underlying). In our study, we have a different objective. We need an analytical formula for first-order



For an option with 10 exercise liabilities and 20 exercise rights in the delivery period January 2021, this figure plots the possible daily numbers of used exercise rights Q_t up to $t - 1$.

FIGURE 4.2: Digital consumption state grid

sensitivities to speed up the root-finding process for the determination of implied volatilities.

We are interested in the local behavior of the value function in the neighborhood of model parameters. Let $p \in \mathbb{R}$ denote a parameter of the spot price model such that the mapping $p \mapsto S_t(p)$ is differentiable for every $t = 0, \dots, T - 1$ and both S_t and its derivative $\frac{\partial S_t}{\partial p}$ have a distribution absolutely continuous with respect to the Lebesgue measure. Furthermore, their density function with respect to the Lebesgue measure is assumed to be bounded. The same assumptions have to hold for $p \mapsto K_t$. They are automatically fulfilled if K_t is constant. For a gas-indexed strike price, $p \mapsto F(s, t)$ has to satisfy the stated requirements for every $s \leq t$. In this context, the following proposition holds.

Proposition 3. *Consider the stochastic dynamic decision problem (4.2.4) and let $p \in \mathbb{R}$ be a parameter of the price model such that the mapping $p \mapsto P_t(p)$ is differentiable for every $t = 0, \dots, T - 1$ and both P_t and its derivative $\frac{\partial P_t}{\partial p}$ have a distribution absolutely continuous with respect to the Lebesgue measure. Further, assume that their density function with respect to the Lebesgue measure is bounded. Then, the optimal value function $v(p)$ is differentiable at almost every $p \in \mathbb{R}$. At the points p where the derivative exists it is given by*

$$v'(p) = \frac{\partial v(p)}{\partial p} = \mathbb{E} \left[\sum_{t=0}^{T-1} \frac{\partial P_t}{\partial p} \cdot q_t^* \right] = \mathbb{E} \left[\sum_{t=0}^{T-1} \left(\frac{\partial S_t}{\partial p} - \frac{\partial K_t}{\partial p} \right) \cdot q_t^* \right], \quad (4.2.13)$$

where $q^* \in \mathcal{Q}^*$ is an arbitrary optimal policy.

Proof. See Bonnans et al. (2012). □

This result justifies a pathwise derivative approach for the unbiased estimation of first-order sensitivities which has been developed in a study of simulation techniques by Broadie and Glasserman (1996) and can be implemented with little additional computation efforts.

It is important to stress that (4.2.13) is not an outcome of the Leibniz integral rule (i.e., interchanging differentiation and taking expectations). It is not allowed in our context because the optimal exercise strategy might change discontinuously on sample paths due to parameter changes. Instead, the optimality of q^* is crucial for the result. In the spirit of the celebrated first-order condition in mathematical optimization, toggling between p and $p + \Delta p$ may cause significant adjustments of the optimal policy q^* but its impact on the optimal value is actually negligible. (4.2.13) does not hold for general exercise strategies (see Chen and Liu, 2014).

This line of argumentation does not rule out second-order effects related to changes in the optimal policy. Because changes in optimal policy cannot be expressed analytically in general, there is no way of deriving a pathwise derivative formula for second-order sensitivities. Nevertheless, we can study the properties of the derivatives. The following proposition, derived from Kohrs et al. (2019), shows that, when using a common additive or multiplicative model to describe the dynamics of the spot price process (S_t) , convexity of the value function emerges for affine parameters of the spot price function.

Proposition 4. *Consider the stochastic dynamic decision problem (4.2.4) and let $p \in \mathbb{R}$ be a parameter of the price model. Assume \mathfrak{X}_t may be written in the form*

$$\mathfrak{X}_{t+1} = f_{t,W_t}(\mathfrak{X}_t, p), \quad 0 \leq t \leq T - 2, \quad (4.2.14)$$

where \mathfrak{X}_0 is deterministic, (W_t) is a sequence of independent squared integrable random variables and $f_{t,w}$ is an affine measurable mapping for fixed w . Then, the value function $v(p)$ is convex and hence twice differentiable almost everywhere.

Proof. See Appendix C. □

We use this result later on to ensure convergence of our root-finding procedure. It holds only for certain price model choices which we discuss in the following.

4.3 Price model

4.3.1 General setup

For the valuation of swing options with gas-indexed strike prices, we have to look at the joint dynamics of the spot price process (S_t) and the forward curve $(F(s, t))$. Because the strike price K_t may depend on past forward prices, modeling only spot prices S_t is not sufficient. Luckily, any specification of the forward curve dynamics implies a process for the spot price because we can derive the spot price by simply taking the t -expiring forward contract at time t , i.e., $S_t := F(t, t)$ (see Clewlow and Strickland, 2000). Therefore, our starting point is the stochastic evolution of the price forward curve (PFC).

In a risk-neutral world, investors price all claims by the expected future value discounted with the riskless rate (see Haugh and Kogan, 2004). Because forward contracts do not require an initial investment, in a risk-neutral world, the expected change in the forward price must be zero. Furthermore, because we wish to translate a given option value to one implied volatility, we assume that forward curve dynamics are driven by a single volatility

factor $\sigma(s, t)$. That is, we assume the returns $R(s, t)$ of the forward price $F(s, t)$ to follow the process

$$R(s, t) = \sigma(s, t)dW_s, \quad (4.3.1)$$

where (W_t) is a Brownian motion. To become applicable, the functional form of the volatility factor and the return type both need to be specified. To incorporate the continuous price model into our discrete time valuation framework, we use standard Euler-Maruyama discretization.

4.3.2 Volatility function

The simplest form of volatility function is a constant, i.e., $\sigma(s, t) = \sigma$. However, it is a suboptimal choice because it ignores certain aspects of swing option flexibility. To understand this, note that, in the case of constant volatility and relative returns, each forward contract $F(\cdot, t)$ follows the famous Black (1976) model, which has been introduced to price plain vanilla options on commodity forward contracts. It is also well-suited for the valuation of certain kinds of swing options which are strips of daily options in nature. We are talking here about swing options with zero liabilities and a number of exercise rights equal to the number of steps (i.e., $0 = l < r = T$). We refer to this kind of optionality as “quantity flexibility” because it allows the purchase of more or less of the underlying commodity. However, a model with constant volatility function is not suitable to value the “structure flexibility” of swing options, i.e., the opportunity to restructure exercise rights in response to the evolution of forward price time-spreads until maturity. Take a 100% ToP swing option (with constraint $l = r$) as an extreme example. It exhibits no quantity flexibility but still offers the right to choose the date of gas consumption if the number of exercise rights is less than the number of steps (i.e., $r < T$). A swing option of this type would exhibit no extrinsic value in a constant volatility model because parallel shifts are the only form of modeled forward curve movement. Such shifts do not provide volatility in forward price time-spreads which drive the value of structure flexibility.

Another fact that could be considered when selecting the volatility function is that natural gas markets exhibit seasonality in form of summer-winter spreads of time-varying magnitude (see Chiarella et al., 2016; Kohrs et al., 2019).¹³ Modeling summer-winter spread movements via trigonometric functions may help to assign a monetary value to structure flexibility. However, because we are limited to one factor, incorporating solely this time-of-maturity effect (and ignoring more important features) leads to misbehaving option time values.

Interestingly, a volatility function which captures a key characteristic of natural gas markets – a decreasing volatility term structure – values quantity and structure flexibility. In contrast to stock and bond markets, production and storage are two major drivers of supply in gas markets. As gas providers are frequently confronted with high demand, supply overcapacity or production deficiency, we can observe high volatility in spot and

¹³In other areas, such as power markets, where swing options are traded as well, summer-winter seasonality is not strongly pronounced (see Weron, 2014).

short-term forward prices. At the same time, we expect a balance between long-run supply and demand. Hence, long-term forward prices are subject to less fluctuation around a rather stable equilibrium price level, which reflects expectations of market production and costs in the long run. Consequently, we need a volatility function which declines with increasing time-to-maturity.

When looking for a function with such features, we are actually quite limited. To obtain a Markovian spot price process, the volatilities of forward prices must have a negative exponential form (see Carverhill, 1994). Therefore, the function

$$\sigma(s, t) = \sigma e^{-\vartheta \cdot (t-s)} \quad (4.3.2)$$

is a natural and common choice (see Clewlow and Strickland, 1999; Fleten et al., 2012). It obviously fits a decreasing volatility term-structure. A higher attenuation rate ϑ leads to lower volatilities until maturity and hence to a lower value of quantity flexibility. In addition, the function generates volatility in forward price spreads because forward contracts near maturity exhibit higher volatility than long-term contracts. The magnitude of these forward spreads is directly linked to the attenuation rate ϑ . Starting from $\vartheta = 0$, a higher ϑ leads to higher forward spreads and hence a higher value of structure flexibility. Thus, in summary, the choice of ϑ involves a trade-off between giving quantity or structure flexibility a higher value. In Section 4.5.2, we discuss this issue in more detail.

Note that the attenuation of the volatility of the forward curve is also directly related to a mean-reverting behavior of the spot price (see Propositions 5 and 6). That is, as often observed in financial markets, prices tend to return to an equilibrium level over time (see Balvers et al., 2000; Poterba, 1988). As far as gas markets are concerned, several factors drive a strong mean-reversion. Its rate depends on, for example, the speed at which the market supply reacts to certain events or how quickly the effects of these events wear off.

4.3.3 Return specification

In general, the returns $R(s, t)$ can be defined in relative or absolute form. Choosing between the two alternatives

$$R(s, t) = \frac{dF(s, t)}{F(s, t)} \quad \text{and} \quad R(s, t) = dF(s, t)$$

leads either to a multiplicative model (with log-normal forward prices and implied volatilities measured in percent) or an additive model (with normal forward prices and implied volatilities measured in monetary units).¹⁴

There are three theoretical reasons for choosing the absolute definition. First, in a constant relative volatility model, a rising price of the underlying goes along with increasing absolute volatility (in terms of the product of relative volatility and price). This causes a higher value of flexibility and hence a higher option value. Because the degrees of

¹⁴Price normality is not necessarily problematic because energy markets can exhibit negative prices (see Zhou et al., 2016) and additive models can perform quite well empirically (see Versluis, 2006).

freedom of our price model are limited (i.e., there is only one volatility parameter to be set) and individual PFCs might exhibit different price levels (due to short-term or seasonal effects), it is important to wipe out this price level effect on the value of flexibility. A constant relative volatility parameter would lead to higher and lower simulated absolute price fluctuations along the time axis; a constant absolute volatility model would not.

Second, the choice of return definition has a crucial influence on option sensitivities and thus on the risk exposures implied by the resulting option price model. To illustrate this important point, we compare the induced spot price processes of multiplicative and additive PFC models and discuss their implications.

Proposition 5 (Multiplicative price model). *Consider a multiplicative single-factor PFC model with volatility function (4.3.2), i.e.,*

$$\frac{dF(s, t)}{F(s, t)} = \sigma e^{-\vartheta \cdot (t-s)} dW_s, \quad (4.3.3)$$

with its solution

$$F(s, t) = F(0, t) \cdot \exp \left(-\frac{1}{2} \int_0^s |\sigma e^{-\vartheta \cdot (t-u)}|^2 du + \int_0^s \sigma e^{-\vartheta \cdot (t-u)} dW_u \right). \quad (4.3.4)$$

This forward price model induces the spot price process

$$\frac{dS_t}{S_t} = \frac{\partial F(0, t)}{\partial t} dt + \vartheta (\log F(0, t) - \log S_t) dt + \frac{\sigma^2}{4} (1 - e^{-2\vartheta t}) dt + \sigma dW_t. \quad (4.3.5)$$

The price forward curve can be reconstructed with the formula

$$F(s, t) = F(0, t) \left(\frac{S_s}{F(0, s)} \right)^{\exp(-\vartheta(t-s))} \exp \left(-\frac{\sigma^2}{4\vartheta} e^{-\vartheta t} (e^{2\vartheta s} - 1) (e^{-\vartheta t} - e^{-\vartheta s}) \right). \quad (4.3.6)$$

Proof. See Clewlow and Strickland (1999). □

Proposition 6 (Additive price model). *Consider an additive single-factor PFC model with volatility function (4.3.2), i.e.,*

$$dF(s, t) = \sigma e^{-\vartheta \cdot (t-s)} dW_s, \quad (4.3.7)$$

with its solution

$$F(s, t) = F(0, t) + \int_0^s \sigma e^{-\vartheta \cdot (t-u)} dW_u. \quad (4.3.8)$$

This forward price model induces the spot price process

$$dS_t = \frac{\partial F(0, t)}{\partial t} dt + \vartheta (F(0, t) - S_t) dt + \sigma dW_t. \quad (4.3.9)$$

The price forward curve can be reconstructed with the formula

$$F(s, t) = F(0, t) + e^{-\vartheta(t-s)} \cdot (S_s - F(0, s)). \quad (4.3.10)$$

Proof. See Appendix C. \square

The delta position of a swing option is defined as the value sensitivity with respect to shifts of forward prices $F(0, t)$. More formally, we consider the mapping $F(0, t) \mapsto v$ and define the delta position δ_t for maturity t as the partial derivative of v with respect to $F(0, t)$. Following Proposition 3, this sensitivity can be calculated as

$$\delta_t := \frac{\partial v}{\partial F(0, t)} = \mathbb{E} \left[\left(\frac{\partial S_t}{\partial F(0, t)} - \frac{\partial K_t}{\partial F(0, t)} \right) \cdot q_t^* \right]. \quad (4.3.11)$$

To delta-hedge a swing option, one has to take a position in the amount of δ_t in each forward contract $F(0, t)$ of the delivery period $t = 0, \dots, T - 1$. The overall delta position δ is defined as the sum of all daily delta positions. It can alternatively be defined as the value sensitivity with respect to parallel shifts of the entire forward curve. That is, we have

$$\delta = \sum_{t=0}^{T-1} \delta_t = \mathbb{E} \left[\sum_{t=0}^{T-1} \left(\frac{\partial S_t}{\partial F(0, t)} - \frac{\partial K_t}{\partial F(0, t)} \right) \cdot q_t^* \right]. \quad (4.3.12)$$

This shows that the delta position is influenced by the way forward curve shifts affect the spot prices. For the multiplicative price model (4.3.3), we have

$$S'_t(F(0, t)) := \frac{\partial S_t}{\partial F(0, t)} = \exp \left(-\frac{1}{2} \int_0^t |\sigma e^{-\vartheta \cdot (t-u)}|^2 du + \int_0^t \sigma e^{-\vartheta \cdot (t-u)} dW_u \right), \quad (4.3.13)$$

whereas the additive price model (4.3.7) yields

$$S'_t(F(0, t)) = 1. \quad (4.3.14)$$

Consequently, for fixed-strike swing options, the delta position δ_t in an additive price model is equal to the expected exercise quantity $\mathbb{E}[q_t^*]$, which is similar to the plain vanilla option pricing model of Bachelier (1900).¹⁵ In contrast, the overall delta position in a multiplicative model is generally higher than its expected total exercise quantity.¹⁶ This

¹⁵For a detailed technical analysis of this model, see Schachermayer and Teichmann (2008).

¹⁶To see more intuitively why the multiplicative model commands higher risk exposure than the additive one, note the following. $\mathbb{E}[S'_t(F(0, t))] = 1$ holds in both models. However, a sample realization ${}^n S'_t(F(0, t)) > 1$ on a path n indicates a relatively high spot price such that exercising is advantageous. There is therefore a positive correlation between the derivative $S'_t(F(0, t))$ and q_t^* which results in $\mathbb{E}[S'_t(F(0, t)) \cdot q_t^*] \geq \mathbb{E}[q_t^*]$. From another perspective, we can argue that the multiplicative delta position captures value gains related to a rise in absolute volatility (see discussion above).

is in line with the Black (1976) model, where the delta of a call option is higher than the exercise probability.

Portfolio managers might prefer a price model where the total delta hedge does not exceed the total contract quantity. For 100% ToP options, where the time of consumption is flexible, the total exercise quantity is predetermined. Nevertheless, a multiplicative price model would suggest to hedge an overall quantity above the total contract quantity. Figure 4.3 illustrates this behavior for a selected swing option. Here, the delta in the multiplicative price model adds up to 123% of the total contract quantity. An analogue effect occurs in gas storage valuation because 100% ToP options can be interpreted as specific types of storage options (see Boogert and de Jong, 2008, 2011; Lai et al., 2010).

Besides the delta argument, which can be an intuitive guide in the return model decision,¹⁷ a third and final theoretical property of the additive model is particularly valuable in our context. The sensitivity with respect to the model volatility is much simpler than in the multiplicative model. In the additive model, we have a linear mapping $\sigma \mapsto F(s, t)$ for every $s \leq t$ leading to a very comfortable computational framework. In Section 4.4.2, we shed more light on this important aspect.

4.4 Implied volatility algorithm

We now introduce our algorithm to translate swing options quotes, which can be embedded into the framework of Section 4.2, into implied volatility parameters σ of the additive price model defined in Section 4.3. To handle the complexity of swing option valuation, we use the Monte Carlo techniques described in Section 4.4.1. That is, we adapt the least squares Monte Carlo algorithm of Longstaff and Schwartz (2001) to general swing contracts. We then utilize the Newton-Rhapson iteration as a root-finding algorithm for implied volatility calculation, as described in Section 4.4.2. Because a first-order derivative is required in every iteration step, we provide an analytical formula to calculate the vega of swing options with no additional computational burden.¹⁸ Finally, we deduce the unconditional convergence of the iteration via derivative property arguments.

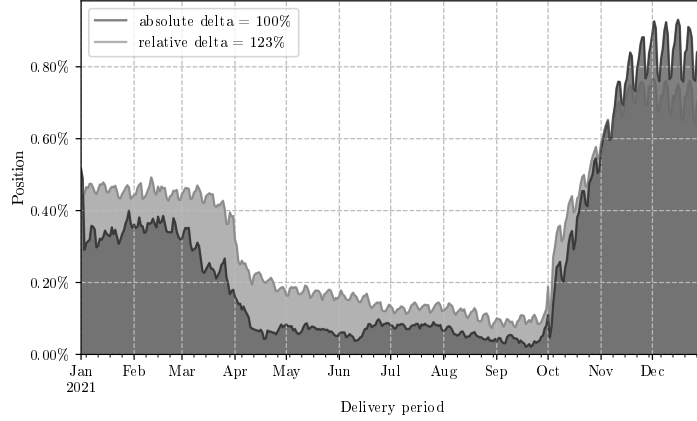
4.4.1 Option valuation algorithm

Longstaff and Schwartz (2001) proposed a simulation-based approach for pricing American-style options by using least squares regression to estimate their continuation values. Least squares Monte Carlo (LSMC) has become a literature standard (see Hanfeld and Schlüter, 2017; Kohler, 2010) and can be considered a state-of-the-art approximate dynamic programming approach. It is applicable to a wide range of path-dependent and multi-factor settings typically encountered when studying complex energy options (see Nadarajah et al., 2017).

We can transfer this approach to swing options with $Q_{\min}, Q_{\max} \in \mathbb{N}_0$ which are exercised in a digital fashion according to Proposition 2. Consider an exercise decision on

¹⁷It should not be considered a definitive knockout argument because even delta hedges based on additive models can be flawed.

¹⁸This means that we can use results from the option valuation process to compute vega.



For a selected fixed-strike 100% ToP swing option, this figure compares the delta position (given as a percentage of the total contract quantity) of the additive price model (4.3.7) to that of the multiplicative model (4.3.3). The delta of the multiplicative model sums up to 123% and hence is not “volume neutral”.

FIGURE 4.3: Delta in additive and multiplicative models

some day $t \in \{0, \dots, T-1\}$ with a given number of used exercise rights $r_t \in \mathbb{N}_0$ up to day $t-1$ and observed price state \mathfrak{X}_t . To decide whether to use one of the remaining exercise rights (if there is one), we have to obtain proper estimates for the continuation values (4.2.11), with $\varrho \in \{r_t, r_t + 1\}$, because they cannot be calculated analytically as a non-trivial conditional expectation. For this purpose, we introduce an orthonormal basis $\phi = (\phi_j)_{j \in \mathbb{N}}$ of $\mathcal{L}^2(\Omega, \mathcal{F}_t, \mathbb{P})$. Then, the continuation value can be represented as

$$\mathcal{C}_t^\varrho(\mathfrak{X}_t) = \sum_{j=1}^{\infty} \alpha_j^{(\varrho, t)} \phi_j(\mathfrak{X}_t) = \langle \alpha^{(\varrho, t)}, \phi(\mathfrak{X}_t) \rangle,$$

where $\alpha^{(\varrho, t)} = (\alpha_j^{(\varrho, t)})_{j \in \mathbb{N}}$ (see Brezis, 2010). Two features of LSMC make this representation applicable in a practical setting. First, only a finite number of basis functions is drawn from the set $(\phi_j)_{j \in \mathbb{N}}$ to estimate \mathcal{C}_t^ϱ as a linear combination thereof. Second, the coefficients $\alpha^{(\varrho, t)}$ are estimated by least squares regression based on information from all simulated paths.

Formally, we consider a sample of daily price states $({}^n\mathfrak{X}_t)_{0 \leq t \leq T-1}$ with $n = 1, \dots, N$ referring to one sample path. Furthermore, we have the exercise decision on day t for a given number of used exercise rights r_t . Our goal is then to derive an optimal decision $\hat{q}_t^*({}^n\mathfrak{X}_t)$ for each path n using (4.2.12). Because we go backwards in time, we have already estimated optimal decisions $\hat{q}_s^*({}^n\mathfrak{X}_s)$ on each path n and for every possible number of used exercise rights r_s on $s = t+1, \dots, T-1$. For fixed $\varrho \in \{r_t, r_t + 1\}$, we therefore obtain the future payoff on each path n using the optimal decisions $(\hat{q}_s^*({}^n\mathfrak{X}_s))_{s \geq t+1}$ as

$${}^nC_t^\varrho = \sum_{s=t+1}^{T-1} P_s({}^n\mathfrak{X}_s) \hat{q}_s^*({}^n\mathfrak{X}_s),$$

which we treat as an observation of the dependent variable $\mathfrak{V}_{t+1}(\varrho, \mathfrak{X}_{t+1})$. These observations can then be used to perform the least squares regression

$${}^n\mathcal{C}_t^{\varrho} = \sum_{j=1}^J \alpha_j^{(\varrho,t)} \phi_j({}^n\mathfrak{X}_t) + \varepsilon_n$$

with error term ε_n . This yields the estimated coefficients $(\hat{\alpha}_j^{(\varrho,t)})_{j=1,\dots,J}$, which finally deliver estimates $\hat{\mathcal{C}}_t^{\varrho}({}^n\mathfrak{X}_t) = \sum_{j=1}^J \hat{\alpha}_j^{(\varrho,t)} \phi_j({}^n\mathfrak{X}_t)$ of the continuation values.

A crucial issue of every Monte Carlo approach is convergence. While the LSMC algorithm is rather intuitive, proving its convergence is far from trivial. The reason for this lies in the variety of potential error sources. For example, when estimating the conditional expectation $\mathcal{C}_t^{\varrho}(\mathfrak{X}_t)$, we use a linear combination of a finite set of J basis functions, i.e., $\mathcal{C}_t^{\varrho}(\mathfrak{X}_t) \approx {}^J\mathcal{C}_t^{\varrho}(\mathfrak{X}_t) := \sum_{j=1}^J \alpha_j^{(\varrho,t)} \phi_j(\mathfrak{X}_t)$. In practical settings, it is impossible to use the entire sequence $(\phi_j)_{j \in \mathbb{N}}$ such that one source of error is naturally linked to the number of used basis functions J . Additional error originates from the simulation technique. When estimating the coefficients $\alpha_j^{(\varrho,t)}$, the estimation error will depend on the number of paths N . Overall, an exercise strategy derived by means of an approximated variable might be suboptimal.

We justify our approach by previous results on LSMC convergence. For a fixed number of paths N , Longstaff and Schwartz (2001) propose to choose the number of basis functions J by increasing J until the option value no longer rises. Clément et al. (2002) provide the first complete proof of convergence under fairly general conditions. While their work considers a single-stopping problem in discrete time, Barrera-Esteve et al. (2006) show that it can be adapted to swing options. Unfortunately, Clément et al. (2002) treat the error sources outlined above separately. Their proof is based on a sequential rather than a joint limit. To ensure the convergence of the estimator in a joint limit, the relationship between N and J has to be analyzed. Glasserman and Yu (2004) take this issue into account and prove convergence of the estimator as the number of paths grows approximately exponentially with the number of basis functions. Under the requirement of a smooth payoff function, Stentoft (2004) shows that convergence to the true value is ensured, if $J \rightarrow \infty$ and $J^3/N \rightarrow \infty$.

In light of this discussion, the number of basis functions has to be chosen carefully. A low J might not capture the true shape of the continuation value function whereas a high J can lead to misbehaving regressions. Both cause poor continuation value estimates. To find an adequate J , we borrow the “greedy heuristic” of Boogert and de Jong (2011) which considers a potentially large basis but only selects a subset monitored for stability.¹⁹ In contrast, as outlined in Section 4.4.2, the number of simulation paths N is set by computational limits.

Within a finite sequence $\phi = (\phi_j)_{j=1}^J$ of basis functions, the greedy heuristic decides in a specific way which ones to include in the regression. To limit the risk of an oversized

¹⁹As we will see, this heuristic is inspired by the increasing J approach of Longstaff and Schwartz (2001).

basis, basis functions are successively added to the regressor matrix X until it misbehaves. We assume a linear relation between the continuation value $\mathcal{C} := ({}^1\mathcal{C}_t^g, \dots, {}^N\mathcal{C}_t^g)'$ and the regressor matrix X with columns $x_j = ({}^1x_j, \dots, {}^Nx_j)'$ and ${}^Nx_j := \phi_j({}^N\mathfrak{X}_t)$. That is, we have $\mathcal{C} = X\alpha^{(g,t)}$ where $\alpha^{(g,t)}$ is the coefficient vector. Consequently, $\alpha^{(g,t)}$ can be estimated via the Moore-Penrose pseudo inverse, i.e., by least squares, as

$$\hat{\alpha}^{(g,t)} = (X'X)^{-1}X'\mathcal{C}. \quad (4.4.1)$$

If $X'X$ is well-conditioned, its inverse can be computed with acceptable accuracy and we can expect stable coefficient estimates.²⁰ Therefore, starting with x_1 , we successively fill X with the vectors x_j , check the condition of $X'X$ in each step and terminate if $X'X$ is no longer well-conditioned.

