Conformal Prediction for Time Series

An application to forecasting French electricity Spot prices

Margaux Zaffran[1,2,3] Aymeric Dieuleveut[3] Olivier Féron[1,4] Yannig Goude[1] Julie Josse[2]

04/11/2021

[1]EDF R&D [2]INRIA [3]CMAP, Ecole Polytechnique [4]FiME

Forecasting French electricity Spot prices

Electricity Spot prices

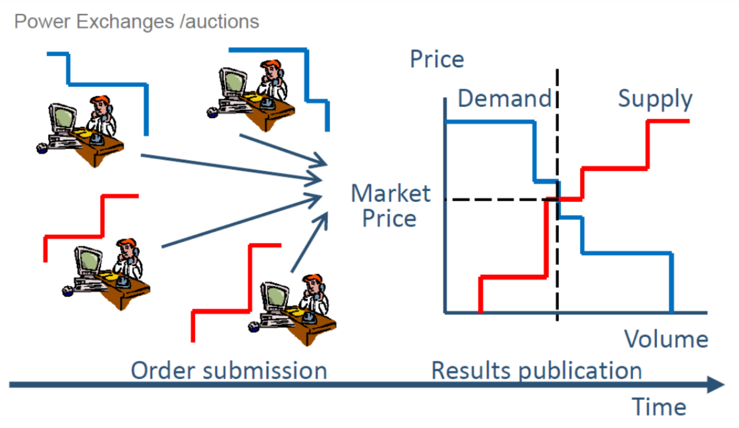


Figure 1: Drawing of spot auctions mechanism

French Electricity Spot prices data set: visualisation

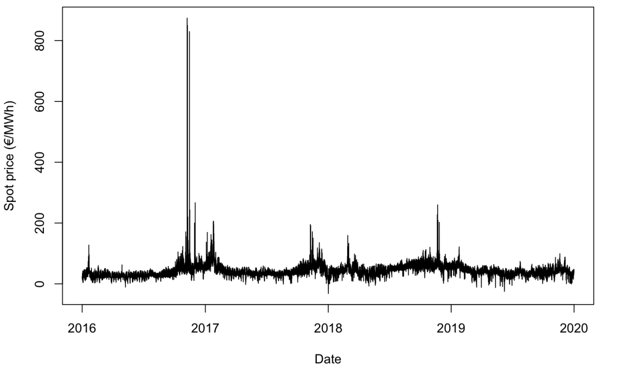


Figure 2: Representation of the French electricity spot price, from 2016 to 2019.

French Electricity Spot prices data set: extract

Date and time	Price	Price D-1	Price D-7	For. cons.	DOW
11/01/16 0PM	21.95	15.58	13.78	58800	Monday
11/01/16 1PM	20.04	19.05	13.44	57600	Monday
⋮	⋮	⋮	⋮	⋮	⋮
12/01/16 0PM	21.51	21.95	25.03	61600	Tuesday
12/01/16 1PM	19.81	20.04	24.42	59800	Tuesday
⋮	⋮	⋮	⋮	⋮	⋮
18/01/16 0PM	38.14	37.86	21.95	70400	Monday
18/01/16 1PM	35.66	34.60	20.04	69500	Monday
⋮	⋮	⋮	⋮	⋮	⋮

Table 1: Extract of the built data set, for French electricity spot price forecasting.

Forecasting French electricity Spot prices

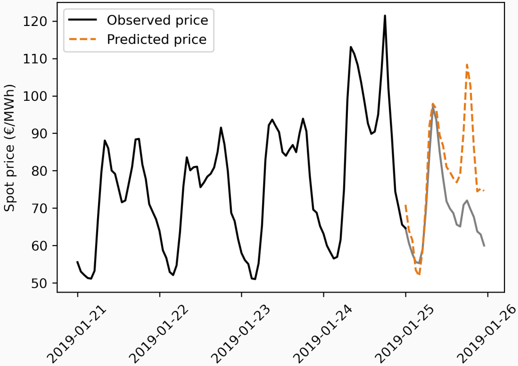


Figure 3: French electricity spot price and its prediction (orange) with random forest.

Forecasting French electricity Spot prices with confidence

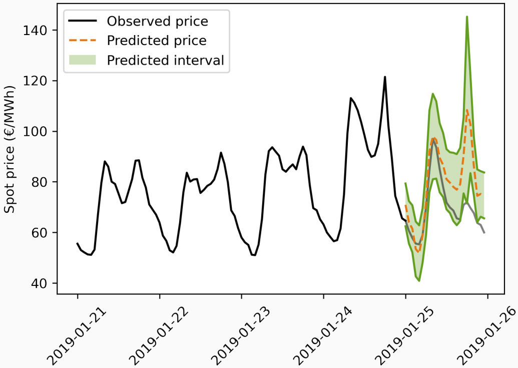


Figure 4: French electricity spot price, its prediction and its confidence (green) with Adaptive Conformal Prediction (Gibbs and Candès, 2021).

Forecasting French electricity Spot prices with confidence: results

- Target coverage: 90%
- Empirical coverage: 90.46%
- Average length: 22.91€/MWh

Introduction to conformal prediction

Conformal prediction: notations and framework

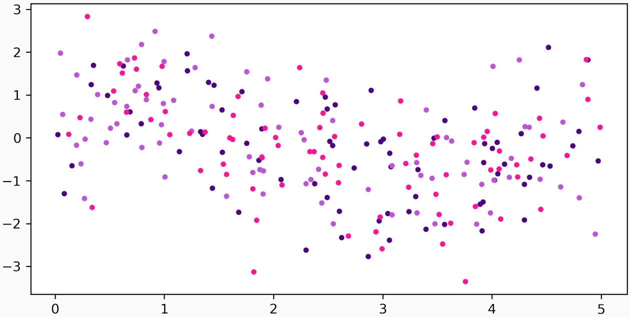
- $(x, y) \in \mathbb{R}^d \times \mathbb{R}$ realization of random variable (X, Y)
- n training samples $(x_i, y_i)_{i=1}^n$
- Goal: predict an unseen point y_{n+1} at x_{n+1} with confidence
- Miscoverage level $\alpha \in [0, 1]$

