Wind Energy Risk Modelling

Wolfgang Karl Härdle
Brenda López Cabrera
Awdesch Melzer

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

Hedging on wind

Motivation

- Energy generation soon dominated by renewables
- Wind power highly depends on wind conditions
- Sudden wind speed changes of e.g. 11m/s $\rightarrow$ power drop by 80%
- Unsteady wind direction may result in zero wind power generation

- Knowledge about variability of wind is essential for
  - Sustainable energy supply
  - Energy management
  - Hedging wind energy futures

Weather futures

- In agriculture: hedging against
  - Rain, drought, snowfall
- In renewable markets: hedging against
  - Unexpected wind conditions, sun duration, clouds

- Market is growing
- Products are still in development
- Site-specific weather futures (e.g. Chicago Weather Brokerage)

Wind futures: soon at EEX

- Contract: A contract settling against the expected on-site power production of future delivery periods
- Underlying: Average wind load factor per contract period. Alternatives: wind direction, wind speed, wind duration indices.
    - Average daily wind speed over past 20 years as basis
    - Index: Expectation of wind energy for specific period

$$CAWS(\tau_1;\tau_2) = \int_{\tau_1}^{\tau_2} W(s)ds$$

Wind risk

- Risk exposure of wind farms depends on
  - Wind speed
  - Wind direction
- Wind speed consists of three components
  1. Seasonal
  2. Intraday-periodicity
  3. Stochastic term
- Understand core drivers of stochastic component
- Incorporate tail risk using measures of tail-events

Seasonal wind direction dependence

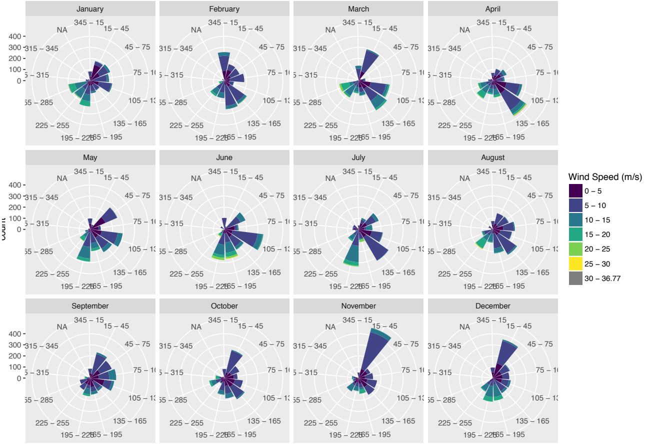


Figure 1: Wind directions for 2014 at Spain, Lugo

Seasonal variance in wind speed

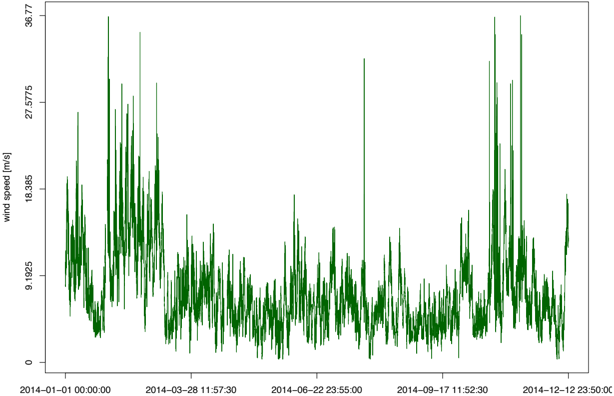


Figure 2: 10-min average Wind speed for 2014 at Spain, Lugo

Understanding volatility: FASTEC

Factorisable Sparse Tail Event Curves on wind speed volatility:

- Common structure
  - Ultra high dimensional (UHD) time series with factors
  - Sparse penalization
- Individual variety
  - Tail behaviour
  - Spread analysis on factor loadings

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

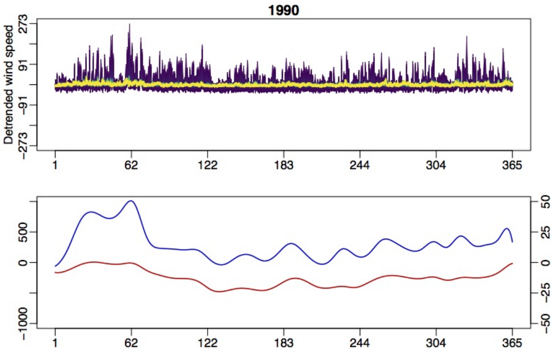


Figure 3: Detrended wind speed and first factor $f_1^\tau(\mathbf{X_i}) = \varphi_{\mathbf{1}}^\top(\tau)\mathbf{X_i}$, 10% and 90% τ-level.

