International Journal of Forecasting 41 (2025) 361–376

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Sparse time-varying parameter VECMs with an application to modeling electricity prices☆

Niko Hauzenberger [a,*], Michael Pfarrhofer [b], Luca Rossini [c,d]

[a] University of Strathclyde, United Kingdom
[b] Vienna University of Economics and Business, Austria
[c] University of Milan, Italy
[d] Fondazione Eni Enrico Mattei, Italy

ARTICLE INFO

Keywords:
Cointegration
Reduced rank regression
Sparsification
Hierarchical shrinkage priors
Error correction models

ABSTRACT

In this paper we propose a time-varying parameter (TVP) vector error correction model (VECM) with heteroskedastic disturbances. We propose tools to carry out dynamic model specification in an automatic fashion. This involves using global–local priors and postprocessing the parameters to achieve truly sparse solutions. Depending on the respective set of coefficients, we achieve this by minimizing auxiliary loss functions. Our two-step approach limits overfitting and reduces parameter estimation uncertainty. We apply this framework to modeling European electricity prices. When considering daily electricity prices for different markets jointly, our model highlights the importance of explicitly addressing cointegration and nonlinearities. In a forecasting exercise focusing on hourly prices for Germany, our approach yields competitive metrics of predictive accuracy.

1. Introduction

This paper discusses econometric tools to achieve dynamic model specification for vector error correction models (VECMs) automatically. Our main idea is to start with a suitably flexible and sophisticated specification amongst this model class and to impose data-driven shrinkage on the parameter space to obtain the simplest adequate nested version. This approach is motivated by our applications, where no clear theoretical guidance is available about how to choose crucial modeling aspects deterministically. Using a very general model guards against underfitting and misspecification while pushing its parameters toward a simpler sparsified solution to avoid overfitting and poor out-of-sample (OOS) predictive performance.

Specifically, we propose a time-varying parameter (TVP) VECM with heteroskedastic errors and apply it to model and forecast European electricity prices.[1] Deregulation and increasingly competitive markets in the power

☆ We thank Pierre Pinson, the Associate Editor, and two anonymous referees, as well as Florian Huber, Matteo Iacopini, Michael McCracken, Francesco Ravazzolo, and Trevor Cory for valuable comments and suggestions. Hauzenberger and Pfarrhofer acknowledge financial support from the Austrian Science Fund (FWF, grant no. ZK 35) and the Oesterreichische Nationalbank, Austria (OeNB, Jubiläumsfonds, projects 18763 and 18765).
The numerical results presented in this manuscript were reproduced by the Editor-in-Chief on 8 September 2024.

* Corresponding author.
E-mail address: niko.hauzenberger@strath.ac.uk (N. Hauzenberger).

[1] TVPs can be useful for improving predictive accuracy in many contexts (see, *e.g.*, Yousuf & Ng, 2021), because they can capture parameter instability from diverse sources. Combining TVPs with homoskedastic errors poses the risk of erroneously attributing changes in the volatility of the shocks to conditional mean dynamics, which is why such specifications are typically equipped with heteroskedastic errors. Indeed, heteroskedasticity has been shown to be a useful model feature in the context of modeling electricity prices; see Gianfreda, Ravazzolo, and Rossini (2023).

https://doi.org/10.1016/j.ijforecast.2024.09.001

sector have led to a surge of interest in statistical methods for modeling and forecasting electricity demand and price dynamics. Competing approaches include both univariate and multivariate time series models in linear and nonlinear settings (*e.g.*, structural breaks in the conditional means and variances). For an overview of the related literature, see Weron (2014). Most directly related to our approach, De Vany and Walls (1999) show that electricity prices in US states are cointegrated, with long-run relationships driven by no-arbitrage conditions. The more recent literature also finds evidence in favor of common dynamics and cointegration between electricity and gas prices for major power exchanges in the European Union (see, among others, Bello & Reneses, 2013; Bosco, Parisio, Pelagatti, & Baldi, 2010; de Marcos, Reneses, & Bello, 2016; Gianfreda, Parisio, & Pelagatti, 2019; Houllier & Menezes, 2012).

We aim to capture these empirical features and regularities in energy markets and model them explicitly. This motivates our proposed TVP-VECM. However, estimating VECMs, particularly with TVPs, poses several econometric challenges (see, *e.g.*, Koop, Leon-Gonzalez, & Strachan, 2011). First, even when it is agreed upon that cointegration should be taken into account, there is often no compelling argument about how a set of (economic) variables are cointegrated (especially in higher dimensions). This complicates introducing reasonable restrictions to identify the long-run behavior of such time series. For electricity prices, this aspect is even more apparent, since there is no clear intuition from economic theory about how to restrict the cointegration space a priori. Second, and relatedly, the cointegration rank is unknown and may be subject to change over time depending on the application. The previous literature often conditions on the rank and then compares measures of model fit ex post (see Geweke, 1996). This may be impractical, either due to computational limits or due to varying the rank requiring additional (possibly ad hoc) identifying restrictions.[2] Third, due to their flexibility, large TVP models are prone to overfitting. Many papers thus propose to restrict the parameter space by relying on hierarchical prior distributions or approximations.[3]

Our approach to solving these interrelated issues combines several recent econometric techniques used for large-scale TVP models and reduced rank regressions. We employ continuous global–local priors for pushing the parameter space towards sparsity. Prüser (2023) is a related paper with a similar approach to dealing with shrinkage in (constant parameter) VECMs. However, as noted by Chakraborty et al. (2020), such priors solely achieve approximate zeroes; the probability of observing exact zeroes is zero. In simple terms, this implies that shrinkage may guide a general model specification towards a simpler one, but it cannot achieve the nested version exactly (in terms of model selection).

As a remedy, we postprocess our posterior by minimizing distinct least absolute shrinkage and selection operator (LASSO)-type loss functions to obtain truly sparse estimates that may feature exact zeroes. The choices of loss functions are due to different implications of varying sparsity patterns across partitions of the parameter space. In particular, we propose to use distinct loss functions for the cointegration matrix (grouped LASSO, to exploit column-wise group structures for matrix rank selection), the autoregressive parameters (element-wise LASSO), and the covariances (graphical LASSO). A key feature of our framework is that it selects the number of cointegration relationships for each period, limiting the need to impose ad hoc restrictions a priori. Moreover, sparsifying the coefficients ex post alleviates overfitting concerns and reduces parameter estimation uncertainty (see also Huber, Koop, & Onorante, 2021).

We apply our model framework in two different energy-related contexts. The first uses daily data from several European markets jointly. Our approach detects distinct patterns in both dynamic and static interdependencies across markets. That is, the procedure selects relevant relationships in a multicountry context; see, *e.g.*, Feldkircher, Huber, Koop, & Pfarrhofer, 2022. The results corroborate previous empirical evidence about the importance of addressing cointegration, nonlinearities, and heteroskedasticity. In our second application, we conduct an extensive OOS forecast exercise for hourly electricity prices for Germany. In this case, our approach uncovers and/or excludes intra-day relationships between energy prices within a single country. We consider forecasts for each hour of the following day and find that multivariate cointegration models with TVPs and heteroskedastic errors provide improvements relative to various simpler benchmarks. In particular, the forecast exercise indicates that our proposed sparsified TVP-VECM yields competitive and in many cases superior forecasts for German hourly electricity prices.

Summarizing, the VECM allows for discriminating between long-run equilibria and short-run adjustment dynamics, which we find to be important for modeling electricity prices. TVPs capture structural breaks in the dynamic relationships of the underlying data. This is particularly useful when addressing complex latent pricing mechanisms and the varying importance of variables such as fuel prices that may be subject to change over time. Moreover, our shrink-then-sparsify approach allows for specification searches and variable selection in high-dimensional data by imposing exact zeroes in the coefficient matrices. Heteroskedastic errors capture large unanticipated shocks in prices, a crucial feature when interest centers on producing accurate density forecasts, which corroborates the findings of Gianfreda et al. (2023).

The paper proceeds as follows. Section 2 presents our flexible TVP-VECM model equipped with global–local shrinkage priors and heteroskedastic errors, while Section 3 discusses our proposed dynamic sparsification techniques. Section 4 applies the TVP-VECM to modeling European electricity prices. Section 5 summarizes and concludes the paper.

[2] Notable exceptions are Jochmann and Koop (2015) and Chua and Tsiaplias (2018), who use regime-switching models to estimate a time-varying cointegration rank.

[3] These three aspects are discussed to a varying extent in, *e.g.*, Bunea, She, and Wegkamp (2012), Chakraborty, Bhattacharya, and Mallick (2020), Eisenstat, Chan, and Strachan (2016), Hauzenberger, Huber, and Koop (2024), Hauzenberger, Huber, Pfarrhofer, and Zörner (2021), Huber, Koop, and Pfarrhofer (2020), Huber and Zörner (2019), Jochmann, Koop, Leon-Gonzalez, and Strachan (2013).

2. Time-varying parameter vector error correction models

We begin by introducing our baseline econometric framework using rather general notation. This reflects the notion that while these methods are developed in light of our applications, they are also applicable in other contexts. Let $\boldsymbol{y}_t$ be an $M \times 1$ vector of endogenous variables for $t = 1, \dots, T$, and denote the first difference operator by Δ such that $\Delta \boldsymbol{x}_t = \boldsymbol{x}_t - \boldsymbol{x}_{t-1}$.

A general specification of the TVP-VECM is:

$$\Delta \boldsymbol{y}_t = \boldsymbol{\Pi}_t \boldsymbol{w}_t + \sum_{p=1}^{P} \boldsymbol{A}_{pt} \Delta \boldsymbol{y}_{t-p} + \boldsymbol{\gamma}_t \boldsymbol{c}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t). \tag{1}$$

Here, $\boldsymbol{w}_t = (\boldsymbol{y}'_{t-1}, \boldsymbol{f}'_{t-1})'$, where $\boldsymbol{y}_{t-1}$ is the first lag of the endogenous variables, and $\boldsymbol{f}_{t-1}$ denotes a set of q_f exogenous factors such that $\boldsymbol{w}_t$ is of size $q \times 1$ with $q = M + q_f$.

The left-hand side of Eq. (1) is integrated of order zero, which we denote as $I(0)$; this in turn requires the product $\boldsymbol{\Pi}_t \boldsymbol{w}_t$ to be $I(0)$. Assuming unit roots of the endogenous variables in levels, this implies that $\boldsymbol{\Pi}_t$ is an $M \times q$ matrix of reduced rank. The rank r_t reflects the number of linearly independent cointegrating relationships, with $r_t < M$. $\boldsymbol{A}_{pt}$ refers to an $M \times M$ time-varying coefficient matrix related to the pth lag $\Delta \boldsymbol{y}_{t-p}$, and $\boldsymbol{\gamma}_t$ of size $M \times N$ relates an $N \times 1$ vector $\boldsymbol{c}_t$ of deterministic terms to $\Delta \boldsymbol{y}_t$. In our baseline version of the model, we assume a zero-mean Gaussian error term $\boldsymbol{\epsilon}_t$ with an $M \times M$ time-varying covariance matrix $\boldsymbol{\Sigma}_t$.

2.1. Cointegration matrix

A more thorough discussion of the matrix $\boldsymbol{\Pi}_t$ of reduced rank r_t, governing the cointegration relationships, is in order. In most applications using VECMs, $r_t = \bar{r}$ is some fixed (time-invariant) integer with $1 \leq \bar{r} \leq (M-1)$. The rank order is commonly motivated either based on economic theory (see, *e.g.*, Giannone, Lenza, & Primiceri, 2019), determined by calculating marginal likelihoods for a set of possible choices (see, *e.g.*, Geweke, 1996). For large-scale models, these approaches are often computationally prohibitive and rather restrictive.[4] As a solution, we adapt the approach of Chakraborty et al. (2020) for the TVP-VECM and estimate r_t for each period.

