Imperial College London

Data Markets for Energy Forecasting

Pierre Pinson

Imperial College London, Dyson School of Design Engineering
mail: `p.pinson@imperial.ac.uk` - **webpage:** `www.pierrepinson.com`

IEEE SmartGridComm 2022
27 October 2022

(Ackn. to collaborators: C. Goncalves, R. Bessa, L. Han, J Kazempour - Partly funded by EU project Smart4RES)

Outline

1. The context and a motivating example
2. Collaborative and market-based analytics platforms
3. Regression market example
4. Concluding thoughts and discussion

1. **The context and a motivating example processes**

Today, everything has to be smart!

Smart Energy

Smart Cities

Industry 4.0

Smart Transport

Smart Agriculture

etc.

Today, everything has to be smart!

Imperial College London

Smart Energy

Smart Cities

Industry 4.0

Smart Transport

Smart Agriculture

"Smart"
=
Data
+
Analytics

1995/2005 – ...

- Data collection is increasing at an astounding rate (order of billions of GB per day!)
- This motivated research efforts towards **big data** analytics

- Data collection and storage is decentralized
- This led to a focus on **edge**, **cloud** and **fog computing**

Challenges and status quo within analytics

1995/2005 – ...

- Data collection is increasing at an astounding rate (order of billions of GB per day!)
- This motivated research efforts towards **big data** analytics

- Data collection and storage is decentralized
- This led to a focus on **edge**, **cloud** and **fog computing**

Remaining gap:

Data ownership is also **distributed**, with agents having **heterogeneous preferences** (privacy, competition, willingness to share, etc.)

Challenges and status quo within analytics

1995/2005 – ...

- Data collection is increasing at an astounding rate (order of billions of GB per day!)
- This motivated research efforts towards **big data** analytics

- Data collection and storage is decentralized
- This led to a focus on **edge**, **cloud** and **fog computing**

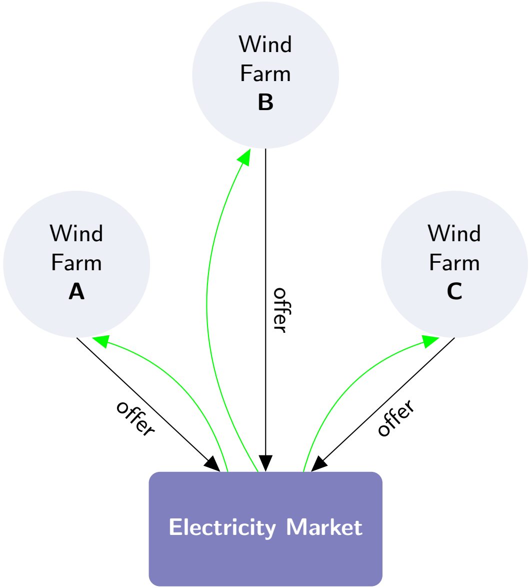
Proposal solution:

Collaborative and market-based analytics!
(or, how can we design systems and mechanims that allow to get the full value from distributed data)

A motivating real-world example

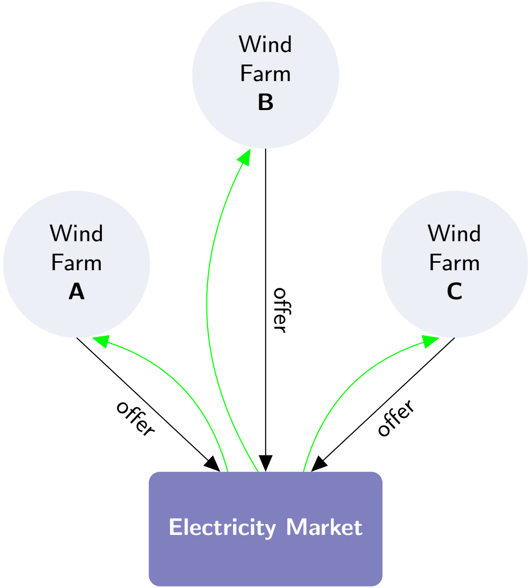
Context:

- Wind farms offer in electricity markets based on their individual (probabilistic) forecasts and private information
- Their revenue is affected by their (lack of) forecast accuracy