WERMtest

WERM: Challenges

- Wind speed at turbine height
- Wind energy index calculation
- FASTEC in UHD
- Forecast tail moments of wind speed
- Wind energy futures

Agenda

1. Motivation ✓
2. Data preparation
3. FASTEC on wind data
4. Wind energy futures
5. Outlook

Meteorological data

- MERRA reanalysis grid data: Germany & Spain
- Hourly wind speed
- Height levels: 2m, 10m, 50m
- Time horizon: 1990-01-01 - 2014-12-31
- Spatial resolution: $0.5^\circ \times 0.67^\circ$

- MERRA hourly temperature data
- MERRA hourly surface pressure data
- MERRA hourly wind direction data

Wind speed at turbine height

- Extrapolate wind speed to higher levels
  - Extended power law (Sen et al. (2012)) Extended power law
  - Robust shear exponent (Istchenko and Turner (2008)) Robust shear exponent $\widehat{\alpha}_t$

	EPL	MERRA50
RMSE	2.992	3.407
MAPE	0.449	0.570

Table 1: RMSE and MAPE for EPL on MERRA10 data and for MERRA50 data from Spain, Lugo

RSME MAPE

Measures of tails events

- Loss function, Breckling and Chambers (1988)

  $\rho_{\tau,\gamma}(u) = |\tau - \boldsymbol{I}\{u < 0\}|\,|u|^{\gamma}, \quad \gamma \geq 1$

  $z_\tau = \arg\min_{\theta} \mathsf{E}\,\rho_{\tau,\gamma}(Y - \theta)$

  - Quantile - ALD location estimate
    $q_\tau = \arg\min_{\theta} \mathsf{E}\,\rho_{\tau,1}(Y - \theta)$

  - Expectile - AND location estimate
    $e_\tau = \arg\min_{\theta} \mathsf{E}\,\rho_{\tau,2}(Y - \theta)$

Loss Function

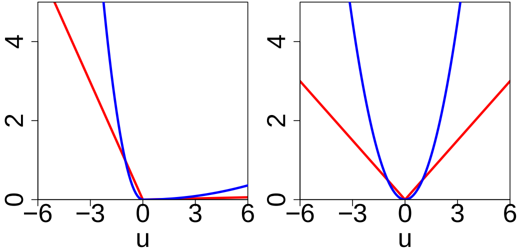


Figure 4: Expectile and quantile loss functions at $\alpha = 0.01$ (left) and $\alpha = 0.50$ (right)

LQRcheck

Tail event structures

- Spread: of $\{e(\tau), e(1-\tau)\}$ τ-range, its changes (expansion, shrinkage, shifting), $0 < \tau < 1$
- Tail: $\tau \approx \{0, 1\}$. Tail event curves (TEC)

Dimension reduction: related structure reduced to lower order factors

FASTEC construction

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

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

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

$\mathbf{Y}_i$: wind power density; $(n \times m)$-matrix

Γ : factor matrix

λ : penalisation parameter Optimal λ

FISTA algorithm

FASTEC: in ultra high dimensional space?

- FISTA: At each iteration SVD
- Singular Value Decomposition, Principal Component Analysis:
  - $n \to \infty, m = const.$
  - $m \to \infty, n = const.$

  not consistent

Solution: matrix approximation with low-rank features

Cost: increased bias

FASTEC in UHD

- Alternating Least Squares SVD
  - fit a low-rank SVD to a matrix by alternating orthogonal ridge regression ▸ ALS

  Hastie et al. (2014)

- Sparse SVD
  - fast iterative thresholding for low-rank SVD ▸ sSVD

  Yang et al. (2011)