It is convenient to consider a reparameterized version of Eq. (1), where $\boldsymbol{\Pi}_t = \boldsymbol{\alpha}_t \boldsymbol{\beta}'$; see also Liu and Wu (1999). Here, the short-run adjustment coefficients are collected in $\boldsymbol{\alpha}_t$ of dimension $M \times q$, and the long-run relationships are captured by $\boldsymbol{\beta}$ which is $q \times q$. Note that we follow Yang and Bauwens (2018) and assume the long-run relations to be constant over time.[5] This amounts to the assumption that long-run fundamental relations do not change over time, since our interest centers on the combined matrix $\boldsymbol{\Pi}_t$, where nonlinearities appear through $\boldsymbol{\alpha}_t$, which is sometimes also referred to as a loadings matrix.

We refrain from restricting the cointegration space by imposing a deterministic structure on $\boldsymbol{\beta}$ (see also Strachan, 2003; Strachan & Inder, 2004; Villani, 2006). Instead, we follow Koop, León-González, and Strachan (2009), Koop et al. (2011) and use the transformations $\tilde{\boldsymbol{\alpha}}_t = \boldsymbol{\alpha}_t \boldsymbol{\zeta}^{-1}$, $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}\boldsymbol{\zeta}$, with $\boldsymbol{\zeta} = (\tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\beta}})^{-0.5}$. This allows for employing a linear state-space modeling approach assuming conditional Gaussianity with a cointegration space prior.

2.2. Time-varying parameters and shrinkage

Using $\tilde{\boldsymbol{w}}_t = \tilde{\boldsymbol{\beta}}' \boldsymbol{w}_t$ and $\boldsymbol{z}_t = (\tilde{\boldsymbol{w}}'_t, \boldsymbol{x}'_t)'$, a more compact version of Eq. (1), providing notational simplicity, is given by:

$$\Delta \boldsymbol{y}_t = \boldsymbol{B}_t \boldsymbol{z}_t + \boldsymbol{L}_t \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}_t), \tag{2}$$

with $\boldsymbol{A}_t = (\boldsymbol{A}_{1t}, \dots, \boldsymbol{A}_{Pt}, \boldsymbol{\gamma}_t)$, $\boldsymbol{B}_t = (\tilde{\boldsymbol{\alpha}}_t, \boldsymbol{A}_t)$ and $\boldsymbol{x}_t = (\Delta \boldsymbol{y}'_{t-1}, \dots, \Delta \boldsymbol{y}'_{t-P}, \boldsymbol{c}'_t)'$, where $\boldsymbol{A}_t$ is of size $M \times J$ and $\boldsymbol{x}_t$ of size $J \times 1$ with $J = (MP + N)$. Furthermore, it is convenient to factor $\boldsymbol{\Sigma}_t = \boldsymbol{L}_t \boldsymbol{H}_t \boldsymbol{L}'_t$, with a diagonal matrix $\boldsymbol{H}_t = \text{diag}(\exp(h_{1t}), \dots, \exp(h_{Mt}))$ and $\boldsymbol{L}_t$ denoting the normalized lower Cholesky factor (*i.e.*, a lower triangular matrix with ones on its diagonal). Note that $\text{Var}(\boldsymbol{L}_t \boldsymbol{\eta}_t) = \text{Var}(\boldsymbol{\epsilon}_t)$; and this triangularization of the multivariate system allows for equation-by-equation estimation (see, *e.g.*, Carriero, Chan, Clark, & Marcellino, 2022; Primiceri, 2005).[6]

We select the ith row of $\boldsymbol{B}_t$, and define $\boldsymbol{b}_{it} = \boldsymbol{B}'_{i\bullet,t}$, which refers to the parameters of the ith equation of the VECM. In addition, we stack all free elements of the matrix $\boldsymbol{L}_t$ in a vector $\boldsymbol{l}_t$. Following the related literature, we then assume a random walk law of motion for these TVPs:[7]

$$\begin{aligned} \boldsymbol{b}_{it} &= \boldsymbol{b}_{it-1} + \boldsymbol{\vartheta}_{it}, \quad \boldsymbol{\vartheta}_{it} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Theta}_{(b)i}), \\ \boldsymbol{l}_t &= \boldsymbol{l}_{t-1} + \boldsymbol{\vartheta}_t, \quad \boldsymbol{\vartheta}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Theta}_{(l)}). \end{aligned} \tag{3}$$

[4] Notable exceptions are Huber and Zörner (2019) and Prüser (2023), who use global–local shrinkage priors to estimate cointegration relations in a data-driven manner.

[5] Assuming both $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ to vary over time further complicates achieving identification. First, note that $\boldsymbol{\Pi}_t = \boldsymbol{\alpha}_t \boldsymbol{\beta}'_t = \boldsymbol{\alpha}_t \boldsymbol{Q} \boldsymbol{Q}^{-1} \boldsymbol{\beta}'_t$ for any non-singular matrix $\boldsymbol{Q}$, which results in the so-called global identification problem. It is common in the literature to use linear normalization schemes such as $\boldsymbol{\beta}_t = (\boldsymbol{I}_{r_t}, \boldsymbol{\beta}'_t)'$; see also Villani (2001) or Strachan (2003). Second, the local identification problem appears for $\boldsymbol{\alpha}_t = \boldsymbol{0}$, which implies that $\text{rank}(\boldsymbol{\Pi}_t) = 0$ (see Kleibergen & Van Dijk, 1994, 1998; Paap & Van Dijk, 2003).

[6] We use the structural form of the VAR merely as a technical device, and do not aim to identify any underlying primitive shocks (where the ordering would impose specific timing restrictions). Potential order-invariance of the system is due to the implied prior on the reduced-form coefficients; see Chan (2022). This poses an issue in large-scale systems. In our empirical application, we use moderately sized models, and limited experiments permuting the order of the variables in the dependent vector show that this has negligible effects on our estimates.

[7] Hauzenberger (2021) assesses different laws of motion for such state equations, and finds that random walks typically provide sufficient flexibility capable of tracing most distinct patterns.

The state innovation variances, which govern the amount of time variation, are collected in diagonal covariance matrices $\boldsymbol{\Theta}_{(b)i}$ and $\boldsymbol{\Theta}_{(l)}$. To impose shrinkage, we use the non-centered parameterization of the TVP model (see Frühwirth-Schnatter & Wagner, 2010, for details), which splits the parameters into a constant and time-varying part:

$$\boldsymbol{b}_{it} = \boldsymbol{b}_{i0} + \sqrt{\boldsymbol{\Theta}}_{(b)i}\tilde{\boldsymbol{b}}_{it}, \quad \tilde{\boldsymbol{b}}_{it} = \tilde{\boldsymbol{b}}_{it-1} + \tilde{\boldsymbol{\vartheta}}_{it}, \quad \tilde{\boldsymbol{\vartheta}}_{it} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \quad \tilde{\boldsymbol{b}}_{i0} = \boldsymbol{0},$$

for $i = 1, \dots, M$, with an analogously transformed state equation for the parameters $\boldsymbol{l}_t$. This allows for imposing shrinkage on the constant part of the coefficients $\boldsymbol{b}_{i0}$, with the jth element $b_{ij,0}$ and the amount of time variation determined by $\sqrt{\theta}_{(b)ij}$, the jth diagonal element of $\sqrt{\boldsymbol{\Theta}}_{(b)i}$.

The sparsification methods proposed in this paper (see Section 3) may be combined with any desired setup from the class of global–local shrinkage priors; see Cadonna, Frühwirth-Schnatter, and Knaus (2020) for a review. The respective parameters—in our case, the constant part of the coefficients and the square root of the state innovation variances—are assumed to follow a Gaussian distribution with zero mean, and a global variance parameter pushes all coefficients strongly towards zero. These global parameters are multiplied with local scalings that can pull prior mass away from zero for specific parameters even in cases where the underlying parameter vector is very sparse. Both of these shrinkage factors are equipped with another prior hierarchy, and shrinkage properties arise from choices about these mixing distributions. From this class of priors, we choose the horseshoe of Carvalho, Polson, and Scott (2010) for its lack of prior tuning parameters and excellent shrinkage properties.

2.3. Heteroskedastic errors

To address possible heteroskedasticity, we use stochastic volatility (SV) models for the structural errors. We follow the literature (see, *e.g.*, Kastner & Frühwirth-Schnatter, 2014) and assume that the logarithms of the diagonal elements of $\boldsymbol{H}_t$ follow independent AR(1) processes:

$$h_{it} = \mu_i + \phi_i(h_{it-1} - \mu_i) + \varsigma_i \xi_{it}, \quad \xi_{it} \sim \mathcal{N}(0, 1), \tag{4}$$

where μ_i is the unconditional mean, ϕ_i is the persistence parameter, and ς_i is the error variance of the log-volatility process. In our applications, we find empirical evidence in favor of using stable processes instead of less densely parameterized state equations such as a random walk.

Note that $\boldsymbol{\eta}_t = (\eta_{1t}, \dots, \eta_{Mt})'$ in Eq. (2) features Gaussian errors. We also consider specifications where we replace this Gaussian with a t-distribution:

$$\eta_{it} \sim t_{\nu_i}(0, \exp(h_{it})),$$

which renders the model even more flexible. Note that as the degrees of freedom $\nu_i \to \infty$, we obtain a Gaussian as the limiting case. We place standard priors on μ_i, ϕ_i, ς_i, and on the degrees of freedom ν_i, in case we use t-distributed errors (see Kastner & Frühwirth-Schnatter, 2014). They are estimated alongside all other parameters. This completes our baseline model specification. Specifics about priors are provided in the appendix.

2.4. Sampling algorithm

The joint posterior of our model is analytically intractable. We thus employ a Markov chain Monte Carlo (MCMC) sampling algorithm to draw from it, and these draws can then be used for posterior and predictive inference. Fortunately, our proposed model structure and prior choices translate into individual sampling steps that are fairly straightforward and well understood. Specifically, our sampling algorithm consists mostly of Gibbs updates:

1. Writing the multivariate system in triangular form allows for updates of the covariances and conditional mean parameters equation-by-equation. Conditional on all other parameters and the TVPs, the constant part of the coefficients and the state innovation variances can be drawn from their conditionally conjugate Gaussian posterior distribution with well-known moments. In fact, these take the form of conditional linear regression models with standard normal errors; see Chan, Koop, Poirier, and Tobias (2020) for a textbook treatment.
2. The TVPs can subsequently be obtained by using a straightforward forward-filter backward-sampling (FFBS) algorithm conditional on all other parameters; see Carter and Kohn (1994), Frühwirth-Schnatter (1994) and specifically Frühwirth-Schnatter and Wagner (2010).
3. The draws for the time-invariant parameters in the conditional mean can then be used to update the global and local shrinkage factors for the horseshoe prior to using the auxiliary representation of Makalic and Schmidt (2015), which involves sampling solely from inverse gamma distributions.
4. The transformed cointegration vectors conditional on all other parameters follow a multivariate Gaussian posterior distribution with standard moments which can again be found in any Bayesian textbook.
5. The SVs and the parameters of their state equations are sampled using the `R` package `stochvol` (Kastner, 2016). This step also involves drawing the parameters that govern the SV state equations, and the degrees-of-freedom parameter for the t-distribution in the case of heavy-tailed errors.
6. Sparsification is performed for each draw as outlined in detail in Section 3 and Ray and Bhattacharya (2018). This implies that we can run the respective specification once and retain draws from the non-sparsified posterior. These may subsequently be postprocessed with dynamic sparsification methods and used to compute predictions outside the MCMC algorithm.

Standard MCMC diagnostics point towards satisfactory convergence properties of this algorithm. In our empirical work, we use a burn-in period of 3000 draws and retain each third of the remaining 9000 draws. This yields a set of 3000 draws for postprocessing and inference. We use the posterior or predictive median as a Bayesian point estimate (see chapters 3–5 in Chan et al., 2020, for further

details), if applicable. For our most demanding specification (with more than 200,000 parameters to estimate), estimation using the proposed algorithm (which is implemented in the software R) takes about 2.5 min per 100 draws on a cluster with Intel E5-2650v3 2.3 GHz cores.

3. Dynamic sparsification

The model proposed above combined with continuous global–local priors yields posterior draws that are pushed towards approximate sparsity.[8] We rely on ex post sparsification of these draws for each point in time to obtain exact sparsity. In cointegration models, particularly with TVPs, it is beneficial to adjust loss functions for specific parts of the parameter space. This is due to the subtle implications of these blocks of parameters for the overall model structure.