▶ Build a predictive interval $\mathcal{C}_\alpha$ such that:

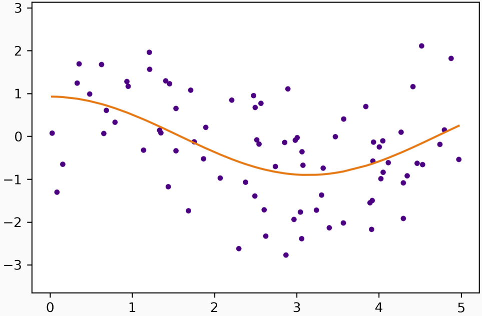
$$\mathbb{P}\left\{Y_{n+1} \in \mathcal{C}_\alpha\left(X_{n+1}\right)\right\} \geq 1 - \alpha, \tag{1}$$

and $\mathcal{C}_\alpha$ should be as small as possible, in order to be informative.

Split conformal prediction: toy example

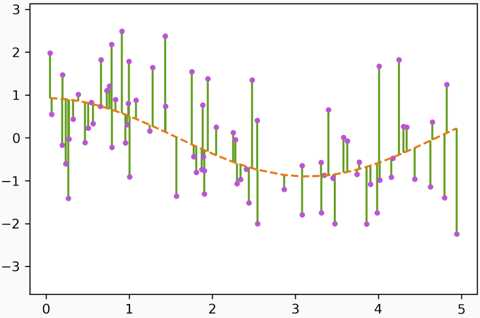


Split conformal prediction: training step



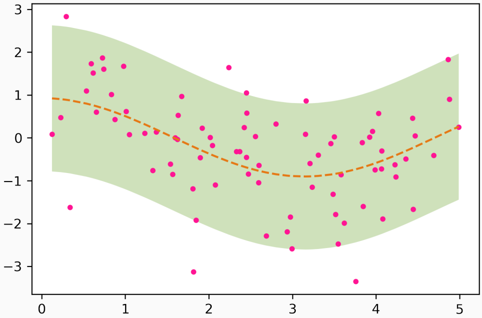
- Learn $\hat{\mu}$

Split conformal prediction: calibration step



- Predict with $\hat{\mu}$
- Get the residuals $\hat{\varepsilon}_i$
- Compute the $1-\alpha$ empirical quantile of the $|\hat{\varepsilon}_i|$, noted $Q_{1-\alpha}\left(|\hat{\varepsilon}_i|\right)$

Split conformal prediction: prediction step



- Predict with $\hat{\mu}$
- Build $\hat{\mathcal{C}}_\alpha(x)$:
  $\left[\hat{\mu}(x) \pm Q_{1-\alpha}\left(|\hat{\varepsilon}_i|\right)\right]$

Split conformal prediction: algorithm

Algorithm 1 Split Conformal Algorithm, with absolute value residuals scores

Input: Regression algorithm $\mathcal{A}$, significance level α, examples $(x_1, y_1), \ldots, (x_n, y_n)$.

Output: Prediction interval $\hat{\mathcal{C}}_\alpha(x)$ for any $x \in R^d$.

Randomly split $\{1, \ldots, n\}$ into two disjoint sets $\mathcal{I}_1$ and $\mathcal{I}_2$.
Fit a mean regression function: $\hat{\mu}(\cdot) \leftarrow \mathcal{A}\left(\{(x_i, y_i) : i \in \mathcal{I}_1\}\right)$
for $j \in \mathcal{I}_2$ **do**
    Set $\hat{\varepsilon}_j = |y_j - \hat{\mu}(x_j)|$, the conformity scores
end for
Compute $Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2)$, the $(1-\alpha)(1 + 1/|\mathcal{I}_2|)$-th empirical quantile of $\{\hat{\varepsilon}_j : j \in \mathcal{I}_2\}$.
Set $\hat{\mathcal{C}}_\alpha(x) = \left[\hat{\mu}(x) - Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2), \hat{\mu}(x) + Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2)\right]$ for any $x \in R^d$.

Split conformal prediction: algorithm

Algorithm 1 Split Conformal Algorithm, with absolute value residuals scores

Input: Regression algorithm $\mathcal{A}$, significance level α, examples $(x_1, y_1), \ldots, (x_n, y_n)$.

Output: Prediction interval $\hat{\mathcal{C}}_\alpha(x)$ for any $x \in R^d$.

1: Randomly split $\{1, \ldots, n\}$ into two disjoint sets $\mathcal{I}_1$ and $\mathcal{I}_2$.

2: Fit a mean regression function: $\hat{\mu}(\cdot) \leftarrow \mathcal{A}\left(\{(x_i, y_i) : i \in \mathcal{I}_1\}\right)$

3: **for** $j \in \mathcal{I}_2$ **do**

4: Set $\hat{\varepsilon}_j = |y_j - \hat{\mu}(x_j)|$, the conformity scores

5: **end for**

6: Compute $Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2)$, the $(1-\alpha)(1 + 1/|\mathcal{I}_2|)$-th empirical quantile of $\{\hat{\varepsilon}_j : j \in \mathcal{I}_2\}$.

7: Set $\hat{\mathcal{C}}_\alpha(x) = [\hat{\mu}(x) - Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2), \hat{\mu}(x) + Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2)]$ for any $x \in R^d$.

Split conformal prediction: algorithm

Algorithm 1 Split Conformal Algorithm, with absolute value residuals scores

Input: Regression algorithm $\mathcal{A}$, significance level α, examples $(x_1, y_1), \ldots, (x_n, y_n)$.

Output: Prediction interval $\hat{\mathcal{C}}_\alpha(x)$ for any $x \in R^d$.

1: Randomly split $\{1, \ldots, n\}$ into two disjoint sets $\mathcal{I}_1$ and $\mathcal{I}_2$.

