Modelling Electricity Prices in 15-Minute Intervals

A Panel Data Approach for a New Market Regime in the Electricity Pricing

AALBORG UNIVERSITY

Master Thesis

Andreas Laurids Pedersen & Elissa Kamil Yammin

Aalborg University
Department of Mathematical Sciences
Thomas Manns Vej 23, 3rd floor
DK-9220 Aalborg Ø

Typeset by the authors in LaTeX. Illustrations made with R.

AALBORG UNIVERSITY
STUDENT REPORT

Fifth year w/
Department of Mathematical Sciences
Mathematics-Economics
Thomas Manns Vej 23,
9220 Aalborg Øst
http://math.aau.dk

Title
Modelling Electricity Prices in 15-Minute Intervals

Topics
Panel Data Models
Factor Models
Power Markets

Project Period
Spring Semester 2026

Project Group
MO10-08
Andreas & Elissa

Participants
Andreas Laurids Pedersen
Elissa Kamil Yammin

Supervisor
J. Eduardo Vera-Valdés

Page Count: 96
Date of Completion:
May 26, 2026

Abstract

This thesis investigates whether panel data models can be used to model and forecast electricity prices under the new 15-minute settlement regime in the Danish DK1 electricity market. Building on panel data econometrics, the project examines dynamic panel data models and factor-augmented specifications for capturing persistence, seasonality, and cross-sectional dependence in electricity prices.

The theoretical part introduces fixed effects, random effects, dynamic panel models, and factor-augmented models. Particular emphasis is placed on modelling temporal persistence and latent common market structures following the transition from hourly to quarter-hourly settlement intervals.

In the application, the methodologies are applied to hourly and quarter-hourly electricity market data. The results indicate that the quarter-hourly specifications improve forecasting performance and capture more detailed market dynamics compared to the hourly models. Furthermore, the factor-augmented specifications successfully account for latent dependence structures, although the resulting forecasts become smoother and less responsive to abrupt market shocks and extreme price spikes.

Preface

This project is written in the Spring semester of 2026 by Andreas & Elissa, tenth-semester students of Mathematics-Economics at the Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark. The document is typeset with LaTeX.

The group would like to acknowledge the project advisor, Assistant Professor **J. Eduardo Vera-Valdés**, and thank him for his encouragement, insightful comments, and guidance throughout the project period.

The following programs were used in the writing of this project:

- **Overleaf** — Writing.
- **Google Drive** — Notes and file sharing.
- **Github** — File sharing.
- **RStudio** — Statistical calculations and data analysis.
- **ChatGPT** — Coding, Grammar & spelling corrections.

Reader's Guide

The table of contents is given on page vi. When the document is viewed as a PDF, entries in the table of contents act as hyperlinks, which allows simple and efficient navigation throughout the project.

The literature used in this project is introduced in the bibliography on page 78. The references in the bibliography follow the APA method shown below:

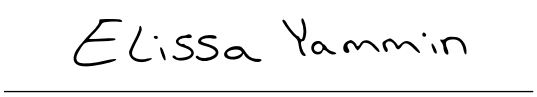
[Author][Year][Title](Publisher)(Edition)(URL).

Fields in [square brackets] are mandatory, while fields in (regular parentheses) are only relevant for certain formats (e.g., scientific articles and web pages). The entries in the bibliography are sorted alphabetically. Source references are typically given at the beginning of chapters or sections. Appendices can be found following the bibliography.

All programming in the project is done in R; the source code can be accessed through the following link: `https://github.com/ElissaShaker/P10_master_thesis`

Signatures

Andreas Laurids Pedersen
<apedee21@student.aau.dk>



Elissa Kamil Yammin
<eyammi21@student.aau.dk>

Contents

1 Introduction

The following chapter is based on [1], [6], [17], and [9].

Electricity markets have undergone substantial structural changes during the last decades. The increasing integration of renewable energy sources, the liberalization of European electricity markets, and the transition toward finer settlement intervals have transformed both the operational and financial dynamics of power systems. Unlike many other commodities, electricity must be balanced continuously in real time, since large-scale storage remains limited. As a consequence, even small forecasting errors in production or consumption can lead to significant price fluctuations and imbalance costs.

These challenges have become increasingly relevant in the Nordic electricity market, where renewable generation from wind and solar power constitutes a growing share of total electricity production. Renewable energy sources are inherently weather-dependent, introducing substantial uncertainty and short-term volatility into the system. At the same time, consumers have become increasingly responsive to electricity prices through variable pricing contracts, electric vehicle charging, and flexible consumption patterns. This has resulted in a higher volatility market.

A major structural reform in the European electricity market is the transition from hourly settlement to 15-minute settlement intervals. In Denmark, this transition was fully implemented in October 2025. Under the previous hourly structure, forecast deviations were averaged over an entire hour, whereas the new market design settles imbalances at a much finer resolution. Consequently, intrahour fluctuations in both consumption and renewable generation have become economically relevant to a much greater extent. This increases the importance of accurate short-term forecasting models capable of capturing rapid market movements and dynamic dependencies between delivery periods.

From a statistical perspective, electricity prices and consumption exhibit several properties that complicate traditional modelling approaches. The series are highly persistent over time, display strong seasonal and intraday patterns, and are often influenced by common latent shocks such as weather systems, transmission congestion, fuel prices, and balancing market conditions. Since these shocks may simultaneously affect multiple delivery periods, substantial cross-sectional dependence arises across the intraday price curve. Standard time series models may therefore fail to adequately capture the underlying dependence structure of the market.

To address these challenges, panel data models provide a natural framework. Panel data combines both cross-sectional and temporal information, allowing the simultaneous modelling of multiple delivery periods observed repeatedly over time. In particular, dynamic panel data models extend this framework by incorporating lagged dependent variables,

thereby capturing persistence and gradual adjustment dynamics. Furthermore, by augmenting the model with latent factor structures, common shocks affecting all delivery periods can be incorporated through unobserved multifactor error components. Such factor-augmented dynamic panel models provide a flexible methodology for modelling both time-persistence and cross-sectional dependence in electricity markets.

1.1 Problem Analysis

The transition toward 15-minute settlement intervals introduces new operational challenges for both market participants and utility companies. Under the finer market resolution, short-term fluctuations in renewable production and consumption patterns become increasingly important, as imbalances are settled more frequently.

To address this, the project considers panel data models. The panel framework allows each hourly or quarter-hourly delivery period to be treated as a cross-sectional unit observed repeatedly over time, while the latent factor structure captures common shocks shared across the market. In addition, instrumental variable and generalized method of moments (GMM) methodologies are considered in expansionism models.

The project further investigates whether the transition from hourly to 15-minute settlement intervals alters the dynamics and predictability of electricity prices. Specifically, simple forecasting performance of models estimated on hourly data is compared with models estimated on quarter-hourly data to evaluate whether the additional granularity can potentially provide improved predictive capability.

1.2 Problem Statement

The central problem of the thesis can be stated as follows:

Can panel data models with lagged regressor structure be used to model and potentially forecast electricity prices under the new 15-minute settlement regime in the Danish(DK1) electricity market?

In answering the problem statement, the following questions will be addressed:

- *How does the transition from hourly to 15-minute settlement intervals affect the dynamics of electricity prices?*
- *Can factor-augmented dynamic panel models adequately capture the cross-sectional dependence induced by common market shocks?*
- *How do hourly and quarter-hourly panel models compare in terms of forecasting performance?*

2 European Power Market

The following chapter is based on [5], [7], [9], [16], [17], [21], and [26]

This chapter presents the current European energy market and its challenges. It provides an overview of how the market is structured, why balancing is necessary, and how this balance is secured. Finally, the chapter outlines the transition from hourly to 15-minute settlement and discusses the challenges arising from the increasing share of renewable energy in the European power system.

2.1 Market Balance and Trading Structure

A fundamental property of electricity markets is that supply and demand must be balanced at all times. Unlike many other commodities, electricity cannot (yet) be stored economically at large scale. Production and consumption therefore need to match continuously, otherwise system imbalances arise that threaten grid stability.

For this reason, the market is organised in layers. The two most relevant are the day-ahead market, and the balancing imbalance market. In the day-ahead market, both producers and consumers submit price-quantity bids for each delivery period. Producers submit supply offers specifying the minimum price at which they are willing to generate electricity, while consumers submit demand bids indicating the maximum price they are willing to pay.

These bids are aggregated into supply and demand curves for each time period as can be seen in Figure 2.1. The market operator then determines a market-clearing price by matching supply and demand such that total accepted generation equals total accepted consumption. All accepted participants are settled at this single clearing price, regardless of their individual bid prices. This uniform pricing mechanism reflects the marginal cost of the last unit required to balance the system.

The bidding starts at 10:00CET each day. The auction then closes at 12:00CET and the day-ahead price is determined and released at around 12:45CET.

In Figure 2.1 we can see that generation technologies with low marginal costs, such as wind and nuclear, are dispatched first, while more expensive units such as condensing plants and gas turbines are only activated when demand is sufficiently high. Demand varies over time, as illustrated by the three different demand curves during a specific day.

During periods of low demand, only low-cost generation is needed, leading to low electricity prices. In contrast, during peak demand periods, higher-cost generation units must be activated, which increases the market price. This pricing mechanism ensures that all

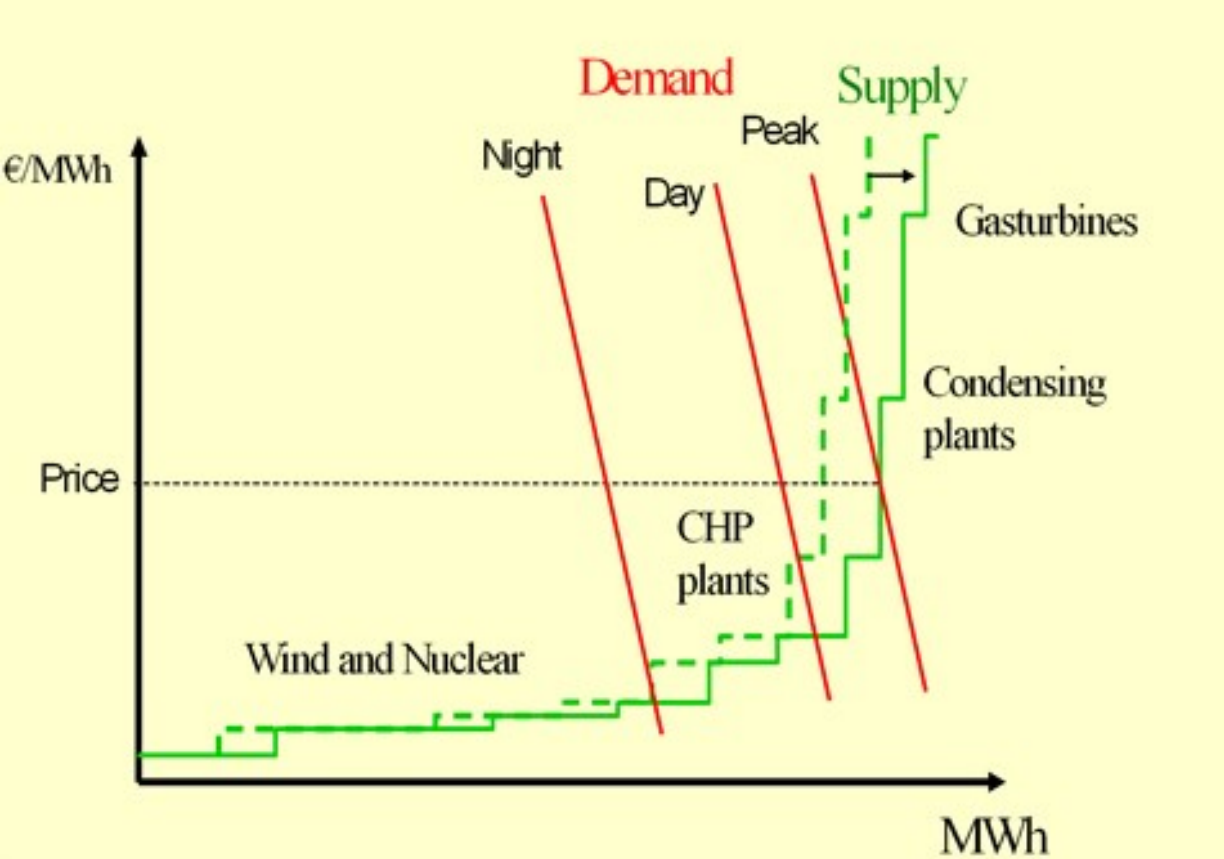


Figure 2.1: Supply and demand representation of the power energy market, for different time periods during the day. Source: [10]

dispatched units receive the same price, equal to the marginal cost of the last unit brought online.

Forecasts are, however, never perfect, a producer may promise to deliver a certain number of MW (Mega Watt) but fail to do so because of unexpected wind conditions or technical issues. In that case, the system operator must procure the missing electricity from other sources in real time to maintain balance. The producer who deviated from their committed position must then settle the deviation financially through the imbalance price.

In this way, the imbalance market acts as the real-time correction mechanism of the system. It ensures physical balance by activating upward or downward regulation and provides economic incentives for accurate forecasting and reliable delivery. The imbalance price reflects the marginal cost of restoring system balance and may differ significantly from the day-ahead price, particularly during periods of system stress.

The increasing share of renewable energy further amplifies these dynamics. As wind and solar generation are inherently variable and uncertain, both surplus and scarcity situations occur more frequently. This increases the importance of short-term markets, flexibility,

and accurate forecasting in maintaining system balance.

2.1.1 Surplus and Scarcity of Power Generation

If the system is not balanced, frequency deviations occur. Excess supply leads to an increase in system frequency, while insufficient supply causes it to drop. In Europe, the nominal system frequency is 50 Hz, with typical operational tolerance bounds of approximately ± 0.050 Hz. Significant deviations from this can damage equipment or, in extreme cases, lead to disconnections. Continuous balancing is therefore a necessity for the market, some examples follow:

A situation may arise where total generation exceeds demand. In terms of the merit-order framework in Figure 2.1, this can be interpreted as either an outward shift in supply, for instance due to high wind or solar production, or a reduction in demand.

In such cases, the market-clearing price decreases and may become negative. Negative prices occur when generators are willing to pay to remain in operation, for example due to high start-up costs or subsidy schemes tied to production.

If market-based mechanisms are insufficient to restore balance, the transmission system operator intervenes by curtailing generation. This includes renewable sources such as wind and solar, despite their low marginal costs. Curtailment is therefore a necessary tool to ensure that the physical balance between supply and demand is maintained.

This situation highlights the importance of system flexibility, as limited storage capacity and inflexible demand increase the likelihood of surplus conditions.

Conversely, periods may occur where available generation is insufficient to meet demand. In such situations, increasingly expensive generation units are activated as demand rises, leading to higher electricity prices.

In such situations, peaking units such as gas turbines are typically dispatched, and electricity may be imported from neighbouring markets if interconnection capacity allows. If supply remains insufficient, the system operator activates reserve capacity to maintain system stability.

In extreme cases, controlled load shedding may be required to prevent system collapse. This reflects the high economic cost of unserved energy, often referred to as the value of lost load (VoLL). As a result, electricity prices can increase significantly during scarcity periods, reflecting the critical importance of maintaining system balance.

2.2 History: Liberalisation of the European Electricity Market

Prior to the mid-1990s, electricity markets in Europe were largely organised as national monopolies. Vertically integrated firms often handled generation, transmission, and retail within the same organisation. Prices were regulated, and cross-border competition was limited.

This structure began to change with the liberalisation process initiated by the European Union in 1996. The objective was to introduce competition, improve efficiency, and integrate national markets. Over time, this led to unbundling requirements separating generation and transmission activities and the establishment of organised power exchanges.

Further regulatory packages, including reforms introduced in 2009, strengthened these separation requirements and reduced conflicts of interest by limiting the degree to which firms could simultaneously control production and transmission assets.

Today, electricity trading across much of Europe takes place on interconnected exchanges, and cross-border market coupling allows electricity to flow to where it is valued most highly, subject to transmission constraints.

2.3 Transition from Hourly to 15-Minute Settlement

Historically, most European day-ahead market products were traded in hourly intervals. Forecasting models, trading strategies, and settlement mechanisms were therefore built around hourly positions.

The transition to 15-minute settlement, fully integrated in Denmark since October 1st, 2025, represents a structural change in how market risk is measured and priced. Although imbalance prices have long been calculated at high temporal resolution, market positions are now defined in 15-minute intervals.

This change makes intrahour dynamics economically relevant. Forecast errors are no longer averaged over the hour but settled at a finer resolution. As a result, short-term volatility, ramping behaviour, and rapid changes in renewable generation play a more direct role in financial outcomes.

For forecasting and trading, this introduces both additional complexity and new opportunities, as more granular modelling may provide improved predictive performance under the new market design.

2.4 Renewable Energy and System Challenges

A central objective of the European Union is the large-scale integration of renewable energy sources. Wind and solar generation are inherently weather-dependent, leading to increased variability and uncertainty in supply.

In this context, 15-minute trading intervals allow the market to better reflect short-term fluctuations in production. Imbalances can be identified and settled closer to real time, reducing system risk and improving price signals.

However, increased renewable penetration also raises structural challenges. Seasonal variation is particularly important: during winter, reduced solar output and higher demand increase the need for alternative generation sources and system flexibility.

Figure 2.2 presents a real-time snapshot of the Danish electricity system on February 10th, 2026. The figure illustrates both the composition of domestic electricity generation and the net exchange of power with neighbouring countries.

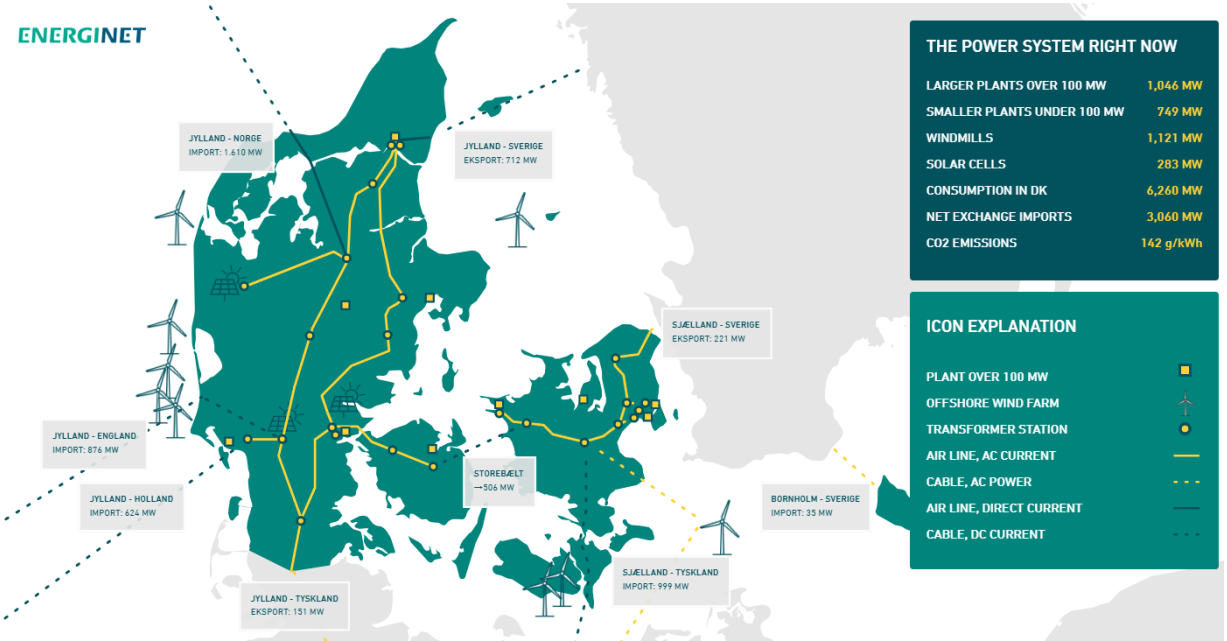


Figure 2.2: Real-time snapshot of Danish electricity consumption, generation mix, and cross-border flows on February 10th, 2026. Source: [6]

Total electricity consumption amounts to 6,260 MW, while domestic production is significantly lower. Wind power constitutes the largest share of generation at 1,121 MW, followed by centralised and decentralised thermal production. Solar generation contributes only marginally, which is consistent with seasonal conditions during winter.

As domestic production is insufficient to meet demand, Denmark operates in a net import position, with total imports of 3,060 MW. Electricity is imported from several neighbour-

ing systems, including Norway, Germany, the Netherlands, and the United Kingdom, while smaller exports occur simultaneously towards Sweden. This reflects the highly interconnected nature of the Danish power system and the role of cross-border trade in maintaining system balance.

From a market perspective, this situation corresponds to a scarcity regime, where domestic supply alone cannot satisfy demand. As described in the merit-order framework, this implies that higher-cost generation units or imports must be activated to clear the market. Consequently, such conditions are typically associated with higher electricity prices and the activation of upward regulation in the balancing market.

Conversely, to Figure 2.2 where we were importing electricity, in Figure 2.3 we can see that on May 4th, 2026 we were exporting electricity due to the Danish high capacity of both solar and wind, producing 2,310 MW and 2,978 MW respectively.

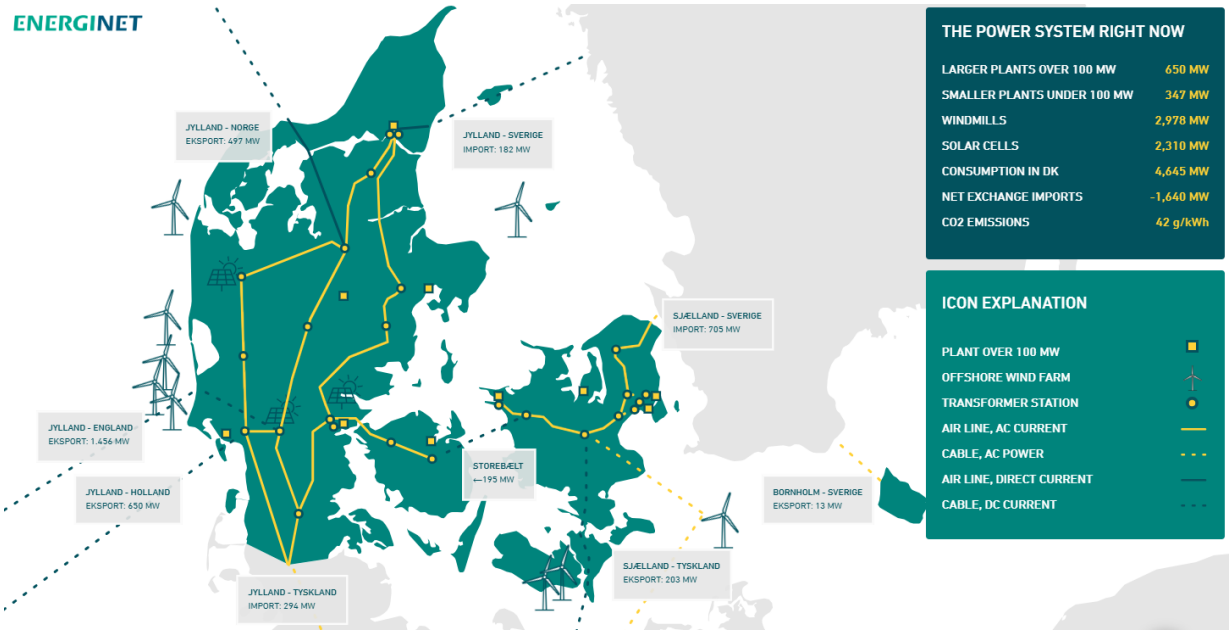


Figure 2.3: Real-time snapshot of Danish electricity consumption, generation mix, and cross-border flows on May 4th, 2026. Source: [6]

Thus, in this case we end up with a excess supply for the Danish demand and end up exporting electricity to other countries like England and Germany.

More broadly, the figures highlights how system balance in Denmark is achieved through a combination of domestic generation and international electricity flows, rather than relying solely on internal production capacity.

Following the energy crisis triggered by the war in Ukraine, Danish households have exhibited a marked shift in electricity consumption behavior. Increased price volatility and heightened public awareness have led many consumers to adopt variable pricing schemes

and actively adjust their electricity usage according to hourly spot prices. A growing share of consumers deliberately shifts flexible consumption, such as electric vehicle charging and household appliances to low-price periods, particularly during nighttime hours [11].

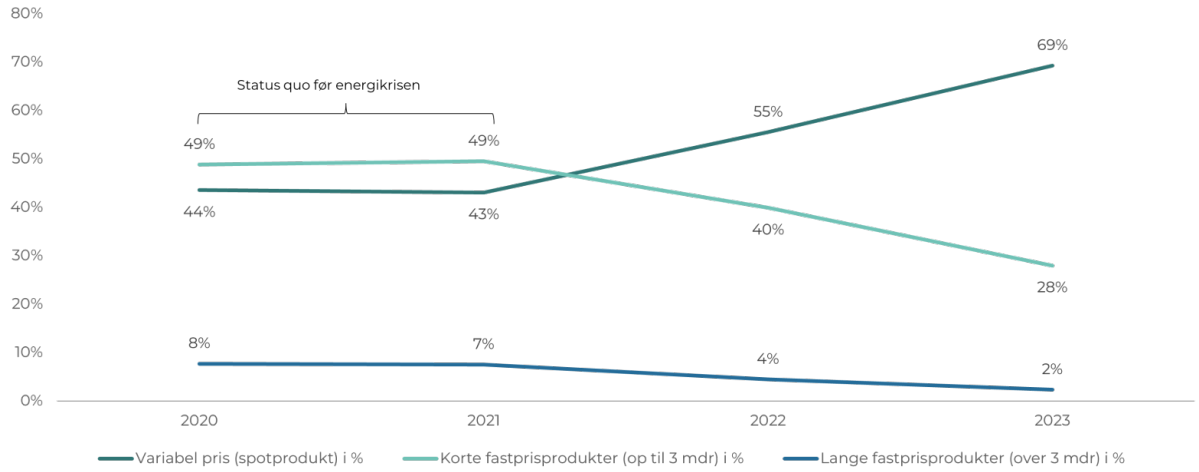


Figure 2.4: Change of payment plan for private consumers 2020-2023. Source: [11]

As can be seen in Figure 2.4 the amount of consumers going on varaiable prices changed from 44% in 2020 to 69% in 2023. Importantly, this shift can be assumed to create a feedback mechanism in the spot market. As more consumers concentrate their demand in low-price periods, the demand profile itself becomes endogenous to price signals.

In equilibrium, large-scale load shifting may flatten price differentials or even relocate peak demand to previously off-peak hours. This highlights a transition from a system with largely exogenous demand to one where consumer behavior actively interacts with price formation.

From a system perspective, large-scale storage solutions and flexible resources are expected to play an increasingly important role in managing these fluctuations. The ability to absorb excess generation and release it during periods of scarcity will be crucial in a future low-carbon energy system, currently the large scale storability is the Northeren fjords where the cheap solar and wind energy during the summer can be used to pump water back up into the fjords that can then be released during the winter months again. Other methods like Power-to-X can be utilized to convert the excess power as an example into splitting water into Hydrogen and Oxygen.

3 | Panel Data

The following chapter is based on [1], [2], [3], [4], and [19].

In this chapter, we present panel data methods that allow us to extract more information from datasets observed across both time and cross-sectional units. Pooled regression, fixed effects, and random effects models are useful in different scenarios, and appropriate tests can help determine which approach is most suitable. Wrapping up the Chapter we introduce to expansions to the modelling, dynamic panel data models, and Factor modelling, which but work with included lagged dependent variables.

Panel data consist of repeated observations on the same cross-sectional units over time, combining the time dimension of time-series data with the cross-sectional dimension of cross sections. Panel data is therefore used for many datasets that are measured across two dimensions. The main advantages of panel data are that they control for cross-sectional heterogeneity, provide greater variability in the data, reduce collinearity among regressors, and yield more efficient estimators. In addition, panel data are well suited for studying dynamic adjustment processes and allow the construction and testing of more complex behavioural models.

The panel data regression for cross-sections $i = 1, \ldots, N$ and time $t = 1, \ldots, T$ is given by

$$Y_{i,t} = \alpha + X_{i,t}\beta + U_{i,t}, \tag{3.1}$$

where α is a scalar, β is a $K \times 1$ parameter vector and $X_{i,t}$ is a $1 \times K$ vector of observations on the explanatory variables for unit i. Given that the same cross-sectional units are observed repeatedly over time, it is generally inappropriate to assume independence of the error terms. As a result, ordinary least squares is no longer efficient and may yield incorrect inference or even inconsistent estimates when regressors are correlated with the composite error term.

The error-components model was developed to model the dependence structure in panel data by decomposing the error into separate shocks, each assumed to be independent of the others. Under the error-components model, the panel data regression can be written as

$$Y_{i,t} = \alpha + X_{i,t}\beta + U_{i,t}, \quad \text{where } U_{i,t} = \mu_i + \nu_{i,t}, \tag{3.2}$$

where μ_i is the individual error, and $\nu_{i,t}$ is the idiosyncratic error. The effect μ_i affects all observations for unit i and is time-invariant while $\nu_{i,t}$ affects only observation (i, t) and is assumed to satisfy the Gauss–Markov assumptions. We assume that μ_i are independent across i, and that the $\nu_{i,t}$ are independent across both i and t and orthogonal to $X_{i,t}$. In addition, we assume that the regressors $X_{i,t}$ are exogenous.