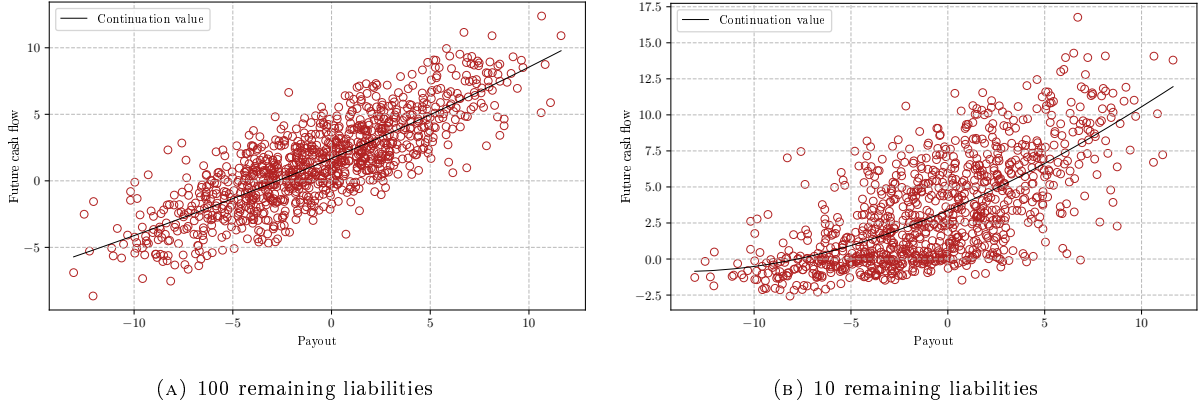
Besides determining the number of basis functions, we have to take into account the potential impact of using different types of basis functions. Moreno and Navas (2003) consider ten families and show that, for American-style options, LSMC is quite robust to the choice of basis functions. For more complex derivatives, however, there can be large differences in option values. In an application of gas storage valuation, Boogert and de Jong (2011) investigate the same families and emphasize that powers should be the preferred choice. Besides the power class, they consider other polynomial families (like Chebyshev, Hermite, Legendre or Laguerre) but find that all of them underperform and/or consume more computation time. Guided by this result, we use the power family with polynomials up to degree three. For fixed-strike swing options, the sequence of basis functions is simply (S_t, S_t^2, S_t^3) . In the case of gas-indexation, we also include the forward strike prices $K(t, J_m)$ for all validity periods J_m .

Figure 4.4 illustrates the continuation value regression for a fixed-strike swing option with a delivery period spanning 366 steps. The option provides 180 exercise rights of which 162 are mandatory. We can clearly see how the number of remaining exercise rights, available 120 steps before delivery, affects the shape of the continuation value function. Figure 4.4a shows a situation where the option holder has 100 remaining exercise liabilities such that there is little opportunity left to pass on exercising. Consequently, the continuation value function exhibits little convexity and is well approximated by the payoff function of a simple forward contract. In contrast, Figure 4.4b shows a situation in which the option holder has only 10 remaining exercise liabilities such that there is ample scope to exercise at the right time. The continuation value hence exhibits stronger convexity and resembles the value function of a plain vanilla call.

4.4.2 Root-finding algorithm

Even for European-style options under the Black and Scholes (1973) model, implied volatility calculation is not trivial because its closed-form representation does not allow obtaining the measure analytically. Initiated by Curtis Jr and Carriker (1988) and Brenner

²⁰The condition number of a matrix A is defined as $\text{cond}(A) = \|A\|_2 \cdot \|A^{-1}\|_2$. A is said to be well-conditioned if $\text{cond}(A)$ is sufficiently low, i.e., below a pre-set barrier. We set this barrier to $(\epsilon \cdot 10^6)^{-1}$, where ϵ is the machine epsilon.



This figure illustrates the continuation value regression for a selected fixed-strike swing option with a 366-step delivery period. Of its 180 exercise rights, 162 are mandatory. The future cash flows (in EUR/MWh) are plotted 120 steps before delivery and dependent on the current payout $S_t - K$ (in EUR/MWh). Subfigures (a) and (b) distinguish between situations where the option holder has 100 and 10 remaining exercise liabilities, respectively.

FIGURE 4.4: Continuation value regression

and Subrahmanyam (1988), some studies have therefore developed approximate closed-form inversion methods which apply some form of Taylor expansion to the pricing formula and then analytically invert the expansion. Unfortunately, there is no closed-form pricing formula for swing options and we have to value them via Monte Carlo techniques. Consequently, we need a solver method. One of the most efficient algorithms to estimate implied volatilities is the classic Newton-Raphson method. This procedure has been shown to work well for standard options, usually producing very accurate estimates with negligible computation time (see Kritzman, 1991), and has also been used in the context of Asian-style options (see Yang et al., 2009). In both cases, it relies on an accurate and fast computation of option vegas. Fortunately, the theoretical results of Section 4.2.3 allow us to derive a vega formula for swing options and to justify the application of the Newton-Raphson method for volatility root-finding.

The Newton-Raphson method produces successively better approximations to the roots of a real-valued function by utilizing the function's derivative. Given a quote \bar{v} for the value of a swing option with a value function $v(\sigma)$ depending on σ , we have to compute a solution of

$$v(\sigma) - \bar{v} = 0.$$

Because v is differentiable by Proposition 3, we can apply Newton-Raphson iteration, i.e.,

$$\sigma_{n+1} = \sigma_n - \frac{v(\sigma_n) - \bar{v}}{v'(\sigma_n)}. \quad (4.4.2)$$

To implement this iteration, in a first step, we have to derive a vega formula for swing options with fixed or gas-indexed strike prices. Recall that, in our additive model (4.3.7), we have

$$F(s, t) = F(0, t) + \int_0^s \sigma e^{-\vartheta(t-u)} dW_u$$

such that differentiating with respect to σ yields

$$\frac{\partial F(s, t)}{\partial \sigma} = \int_0^s e^{-\vartheta(t-u)} dW_u = \frac{F(s, t) - F(0, t)}{\sigma}. \quad (4.4.3)$$

Thus, the first-order payout derivatives are

$$\frac{\partial S_t}{\partial \sigma} = \frac{S_t - F(0, t)}{\sigma} \quad \text{and} \quad \frac{\partial K_{J_m}}{\partial \sigma} = \frac{1}{|I_m|} \sum_{i \in I_m} \frac{F(i, J_m) - F(0, J_m)}{\sigma}, \quad (4.4.4)$$

where $F(i, J_m) = \frac{1}{|J_m|} \sum_{j \in J_m} F(i, j)$. Following Proposition 3, the swing option vega can be calculated by inserting (4.4.4) in

$$v'(\sigma) = \mathbb{E} \left[\sum_{t=0}^{T-1} \left(\frac{\partial S_t}{\partial \sigma} - \frac{\partial K_t}{\partial \sigma} \right) \cdot q_t^* \right]. \quad (4.4.5)$$

In a second step, we ensure the universal convergence of the iteration (4.4.2) by using the classic result on monotone global convergence (see Deuffhard, 2011). It requires v to be twice differentiable in a right neighborhood of the root. If v is (i) increasing and (ii) convex, the sequence σ_n is monotonically decreasing to the root.

With our previous results, we can derive these necessary conditions quite easily. Starting with (i), note that, in case of a gas-indexed strike, we can compactly write

$$\frac{\partial K_{J_m}}{\partial \sigma} = \frac{K_{J_m} - K(0, J_m)}{\sigma}, \quad (4.4.6)$$

where $K(0, J_m) = \mathbb{E}[K_{J_m} | \mathcal{F}_0] = F(0, J_m)$. In case of a fixed strike price, the derivative of K_t is zero which can be expressed as $K_t - K(0, t)$. This leads us to

$$\begin{aligned} v'(\sigma) &= \mathbb{E} \left[\sum_{t=0}^{T-1} \left(\frac{S_t - F(0, t)}{\sigma} - \frac{K_t - K(0, t)}{\sigma} \right) \cdot q_t^* \right] \\ &= \frac{1}{\sigma} \sum_{t=0}^{T-1} \mathbb{E} [(S_t - K_t) \cdot q_t^*] - \mathbb{E}[S_t - K_t] \cdot \mathbb{E}[q_t^*] \\ &= \frac{1}{\sigma} \sum_{t=0}^{T-1} \text{Cov}(S_t - K_t, q_t^*). \end{aligned} \quad (4.4.7)$$

Here, the last term is non-negative because of the optimality of the exercise strategy (q_t^*) . In other words, if there is value in flexibility, the payout $S_t - K_t$ and the quantity q_t^* should be positively correlated on average. The term is zero in edge cases only, e.g., when a contract does not exhibit any flexibility in the “band delivery” configuration ($l = r = T$) or when we have a fixed-strike 100% ToP option in combination with no mean-reversion in prices ($\vartheta = 0$).

With respect to (ii), the pathwise constant first-order derivatives (4.4.3) imply that

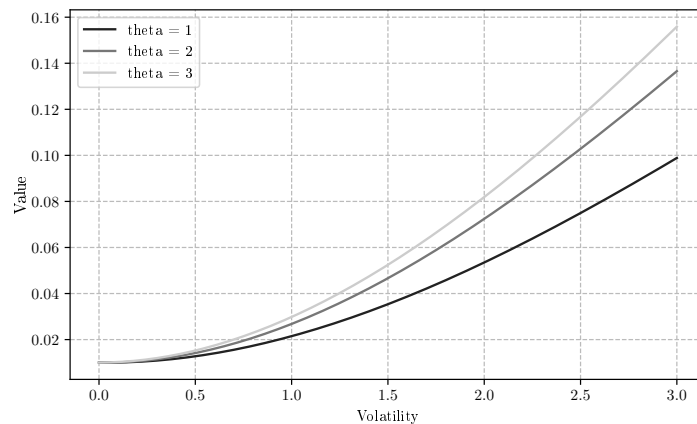
$$\frac{\partial^2 F(s, t)}{\partial \sigma^2} = 0.$$

Thus, the second-order derivative of $P(\sigma)$ is equal to 0 and, according to Proposition 4, the mapping $\sigma \mapsto v(\sigma)$ is convex. Because the second-order effect from the price model is zero, strict convexity can arise from changes in optimal exercise only, i.e.,

$$v''(\sigma) = \lim_{\Delta\sigma \rightarrow 0} \mathbb{E} \left[\sum_{t=0}^{T-1} \left(\frac{\partial S_t}{\partial \sigma} - \frac{\partial K_t}{\partial \sigma} \right) \cdot \frac{\Delta q_t^*(\sigma)}{\Delta \sigma} \right], \quad (4.4.8)$$

where $\Delta q_t^*(\sigma) = q_t^*(\sigma + \Delta\sigma) - q_t^*(\sigma)$. Recall from our discussion of Proposition 3 that jumping between $q_t^*(\sigma)$ and $q_t^*(\sigma + \Delta\sigma)$ has no significant first-order impact on the optimal value v . However, we can have changes in the optimal policy (q_t^*) , which lead to the second-order effects captured in (4.4.8). Without second-order effects from the price model, we expect the overall second-order effect to be rather small. This would result in a slightly curved value function (caused by changes in optimal exercise) and quick Newton-Raphson convergence.

These two presumptions can be confirmed by numerical testing. Figure 4.5, which plots the value function $v(\sigma)$ of a selected swing option, illustrates the expected curvature. As far as the convergence speed is concerned, note that, based on the monotone convergence argument, σ_0 should be high and lie above the root. Otherwise, the method overshoots the root in the first iterative step due to the convexity. For example, with a starting value of $\sigma_0 = 100$, the mean calibration error quickly reaches its minimum which corresponds to the number of sample paths used for Monte Carlo valuation. Specifically, Figure 4.6 shows that, for $N = 1,000$ ($N = 10,000$) sample paths, a stable calibration error below 0.01 EUR/MWh (0.001 EUR/MWh) can be reached after four iterations. With a higher N , the error can be reduced even further. However, this goes along with significantly higher computation time. Because we wish to keep the duration of our numerical analysis in Section 4.5 reasonably low, the remainder of our study sets an error tolerance of 0.01 EUR/MWh by choosing $N = 1,000$. In practice, it is up to the user to decide which levels of accuracy (and thus speed) are required for a given application.



For different choices of $\vartheta \in \{1, 2, 3\}$, this figure shows the value function $v(\sigma)$ in EUR/MWh of a selected fixed-strike 100% ToP swing option with 90 exercise rights during the summer season 2020. The volatility parameter σ is quoted in EUR/MWh p.a.

FIGURE 4.5: Convex value function

4.5 Numerical study

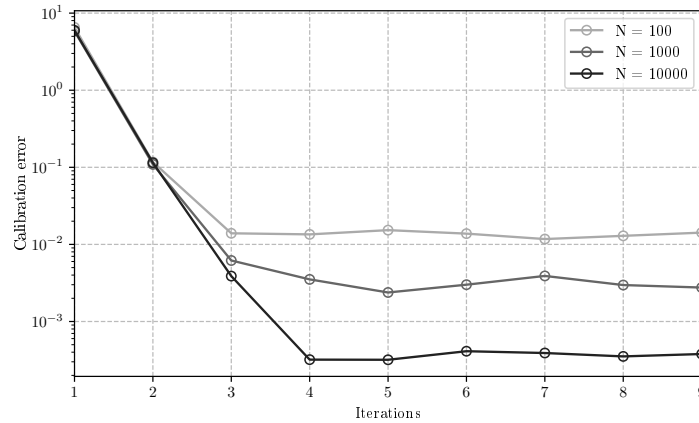
4.5.1 Preliminaries

Our numerical study is based on a quotation data set collected at the VNG Handel & Vertrieb GmbH, which is part of a large company network with more than 60 years of experience in the European energy market.²¹ The data set contains 1,897 natural gas swing option quotations in the two-year period of 2019-2020. We have 962 quotes for fixed-strike options and 935 quotes referring to swings with gas-indexation. The majority of fixed-strike option quotes correspond to 100% ToP options (869 cases). Among the gas-indexed options, a monthly index is the most common form (918 cases). Options with complex floating strikes are rare (17 cases). Besides quotations, we use PFC data with daily granularity of maturity and implied volatility data of plain vanilla options with monthly granularity of maturity based on end-of-day price quotations of natural gas forward and future prices of the Dutch Title Transfer Facility (TTF), the leading virtual trading point for natural gas in continental Europe.²²

Figure 4.7 presents the implied volatilities of the swing option quotations (bid, ask and trade prices) along the time axis. Provisionally, we have set a mean-reversion rate of

²¹For more details, see <https://vng.de/en/newsroom/publications>.

²²Wholesale gas trading at the TTF is largely conducted over-the-counter via interdealer brokers. Physical short-term gas and gas futures contracts are also traded and handled by the ICE Endex Energy Exchange and the European Energy Exchange (EEX) which recently merged with Powernext (see <https://www.eex.com/en/market-data/natural-gas/spot-market/ttf>). All curves are constructed by a curve-building procedure which yields an arbitrage-free PFC, i.e., the curves match all relevant price quotations in the market. When available, price quotations of products with daily granularity (e.g., day-ahead or balance-of-month) are supplemented by products for months, quarters, seasons, calendar and gas years. At the same time, the curves follow daily and seasonal historical shapes.



For a representative subset of our swing option data set (see Section 4.5.1) and different numbers of sample paths $N \in \{100, 1000, 10000\}$, this figure shows the average of the Newton-Raphson iteration calibration error $|v(\sigma_n) - v|$. The parameter ϑ is obtained via volatility curve fitting (see Section 4.5.3).

FIGURE 4.6: Volatility calibration errors

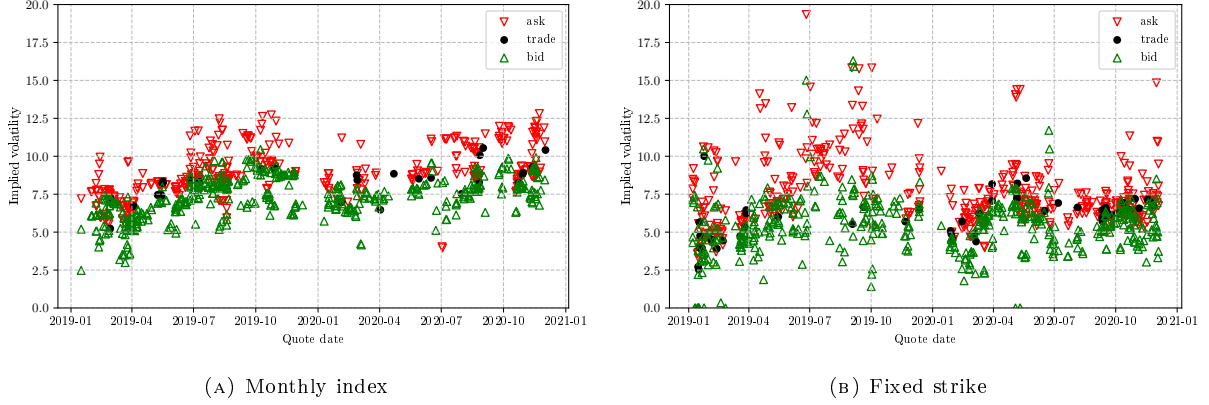
$\vartheta = 2.0$. At given quotation dates, the volatilities vary in a smaller range for swings with monthly indexation than for fixed-strike options. That is, further option specifications aside, traders of monthly index options have a stronger agreement on market volatility. Furthermore, they do not expect volatility to be as extreme as the levels priced by some fixed-strike option traders.

In the following, we deepen this overview by investigating the relationship between implied volatilities and the mean-reversion rate (see Section 4.5.2) and proposing a method to estimate the mean-reversion rate (see Section 4.5.3). In addition, we screen implied volatilities for patterns with respect to seasons (see Section 4.5.4) and moneyness (see Section 4.5.5).

4.5.2 Value drivers

In general, the shape of a swing option's value function is complex and depends on many parameters. Nevertheless, we can derive some interesting features for the two most frequently occurring types of swing options: options with monthly indexation and fixed-strike 100% ToP options. In preparation for Section 4.5.3, where we determine a suitable value for the mean-reversion rate ϑ , we focus on the relationship between ϑ and σ . Specifically, we calculate the implied volatility of each available swing option (via our root-finding procedure) for various $\vartheta \in \{1.0, 1.1, \dots, 5.8, 5.9\}$ and explain the resulting shape of the mapping $\vartheta \mapsto \sigma(\vartheta)$.

Feature 1: The value of a monthly index option is driven by the volatility of the spread between the spot and the monthly index price. In other words, monthly index options provide a market view on spot price volatility in a short future period between $t - i$ and t where i covers approximately a one-month distance. To see this without getting lost in



The figure shows the model-implied volatilities (in EUR/MWh p.a.) for our data set of swing option quotations. While Subfigure (a) refers to options with monthly indexation, Subfigure (b) looks at fixed-strike options. The mean-reversion rate is set to $\vartheta = 2.0$.

FIGURE 4.7: Implied volatilities of swing options

too much technical details, we consider the spread between the spot price $S_t = F(t, t)$ and a past forward price $F(t - i, t)$. Its variance is given by

$$\text{Var}(F(t, t) - F(t - i, t)) = \text{Var}\left(\int_{t-i}^t \sigma e^{-\vartheta(t-s)} dW_s\right) = \frac{\sigma^2}{2\vartheta}(1 - e^{-2\vartheta i}). \quad (4.5.1)$$

We speculate that the value of a monthly index option is determined by this variance. To test this hypothesis, we hold $\nu_i^2 := \text{Var}(F(t, t) - F(t - i, t))$ constant and rearrange

$$\nu_i^2 = \frac{\sigma^2}{2\vartheta}(1 - e^{-2\vartheta i}) \iff \sigma = \sqrt{\frac{2\vartheta\nu_i^2}{1 - e^{-2\vartheta i}}}. \quad (4.5.2)$$

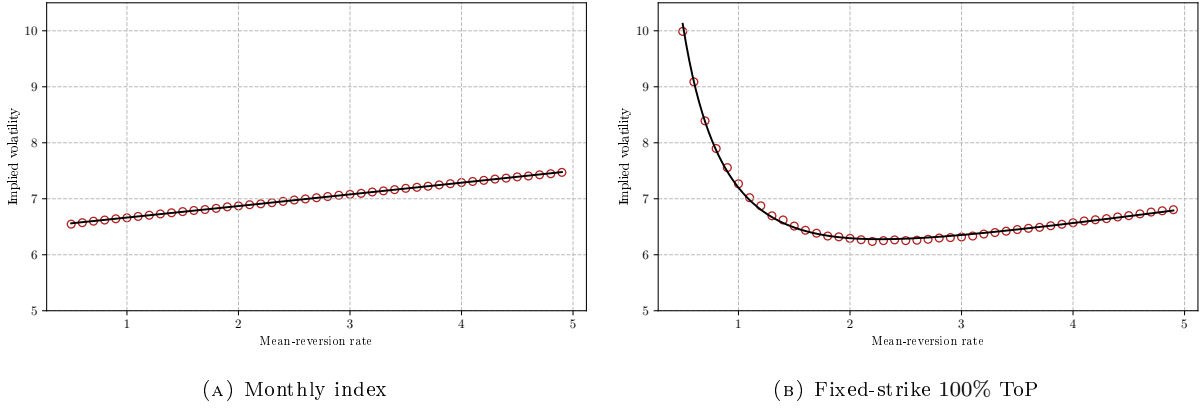
For each option, we then use a non-linear regression to search for optimal parameters ν_i and i such that (4.5.2) fits our (ϑ, σ) pairs. The results in Table 4.1 (Panel A) are in line with our hypothesis. The average MAE of 4.02e-3 over all regressions is lower than our root-finding tolerance of 1e-2 and the minimum R^2 is 0.9842. As expected, with an average of 23.7 days, the relevant time to maturity is almost one month. In this interval, prices move with an average volatility of 1.913 EUR/MWh. As illustrated in Figure 4.8a, for typical values of the mean-reversion rate ϑ , implied volatility σ is approximately an increasing linear function of ϑ .

Feature 2: The flexibility value of 100% ToP options arises from the possibility of switching exercise times during the delivery period. On a day t , we may decide to exercise on day $t + i$ instead of day $t + j$. This is close to a situation considered in Margrabe (1978), where one risky asset is exchanged for another. In his framework, asset prices follow a geometric Brownian motion and the ratio of prices has a constant volatility. Consequently, similar to Black and Scholes (1973), there is a one-to-one link between the volatility of

	Panel A: Monthly index				Panel B: Fixed-strike 100% ToP				
	MAE	R^2	ν_i	i	MAE	R^2	ν_{ij}	i	j
Min	2.27e-4	0.9842	0.622	10.7	8.47e-3	0.9134	0.0004	0.0	67.5
Avg	4.02e-3	0.9995	1.913	23.7	1.35e-1	0.9868	5.2403	1.9	424.1
Max	2.93e-2	1.0000	3.223	27.1	2.52e-0	0.9999	25.6881	75.9	934.6

Panel A of this table summarizes the results of fitting (4.5.2) to each of our monthly index options. Besides the mean absolute error (MAE, in EUR/MWh p.a.) and the coefficient of determination (R^2), we present the estimated parameters of the non-linear regressions: the volatility of the spot-forward spread (ν_i , in EUR/MWh) and the time distance of the spread (i , in days). In a similar vein, Panel B reports the results of fitting (4.5.3) to each of our fixed-strike 100% ToP options. Here, the regression parameters are the volatility of the forward spread (ν_{ij} , in EUR/MWh) and the maturities of the corresponding forward prices (i, j , in days).

TABLE 4.1: Implied volatility function fits



For two selected swing options (i.e., two quotes), this figure shows the fitted relationships between σ and ϑ . Subfigure (a) refers to the implied volatility function of a monthly index option, i.e., the baseline regression (4.5.2), whereas Subfigure (b) considers a fixed-strike 100% ToP option, i.e., regression (4.5.3).

FIGURE 4.8: Exemplary implied volatility functions

the asset price ratio and the option price. Because we use an additive price model, we investigate the difference $F(t, t+i) - F(t, t+j)$ instead. We have

$$\begin{aligned} d(F(t, t+i) - F(t, t+j)) &= dF(t, t+i) - dF(t, t+j) = \sigma e^{-\vartheta i} dW_t - \sigma e^{-\vartheta j} dW_t \\ &= \sigma(e^{-\vartheta i} - e^{-\vartheta j}) dW_t. \end{aligned}$$

In other words, $F(t, t+i) - F(t, t+j)$ follows a Brownian motion with volatility $\sigma(e^{-\vartheta i} - e^{-\vartheta j})$. We wish to check whether the value of flexibility is determined by a characteristic forward spread defined by i and j . That is, again via non-linear regressions, we test the relation

$$\nu_{ij} := \sigma(e^{-\vartheta i} - e^{-\vartheta j}) \iff \sigma = \frac{\nu_{ij}}{e^{-\vartheta i} - e^{-\vartheta j}}. \quad (4.5.3)$$

The results are reported in Table 4.1 (Panel B). With an average MAE (R^2) of 1.35e-1 (0.9868), the fitting errors are higher than for the monthly index options. Nevertheless, the implied volatility functions are well-approximated by the shape in (4.5.3). The characteristic forward spread volatility is on average 5.24 EUR/MWh and refers to a time spread from over a year to a few days ahead. A closer look at the individual regressions, which provide outcomes of the form illustrated in Figure 4.8b, reveals some key properties of the volatility function of structure flexibility options. For $\vartheta \rightarrow 0+$, the implied volatility tends to go to ∞ . That is, for $\vartheta = 0$, a fixed-strike 100% ToP option would have no flexibility value. With increasing ϑ , the implied volatility rises after an initial decline. Because $i \geq 0$, the mapping $\vartheta \mapsto \sigma$ slopes slightly upward for $\vartheta \rightarrow \infty$. Thus, there is a unique ϑ for which σ attains its minimum.

4.5.3 Mean-reversion rate

Because ϑ is a vital input for implied volatility determination, we evaluate three estimation approaches. First, historical day-ahead spot prices tend to exhibit mean-reversion such that they could be a starting point for a historic estimation. Second, because the attenuation rate of the volatility term-structure is directly related to the mean-reversion rate, we could obtain ϑ by fitting the model term-structure to implied volatilities of plain vanilla options. Finally, we might choose ϑ such that it fits empirical swing option quotes.

We start by ruling out historic estimation because it typically requires rather long time series (see Yu, 2012), which are not available in our context, and has a strong backward-looking nature. In contrast, focusing on liquid vanilla options, whose prices are set in a forward-looking manner, allows a more “up-to-date” perspective on the mean-reversion rate in dynamic markets.²³ Recalling that the variance curve of model (4.3.7) can be calculated with

$$\text{Var}(F(t, t)) = \frac{\sigma^2}{2\vartheta} (1 - e^{-2\vartheta t}), \quad (4.5.4)$$

we can obtain a daily estimate of ϑ by fitting (4.5.4), i.e., its square root, to the daily volatility term structure of at-the-money plain vanilla gas options. We refer to this easily implementable method as volatility forward curve (VFC) fitting.

Our final estimation method arises from the fact that ϑ is connected to the value of quantity and structure flexibility of swing options (see Sections 4.3.2 and 4.5.2). Pure structure flexibility is provided by fixed-strike 100% ToP options. Quantity flexibility is strongly exhibited by options with a low ToP percentage and a long time period between strike fixation and maturity. It is somewhat weaker for monthly index options, where this period is rather short. Nevertheless, there is a flexibility trade-off when choosing ϑ for our two main option types. As shown by Figure 4.8, it has the consequence that a given ϑ can lead to quite distinct implied volatilities. To resolve this issue, we could choose ϑ such that the implied volatilities of all individual swing options fluctuate around an “equilibrium”

²³In a sense, this is similar to the literature emphasizing that implied volatilities are better predictors for future volatility than past volatilities (see Christensen and Prabhala, 1998).

level. To achieve this, we suggest looking for uniform values of σ and ϑ which most suitably respect the bid, ask and trade boundaries of our quotes. Specifically, for a fixed ϑ and initial time window, we start by searching for a σ which ideally lies above all bids and below all asks.²⁴ Put differently, we want to know whether we can separate all bid and ask quotes by a constant σ . This is a typical linear classification task which can be tackled by a support-vector machine (SVM; see Vapnik, 1995).