2: Fit a mean regression function: $\hat{\mu}(\cdot) \leftarrow \mathcal{A}(\{(x_i, y_i) : i \in \mathcal{I}_1\})$

3: **for** $j \in \mathcal{I}_2$ **do**

4: Set $\hat{\varepsilon}_j = |y_j - \hat{\mu}(x_j)|$, the conformity scores

5: **end for**

6: Compute $Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2)$, the $(1-\alpha)(1 + 1/|\mathcal{I}_2|)$-th empirical quantile of $\{\hat{\varepsilon}_j : j \in \mathcal{I}_2\}$.

7: Set $\hat{\mathcal{C}}_\alpha(x) = [\hat{\mu}(x) - Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2), \hat{\mu}(x) + Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2)]$ for any $x \in R^d$.

Split conformal prediction: algorithm

Algorithm 1 Split Conformal Algorithm, with absolute value residuals scores

Input: Regression algorithm $\mathcal{A}$, significance level α, examples $(x_1, y_1), \ldots, (x_n, y_n)$.

Output: Prediction interval $\hat{\mathcal{C}}_\alpha(x)$ for any $x \in R^d$.

Randomly split $\{1, \ldots, n\}$ into two disjoint sets $\mathcal{I}_1$ and $\mathcal{I}_2$.
Fit a mean regression function: $\hat{\mu}(\cdot) \leftarrow \mathcal{A}\left(\{(x_i, y_i) : i \in \mathcal{I}_1\}\right)$
for $j \in \mathcal{I}_2$ **do**
    Set $\hat{\varepsilon}_j = |y_j - \hat{\mu}(x_j)|$, the conformity scores
end for
Compute $Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2)$, the $(1-\alpha)(1 + 1/|\mathcal{I}_2|)$-th empirical quantile of $\{\hat{\varepsilon}_j : j \in \mathcal{I}_2\}$.
Set $\hat{\mathcal{C}}_\alpha(x) = \left[\hat{\mu}(x) - Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2), \hat{\mu}(x) + Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2)\right]$ for any $x \in R^d$.

- Conformity scores can be changed (Lei et al. (2018); Romano et al. (2019), among others)

Split conformal prediction: algorithm

Algorithm 1 Split Conformal Algorithm, with absolute value residuals scores

Input: Regression algorithm $\mathcal{A}$, significance level α, examples $(x_1, y_1), \ldots, (x_n, y_n)$.

Output: Prediction interval $\hat{\mathcal{C}}_\alpha(x)$ for any $x \in R^d$.

1: Randomly split $\{1, \ldots, n\}$ into two disjoint sets $\mathcal{I}_1$ and $\mathcal{I}_2$.
2: Fit a mean regression function: $\hat{\mu}(\cdot) \leftarrow \mathcal{A}(\{(x_i, y_i) : i \in \mathcal{I}_1\})$
3: **for** $j \in \mathcal{I}_2$ **do**
4: Set $\hat{\varepsilon}_j = |y_j - \hat{\mu}(x_j)|$, the conformity scores
5: **end for**
6: Compute $Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2)$, the $(1-\alpha)(1 + 1/|\mathcal{I}_2|)$-th empirical quantile of $\{\hat{\varepsilon}_j : j \in \mathcal{I}_2\}$.
7: Set $\hat{\mathcal{C}}_\alpha(x) = [\hat{\mu}(x) - Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2), \hat{\mu}(x) + Q_{1-\alpha}(\hat{\varepsilon}, \mathcal{I}_2)]$ for any $x \in R^d$.

- Conformity scores can be changed (Lei et al. (2018); Romano et al. (2019), among others)
- The correction is needed for finite-sample guarantee

Conformal prediction: theoretical guarantees

This procedure enjoys finite sample guarantee proposed and proved in Lei et al. (2018).

Theorem
Suppose $(X_i, Y_i)_{i=1}^{n+1}$ are exchangeable, and we apply split conformal prediction on $(X_i, Y_i)_{i=1}^{n}$ to predict an interval on X_{n+1}, $\hat{\mathcal{C}}_\alpha(X_{n+1})$. Then we have:

$$\mathbb{P}\left\{Y_{n+1} \in \hat{\mathcal{C}}_\alpha(X_{n+1})\right\} \geq 1 - \alpha.$$

If, in addition, the scores $\hat{\varepsilon}_j$ have a continuous joint distribution, we also have an upper bound:

$$\mathbb{P}\left\{Y_{n+1} \in \hat{\mathcal{C}}_\alpha(X_{n+1})\right\} \leq 1 - \alpha + \frac{2}{n+2}.$$

Conformal prediction: summary

Split conformal prediction is simple to compute and works:

- any regression algorithm (neural nets, random forest...);
- distribution-free as long as the data is exchangeable;

- finite sample.

Conformal prediction: summary

Split conformal prediction is simple to compute and works:

- any regression algorithm (neural nets, random forest...);
- distribution-free as long as the data is exchangeable;
  - $\hookrightarrow$ the scores need to be exchangeable (but then it would not work with any regression algorithm)
- finite sample.

Conformal prediction and time series, what's the issue?

Time series are not exchangeable

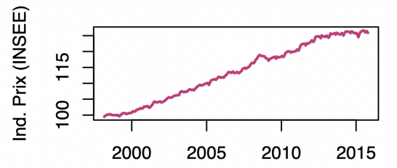


Figure 5: Trend

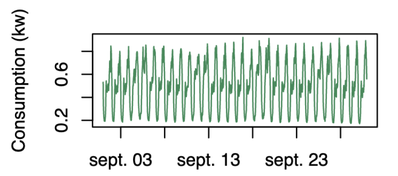


Figure 6: Seasonality

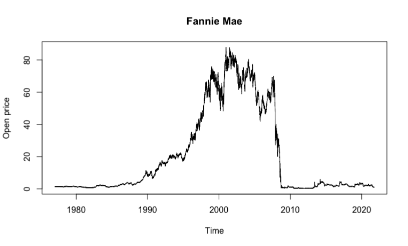


Figure 7: Shift

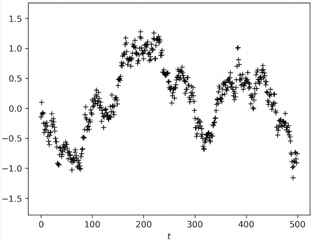


Figure 8: Time dependence

Time dependent noise

Assume the following model:

$$Y_t = f_t(X_t) + \varepsilon_t, \ \text{ for } t \in \mathbb{N}^*,$$

for some function f_t, and some noise ε_t.