3.1. Designing suitable loss functions

To perform variable selection and obtain sparse coefficient matrices we postprocess $\boldsymbol{\Pi}_t$, $\boldsymbol{A}_t$, and $\boldsymbol{\Sigma}_t$ by minimizing three coefficient-type specific LASSO loss functions. In particular, we rely on methods proposed in Friedman, Hastie, and Tibshirani (2008), Hahn and Carvalho (2015), Chakraborty et al. (2020), Ray and Bhattacharya (2018), and Bashir, Carvalho, Hahn, and Jones (2019), which have been successfully used in a range of multivariate and univariate macroeconomic and finance applications (see Hauzenberger, Huber, & Onorante, 2021; Huber et al., 2021; Puelz, Hahn, & Carvalho, 2020; Puelz et al., 2017). Given the properties of the reduced rank matrix $\boldsymbol{\Pi}_t$, we propose modifications when compared to sparsification of the autoregressive coefficients, $\boldsymbol{A}_t$, and the covariance matrix $\boldsymbol{\Sigma}_t$.

As a general remark on notation, draws from the non-sparsified posteriors are indicated by a hat (*e.g.*, $\hat{\boldsymbol{\Pi}}_t$) and sparse estimates are marked with an asterisk (*e.g.*, $\boldsymbol{\Pi}_t^*$). The key difference between the two is that the non-sparsified posterior draws may have many entries close to but not exactly zero, while the sparsified estimate has exact zeros which are imposed using auxiliary loss functions. The general idea is that these loss functions are designed to reward model fit by minimizing a distance measure between the non-sparse and sparse solutions, while an additional tuning parameter penalizes non-zero parameters. Choosing this tuning parameter—which, loosely speaking, governs the number of zeroes—is generally not a trivial task. To avoid extensive pre-estimation tuning procedures, we use the signal adaptive variable selection (SAVS) estimator proposed by Ray and Bhattacharya (2018), which selects an appropriate amount of sparsity automatically.

In a TVP context, dynamic sparsification poses some additional challenges concerning how sparsity is imposed with respect to t:

1. Ex post sparsification is commonly applied to point estimators, such as the posterior median or mean. We deviate from this procedure and solve the respective optimization problem for each draw from the posterior distribution. The procedure of Woody, Carvalho, and Murray (2021) is closely related to this approach. They provide a theoretical motivation for conducting uncertainty quantification of sparse posterior estimates.
2. Our loss functions are defined in terms of full-data matrices instead of t-specific covariates. The latter might be considered the natural candidate when transforming a TVP model to its static representation (for details, see Chan & Jeliazkov, 2009; Hauzenberger, Huber, Koop, & Onorante, 2022). Commonly, sparsification is applied in standard regression frameworks with constant coefficients, implying that all information over time is considered. Using t-by-t draws results in dependence on a single observation in time t.

We illustrate these issues and our solutions in more detail below in the context of sparsifying the cointegration matrix. It is worth noting that these concerns apply to all three blocks of the parameters that we intend to dynamically sparsify.

3.2. Sparsifying the cointegration matrix

Our basic approach to sparsifying the cointegration relationships follows Chakraborty et al. (2020). Ex post sparsifying $\boldsymbol{\Pi}_t$ is of crucial importance to obtain an estimate for the rank. Using solely a global–local prior, some estimates in $\boldsymbol{\Pi}_t$ are pushed towards zero, but they are never exactly zero. In the case of the cointegration matrix, this implies that $\boldsymbol{\Pi}_t$ would always be of full rank (*i.e.*, $r = M$ for all t). The main goal is thus to minimize the predictive loss between a non-sparsified draw $\hat{\boldsymbol{\Pi}}_t$ and a column-sparse solution $\boldsymbol{\Pi}_t^*$. The natural choice for the respective loss function is thus to impose a group structure on the elements of this matrix. We provide details below.

Our interest centers on dynamic sparsification to obtain a sparse cointegration matrix at each point in time. The loss function is specified in terms of the full-data matrix $\boldsymbol{W}$, a $T \times q$ matrix with $\boldsymbol{w}_t'$ in the tth row:

$$\boldsymbol{\Pi}_t^{*'} = \min_{\boldsymbol{\Pi}_t} \left(\| \boldsymbol{W}\hat{\boldsymbol{\Pi}}_t' - \boldsymbol{W}\boldsymbol{\Pi}_t' \|_F^2 + \sum_{j=1}^{q} \kappa_{jt} \| \boldsymbol{\Pi}_{\bullet j,t} \|_2 \right), \quad (5)$$

with $\|\boldsymbol{C}\|_F$ denoting the Frobenius norm of a matrix $\boldsymbol{C}$, $\|\boldsymbol{c}\|_2$ the Euclidean norm of a vector $\boldsymbol{c}$, and $\boldsymbol{\Pi}_{\bullet j,t}$ referring to the jth column of $\boldsymbol{\Pi}_t$ (*i.e.*, the jth row of its transpose). Eq. (5) denotes a grouped LASSO problem with a row (column)-specific penalty κ_{jt} and aims at finding a row (column)-sparse solution of $\boldsymbol{\Pi}_t'$ ($\boldsymbol{\Pi}_t$).[9]

[8] As highlighted by Hahn and Carvalho (2015), the success of the two-step shrink-then-sparsify approach depends on the shrinkage properties of the prior. They note that the horseshoe prior is well suited for such procedures, which is another reason why we illustrate our proposed framework with this specific choice.

[9] The Frobenius norm is a common distance measure between subspaces and is given by $\|\boldsymbol{C}\|_F = \sqrt{\text{tr}(\boldsymbol{C}'\boldsymbol{C})}$ with $\text{tr}(\boldsymbol{C})$ denoting the trace of a matrix $\boldsymbol{C}$. For a detailed discussion on properties of the grouped LASSO, see Yuan and Lin (2006) and Wang and Leng (2008).

The first part controls the distance between an estimate and its sparse solution (measured by the Frobenius norm), while the second part penalizes non-zero elements in $\boldsymbol{\Pi}_t$ (in terms of column-specific Euclidean norms). We use the grouped LASSO as opposed to an element-wise LASSO to establish a loss function that penalizes the cointegration matrix towards a lower rank structure. It is worth noting that using an element-wise LASSO could yield situations where the penalty introduces spurious cointegration relationships. Summarizing in simple terms, this loss function is designed to obtain an adequate reduced rank estimate of $\boldsymbol{\Pi}_t$ that avoids sacrificing model fit relative to the full rank case.

Notice that we rely on the full data matrix $\boldsymbol{W}$ instead of the t-specific covariates $\boldsymbol{w}_t'$. This is in line with our state equations for the TVPs, where all information over time is used for filtering and smoothing. In constant parameter regressions, losses would be based on variation explained by $\boldsymbol{W}\boldsymbol{\Pi}'$. Using the static regression framework to perform dynamic sparsification and solely relying on the tth observation $\boldsymbol{w}_t'$, instead of the full data matrix $\boldsymbol{W}$, renders the penalty highly sensitive to individual observations over time. To illustrate this, we focus on the jth columns in both $\boldsymbol{W}_{\bullet j}$ and $\boldsymbol{\Pi}_{\bullet j}$. The norm of $\boldsymbol{W}_{\bullet j}$ is defined by $\|\boldsymbol{W}_{\bullet j}\|_2 = \sqrt{\sum_{t=1}^{T} w_{tj}^2}$. When using t-by-t estimates independently, the norm of the tth observation is given by $\|w_{tj}\|_2 = \sqrt{w_{tj}^2}$. A simple solution to this issue would be to down-weight the penalty in Eq. (6) by a factor T. However, although theoretically in line with the sparsification techniques proposed in Hahn and Carvalho (2015), this has the disadvantage of being exposed to idiosyncrasies of the tth observation which poses the risk of unstable penalties in Eq. (5). Therefore, using $\boldsymbol{W}$ is arguably the most practicable solution, where each t-specific estimate is used for the entire sample.

Note that Eq. (5) can be interpreted as minimizing the expected loss (see Hahn & Carvalho, 2015). Setting $\hat{\boldsymbol{\Pi}}_t = \boldsymbol{\Pi}_t^{(s)}$ with (s) indicating the sth MCMC draw (rather than a posterior point estimate), has the attractive feature of allowing for uncertainty quantification about the rank of $\boldsymbol{\Pi}_t$ (see also Hauzenberger, Huber, & Onorante, 2021; Huber et al., 2021).

In fact, the resulting reduced rank cointegration matrix can be used to extract a model-based estimate of the number of cointegration relationships. An estimate of this time-varying rank r_t is obtained using:

$$r_t = \sum_{i=1}^{M} \mathbb{I}\left(\mathfrak{s}_{it} > \varphi\right),$$

with $\mathbb{I}(\bullet)$ denoting the usual indicator function, and $\mathfrak{s}_{it}$ for $i = 1, \dots, M$, denoting the singular values of $\boldsymbol{W}\boldsymbol{\Pi}_t^{*\prime}$. We follow Chakraborty et al. (2020) and define the rank as the number of non-thresholded singular values, with φ defined as the largest singular value of the residuals of the full data specification of Eq. (1), which corresponds to the maximum noise level.

To avoid cross-validation for the penalty term (as in Hahn & Carvalho, 2015), we rely on the SAVS estimator proposed by Ray and Bhattacharya (2018) and set the penalty term to $\kappa_{jt} = 1/\|\hat{\boldsymbol{\Pi}}_{\bullet j,t}\|_2^2$. This yields the following soft threshold estimate:[10]

$$\boldsymbol{\Pi}_{\bullet j,t}^{*} = \begin{cases} \boldsymbol{0}_{M\times 1}, & \text{if } \frac{\kappa_{jt}}{2\|\hat{\boldsymbol{\Pi}}_{\bullet j,t}\|_2} \geq \|\boldsymbol{W}_{\bullet j}\|_2^2, \\ \left(1 - \frac{\kappa_{jt}}{2\|\boldsymbol{W}_{\bullet j}\|_2^2 \|\hat{\boldsymbol{\Pi}}_{\bullet j,t}\|_2}\right) \hat{\boldsymbol{\Pi}}_{\bullet j,t}, & \text{otherwise.} \end{cases} \tag{6}$$

As shown by Ray and Bhattacharya (2018), this choice of κ_{jt} has properties similar to the adaptive LASSO proposed by Zou (2006). To extend our discussion about specifying the penalty in terms of full data matrices with respect to SAVS, it is worth noting that the only part that changes is how the norm of the data is calculated ($\|\boldsymbol{W}_{\bullet j}\|_2^2$ instead of w_{jt}^2). In this case, the penalty specified by Chakraborty et al. (2020), $\kappa_{jt} = 1/\|\hat{\boldsymbol{\Pi}}_{\bullet j,t}\|_2^2$, can be used (without correcting for T). Note that this is a practicable solution from an applied perspective, since the norm over the full data matrix is more robust than considering a single observation in period t.

3.3. Sparsifying the autoregressive coefficients

For sparsifying the time-varying autoregressive coefficients, we define a full-data matrix $\boldsymbol{X}$ of dimension $T \times J$ with $\boldsymbol{x}_t'$ in the tth row, an $MJ \times 1$ vector $\boldsymbol{a}_t = \text{vec}(\boldsymbol{A}_t')$ and the corresponding $TM \times MJ$ regressor matrix $\tilde{\boldsymbol{X}} = (\tilde{\boldsymbol{x}}_1', \dots, \tilde{\boldsymbol{x}}_T')'$ with $\tilde{\boldsymbol{x}}_t = (\boldsymbol{I}_M \otimes \boldsymbol{x}_t')$.

We assume a loss function of the form:

$$\boldsymbol{a}_t^{*} = \min_{\boldsymbol{a}_t} \left(\frac{1}{2}\|\tilde{\boldsymbol{X}}\hat{\boldsymbol{a}}_t - \tilde{\boldsymbol{X}}\boldsymbol{a}_t\|_2^2 + \sum_{j=1}^{J} \delta_{jt}|a_{jt}| \right), \tag{7}$$

with a_{jt} denoting the jth element of $\boldsymbol{a}_t$, and δ_{jt} a covariate-specific penalty. It is worth noting that we minimize a standard LASSO-type predictive loss where the first part controls the distance between an estimate and a sparse solution and the second part penalizes non-zero elements in $\boldsymbol{a}_t$, unlike the sparsification of the cointegration relationships.

