Cluster-based short-term wind power forecasting

Wolfgang Karl Härdle[*]
Brenda López Cabrera[*]
Awdesch Melzer[°]

[*]Ladislaus von Bortkiewicz Chair of Statistics
[°]Institute of Finance
Humboldt-Universität zu Berlin

Influence on pricing of VWAP

Assume simple relationship between electricity prices and renewable power production

$$\Delta_p = \alpha + \beta \Delta_r + \varepsilon_i$$

Δ_p : Difference in VWAP (spot) and day ahead price
Δ_r : Difference of actual production and forecast in renewables

- The smaller Δ_r, the smaller the price gap between spot and day ahead markets
- The better the renewable forecasts, the smaller Δ_r

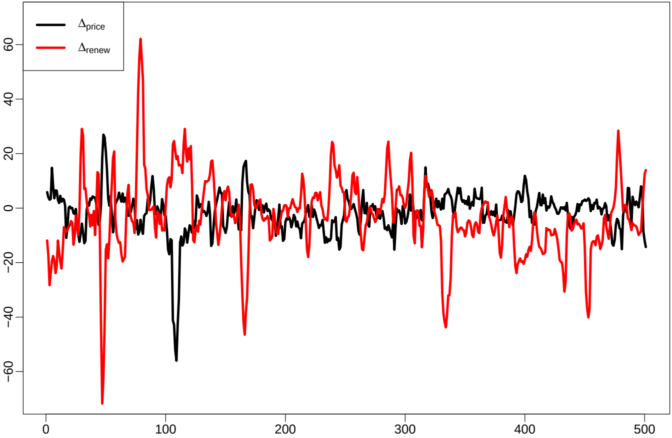


Figure 1: Differences in prices on spot and day ahead market, differences in forecast and production of renewables in Germany

Wind power as function of time

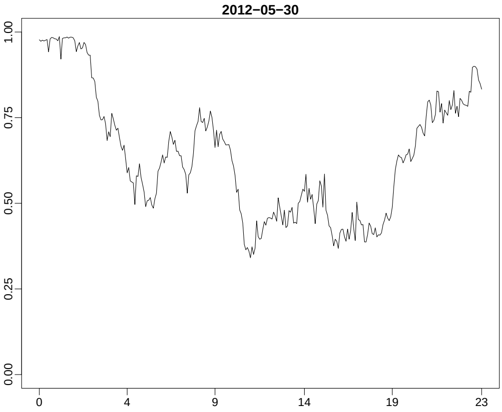


Figure 2: Wind power (WP) generation at HALLWF2.

Research questions

- Can a functional approach increase accuracy of day ahead WP forecasts?
- Does frequency matter?:
  - weather data $\begin{cases}\text{daily average}\\ \text{5 min average}\end{cases}$
- Is there a flexible method wrt underlying data?
  - Turbine-level data vs. TSO-level aggregates
- Why do tail events matter?
  - Data-driven probabilistic forecasts

WP modelling in recent literature

Dowell & Pinsen (2016)

- sparse VAR(p)
  - clusters of correlated turbines
  - probabilistic forecasts (assume logit-normal cdf)

Cavalcante et al. (2017)

- LASSO VAR
  - variable selection to reduce coefficient space
  - point forecast

Browell & Gilbert (2017)

- Cluster-based regime-switching AR (winner WPFC 2017)
  - identify wind regimes by clustering of wind conditions
  - point forecast

Factor model on short term variation

$$\mathbf{U}_t = \Lambda_t + \mathbf{Y}_t$$

$\mathbf{U}_t$: wind power in time t
Λ_t : seasonality component
$\mathbf{Y}_t$: short term variation

Decompose short term variation $\mathbf{Y}_t$ into

- time-invariant common structure ▸ common structure
- time-variant individual components

Employ time-variety of factor loadings in

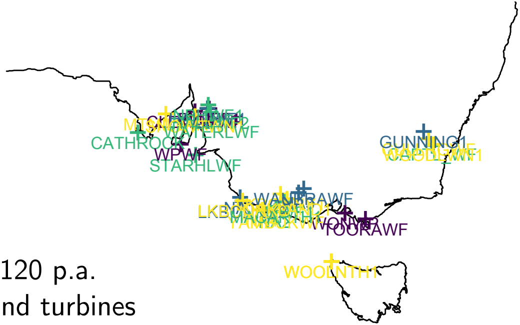
- Time series framework
- Forecasting

López Cabrera & Schulz (2016), Chao et al. (2015), Härdle et al. (2016)

Agenda

1. Motivation ✓
2. Data
3. FASTEC methodology
4. FASTEC-VAR(p) model
5. Forecast evaluation
6. Outlook

AEMO: turbine level wind power data



WP feed-in data:

$\Delta = 5$ min

$N = 288 \cdot 365 = 105120$ p.a.

