On machine learning methods for the estimation of conditional Kendall's tau

Alexis Derumigny, CREST - ENSAE, Palaiseau (France)

Link between Kendall's tau and copulas

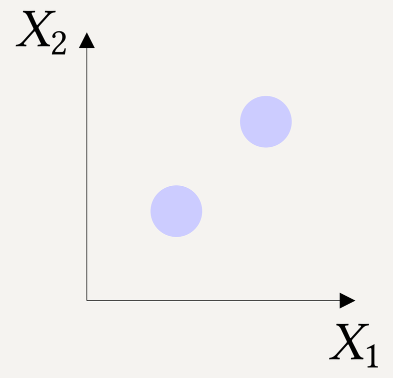
Sklar's theorem:

$$\forall \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2,\ F_{\mathbf{X}}(\mathbf{x}) = C\big(F_1(x_1), F_2(x_2)\big),$$

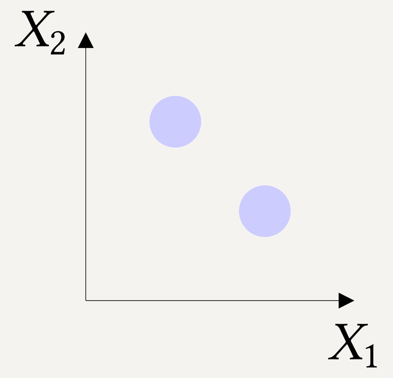
Kendall's tau:

$$\begin{aligned}\tau_{1,2} = \tau(C) &:= 4 \int_{[0,1]^2} C(u,v)\,dC(u,v) - 1 \\ &= \mathbb{P}\big((X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) > 0\big) - \mathbb{P}\big((X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) < 0\big) \\ &= \mathbb{P}\big((\mathbf{X}_{1,1:2}, \mathbf{X}_{2,1:2}) \text{ is a concordant pair}\big) \\ &\qquad - \mathbb{P}\big((\mathbf{X}_{1,1:2}, \mathbf{X}_{2,1:2}) \text{ is a discordant pair}\big),\end{aligned}$$

where $\mathbf{X}_{1,1:2}, \mathbf{X}_{2,1:2} \overset{\text{i.i.d.}}{\sim} F_{\mathbf{X}}$.



A concordant pair



A discordant pair

Conditional Kendall's tau: a measure of conditional dependence

Conditional Kendall's tau between X_1 and X_2 given $\mathbf{Z} = \mathbf{z}$:

$$\begin{aligned}\tau_{1,2|\mathbf{Z}=\mathbf{z}} := \tau(C_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}) &= \mathbb{P}\big((X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) > 0 \,\big|\, \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}\big) \\ &\quad - \mathbb{P}\big((X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) < 0 \,\big|\, \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}\big),\end{aligned}$$

where $(\mathbf{X}_{1,1:2}, \mathbf{Z}_{1,1:p})$ and $(\mathbf{X}_{2,1:2}, \mathbf{Z}_{2,1:p})$ are two i.i.d. copies of a random vector $(\mathbf{X}, \mathbf{Z}) \in \mathbb{R}^{2+p}$.

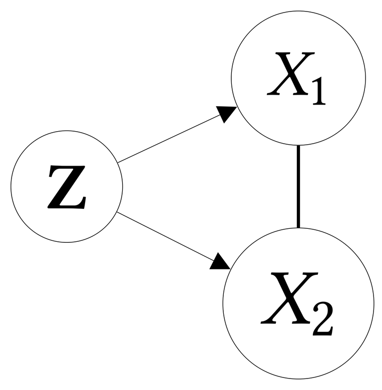


Figure: The "simplifying assumption": $\mathbf{Z}$ has an influence on the conditional margins X_1 and X_2, but not on the conditional dependence between them.

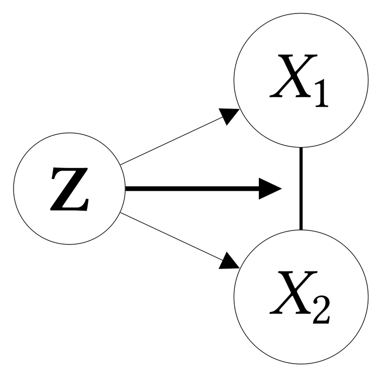


Figure: The general case: $\mathbf{Z}$ has an influence on the conditional margins X_1 and X_2, and also on their conditional dependence.