Before introducing fixed and random effects, we can consider the simplest approach: pooling all cross-sectional uni,ts over time and running ordinary least squares, OLS (see Appendix A.2). This effectively ignores any cross-section-specific effects, i.e., assumes $\mu_i = 0$, and is often referred to as the pooled regression model. While this approach can increase efficiency by using the extra variation from pooling, it is generally unrealistic because cross-section heterogeneity is usually present.

To estimate an error-components model the effects μ_i are considered either fixed or random. If we treat the effects μ_i as fixed then they are treated as parameters to be estimated by OLS via dummy variables. If we treat the effects μ_i as random, the covariance matrix of $U_{i,t}$ must be expressed in terms of the variances of μ_i and $\nu_{i,t}$ and feasible generalized least squares FGLS, (see Appendix A.5) can then be applied. Each approach can be appropriate in some contexts but may be unsuitable in others.

Depending on the assumptions about μ_i, several specific panel data models can be considered:

(i) **Pooled regression:** If $\mathbb{E}[\mu_i \mid X_{i,t}] = 0$, OLS provides consistent but potentially inefficient estimates.

(ii) **Fixed effects:** If $\mathbb{E}[\mu_i \mid X_{i,t}] \neq 0$, OLS is biased and inconsistent due to omitted variable effects. A fixed term is added to account for the cross-section–specific heterogeneity.

(iii) **Random effects:** If $\mathbb{E}[\mu_i \mid X_{i,t}] = 0$, a random-effects term can be included to account for the cross-section correlation, allowing efficient estimation.

In the following, we discuss two common panel data models in detail, each corresponding to different assumptions about μ_i.

3.1 Pooled Regression (PR) Model

The baseline pooled regression model assumes no cross-section-specific effects, $\mu_i = 0$, such that the composite error reduces to the idiosyncratic error

$$Y_{i,t} = \alpha + X_{i,t}\beta + U_{i,t}, \quad \text{where } U_{i,t} = \nu_{i,t}. \tag{3.3}$$

The idiosyncratic error are mean zero conditional on the regressors, i.e. $\mathbb{E}[\nu_{i,t} \mid X_{i,t}] = 0$ for all i and t.

Under these assumptions, the OLS estimator is unbiased and consistent. Moreover, if the idiosyncratic errors are uncorrelated across individuals and time periods, i.e. $\mathbb{E}[\nu_{i,t}\nu_{js}] = 0$ for all $i \neq j$ and $t \neq s$, then the OLS estimator is efficient

In practice, however, these assumptions are often unrealistic. Since we observe the same cross-sectional units over time, it is more realistic to assume

$$\mathbb{E}[\nu_{i,t}\nu_{is}] \neq 0 \quad \text{for } t \neq s,$$

so that the idiosyncratic errors are correlated over time for the same cross-sectional unit. In this case, OLS remains unbiased but is no longer efficient. One can use autocorrelation-robust standard errors to correct for this.

Since the cross-section–specific effect μ_i is constant over time, it can be removed completely by subtracting each observation from its previous period. This *First Difference (FD)* transformation eliminates μ_i from the model, allowing consistent estimation of β even when μ_i is correlated with the regressors.

We take the first difference of the error-components model in (3.2)

$$\Delta Y_{i,t} = Y_{i,t} - Y_{i,t-1} = \Delta X_{i,t}\beta + \Delta \nu_{i,t}, \tag{3.4}$$

thus the time-invariant effect μ_i is eliminated. We can therefore estimate β by running OLS on the model

$$\Delta Y = \Delta X\beta + \Delta \nu,$$

which yields the *First Difference (FD) estimator*

$$\hat{\beta}_{FD} = (\Delta X'\Delta X)^{-1}\Delta X'\Delta Y. \tag{3.5}$$

To derive the covariance structure of the idiosyncratic errors, assume that they satisfy

$$\mathbb{E}[\nu_{i,t}] = 0, \qquad \mathrm{Var}(\nu_{i,t}) = \sigma_\nu^2, \qquad \mathrm{Cov}(\nu_{i,t}, \nu_{is}) = 0 \quad \text{for } t \neq s.$$

After differencing, the transformed error term becomes

$$\Delta \nu_{i,t} = \nu_{i,t} - \nu_{i,t-1}.$$

Even if $\nu_{i,t}$ is serially uncorrelated, $\Delta\nu_{i,t}$ is not. In particular,

$$\begin{aligned} \mathrm{Var}(\Delta\nu_{i,t}) &= 2\sigma_\nu^2, \\ \mathrm{Cov}(\Delta\nu_{i,t}, \Delta\nu_{i,t-1}) &= -\sigma_\nu^2, \\ \mathrm{Cov}(\Delta\nu_{i,t}, \Delta\nu_{i,t-k}) &= 0 \quad \text{for } k \geq 2. \end{aligned}$$

Thus, the differenced errors follow an MA(1) process with known structure. Stacking the transformed idiosyncratic errors in vector form, the covariance matrix can be written as

$$\mathrm{Var}(\Delta\nu) = 2\sigma_\nu^2\Sigma,$$

where Σ is the tridiagonal matrix

$$\Sigma = \begin{bmatrix} 1 & -\frac{1}{2} & 0 & \cdots & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & \cdots & 0 \\ 0 & -\frac{1}{2} & 1 & \ldots & \vdots \\ \vdots & \vdots & \vdots & \ddots & -\frac{1}{2} \\ 0 & 0 & \ldots & -\frac{1}{2} & 1 \end{bmatrix}.$$

Since the FD transformation induces serial correlation of order one, OLS standard errors are no longer valid. Consistent inference can instead be obtained using heteroskedasticity and autocorrelation consistent (HAC) standard errors, such as the Newey-West estimator with lag length one, see Appendix A.6.

3.2 Fixed Effects (FE) Model

Unobserved heterogeneity across cross-sectional units has a possibility to be present when looking at large datasets. If such heterogeneity is correlated with the regressors, pooled OLS becomes inconsistent. The Fixed Effects (FE) model addresses this by allowing each cross-sectional unit to have its own intercept.

If $E[\mu_i \mid X_{i,t}] \neq 0$, the individual-specific effects may be correlated with the regressors. In this case, the model (3.2) can be written as

$$Y_{i,t} = \alpha_i + X_{i,t}\beta + \nu_{i,t}, \tag{3.6}$$

where $\alpha_i = \alpha + \mu_i$ represents the individual-specific intercept and the idiosyncratic errors satisfy

$$E[\nu_{i,t}] = 0, \qquad \text{Var}(\nu) = \sigma_\nu^2 I_n.$$

This implies a different constant for each individual. We then construct a matrix of dummy variables D, and thus our model becomes

$$Y_{i,t} = \alpha + X_{i,t}\beta + \sum_{i=1}^{N} \mu_i D_i + \nu_{i,t}, \tag{3.7}$$

here D_i is a dummy variable for the i-th timeseries. We impose a normalization $\sum_{i=1}^{N} \mu_i = 0$ to avoid perfect multicollinearity. The idiosyncratic errors $\nu_{i,t}$ are assumed to be i.i.d random variables with mean 0 and variance σ_ν^2. This can be further rewritten in matrix form

$$Y = X\beta + D\eta + \nu \tag{3.8}$$

where $\eta = (\alpha+\mu_1, \alpha+\mu_2, \ldots, \alpha+\mu_N)'$ is the vector of fixed effects and D is an $n \times m$ matrix of dummy variables, constructed in such a way that the element in the row corresponding to observation i, t, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, and column j, for $j = 1, \ldots, m$, is equal to 1 if $i = j$ and equal to 0 otherwise.

Equation (3.7) can be estimated using OLS. Under the Gauss–Markov assumptions on $\nu_{i,t}$ [25], the OLS estimator is Best linear unbiased estimator (BLUE), however two problems occur.

Firstly the loss of degrees of freedom since we are estimating $N + K$ parameters. Since N additional parameters are effectively estimated, the FE estimator consumes degrees

of freedom. In panels with large N and small T, this may lead to imprecise estimates. However, in applications with sufficiently long time dimension as is the case with this project, this issue is mitigated.

The second issue arises from the large number of dummy variables. When an intercept is included together with all N dummy variables, perfect multicollinearity occurs because the sum of the dummy variables equals one for every observation. This is known as the dummy variable trap. Estimating the model by including dummy variables for each unit is known as the Least Squares Dummy Variable (LSDV) estimator.

An equivalent and computationally more efficient approach to estimating (3.7) is to eliminate the fixed effects through the within transformation. Since μ_i is constant over time for each unit, we can remove them by subtracting the time average from each observation. Formally, define the sample averages

$$\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{i,t}, \quad \bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{i,t}, \quad \bar{\nu}_i = \frac{1}{T}\sum_{t=1}^{T} \nu_{i,t},$$

Subtracting these averages from the original observations yields the within transformation

$$Y_{i,t} - \bar{Y}_i = (X_{i,t} - \bar{X}_i)\beta + (\nu_{i,t} - \bar{\nu}_i). \tag{3.9}$$

In this form the fixed effects μ_i disappear, leaving a model that can be estimated using standard OLS. This not only reduces the number of parameters to estimate, but also improves computational efficiency, especially when the panel has many cross-sectional units.

From a matrix perspective, we can formalize this transformation using the projection matrix

$$M_D = I - D(D'D)^{-1}D',$$

where D is the same matrix as in (3.8). The projection matrix M_D projects any vector onto the space orthogonal to the columns of D, effectively removing the influence of the fixed effects. Applying M_D to both Y and X, gives the same result as the within transformation, but it highlights that the fixed effects estimator is simply an OLS regression in the subspace orthogonal to the individual-specific effects.

The Frisch-Waugh-Lovell (FWL) Theorem, is outside the scope of this project and can be found with proof [4, p. 68-69]. It states that the OLS coefficient on X is a regression of Y on X and D that can equivalently be obtained by first removing the effect of D from both Y and X, and then regressing the resulting residuals on each other. In other words, the partial effect of X can be estimated after partialling out the contribution of D.

Applying this result, the OLS estimator of β in (3.7) can be obtained by regressing $M_D Y$, the residuals from regressing Y on D and on $M_D X$, which contains the residuals from regressing each column of X on D. The resulting fixed–effects estimator can therefore be written as

$$\hat{\beta}_{FE} = (X'M_D X)^{-1}X'M_D Y. \tag{3.10}$$

It follows that the i,t-th element of $M_D X$ equals $X_{i,t} - \bar{X}_i$, i.e. the deviation from the group mean. Since every variable in (3.10) is premultiplied by M_D, the estimator relies solely on variation within each group around its mean. For this reason, the fixed–effects estimator is often referred to as the *within-groups estimator*.

3.3 Random Effects (RE) Model

In the Fixed Effects (FE) model, a separate parameter is estimated for each cross-sectional unit, which quickly increases the number of parameters and reduces degrees of freedom. To address this, the Random Effects (RE) model treats these individual-specific effects as random variables, allowing for more efficient estimation when certain assumptions hold.

The RE model is based on the same error-components structure as the FE model, but here the cross-section-specific effects μ_i, are treated as random variables rather than fixed parameters. Specifically, the RE model can be written as

$$Y_{i,t} = \alpha + X'_{i,t}\beta + \mu_i + \nu_{i,t}, \tag{3.11}$$

$$U_{i,t} = \mu_i + \nu_{i,t}, \tag{3.12}$$

where the components satisfy that the cross-section-specific effects and the idiosyncratic error have zero mean, i.e. $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ and $\nu_{i,t} \sim \text{IID}(0, \sigma_\nu^2)$ respectively. This means that we only need to estimate the variances. Moreover we assume that μ_i is uncorrelated with $\nu_{i,t}$, i.e. $\text{Cov}(\mu_i, \nu_{i,t}) = 0$, and that both components are independent of the regressors, $\mu_i \perp X_{i,t}$ and $\nu_{i,t} \perp X_{i,t}$ for all i and t.

This independence assumption, also called strict exogeneity, may not always hold in practice. However, if the assumption is satisfied, it follows that

$$\mathbb{E}[U_{i,t} \mid X_{i,t}] = \mathbb{E}[\mu_i + \nu_{i,t} \mid X_{i,t}] = 0, \tag{3.13}$$

since μ_i and $\nu_{i,t}$ are independent of $X_{i,t}$. This condition is exactly what guarantees that OLS estimation of (3.11) produces unbiased estimates of β.

OLS estimation of (3.2) is generally not efficient, since the error terms $U_{i,t}$ are not i.i.d.. Under the assumptions for μ_i and $\nu_{i,t}$, the composite error $U_{i,t}$ has variance

$$\text{Var}(U_{i,t}) = \sigma_\mu^2 + \sigma_\nu^2 \tag{3.14}$$

and exhibits serial correlation over time only within the same individual

$$\text{Cov}(U_{i,t}, U_{i,s}) = \sigma_\mu^2 \quad \text{for } t \neq s, \qquad \text{Cov}(U_{i,t}, U_{j,s}) = 0 \quad \text{for } i \neq j. \tag{3.15}$$

Notice that $U_{i,t}$ exhibits correlation over time within the same individual, since all observations for individual i share the same μ_i.

The key requirement for consistency of the RE estimator is that the cross-section-specific effects μ_i is independent of the regressors $X_{i,t}$ and of the idiosyncratic error $\nu_{i,t}$. Under these conditions, OLS applied to (3.11) is consistent, but generally inefficient. Efficiency can be improved using generalized least squares GLS (see Appendix A.4), which accounts for this covariance structure.

If we stack the data by individual, then the covariance matrix of the error term

$$\mathbf{U} = (U_{11}, \ldots, U_{1T}, U_{21}, \ldots, U_{2T}, \ldots, U_{N1}, \cdots, U_{NT})'$$

is a block-diagonal matrix of dimension $NT \times NT$, given by

$$\Omega = I_N \otimes \Sigma = \begin{bmatrix} \Sigma & 0 & \cdots & 0 \\ 0 & \Sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \Sigma \end{bmatrix},$$

where

$$\Sigma = \sigma_\nu^2 I_T + \sigma_\mu^2 \mathbf{1}_T \mathbf{1}_T'$$

is a $T \times T$ matrix with $\sigma_\nu^2 + \sigma_\mu^2$ on the diagonal and σ_μ^2 on the off-diagonal elements and where I_T is the $T \times T$ identity matrix and $\mathbf{1}_T$ is a $T \times 1$ vector of ones. This structure follows from the variance and covariance expressions in (3.14) and (3.15), respectively. This covariance structure implies that the error terms are correlated within individuals over time.

3.3.1 Variance Estimation

The RE model can be efficiently estimated using GLS if the variance components are known

$$\widehat{\beta}_{\text{GLS}} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y,$$

where Ω is the covariance matrix of the composite error term. In practice, these variance components are unknown and must be estimated from the data.

A common approach is to first obtain consistent estimates of the idiosyncratic error and individual error variances using residuals from simpler regressions, such as the fixed effects and pooled OLS models. Substituting these estimates into Ω leads to the FGLS estimator, which is the standard method for estimating the RE model in applied settings.

The within-group FE regression isolates the idiosyncratic error $\nu_{i,t}$ by removing individual-specific effects, while the pooled regression captures the total variance of the composite error. Subtracting the estimated idiosyncratic variance from the total variance yields an estimate of σ_μ^2.

To estimate the variance component of the idiosyncratic errors, σ_ν^2, we use the within-group FE regression, as discussed in Section 3.2. The residuals $\hat{\nu}_{i,t}$ from the regression in (3.8) reflect only the idiosyncratic errors $\nu_{i,t}$, since the within transformation removes the individual-specific effects μ_i. The variance of these residuals provides an estimate of the idiosyncratic variance

$$\hat{\sigma}_\nu^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\nu}_{i,t}^2}{NT - N - K}, \tag{3.16}$$

where K is the number of regressors in the model. This estimator is consistent for σ_ν^2 because the fixed effects transformation effectively eliminates the influence of the individual effects, as seen in (3.9).

To estimate the variance component of the individual-specific effects, σ_μ^2, we use the pooled regression. Specifically, we run an OLS regression on the stacked panel data while ignoring the individual effects μ_i. The residuals from this regression, $\hat{U}_{i,t}$, contain both the individual-specific effects and the idiosyncratic errors

$$s_{\text{PR}}^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{U}_{i,t}^2}{NT - K - 1} \approx \sigma_\mu^2 + \sigma_\nu^2,$$

where K is the number of regressors and $\hat{U}_{i,t}$ are the pooled OLS residuals. This provides an approximate estimate of the total variance of the composite error term (3.12). Once $\hat{\sigma}_\nu^2$ is known from (3.16), the individual-specific variance can be estimated as

$$\hat{\sigma}_\mu^2 = s_{\text{PR}}^2 - \hat{\sigma}_\nu^2, \tag{3.17}$$

where s_{PR}^2 is the variance of the pooled regression residuals.

3.3.2 Between-Groups Estimator

Once we have consistent estimates of the variance components, $\hat{\sigma}_\nu^2$ and $\hat{\sigma}_\mu^2$, we can define the *between-groups (BG) estimator*. The BG estimator plays a key role in RE estimation because FGLS can be expressed as a weighted combination of FE and BG estimators.

This estimator uses only the variation across group means, ignoring within-group differences. Formally, we write

$$P_D y = P_D X\beta + \text{residuals}, \tag{3.18}$$

where $P_D = I - M_D$ is the projection matrix that removes within-group variation. The BG estimator of β is then

$$\hat{\beta}_{BG} = (X' P_D X)^{-1} X' P_D y. \tag{3.19}$$

Even though (3.18) looks like it uses all $n = mT$ observations, only the m group means actually contribute. The BG estimator uses only variation across group means, which means that if the number of groups is smaller than the number of regressors i.e. $m < K$, $X' P_D X$ may not be invertible and the estimator cannot be computed.

3.3.3 FGLS Estimator

Once consistent estimates of the variance components, $\hat{\sigma}_\nu^2$ and $\hat{\sigma}_\mu^2$, have been obtained, they can be substituted into the individual covariance matrix Σ to obtain the estimated covariance matrix

$$\hat{\Sigma} = \hat{\sigma}_\nu^2 I_T + \hat{\sigma}_\mu^2 \mathbf{1}_T \mathbf{1}_T'. \tag{3.20}$$

If $\hat{\sigma}_\mu^2 = 0$, FGLS reduces to OLS on the pooled data. If $\hat{\sigma}_\nu^2 = 0$, FGLS reduces to the FE (within-group) estimator. In general, FGLS is more efficient than either FE or pooled OLS under the random-effects assumptions.

Since the panel observations are stacked by individual, the covariance matrix of the full NT-dimensional disturbance vector is block diagonal, with each block given by Σ. Replacing Σ with the estimated covariance matrix $\hat{\Sigma}$ therefore gives the estimated covariance matrix $\widehat{\Omega}$ for the stacked error vector. This yields the FGLS estimator

$$\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y. \tag{3.21}$$

Using the FGLS estimator $\hat{\beta}_{FGLS}$ we can also express it as a combination of the within-group (FE) and between-group (BG) estimators. Specifically, consider the decomposition

$$\begin{aligned}\widehat{\beta} &= (X'X)^{-1}X'Y = (X'X)^{-1}(X'M_DY + X'P_DY) \\ &= (X'X)^{-1}X'M_DX\hat{\beta}_{FE} + (X'X)^{-1}X'P_DX\hat{\beta}_{BG}\end{aligned}$$

where $\hat{\beta}_{FE}$ is defined as in (3.10) and $\hat{\beta}_{BG}$ is defines as in (3.19), and M_D and P_D are the within-group and between-group projection matrices, respectively.

This decomposition shows that $\hat{\beta}_{\text{FGLS}}$ (and OLS on the pooled model) is a matrix-weighted average of the FE and between-groups (BG) estimators. Here, $\hat{\beta}_{\text{FE}}$ is consistent under the fixed-effects assumptions, while $\hat{\beta}_{\text{BG}}$ requires the additional random-effects assumptions for consistency. Consequently, the FGLS estimator is consistent only if the BG estimator is consistent, reflecting the standard restriction of the RE model. The weights in this average depend on the relative sizes of the variance components, effectively balancing within-group (FE) and between-group (BG) information according to how much of the total variation is due to individual-specific effects versus idiosyncratic noise.

3.4 Model Selection

The pooled regression, fixed effects, and random effects models provide alternative ways of modelling panel data, each based on different assumptions about unobserved individual-specific effects and their relationship with the regressors. Consequently, the models are estimated using different methods: pooled regression by OLS, fixed effects by OLS on within-transformed data, and random effects by GLS or FGLS.

To determine which specification is most appropriate for the data, a sequence of specification tests is applied. First, an F-test compares the fixed effects and pooled regression models to test for the presence of individual effects. Second, the Breusch–Pagan Lagrange Multiplier test evaluates whether random effects are preferred to pooled regression. Finally, the Hausman test is used to choose between the fixed effects and random effects models by testing whether the individual effects are correlated with the regressors. Together, these tests provide a systematic procedure for selecting the appropriate panel data model.

3.4.1 Fixed Effects vs Pooled Regression

To test whether individual-specific effects are present, we examine the joint significance of the individual dummy variables in the fixed effects model. Formally, the null hypothesis is

$$\mathcal{H}_0 : \mu_1 = \mu_2 = \cdots = \mu_{N-1} = 0,$$

which implies that all individual effects are equal and the pooled regression model is sufficient. Thus, we perform a F-test of the unrestricted model given by FE against the restricted model given by pooled regression. The statistic is given by:

$$F = \frac{(R^2_{\text{FE}} - R^2_{\text{PR}})/(N-1)}{(1 - R^2_{\text{FE}})(NT - N - K)} \tag{3.22}$$

which follows an F-distribution. If the null hypothesis is rejected, the presence of unobserved individual heterogeneity suggests that a random effects model provides a better specification than pooled regression.

3.4.2 Random Effects vs Pooled Regression

To determine whether a random effects specification is preferable to pooled regression, we apply the Breusch–Pagan Lagrange Multiplier (LM) test. The null hypothesis is given by $\mathcal{H}_0 : \sigma^2_\mu = 0$, which implies that there are no individual-specific effects and that the pooled regression model is sufficient.

The LM test statistic is defined as

$$LM = \frac{NT}{2(T-1)} \left[\frac{\sum_{i=1}^{N} \left(\sum_{t=1}^{T} \hat{U}_{i,t} \right)^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{U}^2_{i,t}} - 1 \right]^2, \tag{3.23}$$

where $\hat{U}_{i,t}$ denotes the residuals obtained from the pooled OLS estimation. Under the null hypothesis, the test statistic follows a χ^2 distribution with one degree of freedom.

3.4.3 Fixed Effects vs Random Effects

A key assumption of the RE model is that the unobserved individual effects, μ_i, are uncorrelated with the regressors $X_{i,t}$

$$\mathbb{E}[\mu_i \mid X_{i,t}] = 0$$

If this assumption holds, the RE (GLS) estimator is efficient, consistent, and asymptotically optimal. However, if the individual effects are correlated with any explanatory variables, i.e. $\mathbb{E}[\mu_i \mid X_{i,t}] \neq 0$, the RE estimator becomes inconsistent and biased.

The Fixed Effects (FE) estimator, on the other hand, removes the individual effects by demeaning or differencing, and is consistent regardless of whether μ_i correlates with $X_{i,t}$.

Hausman proposed a formal test to decide between RE and FE by comparing the two estimators. Under the null hypothesis that the RE assumptions are valid $\mathcal{H}_0 : E[\mu_i \mid X_{i,t}] = 0$, both FE and RE are consistent, but RE is more efficient. Under the alternative, $\mathcal{H}_1 : E[\mu_i \mid X_{i,t}] \neq 0$, FE remains consistent while RE becomes inconsistent. The Hausman test statistic is defined as

$$H = (\widehat{\beta}_{\text{FE}} - \widehat{\beta}_{\text{RE}})' \left[\text{Var}(\widehat{\beta}_{\text{FE}}) - \text{Var}(\widehat{\beta}_{\text{RE}})\right]^{-1} (\widehat{\beta}_{\text{FE}} - \widehat{\beta}_{\text{RE}}), \tag{3.24}$$

which asymptotically follows a χ^2 distribution with degrees of freedom equal to the number of regressors.

A large value of H leads us to reject the null hypothesis $\mathcal{H}_0$, that the individual effects are uncorrelated with the regressors, indicating that the RE estimator is inconsistent and that the FE model should be preferred. Conversely, a small value of H means we fail to reject the null hypothesis$\mathcal{H}_0$, suggesting that the RE model is consistent and appropriate.

3.5 Dynamic Panel Models

A dynamic panel model is a panel data model that includes lagged values of the dependent variable as regressors. This extends the standard panel regression framework by allowing the current value of the dependent variable to depend on its own past realizations. Such a specification is particularly useful when adjustment processes are gradual over time.

We state the dynamic ECM as in (3.2), augmented with lagged dependent variables,

$$Y_{i,t} = \alpha + \delta Y_{i,t-1} + X'_{i,t}\beta + \mu_i + \nu_{i,t}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{3.25}$$

where $Y_{i,t}$ denotes the dependent variable for unit i at time t. The coefficient δ measures the persistence of the dependent variable over time. The inclusion of a lagged dependent variable introduces a problem, since $Y_{i,t}$ depends on individual error μ_i, the lagged variable $Y_{i,t-1}$ will also contain μ_i. As a consequence, $Y_{i,t-1}$ is correlated with the composite error

term $U_{i,t} = \mu_i + \nu_{i,t}$, due to the presence of μ_i in both terms, i.e. $\text{Cov}(Y_{i,t-1}, U_{i,t}) \neq 0$. This implies that OLS estimation is biased and inconsistent even when the idiosyncratic errors are serially uncorrelated. The assumptions on the errors as described in the previous sections sustain, i.e. $\mu_i \sim \text{IID}(0, \sigma_\mu^2), \quad \nu_{i,t} \sim \text{IID}(0, \sigma_\nu^2)$.

To eliminate the unobserved individual effects, the within transformation used in FE estimation is applied. However, even after removing μ_i, the transformed lagged dependent variable, $\tilde{Y}_{i,t-1}$, becomes correlated with the transformed error term, $\tilde{\nu}_{i,t}$, even if the original errors $\nu_{i,t}$ are not serially correlated. This arises because the within transformation induces a dependence between $\tilde{Y}_{i,t-1}$ and $\tilde{\nu}_{i,t}$. Nickell shows that the resulting FE estimator is biased in panels with a finite time dimension. The bias decreases at rate $O(1/T)$ and vanishes when the number of time periods is sufficiently large.

To address the endogeneity problem in panels with small T, alternative estimators have been proposed. The source of the problem is that the lagged dependent variable is correlated with the error term, i.e. $\mathbb{E}[Y_{i,t-1}U_{i,t}] \neq 0$. Anderson and Hsiao suggest eliminating the individual effects by first difference modelling as in (3.4)

$$\Delta Y_{i,t} = Y_{i,t} - Y_{i,t-1} = \delta(Y_{i,t-1} - Y_{i,t-2}) + (\nu_{i,t} - \nu_{i,t-1}). \tag{3.26}$$

This transformation removes the individual error μ_i, but the differenced lagged dependent variable $\Delta Y_{i,t-1}$ remains correlated with the differenced error term $\Delta\nu_{i,t}$. Anderson and Hsiao therefore propose using deeper lags of the dependent variable as instruments. For instance, $Y_{i,t-1}$ can serve as an instrument for $\Delta Y_{i,t-3}$, provided that the idiosyncratic errors are not serially correlated.

$$\Delta Y_{i,3} = Y_{i,3} - Y_{i,2} = \delta(Y_{i,2} - Y_{i,1}) + (\nu_{i,3} - \nu_{i,2})$$

Here $Y_{i,1}$ is a valid instrument since it is directly correlated to $Y_{i,3}$ and $Y_{i,2}$ but uncorrelated to either $\nu_{i,3}$ and $\nu_{i,2}$ assuming no serially correlation, we further expand this for $t = 4, \ldots, T-2$

$$\Delta Y_{i,4} = Y_{i,4} - Y_{i,3} = \delta(Y_{i,3} - Y_{i,2}) + (\nu_{i,4} - \nu_{i,3})$$

Where we now see that $Y_{i,1}$ and $Y_{i,2}$ is a valid instrument, Doing this recursively for all $T-2$ we end up with an instrumental matrix,

$$W_i = \begin{bmatrix} [Y_{i,1}] & & & 0 \\ & [Y_{i,1}, Y_{i,2}] & & \\ & & \ddots & \\ 0 & & & [Y_{i,1}, \ldots, Y_{i,T-2}] \end{bmatrix}$$

Matrix of instruments is $W = [W_1, W_2, \ldots, W_N]$, where it is apparent that the following moment condition holds

$$\mathbb{E}[Y_{i,t-s}\, \Delta\nu_{i,t}] = 0, \qquad s \geq 2,\ t = 3, \ldots, T-2 \tag{3.27}$$

Although this does not take into account the differenced error term in (3.26), since in fact

$$\mathbb{E}[\Delta\nu_i \Delta\nu_i'] = \sigma_\nu^2 G \tag{3.28}$$

is apparent, where

$$\Delta\nu_i = (\nu_{i,3} - \nu_{i,2},\, \nu_{i,4} - \nu_{i,3},\, \ldots,\, \nu_{i,T} - \nu_{i,T-1}),$$

and

$$G = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix}.$$

Premultiplying ΔY with W' then gives

$$W'\Delta Y = W'(\Delta Y_{-1})\delta + W'\Delta\nu$$

then beforming GLS we get the one-step consistent estimator

$$\begin{aligned} \widehat{\delta}_1 = {} & [(\Delta Y_{t-1})'W(W'(I_N \otimes G)W)^{-1}W'(\Delta Y_{-1})]^{-1} \\ & \times [(\Delta Y_{-1})'W(W'(I_N \otimes G)W)^{-1}W'(\Delta Y)] \end{aligned} \tag{3.29}$$

The GMM estimator of δ_1 à la Hansen (1982) for $N \to \infty$ and T fixed using only the above moment restrictions can be further expanded and give the same expression as in (3.29) except that

$$W'(I_N \otimes G)W = \sum_{i=1}^{N} W_i' G W_i$$

is replaced by

$$V_N = \sum_{i=1}^{N} W_i'(\Delta\nu_i)(\Delta\nu_i)'W_i.$$

This GMM estimator requires no knowledge concerning the initial conditions or the distributions of ν_i and μ_i. Here, $\Delta\nu$ is replaced by differenced errors obtained from the preliminary consistent estimator $\hat{\delta}_1$. The resulting estimator is the two-step Arellano and Bond (1991) GMM estimator,

$$\hat{\delta}_2 = \left[(\Delta Y_{-1})'W V_N^{-1} W'(\Delta Y_{-1})\right]^{-1} (\Delta Y_{-1})'W V_N^{-1} W'(\Delta Y).$$

A consistent estimate of the asymptotic $\mathrm{var}(\hat{\delta}_2)$ is given by the first term,

$$\widehat{\mathrm{var}}(\hat{\delta}_2) = \left[(\Delta Y_{-1})'W \hat{V}_N^{-1} W'(\Delta Y_{-1})\right]^{-1}.$$

In summary, dynamic panel data models are subject to the Nickell bias when estimated using fixed effects in panels with a small time dimension. The Arellano and Bond estimator addresses this issue by exploiting internal instruments derived from lagged values of the dependent variable, yielding consistent estimates when $N \to \infty$ and T is fixed.