Factor Curves

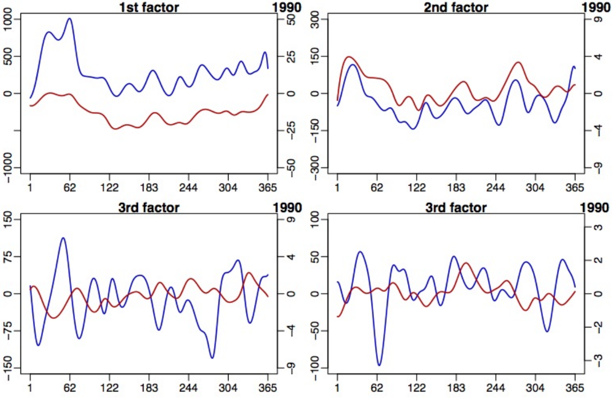


Figure 5: First factors $f_k^\tau(\mathbf{X_i}) = \varphi_\mathbf{k}^\top(\tau)\mathbf{X_i}$, 10% and 90% τ-level.

Wind speed and 1st factor

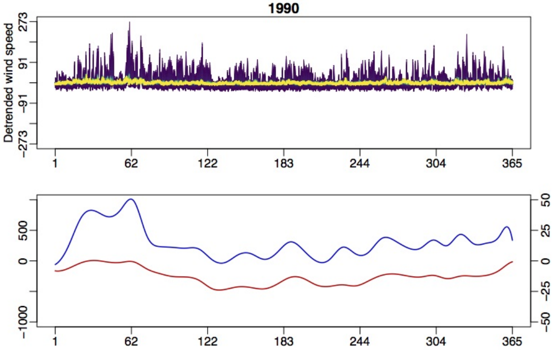


Figure 6: Detrended wind speed and first factor $f_1^{\tau}(\mathbf{X_i}) = \varphi_{\mathbf{1}}^{\top}(\tau)\mathbf{X_i}$, 10% and 90% τ-level.

WERMtest

Spatial loadings

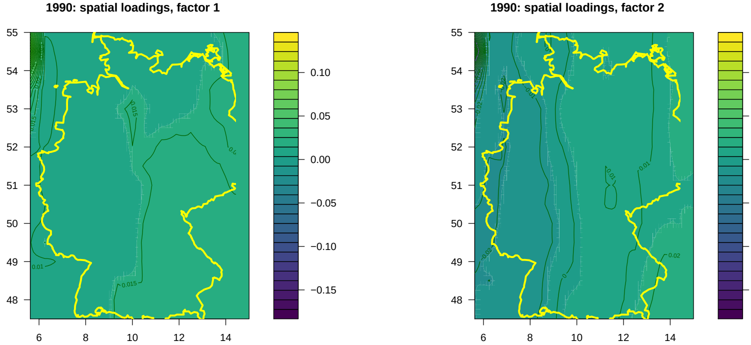


Figure 7: Spatial loadings $\psi_1(\tau)$,$\psi_2(\tau)$ at 10%. WERMtest

Spatial loadings

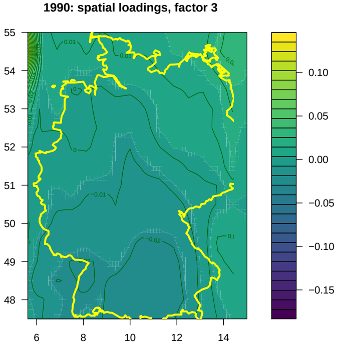


Figure 8: Spatial loadings $\psi_3(\tau)$,$\psi_4(\tau)$ at 10%. WERMtest

Spatial loadings

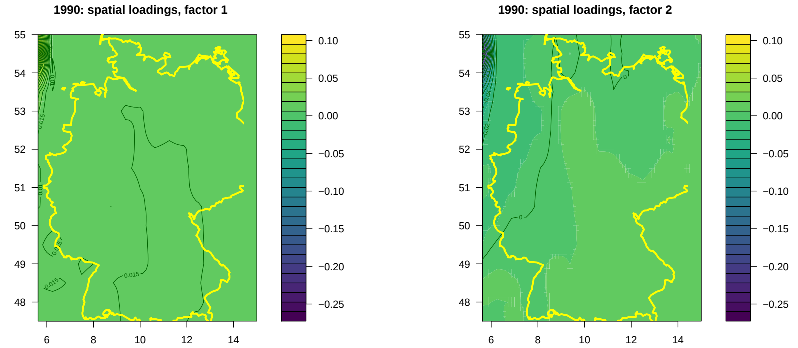
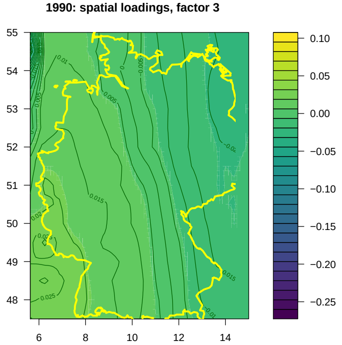


