Statistical modeling in electricity and related markets

Anders Løland

Thesis presented for the degree of Doctor Philosophiae

July 2013

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 1380

ISSN 1501-7710

Cover: Inger Sandved Anfinsen.
Printed in Norway: AIT Oslo AS.

Produced in co-operation with Akademika Publishing.
The thesis is produced by Akademika Publishing merely in connection with the thesis defence. Kindly direct all inquiries regarding the thesis to the copyright holder or the unit which grants the doctorate.

Preface and acknowledgements

> *"Before turning to those moral and mental aspects of the matter which present the greatest difficulties, let the inquirer begin by mastering more elementary problems."*
>
> — Sir Arthur Conan Doyle, *A Study in Scarlet (1887)*

Most of the work presented in this thesis has been carried out in the research centre Statistics for Innovation – $(sfi)^2$ – hosted by the Norwegian Computing Center (NR) and enthusiastically led by Arnoldo Frigessi. I have been fortunate to work with many diverse and interesting problems both in $(sfi)^2$ and otherwise at NR.

First, I would like to thank all my coauthors: Xeni K. Dimakos, Egil Ferkingstad, Arnoldo Frigessi, Nils Lid Hjort, Ingrid Hobæk Haff, Lars Holden, Ragnar Bang Huseby, Ola Lindqvist, Antonio Pievatolo, Fabrizio Ruggeri and Mathilde Wilhelmsen. It has been fun to work with each of you. Being surrounded by clever and, even more important, enjoyable colleagues at NR, has been equally inspiring.

A very encouraging and convincing Fred Espen Benth made me assemble this thesis. Similarly, André Teigland, the head of my department, has been very supportive to this enterprise all the way.

Family and friends; if you happen to try to read this thesis and do not understand much, do not despair! Life is more than work, and you fill my life with joy and meaning.

Oslo, July 2013

Anders Løland

Contents

Chapter 1

Introduction

> *"'Winwood Reade is good upon the subject", said Holmes. "He remarks that, while the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician."'*
>
> — Sir Arthur Conan Doyle, *The Sign of the Four (1890)*

From the early 1990s, markets for electricity and related energy products have been liberalised all over the world (Benth et al., 2008). For electricity, the Nordic market and the England & Wales market were the first. The Nordic Nord Pool (Spot) power exchange was formally established in 1996, with Norway and Sweden as the only members. Finland followed suit in 1998, and Denmark joined the exchange in 2000.

As with the purely financial markets, liberalised energy markets generate an abundance of valuable data, like spot and forward prices, electricity flows and capacities. Hence, the statistician may enter the room. He will start asking questions like *"Is there a risk premium in the market?"*, *"Are there causal links between energy markets?"*, *"What explains local prices?"*, *"Can we forecast transmission congestion?"* and *"What should be the price of complicated gas contracts?"*. In the process, we will notice that many of his questions concern more than one variable (oil and gas prices, prices in more than one electricity market). The basic ingredient to study more than one variable is the concept of correlation, through a correlation (or covariance) matrix. For the modelling not to break down completely, these matrices must satisfy certain criteria; they must be proper. The statistician therefore asks again: *"How should an improper correlation matrix be adjusted to be valid?"*.

In the following, an overview of problems and methods in statistical modelling of electricity markets, statistical cures for invalid correlation matrixes and statistical modelling in gas markets is given.

1.1 Statistical modelling of electricity markets

"It is a capital mistake to theorize before you have all the evidence. It biases the judgment."

— Sir Arthur Conan Doyle, *A Study in Scarlet (1887)*

Figure 1.1: Nord Pool Spot price areas as of January 2013. The green lines denote possible flows between the price areas. Source: `http://www.nordpoolspot.com/`

The Nordic[1] electricity spot (or day ahead) power market, Elspot, is divided into several price areas (Benth et al., 2008; Kristiansen, 2004; Weron, 2006), with the system price being a common reference price (Figure 1.1). The different price areas result from capacity constraints. In theory, if an overall market balance can be achieved without a need to utilise all available capacity between neighbouring areas, the prices are equal in all areas. This theoretical price is

[1]The Nordic electricity spot market is run by Nord Pool Spot.

called the system price. Transmission congestion within the Nord Pool area is not uncommon (Marckhoff and Wimschulte, 2009). During nighttime, the price is often equal in neighbouring areas, while price area differences are seen more often in periods with a high load, like during the winter and daytime.

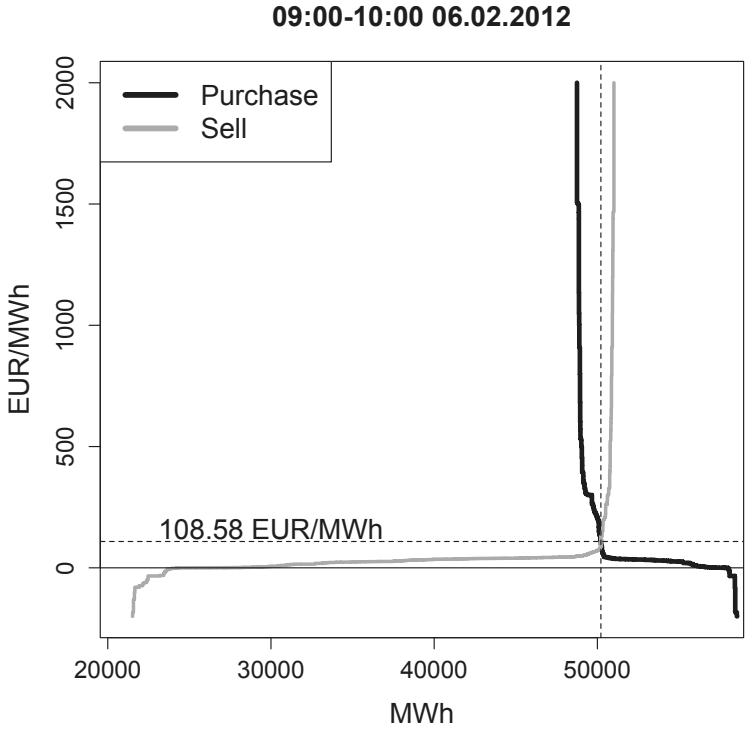


Figure 1.2: A Nord Pool Spot system price curve. The intersection between the purchase (consumers) and sell (producers) curves determines the spot price for this hour.

The Elspot prices are settled once every day for each of the 24 hours of the coming day, based on all bids from market players, buying or selling electricity with a certain volume (see Figure 1.2 for an example). The intersection between the purchase (consumers) and sell (producers) curves is the spot price.

Storing large quantities of electricity is a major challenge. If electricity is produced with gas or coal, the commodity may be stored, and one can sell electricity only when prices are high, and bidding in the spot market is relatively simple.

The Nordic electricity market is dominated by highly flexible hydro power (54% in 2007 according to Fridolfsson and Tangerås (2009), 95% in Norway in 2010 according to Statistics Norway). The German EEX market, being the largest market in Europe, is on the other hand dominated by coal (47%) and nuclear power (23%) (Brunekreeft and Twelemann, 2005). Gas (17%), hydro and an rapidly increasing solar and wind power production complement the picture. The EEX market is generally assumed to be less mature than the Nordic market (Weron, 2006;

Weigt and von Hirschhausen, 2008; Müsgens, 2006; Fridolfsson and Tangerås, 2009).