We may embed our approach into a linear SVM framework. For a given ϑ , we have a data set $(x_1, y_1), \dots, (x_n, y_n)$, where y_i encodes bids as -1 and asks as 1 , and $x_i = (\sigma_i(\vartheta))'$ is a one-dimensional vector which contains only the implied volatility value dependent on ϑ .²⁵ Any hyperplane can be written as the set of points x satisfying $w'x - b = 0$, where w is the normal vector to the hyperplane. If the quotes are linearly separable, we can select two parallel hyperplanes that separate all bids and asks. That is, we have a minimum ask hyperplane $w'x - b = 1$ (i.e., all asks lie on or above this boundary) and a maximum bid hyperplane $w'x - b = -1$ (i.e., all bids lie on or below this boundary). The region bounded by these two hyperplanes is called the “margin”. To prevent quotes from falling into the margin, we require $y_i(w'x - b) \geq 1$ for all $i = 1, \dots, n$. Further, we want the distance $2/||w||$ between these boundaries to be as large as possible. The SVM optimization problem is hence to minimize $||w||$ subject to $y_i(w'x - b) \geq 1$ for $i = 1, \dots, n$. In our simple setting, w is a scalar and the corresponding σ can be calculated as $\sigma = b/|w|$.

Because some quotations corresponding to different options may result in bids lying above asks, our data is not linearly separable. We hence introduce a squared hinge loss

$$l(w, x_i, y_i) = \max(0, 1 - y_i(w'x_i - b))^2 \quad (4.5.5)$$

which penalizes x_i that lie on the wrong side of the margin. The goal of the optimization is then to minimize

$$\frac{1}{2}w'w + C \sum_{i=1}^n l(w, x_i, y_i), \quad (4.5.6)$$

where $C > 0$ is a penalty parameter determining the trade-off between increasing the margin size and ensuring that the x_i lie on the correct side of the margin.²⁶

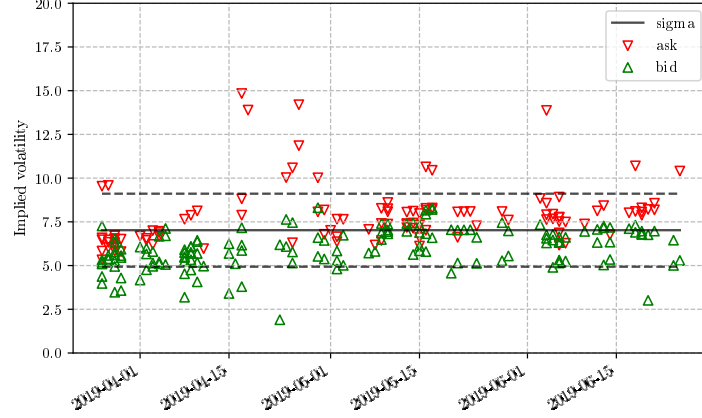
Figure 4.9 shows an exemplary fit for quotes in the three-month period before June 26, 2019 and $\vartheta = 1.8$ (which will turn out to be the optimal choice). The SVM leads to an equilibrium volatility of $\sigma = 7.022$ EUR/MWh p.a. which is accompanied by a classification accuracy of $\alpha = 74.7\%$.²⁷ Because the latter metric measures how well a data set can be separated by a chosen SVM model, we can use it to determine an appropriate mean-reversion rate. Specifically, we repeat the SVM procedure for each mean-reversion

²⁴Trade quotations can be treated as a bid and an ask, i.e., σ should match a trade quotation best.

²⁵In Sections 4.5.4 and 4.5.5, we extend x_i by seasonal and moneyness scores.

²⁶In our procedure, we choose a default value of $C = 1$.

²⁷Classification accuracy α is defined as the fraction of correct model predictions (i.e., the number of quotes i with $y_i = \hat{y}_i$) divided by the number of total quotes n , where $\hat{y}_i = 1$, if $w'x_i - b \geq 0$, and $\hat{y}_i = -1$, otherwise.



For a selected three-month period and a mean-reversion rate $\vartheta = 1.8$, this figure shows the fit of a SVM classification to estimate the equilibrium volatility. The approach yields an optimal choice of $\sigma = 7.022$ EUR/MWh p.a. with a classification accuracy of $\alpha = 74.7\%$.

FIGURE 4.9: Quote fitting procedure

rate $\vartheta \in \{1.0, 1.1, \dots, 5.8, 5.9\}$ and then select the ϑ which maximizes the classification accuracy α . In the exemplary period, this delivers $\vartheta = 1.8$. Overall, we employ a rolling window approach to avoid look-ahead bias. That is, at each day of our sample, we estimate ϑ based on quote data from the previous three months.

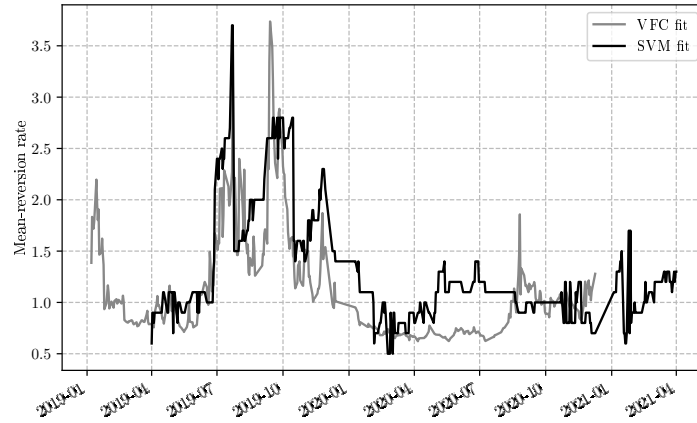
Figure 4.10 illustrates the ϑ estimates of both VFC and SVM fitting. Interestingly, in our application, the mean-reversion rates are quite close to each other. Even though the two approaches have different goals (quick estimation independent from swing option data vs. supplying a unified model for swing option valuation), their results are highly correlated and imply similar parameter choices. Therefore and because the VFC approach yields “seamless” estimates with daily frequency, we use the VFC fit results in the remainder of our analysis and simply write σ_i instead of $\sigma_i(\vartheta)$.

4.5.4 Seasonality

After computing the implied volatilities of our swing options, we now turn to a screening with respect to pronounced empirical patterns. In this context, seasonality is a particularly important issue because day-ahead spot prices of natural gas often exhibit higher volatility in winter than in summer. Monthly indexed swing option quotes are natural candidates for a closer investigation because, in contrast to fixed-strike swing options, their spot linkage (see Section 4.5.2) makes them more susceptible to seasonality.

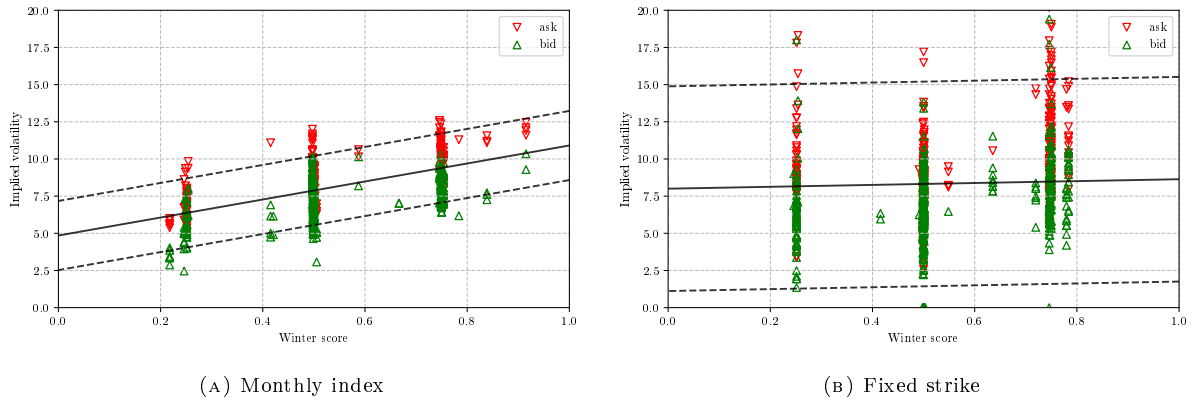
We start by introducing a seasonal score which measures the distance to the middle of a calendar year. For a specific day t of a year (ranging from 1 to 366), we obtain $|t - 183|/183$. For each quote, we then calculate a winter score ρ_i as the average of this distance over all days of the delivery period. Thus, the higher the score, the more of the delivery period falls into the winter season. We finally add ρ_i to the SVM vector x_i of Section 4.5.3, i.e., we now have $x_i = (\sigma_i, \rho_i)'$, and execute the SVM classification.

Figure 4.11a (4.11b) shows the classification results for our entire data set of monthly indexed (fixed-strike) quotes. For fixed-strike options, seasonality is not pronounced. That



This figure shows the daily estimates of ϑ delivered by our two competing methods. VFC fit refers to fitting the model implied volatility forward curve (4.5.4) to plain vanilla option data. SVM fit is a three-month rolling window SVM optimization based on swing option quotes.

FIGURE 4.10: Estimates of the mean-reversion rate

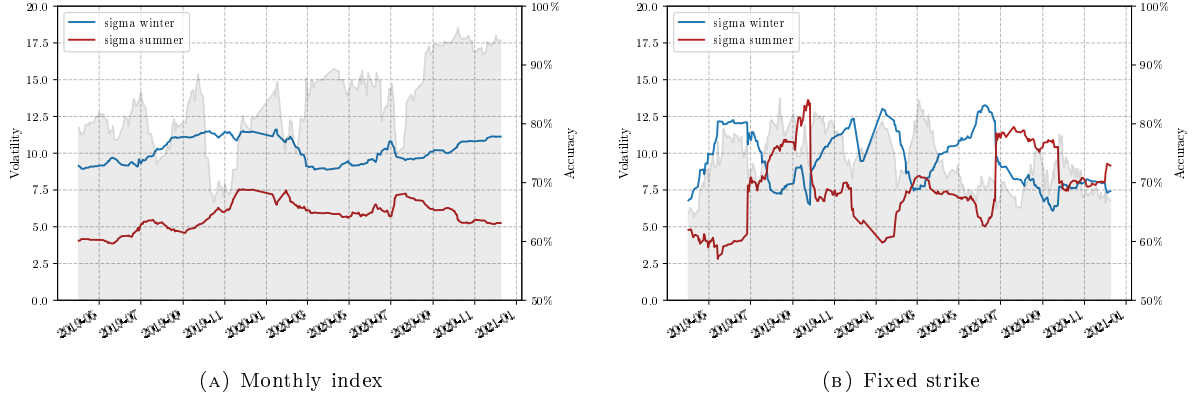


For our entire sample period and subdivided into monthly index and fixed-strike options, this figure presents the seasonal SVM classification of implied volatilities. For both swing option types, the winter score measures the average distance to the middle of a calendar year over all days of the delivery period.

FIGURE 4.11: Full sample seasonal classification

is, we detect only a marginal slope coefficient of 0.636 EUR/MWh p.a. for equilibrium volatility. Furthermore, with $\alpha = 69.5\%$, the classification is rather inaccurate. In contrast, the monthly index classification yields a significant slope of 6.056 EUR/MWh p.a. supplemented by an accuracy of $\alpha = 76.1\%$. As indicated by $\rho = 0.0$ and $\rho = 1.0$, we have low summer volatility of 4.851 EUR/MWh p.a. and high winter volatility of 10.907 EUR/MWh p.a., respectively.

If we conduct the seasonal classification in a three-month rolling window setup similar to Section 4.5.3, we obtain the results of Figure 4.12. For each date, it reports the winter and summer σ as well as the accuracy α of the classification. We see that the seasonal effect for monthly index options robustly occurs over time and is supported by high accuracy scores between roughly 70% and 95%. In contrast, for fixed-strike options, the slope of



This figure shows the three-month rolling window seasonal SVM classification results for monthly index and fixed-strike quotes. Besides the winter and summer σ , we report the accuracy scores α of the classification.

FIGURE 4.12: Rolling window seasonal classification

the seasonal effect changes its sign multiple times, so we have no evidence of a sustained seasonal effect.

4.5.5 Moneyness

We also search for moneyness patterns in implied volatilities. In general, moneyness measures the likelihood of a derivative to expire in-the-money dependent on the relative position of the current price of the underlying and the strike price (see Neftci, 2008). In our setting, we have to consider the distance of each underlying forward price $F(0, t)$ for delivery at day t to the fixed strike price K_t . We do not consider gas-indexed options because their strike building formula is designed to yield at-the-money or near-the-money strikes. In the Bachelier (1900) model, the *standardized moneyness* (in terms of standard deviation units) is given by

$$d_t = \frac{F(0, t) - K_t}{\sigma_t \sqrt{t}}, \quad (4.5.7)$$

where σ_t is the corresponding volatility of the forward price ($F(\cdot, t)$) till maturity. The correct *percentage moneyness* of a plain vanilla call is $\Phi(d_t)$, i.e., the risk-neutral likelihood of expiring in-the-money, which coincides with the option's delta value.²⁸ Unfortunately, the delta position of a swing option (given as a percentage of the total contract quantity) does not reflect its moneyness. This can be seen best when looking at 100% ToP options which exhibit a fixed total delta position regardless of the current intrinsic payouts $F(0, t) - K_t$ of the delivery period.

²⁸This is in contrast to the Black (1976) model, where the risk-neutral probability of exercise deviates from the delta of a vanilla option (see also Section 4.3.3).

Because swing options permit the holder to spread more or less consumption over the delivery period, it is not clear how to weight the moneyness of different dates. We propose to calculate the moneyness corresponding to the maximum intrinsic position. This position is characterized by the optimal execution of all exercise rights with respect to the current intrinsic payouts, i.e.,

$$q^\times := \arg \max_{(q_t)} \sum_{t=0}^{T-1} (F(0, t) - K(0, t)) \cdot q_t, \quad (4.5.8)$$

subject to $q_{\min} \leq q_t \leq q_{\max}$, and $\sum_{t=0}^{T-1} q_t = Q_{\max}$.

The calculation of this intrinsic position is visualized in Figure 4.13. We can approximate the correct percentage moneyness

$$\mathcal{M}^\times := \mathbb{P} \left(\sum_{t=0}^{T-1} (S_t - K_t) \cdot q_t^\times \geq 0 \right) \quad (4.5.9)$$

via the Monte Carlo quantile of the valuation in Section 4.4.1. The quantile can also be calculated analytically because the intrinsic position (q_t^\times) is fixed and model-free. We have

$$\mathbb{E}[S_t] = F(0, t), \quad (4.5.10)$$

$$\text{Cov}(S_s, S_t) = \frac{\sigma^2}{2\vartheta} (e^{-\vartheta|s-t|} - e^{-\vartheta|s+t|}). \quad (4.5.11)$$

We restrict the means $\mu := (\mathbb{E}[S_t] - K_t)_{t=0, \dots, T-1}$ and covariances $\Sigma := (\text{Cov}(S_s, S_t))_{s, t=0, \dots, T-1}$ to the delivery period and calculate

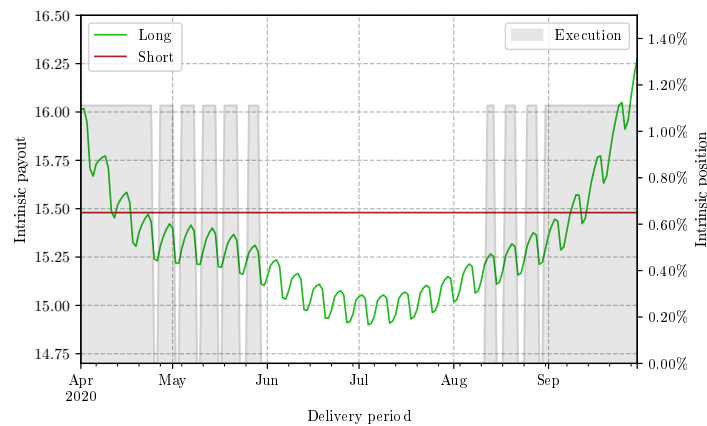
$$\sum_{t=0}^{T-1} (S_t - K_t) \cdot q_t^\times \sim \mathcal{N}((q^\times)' \mu, (q^\times)' \Sigma q^\times). \quad (4.5.12)$$

The intrinsic percentage moneyness is hence given by

$$\mathcal{M}^\times = \Phi(d^\times) \quad \text{with} \quad d^\times = \frac{(q^\times)' \mu}{\sqrt{(q^\times)' \Sigma q^\times}}. \quad (4.5.13)$$

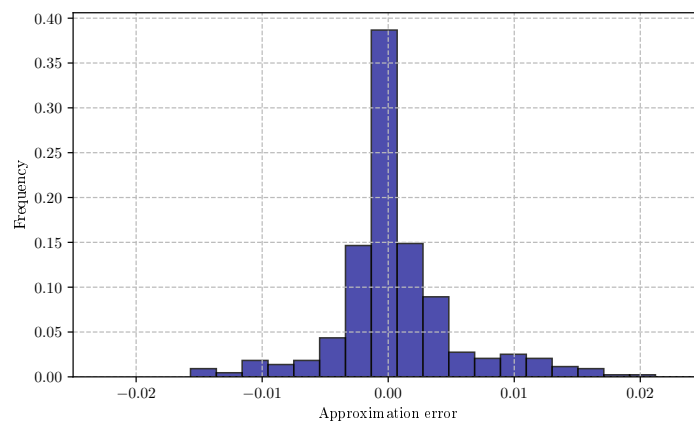
We find only small differences between the analytical and the Monte Carlo approach. The error distribution for $N = 1,000$ sample paths is presented in Figure 4.14.

In our sample, searching for moneyness patterns in implied volatilities is complicated by the fact that most of our quotations refer to swing options which are at-the-money or barely in-the-money. 90% of the swing quotations have a moneyness between 50% and 60%. We do not have many in-the-money and only few out-of-the-money quotations. Nevertheless, we examine our data. That is, we extend the SVM framework of Section 4.5.3 by adding the moneyness \mathcal{M}_i^\times for each quote i to the vector x_i , i.e., $x_i = (\sigma_i, \mathcal{M}_i^\times)'$, and then perform the SVM classification. As with our seasonal classification, we conduct the analysis for the



This figure shows the daily intrinsic exercise loadshape (as a percentage of the total contract quantity) of a selected swing option with 90 exercise rights during the summer season 2019. It addition, it plots the corresponding intrinsic payouts (in EUR/MWh), which are given by the PFC of the valuation date (long intrinsic payout) and the fixed strike price (short intrinsic payout).

FIGURE 4.13: Intrinsic execution



This figure shows the error distribution of the moneyness approximation. Specifically, for all fixed-strike swing options, we have compared the analytical and the Monte Carlo moneyness values which correspond to the derived implied volatilities. The mean absolute approximation error for $N = 1,000$ sample paths is 0.00317.

FIGURE 4.14: Moneyiness approximation error

full data set (of fixed-strike options) and in the form of a rolling window setup. Figure 4.15a, which plots implied volatility against moneyness in the full sample, suggests that implied volatilities decrease with higher moneyness. The rolling window results of Figure 4.15b imply that the form of the volatility smirk changes over time. In periods with higher at-the-money volatility, the smirk is more often upward sloping than in periods with lower at-the-money volatility. However, this should be understood as a preliminary result. More research, which considers non-linear classification kernels (see Lee and Mangasarian, 2001)

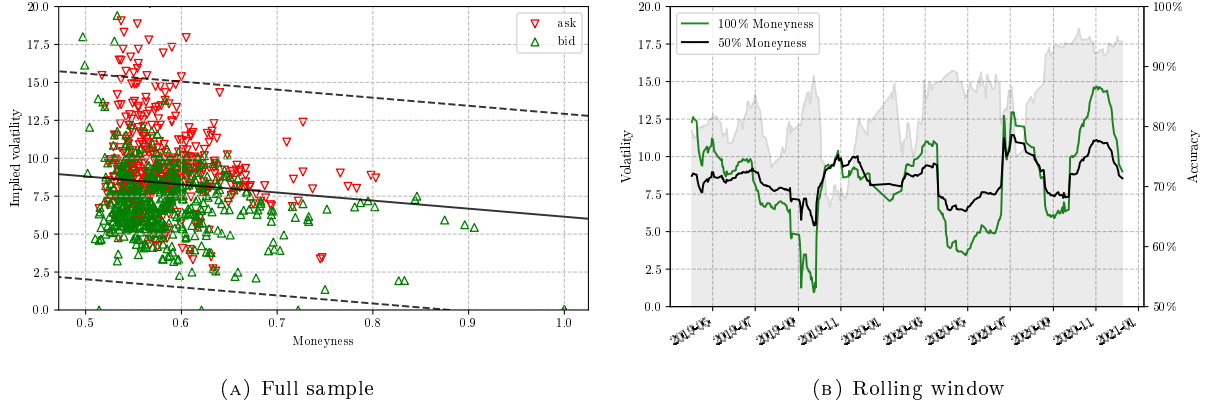


FIGURE 4.15: Moneyness classification

or modifies the components of the input vector x_i , is needed to fully capture the empirical dynamics of swing-option implied volatilities.

4.6 Conclusion

Closing a significant gap in the option pricing literature, we have developed an approach to derive implied volatilities from swing option quotes. Specifically, we have used numerous recent theoretical results to formulate a unified option valuation and sensitivity analysis framework, which, in combination with an adequate price model for the underlying, allows using established Monte Carlo methods and a Newton-Raphson root-finding procedure for an efficient calculation of implied volatilities. Even for just 1,000 sample paths, our universally convergent technique converges after a few iterations and exhibits calibration errors of acceptably low magnitude.

Along the lines of our work, we have made several additional contributions. For example, we use delta-hedging arguments to emphasize that, in the context of energy options, additive price models could be preferred to multiplicative ones. Furthermore, we show that our specific single-factor price model, i.e., its mean-reversion parameter, is linked to the flexibility value of swing options. This relationship helps us to better understand the value dynamics of swing options and can be exploited to estimate the mean-reversion rate (by choosing a parameter value which minimizes implied volatility scattering). Finally, we develop a formula for the vega of swing options (with special emphasis on options with gas-indexed strike prices) and a definition for the moneyness of swing options.

Our study delivers a valuable tool for energy market practice which can be readily understood and implemented. It not only allows a comparison of option market quotes but also informs about the perceptions of traders with respect to future spot market fluctuations. In addition, our empirical study of the European natural gas market presents some first evidence on the properties of swing option-implied volatilities. They serve as an important input for future research. For example, documented seasonalities and partial smirk-indications support a full characterization of the complex high-dimensional swing

option-implied volatility surface (see Fengler et al., 2007). Furthermore, they naturally lead to the question whether information contained in this surface can be used to predict future market movements (see Busch et al., 2011) and improve portfolio selection decisions (see DeMiguel et al., 2013).

Chapter 5

Summary

In this thesis, we analyse selected questions in the field of energy portfolio management. The main results of the three papers can be summarised as follows.

The paper on *portfolio theory* addresses the theoretical and empirical justification of mean-variance analysis in portfolio optimization. Our main contribution is a characterization of distributions that imply mean-variance determined portfolio returns. That is, we focus on returns where the weights over the risk-free asset and the risky assets sum to unity. We show that, in presence of a risk-free asset, a specific class of skew-elliptical distributions is necessary and sufficient for the distribution of all portfolios to be determined by its mean and variance. This finding implies that skewed returns do not allow a rejection of mean-variance analysis, which is a common belief among academics and practitioners. Additionally, we provide auxiliary results which point out links between famous works on this topic and rule out common misunderstandings. Among these are the works of (i) Chamberlain (1983) who provides the basis of our proof but focuses on scaled portfolios rather than on returns, (ii) Meyer and Rasche (1992) who introduces the weight constraint in the context of distributional requirements but only treats some special cases, (iii) Schuhmacher and Auer (2014) whose implications on the irrelevance of the performance measure choice can be expanded to our framework and (iii) Simaan (1993a) whose derivation of the efficient set in his mean-variance-skewness framework can be reduced to our case. To round off our theoretical results, we finish our study by presenting some evidence on the empirical relevance of our skew-elliptical model. To this end, we modify the testing approach proposed by Meyer and Rasche (1992) to be applicable to our specific case.

In the context of *electricity price forecasting* we run a comprehensive comparative study featuring various approaches to tackle dimensionality problems in multivariate short-term forecasting of day-ahead prices. Our study provides several important insights which are relevant for the performance of different forecasting approaches. First, importance scores drawn from random forests reveal the distinct intraday dependencies of hourly prices stating which hours of which past days have the highest predictive power for specific hours in the future. Second, using these scores to perform a feature selection for powerful machine learning tools lead to very promising forecasting results. That is, feeding the identified variables into a support vector machine is an outstanding forecasting technique in our study compared to other approaches (such as dynamic factor models, penalized regressions or Bayesian shrinkage) that are commonly used to resolve dimensionality problems. Third,

we find that some forecasting methods are more accurate for certain hours of the day than others which can be explained by the presence of distinct importance profiles for different hourly prices. Finally, we can cite this observation as a reason for the fact that simple forecast combination schemes tend to outperform the full battery of models in our study.

Dealing with *energy swing options* in the last paper, we develop an approach to derive implied volatilities from swing option quotes. Specifically, we first formulate a unified option valuation and sensitivity analysis framework building on numerous recent theoretical results. Second, we discuss an adequate price model choice for the underlying with particular attention to the volatility function and return specification. Third, we apply the well-known Longstaff and Schwartz (2001) method for the valuation of swing options and a Newton-Raphson root-finding procedure for an efficient calculation of implied volatilities. Working with this concept in our empirical study, we analyse the shape of the value function as well as seasonal and moneyness patterns in implied volatilities of swing option quotes. We provide several insights for the two most often quoted types of options, which are monthly index and 100% ToP fixed price options. For monthly index options, we find that (i) their value is mainly driven by the volatility of the spread between spot and monthly index price and (ii) implied volatilities are higher for delivery in winter months. For 100% ToP fixed price options, we find that (i) their value is mainly determined by the volatility of a characteristic forward time spread, (ii) implied volatilities exhibit no ongoing seasonality pattern (with respect to the delivery period) and (iii) the implied volatility smile is more often upwards sloping in periods with higher market uncertainty, i.e. higher implied volatilities in general.