Time dependent noise

Assume the following model:

$$Y_t = f_t(X_t) + \varepsilon_t, \text{ for } t \in \mathbb{N}^*,$$

for some function f_t, and some noise ε_t.

If the noise ε_t is time dependent, the residuals will be dependent no matter what is the fitted regression function.

Time dependent noise

Assume the following model:

$$Y_t = f_t(X_t) + \varepsilon_t, \text{ for } t \in \mathbb{N}^*,$$

for some function f_t, and some noise ε_t.

If the noise ε_t is time dependent, the residuals will be dependent no matter what is the fitted regression function.

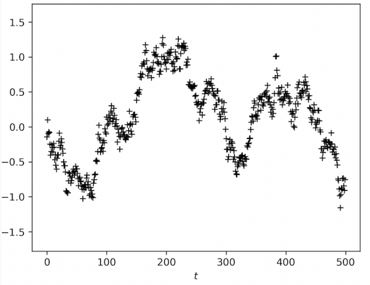


Figure 9: Auto-Regressive noise

Non-exchangeable even if the noise is exchangeable

Even if the noise is exchangeable, we can produce dependent residuals (examples available).

Available methods

Framework and notations

- Data: T_0 observations $(x_1, y_1), \ldots, (x_{T_0}, y_{T_0})$ in $\mathbb{R}^d \times \mathbb{R}$

Framework and notations

- Data: T_0 observations $(x_1, y_1), \dots, (x_{T_0}, y_{T_0})$ in $\mathbb{R}^d \times \mathbb{R}$
- Aim: predict the response values as well as predictive intervals for T_1 subsequent observations $x_{T_0+1}, \dots, x_{T_0+T_1}$

Framework and notations

- Data: T_0 observations $(x_1, y_1), \ldots, (x_{T_0}, y_{T_0})$ in $\mathbb{R}^d \times \mathbb{R}$
- Aim: predict the response values as well as predictive intervals for T_1 subsequent observations $x_{T_0+1}, \ldots, x_{T_0+T_1}$

$\hookrightarrow$ Build the smallest interval $\mathcal{C}_\alpha^t$ such that:

$$\mathbb{P}\left\{Y_t \in \mathcal{C}_\alpha^t\left(X_t\right)\right\} \geq 1 - \alpha, \ \text{ for } t \in]T_0; T_0 + T_1].$$

How to adapt to time series?

Usual ideas from the time series literature:

- Consider an online procedure (for each new data, re-train and re-calibrate)

How to adapt to time series?

Usual ideas from the time series literature:

- Consider an online procedure (for each new data, re-train and re-calibrate)
  - $\hookrightarrow$ update to recent observations (trend impact, period of the seasonality, dependence...)

How to adapt to time series?

Usual ideas from the time series literature:

- Consider an online procedure (for each new data, re-train and re-calibrate)
  - $\hookrightarrow$ update to recent observations (trend impact, period of the seasonality, dependence...)
- Use a sequential split

How to adapt to time series?

Usual ideas from the time series literature:

- Consider an online procedure (for each new data, re-train and re-calibrate)
  - $\hookrightarrow$ update to recent observations (trend impact, period of the seasonality, dependence...)
- Use a sequential split
  - $\hookrightarrow$ use only the past so as to correctly estimate the variance of the residuals (using the future leads to optimistic residuals and underestimation of their variance)

Available methods

- Online sequential split conformal prediction (Wisniewski et al. (2020); Kath and Ziel (2021); and our study);
- Ensemble Prediction Interval (Xu and Xie, 2021);
- Adaptive Conformal Prediction (Gibbs and Candès, 2021).

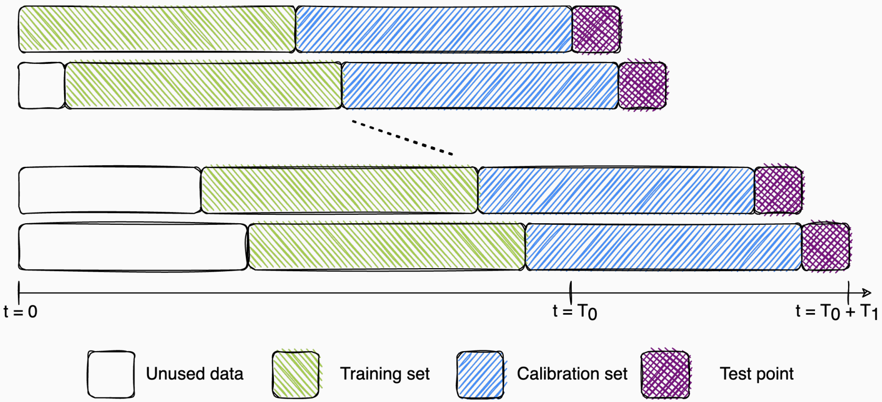


Figure 10: Diagram describing the online sequential split conformal prediction.

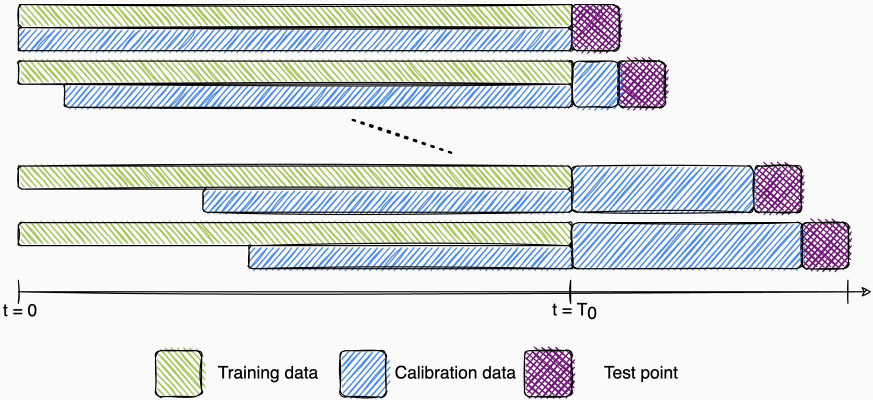
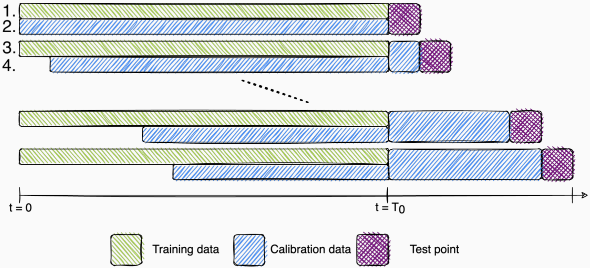


Figure 11: Diagram describing the EnbPI algorithm.