Even though electricity is a non-storable commodity, large water reservoirs makes hydro power partly storable. Run-of-the-river hydroelectric stations have small or no reservoir capacity, while also very large reservoirs exist that can store two or three years of inflow. Production and bidding is often planned using a mixture of medium- and long term optimisation models (Fosso et al., 1999) and very short term forecast models (Weron, 2006).

1.1.1 Regression, model averaging and forecasting

> *"We've long felt that the only value of stock forecasters is to make fortune tellers look good."*
>
> — Warren E. Buffet, *Chairman's Letter, Berkshire Hathaway Inc. (1992)*

Long term generation scheduling models (Pereira, 1989; Pereira and Pinto, 1991; Wolfgang et al., 2005) solve a stochastic optimisation problem: Given a mathematical description of the market and a demand function, the aim is to maximise the socio-economic surplus for consumers and producers. The models can be very detailed. For example, the so-called Samkjøringsmodellen (Wolfgang et al., 2005) contains more than 500 water reservoirs and 250 hydro power plants for Norway. Models like Samkjøringsmodellen implicitly assume that their description of the market is perfect, and that assumption can never be correct.

Alternatively, we can turn to the world of mathematical finance. The classical model for commodity markets is the Schwartz (1997) model, which can be written as

$$S(t) = S(0)\exp(X(t)),$$

where

$$dX(t) = \alpha(\mu - X(t))dt + \sigma dB(t),$$

and $B(t)$ is a Brownian motion. Such time series models from mathematical finance (Benth et al., 2008) will also to some extent provide a superficial description of electricity markets, since their motivation often is mathematical tractability. To avoid to "theorize before you have all the evidence", statistical data analysis therefore has an important role to play here.

Generalised additive models (Hastie and Tibshirani, 1990) has become a very useful tool in exploratory data analysis, but has not been used much in modelling of energy market data. The additive regression model model replaces the linear regression model

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j X_j + \varepsilon \tag{1.1}$$

with

$$Y = \beta_0 + \sum_{j=1}^{p} s_j(X_j) + \varepsilon, \tag{1.2}$$

where the errors ε are independent of the X_js, with $\mathsf{E}[\varepsilon] = 0$ and $\mathsf{VAR}[\varepsilon] = \sigma^2$. The s_js are typically smooth functions formed by splines, one function for each predictor. This enables estimation of non-linear s_js. The generalisation through the choice of link function is similar as for generalised linear models (McCullagh and Nelder, 1989).

Most forecasting methods are based directly on regression models, such as generalised additive models, or indirectly through time series models. A large number of candidate models exist, see for example Hamilton (1994). Finding the best model can be a painstaking task, even though recent advances in model selection can guide us (Claeskens and Hjort, 2008). Luckily, the lazy statistician can seek comfort in methods that just combine forecasts without selecting one (Clemen and Winkler, 1986; Palm and Zellner, 1992; Timmermann, 2006). Even if each forecast model is mediocre, the combination can be very good.

Let X_t be the observed value of the time series X at time t, and $\hat{X}^{(1)}_{t+\delta}, \hat{X}^{(2)}_{t+\delta}, \ldots, \hat{X}^{(n)}_{t+\delta}$ be predictions from n different models of X at time $t+\delta$. The n predictions can be combined into one prediction, $\hat{X}_{t+\delta}$, by taking a weighted average

$$\hat{X}_{t+\delta} = \sum_{i=1}^{n} w_i \hat{X}^{(i)}_{t+\delta}.$$

The simplest combination method is to take the average of the n predictions,

$$w_i = \frac{1}{n}, \quad i = 1, \ldots, n.$$

This method works surprisingly well and is quite robust (Timmermann, 2006). It is surprising, since building a 'super' model that incorporates all of the underlying, simpler forecast models often is expected to be the superior approach. Bates and Granger (1969) argue that this works because discarded forecast models almost always contains some independent information:

1. One forecast model can be based on information or variables not present in another forecast model.

2. The different forecast models may be based on different assumptions about the form of the relationship between the variables.

A more sophisticated combining method is to let the weights be given as

$$w = (w_1, \ldots, w_n) = \frac{u^T \hat{\Sigma}^{-1}}{u^T \hat{\Sigma}^{-1} u},$$

where $u = (1, 1, \ldots, 1)^T$ and $\hat{\Sigma}$ is the estimated covariance matrix of the forecast errors $e = (\hat{X}^{(1)} - X, \ldots, \hat{X}^{(n)} - X)$. This puts lower weights on models with highly variable prediction errors. The covariance matrix at time t is estimated based on the d previous forecast errors. It is common to assume zero correlation between the forecast errors (Timmermann, 2006), since the number of parameters will be large compared to the covariance estimation period d. Also, studies have shown that the prediction results rarely get better when correlations are estimated. Clemen and Winkler (1986) point out that the estimate of Σ "can be quite unstable unless large data sets are available for estimation", particularly when the pairwise correlations are high, as may often be the case with economic forecasts.

1.1.2 Causality?

"If there is a will, there is a way."

— Yasser Arafat, former Chairman of the Palestine Liberation Organization

When faced with multiple time series (such as the ones in Figure 1.5, which we will study later on), we are often interested in the dependence between the series. In (1.1) or (1.2), we might assume or hope that the covariates cause the response, but in general we can not be sure.

The Holy Grail of statistics has been causality. As Aalen and Frigessi (2007) puts it:

> *"For most of the 20th century the dominant attitude in statistics was that, as a statistician, one should shy away from causality. It was firmly stated by the founding fathers, especially Pearson but also to a large extent by Fisher, that statistics is only about association."*

Other fields, such as econometrics (Granger causality, Hamilton (1994)) and machine learning (causal networks, Shimizu et al. (2006)) have not been as shy. We will here take a look at one machine learning approach.

We will assume that the observed variables can be arranged in a causal order, meaning that no variable can cause a preceding variable. This means that it can be represented by a directed acyclic graph (DAG) (Spirtes et al., 2000). Each variable is here a linear function of the values of the preceding variables, plus a noise term and an optional constant term,

$$X_i = \sum_{i,j,\ i>j} \beta_i + \beta_{ij} X_j + \varepsilon_i. \tag{1.3}$$

In Figure 1.3,

$$\begin{aligned} X_3 &= \beta_3 + \beta_{31} X_1 + \beta_{32} X_2 + \varepsilon_3 \\ &= \beta_3 + (\beta_{31} + \beta_{21}) X_1 + \beta_{32}\varepsilon_2 + \varepsilon_3. \end{aligned}$$

Standard causal network analysis is based on the assumption that the variables, the εs in (1.3), are jointly normally distributed. With these assumptions we can estimate DAGs (Chickering,

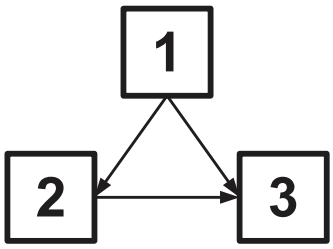


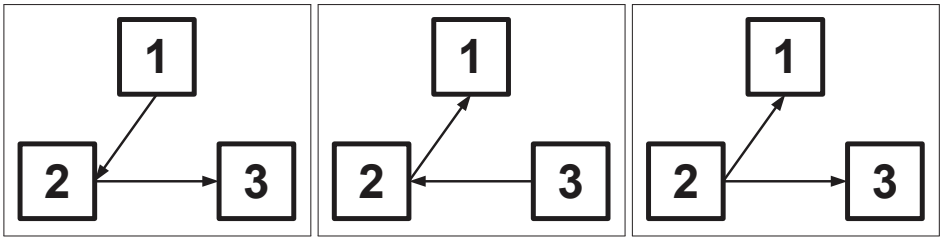
Figure 1.3: An example of a directed acyclical graph, representing a direct and indirect effect of X_1 on X_3.