A motivating real-world example

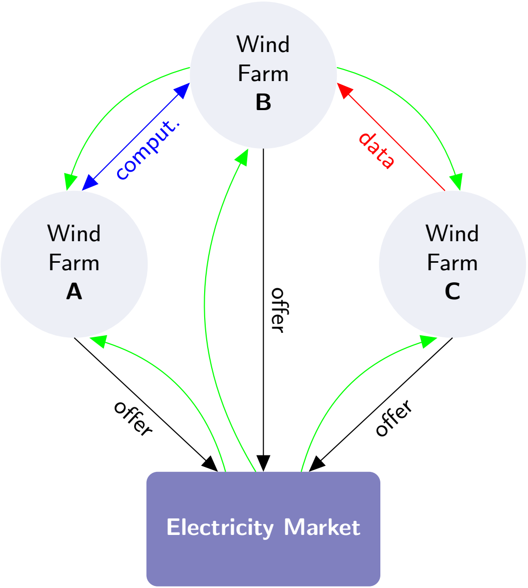
Imperial College London



Opportunity: All *could* benefit from some form of collaboration (e.g., information sharing)

Challenge: They have no interest in doing so

A motivating real-world example

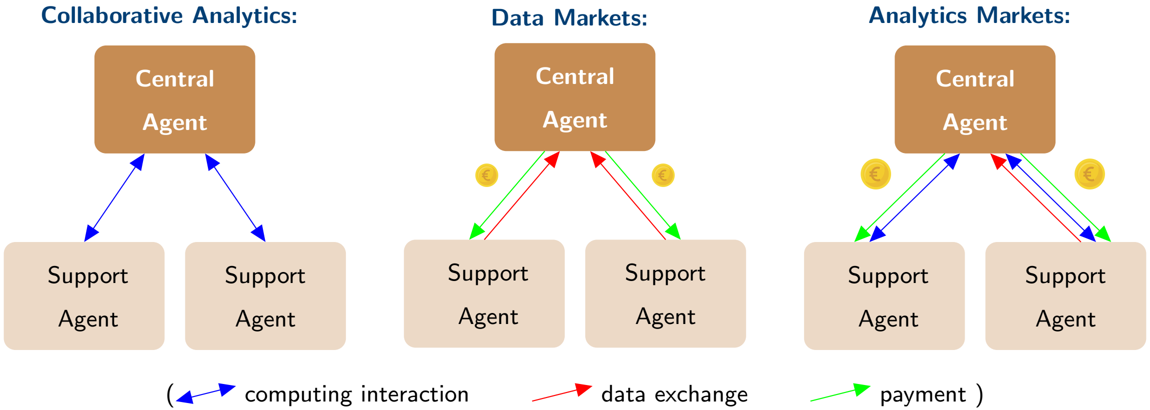


Proposal: Design a framework allowing for all agents to collaborate and benefit from it

2. **Collaborative and market-based analytics platforms**

Various approaches to collaborative and market-based analytics

Agents meet through **analytics platforms** supporting collaborative and market-based analytics



Substantial **methodological research** is needed to design such analytics platforms! (e.g., blending mechanism design, statistical/machine learning, UX/UI, behavioral economics and science, etc.)

Another angle: Prediction markets

Prediction markets have been around for quite a while now...

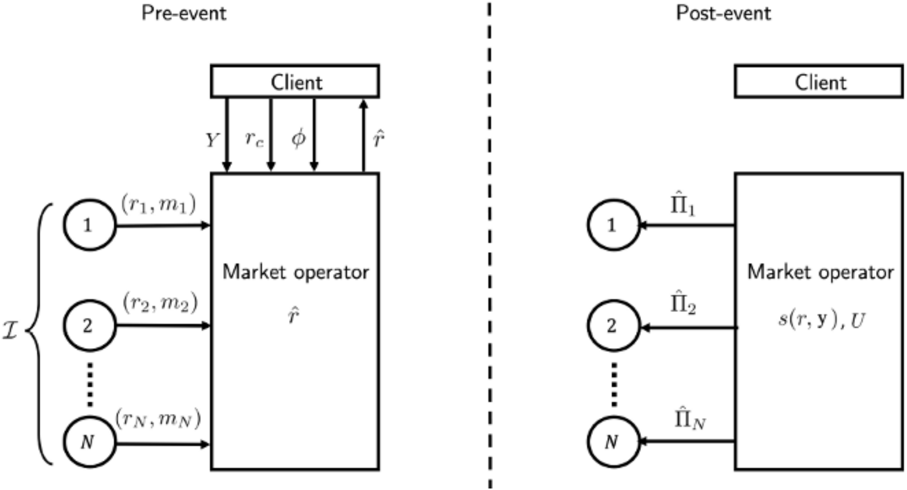
- simply put, you get a number N of agents to bet on a specific outcome (their forecasts for the "event")
- you use N forecasts to produce a final forecast
- agents are monetarily rewarded for their contribution

There has been many interesting applications, e.g.