In the present application, however, the time dimension is relatively large, while the cross-sectional dimension is moderate. Since the Nickell bias decreases at rate $O(1/T)$, the bias of the fixed effects estimator is expected to be small. As noted by Baltagi, this suggests that the fixed effects estimator may still provide reliable estimates in such settings. Consequently, both approaches will be considered, with the choice of estimator ultimately guided by the empirical characteristics of the data.

3.6 Factor-Augmented Dynamic Panel Models

The following section is based on [1], [14], [19], and [20].

To account for the vast power market factors that can affect pricing, we consider a dynamic panel data model structured with factoring, as well as a model with an unobserved multi-factor error structure, heavily following the framework of [14]. The model allows for both autoregressive dynamics and latent common shocks that induce cross-sectional dependence across panels.

We again let $i = 1, \ldots, N$ index the cross-sectional units, corresponding to the hourly or quarter-hourly delivery periods, and let $t = 1, \ldots, T$ index time. The factor-augmented dynamic panel model is specified as

$$Y_{i,t} = \delta Y_{i,t-1} + X'_{i,t}\beta + \lambda'_i f_t + \varepsilon_{i,t},$$

where $Y_{i,t}$ denotes the spot price for delivery period i at time t, $Y_{i,t-1}$ captures autoregressive persistence in spot prices, $X_{i,t}$ is a vector of explanatory variables, β is the associated parameter vector, f_t is an $L \times 1$ vector of unobserved common factors, λ_i is the corresponding vector of factor loadings, and $\varepsilon_{i,t}$ denotes the idiosyncratic error term.

The term $\lambda'_i f_t$ captures latent common shocks that simultaneously influence all i, t. The factor loadings λ_i allow the impact of these shocks to differ across the panel, thereby accommodating heterogeneous responses across the price curve.

The multi factor error structure provides a flexible way of capturing unobserved common variation without explicitly modelling each latent driver individually. Rather than attempting to identify each latent component directly, the factor structure allows these common influences to enter through the error term.

The latent filtration $\mathcal{F}$ contains all common factors F, while still permitting the variables $\{(X_{i,t}, \lambda_i)\}_{t=1}^{T}$ to depend on additional aggregate shocks that are not necessarily generated by a linear factor structure. This introduces further cross-sectional dependence between

units. As an illustration, one may specify

$$X_{i,t} = b\left(\psi_i, g_t, \zeta_{i,t}\right),$$

where $b(\cdot)$ may represent either a linear or nonlinear mapping in its arguments.

Assumption 3.6.1.
The data generating process is assumed to satisfy the following conditions for all $i = 1, \ldots, N$ and $t = 1, \ldots, T$:

(i) $(X_i, \varepsilon_i, \lambda_i)$ are independently and identically distributed across i, conditional on the information set $\mathcal{F}$.

(ii) Each time-varying element $(p_{i,t}^{(1)}, p_{i,t}^{(2)}, \ldots, p_{i,t}^{(K+1)})$ satisfies

$$E\left[\left|p_{i,t}^{(\cdot)}\right|^{4+\gamma}\right] < \infty$$

for some $\gamma > 0$.

(iii) Each factor loading element $\lambda_i^{(\ell)}$ in $\lambda_i = \left(\lambda_i^{(1)}, \ldots, \lambda_i^{(L)}\right)'$ satisfies

$$E\left[\left|\lambda_i^{(\ell)}\right|^{4+\gamma}\right] < \infty.$$

(iv) The idiosyncratic error satisfies

$$E_{\mathcal{F}}\left[\varepsilon_{i,t} \mid \lambda_i, X_{i,1:\tau_1(t)}^{(1)}, \ldots, X_{i,1:\tau_K(t)}^{(K)}\right] = 0,$$

for all t, where $\tau_1(t), \ldots, \tau_K(t)$ are positive integers.

To handle the endogeneity induced by the unobserved common factors, the model introduces a set of instrumental variables. Let $z_i \in \mathbb{R}^{d\times 1}$ denote the vector of instruments available for cross-sectional unit i. These instruments may consist of both internal instruments, constructed from lagged values of the endogenous variables, and external instruments satisfying the required orthogonality assumptions.

As an example a typical instrument vector can be picked as

$$z_i = \begin{pmatrix} Y_{i,t-2} \\ Y_{i,t-3} \\ X_{i,t-2} \\ W_{i,t} \end{pmatrix},$$

where lagged endogenous variables such as $Y_{i,t-2}$ and $X_{i,t-2}$ act as internal instruments,

while $W_{i,t}$ may represent an external instrument assumed to be exogenous. The key requirement is that the instruments are correlated with the endogenous regressors, but uncorrelated with the error term:

$$E(z_i u_i) = 0.$$

Since not all instruments are necessarily valid at every time period, i.e. correlated with the errors, a selection matrix is introduced. Define

$$S = \text{diag}(S_1, \ldots, S_T),$$

where S_t is a $[\zeta_t \times d]$ block-diagonal selection matrix. The role of S_t is to select the valid instruments at time t from the full instrument vector z_i, by picking ζ_t valid instruments. Each row of S_t contains zeros and a single one, such that multiplication by S_t extracts the relevant instruments.

For example, if

$$z_i = \begin{pmatrix} Y_{i,t-2} \\ Y_{i,t-3} \\ X_{i,t-2} \end{pmatrix},$$

and only the first and third instruments are valid at time t, then

$$S_t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

which implies

$$S_t z_i = \begin{pmatrix} Y_{i,t-2} \\ X_{i,t-2} \end{pmatrix}.$$

Hence, the matrix S acts as a bookkeeping device ensuring that only admissible instruments enter the moment conditions at each time period.

Using the selected instruments, we construct the population moment conditions

$$E_F \left[Z_i'(Y_i - X_i\beta_0 - F\lambda_i) \right] = 0, \tag{3.30}$$

where

$$Z_i' \equiv S(I_T \otimes z_i).$$

The matrix Z_i' therefore stacks the valid instruments across all time periods according to the selection structure imposed by S.

Because the factor loadings λ_i are unobserved and potentially correlated with the instruments, the orthogonality condition cannot be imposed directly. Instead, the covariance

between instruments and factor loadings is absorbed into the unknown matrix $G_{z,\lambda} = E_F(z_i\lambda_i')$, which transforms (3.30) into the new moment condition

$$S\left(\text{vec}\left(E_F\left[z_i(Y_i - X_i\beta_0)'\right] - G_{z,\lambda}F'\right)\right) = 0_\zeta. \tag{3.31}$$

This condition forms the core of the estimation strategy. Intuitively, the equation states that once the contribution from the latent factor structure has been removed through the term $G_{z,\lambda}F'$, the remaining component must be orthogonal to the instrument set. The matrix S ensures that only the valid moment conditions are retained at each time period.

The dimension of the resulting moment vector is

$$\zeta = \sum_{t=1}^{T} s_t, \tag{3.32}$$

corresponding to the total number of valid moment restrictions across all time periods.

3.6.1 Construction of Factor Proxies

The central problem in the estimation of the factors lies in replacing the unobserved latent factors with observable factor proxies. Rather than estimating the factor structure directly through nonlinear optimization, the methodology approximates the latent factor space using observable variables that are exposed to the same common shocks as the dependent variable.

Let F_e denote a $T \times L_e$ dimensional matrix satisfying

$$F \in \text{Col}(F_e),$$

where F denotes the true latent factor matrix and $\text{Col}(F_e)$ denotes the column space spanned by the observable factor proxies. Furthermore, let $\hat{F}_e$ denote an estimator of the factor space generated by F_e.

The key requirement imposed on the factor proxies is summarized in the following assumption.

Assumption 3.6.2.
The estimated factor proxies are assumed asymptotically linear such that

$$\sqrt{N}\left(\hat{F}_e - F_e A_N\right) = \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Psi_i\right) + o_p(1), \tag{3.33}$$

where $\Psi_i = (I_T \otimes \psi_i')B_N$, A_N is a rotation matrix, Ψ_i is a $[q \times 1]$ vector, and B_N is a $[L_e \times T_q]$ selection matrix. Furthermore:

(i) Ψ_i is independently and identically distributed across i, conditional on $\mathcal{F}$, with

$$E_{\mathcal{F}}[\Psi_i] = 0,$$

and

$$E\left[|\psi_i^{(j)}|^{4+\gamma}\right] < \infty$$

for some $\gamma > 0$.

(ii) $A_N \xrightarrow{p} A$ and $B_N \xrightarrow{p} B$, where A and B are $\mathcal{F}$-measurable matrices.

(iii) For all N,

$$\text{rank}(A_N) = L_e.$$

(iv) The latent factor matrix satisfies

$$F \in \text{Col}(F_e),$$

with

$$\text{rank}(F_e) = L_e.$$

Assumption 3.6.2 constitutes the fundamental identification condition underlying the entire estimation framework. The assumption states that the observable factor proxies must asymptotically recover the same factor space as the true latent factors. In other words, although the factors themselves are unobserved, the observable proxy variables must contain sufficient information about the common shocks driving the system.

The asymptotic linearity condition in (3.33) implies that the factor proxy estimator admits a standard central limit representation. Consequently, the estimated factor space converges toward the true factor space at rate $\sqrt{N}$, which allows conventional asymptotic inference to be applied within the subsequent GMM estimation procedure.

The rank condition imposed on A_N ensures that the observable proxies span the full latent factor space. If this condition fails, the observable proxies are unable to recover all common shocks affecting the dependent variable, leading to identification failure and inconsistent parameter estimates.

Moment Conditions and GMM Estimation

We let β_0 as the true value of β, and $g_0 \equiv \text{vec}(G_{z,\lambda}(A_N^{-1})')$ being the true value of g with dimensions $[dL_e \times 1]$. We then define $G_{z,\lambda}$ with respect to the vector factor loading $\lambda_{i,e}$ with dimensions $[L_e \times 1]$, where $\lambda_{i,e} \equiv R\lambda_i$ where R is a selection matrix given by $F = F_e R$.

Let θ denote the full parameter vector of the model, i.e. $\theta = (\beta', g')$. Using the estimated factor proxies $\hat{F}_e$, the sample moment conditions, closely resembling (3.31) with

$S(\text{vec}(G_{z,\lambda}F')) = S(F \otimes I_d)g_{z,\lambda}$ and $g_{z,\lambda} = \text{vec}(G_{z,\lambda}) = \text{vec}(E_F(z_i\lambda_i'))$, are constructed as

$$\bar{\mu}_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} Z_i'\left(Y_i - X_i\beta\right) - S\left(\hat{F}_e \otimes I_d\right)g, \tag{3.34}$$

$$\bar{\mu}_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} Z_i'\left(Y_i - F_X\beta\right) - S\left(\hat{F}_e \otimes I_d\right)g, \tag{3.35}$$

where Z_i denotes the matrix of valid instruments and g collects the nuisance parameters associated with the factor loadings.

The corresponding GMM estimator is defined as

$$\hat{\theta} = \arg\min_{\theta} \bar{\mu}_N(\theta)'\Omega_N\bar{\mu}_N(\theta), \tag{3.36}$$

where Ω_N denotes a positive definite weighting matrix.

An important feature of the estimation procedure is that the moment conditions remain linear in the unknown parameters. Consequently, the estimator admits a closed-form solution and avoids the numerical difficulties often associated with nonlinear factor estimators.

The asymptotic distribution of the proposed estimator is determined by the leading term in equation (3.35). In particular, the moment function may be expanded as

$$\bar{\mu}_N(\theta) = \frac{1}{N}\sum_{i=1}^{N}\mu_i(\theta) + o_p(N^{-1/2}),$$

where

$$\mu_i(\theta) = Z_i'\left(Y_i - X_i\beta\right) - S\left((F_eA_N + \Psi_i) \otimes I_d\right)g.$$

To ensure identification and asymptotic validity of the estimator, the following regularity conditions are imposed.

Assumption 3.6.3.
The following conditions are assumed to hold:

(i) The Jacobian matrix associated with the structural parameters satisfies

$$\Gamma_\beta = \text{plim}_{N\to\infty}\left(-\frac{\partial\bar{\mu}_N(\theta)}{\partial\beta'}\right) = E_{\mathcal{F}}\left[Z_i'X_i\right],$$

and has full column rank almost surely.

(ii) The Jacobian matrix associated with the factor component satisfies

$$\Gamma_g = \text{plim}_{N\to\infty}\left(-\frac{\partial \bar{\mu}_N(\theta)}{\partial g'}\right) = S(F_e A \otimes I_d),$$

and has full column rank almost surely.

(iii) The full Jacobian matrix

$$\Gamma = (\Gamma_\beta, \Gamma_g)$$

has full column rank almost surely.

(iv) The covariance matrix

$$\Delta = \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mu_i(\theta_0)\mu_i(\theta_0)'$$

is $\mathcal{F}$-measurable and positive definite almost surely.

Assumption 3.6.3 imposes the regularity conditions required for consistency and asymptotic normality of the estimator. In particular, the first condition ensures that the instruments are sufficiently correlated with the regressors, corresponding to the standard identification requirement in instrumental variable estimation.

The second condition guarantees that the estimated factor proxies span the relevant factor space. Together with Assumption 3.6.2, this ensures that the latent common shocks are consistently approximated by the observable proxies.

The third condition requires that the structural parameters and the factor component are jointly identifiable. Economically, this excludes situations where the explanatory variables are perfectly collinear with the latent factor structure.

Finally, the fourth condition guarantees that the asymptotic covariance matrix is well-defined, thereby ensuring valid asymptotic inference.

The following theorem summarizes the asymptotic properties of the proposed estimator.

Theorem 3.6.4.
Suppose that Assumptions 3.6.1, 3.6.2 and 3.6.3 hold. Then, as $N \to \infty$,

$$\sqrt{N}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \left[(\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega\right]\Delta^{1/2}\pi,$$

where $\pi \sim N(0, I)$ and Ω denotes the probability limit of the weighting matrix Ω_N.

Proof. The proof of Theorem 3.6.4 will not be provided, but can be found in [13][p 33-34.]. □

Theorem 3.6.4 establishes that the estimator is consistent and asymptotically mixed normal. Consequently, conventional asymptotic inference may be conducted using the estimated covariance matrix of the estimator.

To obtain feasible inference, the covariance matrix Δ is estimated using the sample analogue

$$\hat{\Delta} = \frac{1}{N}\sum_{i=1}^{N} \hat{\mu}_i(\hat{\theta})\hat{\mu}_i(\hat{\theta})',$$

where

$$\hat{\mu}_i(\theta) = Z_i'(y_i - X_i\beta) - S\left((\hat{F}_e + \hat{\Psi}_i) \otimes I_d\right) g.$$

Finally, the efficient two-step GMM estimator is obtained by setting

$$\Omega_N = \hat{\Delta}^{-1}.$$

4 | Application

The following chapter is based on the preceding theory and [18], [23], [24], [12], and [22].

This chapter applies the panel data methods introduced in the previous chapters to spot price data observed across both time and cross-sectional units. We begin with an introduction to the dataset and its structure, considering both hourly and quarterly panel setups. Next, pooled regression, fixed effects, and random effects models are estimated using different transformations of the spot price series in order to evaluate their statistical properties and residual behaviour.

Based on these findings, we extend the analysis to dynamic panel data models that incorporate lagged dependent variables, allowing the models to capture the persistence present in spot prices. The dynamic specifications are implemented for both the hourly and quarterly panel structures and are subsequently compared in terms of model performance and forecasting ability.

Finally, the same modelling framework is extended through factor modelling, where latent common factors are introduced to capture shared variation across the cross-sectional units. The forecasting performance of the factor models is then compared with the previously considered panel data approaches.

4.1 Data Introduction

This section introduces the data used in the models. We first describe the data used for the hourly panel model, followed by the data used for the quarterly panel model. Both panel datasets include spot prices, consumption, and renewable energy data obtained from Energinet.

4.1.1 Hourly Panel

The hourly panel dataset is constructed using *Elspot Prices* (hourly data) and *Day-Ahead Prices* (quarterly data) for spot prices. For the consumption dataset, we use *Consumption per Consumer Category, Region and Hour*. For the renewable energy data, we use *Electricity Production and Exchange 5 min Realtime* to obtain offshore wind, onshore wind, and solar power, filtered for DK1. All datasets are obtained from Energinet.

From the spot price dataset we include the spot price in DKK, from the consumption data, we include total consumption measured in kWh, and from the renewable dataset we include offshore wind, onshore wind, and solar power production measured in MW.

Since the data from October 1st, 2025 and onward is recorded at 15-minute intervals, we retain only the first observation of each hour and treat it as representative of the full hour. The consumption data is recorded at an hourly frequency, whereas the renewable energy data is available at a 5-minute frequency and is aggregated to an hourly level to ensure consistency across datasets.

Because of daylight saving time (DST), some dates will have either $i = 25$ or $i = 23$ hourly observations instead of the standard 24. To ensure a consistent panel structure, we first identify the affected days and then reconstruct a balanced 24-hour structure for each date. Missing values are imputed using the previous available hourly observation within the same day.

This spot price dataset spans the period from the 18th of May, 2020 to the 5th of May, 2026. However, in the empirical analysis we restrict attention to the period from January 1st, 2023 onward. The reason for this is the extreme volatility observed between October 2021 and January 2023, where spot prices exhibit frequent and large spikes. This period coincides with the energy market disruptions following the Russian invasion of Ukraine, which led to substantial instability in spot prices across Europe.

In Figure 4.1, the spot prices over time are shown for four selected hours of the day. The figure illustrates pronounced price spikes and high volatility during the 2021–2022 energy crisis, followed by a more stable price pattern from 2023 onward. By focusing on the post-2023 period, we analyse a regime that is less dominated by extreme shocks and therefore more suitable for modelling and eventual forecasting purposes. This is seen in Figure 4.2.

In the renewable energy dataset, a technical issue also related to the transition to DST caused erroneous values to occur between March 29th 2026 at 03:00 and 04:00. To correct this, we remove the affected observation and replace it with the corresponding value from the previous day at the same hour.

Another technical issue occurs on November 22th 2025, where observations are entirely missing from the renewable energy dataset. To address this, we reconstruct the missing day by copying the observations from the previous day (November 21st 2025) and shifting them forward by one day. This ensures that the temporal structure is preserved while providing a reasonable approximation of the missing values.

The final dataset spans the period from the January 1st, 2021 to May 5th, 2026, reflecting the fact that consumption data is published with a delay of approximately 9-15 days. The dataset is then converted into a panel data structure where i denotes the hour (cross-sectional dimension) and t denotes the day (time dimension). This resulting dataset constitutes the final panel used in the analysis, in consistence with the theory.

4.1.2 Quarterly Panel

The quarterly panel dataset is constructed in the same fashion as the hourly using *Day-Ahead Prices* (quarterly data) for spot prices. For the consumption dataset, we use *Con-*

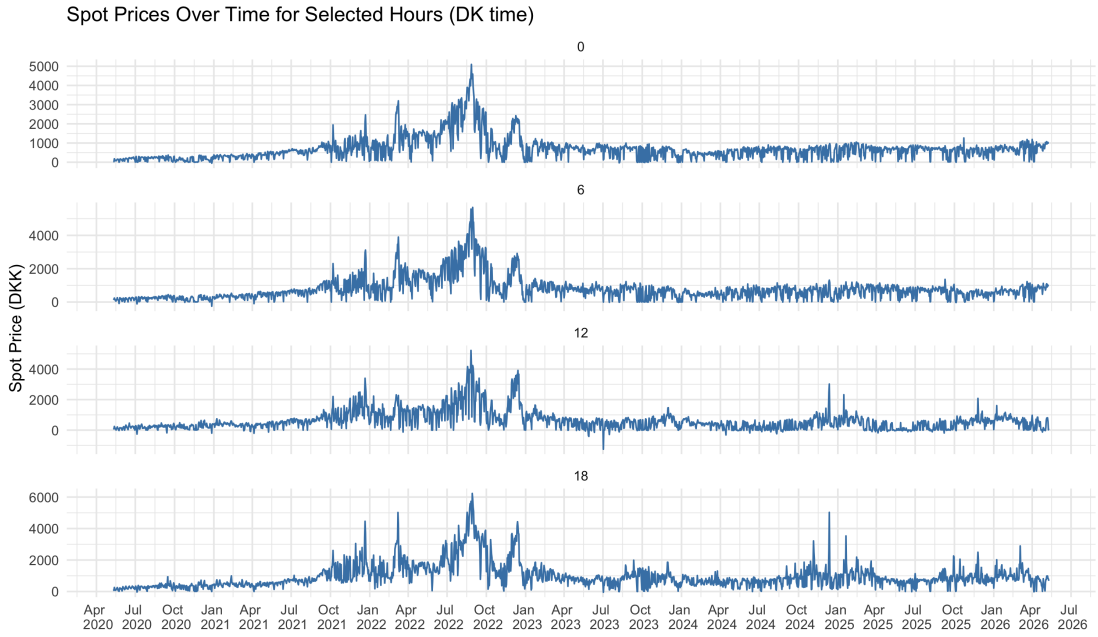


Figure 4.1: Spot prices over time for four selected hours using data from May 2020 to the present. From top to bottom: hour 0, 6, 12, and 18.

sumption per Consumer Category, Region and Hour. For the renewable energy data, we use *Electricity Production and Exchange 5 min Realtime* to obtain offshore wind, onshore wind, and solar power, filtered for DK1. All datasets are obtained from Energinet.

Since the data before October the 1st 2025 is recorded hourly, we only use *Day-Ahead Prices* (quarterly data) for spot prices. The consumption data is originally recorded at an hourly frequency. To align it with the higher-frequency datasets, we convert it to 15-minute intervals by expanding each hourly observation into four equally spaced sub-intervals. Each hourly value is duplicated and assigned to the corresponding 15-minute timestamps within the hour.

The renewable energy data is originally recorded at a 5-minute frequency. To ensure consistency with the other datasets, it is aggregated to 15-minute intervals. This is done by grouping observations into quarters of an hour and computing the average production within each interval for offshore wind, onshore wind, and solar power.

The resulting dataset is structured into quarterly intervals (0–95 per day) thus $i \in (0; 95)$. To maintain a balanced panel, missing quarters arising from DST adjustments are added to ensure that each day contains a full set of observations. Any missing values introduced by this procedure are imputed using the last available observation within the same day.

Just like for hourly dataset a technical issue occurred on November 22nd, 2025, where observations are entirely missing from the renewable energy dataset. To address this, we

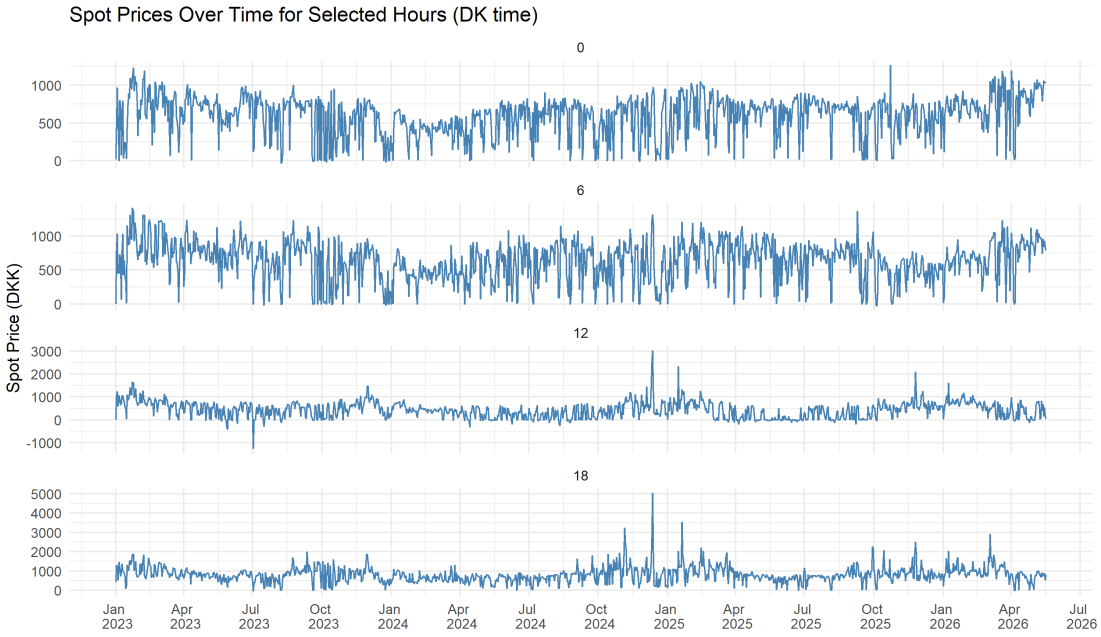


Figure 4.2: Spot prices over time for four selected hours using data from 1 January 2023 to the present. From top to bottom: hour 0, 6, 12, and 18.

followed the same procedure as for hourly. The final dataset spans the period from October 1st 2025 to May 17th 2026.

4.2 Visual Representation of the Spot Price

In this section, we analyse how average spot prices evolve over the course of the day across different weekdays and months. By examining Spot Price market profiles, we aim to identify systematic patterns, such as which days of the week or periods of the year tend to exhibit lower or higher price levels.

Figures 4.3 and 4.4 present these patterns using two datasets with different time resolutions and sample periods. The left panels are based on hourly data covering the period from the 1st of January 2023 to the 5th of May, 2026, and illustrate the Spot Prices by weekday and by month. The right panels are based on quarter-hourly data from the 1st of October 2025 to the 5th of May 2026, and show the corresponding profiles at a finer resolution since we have 4 times as many observations each day.

In Figures 4.3 and 4.4 we observe a pattern, with price spikes around 08:00 each day, followed by a decline towards midday, and a second peak around 19:00. These peaks come at peak hours, when households and businesses predictably need to consume a certain amount at the start and end of the working day.

In contrast, prices tend to decrease around midday. Here we see that electricity is often cheapest between approximately 12:00 and 15:00. During this period, solar power generation is typically at its highest, increasing the supply of low-cost renewable energy.

In Figure 4.3, the average hourly spot price is shown across weekdays. Monday appears to be the cheapest day during the early hours of the day (00:00–04:00); however, from around 06:00 onwards, it becomes the most expensive. In contrast, Saturday and Sunday consistently exhibit the lowest spot prices throughout the day. Spot prices are generally lower on weekends due to reduced demand and renewables still contributing to the system.