Bibliography

- Abhakorn, P., Smith, P., Wickens, M., 2013. What do the Fama-French factors add to C-CAPM? *Journal of Empirical Finance* 22, 113–127.
- Adcock, C. J., 2014. Mean-variance-skewness efficient surfaces, Stein's lemma and the multivariate extended skew-Student distribution. *European Journal of Operational Research* 234 (2), 392–401.
- Agarwal, V., Naik, N., 2004. Risks and portfolio decisions involving hedge funds. *Review of Financial Studies* 17 (1), 63–98.
- Aggarwal, S., Saini, L., Kumar, A., 2009. Short term price forecasting in deregulated electricity markets: A review of statistical models and key issues. *International Journal of Energy Sector Management* 3 (4), 333–358.
- Ahn, S. C., Horenstein, A. R., 2013. Eigenvalue ratio test for the number of factors. *Econometrica* 81 (3), 1203–1227.
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6), 716–723.
- Alessi, L., Barigozzi, M., Capasso, M., 2010. Improved penalization for determining the number of factors in approximate factor models. *Statistics and Probability Letters* 80 (23–24), 1806–1813.
- Alexandrov, A., 1939. Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it. *Leningrad State Univ. Annals [Uchenye Zapiski] Math. Ser.* 37, 3–35.
- Allen, R., Stone, J., 2005. Textbook neglect of the constant coefficient. *Journal of Economic Education* 36 (4), 379–384.
- Alquist, R., Kilian, L., 2010. What do we learn from the price of crude oil futures? *Journal of Applied Econometrics* 25 (5), 539–573.
- Alquist, R., Kilian, L., Vigfusson, R., 2013. Forecasting the price of oil. In: *Handbook of Economic Forecasting*. Vol. 2A. Elsevier, Amsterdam, pp. 427–507.
- Amjady, N., 2006. Day-ahead price forecasting of electricity markets by a new fuzzy neural network. *IEEE Transactions on Power Systems* 21 (2), 887–896.
- Amjady, N., 2007. Short-term bus load forecasting of power systems by a new hybrid method. *IEEE Transactions on Power Systems* 22 (1), 333–341.
- Anbazhagan, S., Kumarappan, N., 2013. Day-ahead deregulated electricity market price forecasting using recurrent neural network. *IEEE Systems Journal* 7 (4), 866–872.
- Andalib, A., Atry, F., 2009. Multi-step ahead forecasts for electricity prices using NARX: A new approach, a critical analysis of one-step ahead forecasts. *Energy Conversion and Management* 50 (3), 739–747.

- Ang, A., 2014. Asset management: A systematic approach to factor investing. Oxford University Press, Oxford.
- Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent measures of risk. *Mathematical Finance* 9 (3), 203–228.
- Auer, B., 2016. How does Germany's green energy policy affect electricity market volatility? An application of conditional autoregressive range models. *Energy Policy* 98, 621–628.
- Auret, L., Aldrich, C., 2012. Interpretation of nonlinear relationships between process variables by use of random forests. *Minerals Engineering* 35, 27–42.
- Avci, E., Ketter, W., van Heck, E., 2018. Managing electricity price modeling risk via ensemble forecasting: The case of Turkey. *Energy Policy* 123, 390–403.
- Bachelier, L., 1900. Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure* 17, 21–86.
- Bacon, C., 2008. Practical portfolio performance measurement and attribution, 2nd Edition. John Wiley & Sons, Hoboken, New Jersey.
- Bai, J., 2003. Testing parametric conditional distributions of dynamic models. *Review of Economics and Statistics* 85 (3), 531–549.
- Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 70 (1), 191–221.
- Ball, C. A., Torous, W. N., Tschoegl, A. E., 1985. On inferring standard deviations from path dependent options. *Economics Letters* 18 (4), 377–380.
- Balvers, R., Wu, Y., Gilliland, E., 2000. Mean reversion across national stock markets and parametric contrarian investment strategies. *Journal of Finance* 55 (2), 745–772.
- Bandi, F. M., Perron, B., 2006. Long memory and the relation between implied and realized volatility. *Journal of Financial Econometrics* 4 (4), 636–670.
- Barasinska, N., Schäfer, D., Stephan, A., 2012. Individual risk attitudes and the composition of financial portfolios: Evidence from German household portfolios. *Quarterly Review of Economics and Finance* 52 (1), 1–14.
- Bardou, O., Bouthemy, S., Pagés, G., 2010. When are swing options bang-bang and how to use it. *International Journal of Theoretical and Applied Finance* 13 (6), 867–899.
- Baron, D., 1977. On the utility theoretic foundations of mean-variance analysis. *Journal of Finance* 32 (5), 1683–1697.
- Barrera-Esteve, C., Bergeret, F., Dossal, C., Gobet, E., Meziou, A., Munos, R., Reboul-Salze, D., 2006. Numerical methods for the pricing of swing options: A stochastic control approach. *Methodology and Computing in Applied Probability* 8 (4), 517–540.
- Baumeister, C., Kilian, L., 2015. Forecasting the real price of oil in a changing world: A forecast combination approach. *Journal of Business and Economic Statistics* 33 (3), 338–351.
- Becker, N., Werft, W., Toedt, G., Lichter, P., Benner, A., 2009. penalizedSVM: A R-package for feature selection SVM classification. *Bioinformatics* 25 (13), 1711–1712.
- Bellman, R., 1952. On the theory of dynamic programming. *Proceedings of the National Academy of Sciences* 38 (8), 716–719.
- Bellman, R., 1957. Dynamic Programming. Princeton University Press, Princeton, New Jersey.
- Berger, B., Dietrich, M., Döttling, R., Heider, P., Spanderen, K., 2018. Semianalytical pricing and hedging of fixed and indexed energy swing contracts. *Journal of Energy*

- Markets 11 (4), 1–26.
- Bernard, C., Vanduffel, S., 2014. Mean-variance optimal portfolios in the presence of a benchmark with applications to fraud detection. *European Journal of Operational Research* 234 (2), 469–480.
- Best, M., Grauer, R., 1991. Sensitivity analysis for mean-variance portfolio problems. *Management Science* 37 (8), 980–989.
- Bigelow, J., 1993. Consistency of mean-variance analysis and expected utility analysis: A complete characterization. *Economics Letters* 43 (2), 187–192.
- Black, F., 1972. Capital market equilibrium with restricted borrowing. *Journal of Business* 45 (3), 444–455.
- Black, F., 1976. The pricing of commodity contracts. *Journal of Financial Economics* 3 (1), 167–179.
- Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81 (3), 637–654.
- Blum, A., 1992. *Neural Networks in C++: An Object-Oriented Framework for Building Connectionist Systems*. John Wiley & Sons, Hoboken.
- Board, J., Sutcliffe, C., 1994. Estimation methods in portfolio selection and the effectiveness of short sales restrictions: UK evidence. *Management Science* 40 (4), 516–534.
- Boger, Z., Guterman, H., 1997. Knowledge extraction from artificial neural networks models. In: *IEEE International Conference on Systems Man and Cybernetics*. Vol. 4. pp. 3030–3035.
- Bonnans, J., Cen, Z., Christel, T., 2012. Sensitivity analysis of energy contracts by stochastic programming techniques. In: Carmona, R. A., Del Moral, P., Hu, P., Oudjane, N. (Eds.), *Numerical Methods in Finance*. Springer, Berlin, Heidelberg, pp. 447–471.
- Boogert, A., de Jong, C., 2008. Gas storage valuation using a Monte Carlo method. *Journal of Derivatives* 15 (3), 81–98.
- Boogert, A., de Jong, C., 2011. Gas storage valuation using a multi-factor price process. *Journal of Energy Markets* 4 (4), 29–52.
- Bordignon, S., Bunn, D. W., Lisi, F., Nan, F., 2013. Combining day-ahead forecasts for British electricity prices. *Energy Economics* 35, 88–103.
- Boser, B. E., Guyon, I. M., Vapnik, V. N., 1992. A training algorithm for optimal margin classifiers. In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. pp. 144–152.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., Ljung, G. M., 2015. *Time Series Analysis: Forecasting and Control*, 5th Edition. John Wiley & Sons, Hoboken.
- Boyd, S., Vandenberghe, L., 2009. *Convex Optimization*, seventh Edition. Cambridge University Press, Cambridge, Massachusetts.
- Breiman, L., 2001. Random forests. *Machine Learning* 45 (1), 5–32.
- Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J., 1984. *Classification and Regression Trees*. Routledge, New York.
- Brenner, M., Subrahmanyam, M. G., 1988. A simple formula to compute the implied standard deviation. *Financial Analysts Journal* 44 (5), 80–83.
- Brezis, H., 2010. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, New York.

- Britten-Jones, M., 1999. The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance* 54 (2), 655–671.
- Broadie, M., Glasserman, P., 1996. Estimating security price derivatives using simulation. *Management Science* 42 (2), 269–285.
- Brockett, P., Golden, L., 1987. A class of utility functions containing all the common utility functions. *Management Science* 33 (8), 955–964.
- Brown, S. J., 1990. Estimating volatility. In: Figlewski, S., Silber, W. L., Subrahmanyam, M. G. (Eds.), *Financial Options: From Theory to Practice*. Business One Irwin, Homewood, Illinois, pp. 516–537.
- Bühlmann, P., van de Geer, S., 2011. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, Berlin, Heidelberg.
- Bunn, D. W., 2004. *Modelling Prices in Competitive Electricity Markets*. John Wiley & Sons, Hoboken.
- Busch, T., Christensen, B. J., Mielsen, M. Ø., 2011. The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics* 160 (1), 48–57.
- Canina, L., Figlewski, S., 1993. The informational content of implied volatility. *Review of Financial Studies* 6 (3), 659–681.
- Carmona, R., Ludkovski, M., 2010. Swing options. In: Cont, R. (Ed.), *Encyclopedia of Quantitative Finance*. John Wiley and Sons, Chichester.
- Carverhill, A., 1994. When is the short rate Markovian? *Mathematical Finance* 4 (4), 305–312.
- Cass, D., Stiglitz, J. E., 1970. The structure of investor preferences and asset returns, and separability in portfolio allocation: A contribution to the pure theory of mutual funds. *Journal of Economic Theory* 2 (2), 122–160.
- Catalão, J., Mariano, S., Mendes, V., Ferreira, L., 2007. Short-term electricity prices forecasting in a competitive market: A neural network approach. *Electric Power Systems Research* 77 (10), 1297–1304.
- Chaâbane, N., 2014. A novel auto-regressive fractionally integrated moving average-least-squares support vector machine model for electricity spot prices prediction. *Journal of Applied Statistics* 41 (3), 635–651.
- Chamberlain, G., 1983. A characterization of the distributions that imply mean-variance utility functions. *Journal of Economic Theory* 29 (1), 185–201.
- Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2 (3), Article 27.
- Che, J., Wang, J., 2010. Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling. *Energy Conversion and Management* 51 (10), 1911–1917.
- Chen, K.-Y., Wang, C.-H., 2007. Support vector regression with genetic algorithms in forecasting tourism demand. *Tourism Management* 28 (1), 215–226.
- Chen, L., He, S., Zhang, S., 2011. When all risk-adjusted performance measures are the same: In praise of the sharpe ratio. *Quantitative Finance* 11 (10), 1439–1447.
- Chen, N., Liu, Y., 2014. American option sensitivities estimation via a generalized infinitesimal perturbation analysis approach. *Operations Research* 62 (3), 616–632.

- Cherkassky, V., Mulier, F., 1998. *Learning From Data: Concepts, Theory, and Methods*. John Wiley & Sons, Hoboken.
- Chiarella, C., Clewlow, L., Kang, B., 2016. The evaluation of multiple year gas sales agreement with regime switching. *International Journal of Theoretical and Applied Finance* 19 (1), 1650005.
- Chriss, N., 1996. *Black Scholes and Beyond: Option Pricing Models*. McGraw-Hill, New York.
- Christensen, B. J., Prabhala, N. R., 1998. The relation between implied and realized volatility. *Journal of Financial Economics* 50 (2), 125–150.
- Clark, T., McCracken, M., 2013. Advances in forecast evaluation. In: *Handbook of Economic Forecasting*. Vol. 2B. Elsevier, Amsterdam, pp. 1107–1201.
- Clemen, R. T., 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5 (4), 559–583.
- Clément, E., Lamberton, D., Protter, P., 2002. An analysis of a least squares regression method for American option pricing. *Finance and Stochastics* 6 (4), 449–471.
- Clewlow, L., Strickland, C., 1999. Valuing energy options in a one factor model fitted to forward prices, QFRC Research Paper No. 10, University of Technology Sydney.
- Clewlow, L., Strickland, C., 2000. *Energy Derivatives: Pricing and Risk Management*. Lacima Publications, London.
- Conejo, A. J., Contreras, J., Espínola, R., Plazas, M. A., 2005. Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting* 21 (3), 435–462.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine Learning* 20 (3), 273–297.
- Cox, J. C., Ross, S. A., Rubinstein, M., 1979. Option pricing: A simplified approach. *Journal of Financial Economics* 7 (3), 229–263.
- Cruz, A., Muñoz, A., Zamora, J., Espinola, R., 2011. The effect of wind generation and weekday on Spanish electricity spot price forecasting. *Electric Power Systems Research* 81 (10), 1924–1935.
- Curtis Jr, C. E., Carriker, G. L., 1988. Estimating implied volatility directly from "nearest-to-the-money" commodity option premiums, DAAE Working Paper No. 116875, Clemson University.
- Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems* 2 (4), 303–314.
- DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies* 22 (5), 1915–1953.
- DeMiguel, V., Plyakha, Y., Uppal, R., Vilkov, G., 2013. Improving portfolio selection using option-implied volatility and skewness. *Journal of Financial and Quantitative Analysis* 48 (6), 1813–1845.
- Deng, H., Runger, G. C., 2013. Gene selection with guided regularized random forest. *Pattern Recognition* 46 (12), 3483–3489.
- Deuffhard, P., 2011. *Newton methods for nonlinear problems: affine invariance and adaptive algorithms*. Vol. 35. Springer Science & Business Media.
- Devaine, M., Gaillard, P., Goude, Y., Stoltz, G., 2013. Forecasting electricity consumption by aggregating specialized experts. *Machine Learning* 90, 231–260.

- Dexter, A., Yu, J., Ziemba, W., 1980. Portfolio selection in a lognormal market when the investor has a power utility function: Computational results. In: Dempster, M. (Ed.), *Stochastic programming*. Academic Press, London, pp. 507–523.
- Díaz-Uriarte, R., Alvarez de Andrés, S., 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics* 7 (1), Article 3.
- Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13 (3), 253–263.
- Edoli, E., Fiorenzani, S., Ravelli, S., Vargiolu, T., 2013. Modeling and valuing make-up clauses in gas swing contracts. *Energy Economics* 35, 58–73.
- Eling, M., Schuhmacher, F., 2007. Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking and Finance* 31 (9), 2632–2647.
- Elton, E., Gruber, M., Brown, S., Goetzmann, W., 2007. *Modern portfolio theory and investment analysis*, 7th Edition. John Wiley & Sons, Hoboken, New Jersey.
- Eydeland, A., Wolyniec, K., 2003. *Energy and Power Risk Management: New Developments in Modeling, Pricing, and Hedging*. John Wiley & Sons, Hoboken.
- Fama, E., French, K., 2012. Size, value, and momentum in international stock returns. *Journal of Financial Economics* 105 (3), 457–472.
- Fan, S., Mao, C., Chen, L., 2007. Next-day electricity-price forecasting using a hybrid network. *IET Proceedings of Generation, Transmission and Distribution* 1 (1), 176–182.
- Fanelli, V., Schmeck, M. D., 2019. On the seasonality in the implied volatility of electricity options. *Quantitative Finance* 19 (8), 1321–1337.
- Felix, B., Woll, O., Weber, C., 2013. Gas storage valuation under limited market liquidity: An application in Germany. *European Journal of Finance* 19 (7-8), 715–733.
- Fengler, M. R., Härdle, W. K., Mammen, E., 2007. A semiparametric factor model for implied volatility surface dynamics. *Journal of Financial Econometrics* 5 (2), 189–218.
- Flack, V. F., Chang, P. C., 1987. Frequency of selecting noise variables in subset regression analysis: A simulation study. *American Statistician* 41 (1), 84–86.
- Fleten, S.-E., Keppo, J., Näsäkkälä, 2012. Risk management in electric utilities. In: Kouvelis, P., Dong, L., Boyabatli, O., Li, R. (Eds.), *Handbook of Integrated Risk Management in Global Supply Chains*. John Wiley and Sons, Hoboken, pp. 495–513.
- Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33 (1), 1–22.
- Fung, W., Hsieh, D., Naik, N., Ramadorai, T., 2008. Hedge funds: Performance, risk, and capital formation. *Journal of Finance* 63 (4), 1777–1803.
- Gao, X., Nardari, F., 2018. Do commodities add economic value in asset allocation? New evidence from time-varying moments. *Journal of Financial and Quantitative Analysis* 53 (1), 365–393.
- García-Ascanio, C., Maté, C., 2010. Electric power demand forecasting using interval time series: A comparison between VAR and iMLP. *Energy Policy* 38 (2), 715–725.
- Gardner, R., Pinder, J., Wood, R., 1980. Monte Carlo estimation of percentiles for the multi-Smirnov test. *Journal of Statistical and Computational Simulation* 10 (3-4), 243–249.
- Genre, V., Kenny, G., Meyler, A., Timmermann, A., 2013. Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting* 29 (1), 108–121.

- Genton, M., Loperfido, N., 2005. Generalized skew-elliptical distributions and their quadratic forms. *Annals of the Institute of Statistical Mathematics* 57 (2), 389–401.
- Genuer, R., Poggi, J.-M., Tuleau, C., 2008. Random forests: Some methodological insights, INRIA Working Paper No. 6729.
- Gilli, M., Këllezzi, E., 2006. An application of extreme value theory for measuring financial risk. *Computational Economics* 27, 207–228.
- Glasserman, P., Yu, B., 2004. Number of paths versus number of basis functions in American option pricing. *Annals of Applied Probability* 14 (4), 2090–2119.
- González, C., Mira, J., Ojeda, J., 2016. Applying multi-output random forest models to electricity price forecast. Preprints 2016090053.
- Green, J., Hand, J., Zhang, X., 2017. The characteristics that provide independent information about average U.S. monthly stock returns. *Review of Financial Studies* 30 (12), 4389–4436.
- Gregorutti, B., Michel, B., Saint-Pierre, P., 2015. Grouped variable importance with random forests and application to multiple functional data analysis. *Computational Statistics and Data Analysis* 90, 15–35.
- Grootveld, H., Hallerbach, W., 1999. Variance vs downside risk: Is there really that much difference? *European Journal of Operational Research* 114 (2), 304–319.
- Grundy, B., Lim, B., Verwijmeren, P., 2012. Do option markets undo restrictions on short sales? Evidence from the 2008 short-sale ban. *Journal of Financial Economics* 106 (2), 331–348.
- Gürtler, M., Paulsen, T., 2018. Forecasting performance of time series models on electricity spot markets: A quasi-meta-analysis. *International Journal of Energy Sector Management* 12 (1), 103–129.
- Hadar, J., Russell, W., 1971. Stochastic dominance and diversification. *Journal of Economic Theory* 3 (3), 288–305.
- Hagan, M. T., Demuth, H. B., Beale, M. H., 1996. *Neural Network Design*. PWS Publishing, Boston.
- Hallin, M., Liška, R., 2007. Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association* 102 (478), 603–617.
- Hanfeld, M., Schlüter, S., 2017. Operating a swing option on today's gas markets: How least squares Monte Carlo works and why it is beneficial. *Zeitschrift für Energiewirtschaft* 41, 137–150.
- Hanoch, G., Levy, H., 1969. The efficiency analysis of choices involving risk. *Review of Economic Studies* 36 (3), 335–346.
- Hansen, P., Lunde, A., 2005. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* 20 (7), 873–889.
- Hansen, P. R., 2005. A test for superior predictive ability. *Journal of Business and Economic Statistics* 23 (4), 365–380.
- Haugh, M. B., Kogan, L., 2004. Pricing American options: A duality approach. *Operations Research* 52 (2), 165–336.
- Heidema, A. G., Boer, J. M. A., Nagelkerke, N., Mariman, E. C. M., van der A, D. L., Feskens, E. J. M., 2006. The challenge for genetic epidemiologists: How to analyze large numbers of SNPs in relation to complex diseases. *BMC Genetics* 7, Article 23.

- Hibon, M., Evgeniou, T., 2005. To combine or not to combine: Selecting among forecasts and their combinations. *International Journal of Forecasting* 21 (1), 15–24.
- Hickernell, F., 1998. A generalized discrepancy and quadrature error bound. *Mathematics of Computation* 67 (221), 299–322.
- Hlawitschka, W., 1994. The empirical nature of Taylor-series approximations to expected utility. *American Economic Review* 84 (3), 713–719.
- Hoerl, A. E., Kennard, R. W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12 (1), 55–67.
- Homm, U., Pigorsch, C., 2012. Beyond the Sharpe ratio: An application of the Aumann-Serrano index to performance measurement. *Journal of Banking and Finance* 36 (8), 2274–2284.
- Hu, Z., Bao, Y., Xiong, T., 2013. Electricity load forecasting using support vector regression with memetic algorithms. *Scientific World Journal* 2013, Article 292575.
- Huffer, F., Park, C., 2007. A test for elliptical symmetry. *Journal of Multivariate Analysis* 98 (2), 256–281.
- Hunter, J., 1972. Independence, conditional expectation, and zero covariance. *American Statistician* 26 (5), 22–24.
- Ibáñez, A., Zapatero, F., 2004. Monte Carlo valuation of American options through computation of the optimal exercise frontier. *Journal of Financial and Quantitative Analysis* 39 (2), 253–275.
- Ingersoll, J. E., 1987. *Theory of financial decision making*. Vol. 3. Rowman & Littlefield, Lanham.
- Islam, S., Anand, S., Hamid, J., Thabane, L., Beyene, J., 2017. Comparing the performance of linear and nonlinear principal components in the context of high-dimensional genomic data integration. *Statistical Applications in Genetics and Molecular Biology* 16 (3), 199–216.
- Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: Why imposing the wrong constraint helps. *Journal of Finance* 58 (4), 1651–1684.
- Jaillet, J., Ronn, E., Tompaidis, S., 2004. Valuation of commodity-based swing options. *Management Science* 50 (7), 909–921.
- Jia, X., Ruan, X., Zang, J. E., 2021. The implied volatility smirk of commodity options. *Journal of Futures Markets* 41, 72–104.
- Jorion, P., 1995. Predicting volatility in the foreign exchange market. *Journal of Finance* 50 (2), 507–528.
- Joutz, F., Maddala, G., Trost, R., 1995. An integrated Bayesian vector auto regression and error correction model for forecasting electricity consumption and prices. *Journal of Forecasting* 14 (3), 287–310.
- Kang-Lin, P., Wu, C.-H., Yeong-Jia, J. G., 2004. The development of a new statistical technique for relating financial information to stock market returns. *International Journal of Management* 21 (4), 492–505.
- Kapsos, M., Christofides, N., Rustem, B., 2014. Worst-case robust Omega ratio. *European Journal of Operational Research* 234 (2), 499–507.
- Karlsson, S., 2013. Forecasting with Bayesian vector autoregression. In: *Handbook of Economic Forecasting*. Vol. 2B. Elsevier, Amsterdam, pp. 791–897.
- Keppo, J., 2004. Pricing of electricity swing options. *Journal of Derivatives* 11 (3), 26–43.

- Kohler, M., 2010. A review of regression-based Monte Carlo methods for pricing American options. In: Devroye, L., Karasözen, B., Kohler, M., Korn, R. (Eds.), *Recent Developments in Applied Probability and Statistics*. Physica, Heidelberg, pp. 37–58.
- Kohrs, H., Mühlichen, H., Auer, B. R., Schuhmacher, F., 2019. Pricing and risk of swing contracts in natural gas markets. *Review of Derivatives Research* 22 (1), 77–167.
- Kolm, P. N., Tütüncü, R., Fabozzi, F. J., 2014. 60 years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research* 234 (2), 356–371.
- Korkie, B., Turtle, H., 2002. A mean-variance analysis of self-financing portfolios. *Management Science* 48 (3), 313–452.
- Kristiansen, T., 2012. Forecasting Nord Pool day-ahead prices with an autoregressive model. *Energy Policy* 49, 328–332.
- Kritzman, M., 1991. What practitioners need to know about estimating volatility - part 1. *Financial Analysts Journal* 47 (4), 22–25.
- Kroll, Y., Levy, H., Markowitz, H. M., 1984. Mean-variance versus direct utility maximization. *Journal of Finance* 39 (1), 47–61.
- Kumar, A., Lim, S., 2008. How do decision frames influence the stock investment choices of individual investors? *Management Science* 54 (6), 1052–1064.
- Kuo, P., Huang, C., 2018. An electricity price forecasting model by hybrid structured deep neural networks. *Sustainability* 10 (4), Article 1280.
- Lago, J., Ridder, F., Vrancx, P., Schutter, B., 2018. Forecasting day-ahead electricity prices in Europe: The importance of considering market integration. *Applied Energy* 211, 890–903.
- Lai, G., Margot, F., Secomandi, N., 2010. An approximate dynamic programming approach to benchmark practice-based heuristics for natural gas storage valuation. *Operations Research* 58 (3), 564–582.
- Lee, Y., Mangasarian, O., 2001. SSVM: A smooth support vector machine for classification. *Computational Optimization and Applications* 20, 5–22.
- Levy, H., Levy, M., 2004. Prospect theory and mean-variance analysis. *Review of Financial Studies* 17 (4), 1015–1041.
- Levy, H., Markowitz, H. M., 1979. Approximating expected utility by a function of mean and variance. *American Economic Review* 69 (3), 308–317.
- Levy, M., Kaplanski, G., 2015. Portfolio selection in a two-regime world. *European Journal of Operational Research* 242 (2), 514–524.
- Li, F., Tkacz, G., 2011. A consistent test for multivariate conditional distributions. *Econometric Reviews* 30 (3), 251–273.
- Liang, J., Fang, K., Hickernell, F., 2008. Some necessary uniform tests for spherical symmetry. *Annals of the Institute of Statistical Mathematics* 60 (3), 679–696.
- Liang, J., Fang, K., Hickernell, F. J., Li, R., 2001. Testing multivariate uniformity and its applications. *Mathematics of Computation* 70 (233), 337–355.
- Liang, J., Ng, K., Tian, G., 2019. A class of uniform tests for goodness-of-fit of the multivariate l_p -norm spherical distributions and the l_p -norm symmetric distributions. *Annals of the Institute of Statistical Mathematics* 71 (1), 137–162.
- Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47 (1), 13–37.