An optimal choice for the penalty is $\delta_{jt} = 1/(|\hat{a}_{jt}|^2)$. We again rely on the soft threshold estimate implied by SAVS to obtain a sparse draw of $\boldsymbol{a}_t$:

$$a_{jt}^{*} = \begin{cases} 0, & \text{if } \frac{\delta_{jt}}{|\hat{a}_{jt}|} \geq \|\tilde{\boldsymbol{X}}_{\bullet j}\|_2^2, \\ \left(1 - \frac{\delta_{jt}}{\|\tilde{\boldsymbol{X}}_{\bullet j}\|_2^2 |\hat{a}_{jt}|}\right) \hat{a}_{jt}, & \text{otherwise.} \end{cases} \tag{8}$$

As shown by Ray and Bhattacharya (2018), this choice of δ_{jt} again has properties similar to the adaptive LASSO proposed by Zou (2006).

[10] To solve the optimization problem in Eq. (5), the SAVS estimator can be interpreted as a special case of the coordinate descent algorithm (Friedman et al., 2007) by relying on a single iteration to obtain a closed-form solution. Ray and Bhattacharya (2018) and Hauzenberger, Huber, and Onorante (2021) both provide evidence that the coordinate descent algorithm already converges after the first pass-through.

3.4. Sparsifying the covariance matrix

A sparse draw of the covariance matrix can be obtained by relying on methods proposed by Friedman et al. (2007) and Bashir et al. (2019). This involves using the precision rather than the covariance matrix, as the precision defines the conditional independence structure of the variables. It is worth noting that directly postprocessing estimates of the covariance matrix can result in a rather dense precision matrix and, hence, induce spurious contemporaneous relationships. Sparse estimates of the precision matrix $\boldsymbol{\Sigma}_t^{-1}$ are based on the graphical LASSO penalty:

$$\boldsymbol{\Sigma}_t^{-1*} = \min_{\boldsymbol{\Sigma}_t^{-1}} \left(\text{tr}\left(\boldsymbol{\Sigma}_t^{-1}\hat{\boldsymbol{\Sigma}}_t\right) - \log\det\left(\boldsymbol{\Sigma}_t^{-1}\right) + \sum_{i\neq j}\lambda_{ij,t}|\sigma_{ij,t}^{-1}| \right), \tag{9}$$

with tr($\boldsymbol{C}$) and log det($\boldsymbol{C}$) respectively denoting the trace and the log determinant of a square matrix $\boldsymbol{C}$, $\lambda_{ij,t}$ an element-specific LASSO penalty, and $\sigma_{ij,t}^{-1}$ the jth element in the ith row of $\boldsymbol{\Sigma}_t^{-1}$. Similar to Eq. (5) and Eq. (7), the parts $\text{tr}\left(\boldsymbol{\Sigma}_t^{-1}\hat{\boldsymbol{\Sigma}}_t\right) - \log\det\left(\boldsymbol{\Sigma}_t^{-1}\right)$ are measures of fit, while the third term penalizes non-zero elements in the precision matrix. Here, it is worth noting that if $\sigma_{ij,t}^{-1*}$ is set to zero, the ith and jth endogenous variable in the system does not feature a contemporaneous relationship. Thus, postprocessing estimates of the precision matrix capture the notion of obtaining a truly sparse set of relationships between elements in $\Delta\boldsymbol{y}_t$.

Following Bashir et al. (2019), the penalty in Eq. (9) is chosen as $\lambda_{ij,t} = 1/|\sigma_{ij,t}^{-1}|^{0.5}$, which constitutes a semi-automatic procedure to circumvent cross-validation. Here, we refrain from showing the exact form of the soft threshold estimates for each element in $\boldsymbol{\Sigma}_t^{-1*}$ and refer to Friedman et al. (2008) instead, who define a set of soft threshold problems, similar to Eq. (8), to solve for an optimal solution for each element in $\boldsymbol{\Sigma}_t^{-1}$ (and thus $\boldsymbol{\Sigma}_t$). We use the coordinate descent algorithm of Friedman, Hastie, and Tibshirani (2019) and iterate once in line with the SAVS estimator.[11]

Estimating the model proposed in Section 2 yields MCMC draws from the posterior distributions of all relevant parameters. These are subsequently postprocessed. The baseline framework may be applied to any dataset where one suspects the presence of cointegration relationships. In the next section, we discuss further specification details in the context of applying this framework to a study of electricity prices.

4. Modeling European electricity prices

We use our proposed model in two different applications related to European electricity prices. First, we use daily data from several European markets jointly. This serves to illustrate our approach in terms of detecting suitable cointegration relationships and nonlinearities among interconnected energy markets in different European countries. Second, we produce forecasts of hourly electricity prices one day ahead. Here, we perform an extensive OOS forecast exercise, selecting German hourly electricity prices as the market of interest. We then evaluate our approach against a large set of competing models. This exercise serves to demonstrate that flexibly controlling for cointegration patterns in data (in this case, between hours during the day), in the absence of prior knowledge of such relationships, is beneficial for forecast accuracy.

4.1. Dataset

For the first application, we use daily prices (in levels, averaged over the hours of the day) to estimate our model jointly for nine different regional markets: the Baltics (BALT), Denmark (DK), Finland (FI), France (FR), Germany (DE), Italy (IT), Norway (NO), Sweden (SE), and Switzerland (CH); that is, $M = 9$. The data are available for the period from January 1st, 2017 to December 31st, 2019, in EUR per megawatt-hour (MWh).[12] We follow the literature and choose day-ahead prices determined on a specific day for delivery at a certain hour on the following day.

Prices for BALT and the Nordic countries (DK, FI, NO, and SE) are obtained from Nord Pool; the German, Swiss, and French hourly auction prices are from the power spot market of the European Energy Exchange (EEX); for the Italian prices, we use the single national prices (PUN) from the Italian system operator Gestore dei Mercati Energetici (GEM). We preprocess the data for daylight saving time changes to exclude the 25th hour in October and to interpolate the 24th hour in March. As additional exogenous factors, we consider daily prices for coal and fuel and interpolate missing values for weekends and holidays. In particular, we use the closing settlement prices for coal (LMCYSPT) and one-month-forward ICE UK natural gas prices (NATBGAS), due to their importance for the dynamic evolution of electricity prices and potential cointegration relationships (*i.e.*, $q = M + 2 = 11$).

In our second application, the forecast comparison, we choose hourly day-ahead prices (in levels) for Germany as our main country of interest, and focus on daylight hours (see Raviv, Bouwman, & Van Dijk, 2015), which refer to 11 brackets starting with 8:00–9:00 a.m. (in the tables, labeled as 8 a.m.) to the slot from 6:00–7:00 p.m (in the tables, labeled as 6 p.m.), and an average of the night hours. We use a hold-out period of approximately a year and a half ranging from July 3rd, 2018 to December 31st, 2019 (in total, 540 observations). We estimate TVP-VECMs with the individual hours per day being treated as dependent variables such that $M = 12$. Our exercise is

[11] Alternatively, one could also regularize the precision matrix by writing $\boldsymbol{\Sigma}_t^{-1}$ as an M-dimensional set of nodewise regressions by using the triangularization decomposition outlined by Meinshausen and Bühlmann (2006). Friedman et al. (2008) and Banerjee, Ghaoui, and d'Aspremont (2008) show that this approach constitutes a special case of Eq. (9) with M independent LASSO problems.

[12] We note that electricity prices may occasionally turn negative in some countries such as Germany, while in others such as Italy, a zero lower bound is applied by law. This aspect of heterogeneity poses no further issues in our flexible multivariate framework.

based on a pseudo-OOS simulation using a rolling window of $T = 365$ observations at a time. We consider one-step-ahead predictions, which implies that we forecast each individual hour for the following day.[13]

For forecasting electricity prices, apart from seasonal day-of-the-week dummies and daily prices for coal and fuel (henceforth labeled FUEL), we also include renewable energy sources (labeled RES) in the form of forecasted average daily demand, forecasted average daily wind generation, and forecasted average daily photovoltaic solar generation as exogenous factors. These factors have been found to carry substantial predictive power for daily electricity prices, in particular in the short term such as one day ahead (see, *e.g.*, Billé, Gianfreda, Del Grosso, & Ravazzolo, 2023; Ziel & Weron, 2018).

4.2. Nonlinearities and cointegration in European electricity prices

One major feature of the proposed approach is that it remains relatively agnostic to the precise form of the cointegration and parameter space spanned by short-run coefficients and cross-country relationships or heterogeneities. We do not rule out complex patterns a priori, but our flexible modeling approach is also capable of supporting a fairly parsimonious specification when the data suggests so. It is worth noting that a more 'structural' analysis of the estimated cointegration patterns would require additional restrictions to allow for economic interpretations. We leave these aspects for future research.

In this subsection, we examine the sparsification patterns and nonlinearities when daily electricity price data from multiple electricity markets are modeled jointly with a sparse TVP-VECM. We first assess whether these electricity prices are indeed cointegrated and whether these cointegration relationships change over time. We roughly gauge the nonlinear features of the cointegration space by assessing estimates for a time-varying cointegration rank. Second, we assess the sparsification patterns and nonlinearities of static and dynamic interdependencies across the European electricity market. To this end, we examine the sparsified estimates of both the short-run adjustment coefficients (dynamic interdependencies) and the contemporaneous relationships (static interdependencies). Third and finally, we investigate heteroskedastic data features and the role of large variance shocks.

Fig. 1 shows the posterior probability of the rank (PPR) over time based on our MCMC output. The probabilities are indicated in various shades of red. Several findings are worth noting. A rank larger than six is hardly ever supported, and while most posterior mass is concentrated on $r_t = 4$ for all t, we detect subtle differences over time. At the beginning of the sample in 2017, our estimates are more dispersed, with non-negligible probabilities for no cointegration. The precision of our rank estimate increases over time, with a much narrower corridor of probabilities starting around 2018. After a brief period in late 2018 and early 2019 with probabilities shifting towards lower ranks, we find increases in cointegration relationships towards the end of the sample. We conjecture that this pattern of the cointegration rank results from the fact that some of the countries (or regions) experienced a large (idiosyncratic) positive shock in electricity prices in early 2018, while others did not. For example, electricity prices in the Baltics and two Nordic countries (Finland and Sweden) increased from around 30 euros to 90 euros. In all other countries, these jumps are either less pronounced (*e.g.*, Denmark and Norway) or almost nonexistent (*e.g.*, Italy) during this episode.

Next, we turn to sparsified estimates of the autoregressive coefficients and the error covariance matrix in panels (a) and (b) of Fig. 2. Rather than showing the magnitudes of the estimated coefficients, we use this exercise to illustrate the sparsification approach. As noted above, conventional shrinkage approaches push coefficients toward zero, but they are never exactly zero. The sparsification approach, on the other hand, introduces exact zeros in these matrices. We use this fact to compute the posterior inclusion probabilities (PIPs) of all coefficients by calculating the relative share of zeroes over the iterations of the algorithm. In other words, these figures showcase the sparsification approach as a variable selection tool (see Hahn & Carvalho, 2015).

We start with the coefficients linked to $\Delta \boldsymbol{y}_{t-p}$, that is, the dynamic interdependencies of the multivariate system in panel (a). A few findings are worth noting. First, we detect different degrees of sparsity across countries. While the Nordic and Baltic countries look rather similar (with comparatively dense coefficient matrices), this is not the case for Switzerland, Germany, Italy, and France (CH, DE, FR, and IT). For these countries, the model estimates rather sparse coefficient matrices. Second, the variables with the highest PIPs are typically the countries' own lagged series. Particularly, France shows extremely sparse estimates, with non-zero inclusion probabilities only for its own lags. Third, we observe several interesting changes in PIPs over time. This implies changes in the importance of predictors over time, a feature that our model detects in a data-based fashion. Fourth, we observe some noteworthy patterns of dynamic interdependencies, namely between Nordic and Baltic countries on one side, and to a lesser degree between continental European economies.