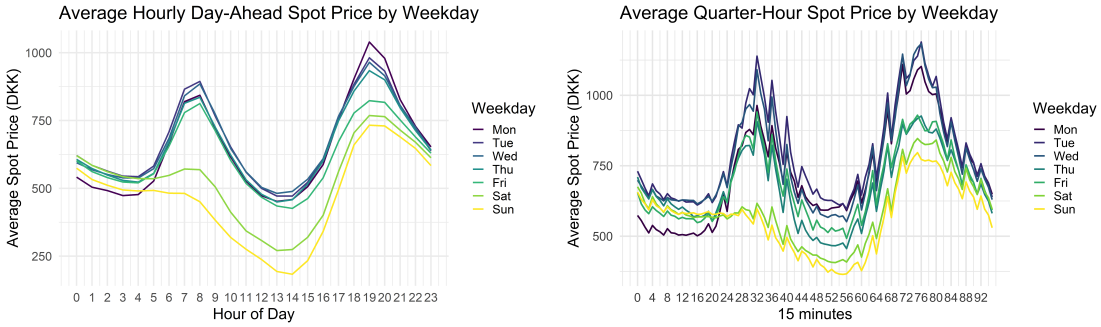


Figure 4.3: The left plot shows the average hourly day-ahead price by month, while the right plot shows the average quarterly day-ahead price by weekday.

In Figure 4.4, the average hourly spot prices are shown across months[1]. We observe that during the early hours, October and December exhibit the lowest prices. However, while October remains relatively cheap throughout the day, December becomes one of the more expensive months during daytime hours because electricity demand is typically higher in winter. Cold temperatures increase heating needs, which raises overall consumption from households and businesses throughout the day. At the same time, renewable generation is often lower or more volatile—especially solar, due to shorter daylight hours and weaker sunlight. This combination of higher demand and less cheap renewable supply means that more expensive production sources are needed to meet demand, which pushes prices up during the day.

Around midday, after the morning peak, January stands out as the most expensive month, as the post-peak decline is less pronounced compared to the other months. In contrast, the cheapest months during midday are May, June, and July, which is likely driven by higher solar generation and lower overall demand during the summer period. This is also consistent with Section 2.4, where it is clear that during winter periods Denmark tends to import electricity due to high demand, and lower supply. In contrast, during spring, the country often becomes a net exporter, as lower demand combined with strong wind and solar generation leads to surplus electricity production.

[1]Some months are missing from the average quarter-hour spot price by month, as the data only spans the period from the 1st of October 2025 to May 2026.

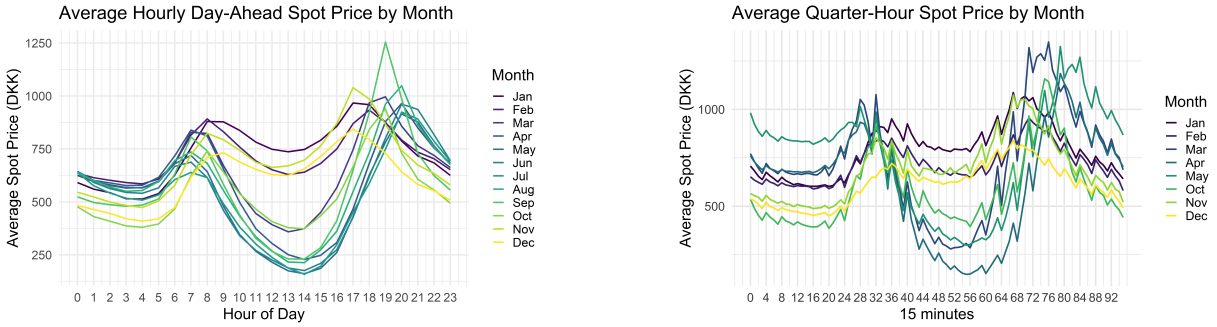


Figure 4.4: The left plot shows the average hourly day-ahead price by month, while the right plot shows the average quarterly day-ahead price by month.

In Figures 4.5 and 4.6, the spot prices at 10:00 a.m. and 4:00 p.m., respectively, are divided into four quarters of the hour and plotted from 1st of October 2025 to the present. At 10:00 a.m., we observe that the first quarter is consistently the most expensive, while the last quarter is the cheapest.

In contrast, at 4:00 p.m., the pattern is reversed: the first quarter is the cheapest, while the last quarter is the most expensive. This reflects the transition into the evening peak hours, when electricity demand typically increases while the availability of low-cost generation decreases. As a result, prices begin to rise again toward the end of the hour. This observation shows how the use of granulating the market to 15-minute intervals makes sense since the possibility to capture this change over the hour now becomes possible.

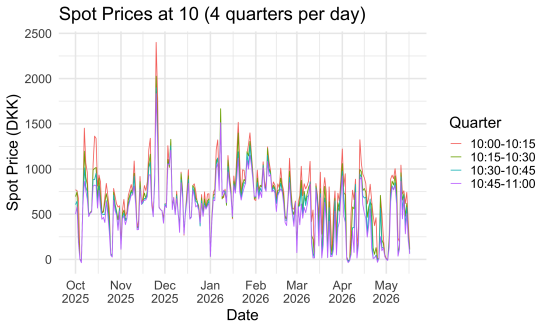


Figure 4.5: Spot prices at hour 10, shown for each 15-minute interval within the hour.

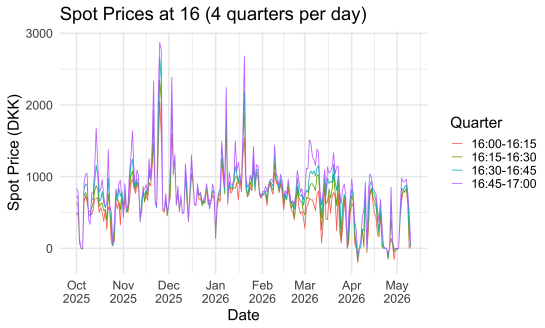


Figure 4.6: Spot prices at hour 16, shown for each 15-minute interval within the hour.

4.3 Quarterly Model Estimation

Before modelling the spot prices, we consider the raw quarterly spot prices together with different transformations, a shifted logarithmic transformation, the inverse hyperbolic sine transformation, and the Yeo-Johnson transformation given by,

$$\widehat{P}_{i,t} = \log\Big(P_{i,t} + \big|\min_{i,s} P_{i,s}\big| + c\Big), \quad c \in \{1, 100\} \tag{4.1}$$

$$\widehat{P}_{i,t} = \operatorname{asinh}(P_{i,t}) = \log\Big(P_{i,t} + \sqrt{P_{i,t}^2 + 1}\Big), \tag{4.2}$$

$$\widehat{P}_{i,t} = \begin{cases} \frac{(P_{i,t}+1)^{\lambda}-1}{\lambda}, & \text{if } \lambda \neq 0 \text{ and } y \geq 0 \\ \log\big(P_{i,t}+1\big), & \text{if } \lambda = 0 \text{ and } y \geq 0 \\ -\frac{(-P_{i,t}+1)^{2-\lambda}-1}{2-\lambda}, & \text{if } \lambda \neq 2 \text{ and } y < 0 \\ -\log\big(-P_{i,t}+1\big), & \text{if } \lambda = 2 \text{ and } y < 0 \end{cases} \tag{4.3}$$

where $P_{i,t}$ denotes the spot price at quarter i, and the minimum is taken over all observed time points. These transformations are used to address the distribution assumptions for the residuals of the models we test, to smooth out strong skewness, heavy tails, and making it still work for negative values observed in our data, while trying to normalize the residuals of the model.

To choose the transformation that makes the residuals normally distributed, we estimate all three standard panel data models introduced in Section 3.1, 3.2, and 3.3, thus a pooled OLS model, a fixed effects model, and a random effects model for each transformation of the spot price.

Therefore the dependent variable is the transformed spot price given in (4.1)-(4.3), while explanatory variables include consumption and production from offshore wind, onshore wind, and solar energy in DK1. Let

$$\begin{aligned} C_{i,t} &= \text{Consumption}_{i,t}, \\ OW_{i,t} &= \text{OffshoreWindPower}_{i,t}, \\ NW_{i,t} &= \text{OnshoreWindPower}_{i,t}, \\ S_{i,t} &= \text{SolarPower}_{i,t}. \end{aligned} \tag{4.4}$$

In addition, we include seasonal controls in the form of weekday and month indicators.

From this, we obtain the QQ-plots of the residuals shown in Figure A.1–A.2. These indicate that the residuals from the shifted logarithmic transformation with $c = 100$ are the closest to being normally distributed. Since approximate normality of the residuals is desirable for statistical inference and model diagnostics, this transformation is chosen as the preferred specification for the remainder of the analysis. Consequently, the following model results are presented using the transformed dependent variable in (4.1) with $c = 100$.

The **pooled regression model** assumes a homogeneous intercept across all panel units and ignores unobserved heterogeneity, and is specified in (3.3). The estimated pooled regression model is given by

$$\widehat{P}_{i,t} = 6.91 + 6.60 \times 10^{-8} C_{i,t} + 1.34 \times 10^{-4} OW_{i,t}$$

$$- 2.27 \times 10^{-4} NW_{i,t} - 3.28 \times 10^{-4} S_{i,t}.$$

The results indicate that higher consumption is associated with higher transformed spot prices, while increased renewable energy production generally lowers prices. In particular, onshore wind power and solar power have negative estimated coefficients, suggesting that higher production from these renewable sources tends to reduce spot prices. Offshore wind power has a surprisingly positive estimated coefficient, indicating a positive association with spot prices during the sample period. All estimated coefficients are statistically significant at the 1% level (see Table 4.1), providing strong evidence of a relationship between the explanatory variables and the transformed spot price. The model explains approximately 26.9% of the variation in the transformed dependent variable, as reflected by the R^2 value.

The **fixed effects model** accounts for unobserved time-invariant heterogeneity across panel units and is specified in (3.9) (the within transformation). The fixed effects model controls for unobserved, time-invariant heterogeneity across individual units by allowing each panel entity to have its own intercept, thereby focusing on within-unit variation over time. The estimated fixed effects model is given by

$$\begin{aligned} \widehat{P}_{i,t} = &-2.64 \times 10^{-8} C_{i,t} + 1.23 \times 10^{-4} OW_{i,t} \\ &- 2.06 \times 10^{-4} NW_{i,t} - 3.59 \times 10^{-4} S_{i,t}. \end{aligned} \tag{4.5}$$

The results indicate that higher consumption is associated with a small but statistically significant decrease in the transformed spot prices within entities over time. Offshore wind power is positively associated with prices, while both onshore wind power and solar power have negative estimated coefficients, suggesting that increased renewable generation from these sources tends to reduce spot prices within the same entity over time. All estimated coefficients are statistically significant at the 1% level (see Table 4.1), providing strong evidence of a relationship between the explanatory variables and the transformed spot price. The model explains approximately 27.3% of the within-entity variation in the transformed dependent variable, as reflected by the within R^2 value.

The **random effects model** accounts for unobserved heterogeneity by treating individual-specific effects as random variables rather than fixed parameters, allowing both within- and between-unit variation to be used in estimation. In addition to the main explanatory variables, the model also includes calendar effects captured by weekday and month dummies. The estimated random effects model is given by

$$\begin{aligned} \widehat{P}_{i,t} = &\, 8.35 - 4.40 \times 10^{-7} C_{i,t} + 4.09 \times 10^{-5} OW_{i,t} \\ &- 1.19 \times 10^{-4} NW_{i,t} - 2.56 \times 10^{-4} S_{i,t} \\ &+ \text{Weekday effects} + \text{Month effects}. \end{aligned}$$

The results indicate that consumption has a negative and highly statistically significant association with the transformed spot prices, while offshore wind power is again positively associated with prices. In contrast, both onshore wind power and solar power have negative

estimated coefficients, suggesting that increased renewable generation from these sources tends to reduce spot prices. The included weekday and month effects are also highly statistically significant, highlighting strong seasonal and intra-week patterns in spot prices over the sample period. All estimated coefficients are statistically significant at the 1% level (see Table 4.1), providing strong evidence of a relationship between the explanatory variables and the transformed spot price. The model explains approximately 47.6% of the variation in the transformed dependent variable, as reflected by the R^2 value, indicating a substantially improved fit compared to the pooled regression and fixed effects models. The results show that approximately 96.7% of the variation is idiosyncratic, while only 3.3% is attributed to individual-specific effects. This indicates that most variation in spot prices is driven by time-varying shocks rather than persistent differences across panel units.

The estimated theta parameter of 0.65 suggests a moderate degree of shrinkage toward the between-entity mean in the random effects transformation.

Table 4.1: Comparison of Panel Data Models for Spot Prices (Quarterly)

	Pooled OLS	**Fixed Effects**	**Random Effects**
Consumption (DK1)	6.598e-08***	-2.635e-08***	-4.396e-07***
	(5.706e-09)	(6.075e-09)	(7.339e-09)
Offshore Wind (DK1)	1.342e-04***	1.233e-04***	4.090e-05***
	(7.801e-06)	(7.334e-06)	(6.574e-06)
Onshore Wind (DK1)	-2.269e-04***	-2.056e-04***	-1.193e-04***
	(4.320e-06)	(4.091e-06)	(3.766e-06)
Solar Power (DK1)	-3.281e-04***	-3.588e-04***	-2.559e-04***
	(5.516e-06)	(6.213e-06)	(6.058e-06)
Weekday Effects	No	No	Yes
Month Effects	No	No	Yes
R^2	0.269	0.273	0.476
Adj. R^2	0.269	0.27	0.475

4.3.1 Model Selection

To determine the appropriate panel data specification for the empirical analysis, we apply a sequence of specification tests comparing the pooled regression, fixed effects, and random effects models as outlined in Section 3.4.

We first test whether unobserved individual heterogeneity is present by comparing the fixed effects model with the pooled regression model using the F-test given in (3.22) for joint significance of individual effects. The null hypothesis states that all individual-specific intercepts are equal, implying that the pooled regression model is sufficient. The test yields $F = 30.012$ with a p-value $< 2.2 \times 10^{-16}$. Hence, the null hypothesis is strongly

rejected. This provides clear evidence of significant unobserved individual heterogeneity, indicating that the pooled regression model is misspecified and that a model accounting for individual-specific effects is required.

Next, we apply the Breusch-Pagan Lagrange Multiplier test given in (3.23) to assess whether random effects are preferred over pooled regression. The null hypothesis assumes no panel-level variance component, i.e. $\sigma_\mu^2 = 0$, which implies that pooled OLS is sufficient. The test statistic is $\chi^2 = 25014$ with a p-value $< 2.2 \times 10^{-16}$. We therefore reject the null hypothesis and conclude that significant unobserved heterogeneity is present. This again suggests that pooled regression is not appropriate and motivates the use of a panel data model.

Finally, we compare the fixed effects and random effects specifications using the Hausman test given in (3.24). The null hypothesis states that the random effects estimator is consistent, i.e. that the individual effects are uncorrelated with the regressors. The test yields a statistic of $\chi^2 = 5004.8$ with a p-value $< 2.2 \times 10^{-16}$. We strongly reject the null hypothesis, implying that the random effects estimator is inconsistent due to correlation between the unobserved individual effects and the explanatory variables.

Based on the results of the three specification tests, both the F-test and the Breusch-Pagan test reject the pooled regression model, while the Hausman test rejects the random effects specification. Consequently, the fixed effects model is selected as the preferred specification for the empirical analysis, as it is robust to correlation between unobserved individual heterogeneity and the regressors.

To assess whether the error variance is constant across observations, we apply the studentized Breusch–Pagan test to all three panel specifications. The results are shown in Table 4.2. The null hypothesis of homoscedasticity is rejected for the pooled regression, fixed effects, and random effects models. This indicates that the variance of the error term is not constant across observations.

Table 4.2: Breusch-Pagan Heteroskedasticity Tests

Model	BP.Statistic	df	p.value
PR	2298.740	4	<2e-16
FE	2298.740	4	<2e-16
RE	2949.203	16	<2e-16

We further test for serial correlation in the idiosyncratic errors using the Breusch–Godfrey test for panel data. The results in Table 4.3 strongly reject the null hypothesis of no serial correlation for both the fixed effects and random effects models. This provides clear evidence of strong temporal dependence in the residuals.

This finding is consistent with the nature of spot prices, which exhibit persistence over time due to recurring demand patterns. The serial correlation implies that shocks to spot

prices are not purely transitory but persist across time periods, which invalidates standard inference procedures based on uncorrected standard errors.

Table 4.3: Breusch–Godfrey Serial Correlation Tests

Model	BG.Statistic	df	p.value
FE	19157.18	212	<2e-16
RE	18525.13	212	<2e-16

To examine whether residuals are independent across panel units, we apply the Pesaran CD test. In the results in Table 4.4 we find strong evidence against the null hypothesis of cross-sectional independence for both the fixed and random effects models, indicating substantial dependence across panel units.

This suggests that common shocks affect all panel entities simultaneously. In the context of electricity markets, such dependence is plausible due to shared exposure to national demand conditions, weather systems, fuel price changes, and interconnected grid dynamics.

The presence of cross-sectional dependence further motivates the use of robust inference methods that account for both temporal and cross-sectional correlation.

Table 4.4: Pesaran's CD or Breusch–Pagan's LM tests for cross sectional dependence

Model	CD.Statistic	p.value
FE	638.1713	<2e-16
RE	574.4507	<2e-16

Given the presence of heteroscedasticity, serial correlation, and cross-sectional dependence, we re-estimate statistical significance using cluster-robust standard errors.

The results in Table 4.5 and 4.6 show that offshore wind power, onshore wind power, and solar power remain statistically significant at conventional levels. In contrast, consumption is no longer statistically significant once robust standard errors are applied.

This indicates that the significance of consumption in the baseline fixed effects specification is not robust to violations of classical error assumptions.

To account simultaneously for heteroscedasticity, serial correlation, and cross-sectional dependence, we further apply Driscoll–Kraay standard errors. The results in Table 4.7 confirm that onshore wind power and solar power remain statistically significant determinants of spot prices. However, the effect of offshore wind power becomes statistically insignificant, and consumption remains insignificant.

While the baseline panel specifications initially indicated statistically significant effects for consumption, offshore wind power, onshore wind power, and solar power, these results

Table 4.5: Clustered FE results

Variable	Estimate	Std. Error	t value	Pr(>|t|)
Consumption_DK1	0e+00	0	-1.5010	0.1334
OffshoreWindPower_DK1	1e-04	0	7.6770	<1e-04
OnshoreWindPower_DK1	-2e-04	0	-33.5856	< 2.2e-16
SolarPower_DK1	-4e-04	0	-47.5750	< 2.2e-16

Table 4.6: Clustered RE results

Variable	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.3473	0.0768	108.7213	< 2.2e-16
Consumption_DK1	0.0000	0.0000	-15.7089	< 2.2e-16
OffshoreWindPower_DK1	0.0000	0.0000	5.8315	< 1e-04
OnshoreWindPower_DK1	-0.0001	0.0000	-22.3003	< 2.2e-16
SolarPower_DK1	-0.0003	0.0000	-17.0876	< 2.2e-16

changed substantially once heteroskedasticity, serial correlation, and cross-sectional dependence were properly accounted for through clustered and Driscoll–Kraay standard errors. This suggests that only the effects of onshore wind and solar generation are robust across all specifications of the error structure.

Overall, this highlights the importance of using fully robust inference methods in panel data settings with strong cross-sectional and temporal dependence.

Table 4.7: Driscoll-Krayy Standard Errors

Variable	Estimate	Std. Error	t value	Pr(>|t|)
Consumption_DK1	0e+00	0e+00	-0.4395	0.6603358
OffshoreWindPower_DK1	1e-04	1e-04	1.2906	0.1968600
OnshoreWindPower_DK1	-2e-04	1e-04	-3.4828	0.0004973
SolarPower_DK1	-4e-04	0e+00	-7.7689	< 1e-04

4.3.2 Dynamic Panel Data Model

We now consider the extended modelling approach with a dynamic modelling, concluding from A.5 we chose the log(SpotPrice_100) transformation, since the residuals behaves the best for that. The dynamic fixed effects model, written like (3.25), is specified as

$$\begin{aligned} P_{i,t} = \alpha_i &+ 0.41\, P_{i,t-1} + 0.12\, P_{i,t-7} - 1.50 \times 10^{-7}\, C_{i,t} \\ &+ 1.90 \times 10^{-5}\, OW_{i,t} - 1.17 \times 10^{-4}\, NW_{i,t} - 2.14 \times 10^{-4}\, S_{i,t} \end{aligned}$$

$$+ \sum_{w \in \mathcal{W}} \delta_w D_{w,i,t}, \tag{4.6}$$

where α_i is a vector of fixed effect affecting each panel, ranging between $(3.768024; 4.05557)$, and a dummy matrix of weekdays is given by

$$\begin{aligned} \delta_{\text{Tuesday}} &= -0.0273, \\ \delta_{\text{Wednesday}} &= -0.0794, \\ \delta_{\text{Thursday}} &= -0.1027, \\ \delta_{\text{Friday}} &= -0.0318, \\ \delta_{\text{Saturday}} &= -0.2016, \\ \delta_{\text{Sunday}} &= -0.1937. \end{aligned} \tag{4.7}$$

Monday serves as the reference category for the weekday dummy variables, as the on average most expensive weekday. The positive and highly significant autoregressive coefficients indicate substantial persistence in transformed spot prices across both daily and weekly horizons. This means that a one-unit increase in the transformed log-price yesterday is associated with a 0.4122 unit increase in the transformed log-price today, conditional on the remaining regressors and fixed effects.

Consumption exhibits a statistically significant negative relationship with prices within the estimated specification. The negative coefficients associated with onshore wind and solar generation are consistent with the merit-order effect, where increased renewable supply reduces equilibrium spot prices following 4.8. In contrast, offshore wind generation again exhibits a positive coefficient, which may reflect structural correlations with broader market conditions, potentially other countries market effects.

The weekday dummy variables reveal clear calendar effects relative to Mondays. In particular, Saturdays and Sundays are associated with the strongest negative deviations in spot prices, suggesting systematically lower demand and market activity during weekends.

Overall, the model explains approximately 49.8% of the within-panel variation in spot prices with an $R^2 = 0.4977$ as can be seen in Figure 4.9, indicating that approximately half of the within-panel variation is captured by the specification.

The diagnostic tests presented in Table 4.9 reveal several violations of the standard panel regression assumptions for the quarterly dynamic panel models. The Pesaran CD tests strongly reject the null hypothesis of cross-sectional independence, indicating the presence of common market shocks and interdependencies between the quarters. Furthermore, the Breusch–Godfrey tests provide strong evidence of serial correlation in the idiosyncratic residuals for all model specifications.

Since the specification includes lagged dependent variables together with fixed effects, the estimator may be subject to the Nickell bias. In dynamic fixed effects models, the within transformation induces a correlation between the transformed lagged dependent variables

Variable	Estimate	Std. Error	t-value	p-value
lag1	0.4122	0.0057	71.7759	< 0.001
lag7	0.1224	0.0055	22.1517	< 0.001
Consumption_DK1	-1.4970×10^{-7}	5.2816×10^{-9}	-28.3435	< 0.001
OffshoreWindPower_DK1	1.9049×10^{-5}	6.2810×10^{-6}	3.0327	0.0024
OnshoreWindPower_DK1	-1.1688×10^{-4}	3.5477×10^{-6}	-32.9442	< 0.001
SolarPower_DK1	-2.1442×10^{-4}	5.5190×10^{-6}	-38.8504	< 0.001
WeekdayTuesday	-0.0273	0.0066	-4.1239	< 0.001
WeekdayWednesday	-0.0794	0.0066	-11.9651	< 0.001
WeekdayThursday	-0.1027	0.0066	-15.5392	< 0.001
WeekdayFriday	-0.0318	0.0066	-4.7957	< 0.001
WeekdaySaturday	-0.2016	0.0068	-29.5918	< 0.001
WeekdaySunday	-0.1937	0.0068	-28.6363	< 0.001

Table 4.8: Quarterly Dynamic Panel Model Estimation Results for price transformation log(SpotPrice_100)

	R^2	F-test	Breush-Pagan	Pesaran CD	BG Serial Correlation
Statistic	0.4977	1616.12	55.11	556.11	18047
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Table 4.9: Diagnostic Tests for the SpotPrice_100 Dynamic Panel Data Model

and the transformed error term, leading to a finite-sample bias of order $\mathcal{O}\left(\frac{1}{T}\right)$ and with $T = 205$ we get

$$\frac{1}{205} \approx 0.0049,$$

which is expected to be relatively small due to the large time dimension of the panel. Consequently, although the dynamic fixed effects estimator may still exhibit finite-sample bias, the impact is likely limited within the current empirical framework.

Furthermore, following the tests results presented in Table 4.9 the potential endogeneity concerns arise within the specification, particularly between spot prices, consumption, and renewable generation variables. Some explanatory variables may be correlated with the error term, potentially leading to biased coefficient estimates. These violations imply that conventional standard errors may be inconsistent, suggesting that robust covariance estimators or common-factor approaches may be preferable.

However, within the scope of the present analysis, the regressors are treated as weakly exogenous, and the model is interpreted in a forecasting-specification.

We construct forecasts for each quarter and plot them together in Figure 4.7. The recursive forecasting procedure produces relatively accurate forecasts during the initial forecasting

day, however, forecast accuracy deteriorates substantially over the subsequent days due to the propagation of errors through the lagged dependent structure, and missing appropriate forecasting methods.

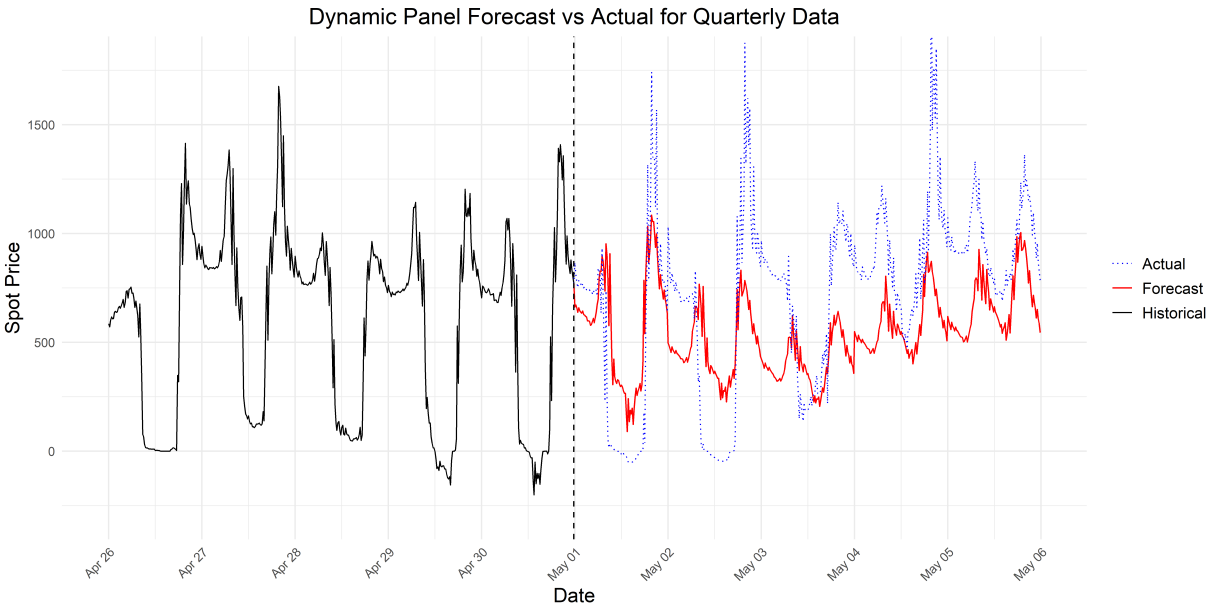


Figure 4.7: Forecast of the Dynamic panel data model for quarterly setup

4.3.3 Factor Model

In this section, we reduce the dimensionality of a large set of explanatory variables using Principal Component Analysis (PCA), extract latent common factors, and subsequently model remaining cross-sectional dependence through an additional factor structure estimated from the residuals.

Before presenting the estimation results, it is necessary to specify the choice of response variable. From the QQ-plots in Figure A.7, it is evident that the shifted logarithmic transformation of the spot price with $c = 100$ yields residuals that are closest to normality. Consequently, we proceed with the factor model estimation using the transformed response variable defined in (4.1) with $c = 100$.

Let $P_{i,t}$ denote the shifted logarithmic transformation of the spot price with $c = 100$ for quarter $i = 1, \ldots, 96$ at time t, and let $X_{i,t}$ denote the corresponding vector of explanatory variables. The explanatory set includes spot prices from neighbouring bidding zones (DK2, Germany, NO2, SE3, and SE4), regional consumption across the five Danish regions, renewable energy production variables (offshore and onshore wind in DK1 and DK2, as well as solar production), and finally exchange flows and aggregate production measures capturing system-wide supply conditions.

The correlation matrix is shown in Figure 4.8. It is shown that offshore and onshore wind production, as well as solar power, are negatively correlated with all spot prices. In contrast, renewable energy production is positively correlated with the consumption variables. Furthermore, a strong positive correlation is observed among the spot prices across all bidding zones. This differentiates itself from the results of the panel data models since now all the renewable energy sources seem to have a clear negative impact as one would expect, and thus giving more reason as to why a different modelling approach might be needed. Because the explanatory variables are correlated, Principal Component Analysis (PCA) is employed to reduce dimensionality and extract the dominant common variation in the data.