- Litterman, R. B., 1986. Forecasting with Bayesian vector autoregressions - five years of experience. *Journal of Business and Economic Statistics* 4 (1), 25–38.
- Longstaff, F., Schwartz, E., 2001. Valuing American options by simulation: A simple least-squares approach. *Review of Financial Studies* 14 (1), 113–147.
- Ludwig, N., Feuerriegel, S., Neumann, D., 2015. Putting big data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests. *Journal of Decision Systems* 24 (1), 19–36.
- Lwin, K. T., Qu, R., MacCarthy, B. L., 2017. Mean-VaR portfolio optimization: A non-parametric approach. *European Journal of Operational Research* 260 (2), 751–766.
- Maciejowska, K., Nowotarski, J., 2016. A hybrid model for GEFCom2014 probabilistic electricity price forecasting. *International Journal of Forecasting* 32 (3), 1051–1056.
- Maciejowska, K., Nowotarski, J., Weron, R., 2016. Probabilistic forecasting of electricity spot prices using factor quantile regression averaging. *International Journal of Forecasting* 32 (3), 957–965.
- Maciejowska, K., Weron, R., 2013. Forecasting of daily electricity spot prices by incorporating intra-day relationships: Evidence from the UK power market, 10th International Conference in the European Energy Market.
- Maciejowska, K., Weron, R., 2016. Short-and mid-term forecasting of baseload electricity prices in the UK: The impact of intra-day price relationships and market fundamentals. *IEEE Transactions on Power Systems* 31 (2), 994–1005.
- Makridakis, S., Spiliotis, E., Assimakopoulos, V., 2018. Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS ONE* 13 (3), Article e0194889.
- Mandelbrot, B., 1963. The variation of certain speculative prices. *Journal of Business* 36 (4), 394–419.
- Manzotti, A., Pérez, F., Quiroz, A., 2002. A statistic for testing the null hypothesis of elliptical symmetry. *Journal of Multivariate Analysis* 81 (2), 274–285.
- Margrabe, W., 1978. The value of an option to exchange one asset for another. *Journal of Finance* 33 (1), 177–186.
- Markowitz, H., 1952. Portfolio selection. *Journal of Finance* 7 (1), 77–91.
- Markowitz, H., 1959. Portfolio selection: Efficient diversification of investments. John Wiley & Sons, New York.
- Markowitz, H., 2000. Mean-variance analysis in portfolio choice and capital markets. Fabozzi Associates, New Hope.
- Markowitz, H., 2014. Mean-variance approximations to expected utility. *European Journal of Operational Research* 234 (2), 346–355.
- Mattera, D., Haykin, S., 1999. Support vector machines for dynamic reconstruction of a chaotic system. In: *Advances in Kernel Methods*. MIT Press, Cambridge, pp. 211–241.
- Mayhew, S., 1995. Implied volatility. *Financial Analysts Journal* 51 (4), 8–20.
- McNeil, A. J., Frey, R., Embrechts, P., 2005. Quantitative risk management: Concepts, techniques and tools. Princeton University Press, Princeton.
- Meyer, D., 2001. Support vector machines. *R News* 1 (3), 23–26.
- Meyer, J., 1987. Two moment decision models and expected utility maximization. *American Economic Review* 77, 421–430.
- Meyer, J., Rasche, R. H., 1992. Sufficient conditions for expected utility to imply mean-standard deviation rankings: Empirical evidence concerning the location and scale con-

- dition. *Economic Journal* 102 (410), 91–106.
- Miller Jr., F., Quesenberry, C., 1979. Power studies of tests for uniformity, ii. *Communications in Statistics - Simulation and Computation* 8 (3), 271–290.
- Misiorek, A., Trueck, S., Weron, R., 2006. Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models. *Studies in Nonlinear Dynamics and Econometrics* 10 (3), Article 2.
- Moreno, M., Navas, J. F., 2003. On the robustness of least-squares Monte Carlo (LSM) for pricing American derivatives. *Review of Derivatives Research* 6 (2), 107–128.
- Mossin, J., 1966. Equilibrium in a capital asset market. *Econometrica* 34 (4), 768–783.
- Nadarajah, S., Margot, F., Secomandi, N., 2017. Comparison of least squares Monte Carlo methods with applications to energy real options. *European Journal of Operational Research* 256 (1), 196–204.
- Nahil, A., Lyhyaoui, A., 2018. Short-term stock price forecasting using kernel principal component analysis and support vector machines: The case of Casablanca stock exchange. *Procedia Computer Science* 127, 161–169.
- Naumzik, C., Feuerriegel, S., 2020. Forecasting electricity prices with machine learning: Predictor sensitivity. *International Journal of Energy Sector Management*, forthcoming.
- Neftci, S. N., 2008. *Principles of Financial Engineering*. Academic Press, Cambridge, Massachusetts.
- Neyman, J., 1937. "Smooth test" for goodness of fit. *Scandinavian Actuarial Journal* 1937 (3-4), 149–199.
- Niu, D., Liu, D., Wu, D., 2010. A soft computing system for day-ahead electricity price forecasting. *Applied Soft Computing Journal* 10 (3), 868–875.
- Nogales, F. J., Contreras, J., Conejo, A. J., Espínola, R., 2002. Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems* 17 (2), 342–348.
- Nowotarski, J., Raviv, E., Trueück, S., Weron, R., 2014. An empirical comparison of alternative schemes for combining electricity spot price forecasts. *Energy Economics* 46, 395–412.
- Nowotarski, J., Weron, R., 2018. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews* 81 (1), 1548–1568.
- Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics* 92 (4), 1004–1016.
- Orlando, G., Taglialatela, G., 2017. A review on implied volatility calculation. *Journal of Computational and Applied Mathematics* 320, 202–220.
- Ortobelli, S., Rachev, S., Stoyanov, S., Fabozzi, F., Biglova, A., 2005. The proper use of risk measures in portfolio theory. *International Journal of Theoretical and Applied Finance* 8 (8), 1107–1133.
- Owen, J., Rabinovitch, R., 1983. On the class of elliptical distributions and their applications to the theory of portfolio choice. *Journal of Finance* 38 (3), 745–752.
- Panagiotelis, A., Smith, M., 2008. Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions. *International Journal of Forecasting* 24 (4), 710–727.

- Paoletta, M., 2019. Linear models and time-series analysis: Regression, ANOVA, ARMA and GARCH. John Wiley & Sons, Hoboken.
- Peiró, A., 1999. Skewness in financial returns. *Journal of Banking and Finance* 23 (6), 847–862.
- Pindyck, R. S., Rubinfeld, D. L., 1998. *Econometric Models and Economic Forecasts*, 4th Edition. McGraw Hill, New York.
- Politis, D. N., Romano, J. P., 1994. The stationary bootstrap. *Journal of the American Statistical Association* 89 (428), 1303–1313.
- Pórtoles, J., Gonzáles, C., Moguerza, J., 2018. Electricity price forecasting with dynamic trees: A benchmark against the random forest approach. *Energies* 11, Article 1588.
- Poterba, J., 1988. Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics* 22 (1), 27–59.
- Pratt, J., 1964. Risk aversion in the small and in the large. *Econometrica* 32 (1/2), 122–136.
- Pulley, L. B., 1983. Mean-variance approximations to expected logarithmic utility. *Operations Research* 31 (4), 685–696.
- Quesenberry, C., Miller Jr., F., 1977. Power studies of some tests for uniformity. *Journal of Statistical Computation and Simulation* 5 (3), 169–191.
- Raviv, E., Bouwman, K. E., van Dijk, D., 2015. Forecasting day-ahead electricity prices: Utilizing hourly prices. *Energy Economics* 50, 227–239.
- Ray, P., Jenamani, M., 2016. Mean-variance analysis of sourcing decision under disruption risk. *European Journal of Operational Research* 250 (2), 679–689.
- Ross, S. A., 1978. Mutual fund separation in financial theory - the separating distributions. *Journal of Economic Theory* 17 (2), 254–286.
- Rothschild, M., Stiglitz, J. E., 1970. Increasing risk: I. a definition. *Journal of Economic Theory* 2 (3), 225–243.
- Sansom, D. C., Downs, T., Saha, T. K., 2003. Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants. *Journal of Electrical and Electronics Engineering, Australia* 22 (3), 227–233.
- Schachermayer, W., Teichmann, J., 2008. How close are the option pricing formulas of Bachelier and Black-Merton-Scholes? *Mathematical Finance* 18 (1), 155–170.
- Schölkopf, B., Burges, C. J., Smola, A. J., 1998. *Advances in Kernel Methods: Support Vector Learning*. MIT Press, Cambridge.
- Schölkopf, B., Smola, A. J., 2002. *Learning with Kernels*. MIT Press, Cambridge.
- Schölkopf, B., Smola, A. J., Williamson, R. C., Bartlett, P. L., 2000. New support vector algorithms. *Neural Computation* 12 (5), 1207–1245.
- Schott, J., 2002. Testing for elliptical symmetry in covariance-matrix-based analyses. *Statistics and Probability Letters* 60 (4), 395–404.
- Schuhmacher, F., 2012. The Sharpe ratio is better than you may think. *Journal of Business Economics* 82 (6), 685–705.
- Schuhmacher, F., Auer, B. R., 2014. Sufficient conditions under which SSD- and MR-efficient sets are identical. *European Journal of Operational Research* 239 (3), 756–763.
- Scrucca, L., 2013. GA: A package for genetic algorithms in R. *Journal of Statistical Software* 53 (4), 1–37.

- Sharpe, W., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19 (3), 425–442.
- Shushi, T., 2016. A proof for the conjecture of characteristic function of the generalized skew-elliptical distributions. *Statistics and Probability Letters* 119, 301–304.
- Shushi, T., 2018. Generalized skew-elliptical distributions are closed under affine transformations. *Statistics and Probability Letters* 134, 1–4.
- Simaan, Y., 1993a. Portfolio selection and asset pricing-three-parameter framework. *Management Science* 39 (5), 568–577.
- Simaan, Y., 1993b. What is the opportunity cost of mean-variance investment strategies? *Management Science* 39 (5), 578–587.
- Simaan, Y., 2014. The opportunity cost of mean-variance choice under estimation risk. *European Journal of Operational Research* 234 (2), 382–391.
- Sims, C. A., Zha, T., 1998. Bayesian methods for dynamic multivariate models. *International Economic Review* 39 (4), 949–968.
- Sinn, H.-W., 1983. *Economic decisions under uncertainty*. North-Holland Publishing Company, New York.
- Sorjamaa, A., Hao, J., Reyhani, N., Ji, Y., Lendasse, A., 2007. Methodology for long-term prediction of time series. *Neurocomputing* 70 (16–18), 2861–2869.
- Statman, M., 1987. How many stocks make a diversified portfolio? *Journal of Financial and Quantitative Analysis* 22 (3), 353–363.
- Stentoft, L., 2004. Convergence of the least squares Monte Carlo approach to American option valuation. *Management Science* 50 (9), 1193–1203.
- Stephens, M., 1970. Use of the Kolmogorov Smirnov, Cramér-von Mises and related statistics without extensive tables. *Journal of the Royal Statistical Society (Series B)* 32 (1), 115–122.
- Stock, J., Watson, M., 2006. Forecasting with many predictors. In: *Handbook of Economic Forecasting*. Vol. 1. Elsevier, Amsterdam, pp. 515–554.
- Stock, J. H., Watson, M. W., 2002. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97 (460), 1167–1179.
- Strang, G., 2009. *Introduction to linear algebra*, 4th Edition. Wellesley-Cambridge Press, Wellesley.
- Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* 8 (1), Article 25.
- Strobl, C., Zeileis, A., 2008. Danger: High power! - Exploring the statistical properties of a test for random forest variable importances, Ludwig-Maximilians-Universität München, Department of Statistics, Technical Report 17.
- Szkuta, B., Sanabria, L. A., Dillon, T. S., 1999. Electricity price short-term forecasting using artificial neural networks. *IEEE Transactions on Power Systems* 14 (3), 851–857.
- Tibshirani, R., 1996. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)* 58 (1), 267–288.
- Timmermann, A., 2006. Forecast combinations. In: *Handbook of Economic Forecasting*. Vol. 1. Elsevier, Amsterdam, pp. 135–196.

- Tobin, J., 1958. Liquidity preference as behavior towards risk. *Review of Economic Studies* 25 (2), 65–86.
- Ugurlu, U., Oksuz, I., Tas, O., 2018. Electricity price forecasting using recurrent neural networks. *Energies* 11 (5), Article 1255.
- Uniejewski, B., Nowotarski, J., Weron, R., 2016. Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies* 9 (8), Article 621.
- Vapnik, V. N., 1995. *The Nature of Statistical Learning Theory*. Springer, New York.
- Vapnik, V. N., 1998. *Statistical Learning Theory*. John Wiley & Sons, Hoboken.
- Vassalos, M., Dillon, C., Childs, P., 2012. Empirically testing for the location-scale condition: A review of the economic literature. *Journal of Risk Model Validation* 6 (3), 51–66.
- Versluis, C., 2006. Option pricing: Back to the thinking of Bachelier. *Applied Financial Economics Letters* 2 (3), 205–209.
- Wahab, M., Lee, C., 2011. Pricing swing options with regime switching. *Annals of Operations Research* 185 (1), 139–160.
- Watson, G., 1961. Goodness-of-fit tests on a circle. *Biometrika* 49 (1-2), 109–114.
- Watson, G., 1962. Goodness-of-fit tests on a circle. ii. *Biometrika* 49 (1/2), 57–63.
- Weron, R., 2006. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. John Wiley & Sons, Hoboken.
- Weron, R., 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting* 30 (4), 1030–1081.
- Weron, R., Misorek, A., 2008. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International Journal of Forecasting* 24 (4), 744–763.
- Whaley, R. E., 1993. Derivatives on market volatility: Hedging tools long overdue. *Journal of Derivatives* 1 (1), 71–84.
- Xing, Y., Zhang, X., Zhao, R., 2010. What does the individual option volatility smirk tell us about future equity returns? *Journal of Financial and Quantitative Analysis* 45 (3), 641–662.
- Yan, X., Chowdhury, N., 2010. Mid-term electricity market clearing price forecasting: A hybrid LSSVM and ARMAX approach. *International Journal of Electrical Power and Energy Systems* 53 (1), 20–26.
- Yang, Z., Ewald, C.-O., Xiao, Y., 2009. Implied volatility from Asian options via Monte Carlo methods. *International Journal of Theoretical and Applied Finance* 12 (02), 153–178.
- Yoon, H., Yang, K., Shahabi, C., 2005. Feature subset selection and feature ranking for multivariate time series. *IEEE Transactions on Knowledge and Data Engineering* 17 (9), 1186–1198.
- Yu, J., 2012. Bias in the estimation of the mean reversion parameter in continuous time models. *Journal of Econometrics* 169 (1), 114–122.
- Zakamouline, V., Koekebakker, S., 2009. Portfolio performance evaluation with generalized Sharpe ratios: Beyond the mean and variance. *Journal of Banking and Finance* 33 (7), 1242–1254.
- Zhou, Y. H., Scheller-Wolf, A., Secomandi, N., Smith, S., 2016. Electricity trading and negative prices: Storage vs. disposal. *Management Science* 62 (3), 880–898.

- Ziel, F., 2016. Forecasting electricity spot prices using LASSO: On capturing the autoregressive intraday structure. *IEEE Transactions on Power Systems* 31 (6), 4977–4987.
- Ziel, F., Steinert, R., Husmann, S., 2014. Efficient modeling and forecasting of electricity spot prices. *Energy Economics* 47, 98–111.
- Ziel, F., Weron, R., 2018. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics* 70, 396–420.
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2), 301–320.

Appendix A

Appendix of Chapter 2

A.1 Proofs

Proof of Lemma 1. Assume that the X_i ($i = 1, \dots, n$) satisfy (2.2.1) where Y is a real-valued random variable and, using Footnote 5, $Z|Y \sim E_n(0, \Sigma, \psi)$ is n -dimensionally elliptically distributed about the origin with covariance matrix Σ and characteristic generator ψ . Note that $\sum_{i=1}^n w_i r = r \sum_{i=1}^n w_i = r$, if $\sum_{i=1}^n w_i = 1$. For $w \in \mathbb{R}^n$ with $w'e = 1$, we hence have $P = w'X = r + Y \sum_{i=1}^n w_i \beta_i + \sum_{i=1}^n w_i \gamma_i Z_i$. Set $\tilde{\beta} := \sum_{i=1}^n w_i \beta_i$ and $\tilde{Z} := \sum_{i=1}^n w_i \gamma_i Z_i$. Because linear combinations of elliptical random vectors remain elliptical with the same characteristic generator (see McNeil et al., 2005, Section 3.3.3), it follows that $\tilde{Z}|Y \sim E_1(0, \sigma^2, \psi)$ is elliptically distributed with variance $\sigma^2 = c'\Sigma c$, where $c := (w_1 \gamma_1, \dots, w_n \gamma_n)'$. \square

Proof of Proposition 1. Assume that X satisfies (2.2.1). From Lemma 1, it follows that portfolio returns $P = w'X$ belong to the same skew-elliptical GLS family for every $w \in \mathbb{R}^n$ with $w'e = 1$. To prove the proposition it is hence sufficient that the distribution of every primary asset X_i ($i = 1, \dots, n$) is determined by its mean and variance. The two equations $\mathbb{E}[X_i] = r + \beta_i \mathbb{E}[Y] + \gamma_i \mathbb{E}[Z_i] = r + \beta_i \mathbb{E}[Y]$ and $\text{Var}[X_i] = \beta_i^2 \text{Var}[Y] + \gamma_i^2 \text{Var}[Z_i] = \beta_i^2 + \gamma_i^2$ can be solved for β_i and $|\gamma_i|$, i. e., $\beta_i = (\mathbb{E}[X_i] - r)/\mathbb{E}[Y]$ and $|\gamma_i| = (\text{Var}[X_i] - ((\mathbb{E}[X_i] - r)/\mathbb{E}[Y])^2)^{\frac{1}{2}}$. Note that $\mathbb{E}[Y] \neq 0$ such that these terms are well-defined. Recall that Z_i and $-Z_i$ have the same distribution such that β_i and $|\gamma_i|$ are in fact enough to determine the distribution of X_i . \square

Proof of Theorem 1. We have to prove that the random vector X satisfies (2.3.1) if and only if X satisfies (2.2.1) with no constant.

" \Leftarrow ": Assume that X satisfies condition (2.2.1) with no constant, i. e., $X_i = \beta_i Y + \gamma_i Z_i$ for $i = 1, \dots, n$. Additionally, we have by definition $\beta_i \neq 0$ for one $i = 1, \dots, n$. Without loss of generality, let $\beta_1 \neq 0$.

FIRST STEP: Find a non-singular matrix T such that $\mathbb{E}[S] = 0$ and $\text{Cov}[S] = I_{n-1}$, where $(m, S')' = TX$.¹ First, choose

$$T_1 := \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\beta_2/\beta_1 & 1 & 0 & \cdots & 0 \\ -\beta_3/\beta_1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\beta_n/\beta_1 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

and set $\tilde{X} := (\tilde{X}_1, \dots, \tilde{X}_n)' := T_1 X$. Note that we have $\tilde{X}_i = \beta_i Y + \gamma_i Z_i - \beta_i/\beta_1 \cdot (\beta_1 Y + \gamma_1 Z_1) = \gamma_i Z_i - \gamma_1 \beta_i/\beta_1 Z_1$ and hence $\mathbb{E}[\tilde{X}_i] = 0$ for $i = 2, \dots, n$. Second, set $\tilde{\Sigma} := \text{Cov}[(\tilde{X}_2, \dots, \tilde{X}_n)]$ and let $\tilde{\Sigma} = Q\Lambda Q'$ be its spectral decomposition, i. e., Q is an orthogonal matrix and $\Lambda = \text{diag}(\lambda_2, \dots, \lambda_n)$ is a diagonal matrix with $\lambda_i > 0$ for $i = 2, \dots, n$ (see Strang, 2009, Section 6.4). Define the block matrix

$$T_2 := \begin{pmatrix} 1 & 0 \\ 0 & \tilde{\Sigma}^{-\frac{1}{2}} \end{pmatrix}$$

where $\tilde{\Sigma}^{-\frac{1}{2}} := Q\Lambda^{-\frac{1}{2}}Q'$ is the inverse of the positive definite square root of $\tilde{\Sigma}$ and we have $\Lambda^{-\frac{1}{2}} := \text{diag}(\lambda_2^{-\frac{1}{2}}, \dots, \lambda_n^{-\frac{1}{2}})$. Let $(m, S')' := TX$, where $T := T_2 T_1$. Note that T is non-singular because both T_1 and T_2 are non-singular. Further, we have $\mathbb{E}[S] = 0$ due to T_1 and $\text{Cov}[S] = I_{n-1}$ due to T_2 . This is because $\text{Cov}[TX] = \text{Cov}[T_2 \tilde{X}] = T_2 \text{Cov}[\tilde{X}] T_2'$ and hence $\text{Cov}[S] = \tilde{\Sigma}^{-\frac{1}{2}} \tilde{\Sigma} (\tilde{\Sigma}^{-\frac{1}{2}})' = Q\Lambda^{-\frac{1}{2}} Q' Q\Lambda Q' Q\Lambda^{-\frac{1}{2}} Q' = Q\Lambda^{-\frac{1}{2}} \Lambda \Lambda^{-\frac{1}{2}} Q' = Q I_{n-1} Q' = I_{n-1}$.

SECOND STEP: Show that S given m is spherically distributed. Similar to the proof of Lemma 1, we can easily verify that, for $r = 0$, all linear transformations of random vectors satisfying condition (2.2.1) inherit the GLS property, i. e., we have the representation $TX = (\tilde{\beta}_i Y + \tilde{\gamma}_i \tilde{Z}_i)_{i=1, \dots, n}$ with some suitable $\tilde{\beta}_i$, $\tilde{\gamma}_i$ and elliptically distributed $\tilde{Z} = (\tilde{Z}_1, \dots, \tilde{Z}_n)'$, conditional on Y . Because $\mathbb{E}[S] = (0, \dots, 0)'$, $\mathbb{E}[Y] \neq 0$ and $\mathbb{E}[\tilde{Z}_2, \dots, \tilde{Z}_n] = (0, \dots, 0)'$, we have $\tilde{\beta}_1 \neq 0$ and $\tilde{\beta}_2 = \dots = \tilde{\beta}_n = 0$, i. e., $S = (\tilde{\gamma}_2 \tilde{Z}_2, \dots, \tilde{\gamma}_n \tilde{Z}_n)'$. We have $\mathbb{E}[m] \neq 0$ as $\tilde{\beta}_1 \neq 0$ and $\mathbb{E}[Y] \neq 0$. Further, as $(\tilde{Z}_1, S')'$ is elliptically distributed (conditional on Y), the conditional distribution of S given \tilde{Z}_1 (and Y) is also elliptical, although generally with different characteristic generator (see McNeil et al., 2005, Section 3.3.3). For $\tilde{\gamma}_1 = 0$, we

¹Chamberlain (1983) does not prove this statement because arguments are drawn from basic linear algebra only. Nonetheless, we provide a proof for the sake of completeness.

have $m = \tilde{\beta}_1 Y$ such that $S|m$ is elliptically distributed. In the case of $\tilde{\gamma}_1 \neq 0$, $S|m$ is also elliptically distributed because S is elliptically distributed conditional on $\{Y = y, \tilde{Z}_1 = z\}$ for every $y, z \in \mathbb{R}$ and $\{m = k\} = \{\tilde{\beta}_1 Y + \tilde{\gamma}_1 \tilde{Z}_1 = k\} = \bigcup_{l \in \mathbb{R}} \{Y = l/\tilde{\beta}_1, \tilde{Z}_1 = (k - l)/\tilde{\gamma}_1\}$. Finally, considering $\text{Cov}[S] = I_{n-1}$, it follows that S given m is spherically distributed.

" \Rightarrow ": Assume that there is a non-singular matrix T such that $TX = (m, S')'$. If we denote $T^{-1} = (\bar{t}_{i,j})_{i,j=1,\dots,n}$, we have $X = T^{-1}(m, S')'$ and hence $X_i = \bar{t}_{i,1}m + \sum_{j=2}^n \bar{t}_{i,j}S_j$ for $i = 1, \dots, n$. Setting $\beta_i := \bar{t}_{i,1}$, $Y := m$ and $\tilde{Z}_i := \sum_{j=2}^n \bar{t}_{i,j}S_j$, we have the representation $X_i = \beta_i Y + \tilde{Z}_i$. As elliptical distributions are obtained by linear transformations of spherical distributions (see McNeil et al., 2005, Section 3.3.2), $\tilde{Z} = (\tilde{Z}_1, \dots, \tilde{Z}_n)$ is elliptically distributed, conditional on Y . \square

Proof of Theorem 2. As we have already proven the sufficient condition in Proposition 1, we now turn to the necessary condition. Assume that $\mathbb{E}[X_i] \neq r$ for at least one $i = 1, \dots, n$ and the distribution of $P = w_0 r + w'X$ is determined by its mean and variance for every $(w_0, w) \in \mathbb{R}^{n+1}$ with $w_0 + w'e = 1$. Note that $\mathbb{E}[P - r] = \mathbb{E}[P] - r$, $\text{Var}[P - r] = \text{Var}[P]$ and $\{(w_0, w) : (w_0, w) \in \mathbb{R}^{n+1} \text{ and } w_0 + w'e = 1\} = \{(1 - w'e, w) : w \in \mathbb{R}^n\}$. Hence, in light of the fact that, for the excess returns \bar{P} and \bar{X} , we have $\bar{P} = w'\bar{X}$ under the full investment constraint, our assumption states that the distribution of $w'\bar{X}$ is determined by its mean and variance for every $w \in \mathbb{R}^n$. Note that $\mathbb{E}[\bar{X}_i] \neq 0$ for at least one $i = 1, \dots, n$. Following from Chamberlain (1983, Theorem 2), \bar{X} must satisfy (2.3.1). By Theorem 1 the family of distributions (2.2.1) with no constant and (2.3.1) are equivalent. \square

Proof of Corollary 1. " \Leftarrow ": Assume the X_1, \dots, X_n in X satisfy (2.3.3) and, furthermore, w_0 and the w_1, \dots, w_n in w are real numbers with $w_0 + w'e = 1$. Then, $P = w_0 X_0 + w'X = (w_0 + w'e)\mathcal{R} + w'\beta Y + w'\xi$ with $\beta := (\beta_1, \dots, \beta_n)'$ and $\xi := (\gamma_1 Z_1, \dots, \gamma_n Z_n)'$. Note that $(w_0 + w'e)\mathcal{R} = \mathcal{R}$, $w'\beta \in \mathbb{R}$ and $w'\xi$ is elliptically distributed about the origin (with the same arguments as before). Thus, in analogy to the proof of Proposition 1, we have mean-variance determination.