A similar exercise of PIPs in the context of the sparse covariance matrix is displayed in panel (b). Here, we show the lower triangular part of the contemporaneous relationships over time. As in the case of the regression coefficients, we detect differences in the degree of sparsity over the cross-section and across time. Strong contemporaneous relationships are detected especially between the Nordic countries. The PIPs in this case are often

[13] The main reason for the use of a rolling rather than expanding window is to limit the computational burden. A rolling window implies quicker parameter change over the holdout than an expanding one. However, the daily frequency provides enough observations for making more abrupt parameter changes—possibly captured by the TVPs—and likely also for the rolling window. The presence of TVPs safeguards our framework to some extent from influencing the results by choices about the length of the rolling window, as discussed by Hubicka, Marcjasz, and Weron (2018).

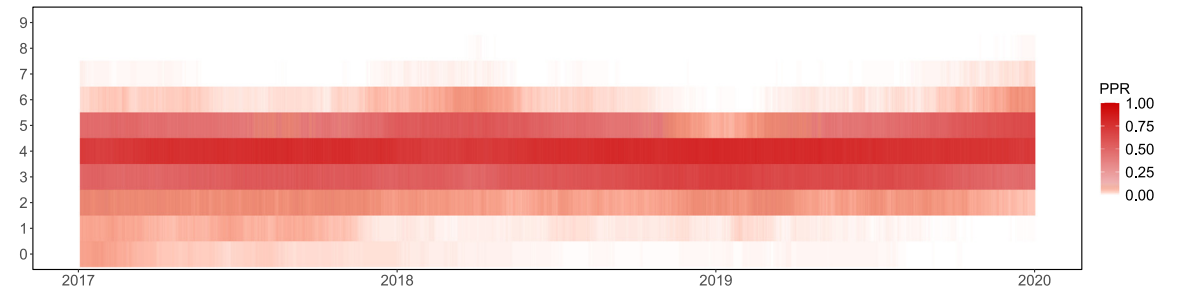

Fig. 1. Posterior probability of the rank (PPR) over time of the most flexible specification.

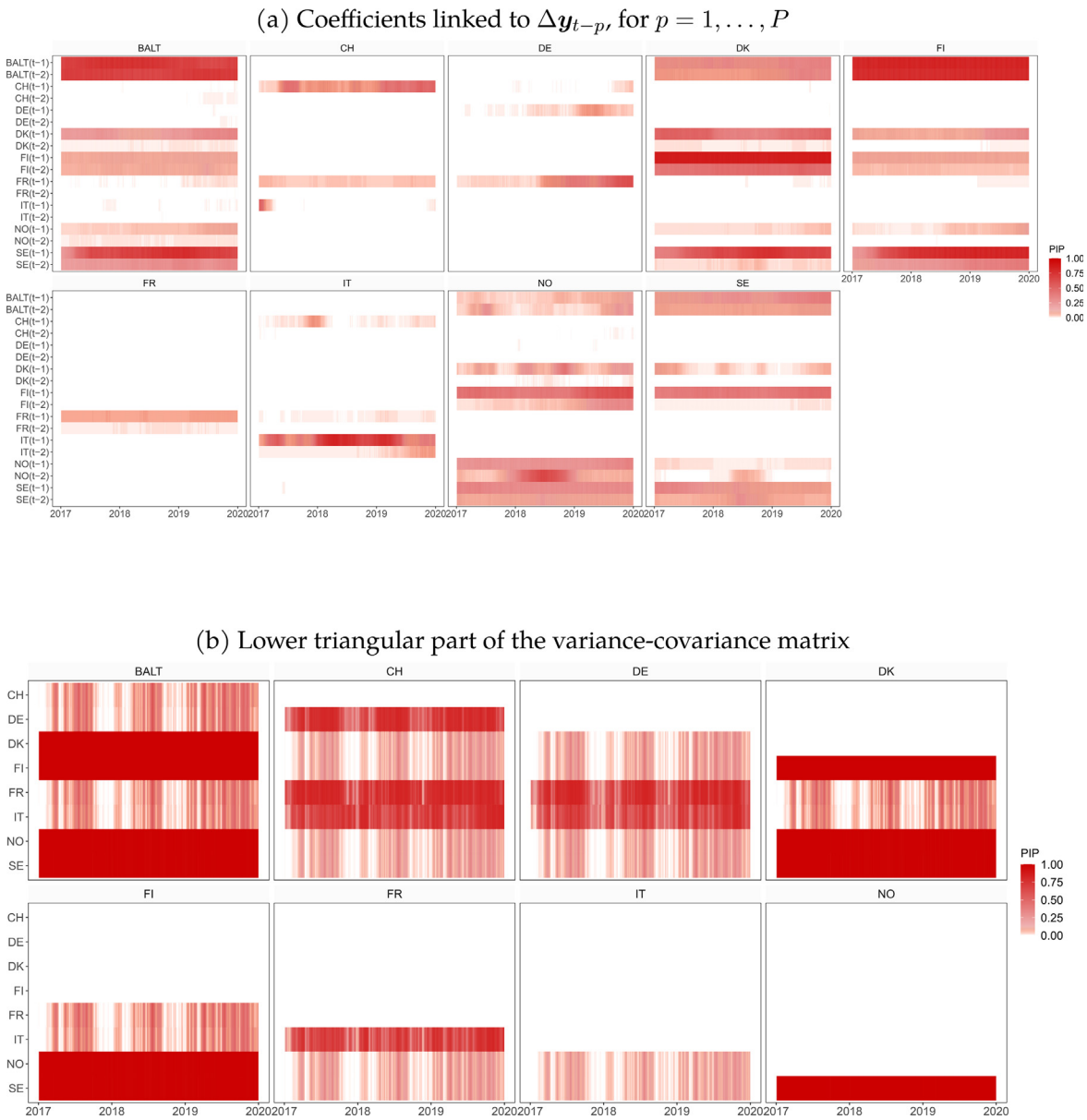

Fig. 2. Posterior inclusion probability (PIP) over time of autoregressive coefficients (panel (a)) and covariance matrices (panel (b)) of the most flexible specification.

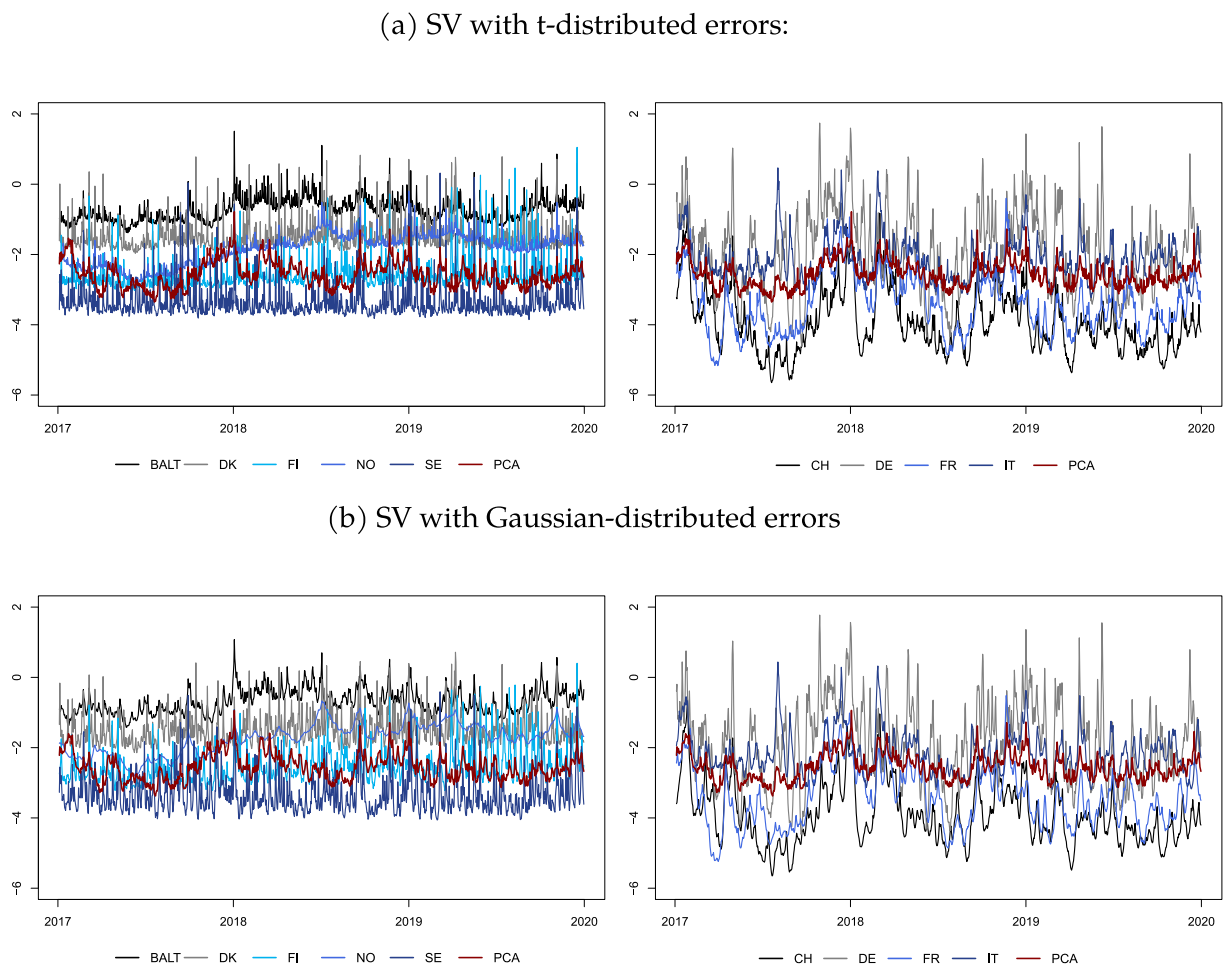

Fig. 3. Posterior median of SV. The red line denotes the first principal component (PCA) of M latent processes (h_{it}).

exactly one, indicating that the respective coefficients feature non-zero draws across all iterations of the sampler. Similarly, albeit with lower inclusion probabilities, we find that covariances appear to be important between continental European electricity prices.

Interestingly, we observe a substantial degree of time variation in the inclusion probabilities. Investigating these patterns in more detail, we find that the overall sparsity of the covariance matrix changes strongly over time, but does so in a specific way. In particular, there are periods where most covariance terms (apart from the always-featured ones across Nordic countries) are sparsified strongly. Examples of such periods are in late 2017 and early 2018, or around the beginning of 2019. Again, it is noteworthy that our model automatically discovers these features.

We complete our discussion of the multi-country application by assessing whether heavy-tailed errors are required to capture energy price fluctuations across the countries. For this purpose, we compare our estimates from our TVP-VECM benchmark model with SV to the same specification with t-distributed errors (see Appendix A). We therefore assess the log-volatilities over time and across countries.

Fig. 3 shows our estimates for t-distributed errors in the top panels (a), while the bottom panels (b) indicate the standard SV specification. A few things are worth noting here. First, while the level of the volatilities varies substantially over the cross-section, the volatilities exhibit a substantial degree of co-movement. Second, even though we detect several differences between heavy-tailed errors and conventional SVs, the first principal component of all volatility processes (marked by the red lines) is almost identical for both error specifications. Third, while t-distributed errors result in numerous high-frequency spikes for Denmark, Finland, and Sweden (*i.e.*, the Nordic countries, which are shown in the left-hand side panels of (a) and (b), except for Norway), Central European countries, shown in the right-hand side panels of (a) and (b), exhibit a higher degree of persistence.

Overall, our results across countries are thus mixed. These findings can be explained by differences in the electricity generation mix of individual countries (see, *e.g.*, Durante, Gianfreda, Ravazzolo, & Rossini, 2022; Ziel,

Steinert, & Husmann, 2015). There is overwhelming empirical evidence for time-varying variances. On the other hand, we find that t-distributed errors are not crucial for Baltic countries, Switzerland, Germany, France, Italy, and Norway, while the data support heavier-than-normal tails for the remaining Nordic countries. Our model detects these features automatically, without the need to take specific a priori decisions about model specification. Next, we discuss whether these features pay off in terms of predictive performance.