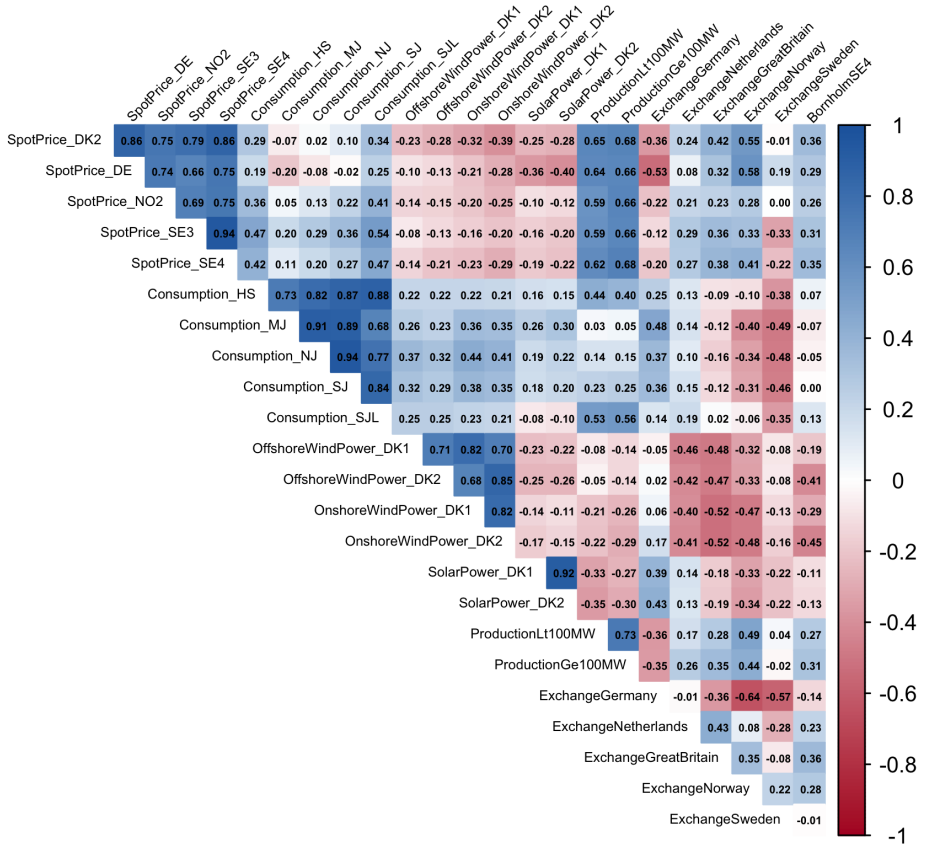


Figure 4.8: Correlation matrix of the explanatory variables used in the factor model for the quarterly data.

Prior to estimation, the explanatory matrix is standardised, and PCA is applied to obtain the factor representation

$$\widetilde{X} = F_X \Lambda_X' + E_X, \tag{4.8}$$

where F_X denotes the matrix of latent common factors, Λ_X contains the corresponding factor loadings, and E_X represents the idiosyncratic component. Based on the scree plot in Figure 4.9 and the proportion of explained variance in Table 4.10 from each principal component, we retain $r_X = 4$ principal components.

Instead of using the full set of regressors, the transformed response variable is projected onto the estimated factors

$$\widehat{P}_t = \alpha + F_{X,t}\beta + u_t. \tag{4.9}$$

The residuals represent variation not explained by the common factors extracted from X. To capture the fact that the residuals may still exhibit latent cross-sectional dependence, we apply a second PCA on the residual matrix,

$$U = F_e \Lambda_e' + E_e, \tag{4.10}$$

where F_e captures additional unobserved common components in the residual structure, and E_e denotes the idiosyncratic component. Based on the scree plot in Figure 4.10 and the PCA summary in Table 4.11, we select $r_e = 2$ residual factors.

To construct the moment conditions, we define the instrument matrix $Z_t = [1, F_{X,t}]$. The corresponding sample moment condition is given by

$$\bar{\mu}_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} Z_t' u_t = \begin{bmatrix} 4.590 \times 10^{-18} \\ -1.658 \times 10^{-17} \\ -5.628 \times 10^{-17} \\ -1.759 \times 10^{-17} \\ -6.496 \times 10^{-18} \end{bmatrix},$$

Table 4.10: PCA Summary on X (First 6 Principal Components)

	Std_Dev	Prop_Var	Cum_Prop
PC1	2.7028137	0.30438	0.30438
PC2	2.4596370	0.25208	0.55646
PC3	1.8896154	0.14878	0.70524
PC4	1.1222696	0.05248	0.75771
PC5	1.0015033	0.04179	0.79951
PC6	0.9095853	0.03447	0.83398

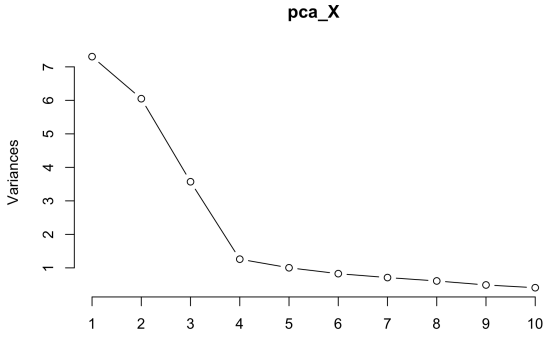


Figure 4.9: Scree plot for explanatory variables.

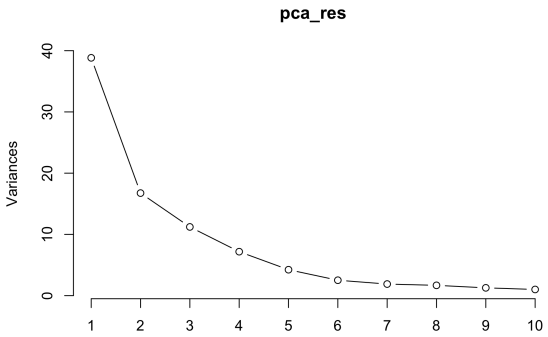


Figure 4.10: Scree plot for residuals.

Figure 4.11: Comparison of scree plots for explanatory variables and residuals.

Table 4.11: PCA Summary on the residuals (First 6 Principal Components)

	Std_Dev	Prop_Var	Cum_Prop
PC1	6.231507	0.40450	0.40450
PC2	4.092324	0.17445	0.57895
PC3	3.349706	0.11688	0.69583
PC4	2.679635	0.07480	0.77062
PC5	2.058321	0.04413	0.81475
PC6	1.587473	0.02625	0.84101

where u_t are first-stage residuals and Z_t contains the PCA factors and intercept. Numerically, these moments are close to zero, indicating approximate orthogonality between instruments and residuals.

However, to further account for remaining latent dependence, we introduce a correction based on the estimated residual factors. Let

$$\widehat{g} = \frac{1}{N}\sum_{t=1}^{N} F_{e,t} = \begin{bmatrix} -3.368 \times 10^{-17} \\ -3.342 \times 10^{-17} \end{bmatrix}$$

which are numerically zero. Consequently, the correction term

$$S(\widehat{F}_e \otimes I_d)\,\widehat{g} = \begin{bmatrix} -3.368 \times 10^{-17} \\ -3.342 \times 10^{-17} \\ -3.368 \times 10^{-17} \\ -3.342 \times 10^{-17} \\ -3.368 \times 10^{-17} \end{bmatrix}.$$

is also negligible (all entries of order 10^{-17}). The correction therefore does not alter the moment conditions in practice.

The idea is that the residual factor structure is removed from the moment conditions, improving robustness against omitted latent dependence. Hence, the corrected moment conditions become

$$\bar{\mu}_N = \frac{1}{N}\sum_{t=1}^{N} Z_t' u_t - S(F_e \otimes I_d)\widehat{g} = \begin{bmatrix} 3.828 \times 10^{-17} \\ 1.684 \times 10^{-17} \\ -2.260 \times 10^{-17} \\ 1.583 \times 10^{-17} \\ 2.718 \times 10^{-17} \end{bmatrix}. \tag{4.11}$$

The corrected moments are all numerically close to zero, indicating that the residual factor adjustment successfully removes the remaining latent dependence from the moment conditions. This suggests that the orthogonality conditions are approximately satisfied after accounting for the residual common factor structure.

It is observed that all corrected moment conditions are extremely close to zero, indicating that the residual factor correction successfully removes the remaining latent dependence from the moment conditions. The magnitudes are on the order of numerical precision, suggesting that the orthogonality conditions between the instruments and the corrected residuals are approximately satisfied. This indicates that the estimated factor structure captures most of the common variation in the residual component, thereby improving the robustness of the estimation procedure against omitted common factors and cross-sectional dependence.

Table 4.12 reports the efficient two-step GMM estimates, standard errors, and significance levels. The model now includes an intercept, which is estimated at 6.86 and is highly significant. All four principal components of the regressor matrix ($F_{X,1}$ to $F_{X,4}$) enter with p-values essentially zero, indicating a very strong in-sample fit. Specifically, the first component has a negative coefficient (-0.123), implying that the dominant common variation in the exogenous variables reduces the log price. The second component also has a negative but much smaller effect (-0.010). The third and fourth components are positive (0.053 and 0.077, respectively), meaning they increase the log price. The large t-statistics (in the hundreds) reflect the high signal-to-noise ratio after extracting the principal components.

Forecast Factor Model

To generate multi-step-ahead forecasts of the transformed spot price, we use the estimated two-layer factor structure. The forecasting procedure combines the dynamics of the common factors extracted from the explanatory variables with the latent factors obtained from the residual component of the model, thereby capturing both observable and unobservable sources of dependence.

Table 4.12: Factor coefficient estimates (GMM with correction)

	Coefficient	Std_Error	t_stat	p_value
Intercept	6.8630786	0.0009115	7529.84595	0
F_X1	-0.1231600	0.0005267	-233.81217	0
F_X2	-0.0099252	0.0003695	-26.86202	0
F_X3	0.0526051	0.0007107	74.02170	0
F_X4	0.0774495	0.0010104	76.65173	0

The first-stage factor model is given by (4.9), where u_t denotes the idiosyncratic error component. In addition, residual cross-sectional dependence is modelled through a second factor structure (4.10). The final forecasting with forecast horizon is five days, $h = 5$, is given by

$$\widehat{P}_{t+h} = \alpha + F'_{X,t+h}\widehat{\beta} + F'_{e,t+h}\widehat{\lambda}, \tag{4.12}$$

where $F_{X,t+h}$ denotes the forecasted principal components of the explanatory variables, $F_{e,t+h}$ denotes the forecasted principal components of the residual component and where both sets of factors are forecasted independently. Because the factors from the explanatory variables are common to all cross-sectional units, the term $F'_{X,t+h}\widehat{\beta}$ is identical for every quarter at a given forecast horizon

Since both $F_{X,t}$ and $F_{e,t}$ are unobserved components estimated from principal component analysis, their dynamics are modelled as reduced-form time series processes. Each factor is treated as an individual univariate time series and forecasted using an ARIMA specification selected via information criteria. This yields forecasts of the form

$$\widehat{F}_{X,t+h} = \mathbb{E}[F_{X,t+h} \mid \mathcal{F}_t], \quad \widehat{F}_{e,t+h} = \mathbb{E}[F_{e,t+h} \mid \mathcal{F}_t]$$

where $\mathcal{F}_t$ denotes the information set available at time t, where the model is estimated using data up to and including 30 April 2026, denoted as the final in-sample observation.

The coefficient vector $\hat{\beta}$ is obtained from the first-stage factor regression of Y_t on $F_{X,t}$. The residual factor loadings $\hat{\lambda}$ are estimated by projecting the residual matrix onto the residual factor space

$$\widehat{\lambda} = (F'_e F_e)^{-1} F'_e U$$

where U denotes the matrix of first-stage residuals. These loadings are treated as time-invariant over the forecast horizon.

Before interpreting the plots and forecast accuracy measures, the predicted values must be transformed back to the original scale, since the response variable was defined using a

shifted logarithmic transformation of the spot prices with $c = 100$. The inverse transformation is given by

$$\widehat{P}_{i,t} = \exp\left(\widehat{Y}_{i,t}\right) - \left(\left|\min\left(P^{\text{train}}\right)\right| + c\right), \quad c \in \{1, 100\} \tag{4.13}$$

where $\widehat{Y}_{i,t}$ denotes the forecasted transformed spot price and P^{train} represents the spot prices in the training sample.

Table 4.13 reports the forecast accuracy measures for both the full forecasting model in (4.12) and the restricted factor model in (4.9). The forecasting results indicate a substantial improvement when augmenting the baseline factor model with residual factors. The full model, which incorporates both observable common factors and latent residual components, achieves an RMSE of 393.25 compared to 820.11 for the restricted specification. This corresponds to a reduction of approximately 52%.

A similar pattern is observed for the MAE, which decreases from 715.38 in the restricted model to 314.89 in the full specification, implying an improvement of approximately 56%.

These results suggest that a significant portion of predictive information is contained in the latent cross-sectional structure captured by the residual factor model. In particular, the residual factors appear to capture additional systematic variation that is not explained by the observed covariates alone.

In Figures A.9–A.14, we present the quarterly forecasts obtained from the full forecasting model in (4.12). Figure 4.12 displays the corresponding 5-day forecasting horizon, comparing the observed and forecasted spot prices generated by the factor model.

4.4 Hourly Model Estimation

In this section we will implement the same application as we did for the Quarterly model estimation. We first estimate a pooled OLS model, a fixed effects model, and a random effects model for each transformation of the spot price and obtain the QQ-plots of the residuals. We then decide the transformation with the most normally distributed residuals for the rest of this analysis.

Table 4.13: Forecast accuracy comparison between full factor model and restricted model.

Model	RMSE	MAE
Full model	393.25	314.89
Restricted model	820.11	715.38
Improvement (%)	52.05	55.98

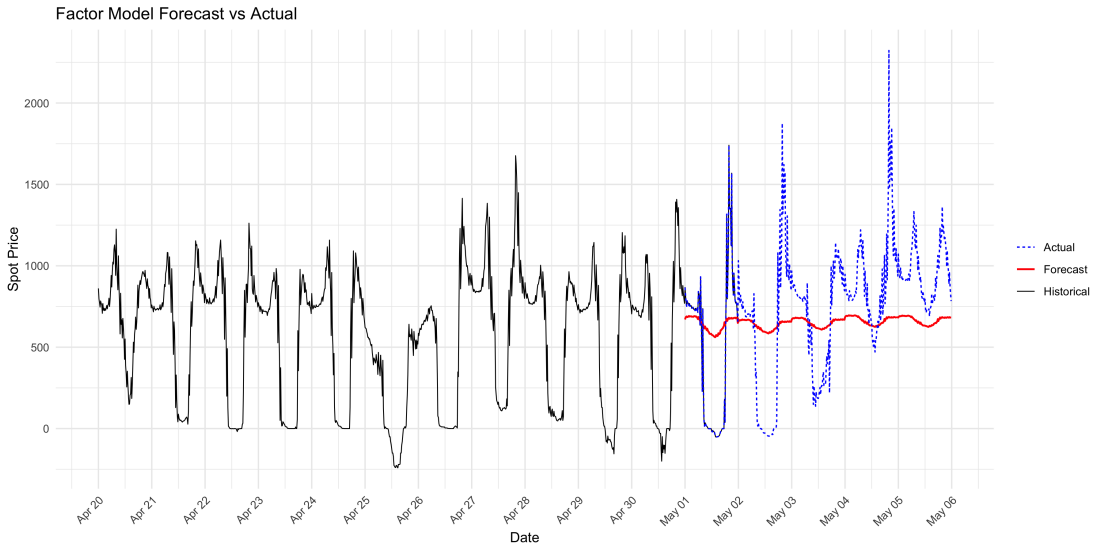


Figure 4.12: Observed and forecasted spot prices from the factor forecasting model in (4.12) over the 5-day forecasting horizon on quarterly data.

The transformation is the same ones presented earlier in (4.1)-(4.3). We again use the explanatory variables introduced in (4.4). When estimating the three models with the different transformation we plotted the residuals in a qq-plot in Figure A.3 and A.4, where we saw that the residuals that looked the most as a normal distribution was the shifted logarithmic transformation (4.1) with $c = 1$. We therefore proceed with the shifted logarithmic transformation $c = 1$ as the response variable in the following estimations.

In Table 4.14, the estimated coefficients and corresponding p-values for the PR, FE, and RE models are reported. The estimated models are given by

$$\begin{aligned}
\widehat{P}_{i,t}^{\mathrm{PR}} &= 8.24 + 3.84 \times 10^{-8} C_{i,t} + 1.11 \times 10^{-5} OW_{i,t} \\
&\quad - 6.37 \times 10^{-5} NW_{i,t} - 7.62 \times 10^{-5} S_{i,t}, \\
\widehat{P}_{i,t}^{\mathrm{FE}} &= 2.82 \times 10^{-8} C_{i,t} + 9.01 \times 10^{-6} OW_{i,t} \\
&\quad - 6.07 \times 10^{-5} NW_{i,t} - 7.28 \times 10^{-5} S_{i,t}, \\
\widehat{P}_{i,t}^{\mathrm{RE}} &= 8.41 - 3.26 \times 10^{-8} C_{i,t} + 8.18 \times 10^{-6} OW_{i,t} \\
&\quad - 5.43 \times 10^{-5} NW_{i,t} - 5.38 \times 10^{-5} S_{i,t} \\
&\quad + \text{Weekday effects} + \text{Month effects}.
\end{aligned} \tag{4.14}$$

The estimated coefficients again indicate that offshore wind production is positively associated with spot prices across all three models, whereas onshore wind production and solar power generation have negative estimated effects on prices.

The estimated effect of consumption is relatively small in all specifications. In the pooled regression and fixed effects models, the coefficient is positive, suggesting that higher elec-

tricity demand is associated with slightly higher prices. In contrast, the random effects model produces a small negative coefficient for consumption.

The goodness-of-fit differs across the models. The pooled regression model achieves an R^2 of 27.7%, the fixed effects model an R^2 of 25.7%, and the random effects model an R^2 of 33.6%, indicating that the random effects specification provides the best overall fit among the three models.

Table 4.14: Comparison of Panel Data Models for Spot Prices (Hourly)

	Pooled OLS	**Fixed Effects**	**Random Effects**
Consumption (DK1)	3.835e-08***	2.819e-08***	-3.259e-08***
	(1.172e-09)	(1.319e-09)	(1.793e-09)
Offshore Wind (DK1)	1.111e-05***	9.009e-06***	8.183e-06***
	(1.924e-06)	(1.848e-06)	(1.759e-06)
Onshore Wind (DK1)	-6.366e-05***	-6.072e-05***	-5.433e-05***
	(1.003e-06)	(9.628e-07)	(9.292e-07)
Solar Power (DK1)	-7.617e-05***	-7.279e-05***	-5.379e-05***
	(1.150e-06)	(1.426e-06)	(1.528e-06)
Weekday Effects	No	No	Yes
Month Effects	No	No	Yes
R^2	0.277	0.257	0.336
Adj. R^2	0.277	0.256	0.335

4.4.1 Model Selection

To determine the appropriate panel data specification for the hourly data, we once again performed the tests described in Section 3.4. First, the F-test comparing the fixed effects model with the pooled regression strongly rejects the null hypothesis of no individual effects ($F = 126.42$, $p < 2.2 \times 10^{-16}$), indicating that the fixed effects model provides a significantly better fit than the pooled model.

Next, the Breusch–Pagan Lagrange Multiplier test comparing the random effects and pooled regression models also strongly rejects the null hypothesis of no random effects ($\chi^2 = 135487$, $p < 2.2 \times 10^{-16}$). This suggests that unobserved individual-specific effects are present in the data, making the pooled regression inappropriate.

Finally, the Hausman test comparing the fixed and random effects models rejects the null hypothesis that the random effects estimator is consistent ($\chi^2 = 945.05$, $p < 2.2 \times 10^{-16}$). Therefore, the fixed effects model is preferred for the hourly data analysis.

4.4.2 Dynamic Panel Data Model

Following the results from Figure A.6, the logarithmic price transformation log(SpotPrice) was selected for the hourly dynamic panel model, as the residual distribution exhibited the most appropriate behaviour among the transformations. The dynamic fixed effects model, written as in 4.3.2, is specified as

$$\begin{aligned} P_{i,t} = \alpha_i &+ 0.30\, P_{i,t-1} + 0.12\, P_{i,t-7} + 6.21 \times 10^{-9}\, C_{i,t} \\ &+ 4.16 \times 10^{-6}\, OW_{i,t} - 4.81 \times 10^{-5}\, NW_{i,t} - 4.83 \times 10^{-5}\, S_{i,t} \\ &+ \sum_{w \in \mathcal{W}} \delta_w D_{w,i,t}, \end{aligned} \tag{4.15}$$

where α_i is ranging between $(4.804224; 4.865078)$, and a the dummy matrix of weekdays is given by

$$\begin{aligned} \delta_{\text{Tuesday}} &= -0.0126, \\ \delta_{\text{Wednesday}} &= -0.0119, \\ \delta_{\text{Thursday}} &= -0.0170, \\ \delta_{\text{Friday}} &= -0.0173, \\ \delta_{\text{Saturday}} &= -0.0411, \\ \delta_{\text{Sunday}} &= -0.0511. \end{aligned} \tag{4.16}$$

As in the quarterly specification, Monday serves as the reference category for the weekday dummy variables. The estimated autoregressive coefficients remain positive and statistically significant, indicating persistence in spot prices across the hourly panel structure. Compared to the quarterly model, the first lag coefficient is moderately smaller, suggesting a weaker short-term persistence effect within the hourly specification.

The renewable generation variables exhibit coefficient signs similar to those observed in the quarterly model. In particular, onshore wind and solar generation maintain negative relationships with spot prices, while offshore wind generation exhibits a small positive coefficient. Consumption now exhibits a weak positive relationship with prices, contrary to the negative coefficient observed in the quarterly specification.

The weekday dummy variables continue to indicate systematic calendar effects similar to those observed in the quarterly model, with weekends associated with the strongest negative deviations in spot prices relative to Mondays.

Overall, the hourly dynamic panel model explains approximately 39.1% of the within-panel variation in spot prices with an $R^2 = 0.3906$, which is lower than the explanatory power obtained in the quarterly specification.

The diagnostic results found in Table 4.16 remain qualitatively similar to those obtained for the quarterly dynamic panel model. In particular, the Pesaran CD test strongly rejects cross-sectional independence, while the Breusch–Godfrey test indicates substantial serial correlation in the residual structure.

Variable	Estimate	Std. Error	t-value	p-value
lag1	0.2998	0.0050	59.8893	< 0.001
lag7	0.1235	0.0049	25.2837	< 0.001
Consumption_DK1	6.2062×10^{-9}	1.2618×10^{-9}	4.9186	< 0.001
OffshoreWindPower_DK1	4.1644×10^{-6}	1.6791×10^{-6}	2.4802	0.0131
OnshoreWindPower_DK1	-4.8070×10^{-5}	8.8976×10^{-7}	-54.0259	< 0.001
SolarPower_DK1	-4.8301×10^{-5}	1.3733×10^{-6}	-35.1709	< 0.001
WeekdayTuesday	-0.0126	0.0017	-7.2437	< 0.001
WeekdayWednesday	-0.0119	0.0017	-6.8517	< 0.001
WeekdayThursday	-0.0170	0.0017	-9.7587	< 0.001
WeekdayFriday	-0.0173	0.0017	-9.9325	< 0.001
WeekdaySaturday	-0.0411	0.0018	-23.2035	< 0.001
WeekdaySunday	-0.0511	0.0018	-28.9112	< 0.001

Table 4.15: Hourly Dynamic Panel Model Estimation Results

	R^2	F-test	Breush-Pagan	Pesaran CD	BG Serial Correlation
Statistic	0.3906	1548.16	125.85	272.40	14248
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Table 4.16: Diagnostic Tests for the LogPrice Hourly Dynamic Panel Data Model

As discussed previously for the quarterly specification, the inclusion of lagged dependent variables together with fixed effects may introduce finite-sample Nickell bias, although with $T = 1209$ it is expected that the practical impact of this bias is limited. Similarly, potential endogeneity concerns may arise between spot prices, consumption, and renewable generation variables, and the specification should therefore primarily be interpreted as a forecasting-oriented model.

Finally, forecasts for the hourly dynamic panel specification are presented in Figure 4.13. Similar to the quarterly forecasting framework, the recursive forecasting procedure performs relatively well over the initial forecasting horizon, while forecast accuracy gradually deteriorates over subsequent periods due to recursive error propagation. When comparing Figure 4.7 and 4.13 there is a clear favourite in the quarterly model with its added granularity, the forecasting accuracy in the quarterly model is preferred.

4.4.3 Factor Model

In this section we follow the same process as we did for the factor model for quarterly data. We start with a large set of explanatory variables, where we apply PCA to reduce the dimensionality, extract latent common factors, and subsequently model remaining cross-sectional dependence through an additional factor structure estimated from the residuals.

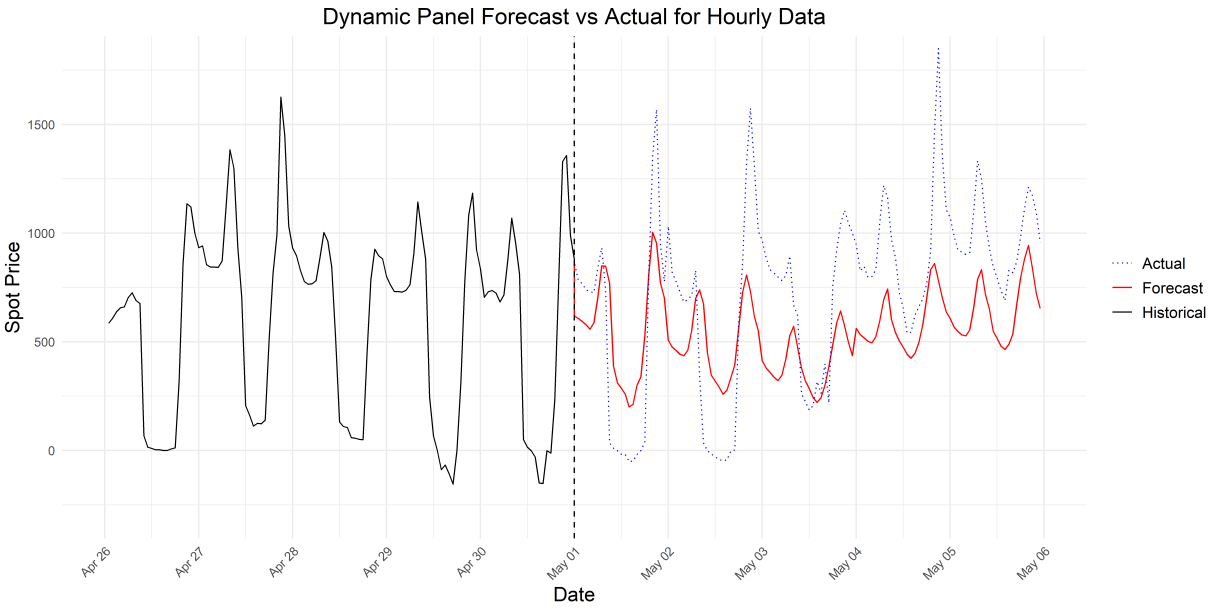


Figure 4.13: Forecast of the Dynamic panel data model for hourly setup

From the QQ-plots in Figure A.8 of the different transformation of the spot price, we choose the shifted logarithm transformation in (4.1) with $c = 1$ as the response variable.

Let $Y_{i,t}$ denote the shifted logarithmic transformation of the spot price with $c = 1$ for quarter $i = 1, \ldots, 96$ at time t, and let $X_{i,t}$ denote the corresponding vector of explanatory variables as in Section 4.3.3. The correlation matrix is shown in Figure 4.14, where we see the same correlation as in the quarterly correlation plot. We see that the renewable energy production is positively correlated with the consumption variables and are negatively correlated with all spot prices.

Since we want to reduce the dimension of the explanatory variables, and we saw correlation between the different explanatory variables, we standardize the variables and apply PCA to obtain (4.8). We then choose $r_X = 4$ principal components, based on the scree plot in Figure 4.15 and Table 4.17.

We again apply PCA on the residuals obtained in (4.9), to obtain the factors of the residuals (4.10). Based on the scree plot in Figure 4.17 and Table 4.18, we select $r_e = 2$ residual factors.