2003), for example with

$$\boldsymbol{X} = (X_1, X_2, X_3) \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}), \tag{1.4}$$

where we assume that the covariance matrix $\boldsymbol{\Sigma}$ is proper. (We will come back to the properness in Section 1.2.)

Standard methods share a fundamental problem: A joint distribution may correspond to several DAGs, since they entail the same conditional independence relations among the observed variables. One therefore only obtains an equivalence class of DAGs that are indistinguishable from data. While some directions of causal influences (edges in the DAG) may be the same for all DAGs in the equivalence class, usually many or most directions are left undetermined.



(a) (b) (c)

Figure 1.4: Three equivalent directed acyclical graphs resulting from a multivariate normal distribution with three variables and $X_1 \perp\!\!\!\perp X_3 \,|\, X_2$.

If $X_1 \perp\!\!\!\perp X_3 \,|\, X_2$ (X_1 and X_3 are independent given X_2) in (1.4) we can only find equivalence classes of causal networks or DAGs (Figure 1.4). Let us assume that all the variables (or all but one variable) of interest are non-Gaussian, which means that the εs in (1.3) are non-Gaussian. Then we can distinguish between the three DAGs in Figure 1.4 using higher-order moments.

Equation (1.3) is then known as the LiNGAM (Shimizu et al., 2006). Ignoring the constant term and writing (1.3) in matrix form gives

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{X} + \varepsilon, \tag{1.5}$$

where $\boldsymbol{X} = (y_1, \ldots, y_m)'$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_m)$ and $\boldsymbol{B}$ is the (permutable to lower triangular) matrix of coefficients β_{ij}. The independence of the elements of ε implies that there are "no unobserved confounders" in the sense of Pearl (2000), so a causal interpretation is valid (Shimizu et al. (2006), Section 2). Letting $\boldsymbol{A} = (\boldsymbol{I} - \boldsymbol{B})^{-1}$, we can rewrite (1.5) as

$$\boldsymbol{X} = \boldsymbol{A}\varepsilon. \tag{1.6}$$

Since the variables in ε are independent and non-Gaussian, (1.6) defines the Independent Component Analysis (ICA) model (Comon, 1994; Hyvärinen and Oja, 2000). For ICA, the goal is to estimate both the so-called mixing matrix $\boldsymbol{A}$ and the independent components ε. We therefore aim to find $\boldsymbol{A}$ and ε such that the entries of ε are as statistically independent as possible.

Non-Gaussianity can be measured by entropy. The entropy of a random vector $\boldsymbol{X}$ with density f is defined as $H(\boldsymbol{X}) = -\int f(\boldsymbol{X}) \log f(\boldsymbol{X}) \mathrm{d}\boldsymbol{X}$. Gaussian variables have the highest possible

entropy among random variables with a given variance. Hence, we can measure non-Gaussianity based on neg-entropy J. Neg-entropy is defined by

$$J(\boldsymbol{X}) = H(\boldsymbol{X}_g) - H(\boldsymbol{X}),$$

where $\boldsymbol{X}_g$ is a Gaussian random vector having the same covariance matrix as $\boldsymbol{X}$. Clearly, $J(\boldsymbol{X})$ is zero for Gaussian $\boldsymbol{X}$ and positive for non-Gaussian $\boldsymbol{X}$. The iterative fixed-point algorithm fastICA (Hyvärinen, 1999) estimates $\boldsymbol{A}$ based on approximations to neg-entropy.

Both $\boldsymbol{A}$ and ε can only be estimated up to a scaling constant and a permutation. However, both the scaling and the permutation can be found in the application of ICA to LiNGAM, as shown by Shimizu et al. (2006). After estimating $\boldsymbol{A}$, the coefficient matrix $\boldsymbol{B}$ is given,

$$\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{A}^{-1}.$$

1.2 Statistical cures for invalid correlation matrixes

"Strictly speaking, there are no "measurements" in the world, only correlations."

— Nick Herbert, *Quantum Reality (1987)*

Let's take a step back an look at the simplest way to describe the dependence of a set of variables: their correlation matrix. This captures the linear component of pairwise dependence.

A correlation matrix is required to be a symmetric and positive semidefinite matrix with unit elements on its diagonal. All the eigenvalues of a positive semidefinite matrix are non-negative. A positive definite matrix has only positive eigenvalues, which is equivalent to the existence of a Cholesky factorisation of the matrix. This also means that the inverse of the matrix exists.

It is non-intuitive what these requirements mean for the dependence between the pairwise correlations in a correlation matrix, but in three dimensions it is possible to envision. Let

$$\boldsymbol{R} = \begin{pmatrix} 1 & \rho_{1,2} & \rho_{1,3} \\ \rho_{2,1} & 1 & \rho_{2,3} \\ \rho_{3,1} & \rho_{3,2} & 1 \end{pmatrix} = \begin{pmatrix} 1 & \rho_{1,2} & \rho_{1,3} \\ \rho_{1,2} & 1 & \rho_{2,3} \\ \rho_{1,3} & \rho_{2,3} & 1 \end{pmatrix}.$$

With a $k \times k$ matrix, the correlation parameters live on a proper subset of the hypercube $(-1, 1)^{k(k-1)/2}$. Rousseeuw and Molenberghs (1994) describe and illustrate this subset when $k = 3$ and to some extent when $k = 4$, and show the simple relation

$$\rho_{1,2}^2 + \rho_{2,3}^2 + \rho_{1,3}^2 - 2\rho_{1,2}\rho_{2,3}\rho_{1,3} = 1$$

on the boundary of $\boldsymbol{R}$, that is with one eigenvalue of zero.

We refer to such matrices as proper correlation matrices. Estimation of a correlation matrix is based on observed time series of n assets, $(Z_{1,t}, Z_{2,t}, ..., Z_{n,t})$, for $t = 1, 2, ..., T$, under appropriate stationarity assumptions.

There are at least two common situations were the construction of a proper correlation matrix is difficult. This is the case when not all assets are observed in the same time points (Higham, 2002; Tchernitser and Rubisov, 2009). A second situation which leads to improper correlation matrices is when some assets are not observed at all, but an expert opinion is obtained on its correlation with the other assets. Relevant credit or operational loss data might for example be hard to obtain, and expert opinions have to be called upon (Medova, 2000; Dimakos and Aas, 2004). In both of these cases, the correlation matrix stems from the estimation or specification of pairwise correlations.

1.2.1 An example from the Nordic electricity market

There is a parallel financial market, NASDAQ OMX Commodities (formerly known as Nord Pool), to the Nordic electricity spot market where players can hedge their positions through futures (days, weeks) and forwards (months, quarters and years) against the system price. Nobody are, however, exposed to the system spot price, but rather to the area spot price. Therefore, to hedge their area price risk, the participants can in addition buy CfDs (Contracts for Difference), to hedge the remaining difference between the system and price area risk. CfD prices can

therefore be negative. The CfDs are typically available for the next two months, the next three quarters and the next three years.