" \Rightarrow ": To prove the necessary condition, we can proceed similar to the proof of Theorem 2. Assume that $P = w_0 X_0 + w'X$ is determined by its mean and variance for every $(w_0, w) \in \mathbb{R}^{n+1}$ with $w_0 + w'e = 1$. Let $\tilde{X}_i := X_i - \mathcal{R}$ and note that $\tilde{P} := P - \mathcal{R} = w'X + (w_0 - 1)\mathcal{R} = (w_0 - 1)\mathcal{R} + w'\tilde{X} + w'e\mathcal{R} = (w'e + w_0 - 1)\mathcal{R} + w'\tilde{X} = w'\tilde{X}$ similar to before. Further, we have $\mathbb{E}[\tilde{P}] = \mathbb{E}[P - \mathcal{R}] = \mathbb{E}[P] - \mathbb{E}[\mathcal{R}]$ and $\text{Var}[\tilde{P}] = \text{Var}[w'\tilde{X}] = \text{Var}[w'(X - e\mathcal{R})] + \text{Var}[\mathcal{R}] - \text{Var}[\mathcal{R}] = \text{Var}[w'(X - e\mathcal{R}) + \mathcal{R}] - \text{Var}[\mathcal{R}] = \text{Var}[w'(X - e\mathcal{R}) + (w_0 + w'e)\mathcal{R}] - \text{Var}[\mathcal{R}] = \text{Var}[w_0 \mathcal{R} + w'X] - \text{Var}[\mathcal{R}] = \text{Var}[P] - \text{Var}[\mathcal{R}]$. Here we used that $X_i - \mathcal{R}$ and \mathcal{R} are independent for each $i = 1, \dots, n$, such that $w'(X - e\mathcal{R})$ is independent of \mathcal{R} . Hence, our assumption states that the distribution of $w'\tilde{X}$ is determined by its

mean and variance for every $w \in \mathbb{R}^n$. Note that $\mathbb{E}[\tilde{X}_i] \neq 0$ for at least one $i = 1, \dots, n$, as $\mathbb{E}[X_i] \neq \mathbb{E}[\mathcal{R}]$ for at least one $i = 1, \dots, n$. Following from Chamberlain (1983, Theorem 2), \tilde{X} must satisfy (2.3.1). In analogy to our Theorem 1, X must follow (2.3.3). \square

Proof of Corollary 5. For a portfolio return $P = w_0 r + w'X$ with $(w_0, w) \in \mathbb{R}^{n+1}$ subject to $w_0 + w'e = 1$, we have $\mathbb{E}[P] = r + w'\beta\mu_Y$ and $\text{Var}[P] = w'Vw$, where $V = \Sigma + \beta\beta'$. Recall that $\mu_Y := \mathbb{E}[Y] \neq 0$ and $\text{Var}(Y) = 1$. Following Simaan (1993a, Theorem 3), the distribution of P is a function of $w'\beta\mu_Y$ and $w'\Sigma w$. Similar to Simaan (1993a, Theorem 4), for given $\mathbb{E}[P] - r = w'\beta\mu_Y$, the expected utility $\mathbb{E}[u(w_0 r + w'X)]$ is a non-increasing function of $w'\Sigma w$, i. e., the elliptical component of the portfolio's variance. Thus, in line with Simaan (1993a, Corollary 4.1), there exists a $\mu \in \mathbb{R}$ which corresponds to the optimal value of $(\mathbb{E}[P] - r)/\mu_Y$ such that the portfolio problem (P) is equivalent to

$$\begin{aligned} \min_{(w_0, w) \in \mathbb{R}^{n+1}} \quad & \frac{1}{2} w' \Sigma w \\ \text{subject to} \quad & w' \beta = \mu \\ & w_0 + w'e = 1. \end{aligned} \tag{P*}$$

The Lagrangian and first-order conditions are

$$\begin{aligned} L &= \frac{1}{2} w' V w - \frac{1}{2} (w' \beta)^2 + \lambda_1 (\mu - w' \beta) + \lambda_2 (1 - w_0 - w'e) \\ \frac{dL}{dw_0} &= -\lambda_2 = 0 \\ \frac{dL}{dw} &= Vw - w' \beta \beta - \lambda_1 \beta - \lambda_2 e = 0 \end{aligned}$$

and consequently we have $0 = Vw - (\lambda_1 + w' \beta) \beta = Vw - (\lambda_1 + \mu) \beta$. Solving for w yields the solution $w^* = \alpha f$, where $\alpha := (\lambda_1 + \mu)(e' V^{-1} \beta)$ and f is the tangency portfolio

$$f = \frac{V^{-1} \beta}{e' V^{-1} \beta}.$$

Further, we have

$$w_0^* = 1 - e' w^* = 1 - e' \left(\alpha \frac{V^{-1} \beta}{e' V^{-1} \beta} \right) = 1 - \alpha \frac{e' V^{-1} \beta}{e' V^{-1} \beta} = 1 - \alpha.$$

\square

A.2 Robustness checks

Because our empirical conclusions may be sensitive to some of the settings in our research design, Table A.1 reports the results of several robustness checks.

	Watson			Liang
		Symmetric	Centered	Star
Survivorship adjustment				
3 portfolios, 2 stocks	0.91	0.76	4.77	6.94
3 portfolios, 4 stocks	0.35	0.97	0.51	4.71
3 portfolios, 6 stocks	0.55	1.45	2.00	5.28
3 portfolios, 8 stocks	0.07	-0.35	0.35	0.87
5 portfolios, 2 stocks	0.13	-0.28	2.70	-2.08
5 portfolios, 4 stocks	0.02	0.78	0.37	-0.56
5 portfolios, 6 stocks	0.04	1.82	1.32	-0.38
5 portfolios, 8 stocks	0.08	1.41	0.65	-1.57
7 portfolios, 2 stocks	0.42	-1.86	4.86	-3.10
7 portfolios, 4 stocks	0.21	-0.57	2.88	-5.00
7 portfolios, 6 stocks	0.32	-1.42	3.94	-3.88
7 portfolios, 8 stocks	0.13	-0.91	2.02	-1.33
25:75 sample split				
FF 6 size/book-to-market	0.14	4.99	0.16	3.03
FF 6 size/momentum	0.21	5.07	-0.61	2.64
FF 6 size/reversal	0.43	5.58	-2.92	1.65
FF 5 industries	0.54	2.22	3.38	-4.48
FF 10 industries	0.07	0.82	2.61	-0.38
75:25 sample split				
FF 6 size/book-to-market	0.36	2.88	3.05	-1.84
FF 6 size/momentum	0.08	2.23	1.77	-1.44
FF 6 size/reversal	0.24	0.57	0.19	-2.31
FF 5 industries	0.15	0.67	-0.72	-1.93
FF 10 industries	0.14	0.79	3.08	-1.58
Non-equity common factor				
GS 5 commodity indices	0.12	1.94	-0.97	-0.85
CS 13 hedge fund indices	0.44	9.28	8.57	5.09

Extending the results of Tables 2.1 and 2.2, this table presents the outcomes of three robustness checks. In the first one, we account for potential survivorship bias by using an equal-weighted portfolio of our stock selection as the common factor. The second sensitivity check changes the sample split rule (estimation sample A vs. test sample B) for Fama-French portfolios from the initially used 50:50 to 25:75 and 75:25. Finally, the last subanalysis for alternative asset classes replaces the stock market index underlying our main calculations by the aggregates of the GSCI and the CSHFI.

TABLE A.1: Robustness checks

First, since following the Meyer and Rasche (1992) portfolio construction approach introduces a potential survivorship bias, we replace the CRSP index with an equal-weighted portfolio of the 503 stocks under consideration.² Second, besides implementing a 50:50 split, we extend our investigation of the Fama-French portfolios by also using 25:75 and 75:25 to determine the sizes of our estimation (A) and test (B) samples. This is because, similar to extreme value research (see Gilli and K llezi, 2006), there may be a trade-off between accurate parameter estimates in the former and the number of available observations for testing in the latter. Finally, while our main calculations follow Fung et al. (2008) by using an equity index as the market proxy when analyzing hedge fund (and commodity) data, we also performed our tests with the aggregate CSHFI (GSCI) instead.

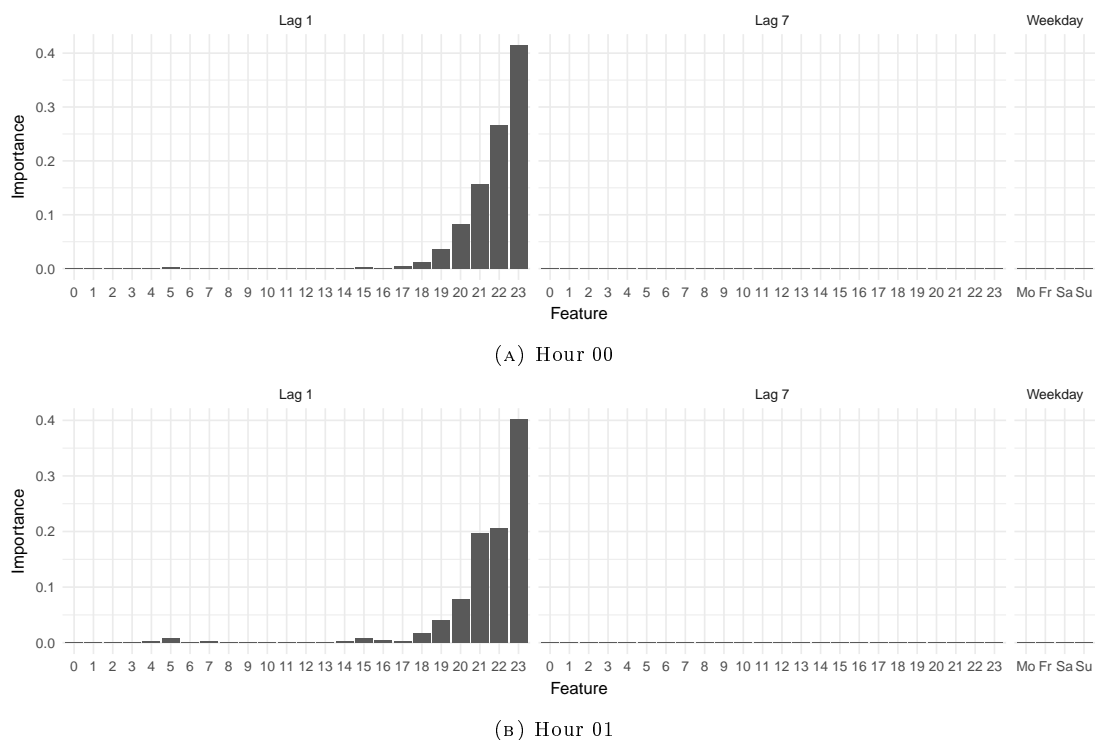
²Using a value-weighted portfolio or even the S&P 500 index yields similar results.

Looking at the outcomes, it is reassuring that the overall picture of mainly insignificant test statistics continues to hold. As far as the survivorship adjustment is concerned, we have slightly more cases of rejection than before. With respect to the sample split, we observe that there are significantly fewer (more) rejections for the Fama-French portfolios when a larger (smaller) estimation sample is used. Finally, the two alternative market indices do not change our conclusions for commodities and hedge funds.

Appendix B

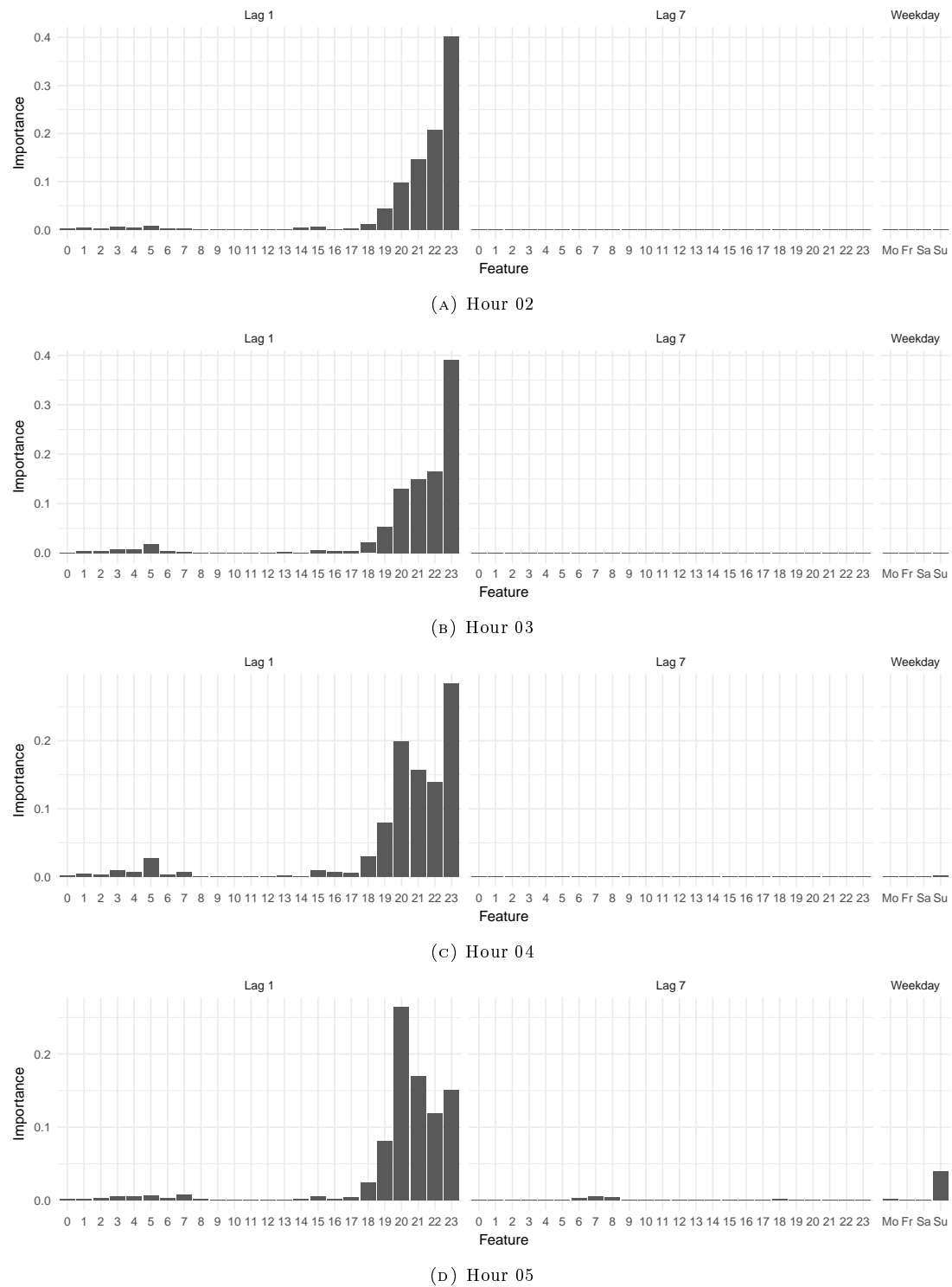
Appendix of Chapter 3

B.1 Additional empirical results



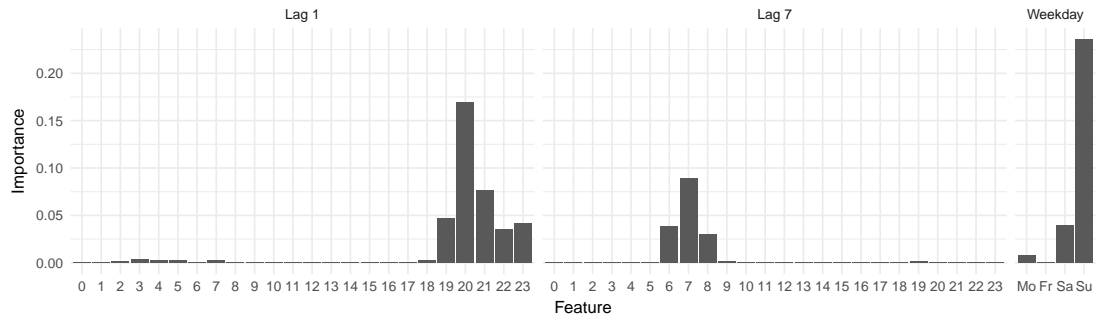
This figure plots the relative permutation importance scores (computed via the random forest algorithm of Section 3.2.3) of lagged price variables $P_{d-l,k}$ (lags $l = 1, 7$ days) and weekday dummies W_i (days $i = 0, 1, 5, 6$) for the explanation of all hourly price variables belonging to the specified daytime ‘night’. It is continued in Figure B.2.

FIGURE B.1: Importance profiles ‘night’ I

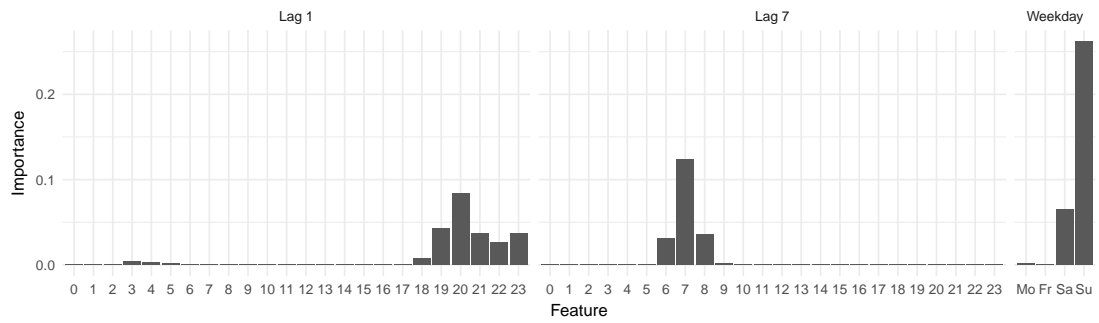


This figure continues Figure B.1.

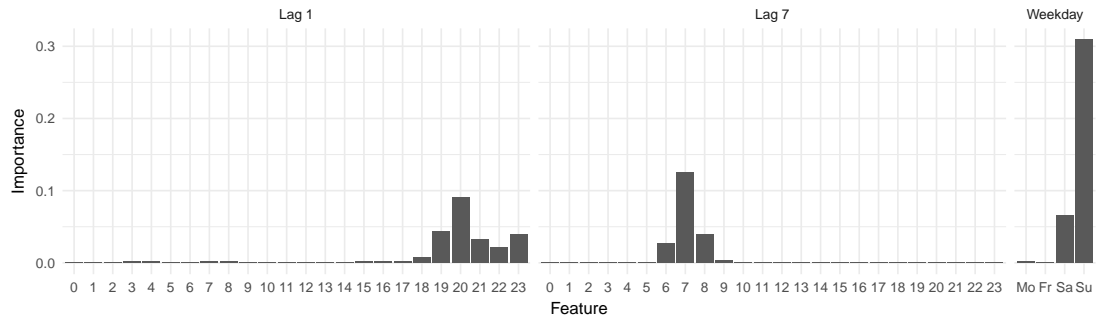
FIGURE B.2: Importance profiles 'night' II



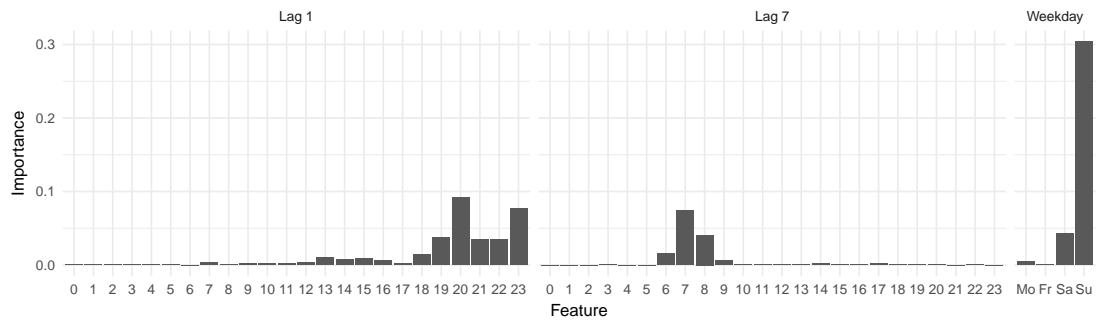
(A) Hour 06



(B) Hour 07



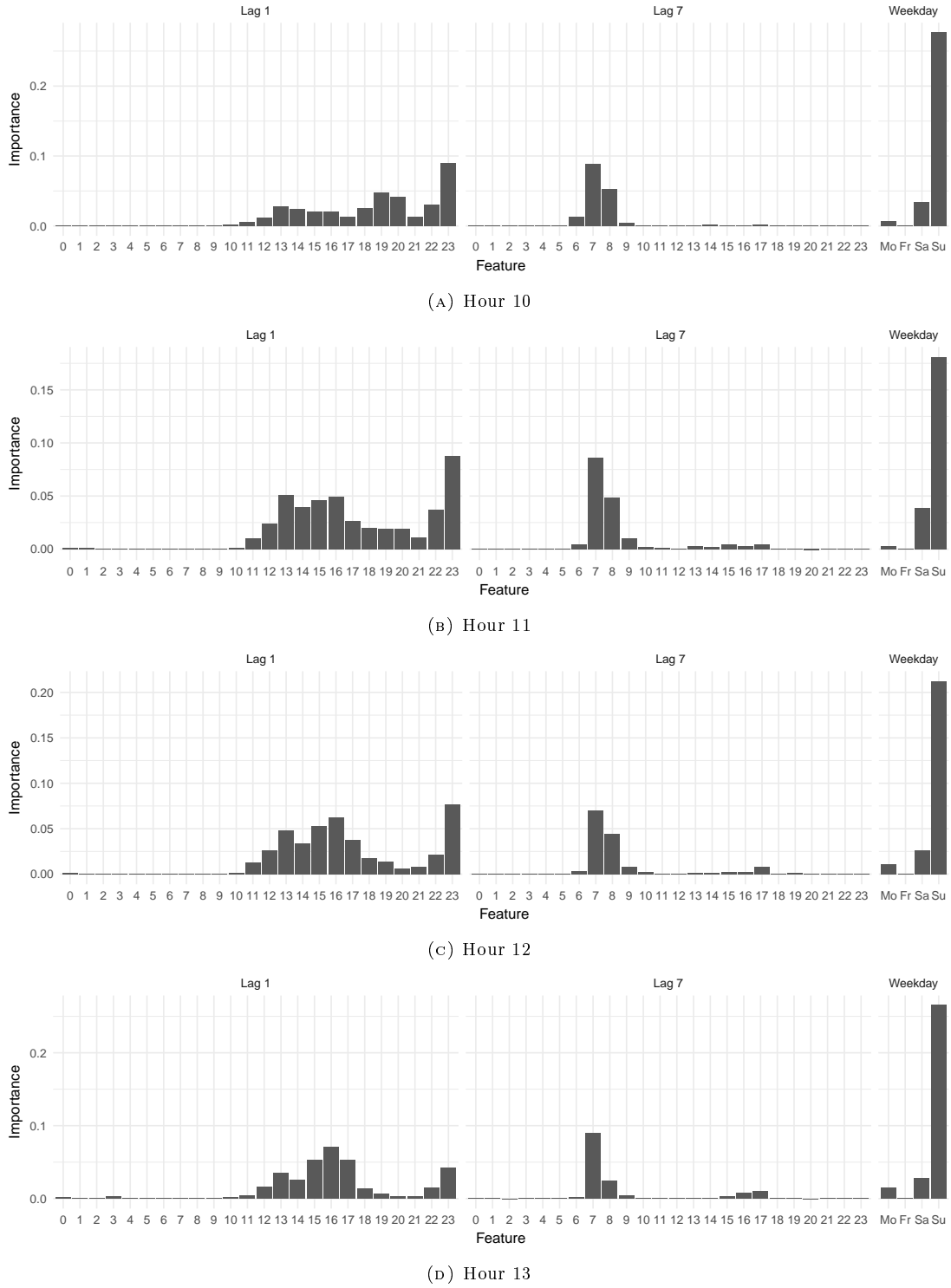
(C) Hour 08



(D) Hour 09

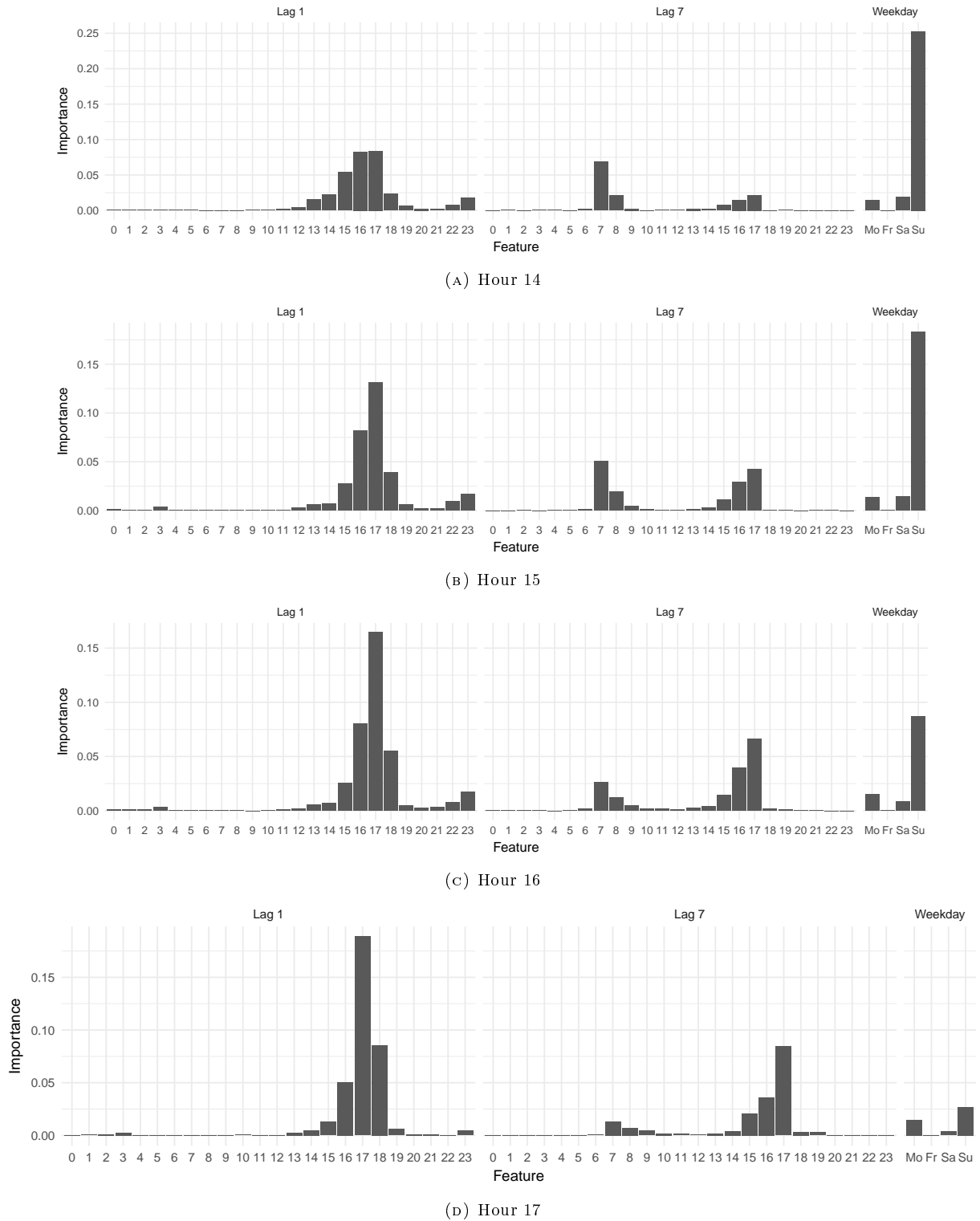
This figure plots the relative permutation importance scores (computed via the random forest algorithm of Section 3.2.3) of lagged price variables $P_{d-l,k}$ (lags $l = 1, 7$ days) and weekday dummies W_i (days $i = 0, 1, 5, 6$) for the explanation of all hourly price variables belonging to the specified daytime ‘morning’.