Goals

- Modeling the influence of $\mathbf{Z}$ on the dependence between X_1 and X_2
- with a conditional dependence parameter that always exists
- and is invariant by changes of margins and scales.

A kernel-based estimator

For a sample $(X_{i,1}, X_{i,2}, \mathbf{Z}_i)$, $i = 1, \dots, n$,

$$\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}} := \sum_{i=1}^n \sum_{j=1}^n w_{i,n}(\mathbf{z}) w_{j,n}(\mathbf{z}) \times \Big(\mathbf{1}\big\{(X_{i,1} - X_{j,1})(X_{i,2} - X_{j,2}) > 0\big\} - \mathbf{1}\big\{(X_{i,1} - X_{j,1})(X_{i,2} - X_{j,2}) < 0\big\}\Big),$$

where $w_{i,n}(\mathbf{z}) := K_h(\mathbf{Z}_i - \mathbf{z}) / \sum_{j=1}^n K_h(\mathbf{Z}_j - \mathbf{z})$ for a kernel K on $\mathbb{R}^p$ and a bandwidth $h > 0$.

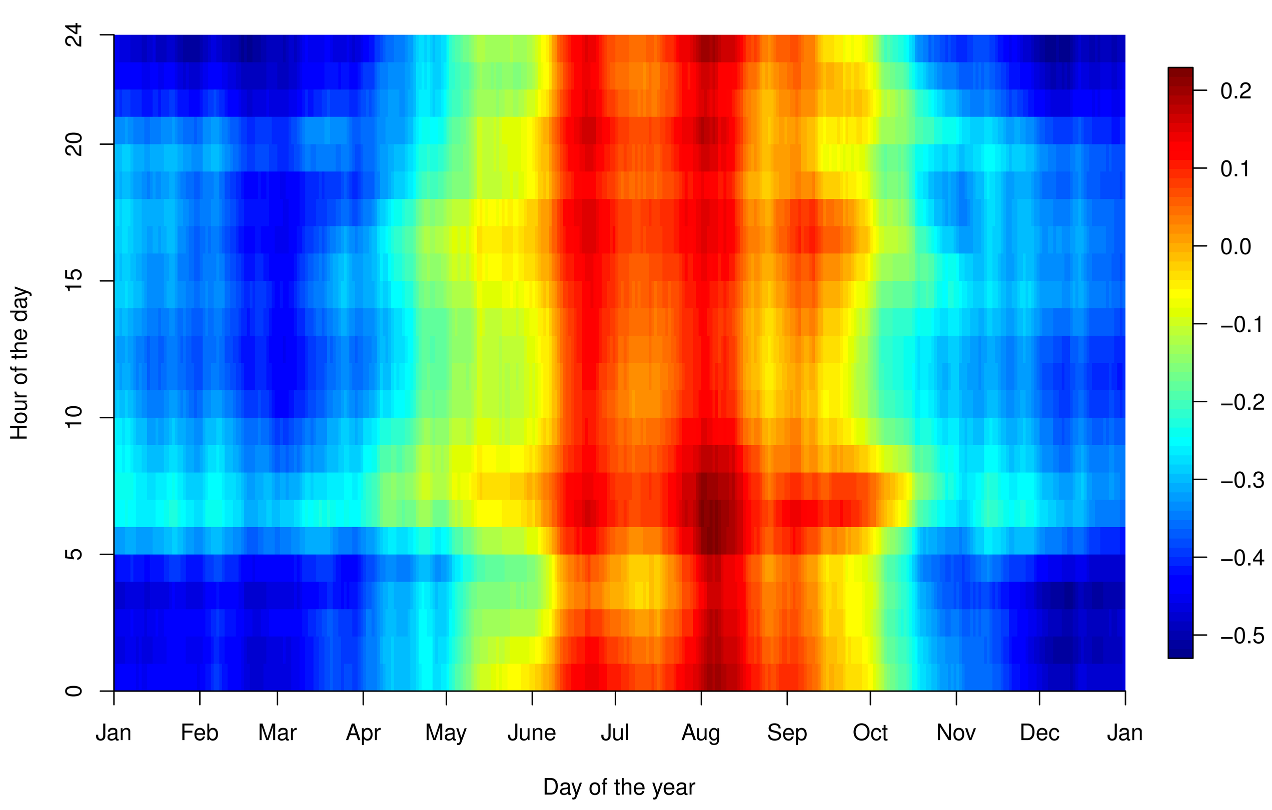


Figure: Conditional Kendall's tau between detrended electricity consumption and detrended temperature given the day of the year ($h_{*,1} = 12$ days) and the time of the day in hours ($h_{*,2} = 1$ hour).

Kendall's regression: a parametric model for the conditional Kendall's tau

- We want to estimate the true parameter $\beta^* \in \mathbb{R}^{p'}$ in our model
$$\forall \mathbf{z} \in \mathcal{Z},\ \Lambda\big(\tau_{1,2|\mathbf{Z}=\mathbf{z}}\big) = \psi(\mathbf{z})^T \beta^*,$$
in the high-dimensional case, i.e. p' is large.
- We observe an iid sample $\mathcal{D} = (X_{i,1}, X_{i,2}, \mathbf{Z}_i)$ but not a sample of $\tau_{1,2|\mathbf{Z}_i}$
$\Rightarrow$ we need estimated Kendall's tau $\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ for some values of $\mathbf{z}$.

Algorithm 1: Estimation algorithm for β^*

First sample: $(\mathbf{X}_i, \mathbf{Z}_i) \overset{\text{i.i.d.}}{\sim} (\mathbf{X}, \mathbf{Z})$, $i = 1, \dots, n$;
Second sample: $\mathbf{Z}'_i$, $i = 1, \dots, n'$ (design points) ;
for $i \leftarrow 1$ ***to*** n' **do**
  Compute the conditional Kendall's tau $\hat\tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}$ on the first sample ;
end
Solve the convex optimization program