- Iowa electronic markets (`iem.uiowa.edu`)
- `numer.ai`
- blockchain-based applications... (e.g., Augur)

A proposal wagering mechanism

The client has a **forecast report** r_c (and a utility function $U(r_c, \hat{r}, \phi, y)$), **can others help in improving it**?



- Y is the event the client is interested in, y is the observation
- (r_i, m_i) are the forecast report and wager for agent i

3. **Regression market example**

P. Pinson, L. Han, J. Kazempour (2022) Regression markets and application to energy forecasting. *TOP*, available online (pdf)

The central agent and the regression problem

- Consider a *central agent* ("**Forecaster**") with a regression problem, e.g., as a basis to forecast renewable power generation for a given site (y_{t+k})
- **Forecaster** owns a set ω of m features, $\omega = \{x_1, \ldots, x_m\}$

The following regression problem could be used as basis for eventual prediction,

$$Y_{t+k} = \beta_0 + \sum_{i=1}^{m} \beta_i x_{i,t} + \varepsilon_t, \quad t = 1, \ldots, T$$

The vector of parameters $\boldsymbol{\beta} = [\beta_0 \; \ldots \; \beta_m]^\top$ can easily be learned by minimizing an appropriate loss function

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \, S_\omega(\boldsymbol{\beta}), \qquad S_\omega(\boldsymbol{\beta}) = \frac{1}{T} \sum_{t=1}^{T} \rho \left(y_{t+k} - (\beta_0 + \sum_{i=1}^{m} \beta_i x_{i,t}) \right)$$

where ρ may be any convex loss function (e.g., quadratic, pinball loss, etc.)

Based on the data available, the minimum loss function value is $S^*_\omega = S_\omega(\hat{\boldsymbol{\beta}})$

- **Forecaster** could post the regression task on an analytics platform, to improve model fit
- **Forecaster** declares a willingness to pay of ϕ =1€ per percent-point improvement in S and per data point provided.

Support agents and the augmented regression problem

- Two *support agents* **Good Data** and **Useful Features** may bring in additional features z_1 and z_2, to be remunerated
- The overall set of features now is $\Omega = \omega \cup \{z_1, z_2\}$

The regression problem can then be augmented, as

$$Y_{t+k} = \underbrace{\beta_0 + \sum_{i=1}^{m} \beta_i x_{i,t}}_{\text{Forecaster}} + \underbrace{\gamma_1 z_{1,t}}_{\text{Good Data}} + \underbrace{\gamma_2 z_{2,t}}_{\text{Useful Features}} + \varepsilon_t, \quad t = 1, \ldots, T$$

where the augmented vector of coefficients $\boldsymbol{\beta}^+ = [\beta_0 \ \ldots \ \beta_m \, \gamma_1 \, \gamma_2]^\top$ can be learned similarly, by minimizing an appropriately chosen convex loss function ρ, i.e.,

$$\hat{\boldsymbol{\beta}}^+ = \operatorname*{argmin}_{\boldsymbol{\beta}^+} S_\Omega(\boldsymbol{\beta}^+), \qquad S_\Omega(\boldsymbol{\beta}^+) = \frac{1}{T} \sum_{t=1}^{T} \rho \left(y_{t+k} - (\beta_0 + \sum_{i=1}^{m} \beta_i x_{i,t} + \gamma_1 z_{1,t} + \gamma_2 z_{2,t}) \right)$$

We eventually write $S_\Omega^* = S_\Omega(\hat{\boldsymbol{\beta}}^+)$

- If z_1 and/or z_2 are informative features, one expects $S_\Omega^* < S_\omega^*$

Revenues, payments and market properties

- How to define **revenues** and **payments** in such a regression market?