We now obtain the corresponding sample moment condition, residual factor correction

Table 4.17: PCA Summary on X (First 6 Principal Components) on hourly data

	Std_Dev	Prop_Var	Cum_Prop
PC1	2.5945684	0.28049	0.28049
PC2	2.5528753	0.27155	0.55204
PC3	1.8360173	0.14046	0.69250
PC4	1.1964293	0.05964	0.75214
PC5	1.0449388	0.04550	0.79764
PC6	0.8998433	0.03374	0.83137

Table 4.18: PCA Summary on the residuals (First 6 Principal Components)

	Std_Dev	Prop_Var	Cum_Prop
PC1	3.2337417	0.43571	0.43571
PC2	2.0458855	0.17440	0.61011
PC3	1.6991528	0.12030	0.73041
PC4	1.2197106	0.06199	0.79240
PC5	1.1312499	0.05332	0.84572
PC6	0.8315032	0.02881	0.87453

term, and the average residual factor

$$\bar{\mu}_N(\theta) = \begin{bmatrix} 1.074321 \times 10^{-17} \\ -1.232110 \times 10^{-17} \\ 5.020008 \times 10^{-18} \\ -9.877322 \times 10^{-18} \\ 1.527427 \times 10^{-17} \end{bmatrix}, \quad S(F_e \otimes I_d)\widehat{g} = \begin{bmatrix} -1.443932 \times 10^{-17} \\ 1.458371 \times 10^{-17} \\ -1.443932 \times 10^{-17} \\ 1.458371 \times 10^{-17} \\ -1.443932 \times 10^{-17} \end{bmatrix},$$

$$\widehat{g} = \begin{bmatrix} -1.443932 \times 10^{-17} \\ 1.458371 \times 10^{-17} \end{bmatrix},$$

Finally, the corrected moment conditions become

$$\bar{\mu}_N = \frac{1}{N}\sum_{t=1}^{N} Z_t' u_t - S(F_e \otimes I_d)\widehat{g} = \begin{bmatrix} -3.696106 \times 10^{-18} \\ 2.262614 \times 10^{-18} \\ -9.419309 \times 10^{-18} \\ 4.706388 \times 10^{-18} \\ 8.349538 \times 10^{-19} \end{bmatrix}. \tag{4.17}$$

The corrected moment conditions are all numerically very close to zero. This indicates that the orthogonality conditions are effectively satisfied after applying the residual factor correction. Hence, the latent dependence structure appears to have been adequately

accounted for in the model. The small residual values further suggest that the estimation procedure is numerically stable and well-specified.

Forecast Factor Model

In this section we will use the same method described in Section 4.3.3. Before interpreting the plots and forecast accuracy measures, the predicted values must be transformed back to the original scale, since the response variable was defined using a shifted logarithmic transformation of the spot prices with $c = 1$. The inverse transformation is given by (4.13) with $c = 1$.

Table 4.19 reports the forecast accuracy measures for both the full forecasting model in (4.12) and the restricted factor model in (4.9). Both RMSE and MAE are substantially lower for the full model, indicating improved accuracy across both squared and absolute error losses. The reduction of approximately 49% in RMSE and 55% in MAE suggests that incorporating the additional latent residual factor structure meaningfully enhances predictive performance. Overall, the results indicate that accounting for both observed and unobserved dependence structures is important.

In Figure A.15, we present the hourly forecasts obtained from the full forecasting model in (4.12). Figure 4.18 displays the corresponding 5-day forecasting horizon, comparing the observed and forecasted spot prices generated by the factor model.

4.5 Comparison

Since our application contains three different methodologies static FE, dynamic FE, and factor models, we will in this section make some comparison on how temporal aggregation (quarterly vs hourly) changes the estimated relationships, how much persistence the dynamic structure captures, and whether the factor correction improves the model structure and forecasting performance.

4.5.1 Fixed Effects Model Comparison

In this section we will focus on coefficient interpretation and temporal aggregation effects.

Table 4.19: Forecast accuracy comparison between full factor model and restricted model.

Model	RMSE	MAE
Full model	425.087	325.531
Restricted model	833.475	727.743
Improvement (%)	48.998	55.268

Both the quarterly and hourly datasets were estimated using pooled regression, fixed effects, and random effects models. In both cases, the Hausman test favoured the fixed effects specification, indicating that the individual effects are correlated with the regressors and that the fixed effects estimator is therefore preferred. The estimated fixed effects model for quarterly and hourly are given in (4.5) and (4.14) respectively. The signs of the coefficients are generally consistent across the two models for offshore wind power, non-wind renewable production, and solar production. Offshore wind power has a positive effect on spot prices, while non-wind renewable production and solar production have negative effects.

However, the magnitudes differ substantially between the quarterly and hourly models. The quarterly model produces considerably larger coefficients in absolute value. This suggests that temporal aggregation smooths short-run fluctuations and amplifies the longer-run relationships between spot prices and the explanatory variables.

A notable difference appears for consumption, where the hourly model yields a small positive coefficient, while the quarterly model produces a small negative coefficient. This may indicate that the relationship between consumption and spot prices differs across time scales. In the hourly setting, higher demand is associated with increased prices, whereas quarterly aggregation may capture broader seasonal or structural effects.

4.5.2 Dynamic Fixed Effects Model Comparison

Dynamic fixed effects models were estimated for both the quarterly and hourly datasets by including lagged prices and weekday dummy variables. The estimated fixed effects model for quarterly and hourly are given in (4.6) and (4.15) respectively.

Both models show strong persistence through the lagged spot price terms. The hourly model estimates

$$0.30P_{i,t-1} \quad \text{and} \quad 0.12P_{i,t-7},$$

while the quarterly model estimates

$$0.41P_{i,t-1} \quad \text{and} \quad 0.12P_{i,t-7}.$$

The coefficients on $P_{i,t-1}$ is larger in the quarterly model, indicating stronger persistence. The coefficients on $P_{i,t-7}$ is similar across both models, suggesting that weekly seasonal dependence remains about the same amount of significance in both models.

The signs of the explanatory variables remain consistent with the static fixed effects models. Offshore wind power contributes positively to prices, whereas on shore-wind renewable production and solar production contribute negatively. Again, the quarterly coefficients are larger in magnitude, reinforcing the conclusion that quarterly data strengthens long-run effects while reducing short-run noise.

The weekday dummy coefficients in (4.7) and (4.16) reveals clear weekly seasonality in spot prices. In both models, weekends exhibit the strongest negative effects, particularly

Sundays. However, the quarterly model produces substantially larger negative weekday coefficients. This suggests that the quarterly data captures more systematic weekly seasonal patterns. For example, the Sunday effect is -0.19 and -0.05 for quarterly and hourly respectively.

The forecast plots for the dynamic model for quarterly and hourly is shown in Figures 4.7 and 4.13 respectively. It it seen that both the dynamic models are able to capture the overall cyclical movement in spot prices during the forecasting period. In particular, both models captures the recurring daily price patterns and the general upward and downward movements observed in the actual prices.

The quarterly dynamic model follows the short-term fluctuations more closely and reacts more rapidly to sudden changes in spot prices. Both models capture the overall trend and periodic structure reasonably well. However, they are both less responsive to abrupt changes in the actual prices changes and fail in high volatility situations. Nevertheless, the inclusion of lagged prices and weekday effects allows both dynamic specifications to reproduce the main temporal dependence structure observed in the spot prices.

4.5.3 Factor Model Comparison

To account for latent cross-sectional dependence not captured by the observed explanatory variables, factor models were estimated for both the hourly and quarterly datasets. The factor correction was implemented by applying PCS to the explanatory variables and the residual matrix and incorporating the resulting latent factors into the moment conditions.

For both datasets, the corrected moment conditions are extremely close to zero after the factor adjustment. This indicates that the residual factor correction successfully removes remaining cross-sectional dependence and improves the orthogonality conditions of the model.

In the quarterly model, the corrected moments in (4.11) are of order 10^{-17}, while the hourly model produces corrected moments in (4.17) of order 10^{-18}. These very small values suggest that both factor models achieve a high degree of moment correction and provide a satisfactory fit to the data.

The forecast plots in Figures 4.12 and 4.18 illustrate that both factor models capture the general level and medium-term movement of spot prices during the forecasting horizon. However, compared with the realized prices, the forecasts are considerably smoother and exhibit substantially lower volatility. In both cases, the forecasts remain within a narrow range, whereas the actual prices exhibit substantial fluctuations and several extreme observations. This suggests that the factor models successfully capture the persistent common component of spot prices but are less effective at modelling sudden market shocks and short-run volatility.

Compared with the dynamic panel models, the factor models provide more stable and less volatile forecasts, reflecting the emphasis on latent common factors and cross-sectional

dependence. However, the dynamic models appear better suited for capturing rapid short-term movements and recurring price spikes in the electricity market.

The dynamic models outperform the static fixed effects models by incorporating temporal dependence through lagged prices and weekday effects. Furthermore, the factor models improve the treatment of latent cross-sectional dependence and produce corrected moment conditions close to zero.

Overall, the choice between hourly and quarterly modelling depends on the forecasting objective. Hourly models are preferable for short-term operational forecasting, whereas quarterly models are more suitable for medium-term trend analysis and structural interpretation.

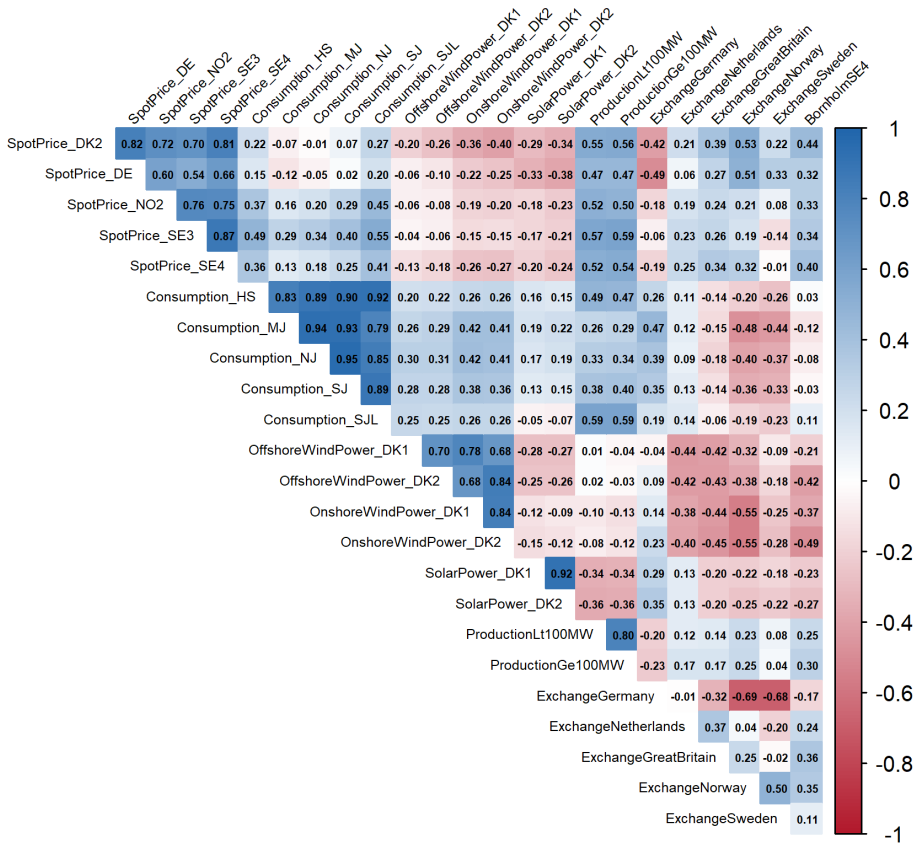


Figure 4.14: Correlation matrix of the explanatory variables used in the factor model for the hourly data.

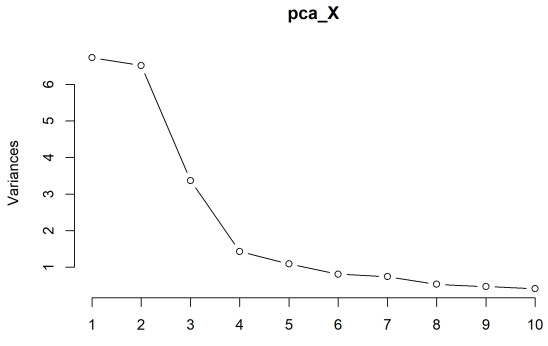


Figure 4.15: Scree plot for explanatory variables.

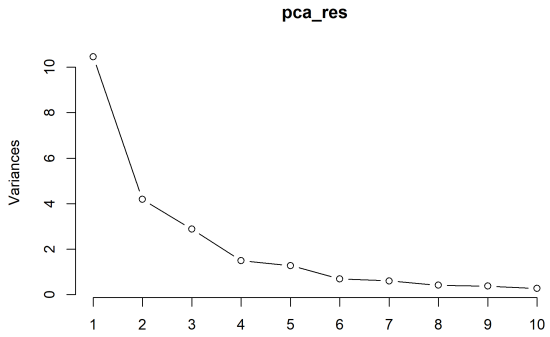


Figure 4.16: Scree plot for residuals.

Figure 4.17: Comparison of scree plots for explanatory variables and residuals.

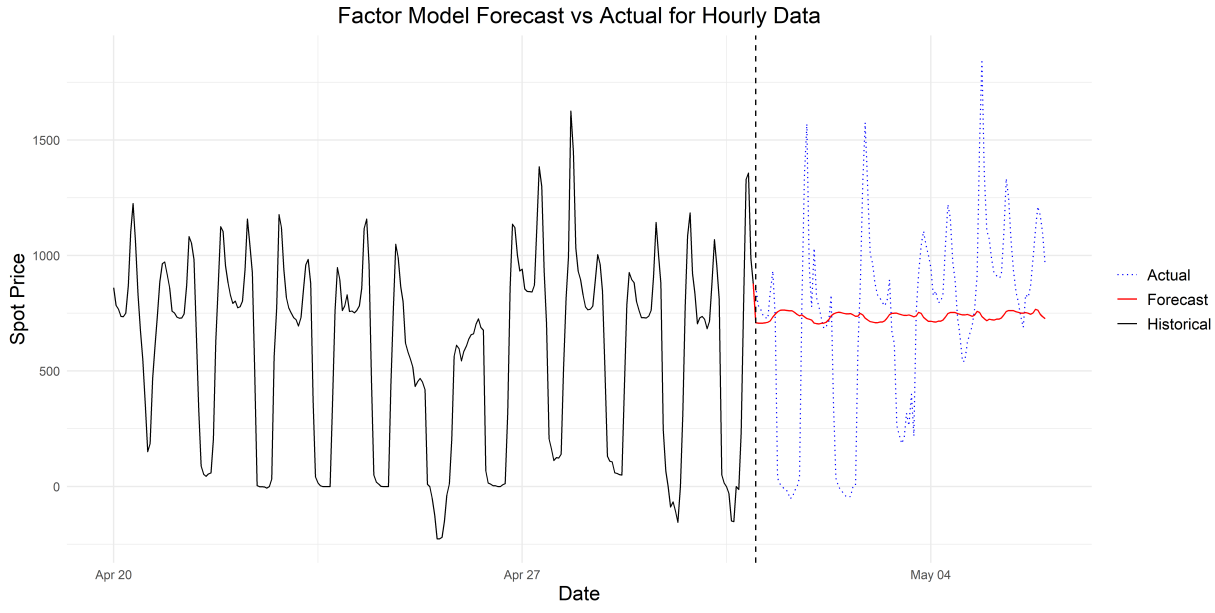


Figure 4.18: Observed and forecasted spot prices from the factor forecasting model in (4.12) over the 5-day forecasting horizon on hourly data.

5 Discussion

In this chapter, the empirical results from the application section are discussed and related to the theoretical framework presented throughout the thesis. In particular, the performance and limitations of the static panel data models, dynamic panel data models, and factor-augmented specifications are evaluated in relation to the modelling of electricity prices under the new 15-minute settlement regime.

Furthermore, the chapter discusses the implications of the obtained results, including the effects of temporal aggregation, cross-sectional dependence, residual persistence, and the choice of data transformations. Finally, potential improvements to the modelling framework and possible directions for future work are considered in order to further strengthen the forecasting and modelling of electricity prices within the Danish electricity market.

The empirical results from both the quarterly and hourly panel specifications indicate that electricity prices exhibit substantial persistence, significant calendar effects, and strong cross-sectional dependencies across time periods. While the dynamic panel data models were able to capture a moderate proportion of the within-panel variation, 0.49 and 0.39, found in Table 4.9 and 4.16 respectively, and produced relatively stable short-horizon forecasts, several limitations remain present within the modelling framework. In particular, the diagnostic results revealed substantial violations of panel data assumptions, including serial correlation and cross-sectional dependence, suggesting that electricity price dynamics are influenced by common latent market factors and evolving market conditions not fully captured within the estimated specifications.

The factor modelling framework further emphasized the importance of latent common structures within electricity markets. By incorporating principal component analysis together with residual factor corrections, the factor models were able to account for substantial cross-sectional dependence remaining within the data. The corrected moment conditions specifications, in (4.11) and (4.17) for quarterly and hourly respectively, were numerically close to zero after the residual factor adjustment, indicating that the latent dependence structure was largely captured by the estimated factors.

Furthermore, the comparison between the hourly and quarterly panel structures highlights the importance of a finer time span with the quarterly prices when modelling electricity prices. While the higher-frequency data contains additional short-term market information, it also introduces increased volatility, and more complex dependency structures. Consequently, the forecasting performance and statistical properties of the models become highly sensitive to the selected temporal resolution and transformation framework. We start by brushing on the main issue with the modelling approach specifically the de-trending and de-seasonalizing considerations.

5.1 De-Trending and De-Seasonalizing Considerations

As can be seen in Figure 5.1, there remains substantial autocorrelation persistence within the residual structure despite the residual distribution appearing approximately Gaussian. This indicates that although the transformations and dynamic specifications improved the statistical properties of the models, important temporal structures may still remain unaccounted for within the data. In particular, the remaining autocorrelation suggests that deterministic seasonal components and evolving trends are still partially present within the residual process.

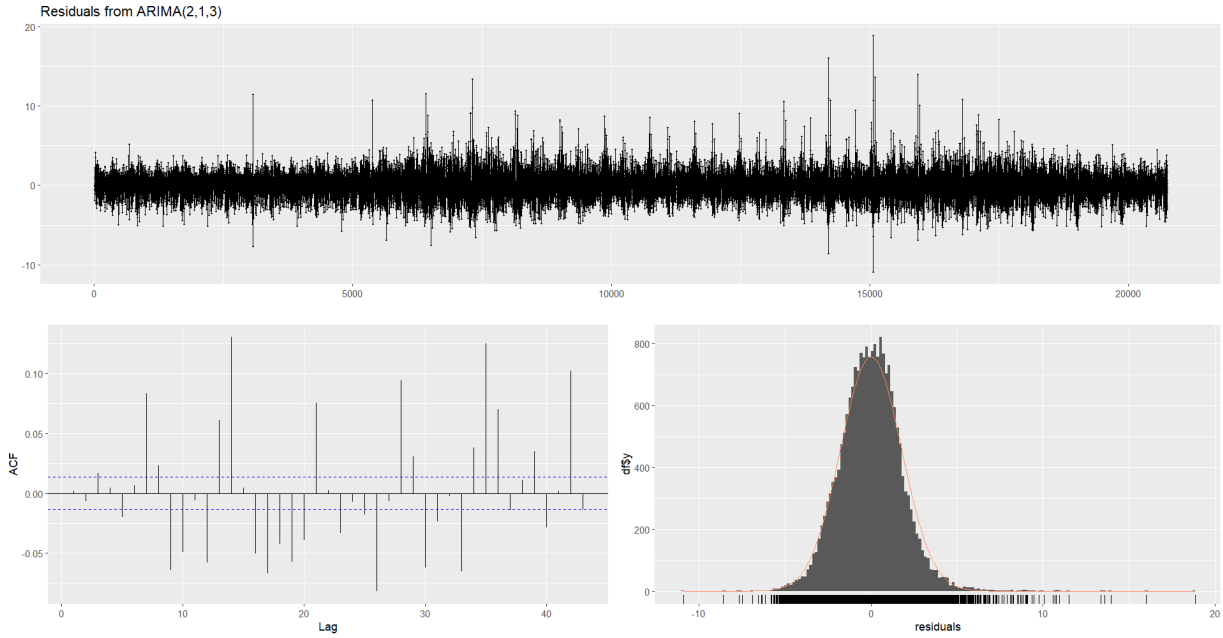


Figure 5.1: Residuals of the ARIMA(2,1,3) are plottet with the ACF and Histogram are plottet, gotten from the hourly factor model.

Consequently, the dynamic panel and factor specifications may unintentionally capture parts of these deterministic effects through the lagged dependent variables rather than isolating the underlying stochastic dynamics directly.

A possible improvement to the modelling framework would therefore have been to explicitly apply de-trending and de-seasonalizing procedures prior to estimation. Removing systematic seasonal structures beforehand may potentially reduce residual persistence and improve stationarity properties, thus strengthen the forecasting stability of the estimated models. This is particularly relevant within the quarterly specification, where the higher-frequency structure may amplify recurring dependencies.

However, due to the scope and time limitations of the present project, the primary focus was placed on comparing the modelling frameworks themselves rather than constructing

a complete preprocessing pipeline. Consequently, the implementation of systematic detrending and de-seasonalizing procedures is considered an important direction for future work and may potentially improve both the interpretability and forecasting performance of the proposed models.

5.2 Frequency Comparison

As explained in Chapter 2, the move toward quarterly price intervals allows for increasingly granular electricity pricing. This is further supported by the empirical application, where the quarterly specification improves the forecasting errors compared to the hourly framework, as seen in Table 5.1. Furthermore, the quarterly specification produces a more detailed forecasting structure that captures the short-term market dynamics more accurately than the hourly model.

Table 5.1: Forecast accuracy comparison between the hourly and quarterly dynamic panel models.

Model	RMSE	MAE
Hourly Dynamic Panel Model	8.260	8.260
Quarterly Dynamic Panel Model	6.879	6.874
Improvement (%)	16.72	16.78

The forecasting improvement is most clearly observed in the dynamic panel forecasts shown in Figure 4.7 and Figure 4.13, respectively. In contrast, the factor model forecasts were substantially more persistent and tended to fluctuate around an average daily level with only limited short-term variation.

The quarterly specification additionally exhibited stronger autoregressive persistence and moderately higher explanatory power compared to the hourly framework. One possible explanation is that the increased temporal resolution allows the model to capture finer market dynamics and short-term dependencies more effectively. However, despite the hourly specification being estimated on a substantially larger data foundation, the overall persistence structure of the recursive forecasting procedure appears qualitatively similar across both sequences.

Furthermore, comparing the hourly and quarterly models directly is not entirely straightforward, since the underlying stochastic properties of the data may differ substantially across temporal resolutions. The persistence structure, volatility clustering, seasonality, data horizons, and dependency patterns may all change depending on the chosen aggregation level. Consequently, the estimated coefficients and forecasting performance should primarily be interpreted within the context of their respective modelling frameworks rather than as directly equivalent model structures.

5.3 Implications of the Logarithmic Transformation and Data Snooping

The use of logarithmic transformations played an important role throughout the modelling framework, particularly due to the highly skewed and volatile nature of electricity prices. The transformation generally improved the residual behaviour, reduced extreme skewness, and stabilized parts of the variance structure within the estimated models. Consequently, the transformed specifications produced more stable estimation results and improved diagnostic properties compared to modelling the raw electricity prices directly.

However, the logarithmic transformation additionally introduces several important limitations. Most notably, the transformation alters the scale and distribution of the original price process, implying that the estimated coefficients no longer directly represent changes in the underlying electricity prices themselves. Instead, the models describe changes within the transformed price structure, making the economic interpretation of coefficient magnitudes less intuitive.

Furthermore, electricity prices are characterized by occasional spikes and abrupt market movements, which become compressed under logarithmic transformations. While this stabilizes the estimation procedure, it may simultaneously reduce the ability of the models to fully capture extreme price events and sudden volatility shifts that are economically important within electricity markets.

The transformation framework additionally complicates the forecasting interpretation. Forecasts generated within the transformed space may not directly correspond to forecasts of the original electricity prices, particularly due to retransformation effects and nonlinearities introduced by the logarithmic scale. Consequently, while the transformation improved statistical stability and residual behaviour, it simultaneously introduces a trade-off between econometric tractability and direct economic interpretability.

An additional limitation within the modelling process concerns the possibility of data snooping and specification selection bias. Since multiple price transformations and model specifications were considered throughout the empirical analysis, there exists a potential risk that parts of the final modelling framework were influenced by characteristics specific to the observed sample period.

For the present project, all considered price transformations were estimated and evaluated systematically through residual diagnostics. In particular, the QQ-plots presented in Figures A.5, A.6, A.7, and A.8, were used as part of the model assessment procedure. The transformations selected for presentation within the thesis were therefore chosen based on their comparatively stronger residual behaviour and distributional properties. Although the remaining model specifications were also computed, they generally produced qualitatively similar results with only moderately weaker residual diagnostics. Consequently, the final model selection should primarily be interpreted as a robustness-based specification

choice rather than the result of substantial differences in forecasting behaviour across the transformations.

The inclusion of lagged dependent variables within the dynamic panel data models was additionally motivated by the strong persistence structure naturally present within electricity markets. Similarly, the inclusion of the seventh lag was motivated by the recurring weekly structure and systematic weekday persistence observed within the spot price dynamics.

5.4 Dynamic Panel Data Performance

The Pesaran CD tests strongly rejected the null hypothesis of cross-sectional independence for both panel structures. Making intuitively sense since both the pricing of electricity as well as the consumption of it is highly dependent on the rest of the panels. For instance if it was sunny at 14:00 the chance of it also being sunny at 14:15 is high. Other market effects could be transmission constraints, fuel prices. Consequently, the residual dependence indicates that latent common drivers may remain omitted from the estimated specifications.

Furthermore, the inclusion of lagged dependent variables together with fixed effects introduces the possibility of finite-sample Nickell bias within the dynamic panel estimators. Although the bias is theoretically of order $\mathcal{O}\left(\frac{1}{T}\right)$, the relatively large time dimension of the panels suggests that the practical impact of this bias is likely limited within the present framework.

Another important limitation concerns potential endogeneity within the electricity market structure. Spot prices, consumption, and renewable generation are inherently interconnected through market equilibrium mechanisms, implying that some explanatory variables may be correlated with the error term. For instance, periods of high electricity demand may simultaneously influence both market prices and generation behaviour, while renewable production is strongly connected to weather conditions that may additionally affect broader market dynamics. This for future modelling could be attempted removed through deterministic stabilization of the data before modelling.

The recursive forecasting procedure revealed that forecast accuracy deteriorates gradually over longer forecasting horizons. While the models were generally capable of producing stable short-term forecasts, the autoregressive structure propagates forecasting errors recursively through future predictions, while failing to capture big spikes in the price movement. Consequently, small misspecifications or shocks within early forecast periods may accumulate over time, reducing forecasting performance over extended horizons.

5.5 Factor Modelling Performance

The factor modelling framework was introduced to explicitly account for the strong cross-sectional dependence observed throughout the empirical analysis of the dynamic panel data models. In both the hourly and quarterly applications, the Pesaran CD tests strongly rejected the null hypothesis of cross-sectional independence. Consequently, the factor models were constructed to capture both the observable common variation through PCA on the explanatory variables and the remaining latent dependence structure through an additional residual factor correction using GMM estimation.

For both the hourly and quarterly models, the corrected moment conditions after the residual factor adjustment were numerically extremely close to zero, indicating that the latent dependence structure was substantially reduced within the estimation procedure. This suggests that the residual factor correction successfully removed a substantial proportion of the remaining cross-sectional dependence and improved the orthogonality conditions underlying the factor estimation. In particular, the quarterly and hourly factor models both produced corrected moments with magnitudes close to numerical precision, indicating that the estimated latent factor structures captured a large proportion of the common residual variation.

The forecasting behaviour of the factor models differed substantially from the dynamic panel data specifications. Whereas the dynamic models reacted more rapidly to short-term fluctuations and recurring price movements, the factor models produced considerably smoother forecasts with substantially lower volatility. In both the hourly and quarterly applications, the forecasts tended to fluctuate around a stable average level while still capturing the broader medium-term movements in spot prices reasonably well.

This behaviour suggests that the factor models were effective at capturing the persistent common component of spot prices and the broader latent market structure shared across periods. However, the factor framework was less effective at reproducing sudden market shocks, abrupt price spikes, and short-term volatility fluctuations. Consequently, while the factor correction improved the treatment of latent cross-sectional dependence and produced highly stable moment conditions, the resulting forecasts became comparatively smoother and less responsive to extreme market events, and thus unreliable for forecasting attempts.

5.6 Future work

One future work improvement for panel data specifically would be to improve the forecasting ability. Recent literature on forecasting in dynamic panel data models emphasizes the use of empirical Bayes shrinkage techniques for stabilizing unit-specific parameter estimates in short panels. Following the intuition of Liu, Moon, and Schorfheide, a potential extension would therefore be to replace fixed effects estimation with shrinkage-based random effects or hierarchical modeling approaches to improve recursive multi-step forecast

stability [15].

Another method for the modelling could be a time-varying approach for the coefficients estimation, either by selection time periods and running the model for those, or making the model time varying in itself.

Given that in the future the amount of observations under the new quarterly regime, since it only started the 01-10-2025, as future work running the model again could give insight, as we conclude from Figure 4.4 there is a clear yearly persistence in the electricity market, and then following the process used in [8].

5.6.1 Machine Learning Approaches

A possible alternative to the dynamic panel data and factor modelling frameworks would have been the use of machine learning methods for spot price forecasting. Machine learning models are often well-suited for electricity markets due to the highly non-linear relationships between prices, demand, renewable production, fuel prices, and weather variables.

Instead of specifying a parametric model directly, machine learning methods attempt to approximate an unknown function

$$y_t = f(X_t) + \varepsilon_t,$$

where X_t contains explanatory variables. A simple example would be a feedforward neural network,

$$\hat{y}_t = \sigma\left(W^{(2)}\sigma\left(W^{(1)}X_t + b^{(1)}\right) + b^{(2)}\right),$$

where $W^{(i)}$ and $b^{(i)}$ denote weight matrices and bias terms, while $\sigma(\cdot)$ represents an activation function.