$P = 22$ wind farms and turbines

U_t on domain of $[0, 1]$

Daily wind speed data at 8 locations

In-sample: 2012-01-02 - 2012-12-31

Out-of-sample: 2013-01-01 - 2013-12-31

DK data

Data-transformation and seasonality

- Transform data from $[0,1]$ domain to $\mathbb{R}$ ▸ Transformation
- Deseasonalize with periodic B-splines: ▸ Seasonality ▸ B-splines

$$\mathbf{U} = \Lambda + \mathbf{Y}$$

$$\Lambda = \Lambda_d \circ \Lambda_a$$

Short-term stochastic component Y

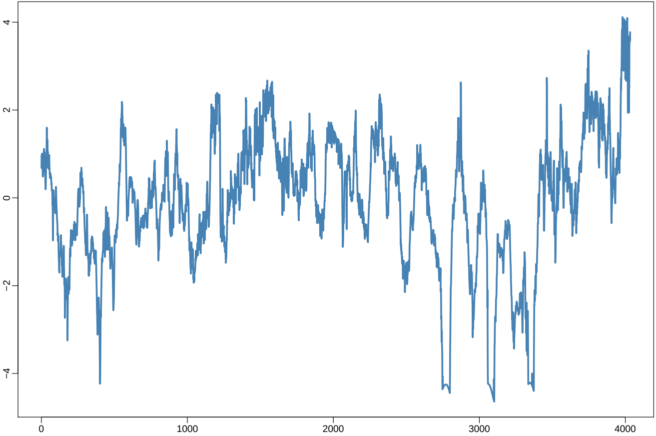


Figure 3: TS 14 days residuals of HALLWF2: June 2012

Factor model at expectile-level

- Apply asymmetric weights to data ▸ Expectiles
- Find factors and loadings by iterative SVD and penalised regression: FASTEC methodology ▸ FASTEC ▸ common structure
- Use loadings for VARX(p) model

FASTEC-VARX model

$$\Psi_t(\tau) = \sum_{k=1}^{P} \Theta_k \Psi_{t-k}(\tau) + \Xi \mathbf{X}_t + \eta_t,$$

$\Psi_t(\tau)$: vector of loadings at τ-level, $\Psi_{(365 \times r)}$
Θ_k: matrix of VAR coefficients
Ξ: matrix of exogenous coefficients
η_t: white noise error term
$\tau = \{1\%, \dots, 50\%, \dots, 99\%\}$

López Cabrera & Schulz (2016)

Markov-switching FASTEC-VARX model

Introduce Markov-switching between states $\mathbf{S}_t$:
high and low production states (strong and weak wind)

$$\Psi_t = \sum_{k=1}^{P} \Theta_k \mathbf{S}_t \Psi_{t-k} + \Xi \mathbf{S}_t \mathbf{X}_t + \Sigma \mathbf{S}_t \eta_t,$$

with a is the transitionprobability matrix $\mathbf{P}$

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{m1} \\ p_{12} & p_{22} & \cdots & p_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1m} & p_{2m} & \cdots & p_{mm} \end{bmatrix}$$

FASTEC-VAR model

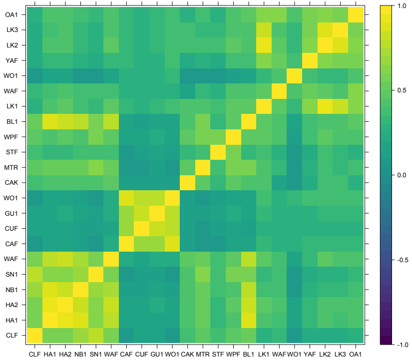
Choice of lag-order p based on AIC for r factors

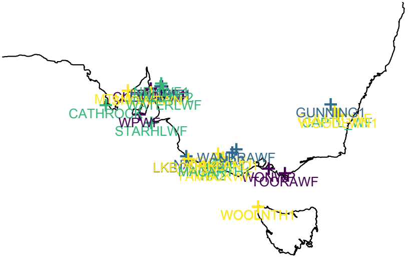
Factor	1	2	3	4	5	6	7	8
$\tau = 0.01$	0.73	0.93	0.99	1.00	1.00	1.00	1.00	1.00
$\tau = 0.5$	0.46	0.72	0.86	0.92	0.96	0.98	1.00	1.00
$\tau = 0.99$	0.75	0.94	0.99	1.00	1.00	1.00	1.00	1.00

Table 1: Choice of number load-curves for VAR modelling. Criterion: lower bound for explained variation by first r factors 95%.