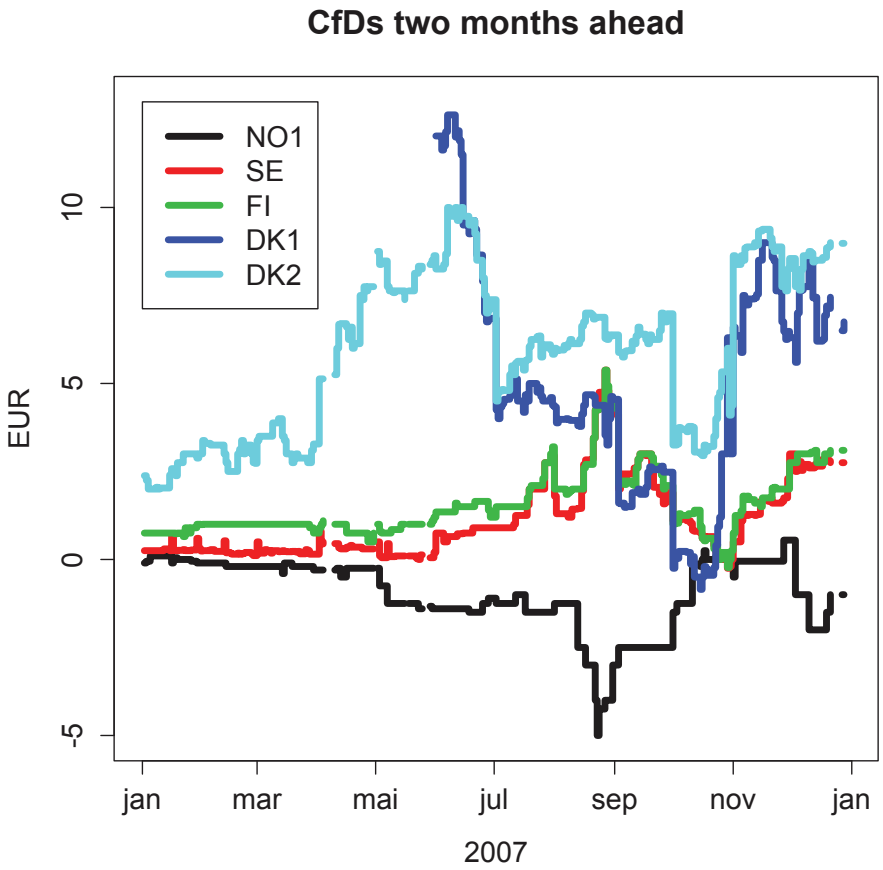


Figure 1.5: Nordic CfD (Contracts for Difference) prices two months ahead for five price areas in 2007. The five price areas are NO1 (Southern Norway), SE (Sweden), FI (Finland), DK1 (Western Denmark), DK2, (Eastern Denmark). Trading takes place only during the work week. The gaps in the data represent public holidays, where trading at NASDAQ OMX Commodities was closed.

The standard approach in finance is to study the (logarithmic) price changes amongst others to approach symmetry or even normality. In this case, and since the CfD prices can be negative, it makes perfect sense to study the correlation between the price levels. Norwegian and Swedish/Finnish CfD prices were clearly negatively correlated in 2007 (Figure 1.5, with pairwise correlations in Table 1.1). Swedish and Finnish CfD prices were highly correlated, while Western and Eastern Denmark were mostly correlated with themselves. Trading of this Western Denmark CfD product was not possible until June 1, 2007.

The eigenvalues of the correlation matrix in Table 1.1 are 2.92, 1.74, 0.33, 0.07 and –0.05,

	NO1	SE	FI	DK1	DK2
NO1	1.00	–0.74	–0.76	*0.13*	-0.36
SE	–0.74	1.00	0.96	*–0.08*	0.38
FI	–0.76	0.96	1.00	*0.08*	0.41
DK1	*0.13*	*-0.08*	*0.08*	1.00	*0.89*
DK2	–0.36	0.38	0.41	*0.89*	1.00

Table 1.1: Estimated pairwise correlations between the CfD prices from Figure 1.5.

so the matrix is clearly not valid. This is due to the late introduction of the DK1 price area. Therefore, the experts in the market may have a different opinion of what the correlations with DK1 should be. We might say that they have chosen the Bayesian path to statistically sound correlation matrices.

1.2.2 Some possible numerical solutions

"Premature optimization is the root of all evil."

— Donald Knuth, *Structured Programming with Goto Statements (1974)*

Suppose we have estimated an invalid correlation matrix $\widehat{\boldsymbol{R}}$, and we want to adjust it to a proper matrix $\boldsymbol{R}^*$. The numerical literature (Higham, 2002; Pietersz and Groenen, 2004; Qi and Sun, 2006; Grubišić and Pietersz, 2007; Borsdorf and Higham, 2010) measures the distance between matrices by means of the Frobenius norm. The problem is then to find the proper correlation matrix $\boldsymbol{R}^*$ which minimises the Frobenius norm of the difference, defined as

$$||\widehat{\boldsymbol{R}} - \boldsymbol{R}^*||_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} |\widehat{\rho}_{ij} - \rho_{ij}^*|^2,$$

or the weighted Frobenius norm

$$||\widehat{\boldsymbol{R}} - \boldsymbol{R}^*||_W^2 = ||\boldsymbol{W}^{1/2}(\widehat{\boldsymbol{R}} - \boldsymbol{R}^*)\boldsymbol{W}^{1/2}||_F^2, \tag{1.7}$$

where $\boldsymbol{W}$ is a given, fixed, symmetric, positive definite weight matrix. The optimal matrix $\boldsymbol{R}^*$ is found by numerical minimisation, and the focus is on reaching the optimum fast also for very large matrices. For example, Qi and Sun (2006) show that their algorithm converges quadratically, which means that

$$\lim_{k\to\infty} \frac{|x_{k+1} - T|}{(x_k - T)^2} \to \phi \mid \phi > 0.$$

Here, T is the number the sequence $\{x_k\}$ should converge to.

The adjusted $\boldsymbol{R}^*$ will lie on the boundary and may in fact be numerically invalid. The solution will lie on the boundary, since pushing the most negative eigenvalue to 0^+ is part of minimising (1.7). With an eigenvalue close to zero, $\boldsymbol{R}^*$ may be invertible in theory, but not with fixed-precision Cholesky factorisation, say.

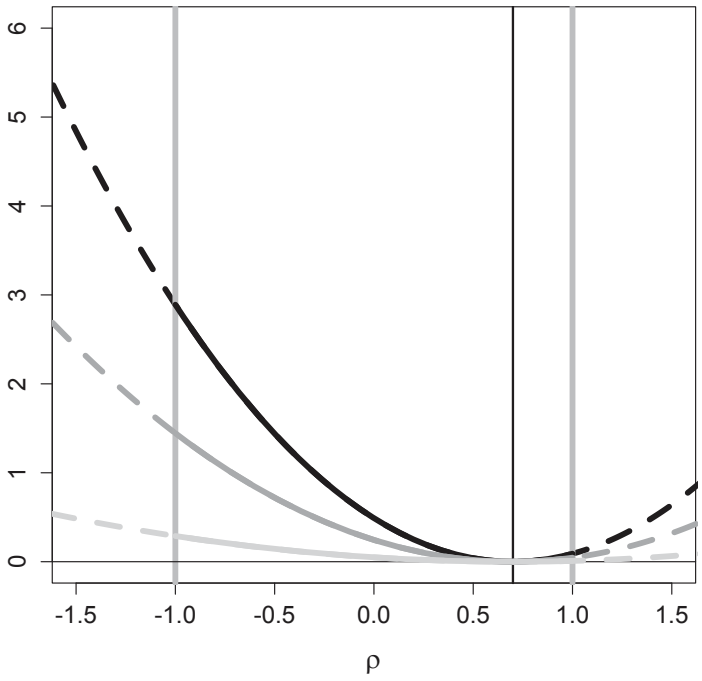


Figure 1.6: The Frobenius norm (1.7) with different weights around a correlation of 0.7.