For sequential data, recurrent neural networks (RNNs) or Long Short-Term Memory (LSTM) models may be more appropriate, since they explicitly incorporate temporal dependencies. An LSTM model recursively updates hidden states through

$$h_t = F(h_{t-1}, X_t),$$

allowing the model to retain information from previous observations over longer time horizons.

Machine learning approaches could potentially capture nonlinear interactions and complex dependencies more effectively than linear panel data models. Furthermore, such models may naturally handle large numbers of explanatory variables without requiring strong distributional assumptions.

A challenge is that electricity prices exhibit spikes and strong seasonality, which may require careful preprocessing. Consequently, while machine learning approaches may improve forecasting performance, they may also reduce the interpretability and economic intuition of the model compared to more traditional stochastic or econometric frameworks.

5.6.2 Ornstein–Uhlenbeck Mean Reversion Models

An alternative approach to modeling electricity prices would have been the use of mean-reverting stochastic processes, such as the Ornstein–Uhlenbeck (OU) process. Instead of modeling prices through dynamic panel data models or latent factor structures, one may directly model the spot price dynamics through a stochastic differential equation. The OU process is defined by

$$dX_t = \kappa(\mu - X_t)dt + \sigma dW_t$$

where μ denotes the long-run equilibrium level, $\kappa > 0$ is the speed of mean reversion, and σ is the volatility parameter. The drift term $\kappa(\mu - X_t)$ ensures that the process reverts toward the equilibrium level μ.

The explicit solution is given by

$$X_t = X_0 e^{-\kappa t} + \mu(1 - e^{-\kappa t}) + \sigma \int_0^t e^{-\kappa(t-s)} dW_s$$

which yields

$$\begin{aligned} \mathbb{E}[X_t] &= X_0 e^{-\kappa t} + \mu(1 - e^{-\kappa t}), \\ \mathrm{Var}(X_t) &= \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa t}\right). \end{aligned}$$

Hence,

$$\lim_{t \to \infty} \mathbb{E}[X_t] = \mu, \tag{5.1}$$

showing the mean-reverting behaviour directly. The OU framework focuses more directly on the stochastic dynamics and mean-reverting properties. However, the panel framework allows for easier incorporation of cross-sectional dependencies, explanatory variables, and interactions between bidding zones.

6 Conclusion

In this thesis, we investigated whether panel data models with lagged regressor structures can be used to model and forecast electricity prices under the new 15-minute settlement regime in the Danish DK1 electricity market. To answer the problem statement proposed in Chapter 1, both static and dynamic panel data models were estimated on hourly and quarter-hourly data, together with factor-augmented specifications designed to account for latent cross-sectional dependence and common market shocks.

The empirical results demonstrated that the quarter-hourly specifications generally improved forecasting performance compared to the hourly models while simultaneously capturing more detailed short-term market dynamics. In particular, the dynamic panel models revealed substantial autoregressive persistence and strong weekly seasonal effects within electricity prices. However, the diagnostic tests additionally revealed significant serial correlation and cross-sectional dependence.

To address these limitations, factor-augmented modelling approaches were introduced through principal component analysis and residual factor corrections. The factor models successfully reduced the remaining latent dependence structure, as indicated by corrected moment conditions numerically close to zero for both temporal resolutions. While the factor models captured medium-term common variation in electricity prices reasonably well, the resulting forecasts became considerably smoother and less responsive to abrupt price spikes and short-term volatility compared to the dynamic panel specifications.

Overall, the thesis answers the problem statement by demonstrating that dynamic panel data models with lagged regressor structures can be used to model and forecast electricity prices under the new 15-minute settlement regime. Furthermore, the analysis showed that factor-augmented panel frameworks provide an effective approach for accounting for the substantial cross-sectional dependence present within electricity markets. Nevertheless, the results additionally highlight that electricity prices contain several overlapping dynamics simultaneously, including persistence, seasonality, latent common dependence, and abrupt nonlinear market shocks, making it difficult for a single modelling framework to fully capture all characteristics of the electricity market.

Consequently, while the dynamic panel models were more effective at modelling short-term fluctuations and recurring temporal structures, the factor models were better suited for capturing persistent latent common components and broader medium-term market behaviour. Future work involving improved preprocessing pipelines, nonlinear forecasting methods, and extended high-frequency datasets may therefore further improve the modelling and forecasting of electricity prices under the evolving quarter-hourly settlement regime.

Bibliography

[1] Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models.

[2] Baltagi, B. H. (2008). *Econometrics*. Springer Texts in Business and Economics. Springer, 4 edition.

[3] Baltagi, B. H. (2011). *Econometrics*. Springer Texts in Business and Economics. Springer, 5 edition.

[4] Davidson, R. and MacKinnon, J. G. (2021). *Econometric Theory and Methods*. Oxford University Press.

[5] Energi Data Service (2026). Imbalance price. https://www.energidataservice.dk/tso-electricity/ImbalancePrice.

[6] Energinet (n.d.). Energinet – danish transmission system operator. Accessed: 2026-04-20.

[7] EPEX SPOT (2026). Basics of the power market. https://www.epexspot.com/en/basicspowermarket.

[8] Ergemen, Y. E., Haldrup, N., and Rodríguez-Caballero, C. V. (2016). Common long-range dependence in a panel of hourly nord pool electricity prices and loads.

[9] European Commission (2024). Renewable energy targets.

[10] European Wind Energy Association (n.d.). The impact of wind power on the power market: Dk case. Accessed: 2026-04-20.

[11] Green Power Denmark (2023). Elkunder flokkes om variable priser. Accessed: 2026-05-04.

[12] Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence.

[13] Juodis, A. and Sarafidis, V. (2020). A linear estimator for factor-augmented fixed-t panels with endogenous regressors: Supplementary appendix. Supplementary Appendix.

[14] Juodis, A. and Sarafidis, V. (2022). A linear estimator for factor-augmented fixed-t panels with endogenous regressors.

[15] Liu, L., Moon, H. R., and Schorfheide, F. (2020). Forecasting with dynamic panel data models.

[16] Next Kraftwerke (2026). Liberalization of energy markets. https://www.next-kraftwerke.com/knowledge/liberalization-energy-markets.

[17] Nord Pool (2026). Day-ahead market. https://www.nordpoolgroup.com/en/the-power-market/Day-ahead-market/.

[18] Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels.

[19] Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure.

[20] Ruppert, D. and Matteson, D. S. (2015). *Statistics and Data Analysis for Financial Engineering: With R Examples*. Springer, 2 edition.

[21] Vattenfall Energy Plaza (2026). Derfor skifter energimarkedet til 15-minutters afregning. https://energyplaza.vattenfall.dk/blog/derfor-skifter-energimarkedet-til-15-minutters-afregning#download-page.

[22] Wikipedia contributors (2026a). Breusch–godfrey test. https://en.wikipedia.org/wiki/Breusch%E2%80%93Godfrey_test. Accessed: 2026-05-26.

[23] Wikipedia contributors (2026b). Breusch–pagan test. https://en.wikipedia.org/wiki/Breusch%E2%80%93Pagan_test. Accessed: 2026-05-26.

[24] Wikipedia contributors (2026c). Clustered standard errors. https://en.wikipedia.org/wiki/Clustered_standard_errors. Accessed: 2026-05-26.

[25] Wikipedia contributors (2026d). Gauss–markov theorem. https://en.wikipedia.org/wiki/Gauss%E2%80%93Markov_theorem.

[26] World Future Council (2026). Global renewable energy goal. https://www.worldfuturecouncil.org/global-renewable-energy-goal/.

Appendices

A.1 Least Squares Estimators

Consider the linear regression model

$$Y = X\beta + U,$$

where Y is a $T \times 1$ vector of observations, X is a $T \times k$ matrix of regressors, β is a $k \times 1$ vector of parameters, and U is a $T \times 1$ vector of errors with covariance matrix

$$\Omega = \mathrm{Var}(U).$$

Depending on the assumptions regarding the covariance matrix Ω, different estimators can be obtained.

A.2 Ordinary Least Squares (OLS)

If the covariance matrix of the errors is

$$\Omega = \sigma^2 I, \tag{A.1}$$

the errors are homoskedastic and uncorrelated. In this case, the ordinary least squares (OLS) estimator is

$$\widehat{\beta}_{\mathrm{OLS}} = (X'X)^{-1}X'Y$$

with variance

$$\mathrm{Var}(\widehat{\beta}_{\mathrm{OLS}}) = \sigma^2(X'X)^{-1}.$$

If the covariance matrix differs from (A.1), the OLS estimator remains unbiased but is no longer efficient.

A.3 Weighted Least Squares (WLS)

Under heteroskedasticity, the covariance matrix Ω is diagonal, with the t-th element equal to w_t^2. To obtain efficient estimates, we can transform the model so that it satisfies the Gauss–Markov assumptions. Consider the transformed regression

$$w_t^{-1}Y_t = w_t^{-1}X_t\beta + w_t^{-1}U_t$$

The transformed model has homoskedastic errors with variance equal to 1. We can then apply OLS to this transformed model, obtaining the weighted least squares (WLS) estimator

$$\widehat{\beta}_{\mathrm{WLS}} = (X'W^{-1}X)^{-1}X'W^{-1}Y$$

where $W = \mathrm{diag}(w_1^2, \ldots, w_t^2)$.

A.4 Generalized Least Squares (GLS)

In the general case, assume that the covariance matrix Ω is known. Since Ω is positive definite, we can find a matrix P such that

$$\Omega^{-1} = PP'.$$

Define the transformed variables

$$Y^* = P'Y, \quad X^* = P'X, \quad U^* = P'U,$$

so that the transformed model becomes

$$Y^* = X^*\beta + U^*.$$

Applying OLS to the transformed regression yields the generalized least squares (GLS) estimator

$$\widehat{\beta}_{\text{GLS}} = (X'^*X^*)^{-1}X'^*Y^* = (X'PP'X)^{-1}X'PP'Y = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y.$$

The variance of the GLS estimator is

$$\begin{aligned}
\text{Var}(\widehat{\beta}_{\text{GLS}}) &= \text{Var}\big((X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y\big) \\
&= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\text{Var}(Y)\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \\
&= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Omega\,\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \\
&= (X'\Omega^{-1}X)^{-1}.
\end{aligned}$$

A.5 Feasible Generalized Least Squares (FGLS)

In practice, the covariance matrix Ω is typically unknown, making GLS infeasible. However, In some cases it may reasonable to suppose that the covariance follows a known form up to some parameters. For example, it may be possible to write

$$\mathbb{E}[u_t^2]\exp(Z_t\gamma),$$

where γ is a vector of unknown parameters and Z_t is a vector of known variables.

In this case, the parameters β and γ can be estimated from the data, leading to the feasible generalized least squares (FGLS) estimator. We estimate the model in several steps:

1. Estimate β using OLS and obtain the residuals $\widehat{U}_t$
2. Estimate γ using OLS in the auxiliary regression

$$\log\left(\widehat{U}_t^2\right) = Z_t\gamma + \nu_t$$

3. For all t, compute the weights

$$\widehat{w}_t^2 = (\exp(Z_t\widehat{\gamma}))^{1/2}$$

4. Obtain FGLS estimates, using weighted least squares

$$\widehat{\beta}_{\text{FGLS}} = (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}Y$$

where $\widehat{\Omega} = \text{diag}(\widehat{w}_1^2, \ldots, \widehat{w}_t^2)$

A.6 Heteroskedasticity and Autocorrelation Consistent Covariance (HAC)

HAC estimators are an extension of Heteroskedasticity-Consistent Covariance Matrix Estimators (HCCME) to allow for autocorrelation. HCCME estimates the covariance matrix under heteroskedasticity using only residuals; HAC generalizes this to errors that may also be correlated over time.

Let Ω be the unknown error covariance. For processes with temporally limited dependence, HAC approximates

$$\Sigma = X'\Omega X = \Gamma(0) + \sum_{j=1}^{p} \omega_j(\Gamma(j) + \Gamma'(j))$$

where

$$\Gamma(j) = \sum_{t=j+1}^{T} \widehat{u}_j\widehat{u}_{t-j}X_t'X_{t-j}$$

and ω_j is a weight assigned to lag j. Newey and West (1987) proposed a Bartlett kernel

$$K(x) = \begin{cases} 1 - |x|, & \text{if } |x| \leq 1, \\ 1, & \text{if } |x| > 1. \end{cases}$$

which guarantees positive semi-definiteness and provides a Heteroskedasticity and Autocorrelation Consistent (HAC) variance estimate.

A.7 Generalized Method of Moments (GMM)

The generalized method of moments (GMM) is an estimation framework based on population moment restrictions implied by the underlying economic model. In the present

framework, the orthogonality conditions arise from the assumption that the selected instruments are uncorrelated with the idiosyncratic error component.

Let θ_0 denote the true parameter vector, where

$$\theta_0 = (\beta_0', g_0')',$$

with β_0 containing the structural parameters and g_0 containing the nuisance parameters associated with the factor structure.

Suppose that the model satisfies the population moment condition

$$E\left[\mu_i(\theta_0)\right] = 0,$$

where $\mu_i(\theta)$ denotes an $[\zeta \times 1]$ vector of moment functions associated with cross-sectional unit i.

In the factor-augmented panel framework considered in Chapter 3.6, the moment conditions are given by

$$\mu_i(\theta) = Z_i' \left(Y_i - X_i\beta\right) - S\left(\hat{F}_e \otimes I_d\right) g,$$

where Z_i denotes the matrix of valid instruments, $\hat{F}_e$ denotes the estimated factor proxy matrix obtained from PCA, and S denotes the instrument selection matrix.

Since the population moments are unknown, GMM replaces them with their sample counterparts,

$$\bar{\mu}_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} \mu_i(\theta).$$

If the model is correctly specified, the sample moment vector should be close to zero for parameter values near the true parameter vector θ_0.

The GMM estimator is therefore defined as

$$\hat{\theta} = \arg\min_{\theta} \bar{\mu}_N(\theta)' \Omega_N \bar{\mu}_N(\theta),$$

where Ω_N denotes a positive definite weighting matrix.

If the number of moment conditions equals the number of unknown parameters, the estimator solves

$$\bar{\mu}_N(\theta) = 0.$$

When the number of moment conditions exceeds the number of parameters, the model becomes overidentified. In this case, the GMM estimator selects the parameter vector minimizing the weighted distance between the sample moments and zero.

The asymptotic distribution of the estimator depends on the Jacobian matrix

$$\Gamma = \operatorname*{plim}_{N\to\infty}\left(-\frac{\partial \bar{\mu}_N(\theta_0)}{\partial \theta'}\right),$$

together with the asymptotic covariance matrix of the moment conditions,

$$\Delta = \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mu_i(\theta_0)\mu_i(\theta_0)'.$$

Under standard regularity conditions, the GMM estimator satisfies

$$\sqrt{N}(\hat{\theta}-\theta_0) \xrightarrow{d} N\left(0, (\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega\Delta\Omega\Gamma(\Gamma'\Omega\Gamma)^{-1}\right),$$

where Ω denotes the probability limit of the weighting matrix Ω_N.

The efficient GMM estimator is obtained by selecting

$$\Omega_N = \hat{\Delta}^{-1},$$

where

$$\hat{\Delta} = \frac{1}{N}\sum_{i=1}^{N}\hat{\mu}_i(\hat{\theta})\hat{\mu}_i(\hat{\theta})'$$

is the sample analogue of the covariance matrix of the moment conditions.

Using the efficient weighting matrix yields the asymptotic covariance matrix

$$\operatorname{Var}(\hat{\theta}) = (\Gamma'\Delta^{-1}\Gamma)^{-1}.$$

The GMM framework is particularly well suited for the present setting since it allows estimation in the presence of endogenous regressors, heteroscedasticity, serial correlation, and latent common factor structures. Furthermore, the methodology remains applicable even when the number of valid moment conditions differs across time periods through the selection matrix S.

A.8 Principal Component Analysis

To construct observable factor proxies for the latent common factor structure introduced in Section 3.6, we employ Principal Component Analysis (PCA). The purpose of PCA is to recover a low-dimensional representation of the common variation present across the candidate factor proxies.

Let

$$\hat{F}_R = \begin{pmatrix} \hat{f}_{1,1} & \cdots & \hat{f}_{1,T} \\ \vdots & \ddots & \vdots \\ \hat{f}_{R,1} & \cdots & \hat{f}_{R,T} \end{pmatrix}$$

denote the matrix of candidate factor proxies, where each row corresponds to a candidate proxy variable observed over time. Throughout the analysis, the variables are assumed centered such that the sample mean has been removed from each series.

The covariance matrix associated with the candidate factor proxies is defined as

$$\Sigma_F = T^{-1}\hat{F}_R\hat{F}_R'.$$

Following the standard PCA methodology, the covariance matrix admits the spectral decomposition

$$\Sigma_F = O\Lambda O',$$

where

$$O = (o_1, \ldots, o_R)$$

is an orthogonal matrix containing the eigenvectors of Σ_F, and

$$\Lambda = \text{diag}(\lambda_1(\Sigma_F), \ldots, \lambda_R(\Sigma_F))$$

contains the ordered eigenvalues satisfying

$$\lambda_1(\Sigma_F) \geq \cdots \geq \lambda_R(\Sigma_F).$$

The first principal component is obtained by solving

$$\max_{o_1} \quad o_1'\Sigma_F o_1$$

subject to

$$o_1'o_1 = 1.$$

The solution is given by the eigenvector corresponding to the largest eigenvalue $\lambda_1(\Sigma_F)$. More generally, the j'th principal component is obtained from the projection

$$o_j'\hat{F}_R,$$

subject to the orthogonality conditions

$$o_j'o_k = 0, \quad j \neq k.$$

The principal components are therefore mutually uncorrelated since

$$\text{Cov}(o_j'\hat{F}_R, o_k'\hat{F}_R) = o_j'\Sigma_F o_k = 0,$$

for all $j \neq k$.

The eigenvalues measure the amount of variation explained by each principal component, where

$$\frac{\lambda_j(\Sigma_F)}{\sum_{\ell=1}^{R} \lambda_\ell(\Sigma_F)},$$

denotes the proportion of total variation explained by the j'th component.

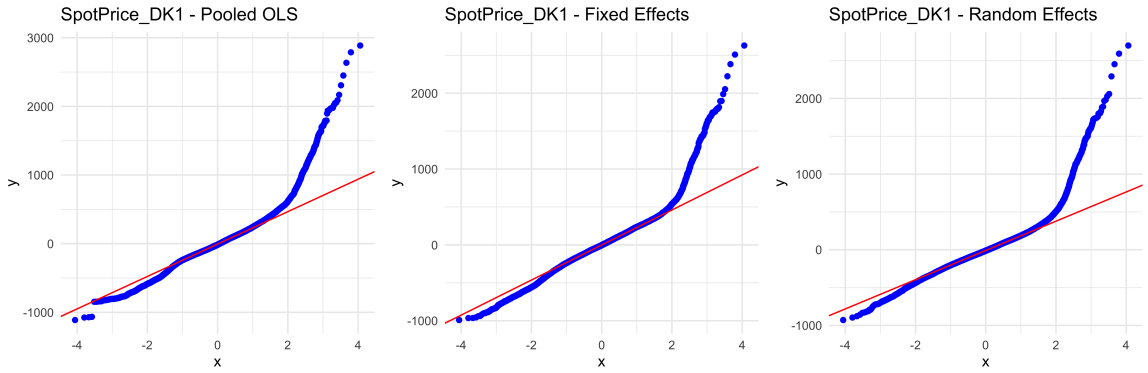
To determine the number of latent factors, we employ the eigenvalue ratio criterion given by

$$\hat{L}_e = \arg \max_{r \in \{1,\ldots,r_{\max}\}} \frac{\lambda_r(\Sigma_F)}{\lambda_{r+1}(\Sigma_F)},$$

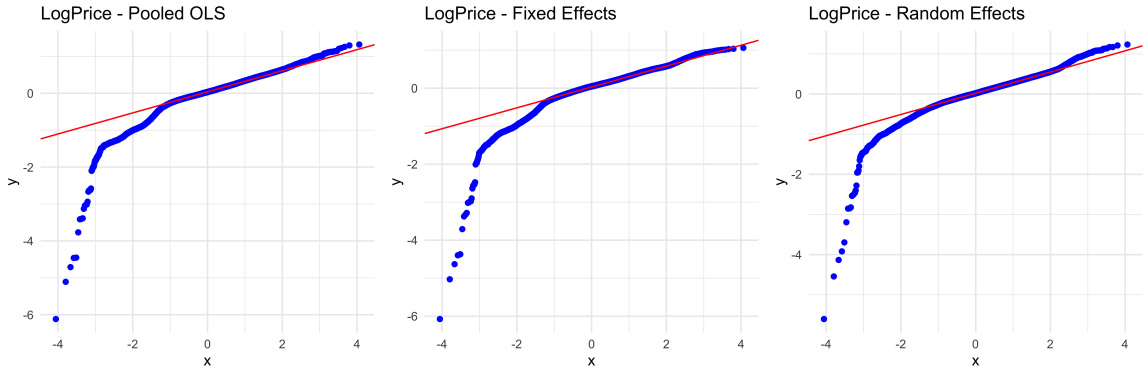
where $\lambda_r(\Sigma_F)$ denotes the r-th largest eigenvalue of Σ_F.

Intuitively, the criterion selects the dimension at which the eigenvalues exhibit the largest relative drop, corresponding to the point where the dominant common variation transitions into idiosyncratic noise. The first $\hat{L}_e$ principal components are subsequently collected into the estimated factor proxy matrix $\hat{F}_e$, which is used in the factor-augmented GMM estimation procedure.

A.9 Plots



(a)

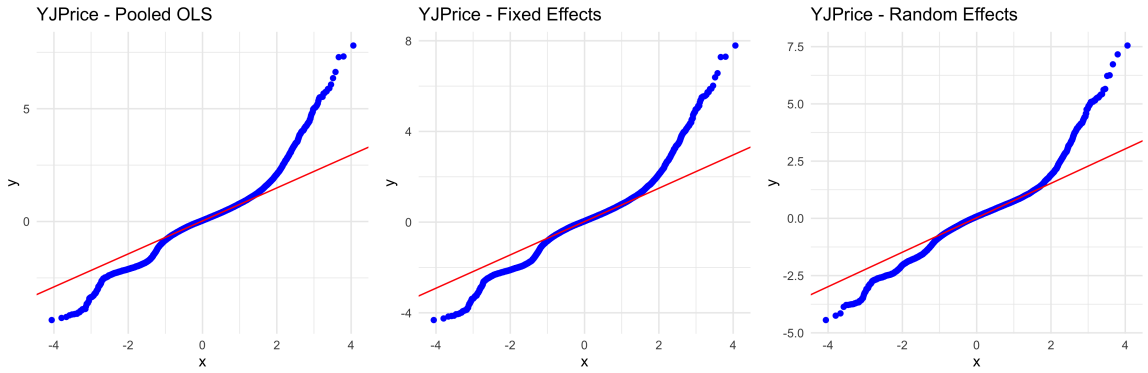


(b)

(c)

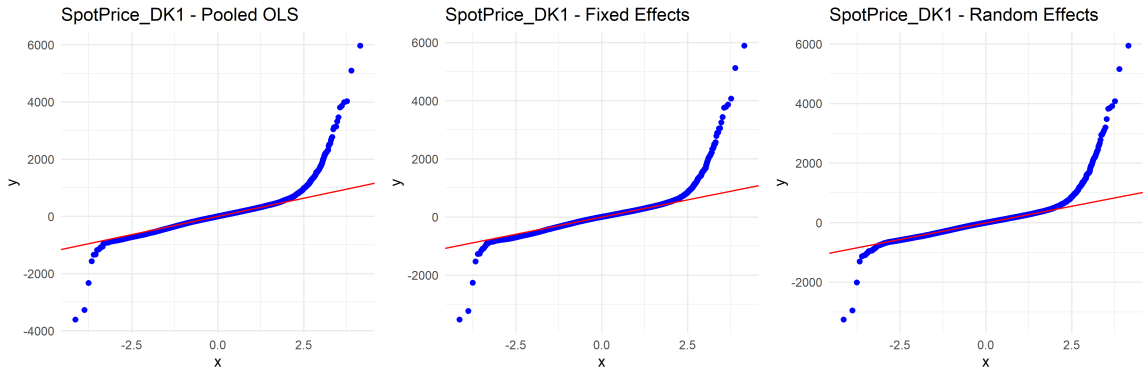
Figure A.1: QQ-plots comparing the residuals distribution of transformations of quarterly spot prices: (a) raw spot prices, (b) logarithmic transformation with $c = 1$, and (c) logarithmic transformation with $c = 100$.

(d)

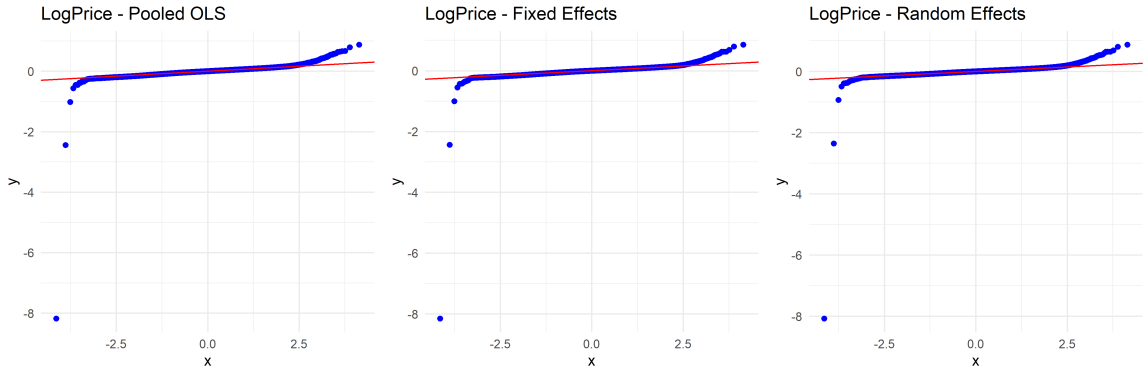


(e)

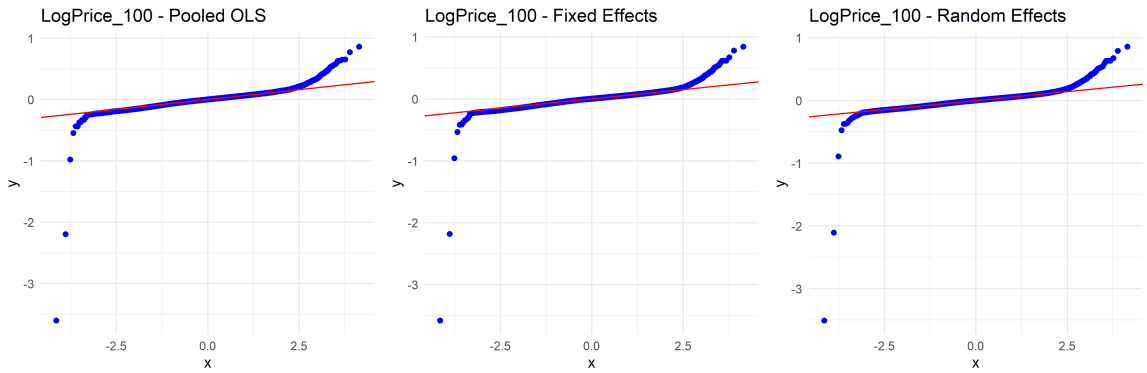
Figure A.2: QQ-plots comparing the residuals distribution of transformations of quarterly spot prices: (d) inverse hyperbolic sine transformation, and (e) Yeo–Johnson transformation.