Use cluster information (AEMO data)

- Perform k-means clustering on correlation matrix with $k = 4$
- Add first r loadings of each cluster member to the VAR-load model





Day ahead forecast

1 Having $\widehat{\Psi}_{T+h}(\tau)$, calculate WP curve at each τ-level:
Deploy $f_k(X_i) = \varphi_k^\top(\tau)\mathbf{X}_i$, and $\varphi_{k,s}(\tau) = \sigma_k U_{*k}^\top$

$$\widehat{\mathbf{Y}}_{T+h}(\tau) = \mathbf{XU}(\tau)\mathbf{D}(\tau)\widehat{\Psi}^\top_{T+h}(\tau).$$

2 Add seasonality Λ_{T+h}

3 Backtransform: inverse logit-transformation

$$\widehat{U}_{T+h} = [1 + \exp\{-(\widehat{\Lambda}_{T+h} + \widehat{Y}_{T+h})\}]^{-1},$$
$$\widehat{U}_{T+h} \in (0,1), \widehat{\Lambda}_{T+h}, \widehat{Y}_{T+h} \in \mathbb{R}$$

Results: one day ahead forecast

Noisy turbine-level data: ▸ AEMO

- Increase in accuracy by **13.08%**
  Cluster based FASTEC-VARX(p) with daily weather data
  VAR(p) no 5min weather data available

Aggregated TSO-level data: ▸ DK

- Increase in accuracy by **5.17%**
  FASTEC-VARX(p) with daily weather data
  ARX(p) with 5min weather data
- Increase in accuracy by **10.71%**
  MS-FASTEC-VARX(p) with daily weather data

Diebold-Mariano tests indicate significant improvement.

Discussion and sensitivity

- Cluster-based FASTEC-VARX increases accuracy wrt benchmark models ▸ Benchmark
- Probabilistic forecast: information on density of wind power generation is given via expectiles
- No need for expensive (detailed) meteorological data

Sensitivity analysis in forecasts: very short term (5 min ahead)

- equivalent level of accurary (wrt ARX)
- lower accuracy (wrt VAR)

Sensitivity analysis in forecasts: short term (12 hours ahead)

- equivalent level of accurary (wrt ARX and VAR)

For Further Reading

Browell, J and Gilbert, C (2017)
Cluster-based regime-switching AR for the EEM 2017 Wind Power Forecasting Competition
14th International Conference on the European Energy Market (EEM), Dresden, Germany, 2017

Cavalcante, L, Bessa, RJ, Reis, M and Browell, J (2017)
LASSO vector autoregression structures for very short-term wind power forecasting
Wind Energy, 20, 657-675

Chao, SK, Härdle, W and Yuan, M (2015)
Factorisable Sparse Tail Event Curves
SFB 649 Discussion Paper, **2015-034**

Dowell, J and Pinson, P (2016)
Very-Short-Term Probabilistic Wind Power Forecasts by Sparse Vector Autoregression
IEEE Transactions on Smart Grid, vol. 7, no. 2, 763-770

Ehm, W, Gneiting, T, Jordan, A and Krüger, F (2016)
Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings
J. R. Statist. Soc. B **78**,(3), 505-562

Guler, k, Ng, PT and Xiao, Z (2014)
Functional Data Analysis of Generalized Regressions for Forecast Evaluations under Asymmetric Loss Functions
Working Paper Series No. 14-01, Northern Arizona University

Härdle, W K, Huang, C and Chao, SK (2016)
Multivariate Factorisable Sparse Asymmetric Least Squares Regression
SFB 649 Discussion Paper **2016-058**

Hastie, T., Mazumder, R., Lee, J.D. and Zadeh, R. (2014)
Matrix Completion and Low-Rank SVD via Fast Alternating Least Squares
eprint arXiv:1410.2596

Lew, D, Milligan, M, Jordan, G and Piwko, R (2011)
The Value of Wind Power Forecasting
NREL Conference Paper NREL/CP-5500-50814

López Cabrera, B and Schulz, F (2016)
Time-Adaptive Probabilistic Forecasts of Electricity Spot Prices with Application to Risk Management
SFB 649 Discussion Paper **2016-035**

López Cabrera, B and Schulz, F (2016)
Forecasting Generalized Quantiles of Electricity Demand: A Functional Data Approach
Journal of the American Statistical Association, 112:517, 127-136

Yang, D, Ma, Z and Buja, A (2011)
A Sparse SVD Method for High-dimensional Data
eprint arXiv:1112.2433v1

Ziel, F, Croonenbroeck, C and Ambach, D (2016)
Forecasting wind power - Modeling periodic and non-linear effects under conditional heteroscedasticity
Applied Energy **177**, 285-297

Cluster-based short-term wind power forecasting

Wolfgang Karl Härdle*
Brenda López Cabrera*
Awdesch Melzer°

*Ladislaus von Bortkiewicz Chair of Statistics
°Institute of Finance
Humboldt-Universität zu Berlin