There is no interpretable connection between the minimisation procedure of these norms and the pairwise correlations. The norm (1.7) is symmetric around $\widehat{\boldsymbol{R}}$ (Figure 1.6), but for strong correlations (close to –1 or 1), it is, however, natural to use a non-symmetric norm. More generally, the modification which the original estimated correlations $\widehat{\boldsymbol{R}}$ shall undergo should be interpretable.

Any bounded probability distribution can act as a non-symmetric norm. With non-symmetric norms, we assume that confidence in every pairwise correlation of the improper matrix is described by a distribution with mean in the current value. The beta distribution, scaled to $[-1, +1]$, is sufficiently flexible, though easy to utilise in practice (Figure 1.7). In this way we can express the problem by means of a probability density, where the unknown parameters, here the elements of the proper correlation matrix, each follow a beta distribution. We propose an algorithm that maximises the product of such beta densities within the set of proper correlation matrices. This is in general different from minimising (1.7), which can be seen as a special case.

1.2.3 "So says the statistician."

So far we have considered norm based methods, either the Frobenius norm or the product of beta densities. Neither of them take properly into account how these pairwise correlations arose

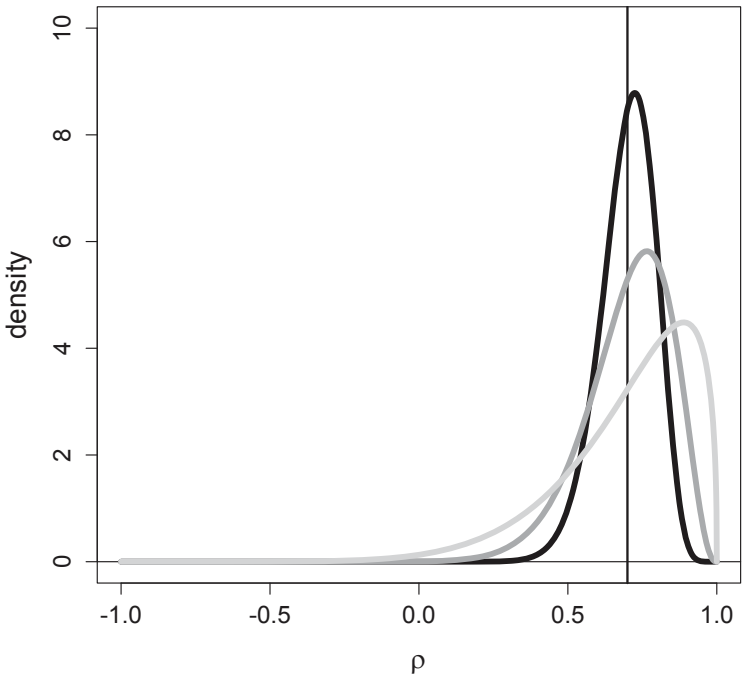


Figure 1.7: Scaled Beta distributions with a mean of 0.7.

or the number of data points behind each pairwise correlations. There is therefore a need for inferential methods.

Statistical adjustment of invalid correlation matrices is not a common theme in the literature, but especially Bayesian models for (valid) covariance matrices are more common. Modelling the covariance matrix is a more general statistical problem, and the standard approach is to assume observations from a multivariate normal distribution. Chan and Jeliazkov (2009) and Barnard et al. (2000) explicitly or implicitly argue for a diffuse prior for the correlation matrix, for example by using the inverse Wishart distribution or variants thereof as a prior for the correlation matrix. Barnard et al. (2000) model the covariance matrix $\boldsymbol{\Sigma}$ in terms of its standards deviations ($\boldsymbol{S}$) and correlation matrix

$$\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{S})\ \boldsymbol{R}\ \text{diag}(\boldsymbol{S}).$$

Barnard et al. argue that we tend to be willing to express prior beliefs on $\boldsymbol{S}$, and less willing on $\boldsymbol{R}$, and then write the prior on $\boldsymbol{\Sigma}$ as

$$\pi(\boldsymbol{S}, \boldsymbol{R}) = \pi(\boldsymbol{S})\pi(\boldsymbol{R}|\boldsymbol{S}),$$

where $\pi(\boldsymbol{R}|\boldsymbol{S})$ is diffuse (marginally or jointly uniform). The authors state that "particularly in

high dimensional problems, priors are never really "non-informative", so some care is needed." As we later will argue, priors could in many cases be informative, and a pairwise specification of them is preferable. The alternative is difficult for a user: How should she specify an informative prior distribution for correlations between correlations?

Turning to the frequentist view, if we assume that the correlation estimates originate from separate bivariate studies and that the distribution of a given pair (Y_i, Y_j) is well approximated by a bivariate normal distribution, Anderson (1984) shows that marginally

$$\widehat{\rho}_{ij} \sim \mathrm{N}\left(\rho_{ij}, \frac{(1-\rho_{ij}^2)^2}{n_{ij}}\right) \tag{1.8}$$

is a good approximation. Here, n_{ij} is the sample size involved. The distribution is centred around the actual correlation, with the scaling of the variance in the sample size (Figure 1.8).

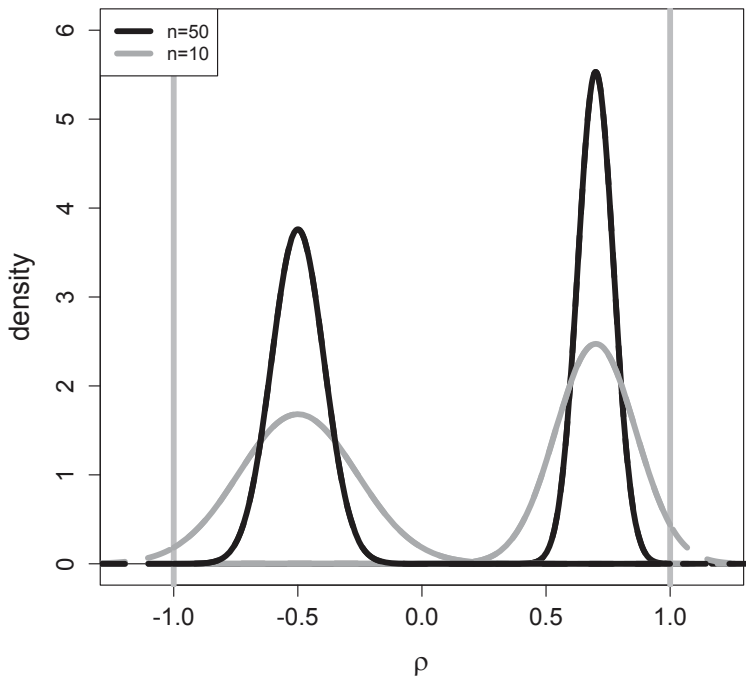


Figure 1.8: Bivariate normal approximation for the estimated pairwise correlation from Equation (1.8), here with means of –0.5 and 0.7.

The pseudo maximum likelihood estimate can then be obtained by maximising the log-likelihood function under the constraint that the parameters ρ_{ij} are such that $\boldsymbol{R} \in \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is the set of proper matrices, i.e. symmetric, with unit diagonal and only nonnegative eigenvalues.

1.3 Statistical models in gas markets

"The good Lord didn't see fit to put oil and gas only where there are democratically elected regimes friendly to the United States. Occasionally we have to operate in places where, all things considered, one would not normally choose to go. But, we go where the business is."