For each support agent j ($j = 1, 2$), the revenue is given by

$$\pi_j = (S^*_\omega - S^*_\Omega)\, T\, \phi\, \psi_j, \quad j = 1, 2$$

where ψ_j is an allocation policy based on feature valuation (can be obtained with, e.g., leave-one-out or Shapley-based allocation), such that $\sum_j \psi_j = 1$

For **Forecaster**, the payment is

$$\pi_c = \phi(S^*_\omega - S^*_\Omega)\, T$$

Such a simple approach actually yields a market with a wealth of good properties, i.e.,

- budget balance
- symmetry (or anonymity)
- zero element
- incentive compatibility
- individual rationality (truthfulness)

In-sample and out-of-sample

Typically,

- we learn in-sample (batch or online)
- we predict out-of-sample...

Can we use the above concepts more generally?

In-sample and out-of-sample

Typically,

- we learn in-sample (batch or online)
- we predict out-of-sample...

Can we use the above concepts more generally?

In **online regression** markets, the payments can be generalized with

$$\pi_{j,t} = (S^*_{\omega,t} - S^*_{\Omega,t})\, \phi\, \psi_{j,t}, \quad j = 1, 2$$

where $S^*_{\omega,t}$ and $S^*_{\Omega,t}$ are time-varying estimator of the loss function, and $\psi_{j,t}$ is a time-varying estimate of allocation policies (profiting of their linearity property)

And, in **out-of-sample regression** markets (i.e., for genuine forecasting),

$$\pi_{j,t} = (s^*_{\omega,t} - s^*_{\Omega,t})\, \phi\, \psi_{j,t}, \quad j = 1, 2$$

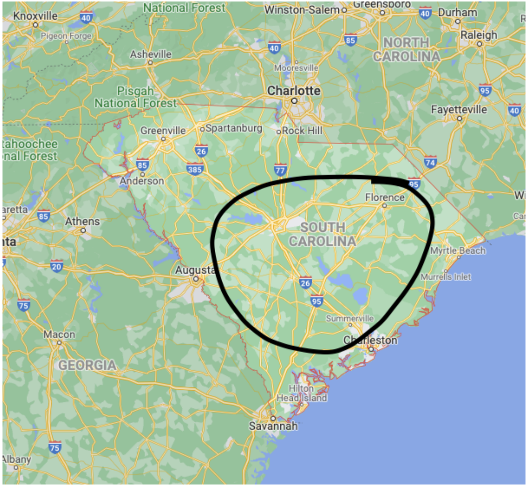
where $s^*_{\omega,t}$ and $s^*_{\Omega,t}$ are time-varying estimator of the loss function, and $\psi_{j,t}$ is the instantaneous allocation policies (i.e., readily Shapley additive explanation)

Batch, online, and out-of-sample regression markets all enjoy the same properties.
(given convex loss functions and models that are linear in their parameters)

A real world case study in South Carolina (USA)

Imperial College London

Wind power generation for 9 locations in South Carolina (US) – 7 years of data with hourly resolution



Agent	P_n [MW]	Lat./Long.	County
a_1	1.75	34.248/-79.75	Florence
a_2	2.96	34.02/-79.537	Florence
a_3	3.38	33.925/-79.958	Florence
a_4	16.11	34.732/-82.122	Laurens
a_5	37.98	34.556/-81.889	Laurens
a_6	30.06	34.334/-82.133	Laurens
a_7	2.53	33.136/-80.857	Colleton
a_7	2.6	33.112/-80.665	Colleton
a_9	1.24	32.641/-80.504	Colleton

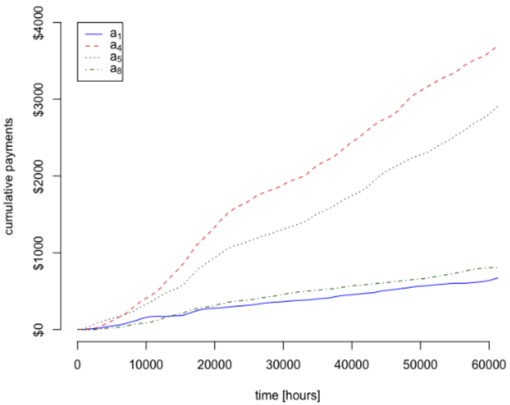
1-hour ahead forecasting based on $Y_{i,t} = \beta_0 + \sum_{\delta=1}^{\Delta} y_{i,t-\delta} + \sum_{j\neq i}\sum_{\delta=1}^{\Delta} y_{j,t-\delta} + \varepsilon_{i,t}$