$$\hat\beta := \arg\min_{\beta \in \mathbb{R}^{p'}} \left[\frac{1}{n'} \sum_{i=1}^{n'} \big(\Lambda(\hat\tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}) - \psi(\mathbf{Z}'_i)^T \beta\big)^2 + \lambda |\beta|_1 \right],$$

Summary of theoretical results

- Nonasymptotic bounds on $|\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}} - \tau_{1,2|\mathbf{Z}=\mathbf{z}}|$ for a given $\mathbf{z}$ (resp. uniform in $\mathbf{z}$ under stronger assumptions)
- Consistency, uniform consistency and asymptotic normality of $\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ as $n \to \infty$
- Nonasymptotic bounds on $|\hat\beta - \beta|$
- Consistency and asymptotic normality of $\hat\beta$ in two different regimes: $(n \to \infty, n' \text{ fixed})$ and $(n, n') \to (\infty, \infty)$.

A classification point-of-view

- Defining $W_{1,2} := 2 \times \mathbf{1}\{(X_{2,1} - X_{1,1})(X_{2,2} - X_{1,2}) > 0\} - 1$,
we have $\tau_{1,2|\mathbf{Z}=\mathbf{z}} = 2 \times p(\mathbf{z}) - 1$, where $p(\mathbf{z}) := \mathbb{P}(W = 1 | \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z})$.
- Prediction of concordance/discordance among pairs of observations $(\mathbf{X}_1, \mathbf{X}_2)$ given $\mathbf{Z} \simeq$ a classification task of such pairs.
- Evaluate conditional probabilities of observing concordant pairs of observations $\simeq$ evaluate conditional Kendall's tau: $\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}} = 2\hat p(\mathbf{z}) - 1$
- $\Rightarrow$ most classifiers can potentially be invoked, but applied here to a dataset $\tilde{\mathcal{D}}$ of (weighted) pairs of observations:
$$\tilde{\mathcal{D}} := (W_k, \tilde{\mathbf{Z}}_k, V_k)_{k \in \{1, \dots, n(n-1)/2\}} \in \big(\{-1, 1\} \times \mathbb{R}^p \times \mathbb{R}_+\big)^{n(n-1)/2}$$
  - Binary variable: $W_k = W_{i,j} := 2 \times \mathbf{1}\{(X_{i,1} - X_{j,1})(X_{i,2} - X_{j,2}) > 0\} - 1$
  - Average covariate: $\tilde{\mathbf{Z}}_k = (\mathbf{Z}_i + \mathbf{Z}_j)/2$
  - Weight of the pair: $V_k = K_h(\mathbf{Z}_i - \mathbf{Z}_j)$