— Dick Cheney, *Speech delivered at the Cato Institute (June 23, 1998)*

Dick Cheney's sometimes seemingly malicious agenda is not a topic for this thesis, but natural gas markets are. When the large gas fields in the North Sea (like the Troll field) first were explored, there was a need to enter long term contracts to finance the long term commitments in platforms and pipelines. These contracts were typically linked to a corresponding commodity with a more or less functioning market, namely oil. In the European natural gas market, there are many such long term contracts that also are designed to allow flexibility of delivery (Asche et al., 2002). The timing and the amount of energy are allowed to be flexible, within certain constraints. These contracts can last for ten years or more and the volumes are large. Especially in continental Europe, these contracts constitute more than half of the turn over in the markets. The rest of the gas is traded on spot (read: "day-ahead") markets, like the British NBP (National Balancing Point) market. The NBP is a notional point in the UK Transmission System (NTS). For accounting and balancing purposes all gas is said to flow through this point.

1.3.1 Risk premiums

"Audentes fortuna iuvat." ("Fortune favours the brave.")

— Publius Vergilius Maro ("Virgil"), *The Aeneid, Book X, line 284*

In the NBP gas market, and any commodity market where there is a separate spot and financial market, the players have the opportunity to sell or buy their production or consumption in advance, like with the CfD prices in Section 1.2.1. Advanced players may base their potentially risky positioning on a thorough investigation of (another) Holy Grail called the risk premium. As with the actual Holy Grail, this risk premium is not directly observable, and we are left to speculate on its whereabout and origin.

If the producers are the most compelled to hedge, they have to sell forward gas cheaper than the expected value of future spot prices (Keynes, 1930; Hicks, 1939). Hence, the net position is short and the price of forward contracts will decrease. If, on the other hand, consumers are constrained to secure future prices, the opposite will occur, pushing forward prices above expected future spot prices (Duffie, 1989; Longstaff and Wang, 2004).

Different risk premium definitions exist (Fama and French, 1987; Benth et al., 2008), but the following will serve our purpose. Let ${}_{t-k}F_t$ denote the forward or futures price quoted at time $t-k$ for delivery at time t, and S_t the spot price at time t. In this if there is no risk premium,

$$\mathsf{E}_{t-k}[S_t] = {}_{t-k}F_t,$$

where E_{t-k} is the expectation conditioned on all relevant information available at time $t-k$.

Many of the statistical tests used in this field require stationary and normally distributed data. The augmented Dickey-Fuller (ADF) and the Phillips-Perron (PP) unit-root tests (Dickey and Fuller, 1979; Phillips and Perron, 1988) assist us here.

If a risk premium exists, a useful, linear generalisation is the regression

$$S_t = \alpha + \beta({}_{t-k}F_t) + e_{k,t}, \tag{1.9}$$

where α and β are constants, and the error term $e_{k,t}$ typically follows a moving average process of order $k-1$. Our null hypothesis (no risk premium) is then

$$H_0\text{: } \alpha = 0 \text{ and } \beta = 0.$$

If $\alpha \neq 0$ and $\beta = 1$, there is a constant risk premium around the forward price. If, in addition, $\beta \neq 1$, the risk premium varies with the forward price. Equation (1.9) describes one of the simplest possible cointegration relationships.

Natural gas is a storable commodity, even though the storage capacity is limited. Therefore, storage may also affect the difference between forward and spot prices. The theory of storage (Fama and French, 1987) says that this price difference depends linearly on the interest rate, marginal cost of storage and convenience yield. The latter two are, in turn, functions of the inventory. The difference between forward and spot prices may be influenced by inventory even if the theory of storage does not apply to the natural gas market. In particular, this difference may depend on other factors as well as storage. Moreover, the relationship could be non-linear, and the generalised additive model machinery of Section 1.1.1 can be applied.

1.3.2 Pricing flexible contracts

Long term gas contracts are also called take-or-pay contracts or swing options, since the buyer either takes the gas volume from the supplier or pays the supplier a penalty. Such contracts may have been beneficial for the producers when they entered them, but with their long term flexibility and constraints they are very difficult to price. To price a flexible contract you need two key ingredients: a price process and a pricing machinery.

We will merely touch upon the price process here. For contracts lasting e.g. ten years, the difference between a stationary and non-stationary price process can be enormous. Likewise, knowledge of the existence and nature of a time-varying risk premium is key.

When the first long term gas contracts were entered, we suspect that the true contract value was unknown. The seminal paper of Longstaff and Schwartz (2001) describes a method for valuing American options using least-squares Monte Carlo simulations (LSM). The method uses scenarios from any price process, giving full flexibility on the underlying price process. American options are a special case of the long term gas contracts, and the methodology can be extended to swing options. Generally, a swing option is equal to N nested American-style call options (N being the number of exercise rights), similar to a Bermudan option. But where the Bermudan option has predetermined exercise dates, the swing option has further optionality.

The idea with the LSM is to work backwards in time. At the last time step (i.e. the last day of the contract period), the option is exercised if the option is in the money and expired if not.

At the time step prior to the last time step, the holder has two possibilities. If the option is not in the money, the option is not exercised. If the option is in the money, the option is exercised if the instant payoff is higher than the expected value of waiting, which is referred to as the continuation value. The idea behind the least-squares Monte Carlo method is to use least-squares regression to find the continuation values.

The method starts with defining a set of basis functions and use a regression to find the parameter vector β_i for each time step i in (1.10) below. The idea is based on the equality

$$\mathsf{E}[V_{i+1}(S_{i+1})|S_i = x] = \sum_{r=1}^{M} \beta_{ir}\psi_r(x). \tag{1.10}$$

Here, V_i is the value of the option (which is known at the final time T). S_i is the price of the underlying, $\psi_r(x)$, $r = 1, \ldots, M$, are the basis functions and β_{ir} are the parameter vectors to be estimated. It can be shown that (1.10) is true when $M \to \infty$ (Stentoft, 2004). The method assumes that for a finite M, we have an appropriate approximation. β_{ir} are estimated by least-squares regression, hence the name of the method.

The basis functions can be functions of the underlying asset, or different states of factors included in the price process. Typically they are polynomials of degree $M - 1$. The regression is done with price paths that are in the money only and each continuation value, $C_i(x)$, is computed as

$$C_i(x) = \psi(x)^{\mathrm{T}} \cdot \beta_i. \tag{1.11}$$

The equation for the regression to estimate β_i is

$$\psi(S_i)^{\mathrm{T}} \cdot \beta_i = e^{-r\Delta t} \cdot V_{i+1}, \tag{1.12}$$

where $r\Delta_t$ is the discounting factor with interest rate r.

The simulation algorithm is as follows:

- Simulate b price paths
- Set $\hat{V}_{Tj} = h(S_{Tj})$ at the end points of each price path, $j = 1, ..., b$, where $h(S_{Tj})$ is the payoff at time T
- For each time step, $i = T - 1, ..., 1$, work backwards in time
  - Calculate $\hat{\beta}_i$ with estimated values $\hat{V}_{i+1,j}$ from (1.12)
  - Calculate continuation values $\hat{C}_i(S_{ij})$ using $\hat{\beta}_i$ as in (1.11)
  - Set
$$\hat{V}_{ij} = \begin{cases} h(S_{ij}), & \text{if } h(S_{ij}) \geq \hat{C}_i(S_{ij}) \\ e^{-r\Delta t}\hat{V}_{i+1,j}, & \text{if } h(S_{ij}) < \hat{C}_i(S_{ij}) \end{cases}$$
- The option value is given as $\hat{V}_0 = (\hat{V}_{11} + \cdots + \hat{V}_{1b})/b$

Our challenge has been to extend the LSM method to be able to include a so-called carry forward right. A carry forward right gives the holder of the option the possibility to transfer unused flexibility to future years.