Figure 9: Spatial loadings $\psi_1(\tau)$,$\psi_2(\tau)$ at 90%. WERMtest

Spatial loadings



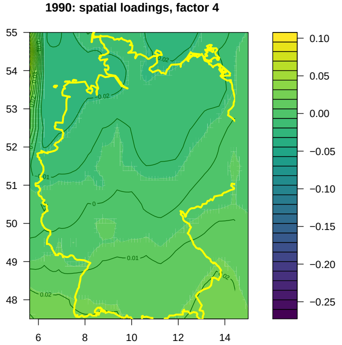


Figure 10: Spatial loadings $\psi_3(\tau)$,$\psi_4(\tau)$ at 90%.

Computation time and error

	ALS	sSVD	eFASTEC	FASTEC
Loss [$\tau = 10\%$]	-4142130	-4142160	-4142130	-
Error	-2183345	-2183336	-2183345	-
Time [s]	2865	2798	3442	>25000
Loss [$\tau = 90\%$]	9683058	9683140	9683058	-
Error	-18639.96	-18640.34	-18639.96	-
Time [s]	11997	11377	14178	>25000

Table 2: Loss, error and computation time for ALS-FASTEC, sSVD-FASTEC, efficient-FASTEC and original expectile FASTEC

Wind energy futures

- Risk products: wind futures

$$CAWS(\tau_1;\tau_2) = \int_{\tau_1}^{\tau_2} W(s)ds$$

- How to smooth seasonal mean, variance?
- How close are the residuals to $N(0,1)$?
- How to infer the market price of wind risk

Algorithm

$$
\begin{gathered}
W_t \\
\downarrow \\
X_t = W_t - \Lambda_t \\
\downarrow \\
X_t = \beta^\top X_{t-L} + \varepsilon_t, \quad \varepsilon_t = \sigma_t e_t \\
\downarrow \\
e_t = \frac{\widehat{X}_t}{\widehat{\sigma}_t} \sim \mathrm{N}(0,1)
\end{gathered}
$$

Härdle et al. (2011)

First results: the time series

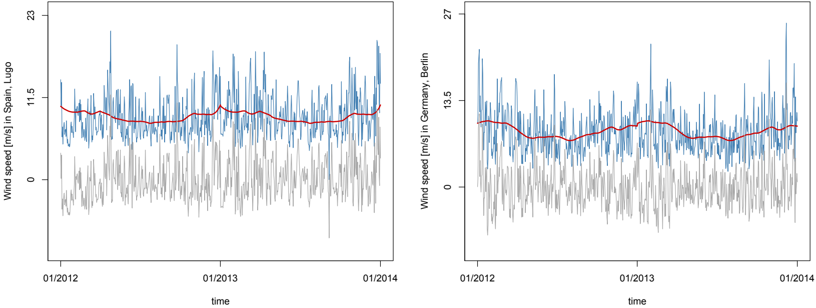


Figure 11: Time series of daily average wind speed of Spain, Lugo and Germany, Berlin.

Seasonal variance

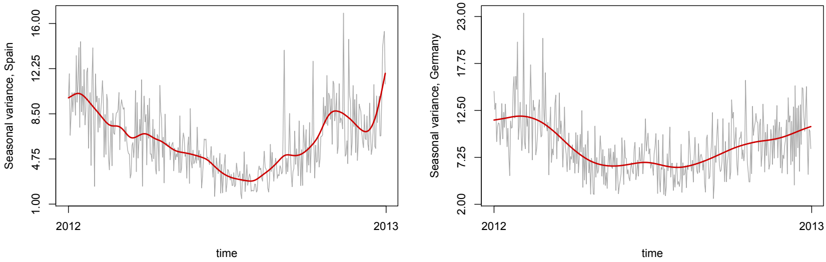


Figure 12: Seasonal variance of wind speed of Spain, Lugo and Germany, Berlin.