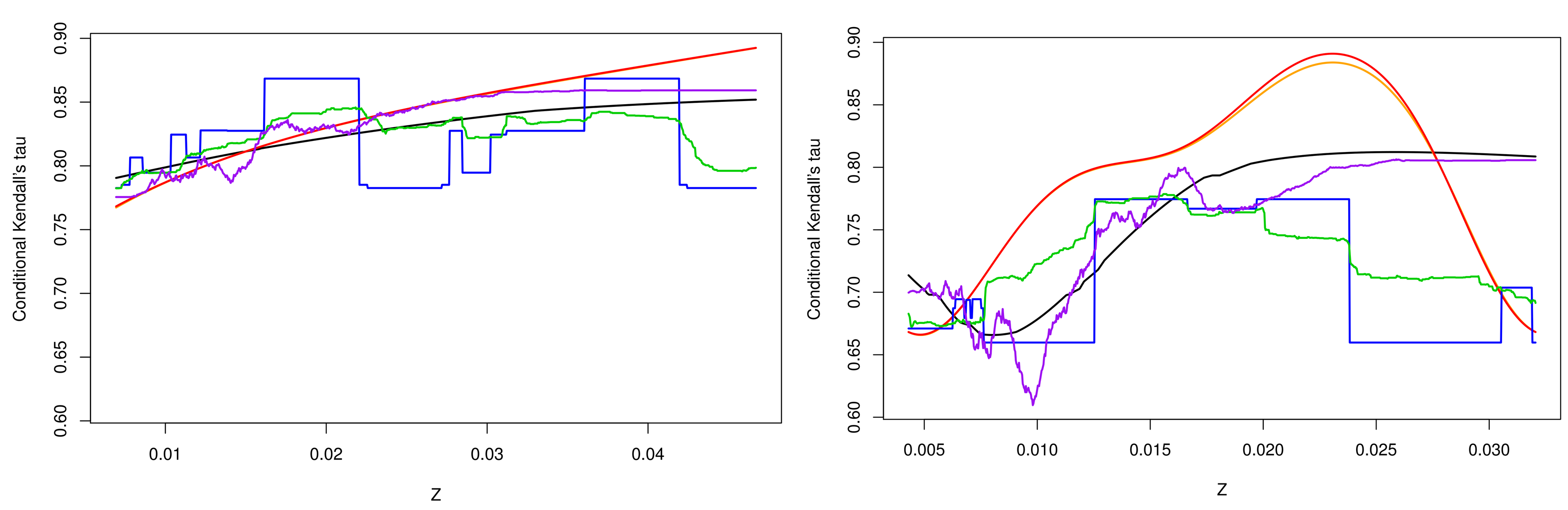


Figure: Estimated conditional Kendall's tau between French and German daily stock returns given the intraday volatility $\sigma := (High - Low)/Close$ during the European debt crisis (2009-2012) (left) and during the following period (2012-2019).

References

- Derumigny, A., & Fermanian, J. D., A classification point-of-view about conditional Kendall's tau. *Computational Statistics & Data Analysis*, 135, 70-94, 2019.
- Derumigny, A., & Fermanian, J. D., About Kendall's regression. *ArXiv preprint*, arXiv:1802.07613, 2018.
- Derumigny, A., & Fermanian, J. D., About kernel-based estimation of conditional Kendall's tau: finite-distance bounds and asymptotic behavior. *ArXiv preprint*, arXiv:1810.06234, 2018.