4.3. Forecast results

In our forecasting comparison, we zoom in on a single country to conduct a thorough analysis featuring many competing models that have been shown to work well in the previous literature. In particular, we select hourly day-ahead prices for Germany and evaluate point and density forecasts for a set of models that are described in detail below.

4.3.1. Competing models

The competing specifications include a large set of univariate and multivariate models. We provide additional details about key modeling aspects below. At a glance, the major dimensions of differentiation (which we consider in various combinations) are the following:

- **Dynamic process**: Multivariate vector error correction model (VECM), vector autoregression (VAR), and equation-specific univariate autoregression (AR), with all three variants featuring exogenous/deterministic variables (see details below).
- **Conditional mean**: We consider both time-varying parameter (TVP) and time-invariant (TIV), *i.e.*, constant parameter, versions of our model.
- **Conditional variance**: We vary between stochastic volatility (SV) with Gaussian or t-distributed errors, and homoskedastic variants.
- **Postprocessing**: Sparsified posteriors, as described in Section 3, and non-sparsified, exact posteriors from the MCMC algorithm are considered.

Our main interest centers on VECMs that are either sparsified or non-sparsified. For all these VECM specifications, we remain agnostic regarding the cointegration relations and estimate them from the data. Recall that these VECMs feature two lags, *i.e.*, two full days' worth of hourly lags ($P = 2$), and are equipped with a horseshoe (HS, Carvalho et al., 2010) prior. It is also noteworthy that the VECM specifications all include *expert* information, as discussed in Section 4.1. In this sense, our VECM models can be thought of as flexible multivariate extensions of the expert autoregressive distributed lag (ADL or AR-X) models (see, *e.g.*, Billé et al., 2023; Ziel & Weron, 2018). These VECM specifications explicitly account for any cointegration relationships and — in the case of the sparsified versions — consider not only shrinkage but also sparsification to obtain truly sparse solutions.[14]

As natural (non-sparsified) multivariate competitors, we consider VAR-X models in levels. Natural (non-sparsified) univariate competitors are AR-X and AR models in levels. The VAR-X models feature the same set of exogenous covariates as the VECMs (*i.e.*, seasonal day-of-the-week dummies, FUEL, and RES), while for the AR-X model, we differentiate between two variants. One variant includes both FUEL and RES, while a smaller specification includes just RES as exogenous indicators (apart from the dummies). In this sense, all these competitors are among the class of *expert* models. Moreover, to facilitate comparisons across the different model classes, all specifications feature HS shrinkage for regularization, which mitigates the risk of specifics about prior choices affecting our results.

The univariate AR-X competitors can be interpreted as Bayesian implementations of Lasso-Expert-AR (LEAR, Lago, Marcjasz, De Schutter, & Weron, 2021) models, which have been found to be very successful at forecasting daily electricity prices. The VAR-X is the respective multivariate companion. For the univariate AR-X and AR models, we also vary the lag length. We consider two lags ($P = 2$) and seven lags, *i.e.*, seven full days' worth of hourly lags ($P = 7$), while the VAR-X specifications, similar to the VECMs, feature two lags. All the multivariate model variants are much more flexible and are estimated with constant/time-invariant (`TIV`) and time-varying (`TVP`) parameters, while for the univariate benchmarks we consider just constant coefficients. In the following, the AR(2) in levels with constant coefficients serves as an overall benchmark.

Concerning the stochastic disturbances, we account for heteroskedasticity by relying either on a conventional SV specification with Gaussian errors (`SV-n` in the tables) as in Eq. (4), or the extension with t-distributed errors (labeled `SV-t` in the tables). For VECMs, we also estimate additional competing variants with homoskedastic error variances (labeled `homosk.` in the tables). These competitors, alongside all other VECM specifications discussed below, allow us to assess whether changes in conditional variances and/or changes in the conditional means improve forecast accuracy for VECMs (our main model class of interest).

4.3.2. Point and density forecasts

Root mean squared errors (RMSEs) serve to evaluate the point predictions (which are based on the predictive median) across our models. As a density forecast measure, we rely on the continuous ranked probability score (CRPS, Gneiting & Raftery, 2007). It captures not only the first but also higher-order moments of the predictive distribution. The CRPS can be interpreted in the scale of the data and

[14] Another flexible approach worth mentioning, although among a different class of machine learning models, is the neural network architecture proposed by Marcjasz, Narajewski, Weron, and Ziel (2023).

Table 1
Forecast performance for point forecasts (and density forecasts in parentheses) relative to the benchmark.

Specification		One-day ahead												
		Total	8 a.m.	9 a.m.	10 a.m.	11 a.m.	12 noon	1 p.m.	2 p.m.	3 p.m.	4 p.m.	5 p.m.	6 p.m.	Night
Sparsified VECM														
TVP	SV-t	0.965	0.960	0.952	0.962	0.940	0.977	0.962	**0.953**	**0.908**	0.967	1.031	1.035	**0.925**
		(0.902)	(0.948)	(0.943)	(0.880)	(0.869)	(0.859)	(0.873)	(0.865)	(**0.855**)	(0.903)	(0.963)	(0.984)	(0.871)
TVP	SV-n	0.968	0.949	0.944	0.953	0.952	0.979	0.968	0.976	0.934	0.976	1.023	1.040	0.932
		(0.902)	(0.942)	(0.945)	(0.880)	(**0.861**)	(0.857)	(0.884)	(0.861)	(0.869)	(0.904)	(0.964)	(0.984)	(0.864)
TVP	homosk.	0.982	1.032	0.991	0.944	0.951	0.951	0.996	0.966	0.967	0.984	1.019	1.006	0.954
		(0.925)	(1.093)	(1.008)	(0.897)	(0.883)	(0.860)	(0.891)	(0.889)	(0.907)	(0.910)	(0.912)	(0.945)	(0.878)
TIV	SV-t	0.984	0.952	0.959	0.960	0.956	1.007	0.973	0.985	0.936	0.992	1.079	1.064	0.960
		(0.923)	(0.928)	(0.932)	(0.892)	(0.881)	(0.869)	(0.890)	(0.893)	(0.886)	(0.930)	(1.022)	(1.051)	(0.907)
TIV	SV-n	0.974	0.941	0.922	0.962	0.951	0.983	0.967	0.998	0.939	0.985	1.062	1.055	0.945
		(0.921)	(0.925)	(0.927)	(0.893)	(0.867)	(0.860)	(0.888)	(0.890)	(0.892)	(0.934)	(1.022)	(1.051)	(0.907)
TIV	homosk.	0.996	0.961	0.975	0.984	0.990	0.979	0.994	1.026	0.965	0.999	1.056	1.075	0.968
		(0.927)	(0.949)	(0.939)	(0.928)	(0.923)	(0.869)	(0.897)	(0.923)	(0.909)	(0.922)	(0.950)	(1.010)	(0.909)
Non-sparsified VECM														
TVP	SV-t	0.959	0.919	0.946	0.934	0.957	0.961	0.984	0.986	0.937	0.949	1.010	1.009	0.949
		(0.890)	(0.902)	(0.912)	(**0.871**)	(0.890)	(0.845)	(0.864)	(0.875)	(0.887)	(**0.873**)	(0.910)	(0.972)	(0.864)
TVP	SV-n	0.962	0.916	0.953	0.939	0.950	0.971	0.999	0.991	0.950	0.950	0.990	1.019	0.949
		(**0.889**)	(0.901)	(0.910)	(0.874)	(0.886)	(0.840)	(**0.861**)	(0.881)	(0.900)	(0.881)	(**0.904**)	(0.975)	(**0.860**)
TVP	homosk.	0.966	0.961	0.964	0.935	0.938	0.973	0.973	0.991	0.968	0.961	0.987	1.002	0.948
		(0.907)	(0.944)	(0.965)	(0.887)	(0.867)	(0.854)	(0.892)	(0.886)	(0.899)	(0.908)	(0.925)	(0.951)	(0.894)
TIV	SV-t	0.969	0.928	0.946	0.931	0.967	0.972	0.985	1.005	0.948	0.958	1.013	1.051	0.963
		(0.907)	(0.880)	(0.892)	(0.897)	(0.904)	(0.848)	(0.885)	(0.873)	(0.903)	(0.905)	(0.965)	(1.050)	(0.901)
TIV	SV-n	0.968	0.923	0.945	0.941	0.964	0.950	0.988	0.999	0.955	0.948	1.011	1.063	0.960
		(0.904)	(0.880)	(0.895)	(0.898)	(0.901)	(**0.832**)	(0.872)	(**0.857**)	(0.893)	(0.913)	(0.972)	(1.051)	(0.894)
TIV	homosk.	0.994	0.937	0.976	0.963	0.987	1.005	1.032	1.016	0.978	0.985	1.057	1.031	1.000
		(0.931)	(0.894)	(0.906)	(0.909)	(0.919)	(0.906)	(0.939)	(0.920)	(0.949)	(0.943)	(1.009)	(0.968)	(0.918)
Non-sparsified VAR-X, estimated in levels														
TVP	SV-n	0.967	0.943	0.971	0.940	0.937	0.973	0.987	0.987	0.960	0.986	0.997	0.985	0.949
		(0.913)	(0.909)	(0.906)	(0.925)	(0.874)	(0.865)	(0.912)	(0.906)	(0.902)	(0.952)	(0.966)	(0.949)	(0.888)
TIV	SV-n	0.968	0.955	0.966	0.955	**0.936**	0.956	0.984	0.993	0.947	0.995	0.995	0.994	0.949
		(0.910)	(0.924)	(0.915)	(0.926)	(0.877)	(0.858)	(0.896)	(0.883)	(0.910)	(0.945)	(0.952)	(0.953)	(0.884)
Exo. fac.	**No. of lags**	**AR-X with TIV and SV-n, estimated in levels**												
RES & FUEL	7 lags	0.943	0.845	0.877	0.916	0.950	**0.948**	0.999	1.012	0.968	0.991	0.999	0.932	0.943
		(0.919)	(0.820)	(0.840)	(0.916)	(0.936)	(0.909)	(0.980)	(0.954)	(0.959)	(0.948)	(0.947)	(0.920)	(0.924)
RES & FUEL	2 lags	1.016	1.017	1.006	1.001	1.001	1.003	1.044	1.056	1.009	1.036	1.036	0.972	1.017
		(0.991)	(0.993)	(0.966)	(0.979)	(0.978)	(0.960)	(1.039)	(1.010)	(0.998)	(1.011)	(1.014)	(0.966)	(0.988)
RES	7 lags	0.934	0.843	0.882	0.960	0.966	0.973	0.973	0.979	0.956	0.946	**0.945**	**0.895**	0.930
		(0.917)	(0.835)	(0.837)	(0.923)	(0.945)	(0.939)	(0.990)	(0.958)	(0.932)	(0.911)	(0.913)	(**0.902**)	(0.945)
RES	2 lags	1.009	1.009	1.047	1.039	1.028	1.027	0.986	1.007	0.993	0.982	1.006	0.971	0.973
		(1.007)	(1.032)	(1.014)	(1.042)	(1.040)	(0.999)	(1.049)	(1.009)	(0.996)	(0.981)	(0.975)	(0.956)	(0.981)
AR with TIV and SV-n, estimated in levels														
	7 lags	**0.916**	**0.793**	**0.845**	**0.901**	0.948	0.949	**0.955**	0.961	0.949	**0.936**	0.960	0.942	0.930
		(0.898)	(**0.817**)	(**0.816**)	(0.893)	(0.921)	(0.916)	(0.948)	(0.919)	(0.918)	(0.900)	(0.905)	(0.922)	(0.933)
	2 lags	8.55	10.32	9.78	8.77	8.56	8.40	7.96	8.84	9.86	9.32	8.44	7.62	4.58
		(5.67)	(6.40)	(6.27)	(5.76)	(5.67)	(5.78)	(5.43)	(6.05)	(6.53)	(6.20)	(5.66)	(4.94)	(3.02)

Notes: Point forecasts are evaluated using root mean squared errors (RMSEs). Density forecasts are continuous ranked probability scores (CRPSs). Values in the first row per model indicate RMSEs. Those in parentheses in the second row are CRPSs. The red shaded rows denote the benchmark and display actual values for RMSEs and CRPSs. All other models are shown as ratios to the benchmark. Relative numbers below one indicate superior forecast performance, with the best performing specification in bold. For the VECM and VAR-X models, we always include seasonal day-of-the-week dummies, daily prices for coal and fuel (FUEL), and forecasts of average daily renewable energy sources (RES) as exogenous factors. For the AR-X model, we differentiate between two variants. One variant includes seasonal dummies, FUEL, and RES, while a smaller specification just includes RES and seasonal dummies as exogenous indicators. Moreover, for the univariate benchmarks, we consider a lag length of two and seven (along with seasonal dummies).

is defined such that lower values are better:

$$\text{CRPS}_{h,t}(y_{h,t+1}) = \int_{-\infty}^{\infty} \left(F(z) - \mathbb{I}\{y_{h,t+1} \leq z\}\right)^2 dz = \\ = E_f|\hat{y}_{h,t+1} - y_{h,t+1}| - 0.5E_f|\hat{y}_{h,t+1} - \hat{y}'_{h,t+1}|,$$

for hour h. $F(\cdot)$ denotes the cumulative distribution function associated with the predictive density f, $\mathbb{I}\{y_{h,t+1} \leq z\}$ denotes an indicator function taking the value 1 if $y_{h,t+1} \leq z$ and 0 otherwise, and $\hat{y}_{h,t+1}$ and $\hat{y}'_{h,t+1}$ are independent draws from the posterior predictive density.