Chapter 2

Summary of papers

Paper I
Modeling Nord Pool's NO1 area price

Authors: Løland, Anders; Dimakos, Xeni K.

Status: Published in Journal of Energy Markets, 2010, Vol. 3, Number 1, pp. 73-92.

Synopsis: We investigate how the daily Nord Pool Spot NO1 area price could be explained by using historical reservoir levels, flow and capacities, as well as seasonal terms

Flow and capacities were aggregated from hourly to daily values. Combining flow and capacity, we defined the net capacity utilisation. Using the generalised additive model (GAM) framework, we estimated a model for the relative price difference between NO1 and the Nord Pool system price. The model provides sensible results, and supports the notion that NO1 price is below the system price when the water reservoir levels are high relative to normal levels, and when the export capacity is limited. A more thorough investigation could validate the model historically, by some sort of cross-validation or out-of-sample testing. The historic data period makes such a fair validation difficult, since really only the last year of data include large, systematic price differences.

Our results indicate that the models are able to explain a substantial amount of the variability in the data, but there is room for improvement. Explanatory variables not included here, like CO_2 emission prices and snow reservoir levels, might improve the model. Also, including lagged effects could improve the model. However, the model is a good starting point for further modelling, and the results are interesting and plausible.

Paper II
Forecasting transmission congestion

Authors: Løland, Anders; Ferkingstad, Egil; Wilhelmsen, Mathilde

Status: Published in Journal of Energy Markets, 2012, Vol. 5, Number 3, pp. 65-83.

Synopsis: We re-visit the NO1 price area, and analyse it from a different angle. Transmission congestion is the reason for area price differences. We therefore investigate models for forecasting hourly day-ahead transmission congestion. Using the net capacity definition for a price area as a whole, we avoid restricting our analysis to pairs of neighbouring price areas. Still, our method can also be applied for other price areas, and for two neighbouring price areas, if that is of interest.

The best out-of-sample results were found by adaptive model combinations. The best combination method was to take the weighted sum with respect to previous performance. For high absolute values of the net capacity utilisation, our combination method is not superior to the other methods.

Our approach is a first attempt at predicting transmission congestion, and there are plenty of opportunities for future work. First, we have only provided point forecasts. These should be accompanied by proper uncertainty measures. Second, we would like to predict beyond one day ahead. Our preliminary investigations suggest that including congestion forecasts in simple price models does not improve price predictions, and more sophisticated price forecast models are needed.

Paper III
Causal modeling and inference for electricity markets

Authors: Ferkingstad, Egil; Løland, Anders; Wilhelmsen, Mathilde

Status: Published in Energy Economics, 2011, Vol. 33, Issue 3, pp. 404-412.

Synopsis: How does dynamic price information flow among Northern European electricity spot prices and prices of major electricity generation fuel sources? We use time series models combined with new advances in causal inference to answer these questions. Applying our methods to weekly Nordic and German electricity prices, and oil, gas and coal prices, with German wind power and Nordic water reservoir levels as exogenous variables, we estimate a causal model for the price dynamics, both for contemporaneous and lagged relationships. In contemporaneous time, Nordic and German electricity prices are interlinked through gas prices. In the long run, electricity prices and British gas prices adjust themselves to establish the equilibrium price level, since oil, coal, continental gas and EUR/USD are found to be weakly exogenous. The results are compared to a previous US study.

There are two main methodological advantages of our approach. First, we are able to identify one unique contemporaneous graph, as opposed to a Markov equivalence class (which might be large). Second, we are able to properly and coherently deal with both instantaneous and time-lagged effects in the same analysis. We have shown that directed acyclical graphs are in fact useful for combining time-lagged and instantaneous effects.

Paper IV
Statistical rehabilitation of improper correlation matrices

Authors: Frigessi, Arnoldo; Løland, Anders; Pievatolo, Antonio; Ruggeri, Fabrizio

Status: Published in Quantitative Finance, 2011, Vol. 11, Issue 7, pp. 1081-1090.

Synopsis: The standard approach to describe the dependence for a set of financial assets is their correlation matrix. This correlation matrix can be improper when it is specified element-wise or is estimated from incomplete or partly overlapping data. We describe a new method for obtaining a positive definite correlation matrix starting from an improper one. The expert's opinion and trust in each pairwise correlation is described by a beta distribution. Then, by combining these individual distributions, a joint distribution over the space of positive definite correlation matrices is obtained using the Cholesky factorisation, and its mode constitutes the new proper correlation matrix. We also sketch a Bayesian approach to the same problem.

We present new ways of visualising what we call transformation hotspots, that is positions where the correlation has been adjusted significantly. This allows a detailed monitoring of the effects of the procedure.

Paper V
Statistical corrections of invalid correlation matrices

Authors: Løland, Anders; Huseby, Ragnar Bang; Hjort, Nils Lid; Frigessi, Arnoldo

Status: Accepted for publication in Scandinavian Journal of Statistics.

Synopsis: This is a generalisation and improvement of the approach from Paper VI. We propose frequentist and Bayesian solutions to the problem of finding a valid correlation matrix, given correlation estimates obtained from separate data. We suggest two likelihood solutions: either a binormal approximation for the data, resulting in a normal approximation for each correlation, or assuming that each correlation is beta distributed. We suggest a range of possible priors. With the beta prior, an expert may express her opinions on each correlation in a clear fashion.

From a practical point of view, we support the use of the posterior means. The posterior mean will, for proper priors, generally have truly positive eigenvalues. This is in contrast to the posterior mode and the maximum likelihood solution, which may very well have at least one eigenvalue very close to zero.

To sample from the posterior distribution, the standard approach is to work on the Cholesky decomposition of the correlation matrix, to ensure positive definiteness, and constrain the Cholesky elements to obtain a correlation matrix. We transform the correlations first to the Cholesky representation, and then further to hyperspherical coordinates. In that way the correlations lie in a hypercube, and we are guaranteed to find a valid correlation matrix.

A Metropolis-Hastings sampler was implemented to obtain posterior matrices. We have combined a Metropolis-Hastings sampler (MCMC) with simulated annealing to find the Maximum likelihood solution and the posterior mode, but a proper numerical optimiser will also suffice, and in many cases be faster.

Paper VI
Risk premium in the UK natural gas forward market

Authors: Hobæk Haff, Ingrid; Lindqvist, Ola; Løland, Anders

Status: Published in Energy Economics, 2008, Vol. 30, Issue 5, pp. 2420-2440.

Synopsis: We have analysed the UK gas forward market, represented by the five monthly forward contracts that are traded in this market. Mainly, we have investigated whether the forward prices are unbiased predictors of the future spot prices.

The forward and spot price series are all non-stationary, but they are cointegrated with significant parameters, as in the US market. As one would expect, the contracts with longer time to delivery are less sensitive to new market information.

Furthermore, the forward prices overestimate the spot prices. This means that there is a positive risk premium in the forward contracts, which is consistent with previous results. Moreover, this risk premium appears to be time-varying, in the sense that it varies with the time to delivery. The price changes from one month to the next are highest the last month before delivery.

We also tested whether the theory of storage is applicable to the UK natural gas market. Although storage seems to have an impact on the difference between the spot and the forward price, the storage model clearly is not a complete model for this market. One reason may be that the storage capacity for gas is rather limited compared with other commodities.