FISTA algorithm

1 Initialise: $\Gamma_0 = 0, \Omega_1 = 0$, step size $\delta_1 = 1$

2 For $t = 1, 2, \ldots, T$

- $\Gamma_t = \arg\min_\Gamma \left[\frac{g(\Gamma)}{L_{\bigtriangledown g}} + \frac{1}{2} \left\| \Gamma - \left\{ \Omega_t - \frac{1}{L_{\bigtriangledown g}} \bigtriangledown g(\Omega_t) \right\} \right\|^2 \right]$
- when penalising nuclear norm $\Gamma_t = \mathbf{P} \left(\mathbf{R} - \frac{\lambda}{L_{\bigtriangledown g}} \mathbf{I}_{p \times m} \right) \mathbf{Q}^\top$, and $\Omega_t - \frac{1}{L_{\bigtriangledown g}} \bigtriangledown g(\Omega_t) = \mathbf{PRQ}^\top$ with ALS-SVD (Hastie et al. (2014)) or sparse SVD (Yang et al. (2011))
- $\delta_{t+1} = \frac{1 + \sqrt{1 + 4\delta_t^2}}{2}$
- $\Omega_{t+1} = \Gamma_t + \frac{\delta_{t-1}}{t+1} (\Gamma_t - \Gamma_{t-1})$

3 $\widehat{\Gamma} = \Gamma_T$

Return

ALS-SVD algorithm

1. Initialise $A = UD$, $U_{m\times r}$ is randomly chosen matrix with orthonormal columns and $D = I_r$
2. Given A, solve for B
   - $\min_B \|X - AB^\top\|_F^\top + \lambda\|B\|_F^2$
   - $\widetilde{B}^\top = (D^2 + \lambda I)^{-1} DU^\top X$
3. Compute SVD $\widetilde{B}D = \widetilde{V}\widetilde{D}^2\widetilde{R}^\top$, let $V \leftarrow \widetilde{V}, D \leftarrow \widetilde{D}, B = VD$
4. Given B, solve for A
   - $\min_A \|X - AB^\top\|_F^\top + \lambda\|A\|_F^2$
   - $\widetilde{A}^\top = XVD(D^2 + \lambda I)^{-1}$
5. Compute SVD $\widetilde{A}D = \widetilde{U}\widetilde{D}^2\widetilde{R}^\top$, let $U \leftarrow \widetilde{U}, D \leftarrow \widetilde{D}, A = UD$
6. Repeat (2)-(5) until convergence of $AB^\top$
7. Compute $M = XV$, its SVD $M = UD_\sigma R^\top$, output: $U, V \leftarrow VR, \mathcal{S}_\lambda(D_\sigma) = diag\{(\sigma_1 - \lambda)+, \ldots, (\sigma_r - \lambda)\}$

Return

FIT-SSVD algorithm

input Observed X, target rank r, threshold function η ▸ η, initial orthonormal matrix $V^{(0)} \in \mathbb{R}^{p\times r}$, threshold level γ ▸ γ,

output $\widehat{U}, \widehat{V}$

repeat

- R-2-L-$\times$: $U^{(k),\times} = XV^{(k-1)}$
- L-thresholding: $U^{(k),thr} = \eta\left(U_{il}^{(k),\times}, \gamma_{ul}\right)$, where $\gamma_{ul} = f(X, U^{(k-1)}, V^{(k-1)}, \widehat{\sigma}), \widehat{\sigma} = 1.4826 \cdot MAD(X)$
- L-orthonormalisation QR-decomposition: $U^{(k)} R_u^{(k)} = U^{(k),thr}$
- L-2-R-$\times$: $V^{(k),\times} = X^\top U^{(k)}$
- R-thresholding: $V^{(k),thr} = \eta\left(v_{jl}^{(k),\times}, \gamma_{vl}\right)$, where $\gamma_v = f(X^\top, V^{(k-1)}, U^{(k)}, \widehat{\sigma})$
- R-orthonormalisation QR-decomposition: $V^{(k)} R_v^{(k)} = V^{(k),thr}$

until convergence

▸ Return

Thresholding function η

- Allow any thresholding function $\eta(x,\gamma)$, satisfying $|\eta(x,\gamma) - x| \leq \gamma$ and $\eta(x,\gamma)\mathbf{1}_{|x|\leq\gamma}$
  - Soft-thresholding: $\eta_{soft}(x,\gamma) = sign(x)(|x| - \gamma)_+$
  - Hard-thresholding: $\eta_{hard}(x,\gamma) = x\mathbf{1}_{|x|>\gamma}$
  - SCAD-thresholding: $\mathrm{sign}(x)\mathbf{1}_{|x|\geq\gamma}\mathbf{1}_{|x|\leq 2\gamma}\frac{((\alpha-1)x-\mathrm{sign}(x)\alpha\gamma)}{(\alpha-2)}\mathbf{1}_{2\gamma<|x|}\mathbf{1}_{|x|\geq\alpha\gamma}+x\mathbf{1}_{|x|>\alpha\gamma}$