Table 1 displays RMSEs and CRPSs across all model types (rows) and over the respective hours of the day (columns). Forecasts are produced hourly for a full day ahead (8 a.m. until 6 p.m., and aggregate 'Night' hours). Besides individual hours, we produce a summary figure for overall forecast performance across a full day (labeled 'Total'). Model abbreviations and variants of features such as TVPs or the error variances specification are indicated in the previous subsection. Values in the first row per model indicate RMSEs. Those in parentheses in the second row are CRPSs. They are benchmarked relative (as ratios) to the AR(P) model in differences with constant parameters and a standard SV specification (red shaded row, indicating raw values of RMSEs and CRPSs). In both cases,

relative numbers below one indicate superior forecast performance, with the best-performing specification in bold.

The upper panel contains results for our main model variant (*i.e.*, a VECM equipped with shrinkage priors) with both sparsified and non-sparsified estimates. The lower panel shows several non-sparsified benchmarks. It is worth mentioning that many of these benchmark specifications are nested in our proposed model. For example, in the case where our approach selects the cointegration matrix to be of full rank, we obtain a VAR in levels. Such specifications in essence differ in terms of the implied shrinkage prior of the reduced-form coefficients (see, for instance, Eisenstat et al., 2016; Giannone et al., 2019; Villani, 2009).

Our results indicate the lack of a single best-performing specification. However, some intriguing patterns still emerge across the model variants. We find that predictive accuracy is rather heterogeneous with respect to hours of the day, and that some differences occur depending on whether the focus is on point or density forecasts. Table 1 reports that univariate expert models combined with shrinkage, particularly those with higher lag-orders, perform very well and virtually always improve upon the naive AR(2) benchmark, as expected. Specifically, the relative performance of these models is strongest in the early hours of the day, which is the main reason for their overall strong performance on average.

Versions of the VECM perform particularly well in the afternoon and evening, where they are the best-performing specification in several cases, reducing losses in density forecasts by more than 10% relative to the benchmark in some instances. Focusing on the column `Total`, which summarizes overall forecast performance across all hours over the day ahead, indicates that the best-performing model in terms of density forecast accuracy is a non-sparsified TVP-VECM with conventional SV. Hence, flexible expert VECMs with TVPs and SV produce competitive forecasts and have the potential to improve density forecast accuracy upon strong benchmarks.

Turning to predictive accuracy premia that arise from explicitly modeling cointegration relationships, we find that our proposed VECMs typically have a slight edge over the VAR specifications. This points towards the usefulness of our framework to forecast electricity prices. The same is true when comparing sparsified to non-sparsified specifications. The sparsified and non-sparsified predictive metrics are often close, but there is some evidence that sparsification on average offers bigger gains than potential losses. Improvements are usually more substantial for density forecasts as evaluated with CRPSs, when compared to RMSE metrics for point forecasts. Moreover, for our VECM specifications, when examining in greater detail whether allowing for TVPs and/or time-varying variances (heteroskedasticity through SV) improves predictive accuracy, we generally find that introducing flexibility along both dimensions yields predictive gains (although these gains are sometimes small). Specifically, around noon and in the afternoon hours, we find systematic gains in predictive accuracy for specifications that feature both TVPs and SV. We do not find systematic improvements when contrasting Gaussian errors with t-distributed ones, consistent with our findings for Germany from the cross-country application. However, occasionally, during the afternoon hours, we even find support for t-distributed errors, suggesting that during these times of the day, it is important to explicitly model potential outliers. Conversely, the worst VECM specification (on average) is a model that neither allows for TVPs nor features heteroskedasticity (*i.e.*, the homoskedastic VECM-TIV specification).

Next, we gauge the statistical significance of our results. Our set of competitors is large, and conventional approaches—such as the common (Diebold & Mariano, 1995) test—are typically pairwise tests that require choosing a distinct benchmark to formulate a hypothesis. While this allows us to check whether any specific model is statistically more accurate than the benchmark, it is not possible to comment on two non-benchmark specifications in terms of their relative significance. Consequently, we employ the model confidence set (MCS) procedure of Hansen, Lunde, and Nason (2011), which has been designed specifically for such large-scale comparisons.

The MCS procedure yields a set of specifications, a collection of models, that contains the best one with a given level of confidence. In essence, it allows us to obtain a set of models that are statistically indistinguishable from one another, and thus is informative both in terms of which models consistently perform well and which ones are strictly dominated by others. It is worth mentioning that, as in any testing procedure, several assumptions (that may or may not be fulfilled in finite samples) underpin the MCS (see Diebold, 2015, for a discussion of related issues). This is why we employ the MCS procedure only as a rough gauge of statistical significance in a model-selection context and stress that these results must be interpreted with some caution. In the following, we opt for an MCS setting that uses the T_R test statistic and a pre-defined level of 90% confidence for our loss of choice, which is the CRPS. Hansen et al. (2011) derive another test statistic, $T_{\max}$, that can also be used for this procedure.[15] In Appendix B, we show additional results for MCS, where we vary not only the pre-defined confidence level (75% versus 90%) but also the employed test statistic (T_R versus $T_{\max}$). Overall, we find that the obtained rankings are robust with respect to these different settings.

Given the informativeness of the data, this procedure may either select a single best-performing specification or a ranking of several comparable models on our chosen level of 10% significance (see also Hansen et al., 2011). The results of this MCS exercise are shown in Table 2 and empty cells indicate that the corresponding model was eliminated. We again stress that models included in the MCS are statistically indistinguishable from one another. This also translates to the model ranks derived from the two common test statistics in the context of the procedure, which must be interpreted as likely imperfectly measured and approximate indicators of relative performance.

[15] For details about MCS and how these two key test statistics are derived, refer to Hansen et al. (2011).

Table 2
Model confidence set (MCS) for density forecasts.

Specification		One-day ahead												
		Total	8 a.m.	9 a.m.	10 a.m.	11 a.m.	12 noon	1 p.m.	2 p.m.	3 p.m.	4 p.m.	5 p.m.	6 p.m.	Night
Sparsified VECM														
TVP	SV-t	4			**3**	9	9	**3**	**3**	**1**	4	9	5	7
TVP	SV-n	5			**2**	**1**	7	10	**2**	6	7	10	7	**2**
TVP	homosk.	10			7	7	5	6	6	13	9	**3**	**3**	**3**
TIV	SV-t				16	13		12			13			13
TIV	SV-n				14	8	8	11	14		14			12
TIV	homosk.	11			11		10	8	12	11	8	7		9
Non-sparsified VECM														
TVP	SV-t	**2**			4	6	11	**2**	10	**2**	**2**	4	11	5
TVP	SV-n	**1**			**1**	5	4	**1**	9	4	**1**	**2**	12	**1**
TVP	homosk.	8			5	**2**	**3**	7	5	8	5	6	9	8
TIV	SV-t	14	4	4	12	12	12	13	8	5	12	14		14
TIV	SV-n	13	5	5	17	11	**1**	5	**1**	**3**	16			10
TIV	homosk.	12			15	10	15	15	13	10	10	13	6	11
Non-sparsified VAR-X, estimated in levels														
TVP	SV-n	7			9	**3**	6	9	7	7	15	11	10	6
TIV	SV-n	6			10	4	**2**	4	4	9	11	8	8	4
Exo. fac.	**No. of lags**	**AR-X with TIV and SV-n, estimated in levels**												
RES & FUEL	7 lags	9	**2**	**2**	8	14	14				17	12	**2**	
RES & FUEL	2 lags													
RES	7 lags	15	**3**	**3**	13	15				14	6	5	**1**	
RES	2 lags												13	
AR with TIV and SV-n, estimated in levels														
	7 lags	**3**	**1**	**1**	6	16	13	14	11	12	**3**	**1**	4	
	2 lags													

Notes: Results for the model confidence set (MCS) procedure of Hansen et al. (2011) at a 10% significance level using the T_R test statistic. The loss function is specified in terms of continuous ranked probability scores (CRPSs) as a density forecast measure. Empty cells indicate that the model is **not** part of the MCS. The top three ranked models are marked in bold. For the VECM and VAR-X models, we always include seasonal day-of-the-week dummies, daily prices for coal and fuel (FUEL), and forecasts of average daily renewable energy sources (RES) as exogenous factors. For the AR-X model, we differentiate between two variants. One variant includes seasonal dummies, FUEL, and RES, while a smaller specification just includes RES and seasonal dummies as exogenous indicators. Moreover, for the univariate benchmarks, we consider a lag length of two and seven (along with seasonal dummies).

We find that our proposed models perform consistently well overall. The TVP-VECMs are always included in the MCS (*i.e.*, among the statistically significant superior model set), and a single TVP-VECM variant appears among the top three ranked models for density forecasts in all hours (except for 8 a.m. and 9 a.m.). In terms of the univariate models, it is noteworthy that the two-lag specifications are consistently eliminated by MCS. However, the seven-lag specifications perform quite well, as already suggested by the overall forecast performance in Table 1. Overall, Table 2 indicates that simple univariate benchmarks (in the form of AR variants) dominate before noon, while in the afternoon and evening, multivariate expert models dominated—specifically those that explicitly model cointegration relationships as well as shrink and sparsify.

It is also worth reiterating that VECMs appear to yield superior forecasts to VARs. However, compared to our previous discussion of CRPSs, we observe some differences. The MCS procedure yields a different performance ranking in terms of significance when compared to evaluating solely CRPSs and producing a ranking in absolute terms. We conjecture that this is due to several periods in our holdout sample that affect the end-of-sample metrics shown in parentheses in Table 1, whereas the MCS procedure is more robust to such idiosyncrasies. In particular, for the sparsified TVP-VECMs, we observe high ranks particularly during the afternoon in terms of density forecasts.

5. Concluding remarks

In this paper, we proposed a TVP-VECM equipped with shrinkage priors and heteroskedastic errors. We then discussed methods for inducing sparsity. This framework is capable of introducing exact zeroes in cointegration relationships, autoregressive coefficients, and the covariance matrix. The main idea is to start with a suitably flexible and sophisticated specification and to impose data-driven sparsity on the parameter space to obtain the simplest adequate nested version. Moreover, our procedure yields estimates for a time-varying cointegration rank, without the need for introducing prior information on the cointegration relationships.