Outlook

- Incorporate the spread of factors from FASTEC
- Joint distribution (e.g. with copulae) of wind and solar indices

For Further Reading

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

Chao, S. K., Proksch, K., Dette, H.and Härdle, W. (2015)
Confidence Corridors for Multivariate Generalized Quantile Regression
Journal of Business & Economic Statistics, accepted 20150611

Ehm, W., Gneiting, T., Jordan, A., Krüger, F. (2015)
Of Quantiles and Expectiles: Consistent Scoring Functions, Choquet Representations, and Forecast Rankings
eprint: http://arxiv.org/pdf/1503.08195.pdf

George, S.O., George, H.B. and Nguyen, S.V. (2011)
Risk Quantification Associated With Wind Energy Intermittency in California
Power Systems, IEEE Transactions on 26 (4), 1937–1944

Härdle, W. K., Huang, C. and Chao, S.K. (2016)
FASTEC with expectiles
SFB 649 Discussion Paper **2016-008**

Härdle, W. K., Müller, M., Sperlich, S., Werwatz, A. (2004)
Nonparametric and Semiparametric Models
Springer, Berlin

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

Istchenko, R. and Turner, B. (2008)
Extrapolation of Wind Profiles Using Indirect Measures of Stability
Wind Engineering 32 (5), 433–438

Jones, F.E. (1978)
The Air Density Equation and the Transfer of the Mass Unit
Journal of Research of the National Bureau of Standards, **83** (5), 419–428

Ritter, M., Shen, Z. López Cabrera, B., Odening, M., Deckert, L. (2015a)
A New Approach to Assess Wind Energy Potential
Energy Procedia, 75, 671–676

Ritter, M., Shen, Z. López Cabrera, B., Odening, M., Deckert, L. (2015b)
Designing an index for assessing wind energy potential
Renewable Energy 83, 416–424

Şen, Z., Altunkaynak, A. and Erdik, T. (2012)
Wind Velocity Vertical Extrapolation by Extended Power Law
Advances in Meteorology, doi:10.1155/2012/178623

Ward, K. and Boland,J. (2007)
Modelling the Volatility in Wind Farm Output
eprint:
http://refman.et-model.com/publications/1808/download

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

Wind Energy Risk Modelling

Wolfgang Karl Härdle
Brenda López Cabrera
Awdesch Melzer

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt-Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

Data

- MERRA for wind profile extrapolation from $(\nu_2, \nu_{10}, \nu_{50})$
  - 01.01.1980-31.12.2014, hourly, 0.5×0.67 deg resolution
- ERA-Interim for wind power density (WPD) estimation
  - 01.01.2010-31.12.2014, 3 hourly, 0.125×0.125 deg resolution
- MARS forecasts for multivariate expectile regression (FASTEC)
  - 01.01.2010-31.12.2014, 6 hourly, 0.5×0.5 deg resolution

WPE: logarithmic law

Logarithmic law

$$\nu_z = \left(\frac{\nu_*}{\kappa}\right) \log\left\{\frac{(z-d)}{z_0}\right\}$$

ν_z : velocity at height z
ν_* : const. friction velocity
d : displacement height
κ : Kármán const. ≈ 0.41
z_0 : surface roughness Return

WPE: power law

Power law (Istchenko & Turner (2008))

$$\nu_{z_2} = \nu_{z_1} \left(\frac{z_2}{z_1} \right)^{\alpha_t}$$

ν_{z_2} : velocity at height z_2
ν_{z_1} : velocity at reference height z_1, $z_2 > z_1$
α_t : power law coefficient Return

WPE: extended power law

Extended power law (Sen et al. (2012))

$$\left(\frac{z_2}{z_1}\right)^{\alpha_t} = \left(\frac{\nu_{z_2}}{\nu_{z_1}}\right)$$

rewritten in terms of time averages $\bar{\nu}_{z_i} = T^{-1}\sum_{t=1}^{T}\nu_{z_i,t}$ and perturbation $s_{\nu_{z_i}} = T^{-1}\sum_{t=1}^{T}(\nu_{z_i,t} - \bar{\nu}_{z_i})^2$ at height z_i and 4th order approximation of Binomial expansion.