Return

Data: $X \in \mathbb{R}^{n\times p}$, $U^{(k)} \in \mathbb{R}^{n\times r}$, $V^{(k)} \in \mathbb{R}^{p\times r}$, number of bootstraps M, standard deviation of noise $\widehat{\sigma}$

Result: Threshold level $\gamma \in \mathbb{R}^r$

Subset: $L_u = \{i : u_{i1}^{(k)} = \ldots = u_{ir}^{(k)} = 0\}, H_u = L_u^c, H_v = L_v^c;$

$L_v = \{j : v_{j1}^{(k)} = \ldots = v_{jr}^{(k)} = 0\};$

if $|L_v||L_u| < n|H_v|\log(n|H_v|)$ **then**

  return $\gamma = \widehat{\sigma}\sqrt{2\log(n)}\mathbf{1} \in \mathbb{R}^r$;

else

  for $i \leftarrow 1$ **to** M **do**

    Sample $n|H_v|$ entries from $X_{L_u L_v}$ and reshape them into a matrix $\widetilde{Z} \in \mathbb{R}^{n\times |H_v|}$;

    $B = \widetilde{Z} V_{H_{v:}}^{(k)} \in \mathbb{R}^{n\times r}$; $C_{i:} = (\|B_{:1}\|_\infty, \|B_{:2}\|_\infty, \|B_{:r}\|_\infty)^\top$;

  end

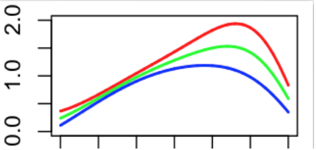
  $\gamma_l =$ median$(C_{:l})$; **return** $\gamma = (\gamma_1, \ldots, \gamma_r)^\top$.

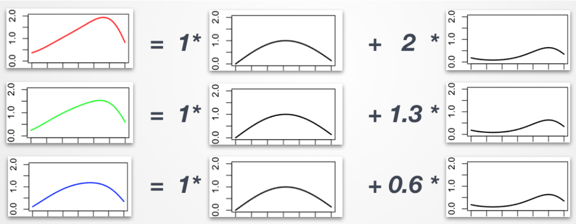
end

Return

How to find the common structure?

Set functions $X_n, n = 1, \dots, N$
Decomposed with help of
functional principal components





$$U^{\text{day}} = \Lambda^{\text{day}} + \sum_i \text{loadings}_i^{\text{day}} \cdot \text{factor}_i$$

FM Motivation
exp Factor Model

Denmark (DK) TSO-level data

- Denmark TSO area data
- 2 network areas: DK1, DK2
- Measurments:
  - 5min avg. wind power
  - 5min avg. wind speed at DK1
  - 5min avg. temperature at DK1

In-sample: 2012-01-02 - 2013-12-31
Out-of-sample: 2014-01-01 - 2014-12-31
$\Delta = 5$ min
$N = 288 \cdot 365 = 105120$
$P = 2$ areas
U_t on domain of $[0, 1]$

Data

Denmark (DK) TSO data

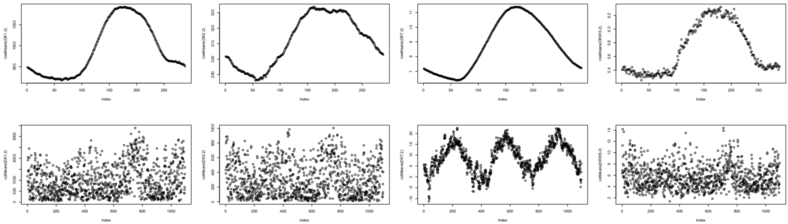


Figure 4: Within day structures, seasonality for WP in DK1, WP in DK2, WSP, TMP.

▸ Data

Seasonality: periodic B-splines

$$\mathbf{U} = \Lambda + \mathbf{Y}$$

$$\Lambda = \Lambda_d \circ \Lambda_a$$

$$\Lambda_{a,t} = \arg\min_{\alpha_j} \sum_{t=1}^{365} \left\{ \bar{U}_t - \sum_{j=1}^{J} \alpha_{a,j} \Psi_{a,j}(s_t) \right\}^2,$$

$$\Lambda_{d,t} = \arg\min_{\alpha_j} \sum_{t=1}^{288} \left\{ \tilde{U}_t - \sum_{j=1}^{J} \alpha_{d,j} \Psi_{d,j}(s_t) \right\}^2,$$

where $\Psi_j(s_t)$ is a vector of known basis functions, $\alpha_{i,j}$ are coefficients of daily d or annual a seasonality, J is the number of knots, $\bar{U}_t$: average pooled over min., $\tilde{U}_t$: average pooled over days. Ziel et al. (2016)

Return

Logit-normal adjustment

For the logit-normal transformation define an additive wind power model

$$\widetilde{U}_t = \gamma(U) \stackrel{\text{def}}{=} \log\left(\frac{U_t}{1-U_t}\right) = \Lambda_t + Y_t, \quad U_t \in (0,1),$$