Paper VII
Valuation of long term, flexible gas contracts

Authors: Holden, Lars; Løland, Anders; Lindqvist, Ola

Status: Published in Journal of Derivatives, 2011, Vol. 18, Number 3, pp. 75-85.

Synopsis: Many contracts in the energy markets are designed to allow flexibility of delivery. Both the timing and the amount of energy are allowed to be flexible, within certain constraints. There are many long term contracts with such flexibility in the European natural gas market, which typically last for ten years or more and the volumes are large. Correct valuation of long term contracts is important not only for pricing when entering the contract and possibly hedging, but also when renegotiations of contract parameters take place.

We discuss the major types of flexibility and constraints related to long term gas contracts, also known as take-or-pay contracts or swing options. By adapting the Least Squares Monte Carlo method, we introduce a method to price contracts which include more flexibility constraints than previous authors have done. We focus on the carry forward option, which allows flexibility between years. This is considered to be the most important part of the contract that has not been quantified earlier. It is shown that carry forward options give a significant increase in the value of the contract.

Bibliography

Aalen, O. O. and Frigessi, A. (2007). What can statistics contribute to a causal understanding? *Scandinavian Journal of Statistics*, 34(1):155–168.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley, second edition.

Asche, F., Osmundsen, P., and Tveterås, R. (2002). European market integration for gas? Volume flexibility and political risk. *Energy Economics*, 24(3):249–265.

Barnard, J., McCulloch, R., and Meng, X. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4):1281–1311.

Bates, J. M. and Granger, C. W. J. (1969). The Combination of Forecasts. *Operational Research Quarterly*, 20(4):451–468.

Benth, F. E., Benth, J. S., and Koekebakker, S. (2008). *Stochastic modelling of electricity markets*, volume 11 of *Advanced Series on Statistical Science and Applied Probability*. World Scientific Publishing.

Borsdorf, R. and Higham, N. J. (2010). A preconditioned newton algorithm for the nearest correlation matrix. *IMA Journal of Numerical Analysis*, 30(1):94–107.

Brunekreeft, G. and Twelemann, S. (2005). Regulation, competition and investment in the German electricity market: RegTP or REGTP. *Energy Journal*, pages 99–126.

Chan, J. C.-C. and Jeliazkov, I. (2009). MCMC Estimation of Restricted Covariance Matrices. *Journal of Computational and Graphical Statistics*, 18(2):457–480.

Chickering, D. M. (2003). Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3(3):507–554.

Claeskens, G. and Hjort, N. L. (2008). *Model Selection and Model Averaging*. Cambridge series in statistical and probibalistic mathematics. Cambridge University Press.

Clemen, R. T. and Winkler, R. L. (1986). Combining Economic Forecasts. *Journal of Business & Economic Statistics*, 4(1):39–46.

Comon, P. (1994). Independent component analysis – a new concept? *Signal Processing*, 36:287–314.

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

Dimakos, X. K. and Aas, K. (2004). Integrated risk modelling. *Statistical modelling*, 4(4):265–277.

Duffie, D. (1989). *Futures markets*. Prentice Hall.

Fama, E. F. and French, K. R. (1987). Commodity futures prices: some evidence on forecast power, premiums, and the theory of storage. *The Journal of Business*, 60(1):55–73.

Fosso, O., Gjelsvik, A., Haugstad, A., Mo, B., and Wangensteen, I. (1999). Generation scheduling in a deregulated system. The Norwegian case. *IEEE Transactions on Power Systems*, 14(1):75–80.

Fridolfsson, S. O. and Tangerås, T. P. (2009). Market power in the Nordic electricity wholesale market: A survey of the empirical evidence. *Energy Policy*, 37(9):3681–3692.

Grubišić, I. and Pietersz, R. (2007). Efficient rank reduction of correlation matrices. *Linear algebra and its applications*, 422(2-3):629–653.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*. Chapman & Hall.

Hicks (1939). *Value and Capital*. Oxford University Press.

Higham, N. J. (2002). Computing the nearest correlation matrix – a problem from finance. *IMA Journal of Numerical Analysis*, 22:329–343.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634.

Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430.

Keynes, J. M. (1930). *A treatise on money. Volume II: The applied theory of money*. MacMillan.

Kristiansen, T. (2004). Congestion management, transmission pricing and area price hedging in the Nordic region. *International Journal Of Electrical Power & Energy Systems*, 26(9):685–695.

Longstaff, F. A. and Schwartz, E. S. (2001). Valuing american options by simulation: A simple least-squares approach. *The Review of Financial Studies*, 14(1):113–147.

Longstaff, F. A. and Wang, A. W. (2004). Electricity forward prices: A high-frequency empirical analysis. *Journal of Finance*, 59(4):1877–1900.

Marckhoff, J. and Wimschulte, J. (2009). Locational price spreads and the pricing of contracts for difference: Evidence from the Nordic market. *Energy Economics*, 31(2):257–268.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, second edition.

Medova, E. A. (2000). Extreme value theory: extreme values and the measurement of operational risk. *Operational Risk*, 1(7):13–17.

Müsgens, F. (2006). Quantifying market power in the German wholesale electricity market using a dynamic multi-regional dispatch model. *Journal of Industrial Economics*, 54(4):471–498.

Palm, F. C. and Zellner, A. (1992). To Combine or not to Combine? Issues of Combining Forecasts. *Journal of Forecasting*, 11:687–701.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge University Press.

Pereira, M. V. F. (1989). Optimal stochastic operations scheduling of large hydroelectric systems. *Electrical Power and Engineering Systems*, 11(3):161–169.

Pereira, M. V. F. and Pinto, L. M. V. G. (1991). Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375.

Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regressions. *Biometrika*, 75:335–346.

Pietersz, R. and Groenen, P. J. F. (2004). Rank reduction of correlation matrices by majorization. *Quantitative Finance*, 4:649–662.

Qi, H. and Sun, D. (2006). A quadratically convergent newton method for computing the nearest correlation matrix. *SIAM Journal on Matrix Analysis and Applications*, 28(2).

Rousseeuw, P. J. and Molenberghs, G. (1994). The Shape of Correlation Matrices. *The American Statistician*, 48:276–279.

Schwartz, E. (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance*, 52(3):923–973.

Shimizu, S., Hoyer, P., Hyvärinen, A., and Kerminen, A. (2006). A linear non-Gaussian acyclic model for causal discovery. *The Journal of Machine Learning Research*, 7:2003–2030.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. The MIT Press, Cambridge, MA.

Stentoft, L. (2004). Convergence of the least squares monte-carlo approach to american option valuation. *Management Science*, 50(9).

Tchernitser, A. and Rubisov, D. H. (2009). Robust estimation of historical volatility and correlations in risk management. *Quantitative Finance*, 9(1):43–54.

Timmermann, A. (2006). *Handbook of Economic Forecasting*, volume 1, chapter Forecast Combinations, pages 135–194. Elsevier.

Weigt, H. and von Hirschhausen, C. (2008). Price formation and market power in the German wholesale electricity market in 2006. *Energy Policy*, 36(11):4227–4234.

Weron, R. (2006). *Modelling and Forecasting Electricity Loads and Prices, A Statistical Approach*. John Wiley & Sons Ltd.

Wolfgang, O., Doorman, G., Gjelsvik, A., Mo, B., Haugstad, A., Førsund, F. R., and Singh, B. (2005). Samkjøringsmodellen: En oversikt med vekt på økonomisk tolkning. SINTEF/SNF.

I

II

III

IV

V

VI

VII