EnbPI, Xu and Xie (2021)



1. Train B bootstrap predictors ($\rightarrow$ learn better $\hat{\mu}$ hoping to have less dependent residuals);
2. Obtain out-of-bootstrap residuals by aggregating the corresponding predictors;
3. Do not re-train the B bootstrap predictors;
4. Obtain new residual by aggregating all the predictors. Forget the first residuals.

Adaptive Conformal Prediction (ACP), Gibbs and Candès (2021)

Refitting the model may be insufficient $\Rightarrow$ adapt the quantile level used on the calibration's scores.

Adaptive Conformal Prediction (ACP), Gibbs and Candès (2021)

Refitting the model may be insufficient $\Rightarrow$ adapt the quantile level used on the calibration's scores. **(Distribution shift)**

Adaptive Conformal Prediction (ACP), Gibbs and Candès (2021)

Refitting the model may be insufficient $\Rightarrow$ adapt the quantile level used on the calibration's scores. **(Distribution shift)**

The proposed update scheme is the following:

$$\alpha_{t+1} := \alpha_t + \gamma \left(\alpha - \mathrm{err}_t \right) \tag{2}$$

with:

$$\mathrm{err}_t := \begin{cases} 1, & \text{if } y_t \notin \hat{\mathcal{C}}_{\alpha_t}\left(x_t\right), \\ 0, & \text{otherwise} \end{cases},$$

and $\alpha_1 = \alpha$.

Adaptive Conformal Prediction (ACP), Gibbs and Candès (2021)

Refitting the model may be insufficient $\Rightarrow$ adapt the quantile level used on the calibration's scores. **(Distribution shift)**

The proposed update scheme is the following:

$$\alpha_{t+1} := \alpha_t + \gamma \left(\alpha - \mathrm{err}_t \right) \tag{2}$$

with:

$$\mathrm{err}_t := \begin{cases} 1, & \text{if } y_t \notin \hat{\mathcal{C}}_{\alpha_t}\left(x_t\right), \\ 0, & \text{otherwise} , \end{cases}$$

and $\alpha_1 = \alpha$.

Intuition: if we did make an error, the interval was too small so we want to increase its length by taking a higher quantile (a smaller α_t). Reversely if we included the point.

Visualisation of the procedure

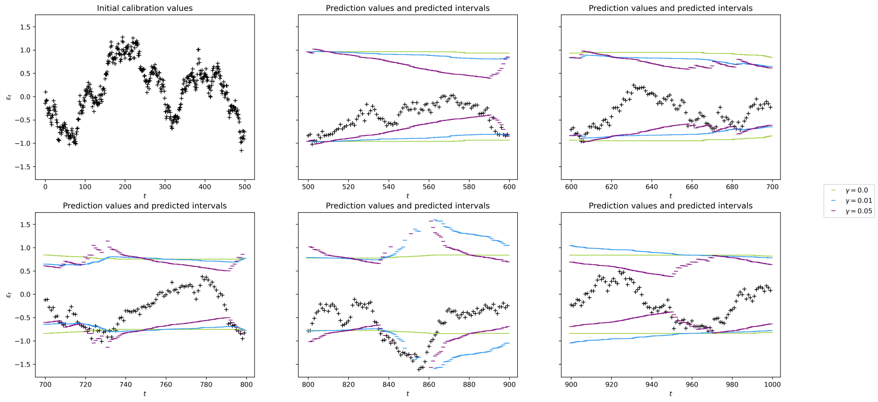


Figure 11: Visualisation of ACP with different values of γ

Visualisation of the procedure

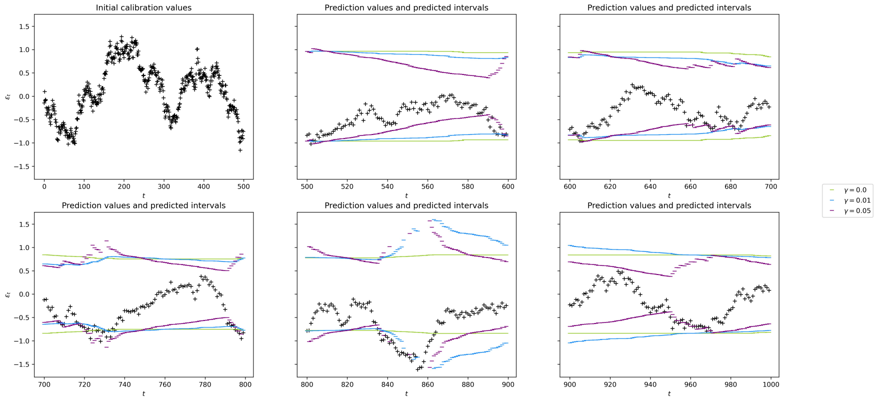


Figure 11: Visualisation of ACP with different values of γ

ACP originally splitted randomly. We use ACP with a **sequential split**.

Summary of the methods

	Scores distribution		
Methods	Exchangeable	Strongly mixing	No assumption
Conformal Prediction	✓	✓	✗
EnbPI	✗	✗	✗
Adaptive Conformal Prediction	✓	✓	✓

Table 2: Methods validity with respect to the conformity scores distribution. Green marks indicates finite-sample validity, orange long-term validity and red no theoretical validity.