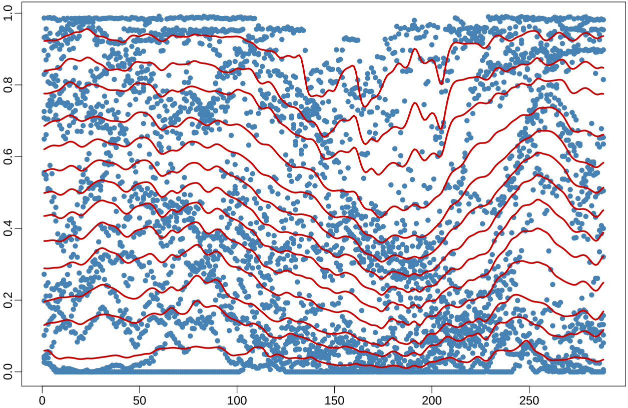
$$U_t = \gamma^{-1}(\widetilde{U}_t) \stackrel{\text{def}}{=} \{1+\exp(-\widetilde{U}_t)\}^{-1} = [1+\exp\{-(\Lambda_t+Y_t)\}]^{-1},$$

$$\widetilde{U}_t \in \mathbb{R}$$

Pinsen (2012)

Return

Expectiles



Return

Location model

$$
\begin{aligned}
&Y_i = \theta + \varepsilon_i, \quad i = 1, \ldots, n \\
&\varepsilon_i \sim F_\varepsilon \\
&\quad \text{observe data} \quad \{y_i\}_{i=1}^n = \mathbf{y}
\end{aligned}
$$

- M-Estimator: Huber (1964)
- M-Quantile: Breckling and Chambers (1988)

Estimation

$$\theta_n = \arg\min_{\theta} \sum_{i=1}^{n} \rho(Y_i - \theta)$$

Different possibilities for the loss function $\rho(u)$

M-Estimator		M-Quantile	
$\rho(u) = u^2$	Mean	$\rho(u) = \lvert\tau - \mathrm{I}\{u < 0\}\rvert u^2$	Expectile
$\rho(u) = \lvert u\rvert$	Median	$\rho(u) = \lvert\tau - \mathrm{I}\{u < 0\}\rvert \lvert u\rvert$	Quantile

Table 2: M-Quantiles use asymmetric weights

Return

FASTEC construction

- Data: $\{\mathbf{X}_i\}_{i=1}^n \in \mathbb{R}^p, \{\mathbf{Y}_i\}_{i=1}^n \in \mathbb{R}^m$ i.i.d.
- Linear model for τ-expectile curve of Y_j, $j = 1, ..., m, 0 < \tau < 1$:

$$Y_j = e_j(\tau|\mathbf{X}_i) + u_{ij,\tau} = \mathbf{X}_i^\top \Gamma_{*j}(\tau) + u_{ij,\tau}, \tag{1}$$

  where coefficients for j response: $\Gamma_{*j}(\tau) \in \mathbb{R}^p$
- Sparse factorisation: $f_k^\tau(\mathbf{X_i}) = \varphi_{\mathbf{k}}^\top(\tau)\mathbf{X_i}$ factors

$$e_j(\tau|\mathbf{X}_i) = \sum_{k=1}^r \psi_{j,k}(\tau) f_k^\tau(\mathbf{X}_i), \tag{2}$$

  where r : number of factors;
  $\Gamma_{*j}(\tau) = (\sum_{k=1}^r \psi_{j,k}(\tau)\varphi_{k,1}(\tau), ..., \sum_{k=1}^r \psi_{j,k}(\tau)\varphi_{k,p}(\tau))$

FASTEC_with_Expectiles

MER formulation: penalised loss

$$\widehat{\Gamma}_\lambda(\tau) = \arg \min_{\Gamma \in \mathbb{R}^{p \times m}} \left\{ (mn)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \rho_\tau \left(Y_{ij} - \mathbf{X}_i^\top \Gamma_{*j} \right) + \lambda \|\Gamma\|_* \right\},$$

$\|\Gamma\|_* = \sum_{j=1}^{\min(p,m)} \sigma_j(\Gamma)$ nuclear norm of Γ

$\mathbf{X}_i$: B-splines ▸ B-splines

$\mathbf{Y}_i$: detrended wind power

Γ : factor matrix

λ : penalisation parameter (optimality via CV)

Chao et al. (2015), Härdle et al (2016)

▸ FISTA algorithm ▸ Return

Factor Curves

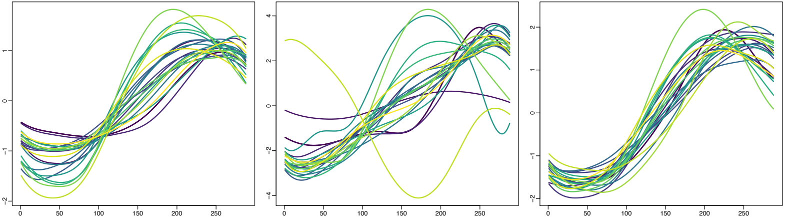


Figure 5: First five factors $f_j^{\tau}(\mathbf{X_i}) = \varphi_{\mathbf{j}}^{\top}(\tau)\mathbf{X_i}$, at 1%, 50% and 99% τ-level (from left to right).