$$\left(\frac{z_2}{z_1}\right)^{\alpha_t} = \left(\frac{\bar{\nu}_{z_2}}{\bar{\nu}_{z_1}}\right)\left\{1 - \frac{\mathrm{Cov}(\nu_{z_2}\nu_{z_1})}{\bar{\nu}_{z_2}\bar{\nu}_{z_1}} + \frac{s^2_{\nu_{z_1}}}{\bar{\nu}^2_{z_1}} + \frac{s^4_{\nu_{z_1}}}{\bar{\nu}^4_{z_1}}\right\}$$

$$\alpha_t = \frac{\log\left(\frac{\bar{\nu}_{z_2}}{\bar{\nu}_{z_1}}\right) + \log\left\{1 - \frac{\mathrm{Cov}(\nu_{z_2}\nu_{z_1})}{\bar{\nu}_{z_2}\bar{\nu}_{z_1}} + \frac{s^2_{\nu_{z_1}}}{\bar{\nu}^2_{z_1}} + \frac{s^4_{\nu_{z_1}}}{\bar{\nu}^4_{z_1}}\right\}}{\log\left(\frac{z_2}{z_1}\right)}$$

with $t = 1, \ldots, T,\ T = 25$ years $\times$ 365 days $\times$ 24 hours.

Return

Robust shear exponent

Istchenko and Turner (2008) propose

- Median $\widehat{\alpha}_t$ over time of day
- $\widehat{\alpha}_t$ estimated by
  - Power law Power law
  - Extended power law Extended power law

Return

Local linear smoothing on $\alpha_{d,h}$

Solving (Härdle et al. (2004))

$$\min_{m,\beta} \sum_{d=1}^{D} \{Y_d - m - (x - x_d)^\top \beta\}^2 K_h(x - x_d),$$

where

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_D \end{pmatrix}, X = \begin{Bmatrix} 1 & (x - x_1)^\top \\ \vdots & \vdots \\ 1 & (x - x_D)^\top \end{Bmatrix}$$

$$W = \operatorname{diag}\{K_h(x - x_1), \dots, K_h(x - x_D)\}$$

leads to

$$\widehat{m}_{1,h}(x) = e_0^\top (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{Y}$$

with $e_0 = (1, 0, \dots, 0)^\top$.

Optimal bandwidth h for LLS

Estimate h by minimising the cross-validation equation (Härdle et al. (2004))

$$\arg\min_h CV(h;X) = \arg\min_h \sum_{d=1}^{D} \{Y_d - \widehat{m}_{h,-d}(X_d)\}^2 ,$$

where

$$\widehat{m}_{h,-d}(X_d) = \sum_{j\neq d} \frac{K_h(X_d - X_j)Y_j}{\sum_{j\neq d} K_h(X_d - X_j)}$$

Return

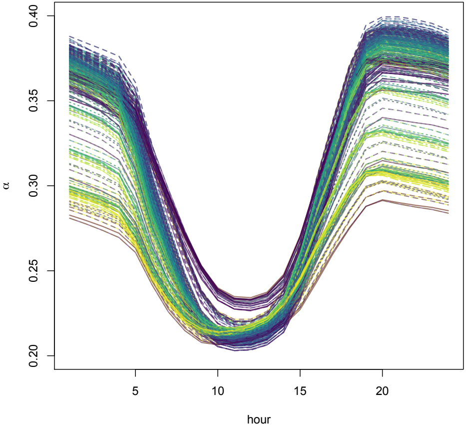
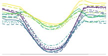


Figure 13: Extended power law: averaged daily $\widehat{\alpha}_t$, based on MERRA velocity data from 01.01.1980-31.12.2014. ▸ Return



Air density estimation

Saturation pressure of water vapour E_s based on dewpoint temperature T_d and coefficients c_0, c_1, c_2

$$E_s = c_0 \cdot 10^{c_1 \frac{T_d}{c_2 + T_d}}$$

Use relationship between actual vapour pressure P_v with the saturation pressure E_s at dew point

$$P_v = E_s$$

Dry air pressure P_d, decomposed into total pressure P and pressure due to water vapour P_v

$$P_d = P - P_v$$

Air density estimation

Substitution of above equations leads to air density ρ given temperature T gas constants dry air R_d and water vapour R_v

$$\rho = \frac{P_d}{R_d \cdot T} + \frac{P_v}{R_v \cdot T}$$

Jones (1978)