Comparison on simulated data

Data generation

$$Y_t = 10 \sin\left(\pi X_{t,1} X_{t,2}\right) + 20\left(X_{t,3} - 0.5\right)^2 + 10 X_{t,4} + 5 X_{t,5} + \varepsilon_t$$

where the X_t are multivariate uniformly distributed on $[0, 1]$ and ε_t are generated from an ARMA(1,1) process.

Data generation

$$Y_t = 10\sin\left(\pi X_{t,1} X_{t,2}\right) + 20\left(X_{t,3} - 0.5\right)^2 + 10X_{t,4} + 5X_{t,5} + \varepsilon_t$$

where the X_t are multivariate uniformly distributed on $[0, 1]$ and ε_t are generated from an ARMA(1,1) process.

$\Rightarrow$ dependence structure in the noise in order to:

- control the strength of the scores dependence,
- evaluate the impact of this temporal dependence structure of the results.

Definition (ARMA(1,1) process)
We say that ε_t is an ARMA(1,1) process if for any t:

$$\varepsilon_{t+1} = \varphi \varepsilon_t + \xi_{t+1} + \theta \xi_t,$$

with:

- $\theta + \varphi \neq 0$, $|\varphi| < 1$ and $|\theta| < 1$;
- ξ_t is a white noise of variance σ^2, called the innovation.

Comments on the ARMA(1,1) process

- The higher φ and θ, the stronger the dependence.

Comments on the ARMA(1,1) process

- The higher φ and θ, the stronger the dependence.
- The asymptotic variance of this process is:

$$\mathsf{Var}(\varepsilon_t) = \sigma^2 \frac{1 - 2\varphi\theta + \theta^2}{1 - \varphi^2}. \tag{3}$$

Comments on the ARMA(1,1) process

- The higher φ and θ, the stronger the dependence.
- The asymptotic variance of this process is:

$$\text{Var}(\varepsilon_t) = \sigma^2 \frac{1 - 2\varphi\theta + \theta^2}{1 - \varphi^2}. \tag{3}$$

- If $\theta = 0$ we only have the auto-regressive part, we talk about an AR(1).

Comments on the ARMA(1,1) process

- The higher φ and θ, the stronger the dependence.
- The asymptotic variance of this process is:

$$\mathrm{Var}(\varepsilon_t) = \sigma^2 \frac{1 - 2\varphi\theta + \theta^2}{1 - \varphi^2}. \tag{3}$$

- If $\theta = 0$ we only have the auto-regressive part, we talk about an AR(1).
- If $\varphi = 0$ we only have the moving-average part, we talk about an MA(1).

Simulation settings

- φ and θ range in $[0.1, 0.8, 0.9, 0.95, 0.99]$.

Simulation settings

- φ and θ range in $[0.1, 0.8, 0.9, 0.95, 0.99]$.
- We fix σ so as to keep the variance $\mathrm{Var}(\varepsilon_t)$ in equation (3) constant to 1 or 10.

Simulation settings

- φ and θ range in $[0.1, 0.8, 0.9, 0.95, 0.99]$.
- We fix σ so as to keep the variance $\mathrm{Var}(\varepsilon_t)$ in equation (3) constant to 1 or 10.
- We use random forest as regressor.

Simulation settings

- φ and θ range in $[0.1, 0.8, 0.9, 0.95, 0.99]$.
- We fix σ so as to keep the variance $\text{Var}(\varepsilon_t)$ in equation (3) constant to 1 or 10.
- We use random forest as regressor.

For each setting:

- 300 points, the last 100 kept for prediction and evaluation,

Simulation settings

- φ and θ range in $[0.1, 0.8, 0.9, 0.95, 0.99]$.
- We fix σ so as to keep the variance $\mathrm{Var}(\varepsilon_t)$ in equation (3) constant to 1 or 10.
- We use random forest as regressor.

For each setting:

- 300 points, the last 100 kept for prediction and evaluation,
- 500 repetitions,

Simulation settings

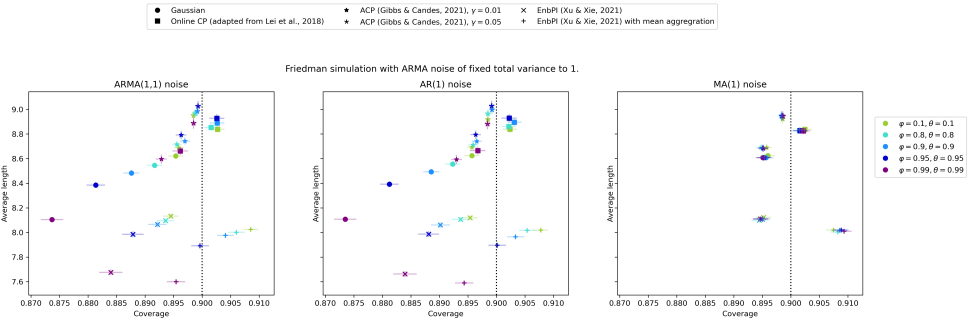
- φ and θ range in $[0.1, 0.8, 0.9, 0.95, 0.99]$.
- We fix σ so as to keep the variance $\text{Var}(\varepsilon_t)$ in equation (3) constant to 1 or 10.
- We use random forest as regressor.

For each setting:

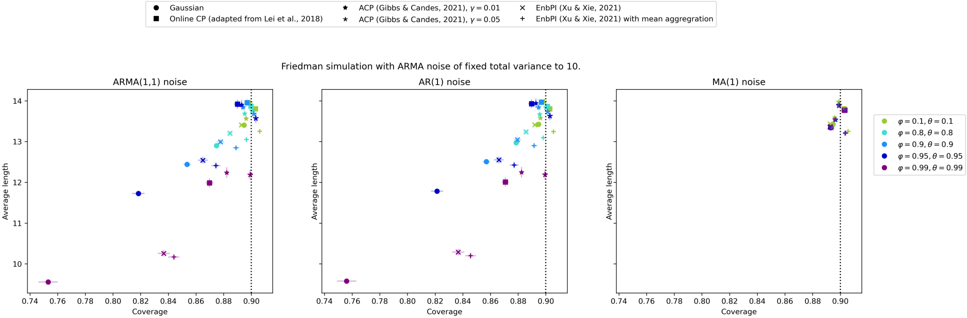
- 300 points, the last 100 kept for prediction and evaluation,
- 500 repetitions,

$\Rightarrow$ in total, $100 \times 500 = 50000$ predictions are evaluated.