WERMVAR

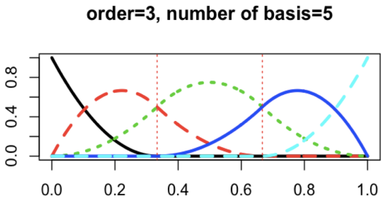
Return

B-spline basis

Knot vector $t = (t_1, \ldots, t_M)$ as non-decreasing sequence in $[0, 1]$
Define ith B-spline basis function $N_{i,j}$ of order j as

$$N_{i,0}(t) = \begin{cases} 1 & \text{if } t_i < t < t_{i+1} \\ 0 & \text{otherwise} \end{cases}$$

$$N_{i,j}(t) = \frac{t - t_i}{t_{i+j} - t_i} N_{i,j-1}(t) + \frac{t_{i+j+1} - t}{t_{i+j+1} - t_{i+1}} N_{i+1,j+1}(t)$$



MER Adjustments

Benchmark: SCAD FASTEC-VARX model

Introduce SCAD variable selection by minimising

$$\arg\min_{\Theta,\lambda} \frac{1}{2}(\mathbf{Y} - \Theta\mathbf{Z})^\top(\mathbf{Y} - \Theta\mathbf{Z}) + T^* \sum_{j=1}^{K\cdot P} \mathfrak{p}_\lambda(|\theta|_j).$$

$$\mathfrak{p}_\lambda(|\theta|) = \begin{cases} \lambda|\theta| & \text{if} |\theta| \leq \lambda \\ -\frac{|\theta|^2 - 2\alpha\lambda|\theta| + \lambda^2}{2(\alpha-1)} & \text{if} \lambda \leq |\theta| \leq \alpha\lambda \\ \frac{(\alpha+1)\lambda^2}{2} & \text{if } |\theta| \geq \alpha\lambda, \end{cases}$$

Return

Benchmark: Sparse VAR model

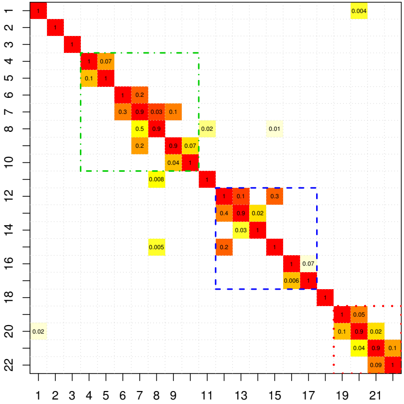
1. Stage: partial spectral coherence (PSC):
neg. scaled inverse of spectral density

$$PSC_{ij}(\omega) = -\frac{g^Y_{ij}(\omega)}{\sqrt{g^Y_{ii}(\omega) g^Y_{jj}(\omega)}}, \quad \omega \in (-\pi, \pi]$$

$$S_{ij} = \sup_{\omega} |PSC_{ij}(\omega)|$$

$g^Y(\omega) = f^Y(\omega)^{-1}$: inverse density
$S_{i,j}$: conditional correlation of turbine i with j

Sparse VAR model



2. Stage: variable selection by ranking according to t-stat

Dowell & Pinsen (2016) ▸ Return

Benchmark: VAR(p)

Assuming all intra-day observations being "variables" of each day

$$\mathbf{Y}_k = \sum_{i=1}^{p} \Theta_i \mathbf{Y}_{k-i}(\tau) + \eta_k$$

Y_k: matrix of deseasonalised utilisation
Θ_i: matrix of VAR coefficients
η_k: white noise error term
Pre-apply logit-normal transform due to support on (0,1):

$$y = \gamma(x) = \log\left(\frac{x}{1-x}\right)$$
$$x = \gamma^{-1}(y) = \{1 + \exp(-y)\}^{-1}$$