where a_i is the central agent and a_j $(j \neq i)$ are the support agents

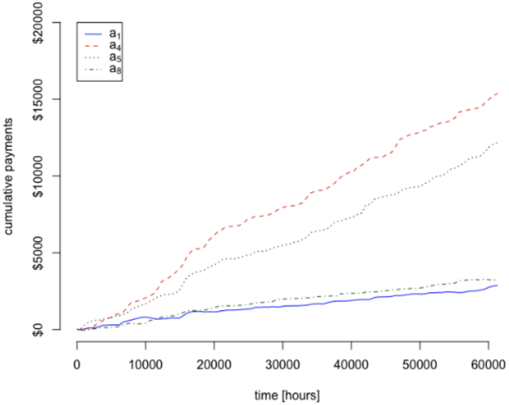
Online and out-of sample regression markets

- Online quantile regression ($\tau = 0.55$) in models with 2 lags for a_i and 1 lag for a_j ($j \neq i$)
- $\phi = 0.2\$$ in-sample, and $\phi = 0.8\$$ out-of-sample (per unit loss, per data point)

Cumulative payments of a_6 towards others:



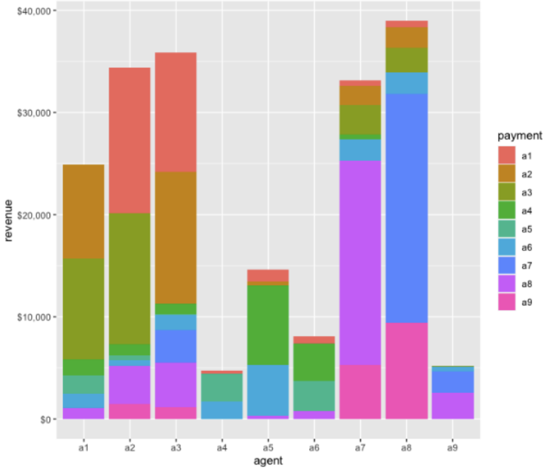
(a) Online regression market.



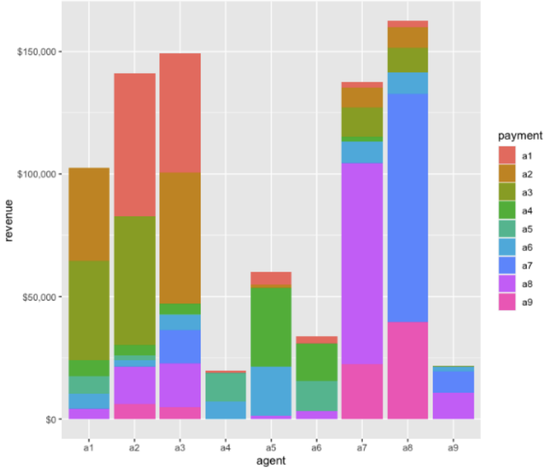
(b) Out-of-sample regression market.

Overall payments

Let's see what happens if they all pay each other for data to improve forecasts...



(a) Online regression market.



(b) Out-of-sample regression market.

4. **Concluding thoughts and discussion**

Collaborative and market-based analytics have a bright future...

- need for many methodological developments
- and focus on relevant business cases and models

In the broader picture, privacy, competition and ethics eventually kick in...!

Thanks for your attention!

Additional readings

1. P. Pinson, L. Han, J. Kazempour (2022) Regression markets and application to energy forecasting. *TOP* **30**: 533–573 (pdf)
2. L. Han, P. Pinson, J. Kazempour (2022) Trading data for wind power forecasting: A regression market with Lasso regularization. Proc. of the Power System Computation (PSCC) conference 2022, Porto, Portugal (arxiv.org/pdf/2110.07432)
3. C. Goncalves, P. Pinson, R. Bessa (2020) Towards data markets in renewable energy forecasting. *IEEE Transactions on Sustainable Energy* **12**(1): 533-542 (pdf)
4. A. M. Kharman, C. Jursitzky, Q. Zhao, P. Ferraro, J. Marecek, P. Pinson, R. Shorten (2022) On the design of decentralised data markets. Preprint, under review (arxiv.org/abs/2206.06299)
5. S. R. Pandey, P. Pinson, P. Popovski (2022) Participation and data valuation in IoT data markets through distributed coalitions. Preprint, under review (arxiv.org/abs/2206.07785)

... among many other papers appearing lately about data markets!