Results: impact of the temporal dependence, variance 1



Results: impact of the temporal dependence, variance 10



Methods ranking

1. ACP: achieves valid coverage with $\gamma = 0.05$. Nevertheless, the choice of γ is important.

Methods ranking

1. ACP: achieves valid coverage with $\gamma = 0.05$. Nevertheless, the choice of γ is important.
2. Online CP: achieves valid coverage for values of φ and θ smaller than 0.99.

Methods ranking

1. ACP: achieves valid coverage with $\gamma = 0.05$. Nevertheless, the choice of γ is important.
2. Online CP: achieves valid coverage for values of φ and θ smaller than 0.99.
3. EnbPI: for small variance, really competitive (small lengths). But for strong dependence and/or high variance, fails to attain coverage.

Methods ranking

1. ACP: achieves valid coverage with $\gamma = 0.05$. Nevertheless, the choice of γ is important.
2. Online CP: achieves valid coverage for values of φ and θ smaller than 0.99.
3. EnbPI: for small variance, really competitive (small lengths). But for strong dependence and/or high variance, fails to attain coverage.
4. Gaussian: the behavior is similar to EnbPI, with larger interval and more under-coverage.

Price prediction with confidence in 2019

Settings

- Forecast for the year 2019.

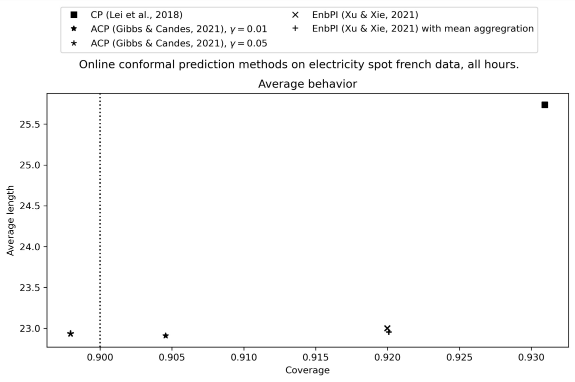
Settings

- Forecast for the year 2019.
- Random forest regressor.

Settings

- Forecast for the year 2019.
- Random forest regressor.
- One model per hour, we concatenate the predictions afterwards.

Performance on predicted French electricity Spot price for the year 2019



Performance on predicted French electricity Spot price: visualisation of a day

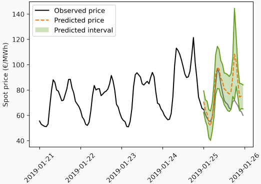


Figure 12: Online seq. split CP

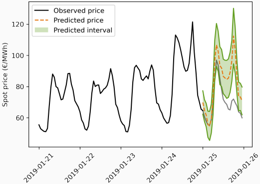


Figure 13: EnbPI

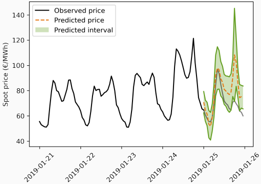


Figure 14: ACP with $\gamma = 0.01$

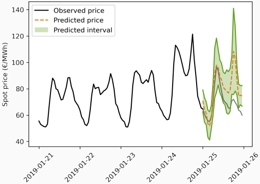


Figure 15: ACP with $\gamma = 0.05$

Concluding remarks

Conclusion

- Online sequential split conformal prediction achieves correct performances, but can lead to really large intervals
- ACP is sensitive to γ choice (ongoing work, empirical and theoretical)
- ACP obtains valid coverage in the time dependent settings, whilst designed initially for shifts
- EnbPI is highly competitive in some regimes, but its performanced depends a lot on the regime

References i

Gibbs, I. and Candès, E. (2021). Adaptive Conformal Inference Under Distribution Shift. *arXiv:2106.00170 [stat]*. arXiv: 2106.00170.

Kath, C. and Ziel, F. (2021). Conformal prediction interval estimation and applications to day-ahead and intraday power markets. *International Journal of Forecasting*, 37(2):777–799.

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. (2018). Distribution-Free Predictive Inference for Regression. *Journal of the American Statistical Association*, 113(523):1094–1111. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2017.1307116.

Romano, Y., Patterson, E., and Candes, E. (2019). Conformalized Quantile Regression. *Advances in Neural Information Processing Systems*, 32.

Wisniewski, W., Lindsay, D., and Lindsay, S. (2020). Application of conformal prediction interval estimations to market makers' net positions. In Gammerman, A., Vovk, V., Luo, Z., Smirnov, E., and Cherubin, G., editors, *Proceedings of the Ninth Symposium on Conformal and Probabilistic Prediction and Applications*, volume 128 of *Proceedings of Machine Learning Research*, pages 285–301. PMLR.

References iii

Xu, C. and Xie, Y. (2021). Conformal prediction interval for dynamic time-series. In Meila, M. and Zhang, T., editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 11559–11569. PMLR.

References i

Gibbs, I. and Candès, E. (2021). Adaptive Conformal Inference Under Distribution Shift. *arXiv:2106.00170 [stat]*. arXiv: 2106.00170.

Kath, C. and Ziel, F. (2021). Conformal prediction interval estimation and applications to day-ahead and intraday power markets. *International Journal of Forecasting*, 37(2):777–799.

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. (2018). Distribution-Free Predictive Inference for Regression. *Journal of the American Statistical Association*, 113(523):1094–1111. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2017.1307116.

References ii

Romano, Y., Patterson, E., and Candes, E. (2019). Conformalized Quantile Regression. *Advances in Neural Information Processing Systems*, 32.

Wisniewski, W., Lindsay, D., and Lindsay, S. (2020). Application of conformal prediction interval estimations to market makers' net positions. In Gammerman, A., Vovk, V., Luo, Z., Smirnov, E., and Cherubin, G., editors, *Proceedings of the Ninth Symposium on Conformal and Probabilistic Prediction and Applications*, volume 128 of *Proceedings of Machine Learning Research*, pages 285–301. PMLR.