Return

Cluster-based: AEMO

	Cluster	VAR(X)				FASTEC-VAR(X)					
		as.fun	as.fun X	plain	sparse	plain	VARX	SCAD	SCADX	MS	MSX
RMSE	1	0.303	0.314	0.306	0.300	0.305	**0.269**	0.298	0.334	0.351	0.303
	2	0.280	0.274	0.291	0.263	0.282	**0.247**	0.284	0.348	0.337	0.287
	3	0.276	0.299	0.275	0.271	0.279	**0.247**	0.275	0.330	0.302	**0.268**
	4	0.277	0.267	0.265	0.278	0.267	**0.245**	0.264	0.316	0.300	0.273
MSE	1	0.092	0.099	0.094	0.090	0.093	**0.073**	0.089	0.112	0.124	0.092
	2	0.079	0.075	0.085	0.070	0.080	**0.061**	0.081	0.121	0.113	0.083
	3	0.077	0.091	0.076	0.074	0.078	**0.061**	0.076	0.109	0.092	**0.072**
	4	0.077	0.072	0.071	0.078	0.072	**0.060**	0.070	0.100	0.090	0.075
MAE	1	0.227	0.232	0.230	0.225	0.234	**0.202**	0.230	0.276	0.280	0.226
	2	0.204	0.200	0.219	0.192	0.211	**0.179**	0.216	0.274	0.244	0.211
	3	0.209	0.230	0.211	0.201	0.209	**0.182**	0.208	0.261	0.236	0.201
	4	0.208	0.196	0.195	0.209	0.200	**0.179**	0.198	0.247	0.217	0.203
MAPE	1	4.098	**2.463**	3.270	2.930	3.835	3.110	3.806	5.156	7.264	3.016
	2	2.891	3.824	5.504	3.183	3.387	**2.533**	3.443	4.448	**2.191**	3.611
	3	3.799	4.951	3.769	2.970	2.923	**2.520**	2.990	4.248	4.879	3.534
	4	3.670	2.376	2.747	3.732	2.720	**2.259**	2.642	3.512	**2.001**	3.377
MASE	1	0.763	0.779	0.771	0.755	0.786	**0.678**	0.772	0.924	0.941	**0.758**
	2	0.746	0.734	0.803	0.702	0.777	**0.656**	0.792	1.006	0.896	0.776
	3	0.765	0.842	0.769	0.735	0.761	**0.665**	0.759	0.951	0.861	**0.736**
	4	0.767	0.719	0.715	0.771	0.737	**0.659**	0.729	0.907	0.802	0.751

Return

	DK 1					DK 2				
	RMSE	MSE	MAE	MAPE	MASE	RMSE	MSE	MAE	MAPE	MASE
VAR	0.2203	0.0485	0.1555	1.1577	0.8576	0.2203	0.0485	0.1563	1.5567	0.8601
VAR+WSP	0.2167	0.0470	0.1524	1.2409	0.8405	0.2201	0.0485	0.1555	1.7467	0.8557
VAR+TMP	0.2203	0.0486	0.1556	1.1725	0.8583	0.2205	0.0486	0.1565	1.5713	0.8610
VAR+TMP+WSP	0.2147	0.0461	0.1491	1.1365	0.8226	0.2135	0.0456	0.1498	1.5749	0.8240
AR	0.1775	0.0315	0.1257	0.9286	0.6934	0.1908	0.0364	0.1379	1.2292	0.7593
AR+WSP	0.1775	0.0315	0.1257	0.9285	0.6933	0.1908	0.0364	0.1379	1.2290	0.7591
AR+TMP	0.1776	0.0315	0.1255	0.9304	0.6926	0.1907	0.0364	0.1378	1.2311	0.7591
AR+TMP+WSP	0.1760	0.0310	0.1247	0.9300	0.6880	0.1906	0.0363	0.1378	1.2312	0.7587
FAST	0.1848	0.0341	0.1320	0.9062	0.7281	0.1869	0.0349	0.1382	1.3995	0.7607
FAST+WSP	**0.1706**	**0.0291**	**0.1245**	0.9835	0.6868	**0.1738**	**0.0302**	**0.1310**	1.5989	**0.7208**
FAST+TMP	0.1836	0.0337	0.1316	0.9237	0.7260	0.1864	0.0347	0.1381	1.4142	0.7599
FAST+TMP+WSP	**0.1686**	**0.0284**	**0.1218**	**0.8876**	**0.6720**	**0.1734**	**0.0301**	**0.1272**	1.4239	**0.6999**
FAST SCAD	0.1848	0.0341	0.1320	0.9062	0.7281	0.1870	0.0350	0.1383	1.4029	0.7611
FAST MS	0.1899	0.0360	0.1355	0.9254	0.7472	0.1936	0.0375	0.1433	1.4545	0.7885
FAST MS TMP+WS	**0.1551**	**0.0241**	**0.1107**	**0.6813**	**0.6104**	**0.1762**	**0.0310**	**0.1238**	**0.7561**	**0.6813**

Table 3: Left DK 1, right DK 2. Evaluation of different forecasting methods with various accuracy measures.

Return

Accuracy measures

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\widehat{y}_t - y_t)^2}$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_t - \widehat{y}_t}{y_t}\right|$$

$$q_t = \frac{y_t - \widehat{y}_t}{\frac{1}{T-1}\sum_{s=2}^{T}|y_s - y_{s-1}|}$$

$$MASE = \frac{1}{T}\sum_{t=1}^{T}|q_t|$$

Return