(a)

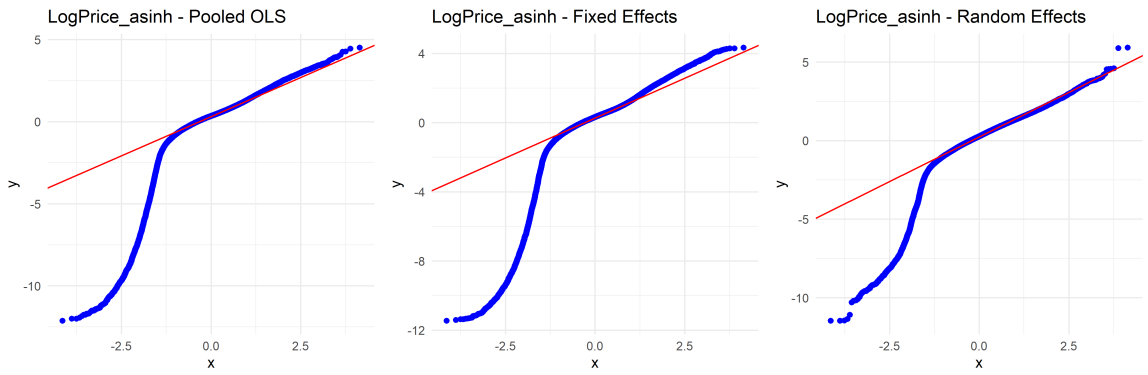


(b)

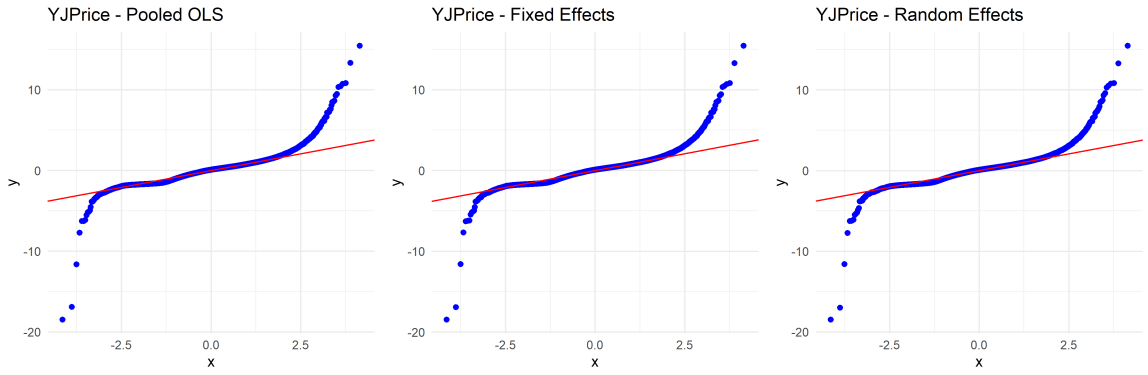


(c)

Figure A.3: QQ-plots comparing the residuals distribution of transformations of hourly spot prices: (a) raw spot prices, (b) logarithmic transformation with $c = 1$, and (c) logarithmic transformation with $c = 100$.



(d)



(e)

Figure A.4: QQ-plots comparing the residuals distribution of transformations of hourly spot prices: (d) inverse hyperbolic sine transformation, and (e) Yeo–Johnson transformation.

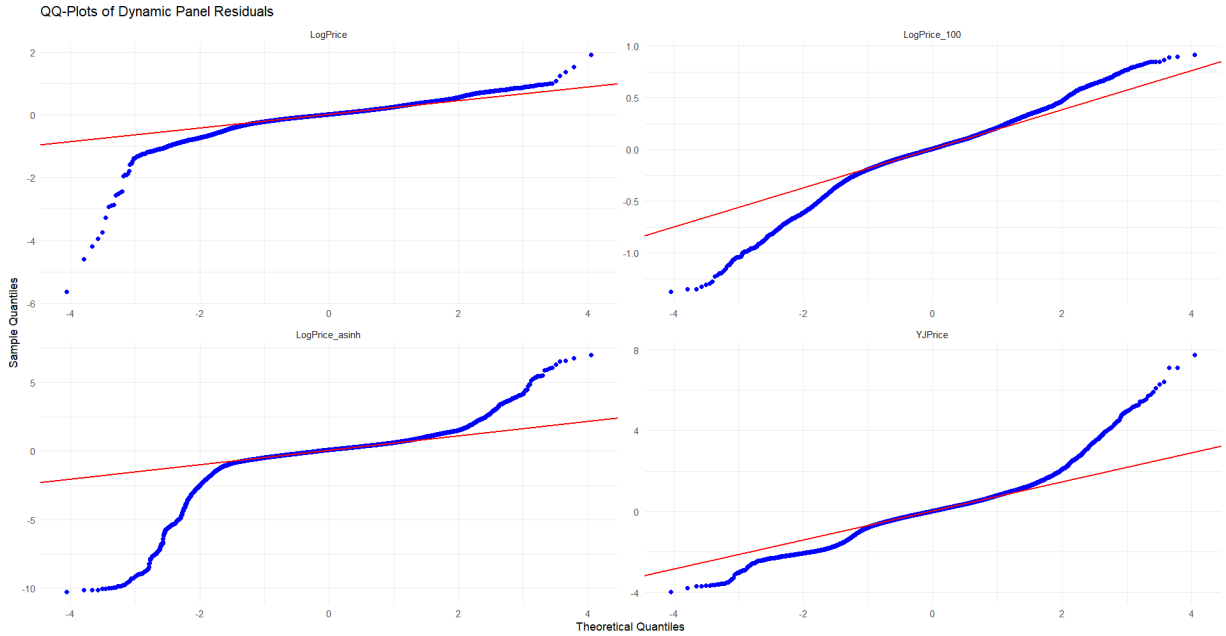


Figure A.5: QQ-plots comparing the residuals distribution of transformations of quarterly spot prices: logarithmic transformation with $c = 1$, logarithmic transformation with $c = 100$, inverse hyperbolic sine transformation, and Yeo–Johnson transformation, for the dynamic models residuals.

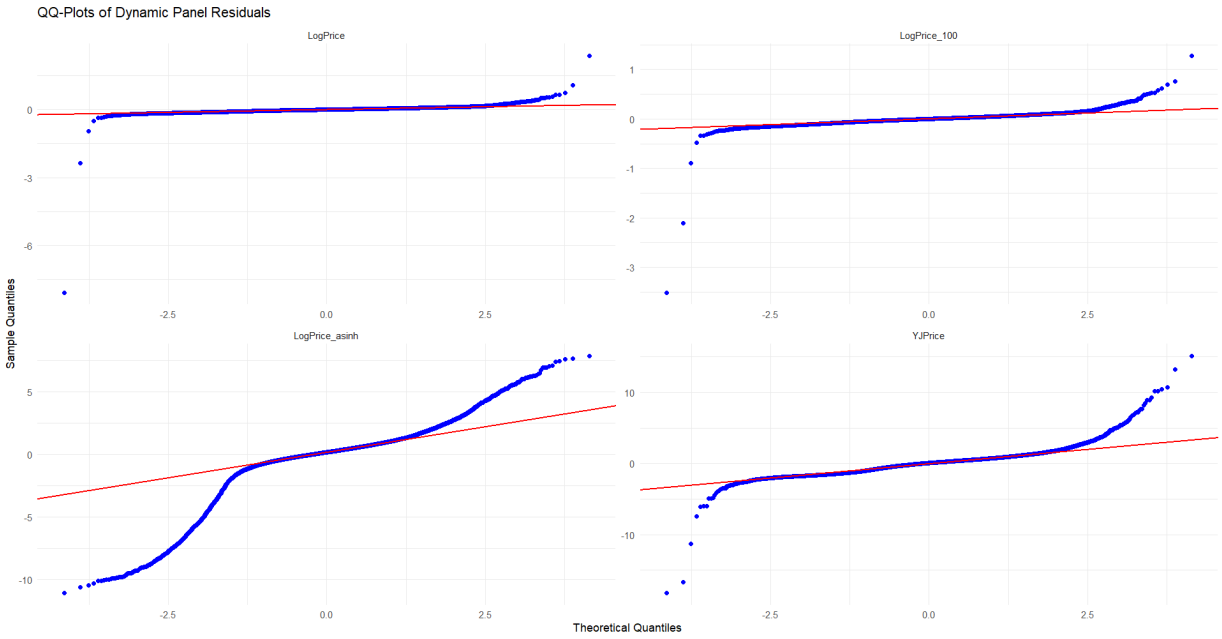


Figure A.6: QQ-plots comparing the residuals distribution of transformations of hourly spot prices: logarithmic transformation with $c = 1$, logarithmic transformation with $c = 100$, inverse hyperbolic sine transformation, and Yeo–Johnson transformation, for the dynamic models residuals.

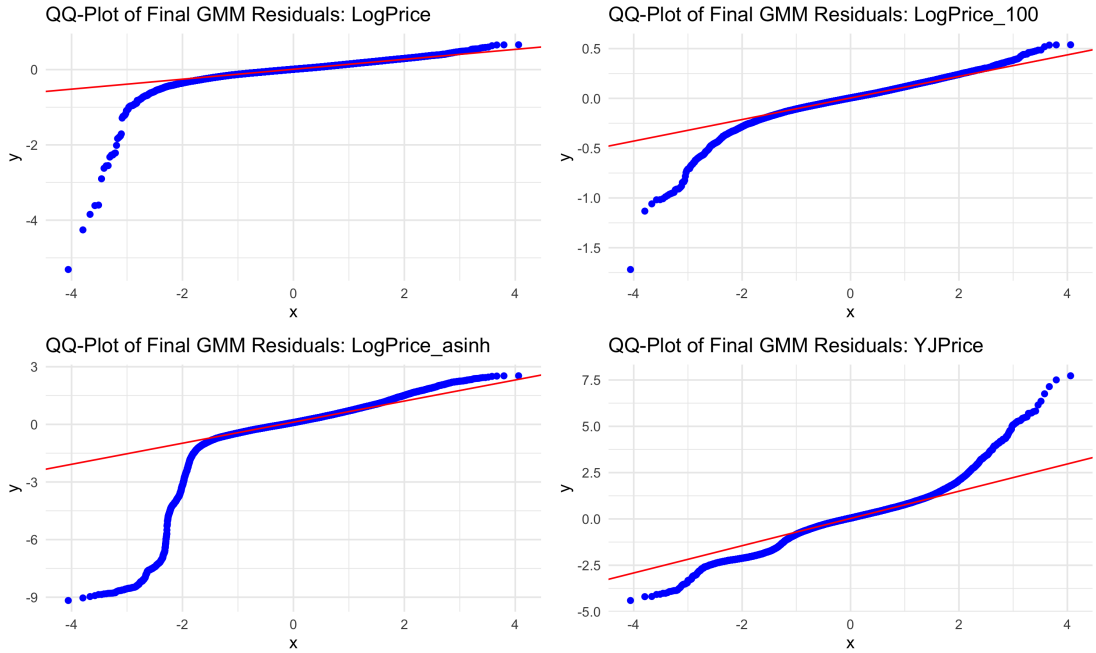


Figure A.7: QQ-plots comparing the residuals distribution of transformations of quarterly spot prices: logarithmic transformation with $c = 1$, logarithmic transformation with $c = 100$, inverse hyperbolic sine transformation, and Yeo–Johnson transformation.

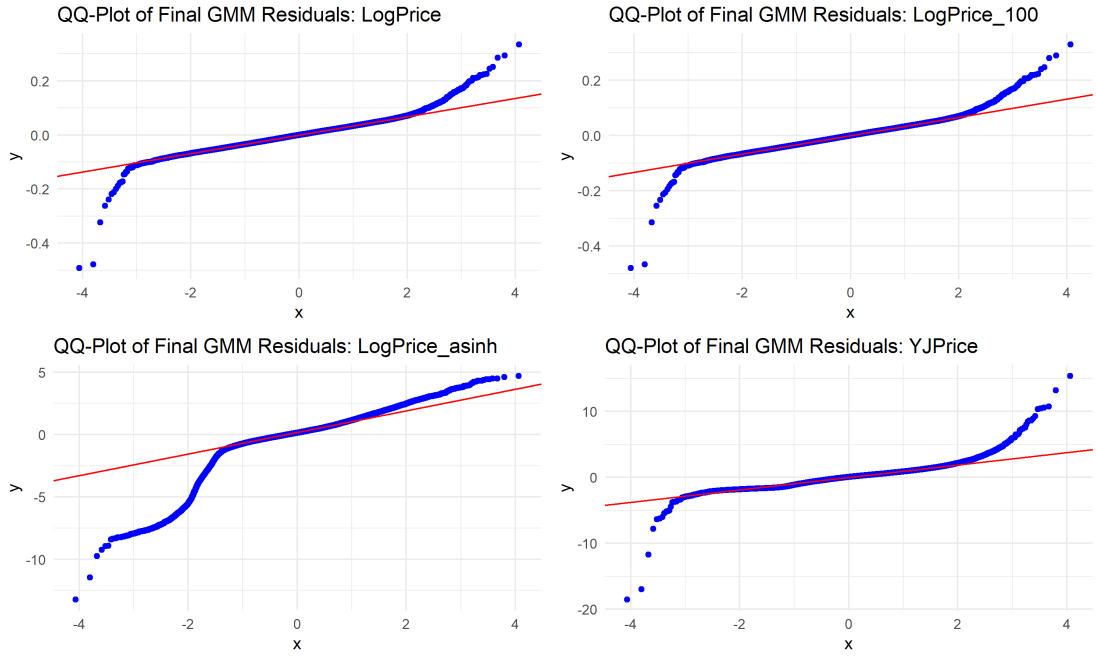


Figure A.8: QQ-plots comparing the residuals distribution of transformations of hourly spot prices: logarithmic transformation with $c = 1$, logarithmic transformation with $c = 100$, inverse hyperbolic sine transformation, and Yeo–Johnson transformation.

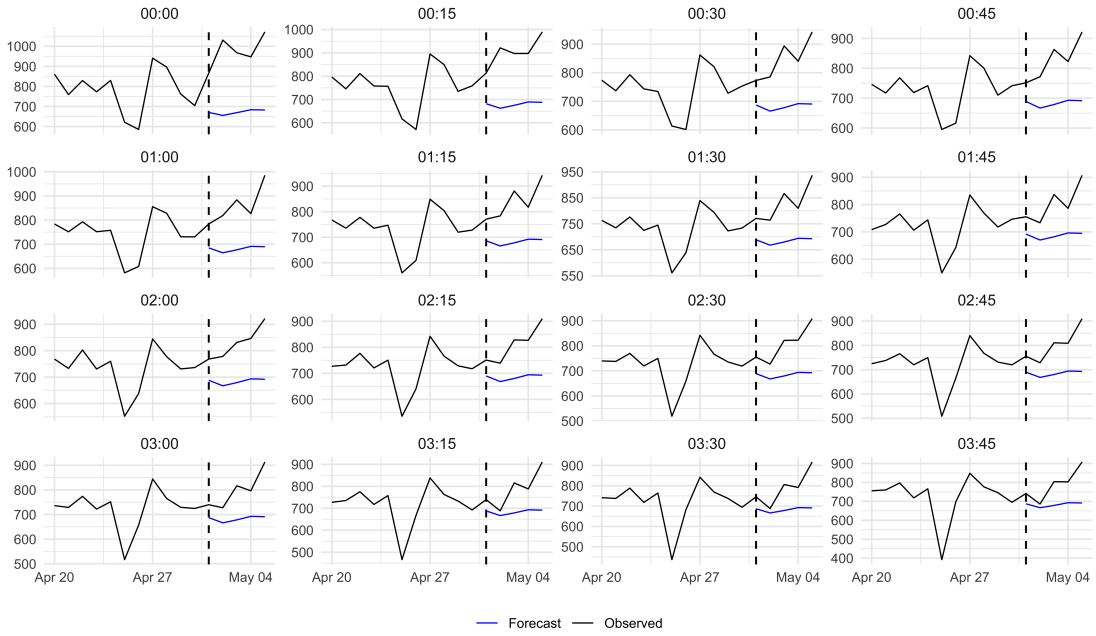


Figure A.9: Multi-step-ahead forecasts of the transformed spot price for quarters Q0–15.

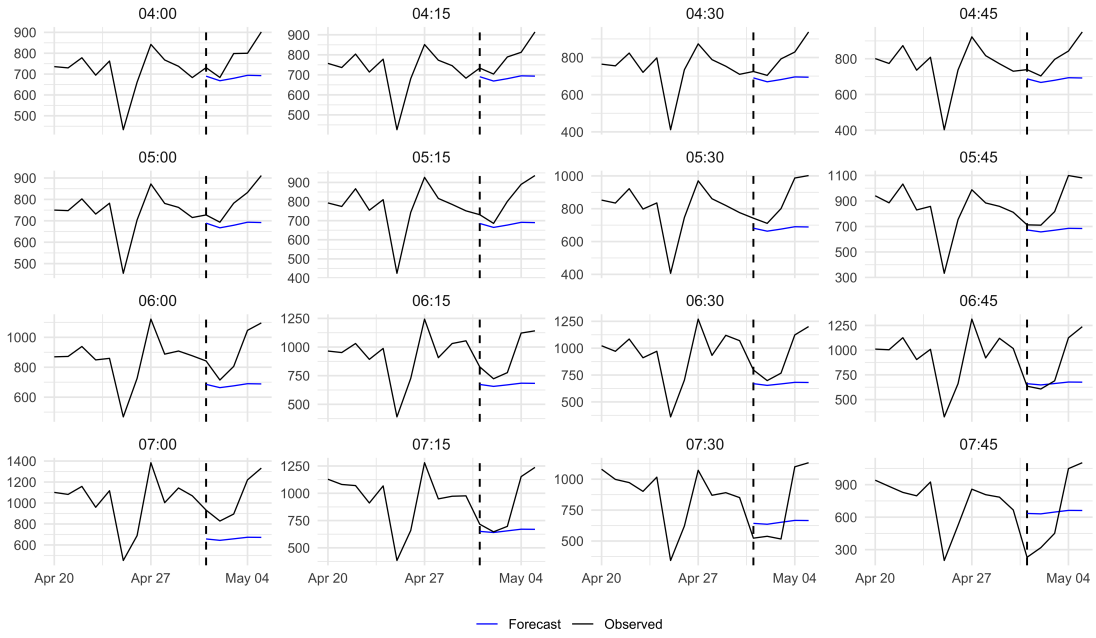


Figure A.10: Multi-step-ahead forecasts of the transformed spot price for quarters Q16–31.

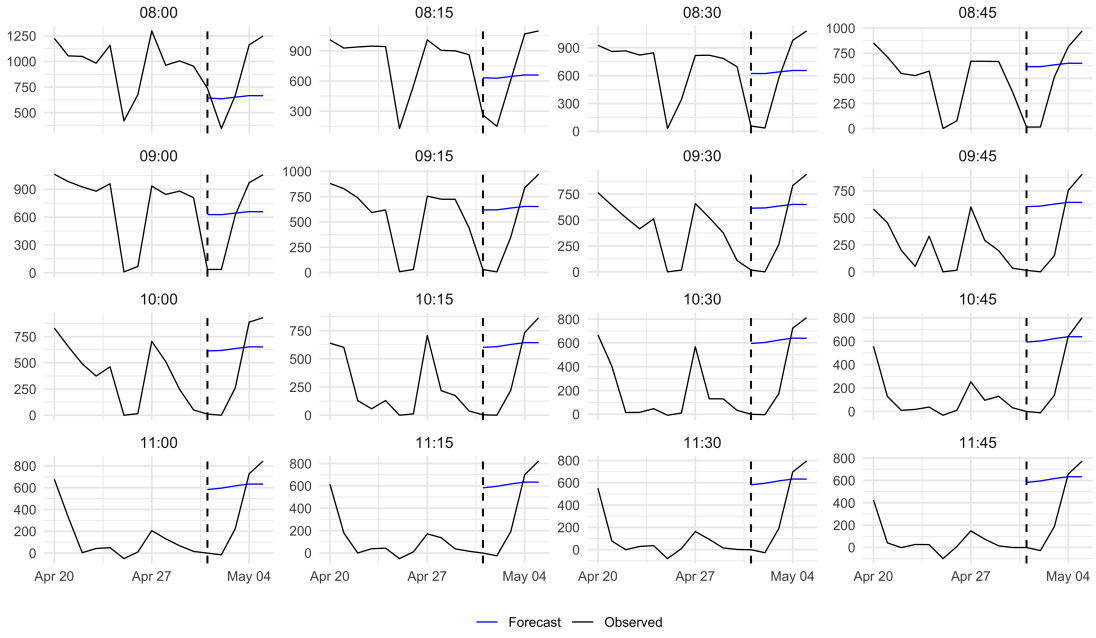


Figure A.11: Multi-step-ahead forecasts of the transformed spot price for quarters Q32–47.

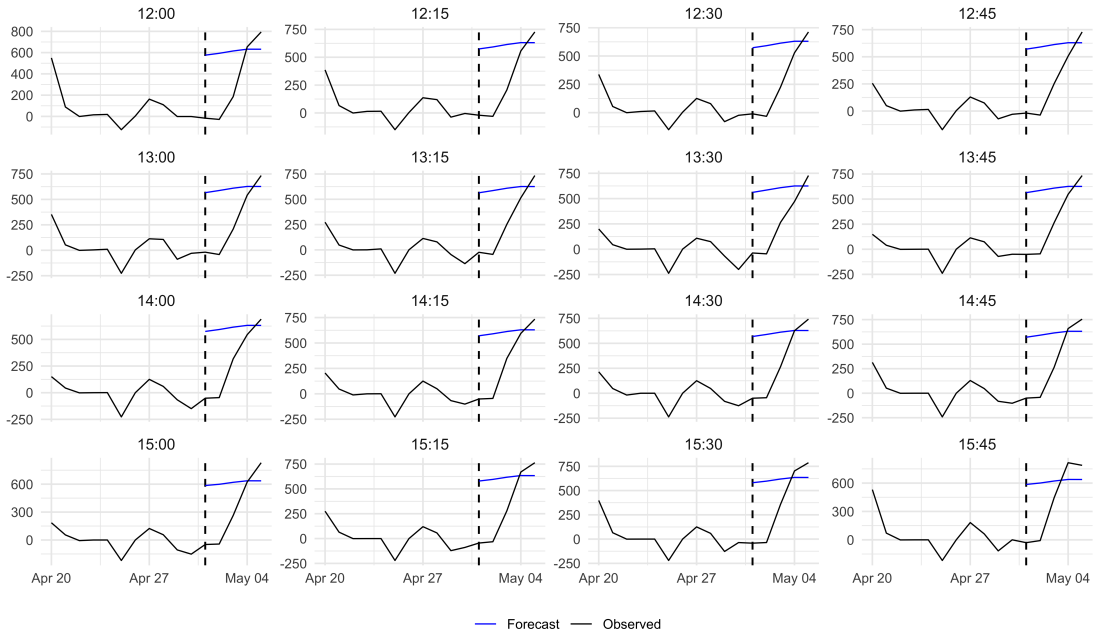


Figure A.12: Multi-step-ahead forecasts of the transformed spot price for quarters Q48–63.

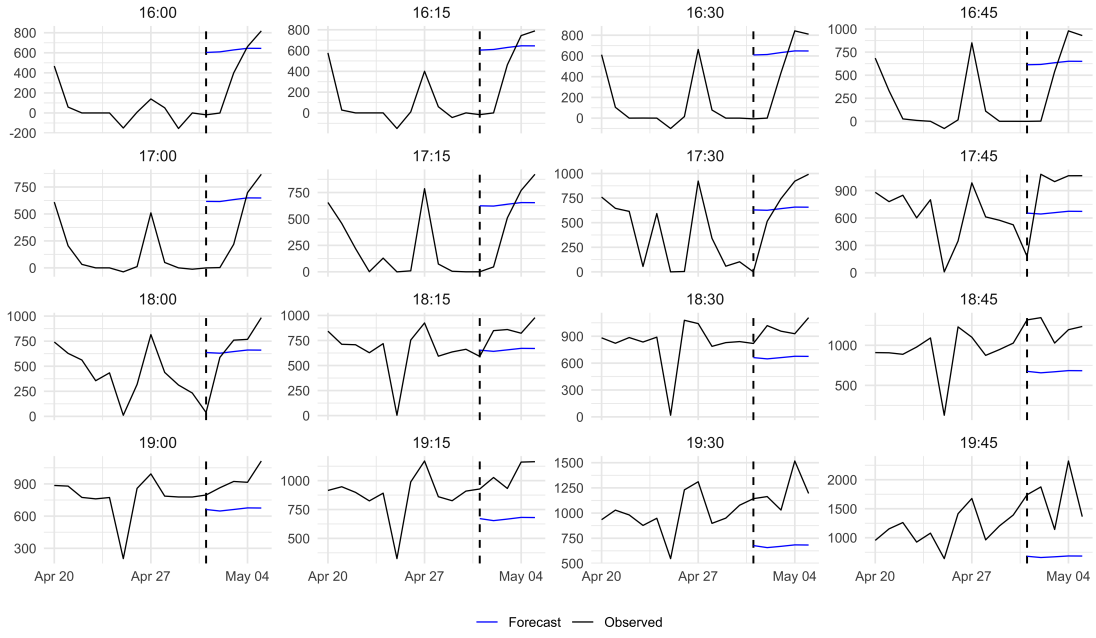


Figure A.13: Multi-step-ahead forecasts of the transformed spot price for quarters Q64–79.

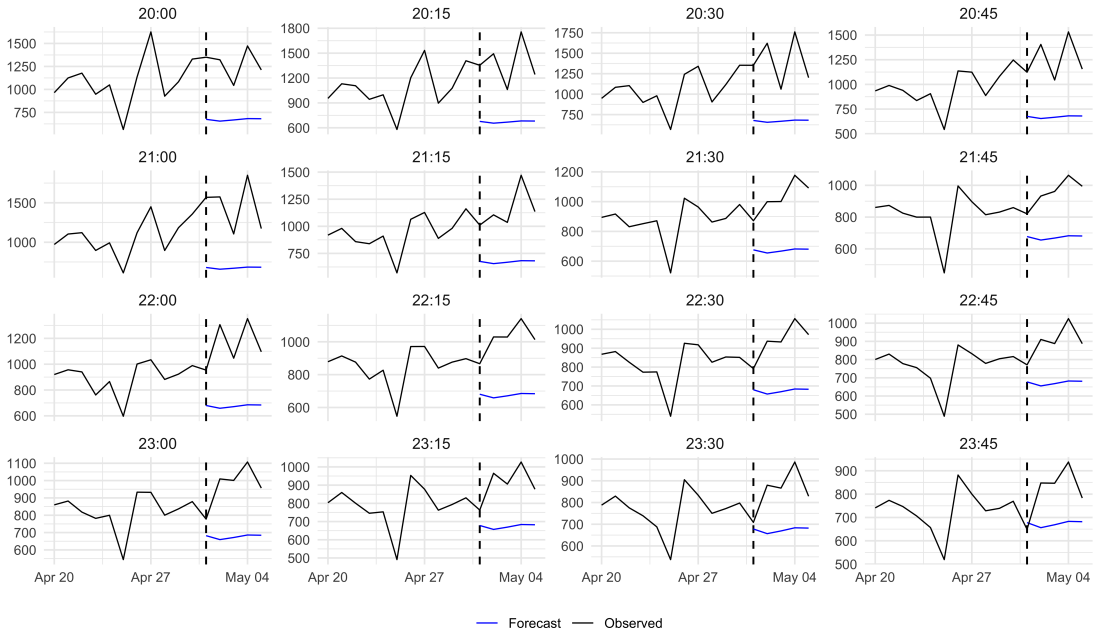


Figure A.14: Multi-step-ahead forecasts of the transformed spot price for quarters Q80–95.

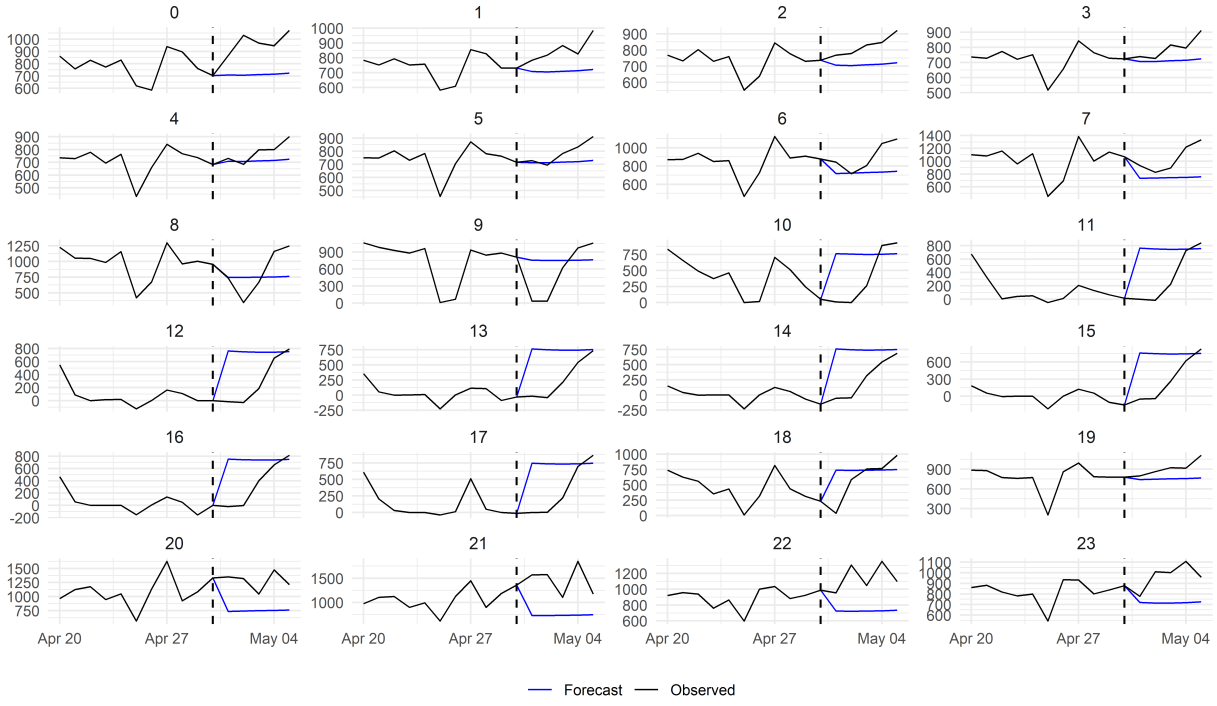


Figure A.15: Forecast of hourly prices with Log_100 transformation split into hours modelled using factor approach.