FIGURE B.3: Importance profiles ‘morning’



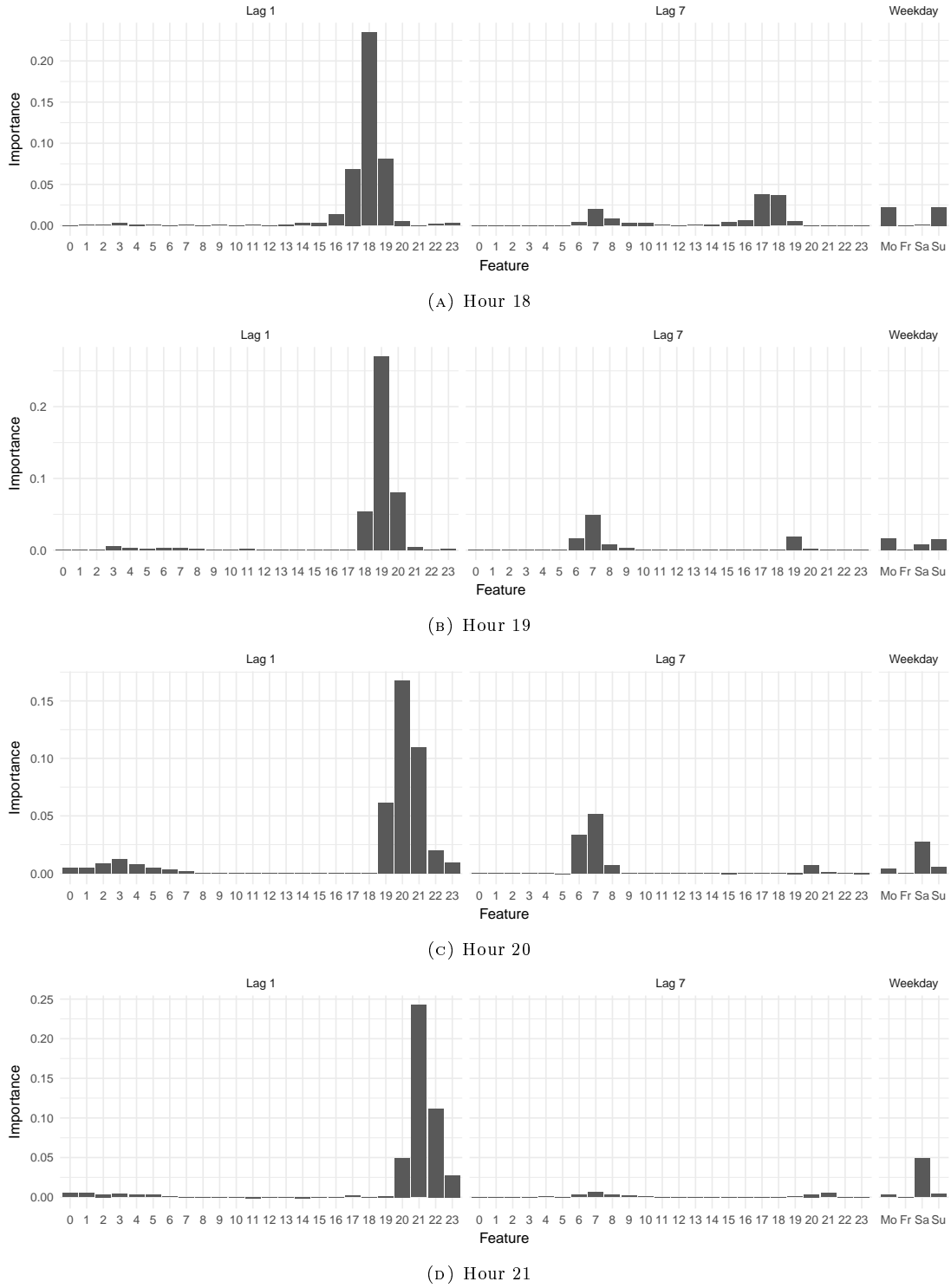
This figure plots the relative permutation importance scores (computed via the random forest algorithm of Section 3.2.3) of lagged price variables $P_{d-l,k}$ (lags $l = 1, 7$ days) and weekday dummies W_i (days $i = 0, 1, 5, 6$) for the explanation of all hourly price variables belonging to the specified daytime ‘high noon’.

FIGURE B.4: Importance profiles ‘high noon’



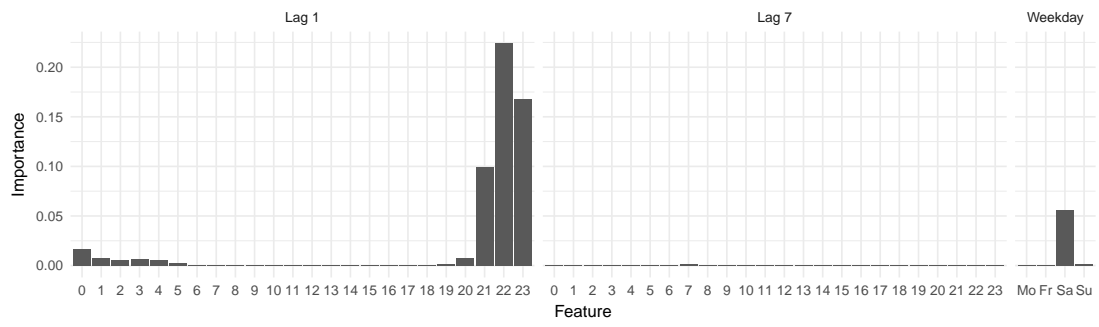
This figure plots the relative permutation importance scores (computed via the random forest algorithm of Section 3.2.3) of lagged price variables $P_{d-l,k}$ (lags $l = 1, 7$ days) and weekday dummies W_i (days $i = 0, 1, 5, 6$) for the explanation of all hourly price variables belonging to the specified daytime ‘afternoon’.

FIGURE B.5: Importance profiles ‘afternoon’

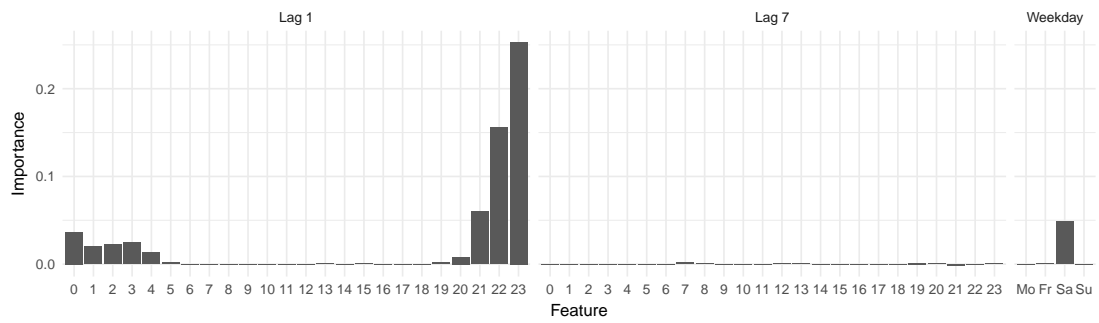


This figure plots the relative permutation importance scores (computed via the random forest algorithm of Section 3.2.3) of lagged price variables $P_{d-l,k}$ (lags $l = 1, 7$ days) and weekday dummies W_i (days $i = 0, 1, 5, 6$) for the explanation of all hourly price variables belonging to the specified daytime ‘evening’. It is continued in Figure B.7.

FIGURE B.6: Importance profiles ‘evening’ I



(A) Hour 22



(B) Hour 23

This figure continues Figure B.6.

FIGURE B.7: Importance profiles ‘evening’ II

B.2 Benchmark models

B.2.1 Persistent model

Simple persistent forecasts, which assume that price changes are completely random and thus the best forecast is to use an appropriate past hourly price, are probably the most often-used benchmarks in EPF (see Nogales et al., 2002). In our application, we generate such forecasts as follows: For a working day (Monday through Friday, ignoring holidays), we use the hourly prices of the previous working day. For a weekend day (Saturday through Sunday), we use the hourly prices of the previous week. Thus, we have

$$P_{d,h} = \begin{cases} P_{d-3,h} + \varepsilon_{d,h}, & \mathcal{W}(d) = 1, \\ P_{d-1,h} + \varepsilon_{d,h}, & \mathcal{W}(d) \in \{2, \dots, 5\}, \\ P_{d-7,h} + \varepsilon_{d,h}, & \mathcal{W}(d) \in \{0, 7\}. \end{cases} \quad (\text{B.2.1})$$

Forecasts via this model are also called naive forecasts and, in many fields, have been shown to be superior to more complex forecasting methods (see Alquist and Kilian, 2010; Alquist et al., 2013).

B.2.2 24 univariate AR models

Ziel (2016) suggests another benchmark for evaluating new electricity price models: 24 univariate AR(p) models (henceforth abbreviated with 24AR).¹ That is, for each hour, we have

$$P_{d,h} = \phi_{h,0} + \sum_{i=0}^6 \delta_{h,i} W_i(d) + \sum_{l=1}^{p_h} \phi_{h,l} \left(P_{d-l,h} - \sum_{i=0}^6 \delta_{h,i} W_i(d-l) - \phi_{h,0} \right) + \varepsilon_{d,h}. \quad (\text{B.2.2})$$

To estimate these models, we follow Ziel (2016) by first regressing prices on weekday dummies to produce centered hourly prices. Afterward, we solve the Yule-Walker equations to estimate the parameters $\phi_{h,l}$ because their solution guarantees stationarity (see Pindyck and Rubinfeld, 1998). We select the model orders p_h by minimizing the Akaike (1974) information criterion (AIC) with a maximum upper bound of $p_{\max} = 21$.

¹One might also think of considering smoothing models (see Makridakis et al., 2018).

B.2.3 Unrestricted VAR

As a more advanced benchmark, we consider an unrestricted VAR(p) model (henceforth VAR) with exogenous weekday dummies, which has been used by, for example, Raviv et al. (2015) and Ziel and Weron (2018). That is, we specify

$$P_d = \phi_0 + \sum_{l=1}^p \Phi_l P_{d-l} + \sum_{i=0}^6 D_i W_i(d) + \varepsilon_d, \quad (\text{B.2.3})$$

with (24×1) intercept vector ϕ_0 , (24×24) parameter matrices Φ_l and (24×1) weekday dummy parameter vectors D_i . We estimate the VAR model via equation-by-equation ordinary least squares (OLS) and choose the lag order p by minimizing the AIC for a maximum lag order of $p_{\max} = 21$. For our data, this yields $p = 2$.

B.2.4 Expert VAR

Expert models use a fixed parsimonious autoregressive structure based on prior knowledge of experts (see Maciejowska et al., 2016; Misiolek et al., 2006). Because the one- and seven-day lags are often considered the most important (see Weron and Misiolek, 2008), a straightforward expert model would include all lagged price variables $P_{d-l,h}$ for the lags $l = 1, 7$. Moreover, experts would add weekday dummy variables W_i for Monday, Friday, Saturday and Sunday, i.e., $i = 0, 1, 5, 6$. In the following, we refer to this selection of lagged and dummy variables as expert choice and note that it could also have been derived from our random forest importance scores. Additionally considering that the most common expert models are linear, we come up with the expert VAR model

$$P_d = \phi_0 + \Phi_1 P_{d-1} + \Phi_7 P_{d-7} + \sum_{i=0,1,5,6} D_i W_i(d) + \varepsilon_d, \quad (\text{B.2.4})$$

with (24×1) intercept vector ϕ_0 , (24×24) parameter matrices Φ_l for lags $l = 1, 7$ and (24×1) weekday dummy parameter vectors D_i . We estimate via equation-by-equation OLS.

B.2.5 Expert neural network

Besides the standard time series models of the previous sections, we use well-established machine learning methods as benchmarks. Because energy market research (including Catalão et al., 2007; Kuo and Huang, 2018; Lago et al., 2018; Szkuta et al., 1999) has paid much attention to neural networks (NNs), we set up an expert NN. The network architecture of such models is inspired by nature and consists of connected network nodes

(or neurons). The simplest network, a single-layer perceptron (SLP), covers only an input and an output layer of nodes and is equivalent to a linear regression. Each node value of the output layer is derived by a linear combination of the input node values. The weights connecting the nodes correspond to the coefficients of the regression and are selected using a learning algorithm that minimizes, for example, the mean square error. By adding intermediate layers with hidden nodes, we obtain the multi-layer perceptron (MLP). MLPs (and SLPs) belong to the family of feed-forward networks which have no loops and are often preferred in forecasting (see Weron, 2014).² The outputs of the nodes in one layer are inputs for the next. The transfer function for the nodes in the hidden layers is often non-linear because otherwise the MLP would collapse to a SLP (see Cybenko, 1989).

Similar to García-Ascanio and Maté (2010) and Cruz et al. (2011), for each hour h , we train a MLP with one hidden layer. As input variables, we use the expert choice of Section B.2.4. The transfer function is chosen to be a tangent-sigmoid function. For the implementation, we resort to the `feedforwardnet` function provided in `Matlab`, where we set training and validation ratios of 2/3 and 1/3 for our calibration period. Learning happens via the Levenberg-Marquardt algorithm, which is a reasonable choice because it trains a network 10 to 100 times faster than back-propagation (see Amjady, 2007). The ideal number of hidden nodes complexly depends on various factors such as the number of input and output nodes, the number of training cases and the magnitude of noise in the variables. However, there are some rules of thumb for deriving a reasonable number. Blum (1992) argues that the size of the hidden layer should be somewhere between the input layer size and the output layer size. Boger and Guterman (1997) propose an interesting heuristic. They specify as many hidden nodes as there are principal components needed to capture between 70% and 90% of the input data variance. In our case, this rule suggests a number between 2 and 6. If we require 95% of the variance to be explained, the number increases up to 10. An additional rule of thumb, which can be derived from Hagan et al. (1996), states that the ratio $N_{\text{train}}/[\alpha \cdot (N_{\text{input}} + N_{\text{output}})]$ gives an upper bound on the number of hidden nodes, which will not result in overfitting. Here, N_{train} is the number of training samples, N_{input} (N_{output}) is the number of input (output) nodes and α is an arbitrary scaling factor (usually between 2 and 10). If we start with $\alpha = 2$, we obtain a node number around 7. To avoid relying exclusively on rules of thumb, we train several

²We focus on a MLP because it is frequently used in the literature and a popular parsimonious benchmark. Alternatives would be feedback networks of simple/fully recurrent nature (see Anbazhagan and Kumarappan, 2013; Andalib and Atry, 2009; Ugurlu et al., 2018) or self-organizing type (see Fan et al., 2007; Niu et al., 2010) as well as networks using fuzzy logic (see Amjady, 2006).

networks with a number of hidden nodes ranging from 1 to 12. As far as the network generalization error (the mean square error in the validation sample) is concerned, we find 8 hidden nodes to be a reasonable layer size.³

B.2.6 Expert support vector machine

Support vector machines (SVM) are another popular machine learning technique, typically occur as elements of hybrid EPF systems (see Weron, 2014) and, depending on the market, have been shown to be more accurate than NNs (see Che and Wang, 2010; Sansom et al., 2003). While, in the more advanced sections of our study, we utilize SVMs supplemented by feature selection and feature extraction techniques, we start by specifying a simple expert SVM for benchmarking. Again, we use the expert choice of Section B.2.4 to determine the input variables of the model.

SVMs were originally developed to solve pattern recognition problems (see Boser et al., 1992; Cortes and Vapnik, 1995) but can also be used as a regression method. The goal of SVM regression is to find a function of the input variables that deviates from the response variable by a value no greater than a pre-set value $\epsilon > 0$ for each training data point, and at the same time is as smooth as possible. Because there may be no function to satisfy this constraint for all points, slack variables ξ and ξ^* are introduced for each point. This allows regression errors to exist up to the value of ξ and ξ^* , yet still satisfy the required conditions. Finally, a parameter $C > 0$ is used to control the penalty imposed on observations outside the ϵ -margin.

For each day d , let the row vector x_d contain the relevant values of different features and y_d be the corresponding target output, both originating from y and X defined in Section 3.2.2. Then, the standard form of SVM regression involves solving the optimization problem

$$\begin{aligned} \min_{w,b,\xi,\xi^*} \quad & \frac{1}{2}w'w + C \sum_{d=1}^N (\xi_d + \xi_d^*) \\ \text{subject to} \quad & (w'\phi(x_d) + b) - y_d \leq \epsilon + \xi_d, \\ & y_d - (w'\phi(x_d) + b) \leq \epsilon + \xi_d^* \\ & \xi_d, \xi_d^* \geq 0, d = 1, \dots, N, \end{aligned} \tag{B.2.5}$$

³Even though Lago et al. (2018) propose some new specification rules to boost the performance of NNs, our focus is a simple benchmark derived from currently established specification standards.

where the function $\phi(x_d)$, if non-linear, maps x_d into a higher-dimensional space, and w , b are regression parameters (see Vapnik, 1998). The algorithm of Schölkopf et al. (2000) introduces a parameter $\nu \in (0, 1]$ to control the number of support vectors (see below). Also, ϵ becomes a parameter to be determined in the optimization. This setting is called SVM with ν -regression and solves

$$\min_{w, b, \xi, \xi^*, \epsilon} \frac{1}{2} w' w + C \left(\nu \epsilon + \frac{1}{N} \sum_{i=d}^N (\xi_d + \xi_d^*) \right) \quad (\text{B.2.6})$$

subject to the same constraints used in (B.2.5). This problem (also called primal problem) is computationally simpler to solve in its Lagrange dual formulation.⁴ That is, we solve

$$\min_{\alpha, \alpha^*} \frac{1}{2} (\alpha - \alpha^*)' Q (\alpha - \alpha^*) + y' (\alpha - \alpha^*) \quad (\text{B.2.7})$$

$$\text{subject to } e' (\alpha - \alpha^*) = 0, e' (\alpha + \alpha^*) \leq C\nu, \quad (\text{B.2.8})$$

$$0 \leq \alpha_d, \alpha_d^* \leq C/N, d = 1, \dots, N, \quad (\text{B.2.9})$$

where Q is a $(N \times N)$ positive semidefinite matrix and $Q_{d,l} = K(x_d, x_l) = \phi(x_d)' \phi(x_l)$ is a kernel function. After the problem is solved, the approximate function is

$$\hat{y}(x) = \sum_{d=1}^N (-\alpha_d + \alpha_d^*) K(x_d, x) + b. \quad (\text{B.2.10})$$

If either α_d or α_d^* is not 0, then the corresponding observation is called a support vector.⁵ In other words, the function to predict new values depends only on support vectors. For the kernel, we use a Gaussian radial basis function, i.e., $K(x_d, x_l) = e^{-\gamma \|x_d - x_l\|^2}$ with bandwidth parameter γ . This way, we can model a non-linear input-output relationship.

To implement our SVM regression, we have to choose suitable values for the parameters C , ν and γ . This is not easy because, for good regression performance, they have to fit the sample data. Several authors have proposed selecting parameters using cross-validation (see Schölkopf and Smola, 2002) or grid search optimization (see Cherkassky and Mulier, 1998; Schölkopf et al., 1998). However, these methods are computationally intensive and

⁴In general, the optimal values of the primal and the dual problem need not be equal. However, when the problem is convex and satisfies a constraint qualification condition, the value of the optimal solution to the primal problem is given by the solution of the dual problem.

⁵ $\alpha_d = \alpha_d^* = 0$ holds for all observations which lie strictly inside the ϵ -insensitive zone.

time-consuming. We therefore use a genetic algorithm (similar to Chen and Wang, 2007; Kang-Lin et al., 2004) to find optimal parameters for SVM regression.⁶ In a first step, we search for reasonable parameter ranges by heuristics, empirical regularities, and trial and error. In a second step, we select parameters from the derived ranges by using the real-value genetic algorithm of Scrucca (2013).

The parameter ν is a lower bound for the share of support vectors in the training sample and an upper bound on the fraction of poorly predicted training observations. Because electricity prices have a complex importance profile and, in addition, exhibit spikes and high volatility (especially for German/Austrian data), we search for relatively high parameter values in the range $\nu \in [0.5, 1]$. The cost parameter C determines the trade-off between the smoothness of the function of interest and the amount up to which deviations larger than ϵ are tolerated. Increasing the cost value achieves a closer fit to the calibration data but presents the danger of overfitting. Mattera and Haykin (1999) propose setting C based on the output values, i.e., $C = \max(y) - \min(y)$. This is reasonable but sensitive to outliers in the training data. We consider the similar heuristic $C = |\text{mean}(y)| + 3 \cdot |\text{sd}(y)|$ which, in our application, takes values between 50 and 100 across different hours h . To limit the risk of faulty heuristic-based selection, we enlarge this range and search within $C \in [10, 200]$. The kernel parameter γ controls the shape of the separating hyperplane. Increasing γ typically increases the number of support vectors. A standard value for γ is the reciprocal of the number of features. We find reasonable values (below this standard) in the range $\gamma \in [10^{-4}, 10^{-3}]$ by trial and error.

For the search algorithm of Scrucca (2013), we use a maximal number of 10 generations. The population size is 50, the elitism parameter is 2, the crossover probability 0.8 and the mutation probability 0.1. We further have to specify a fitness function for parameter evaluation. Since the goal of our SVM regression is to forecast future electricity prices, it is vital to choose a fitness function capturing generalizability. Recall that the training sample is used to estimate the SVM model for a certain parameter set, while the validation sample is used to assess the generalizability of the estimated model. We use the mean square error in the validation sample as the fitness function.

For our SVM regression, we rely on the implementation LIBSVM of Chang and Lin (2011) provided in the R package `e1071` by Meyer (2001). We set a termination criterion tolerance of 0.01 and use shrinking heuristics following Chang and Lin (2011).

⁶Hu et al. (2013) follow a related approach based on a memetic algorithm for parameter selection.

B.3 Feature extraction

While our random forest approach concentrates on selecting the most important features, another approach to dimension reduction is to extract (or generate) new features from the set of input variables and to use these new variables instead of the original ones. Maciejowska and Weron (2013) have proposed reducing the dimension of the 24-dimensional electricity price process by specifying a dynamic factor model, i.e., performing principal component analysis (PCA).

The main assumption is that all hourly prices $P_{d,h}$ co-move and depend on a small set of common factors $F_d^{(K)} = (F_{1,d}, \dots, F_{K,d})'$. The individual series $P_{d,h}$ can be modeled as a linear function of the $K \leq 24$ principal components $F_{k,d}$ and the idiosyncratic components $\nu_{d,h}$. Specifically, we have

$$P_{d,h} = \mu_h + \sum_{k=1}^K \gamma_{h,k} F_{k,d} + \nu_{d,h}, \quad (\text{B.3.1})$$

where the factor loading $\gamma_{h,k}$ describes the effect of the k th factor $F_{k,d}$ on the price $P_{d,h}$ of hour h .⁷ This model can be rewritten as a 24-dimensional time series

$$P_d = \mu + \Gamma^{(K)} F_d^{(K)} + \nu_d, \quad (\text{B.3.2})$$

where $P_d = (P_{d,0}, \dots, P_{d,23})'$ is the (24×1) price vector, $\Gamma^{(K)} = (\gamma_{h,k})_{h=0,\dots,23;k=1,\dots,K}$ the $(24 \times K)$ loading matrix, $F_d^{(K)} = (F_{1,d}, \dots, F_{K,d})'$ the $(K \times 1)$ factor score vector and $\nu_d = (\nu_{d,0}, \dots, \nu_{d,23})'$ the (24×1) vector of idiosyncratic components.

Let $P = (\bar{P}_{d,h})_{d=1,\dots,N;h=0,\dots,23}$ be the $(N \times 24)$ centered price matrix, where $\bar{P}_{d,h} = P_{d,h} - \hat{\mu}_h$ and $\hat{\mu}_h$ is the sample mean of $(P_{d,h})_{d=1,\dots,N}$. The sample covariance matrix of prices is then given by $\Sigma = P'P/N$. We decompose $\Sigma = \Theta\Lambda\Theta'$, where $\Theta = (\vartheta_0, \dots, \vartheta_{23})$ is the matrix containing the eigenvectors as column vectors and Λ is the diagonal matrix containing the corresponding eigenvalues $\lambda_0, \dots, \lambda_{23}$ ordered descending by magnitude. We calculate $\hat{\Gamma} = (\sqrt{\lambda_0} \cdot \vartheta_0, \dots, \sqrt{\lambda_{23}} \cdot \vartheta_{23})$ and truncate $\hat{\Gamma}^{(K)} = (\Gamma_1, \dots, \Gamma_K)$. The factor scores can be reconstructed by taking $\hat{F}_d = \hat{\Gamma}^{-1} P_d$ for each day d .⁸ Note that the estimated

⁷The factor loadings $\gamma_{h,k}$ are model parameters which should not be confused with power system loads.

⁸Stock and Watson (2002) show that multiplying the tuple of eigenvectors of the matrix PP' by \sqrt{N} yields a consistent estimator of the factor scores. Our calculation is equivalent to theirs.

factor scores \hat{F}_d are standardized to a variance of 1. The idiosyncratic component scores are calculated via (B.3.2) using the truncated factor scores and loading matrix.

To determine the ideal number of factors K , we compare several statistical tests. The popular approach of Bai and Ng (2002) suggests to find K by minimizing a variance criterion with a penalty term depending on the number of used factors. For their criteria IC1 and IC2, Alessi et al. (2010) have improved the finite sample performance via the calibration strategy of Hallin and Liška (2007). Applying this approach in our setting, both IC1 and IC2 advise to use $\hat{K} = 7$. Ahn and Horenstein (2013) propose selecting K by maximizing the ratio of adjacent eigenvalues. Their eigenvalue ratio test yields $\hat{K} = 6$. Alternatively, Onatski (2010) has introduced a threshold approach based on the empirical distribution of the sample covariance eigenvalues, which can be used for both stationary and non-stationary factors. For the estimated dimension, this technique provides $\hat{K} = 7$. Given these results, we use $K = 7$ factors which explain 97.3% of the intraday variability in our data.⁹

To predict future hourly prices, we have to forecast both factor scores and the idiosyncratic components. The idiosyncratic components can only be weakly correlated across periods and should therefore be modeled separately for each hour. Maciejowska and Weron (2016) use an AR(q) model to describe and forecast the hourly idiosyncratic component. That is, we specify

$$\nu_{d,h} = \sum_{l=1}^{q_h} \theta_{l,h} \nu_{d-l,h} + \varepsilon_{d,h}, \quad (\text{B.3.3})$$

where $\theta_{l,h}$ are the autoregressive parameters and $\varepsilon_{d,h}$ are the error terms. For each hour h , we solve the Yule-Walker equations where the lag order q_h is chosen by minimizing the AIC for a maximum lag order of $q_{\max} = 21$. To forecast factor scores, Maciejowska and Weron (2016) assign an unrestricted VAR model to the truncated factors. Incorporating weekday effects, this yields

$$F_d^{(K)} = \phi_0 + \sum_{l=1}^p \Phi_l F_{d-l}^{(K)} + \sum_{i=0}^6 D_i W_i(d) + \xi_d, \quad (\text{B.3.4})$$

where ϕ_0 is the $(K \times 1)$ intercept, Φ_l the $(K \times K)$ autoregressive parameter matrices and ξ_d the $(K \times 1)$ error process. The lag order p is chosen by minimizing the AIC for a maximum

⁹A smaller number of factors yields worse out-of-sample forecasts.

order of $p_{\max} = 21$. In our case, this yields $p = 7$. We refer to the overall price forecasting approach as PC(K)-VAR.

Besides this setting we can think of two related forecasting techniques. First, we can modify the previous PCA approach by applying the least absolute shrinkage and selection operator (LASSO) method (as specified in Section B.4.1) to (B.3.4). Even though PCA already reduces dimensionality, LASSO may identify additional simplification potential. Second, we may combine PCA and the SVM with ν -regression and radial basis function kernel (as specified in Section B.2.6). However, in this case, a consistent implementation (linear vs. non-linear) and a fair comparison to our random forest counterpart require kernel PCA (as in Islam et al., 2017; Nahil and Lyhyaoui, 2018). For both alternatives, we set $p = 7$ to (i) ensure comparability with the PC-VAR model estimated via equation-by-equation OLS and to (ii) avoid introducing new dimensionality problems with higher p . We call the additional settings PC(K)-LASSO and PC(K)-SVM and use $K = 7$ as well as $K = 11$ because the latter number of factors yields the best forecasts for PC-LASSO and PC-SVM regression.