We estimated our model using daily and hourly day-ahead prices for different European electricity markets. In the empirical section, we illustrated some features of our approach in-sample using a multicountry dataset and conducted an extensive OOS forecast comparison for Germany. Regarding the in-sample analysis, we detected several interesting time-varying patterns of sparsity in the autoregressive coefficients and the covariance matrix. Our proposal detects such features of the data automatically. In addition, we found that our approach is competitive when forecasting hourly one-day-ahead electricity prices

compared to a large set of univariate and multivariate benchmarks.

Data and code availability

The replication package is available at: github.com/nhauzenb/hpr-ijof-tvpvecm.

CRediT authorship contribution statement

Niko Hauzenberger: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Michael Pfarrhofer:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Luca Rossini:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2024.09.001.

References

Banerjee, O., Ghaoui, L. E., & d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *Journal of Machine Learning Research*, *9*(Mar), 485–516.

Bashir, A., Carvalho, C. M., Hahn, P. R., & Jones, M. B. (2019). Post-processing posteriors over precision matrices to produce sparse graph estimates. *Bayesian Analysis*, *14*(4), 1075–1090.

Bello, A., & Reneses, J. (2013). Electricity price forecasting in the Spanish market using cointegration techniques. In *33rd annual international symposium on forecasting (ISF 2013) forecasting with big data* (pp. 23–26).

Billé, A. G., Gianfreda, A., Del Grosso, F., & Ravazzolo, F. (2023). Forecasting electricity prices with expert, linear, and nonlinear models. *International Journal of Forecasting*, *39*(2), 570–586.

Bosco, B., Parisio, L., Pelagatti, M., & Baldi, F. (2010). Long-run relations in European electricity prices. *Journal of Applied Econometrics*, *25*, 805–832.

Bunea, F., She, Y., & Wegkamp, M. H. (2012). Joint variable and rank selection for parsimonious estimation of high-dimensional matrices. *The Annals of Statistics*, *40*(5), 2359–2388.

Cadonna, A., Frühwirth-Schnatter, S., & Knaus, P. (2020). Triple the gamma – A unifying shrinkage prior for variance and variable selection in sparse state space and TVP models. *Econometrics*, *8*(2), 20.

Carriero, A., Chan, J., Clark, T. E., & Marcellino, M. (2022). Corrigendum to 'large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors' [J. Econometrics 212 (1)(2019) 137–154]. *Journal of Econometrics*, *227*(2), 506–512.

Carter, C., & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, *81*(3), 541–553.

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, *97*(2), 465–480.

Chakraborty, A., Bhattacharya, A., & Mallick, B. K. (2020). Bayesian sparse multiple regression for simultaneous rank reduction and variable selection. *Biometrika*, *107*(1), 205–221.

Chan, J. C. (2022). Asymmetric conjugate priors for large Bayesian VARs. *Quant. Econ.*, *13*(3), 1145–1169.

Chan, J. C., & Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, *1*(1–2), 101–120.

Chan, J., Koop, G., Poirier, D. J., & Tobias, J. L. (2020). *Bayesian econometric methods*: *vol. 2*, Cambridge University Press.

Chua, C. L., & Tsiaplias, S. (2018). A Bayesian approach to modeling time-varying cointegration and cointegrating rank. *Journal of Business & Economic Statistics*, *36*(2), 267–277.

de Marcos, R. A., Reneses, J., & Bello, A. (2016). Long-term spanish electricity market price forecasting with cointegration and VEC models. In *2016 international conference on probabilistic methods applied to power systems* (pp. 1–7).

De Vany, A., & Walls, W. (1999). Cointegration analysis of spot electricity prices: Insights on transmission efficiency in the western US. *Energy Economics*, *21*(5), 435–448.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. *Journal of Business & Economic Statistics*, *33*(1), 1.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

Durante, F., Gianfreda, A., Ravazzolo, F., & Rossini, L. (2022). A multivariate dependence analysis for electricity prices, demand and renewable energy sources. *Information Sciences*, *590*, 74–89.

Eisenstat, E., Chan, J. C., & Strachan, R. W. (2016). Stochastic model specification search for time-varying parameter VARs. *Econometric Reviews*, *35*(8–10), 1638–1665.

Feldkircher, M., Huber, F., Koop, G., & Pfarrhofer, M. (2022). Approximate Bayesian inference and forecasting in huge-dimensional multicountry VARs. *International Economic Review*, *63*(4), 1625–1658.

Friedman, J., Hastie, T., Höfling, H., Tibshirani, R., et al. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics*, *1*(2), 302–332.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441.

Friedman, J., Hastie, T., & Tibshirani, R. (2019). glasso: Graphical lasso: Estimation of Gaussian graphical models. R package version 1.11.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, *15*(2), 183–202.

Frühwirth-Schnatter, S., & Wagner, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, *154*(1), 85–100.

Geweke, J. (1996). Bayesian reduced rank regression in econometrics. *Journal of Econometrics*, *75*(1), 121–146.

Gianfreda, A., Parisio, L., & Pelagatti, M. (2019). The RES-induced switching effect across fossil fuels: An analysis of day-ahead and balancing prices. *The Energy Journal*, *40*, 1–22.

Gianfreda, A., Ravazzolo, F., & Rossini, L. (2023). Large time-varying volatility models for electricity prices. *Oxford Bulletin of Economics and Statistics*, *85*, 545–573.

Giannone, D., Lenza, M., & Primiceri, G. E. (2019). Priors for the long run. *Journal of the American Statistical Association*, *114*(526), 565–580.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.

Hahn, P. R., & Carvalho, C. M. (2015). Decoupling shrinkage and selection in Bayesian linear models: A posterior summary perspective. *Journal of the American Statistical Association*, *110*(509), 435–448.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Hauzenberger, N. (2021). Flexible mixture priors for large time-varying parameter models. *Econometrics and Statistics*, *20*, 87–108.

Hauzenberger, N., Huber, F., & Koop, G. (2024). Dynamic shrinkage priors for large time-varying parameter regressions using scalable Markov chain Monte Carlo methods. *Studies in Nonlinear Dynamics & Econometrics*, *28*(2), 201–225.

Hauzenberger, N., Huber, F., Koop, G., & Onorante, L. (2022). Fast and flexible Bayesian inference in time-varying parameter regression models. *Journal of Business & Economic Statistics*, *40*(4), 1904–1918.
Hauzenberger, N., Huber, F., & Onorante, L. (2021). Combining shrinkage and sparsity in conjugate vector autoregressive models. *Journal of Applied Econometrics*, *36*(3), 304–327.
Hauzenberger, N., Huber, F., Pfarrhofer, M., & Zörner, T. O. (2021). Stochastic model specification in Markov switching vector error correction models. *Studies in Nonlinear Dynamics & Econometrics*, *25*(2), Article 20180069.
Houllier, M. A., & Menezes, L. M. De (2012). A fractional cointegration analysis of European electricity spot prices. In *9th international conference on the European energy market*.
Huber, F., Koop, G., & Onorante, L. (2021). Inducing sparsity and shrinkage in time-varying parameter models. *Journal of Business & Economic Statistics*, *39*(3), 669–683.
Huber, F., Koop, G., & Pfarrhofer, M. (2020). Bayesian inference in high-dimensional time-varying parameter models using integrated rotated Gaussian approximations. arXiv:2002.10274.
Huber, F., & Zörner, T. O. (2019). Threshold cointegration in international exchange rates: A Bayesian approach. *International Journal of Forecasting*, *35*(2), 458–473.
Hubicka, K., Marcjasz, G., & Weron, R. (2018). A note on averaging day-ahead electricity price forecasts across calibration windows. *IEEE Transactions on Sustainable Energy*, *10*(1), 321–323.
Jochmann, M., & Koop, G. (2015). Regime-switching cointegration. *Studies in Nonlinear Dynamics & Econometrics*, *19*(1), 35–48.
Jochmann, M., Koop, G., Leon-Gonzalez, R., & Strachan, R. W. (2013). Stochastic search variable selection in vector error correction models with an application to a model of the UK macroeconomy. *Journal of Applied Econometrics*, *28*(1), 62–81.
Kastner, G. (2016). Dealing with stochastic volatility in time series using the R package stochvol. *Journal of Statistical Software*, *69*(5) 1–30.
Kastner, G., & Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, *76*, 408–423.
Kleibergen, F., & Van Dijk, H. K. (1994). On the shape of the likelihood/posterior in cointegration models. *Econometric Theory*, *10*(3–4), 514–551.
Kleibergen, F., & Van Dijk, H. K. (1998). Bayesian simultaneous equations analysis using reduced rank structures. *Econometric Theory*, *14*(6), 701–743.
Koop, G., León-González, R., & Strachan, R. W. (2009). Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econometric Reviews*, *29*(2), 224–242.
Koop, G., Leon-Gonzalez, R., & Strachan, R. W. (2011). Bayesian inference in a time varying cointegration model. *Journal of Econometrics*, *165*(2), 210–220.
Lago, J., Marcjasz, G., De Schutter, B., & Weron, R. (2021). Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, *293*, Article 116983.
Liu, J. S., & Wu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association*, *94*(448), 1264–1274.
Makalic, E., & Schmidt, D. F. (2015). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, *23*(1), 179–182.
Marcjasz, G., Narajewski, M., Weron, R., & Ziel, F. (2023). Distributional neural networks for electricity price forecasting. *Energy Economics*, *125*, Article 106843.
Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics*, *34*(3), 1436–1462.
Paap, R., & Van Dijk, H. K. (2003). Bayes estimates of Markov trends in possibly cointegrated series: An application to US consumption and income. *Journal of Business & Economic Statistics*, *21*(4), 547–563.
Primiceri, G. (2005). *Time varying structural autoregressions and monetary policy*: *vol. 72*, (3), (pp. 821–852). Oxford University Press.
Prüser, J. (2023). Data-based priors for vector error correction models. *International Journal of Forecasting*, *39*(1), 209–227.
Puelz, D., Hahn, P. R., & Carvalho, C. M. (2020). Portfolio selection for individual passive investing. *Applied Stochastic Models in Business and Industry*, *36*(1), 124–142.
Puelz, D., Hahn, P. R., Carvalho, C. M., et al. (2017). Variable selection in seemingly unrelated regressions with random predictors. *Bayesian Analysis*, *12*(4), 969–989.
Raviv, E., Bouwman, K. E., & Van Dijk, D. (2015). Forecasting day-ahead electricity prices: Utilizing hourly prices. *Energy Economics*, *50*, 227–239.
Ray, P., & Bhattacharya, A. (2018). Signal adaptive variable selector for the horseshoe prior. arXiv:1810.09004.
Strachan, R. W. (2003). Valid Bayesian estimation of the cointegrating error correction model. *Journal of Business & Economic Statistics*, *21*(1), 185–195.
Strachan, R. W., & Inder, B. (2004). Bayesian analysis of the error correction model. *Journal of Econometrics*, *123*(2), 307–325.
Villani, M. (2001). Bayesian prediction with cointegrated vector autoregressions. *International Journal of Forecasting*, *17*(4), 585–605.
Villani, M. (2006). Bayesian point estimation of the cointegration space. *Journal of Econometrics*, *134*(2), 645–664.
Villani, M. (2009). Steady-state priors for vector autoregressions. *Journal of Applied Econometrics*, *24*(4), 630–650.
Wang, H., & Leng, C. (2008). A note on adaptive group Lasso. *Computational Statistics & Data Analysis*, *52*(12), 5277–5286.
Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, *30*(4), 1030–1081.
Woody, S., Carvalho, C. M., & Murray, J. S. (2021). Model interpretation through lower-dimensional posterior summarization. *Journal of Computational and Graphical Statistics*, *30*(1), 144–161.
Yang, Y., & Bauwens, L. (2018). State-space models on the Stiefel manifold with a new approach to nonlinear filtering. *Econometrics*, *6*(4), 48.
Yousuf, K., & Ng, S. (2021). Boosting high dimensional predictive regressions with time varying parameters. *Journal of Econometrics*, *224*(1), 60–87.
Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *68*(1), 49–67.
Ziel, F., Steinert, R., & Husmann, S. (2015). Efficient modeling and forecasting of electricity spot prices. *Energy Economics*, *47*, 98–111.
Ziel, F., & Weron, R. (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics*, *70*, 396–420.
Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429.