Return

Potential wind power density

Theoretical wind power density for a fully efficient $\xi = 1$, (standard efficiencies $\xi = 0.2 \sim 0.4$) wind mill

$$WPD_{i,j,t} = \frac{1}{2} \cdot \rho_{i,j,t} \cdot A \cdot \nu^3_{i,j,t}, \quad i \times j : \text{ lat-lon-grid}, t = d \cdot s$$

where A is the area covered by the rotor blades, ρ is air density, ν is velocity.

Return

Fast Iterative Shrinkage Thresholding Algorithm

- Objective: $\min_\Gamma \left\{ F(\Gamma) \overset{def}{=} g(\Gamma) + h(\Gamma) \right\}$
- g: smooth convex function with Lipschitz continuous gradient

$$\| \bigtriangledown g(\Gamma_1) - \bigtriangledown g(\Gamma_2) \|_F \leq L_{\bigtriangledown g} \| \Gamma_1 - \Gamma_2 \|_F, \quad \forall \Gamma_1, \Gamma_2$$

  where $L_{\bigtriangledown g} = 2(mn)^{-1} \max(\tau, 1-\tau) \|X\|_F^2$ is the Lipschitz constant of $\bigtriangledown g$
- h: continuous convex function, possibly nonsmooth
- $|F(\Gamma_t) - F(\Gamma^*)| \leq \frac{2L_{\bigtriangledown g} \|\Gamma_0 - \Gamma^*\|_F^2}{(t+1)^2}$

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: $\text{sign}(x)\mathbf{1}_{|x|\geq\gamma}\mathbf{1}_{|x|\leq 2\gamma}\frac{((\alpha-1)x-\text{sign}(x)\alpha\gamma)}{(\alpha-2)}\mathbf{1}_{2\gamma<|x|}\mathbf{1}_{|x|\geq\alpha\gamma}+x\mathbf{1}_{|x|>\alpha\gamma}$

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

Some assumptions

- $\{(\mathbf{X}_i, \mathbf{Y}_i)\}_{i=1}^n \in \mathbb{R}^{p+m}$ are i.i.d., $\{X_i\}_{i=1}^n \in \mathbb{R}^p \sim \mathrm{N}(0, \Sigma)$
- Conditional on $\mathbf{X}_i$, $u_{ij} = \{Y_{ij} - \mathbf{X}_i^\top \Gamma_{\cdot j}\}_{j=1}^m$ are cross-sectional independent over j
- $u_{i1}, \ldots, u_{im}$ are sub-gaussian: $\exists C > 0$ s.t. $\mathrm{P}(|u_{ij}| > s) \leq \exp\{1 - \left(\frac{s}{C}\right)^2\}, j \in \{1, \ldots, m\}$
- $K_u \overset{def}{=} \max_{1 \leq j \leq m} \|U_{ij}\|_{\psi_2} = \max_{1 \leq j \leq m} \sup_{p \leq 1} p^{-\frac{1}{2}} (\mathrm{E}\,|u_{ij}|^p)^{\frac{1}{p}}$

Optimal choice of tuning parameter λ

- Under the assumptions of sample setting, selecting

$$\lambda = \frac{2}{m} S \max(\tau, 1-\tau) \sqrt{K_u^2 \|\Sigma\| \frac{p+m}{n}}, \quad \text{for } n \leq 2\min(m, p)$$

  any optimal solution $\widehat{\Gamma}_\lambda$ satisfies error bound conditions (Huang et al. (2016))

- Then

$$P\left\{ \| \triangledown g(\Gamma) \| \leq m^{-1} S \max(\tau, 1-\tau) \sqrt{K_u^2 \|\Sigma\| \frac{p+m}{n}} \right\} \leq 1 - 3 \cdot 8^{-(p+m)} - 4\exp(-n/2),$$

  where S is an absolute constant.

Return

Root mean squared error

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

▸ Return

Mean absolute percentage error

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

Return