B.4 Regularization

Regularization is a process of adding information to solve ill-posed optimization problems or to prevent overfitting (see Bühlmann and van de Geer, 2011). This information can be quite different across different methods. For example, in elastic net regressions (see Section B.4.1), the objective function is equipped with additional penalties, whereas, in a Bayesian framework (see Section B.4.2), additional (inexact) restrictions are imposed.

B.4.1 Elastic net

Introduced by Zou and Hastie (2005), the elastic net penalizes large coefficient values in linear regression models and shrinks them to 0. To this end, it combines the l_1 - and l_2 -penalties of ridge and LASSO methods. For a better understanding of these techniques, two aspects are important. First, in their standard forms, they are all applied to a scaled version of the linear model (3.2.2), where we again use a maximum lag order of $p_{\max} = 21$.¹⁰ The scaled model has the form

$$\tilde{y} = \tilde{X}\beta + \varepsilon, \tag{B.4.1}$$

¹⁰As discussed in Sections B.3 and B.5.2, we also use LASSO in conjunction with estimated factors and a SVM.

where \tilde{X} is the $(N \times J)$ scaled regressor matrix, \tilde{y} is the $(N \times 1)$ scaled dependent variable vector and β is the $(J \times 1)$ standardized coefficient vector. The columns of \tilde{X} and \tilde{y} are scaled to a mean of 0 and a variance of 1. Second, recall that OLS minimizes the sum of squared residuals, such that the OLS estimator of the scaled model can be compactly written as

$$\hat{\beta} = \arg \min_{\beta} \|\tilde{y} - \tilde{X}\beta\|_2^2. \quad (\text{B.4.2})$$

Ridge regression is a regularization method developed by Hoerl and Kennard (1970) and its first EPF application can be found in Uniejewski et al. (2016). In contrast to OLS, ridge regression has a preference for smaller coefficient values and minimizes the sum of squared residuals penalized by a quadratic shrinkage factor. That is, we have

$$\hat{\beta} = \arg \min_{\beta} \|\tilde{y} - \tilde{X}\beta\|_2^2 + \lambda \|\beta\|_2^2, \quad (\text{B.4.3})$$

where $\|\beta\|_2^2 = \sum_{j=1}^J |\beta_j|^2$ is the squared l_2 -norm and λ is a pre-set tuning parameter. Note that, for $\lambda = 0$, we get the standard OLS estimator. Ridge regression will include all predictors in the final model. While the quadratic factor will shrink all coefficient values towards 0, it will not set any of them exactly to 0. This issue is resolved by the LASSO.

The LASSO method has been suggested by Tibshirani (1996) as a regularized estimation technique and also found its way to EPF (see Ziel, 2016). Due to its shrinkage property, the estimator can handle models with many parameters from which only a few are included in the final model. Hence, LASSO regression performs a feature selection within model estimation. The LASSO estimator is given by

$$\hat{\beta} = \arg \min_{\beta} \|\tilde{y} - \tilde{X}\beta\|_2^2 + \lambda \|\beta\|_1, \quad (\text{B.4.4})$$

where $\|\beta\|_1 = \sum_{j=1}^J |\beta_j|$ denotes the l_1 -norm and λ is again a tuning parameter. While the ridge penalty shrinks the coefficients of correlated predictors towards each other, the LASSO tends to pick one of them and discard the others.

Zou and Hastie (2005) proposed the elastic net, a regularization and variable selection method, which often outperforms the LASSO, while exhibiting a similar sparsity of representation. The elastic net uses a mixture of linear and quadratic penalty factors, i.e.,

$$\hat{\beta} = \arg \min_{\beta} \|\tilde{y} - \tilde{X}\beta\|_2^2 + \lambda(\alpha \|\beta\|_1 + (1 - \alpha) \|\beta\|_2^2/2), \quad (\text{B.4.5})$$

where $\alpha \in [0, 1]$ controls the elastic net penalty bridging the gap between ridge ($\alpha = 0$) and LASSO ($\alpha = 1$) regression. The tuning parameter λ controls the overall strength of the penalty.

For elastic net estimation, we adopt the very fast pathwise coordinate descent algorithm of Friedman et al. (2010), which is available in the R package `glmnet`. It efficiently solves the elastic net problem on a grid Λ of λ -values. We define $\Lambda = \{2^j : j \in \mathcal{G}\}$, where \mathcal{G} is an equidistant grid from 4 to -12 with step size -0.2 . We then set the tuning parameter $\lambda \in \Lambda$ for each hour h by 10-fold cross-validation minimizing the mean square error in the calibration set.¹¹ For ridge and LASSO, we proceed analogously. The elastic net mixture parameter α is tuned (simultaneously to λ) on the grid $[0, 0.1, \dots, 0.9, 1]$. Table B.1 presents its optimal value for each hour h . Interestingly, most of the numbers correspond to a pure LASSO approach ($\alpha = 1.0$). Thus, we expect the forecasting performance of elastic net and LASSO to be similar.

h	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
α	1.0	1.0	1.0	1.0	0.8	0.6	1.0	0.4	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.4	0.5	0.9	0.6	0.5	0.2	1.0

For each our h of the day and based on our calibration dataset, this table reports the optimal mixture parameter α for elastic net regression.

TABLE B.1: Optimal elastic net mixture parameters

B.4.2 Bayesian VAR

Rather than using exact restrictions (such as zeroing out lags or deleting variables altogether), a Bayesian VAR model imposes a set of inexact restrictions on model coefficients. It treats the model parameters as random variables and assigns prior probabilities to them. Thus, by using informative priors, an unrestricted model can be shrunk towards a parsimonious model, thereby reducing parameter uncertainty and improving forecast accuracy (see Karlsson, 2013). We rely on the reduced form Bayesian VAR model with Minnesota prior of Litterman (1986), which is a special case of the general Sims and Zha (1998) framework.

We start with the VAR model (B.2.3) in reduced form, where $\varepsilon_d \sim N(0, \Sigma)$, $d = 1, \dots, N$, are (24×1) vectors of independent Gaussian shocks. To describe the prior distributions, we write the model in stacked matrix form similar to (3.2.2). That is, we have

$$P = X\phi + \varepsilon \tag{B.4.6}$$

¹¹Using the mean absolute prediction error instead does not improve out-of-sample forecasts.

with $(N \times 24)$ price matrix $P = (P_{d,h})_{d=1,\dots,N;h=0,\dots,23}$, $(N \times J)$ regressor matrix $X = (x^j)_{j=1,\dots,J}$ and $(J \times 24)$ stacked coefficient matrix $\phi = (\psi_0, \dots, \psi_{23})$, which is formed by stacking the coefficient vectors ψ_h for each hour h . Note that, in this section, we add the index h to ψ for clarity. The error term ε has dimension $(N \times 24)$.

In this model, the likelihood function becomes

$$L(P|\phi, \Sigma) \propto |\Sigma|^{-N/2} \cdot \exp \left(-\frac{1}{2} \text{tr}((P - X\phi)' \Sigma^{-1} (P - X\phi)) \right). \quad (\text{B.4.7})$$

Given the joint prior distribution of the parameters, denoted $p(\phi, \Sigma)$, their joint posterior distribution conditional on the data P can be obtained via the Bayes rule

$$p(\phi, \Sigma|P) \propto L(P|\phi, \Sigma)p(\phi, \Sigma). \quad (\text{B.4.8})$$

In the Sims and Zha (1998) system of priors, we assume $\Sigma = \text{diag}(\sigma_{\varepsilon,0}^2, \dots, \sigma_{\varepsilon,23}^2)$ to be fixed and diagonal. Further, $p(\psi_h) = N(\bar{\psi}_h, \Omega_h)$ is a normal probability density function with yet to define mean $\bar{\psi}_h$ and diagonal covariance matrix Ω_h . In the process of supplying these definitions, we are guided by the Minnesota prior which assumes that a random walk model for each variable is a reasonable ‘center’ of beliefs about variable behavior. This way, the prior mean is 1 for the first own lag of each equation and 0 for the other lags, the intercept and exogenous variable parameters. Formally, we have $\bar{\phi} = \text{vec}(I_{24}, 0)$. Priors on variances are calculated from hyperparameters $\lambda_0, \lambda_1, \lambda_3, \lambda_4, \lambda_5$. Specifically, the variance $\nu_{l,k,h}$ of the parameter at lag l of the price variable $P_{.,k}$ in equation for hour h , the variance ν_c of the intercept and the variance ν_w for exogenous weekday parameters are

$$\nu_{l,k,h} = \left(\frac{\lambda_0 \lambda_1}{\sigma_k l^{\lambda_3}} \right)^2, \quad \nu_c = (\lambda_0 \lambda_4)^2, \quad \nu_w = (\lambda_0 \lambda_5)^2, \quad (\text{B.4.9})$$

where the prior variances differ between dependent variables by a scaling factor σ_k which we set equal to the calibration sample standard deviation of $P_{.,k}$.

We use the R package **MSBVAR** for estimation. In the hyperparameter calibration, we prefer a low, harmonic decay of variances with rising lags of the independent variables, i.e., we follow Sims and Zha (1998) by setting $\lambda_3 = 1$. We ‘forbid’ a shrinkage of the coefficients of exogenous variables via $\lambda_5 = 1,000$. In other words, the estimated dummy coefficients shall be similar to OLS estimates. To allow loose shrinkage models, the overall tightness of the prior is set to $\lambda_0 = 0.6$. The remaining λ_1 and λ_4 , which control the tightness around the AR(1) parameters and the intercept, are chosen by a 10-fold cross-

validation grid search to minimize the mean absolute forecast error in the calibration set.¹² We search in the parameter space subset spanned by the finite sets $\lambda_1 \in \{0.1, 0.2, \dots, 0.9\}$ and $\lambda_4 \in \{0.2, 0.4, \dots, 2.0\}$. For the lag order, we choose $p = 7$ only, such that computation time remains within reasonable limits. In summary, we find the following optimal hyperparameters for our setting: $\lambda_0 = 0.6$, $\lambda_1 = 0.3$, $\lambda_3 = 1$, $\lambda_4 = 1$, $\lambda_5 = 1,000$.¹³

B.4.3 Guided regularized random forest

As a last regularization technique, we consider the regularized random forest (RRF) algorithm which applies a tree regularization framework to a RF. While it is very similar to a standard RF, the main difference is that a regularized impurity reduction

$$\Delta_{\iota_R}(x^j, v) = \begin{cases} \lambda \cdot \Delta_{\iota}(x^j, v), & j \notin F \\ \Delta_{\iota}(x^j, v), & j \in F \end{cases} \quad (\text{B.4.10})$$

is used, where F is the set of indices corresponding to features used for splitting in previous nodes and is an empty set at the root node in the first tree. $\lambda \in (0, 1]$ is a penalty coefficient. When $j \notin F$, it penalizes the j th feature for splitting the node v . More intuitively, the smaller λ , the more predictive power a new feature needs to be selected for splitting a node. A RRF with $\lambda = 1$ has the minimum regularization and refers to a standard RF.

An enhanced RRF, referred to as the guided RRF (GRRF), has been proposed by Deng and Runger (2013). It uses the importance scores of a preliminary standard RF to guide the feature selection process of the RRF. A normalized importance score is defined as $\tilde{I}(x^j) = I(x^j) / \max_i I(x^i)$, where $I(x^j)$ is the importance score delivered by the standard RF. Instead of assigning the same penalty coefficient to all features, a GRRF assigns an individual penalty coefficient

$$\lambda_j = (1 - \gamma) + \gamma \tilde{I}(x^j) \quad (\text{B.4.11})$$

to each feature. Here, $\gamma \in [0, 1]$ controls the degree of regularization. We tested the values $\gamma \in \{0, 0.25, 0.5, 0.75, 1\}$ and found that, for $\gamma = 0.75$, the model performs best. Furthermore, following Section 3.2.3, we use $N_{tree} = 500$ and $m = \lfloor J/3 \rfloor$ as RF parameters.

¹²In empirical Bayes fashion, we could alternatively calibrate by maximizing the marginal likelihood with respect to the hyperparameters. However, we find that this is rather unreliable for the determination of the best forecasting model.

¹³Because electricity prices exhibit a strong weekly persistence, we also used a higher lag decay term $\lambda_3 = 2$ for the variance $\nu_{7,k,h}$ of parameters at lag $l = 7$. However, this did not improve out-of-sample performance.

B.5 Feature selection

We now summarize the specifications of our feature selection approaches. Besides introducing a popular stepwise regression procedure, we discuss a LASSO-extended non-linear model and details on our proposal, i.e., the combination of random forest selection with a SVM.

B.5.1 Stepwise regression

Stepwise regression is a systematic method for adding and removing terms from a regression model based on some prespecified criterion. In our study, it has a benchmark character because it is one of the simplest variable selection procedures. Following the EPF application of Uniejewski et al. (2016), we perform a linear model selection by minimizing the exact AIC (henceforth abbreviated with Step-Linear) because this measure directly penalizes a model for unnecessary complexity. With respect to model (3.2.2), the stepwise procedure starts with the intercept as a ‘lower bound model’ and then iteratively adds (removes) input variables x^j to (from) the model. At each step, the algorithm searches for the greatest improvement of the model fit (measured by the AIC), and terminates if no further improvement can be made. Although leading to longer computation times, we have a look at lagged variables up to order $p = 21$ to ensure a fair comparison.

B.5.2 LASSO selection

LASSO regression encourages shrinking coefficients to 0, i.e., dropping the corresponding variates from a model. Therefore, it is a powerful special case of regularization which allows feature selection. While, in Section B.4.1, a linear model is shrunk via the LASSO and used for prediction afterward, the approach of this section is different because we wish to define a close counterpart to our own prediction approach based on importance scores (see Section B.5.3). Specifically, we combine LASSO feature selection with SVM regression (similar to Becker et al., 2009). In this LASSO-SVM model, we set the LASSO parameter $\lambda \in \Lambda$ for each hour h similar to Section B.4.1 and select all features with non-zero coefficients as inputs for a non-linear SVM with ν -regression specified as in Section B.2.6.

B.5.3 Random forest scores

In the presence of a significant number of input variables, classic feature selection methods, which evaluate many feature subsets by building forecasting models for each subset, are error-prone (see Flack and Chang, 1987). In what follows, we demonstrate that we can find a good feature subset by solely relying on the importance scores provided in Section 3.2. That is, we simply select the K most relevant features according to their

importance score ranking. After that, we feed the selected variables into a linear and a non-linear regression model. Recall that the importance score $I(x^j)$ derived from the random forest algorithm is the mean decrease in prediction accuracy caused by removing the link between feature x^j and the target. The score is therefore directly related to the explanatory power of x^j and measures how much we would regret not including it in the feature subset. In addition, even though they are based on linear calculations, an important feature of random forests is that they can detect both linear and non-linear relationships between input and target variables (see Auret and Aldrich, 2012).¹⁴ Thus, if relationships were actually non-linear, it would not be surprising to observe that a random forest selection used in a linear prediction model performs poorly whereas a strong non-linear regression tool makes more efficient use of a random forest feature choice.

As for the number K of features to select, we consider two settings. First, we set $K = 50$ for each hour h to allow a comparison to our expert models which consider $24 + 24 + 4 = 52$ features, i.e., all lagged variables for the orders $l = 1, 7$ plus our weekday dummies. This is because we are interested in answering the question of whether the exact importance score rankings have superior predictive abilities than the rough expert choice. Second, we optimize K . In general, K should be chosen such that we can expect a good performance of the favored regression model. Therefore, we determine the K between 10 and 100 which minimizes the mean absolute prediction error in the validation set after estimating the model in the training set.¹⁵

We combine random forest feature selection with a standard linear model and a non-linear SVM with ν -regression (see Section B.2.6). The models using $K = 50$ features will be denoted RF(50)-Linear and RF(50)-SVM. For optimized K , we simply write RF-Linear and RF-SVM.

B.6 Forecast combinations

Besides forecasting via individual models, we look at simple forecast combinations because (i) many authors emphasize that they can substantially improve forecast accuracy (see Stock and Watson, 2006; Timmermann, 2006) and (ii) rudimental combination schemes often work reasonably well in comparison to more complex ones (see Genre et al., 2013). By analyzing combinations containing machine learning techniques, we can extend studies

¹⁴We might alternatively think of using a non-linear model within the random forest calculations.

¹⁵Minimizing the mean square error instead delivers worse results.

combining electricity price forecasts of less sophisticated models (see Nowotarski et al., 2014; Weron, 2014).

We consider two popular weighting schemes for forecast combination. A weighting scheme can be formally introduced as

$$\hat{P}_{d,h} = \sum_{m=1}^M w_{d,h}^{(m)} \hat{P}_{d,h}^{(m)}, \quad (\text{B.6.1})$$

where M is the number of considered forecasting methods and $\hat{P}_{d,h}^{(m)}$ is the forecast of the hourly price obtained from method m . For each hour h , $w_{d,h}^{(m)}$ assigns a weight to forecast method m . This weight may vary over time d . For each d - h -tuple, we have $w_{d,h}^{(m)} \geq 0 \forall m = 1, \dots, M$ and $\sum_{m=1}^M w_{d,h}^{(m)} = 1$.

The first combination scheme is the simple average which uses equal weights $w_{d,h}^{(m)} = 1/M$ for each d , h and m . While this approach can provide some insurance against crucial forecasting failures and thus makes combined forecasts potentially less risky than selecting an individual forecasting model (see Hibon and Evgeniou, 2005), it does not consider the fact that the performance of different models can vary with market conditions (see Weron and Misiorek, 2008). In contrast, the second weighting approach uses inverse mean square prediction error (MSPE) weights based on the recent forecasting performance of each model (see Baumeister and Kilian, 2015).¹⁶ We have

$$w_{d,h}^{(m)} = \frac{1/\text{MSPE}_{d,h}^{(m)}}{\sum_{m=1}^M 1/\text{MSPE}_{d,h}^{(m)}}, \quad (\text{B.6.2})$$

where $\text{MSPE}_{d,h}^{(m)} = \frac{1}{Q} \sum_{j=1}^Q (P_{d-j,h} - \hat{P}_{d-j,h}^{(m)})^2$. Thus, the smaller the MSPE over the most recent Q periods, the larger the weight a model receives in the combined forecast. We test three different window lengths to determine combination weights: a very short period of one week ($Q = 7$ days), a mid-term period of one month ($Q = 30$ days) and a longer period of one year ($Q = 365$ days). These forecast combination settings are labeled Inv-MSPE(Q). Combinations generally contain the individual forecasts of all models discussed above.

¹⁶Other popular alternatives are exponentially weighted averages or use a ‘gradient trick’ to take into account dependencies between models (see Devaine et al., 2013).

Appendix C

Appendix of Chapter 4

C.1 Proofs

Proof of Proposition 4. (based on Kohrs et al., 2019, Proposition 3.9)

We show by backward induction that $p \mapsto \mathfrak{V}_t(Q_t, \mathfrak{X}_t(p))$ is convex for every $t \in \{0, \dots, T-1\}$ and every feasible quantity Q_t . Note that, on day t , subject to the local constraint $q_t \in [0, 1]$ and the global constraint $Q_T \in [Q_{\min}, Q_{\max}]$, a feasible policy requires $Q_t \in [\underline{Q}_t, \overline{Q}_t] := [0 \vee (Q_{\min} - (T - t)), Q_{\max} \wedge t]$ and $q_t \in [\underline{q}_t, \overline{q}_t] := [0 \vee (Q_{\min} - Q_t - (T - t - 1)), 1 \wedge (Q_{\max} - Q_t)]$. Further note that, by forward induction, it is easy to see that $p \mapsto \mathfrak{X}_t(p)$ is an affine mapping and hence convex (but not strictly convex).

For $t = T - 1$ and $Q_{T-1} \in [\underline{Q}_{T-1}, \overline{Q}_{T-1}]$, we have

$$\begin{aligned} \mathfrak{V}_{T-1}(Q_{T-1}, \mathfrak{X}_{T-1}(p)) &= \max_{q_{T-1} \in [0, 1]} \mathfrak{X}_{T-1}(p) \cdot q_{T-1} + \mathbb{E}[\mathfrak{V}_T(Q_T, \mathfrak{X}_T(p)) | \mathcal{F}_{T-1}] \\ &= \max_{q_{T-1} \in [\underline{q}_{T-1}, \overline{q}_{T-1}]} \mathfrak{X}_{T-1}(p) \cdot q_{T-1}, \end{aligned}$$

which is convex in p as a pointwise maximum over a set of convex functions.¹

We now pass from t to $t - 1$. Let $Q_{t-1} \in [\underline{Q}_{t-1}, \overline{Q}_{t-1}]$. By induction assumption, $p \mapsto \mathfrak{V}_t(Q_{t-1} + q_{t-1}, \mathfrak{X}_t(p))$ is convex for every $q_{t-1} \in [\underline{q}_{t-1}, \overline{q}_{t-1}]$. For fixed w , the function $f_{t-1,w}$ is affine in its arguments by assumption, such that the mapping $p \mapsto \mathfrak{V}_t(Q_{t-1} +$

¹Operations preserving convexity are discussed in, for example, Boyd and Vandenberghe (2009).

$q_{t-1}, f_{t-1,w}(\mathfrak{X}_{t-1}(p), p))$ is convex. Denote by φ_{t-1} the density function of W_{t-1} . Then, it follows that

$$p \mapsto \mathbb{E}[\mathfrak{V}_t(Q_{t-1} + q_{t-1}, \mathfrak{X}_t(p)) | \mathcal{F}_{t-1}] = \int_x \mathfrak{V}_t(Q_{t-1} + q_{t-1}, f_{t-1,w}(\mathfrak{X}_{t-1}(p), p)) \varphi_{t-1}(w) dw$$

is convex. Because $p \mapsto \mathfrak{X}_{t-1}(p)$ is convex and taking the pointwise maximum or adding functions of this kind are operations preserving convexity,

$$\mathfrak{V}_{t-1}(Q_{t-1}, \mathfrak{X}_{t-1}(p)) = \max_{q_{t-1} \in [\underline{q}_{t-1}, \bar{q}_{t-1}]} \mathfrak{X}_{t-1}(p) \cdot q_{t-1} + \mathbb{E}[\mathfrak{V}_t(Q_{t-1} + q_{t-1}, \mathfrak{X}_t(p)) | \mathcal{F}_{t-1}]$$

is convex, which concludes the induction step.

As v is convex, the theorem of Alexandrov (1939) ensures the twice differentiability almost everywhere. \square

Proof of Proposition 6. We start with the forward curve at time s in terms of its initially observed state at time 0:

$$F(s, t) = F(0, t) + \int_0^s \sigma(u, t) dW_u. \quad (\text{C.1.1})$$

Hence, the spot price $S_t = F(t, t)$ is given by

$$S_t = F(0, t) + \int_0^t \sigma(u, t) dW_u. \quad (\text{C.1.2})$$

Differentiating this equation yields the stochastic differential equation for the spot price process:

$$dS_t = \frac{\partial F(0, t)}{\partial t} dt + \int_0^t \frac{\partial \sigma(s, t)}{\partial t} dW_s dt + \sigma(t, t) dW_t.$$

Considering our specific volatility function (4.3.2), we have

$$\frac{\partial \sigma(s, t)}{\partial t} = -\vartheta \sigma e^{-\vartheta(t-s)} = -\vartheta \sigma(s, t)$$

and hence

$$dS_t = \frac{\partial F(0, t)}{\partial t} dt - \vartheta \left(\int_0^t \sigma(s, t) dW_s \right) dt + \sigma dW_t.$$

Using the rearranged (C.1.2), i.e.,

$$\int_0^t \sigma(s, t) dW_s = S_t - F(0, t), \quad (\text{C.1.3})$$

we finally receive (4.3.9):

$$dS_t = \frac{\partial F(0, t)}{\partial t} dt + \vartheta(F(0, t) - S_t) dt + \sigma dW_t.$$

Furthermore, (4.3.10) can be obtained by using (4.3.2) and (C.1.3) in (C.1.1):

$$\begin{aligned} F(s, t) &= F(0, t) + \int_0^s \sigma e^{-\vartheta(t-u)} dW_u \\ &= F(0, t) + e^{-\vartheta(t-s)} \int_0^s \sigma e^{-\vartheta(s-u)} dW_u \\ &= F(0, t) + e^{-\vartheta(t-s)} (S_s - F(0, s)). \end{aligned}$$

□

Kurzlebenslauf - Hendrik Kohrs

Persönliche Daten

Geburtstag: 15.07.1991
Geburtsort: Bremervörde
Nationalität: deutsch

Forschungsgebiete

Komplexe Energiederivate, Portfoliotheorie, Strompreisprognose

Ausbildungsweg

06/2017 - 04/2022: Promotion im Bereich Finanzierung & Investition, Universität Leipzig, Wirtschaftswissenschaftliche Fakultät (Thema: "Essays on Energy Portfolio Management"; Gutachter: Prof. Dr. Frank Schuhmacher, Prof. Dr. Benjamin R. Auer)
10/2011 - 04/2017: Studium der Wirtschaftsmathematik, Universität Leipzig, Fakultät für Mathematik und Informatik (Diplomarbeit: "Pricing and Risk of Swing Contracts in Natural Gas Markets"; Abschluss: Diplom mit Gesamtnote "mit Auszeichnung" (1,0))
07/2004 - 06/2011: Erwerb der Allgemeinen Hochschulreife, Gymnasium Warstade Hemmoor (Abschluss: Abitur mit Gesamtnote "sehr gut" (1,3))

Berufserfahrung

04/2018 - heute: Hauptreferent Quantitative Analyse / Modellierung, VNG Handel & Vertrieb GmbH, Leipzig
06/2017 - 08/2018: Wissenschaftlicher Mitarbeiter, Universität Leipzig, Lehrstuhl für Finanzierung und Investition
09/2015 - 11/2015: Praktikant Konzernrisikomanagement, VNG - Verbundnetz Gas AG, Leipzig
10/2014 - 03/2015: Tutor für "Analysis für Informatiker", Universität Leipzig, Fakultät für Mathematik und Informatik
04/2014 - 09/2016: Tutor für "Mikroökonomik", Universität Leipzig, Wirtschaftswissenschaftliche Fakultät

Leipzig, den 27. April 2022

Hendrik Kohrs

Selbständigkeitserklärung

Hiermit erkläre ich, die von mir vorgelegte Dissertation selbständig und ohne unzulässige Hilfe Dritter angefertigt zu haben. Der Anteil von Koautoren an einzelnen Aufsätzen der kumulierten Schrift wurde in einem separaten Dokument im Antrag zur Eröffnung des Promotionsverfahrens dargelegt. Ich habe keine anderen als die angeführten Quellen und Hilfsmittel benutzt und sämtliche Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen wurden, und alle Angaben, die auf mündlichen Auskünften beruhen, als solche kenntlich gemacht. Ebenfalls sind alle von anderen Personen bereitgestellten Materialien oder erbrachten Dienstleistungen als solche gekennzeichnet.

Weiterhin versichere ich, dass die vorgelegte Dissertation von mir weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zwecke einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde.

Leipzig, den 27. April 2022

Hendrik Kohrs