Xu, C. and Xie, Y. (2021). Conformal prediction interval for dynamic time-series. In Meila, M. and Zhang, T., editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 11559–11569. PMLR.

Definition (Exchangeability)
$(X_t, Y_t)_{t=1}^{T}$ are exchangeable if for any permutation σ of $[1, T]$ we have:

$$\mathcal{L}\left((X_1, Y_1), \ldots, (X_T, Y_T)\right) = \mathcal{L}\left(\left(X_{\sigma(1)}, Y_{\sigma(1)}\right), \ldots, \left(X_{\sigma(T)}, Y_{\sigma(T)}\right)\right),$$

where $\mathcal{L}$ designates the joint distribution.

Endogenous and not perfectly estimated

Assume $X_t = Y_{t-1} \in \mathbb{R}$ and that

$$Y_t = aY_{t-1} + \varepsilon_t,$$

where ε_t is a white noise.

Endogenous and not perfectly estimated

Assume $X_t = Y_{t-1} \in \mathbb{R}$ and that

$$Y_t = aY_{t-1} + \varepsilon_t,$$

where ε_t is a white noise.

Assume that the fitted model is $\hat{f}_t(x) = \hat{a}x$, with $\hat{a} \neq a$.

Endogenous and not perfectly estimated

Assume $X_t = Y_{t-1} \in \mathbb{R}$ and that

$$Y_t = aY_{t-1} + \varepsilon_t,$$

where ε_t is a white noise.

Assume that the fitted model is $\hat{f}_t(x) = \hat{a}x$, with $\hat{a} \neq a$.

Then, for any t, we have that:

$$\hat{\varepsilon}_t = Y_t - \hat{Y}_t = (a - \hat{a})\, Y_{t-1} + \varepsilon_t$$

$$\hat{\varepsilon}_t = a\hat{\varepsilon}_{t-1} + \xi_t$$

with $\xi_t = \varepsilon_t - \hat{a}\varepsilon_{t-1}$.

Endogenous and not perfectly estimated

Assume $X_t = Y_{t-1} \in \mathbb{R}$ and that

$$Y_t = aY_{t-1} + \varepsilon_t,$$

where ε_t is a white noise.

Assume that the fitted model is $\hat{f}_t(x) = \hat{a}x$, with $\hat{a} \neq a$.

Then, for any t, we have that:

$$\hat{\varepsilon}_t = Y_t - \hat{Y}_t = (a - \hat{a})\, Y_{t-1} + \varepsilon_t$$
$$\hat{\varepsilon}_t = a\hat{\varepsilon}_{t-1} + \xi_t$$

with $\xi_t = \varepsilon_t - \hat{a}\varepsilon_{t-1}$.

$\hat{\varepsilon}_t$ is an ARMA process of parameters $\varphi = a$ and $\theta = -\hat{a}$.

Endogenous and not perfectly estimated

Assume $X_t = Y_{t-1} \in \mathbb{R}$ and that

$$Y_t = aY_{t-1} + \varepsilon_t,$$

where ε_t is a white noise.

Assume that the fitted model is $\hat{f}_t(x) = \hat{a}x$, with $\hat{a} \neq a$.

Then, for any t, we have that:

$$\hat{\varepsilon}_t = Y_t - \hat{Y}_t = (a - \hat{a})\, Y_{t-1} + \varepsilon_t$$
$$\hat{\varepsilon}_t = a\hat{\varepsilon}_{t-1} + \xi_t$$

with $\xi_t = \varepsilon_t - \hat{a}\varepsilon_{t-1}$.

$\hat{\varepsilon}_t$ is an ARMA process of parameters $\varphi = a$ and $\theta = -\hat{a}$.

Thus, we have generated dependent residuals (ARMA residuals) even if the underlying model only had white noise.

Exogenous and misspecified

Assume $X_t \in \mathbb{R}^2$ and that:

$$Y_t = aX_{1,t} + bX_{2,t} + \varepsilon_t,$$

with $\varepsilon_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $X_{2,t+1} = \varphi X_{2,t} + \xi_t$, $\xi_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $X_{1,t}$ can be any random variable.

Exogenous and misspecified

Assume $X_t \in \mathbb{R}^2$ and that:

$$Y_t = aX_{1,t} + bX_{2,t} + \varepsilon_t,$$

with $\varepsilon_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $X_{2,t+1} = \varphi X_{2,t} + \xi_t,\ \ \xi_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $X_{1,t}$ can be any random variable.

Assume that we misspecify the model such that the fitted model is $\hat{f}_t(x) = ax_1$.

Exogenous and misspecified

Assume $X_t \in \mathbb{R}^2$ and that:

$$Y_t = aX_{1,t} + bX_{2,t} + \varepsilon_t,$$

with $\varepsilon_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $X_{2,t+1} = \varphi X_{2,t} + \xi_t, \;\; \xi_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $X_{1,t}$ can be any random variable.

Assume that we misspecify the model such that the fitted model is $\hat{f}_t(x) = ax_1$.

Then, for any t, we have that

$$\hat{\varepsilon}_t = Y_t - \hat{Y}_t = bX_{2,t} + \varepsilon_t.$$

Exogenous and misspecified

Assume $X_t \in \mathbb{R}^2$ and that:

$$Y_t = aX_{1,t} + bX_{2,t} + \varepsilon_t,$$

with $\varepsilon_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $X_{2,t+1} = \varphi X_{2,t} + \xi_t$, $\xi_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $X_{1,t}$ can be any random variable.

Assume that we misspecify the model such that the fitted model is $\hat{f}_t(x) = ax_1$.

Then, for any t, we have that

$$\hat{\varepsilon}_t = Y_t - \hat{Y}_t = bX_{2,t} + \varepsilon_t.$$

Thus, we have generated dependent residuals (auto-regressive residuals) even if the underlying model only had i.i.d. Gaussian noise.

ACP sensitivity to γ