Is more data always better?
Optimal data usage in non-stationary systems

Jakob Krause

Martin Luther University Halle-Wittenberg
School of Economics and Business

March 27, 2018

Updated on March 27, 2018

Connection: Session Title - Paper?

Connection: Session Title - Paper?

- This is primarily a paper on **statistical methodology** with a **risk measurement application**.

Connection: Session Title - Paper?

- This is primarily a paper on **statistical methodology** with a **risk measurement application**.
- **Premise:** The data generating process is changing over time, i.e. that there is genuine parameter instability!

Connection: Session Title - Paper?

- This is primarily a paper on **statistical methodology** with a **risk measurement application**.
- **Premise:** The data generating process is changing over time, i.e. that there is genuine parameter instability!
- **Consequence:** Answers to questions can change over time.

Connection: Session Title - Paper?

- This is primarily a paper on **statistical methodology** with a **risk measurement application**.
- **Premise:** The data generating process is changing over time, i.e. that there is genuine parameter instability!
- **Consequence:** Answers to questions can change over time.
- **Objective:** Identify 'optimal data sets' by trading off

Connection: Session Title - Paper?

- This is primarily a paper on **statistical methodology** with a **risk measurement application**.
- **Premise:** The data generating process is changing over time, i.e. that there is genuine parameter instability!
- **Consequence:** Answers to questions can change over time.
- **Objective:** Identify 'optimal data sets' by trading off
  - estimator convergence errors (incentive: Use as much data as possible)
  - representativeness errors (incentive: Use as few data as possible)

Connection: Session Title - Paper?

- This is primarily a paper on **statistical methodology** with a **risk measurement application**.
- **Premise:** The data generating process is changing over time, i.e. that there is genuine parameter instability!
- **Consequence:** Answers to questions can change over time.
- **Objective:** Identify 'optimal data sets' by trading off
  - estimator convergence errors (incentive: Use as much data as possible)
  - representativeness errors (incentive: Use as few data as possible)
- **Application:** Impact Analysis of a paragraph in the Basel 3 framework (risk measurement).

Roadmap

- Introduction and Motivation
- Setup, Definitions (Representativeness Metrics)
- Error 1: Non-Representativeness
- Error 2: Estimator Convergence
- Tradeoff: Minimal bias estimates
- Application: Impact Study of Basel III paragraph
- Conclusion
- Further Research

Introduction & Motivation

- Q1: Why do Physics and Economic (statistical) methodology look so similar? [1] , [2]

[1] George Soros, (2013). *Fallibility, Reflexivity, and the human uncertainty principle* Journal of Economic Methodology , 2013 Vol. 20, No. 4, 309-329

[2] Mark Kirstein, (2015). *From the Ergodic Hypothesis in Physics to the Ergodic Axiom in Economics*, Prepared for the 7. Wintertagung des ICAE Linz

[3] Atsushi Inoue, Lu Jin, Barbara Rossi, (2017). *Rolling window selection for out-of-sample forecasting with time-varying parameters* Journal of Econometrics 196 (2017) 55-67

[4] Carol Alexander (2008), *Market Risk Analysis, Vol. IV*, Wiley

Introduction & Motivation

- Q1: Why do Physics and Economic (statistical) methodology look so similar? [1], [2]
- Q2: A lot of Economics Studies use rolling window estimation.
  This is an acknowledgement of parameter instability.
  Why has the choice of the rolling window size nothing to do with the objective it wants to achieve? [3]

---

[1] George Soros, (2013). *Fallibility, Reflexivity, and the human uncertainty principle* Journal of Economic Methodology , 2013 Vol. 20, No. 4, 309-329

[2] Mark Kirstein, (2015). *From the Ergodic Hypothesis in Physics to the Ergodic Axiom in Economics*, Prepared for the 7. Wintertagung des ICAE Linz

[3] Atsushi Inoue, Lu Jin, Barbara Rossi, (2017). *Rolling window selection for out-of-sample forecasting with time-varying parameters* Journal of Econometrics 196 (2017) 55-67

[4] Carol Alexander (2008), *Market Risk Analysis, Vol. IV*, Wiley

Introduction & Motivation

- Q1: Why do Physics and Economic (statistical) methodology look so similar? [1], [2]
- Q2: A lot of Economics Studies use rolling window estimation.
  This is an acknowledgement of parameter instability.
  Why has the choice of the rolling window size nothing to do with the objective it wants to achieve? [3]
  - Standard in the Literature
  - 'Experience'
  - Significance [4]

---

[1] George Soros, (2013). *Fallibility, Reflexivity, and the human uncertainty principle* Journal of Economic Methodology , 2013 Vol. 20, No. 4, 309-329

[2] Mark Kirstein, (2015). *From the Ergodic Hypothesis in Physics to the Ergodic Axiom in Economics*, Prepared for the 7. Wintertagung des ICAE Linz

[3] Atsushi Inoue, Lu Jin, Barbara Rossi, (2017). *Rolling window selection for out-of-sample forecasting with time-varying parameters* Journal of Econometrics 196 (2017) 55-67

[4] Carol Alexander (2008), *Market Risk Analysis, Vol. IV*, Wiley

Introduction & Motivation

- Q3: Why is noone correcting course although things went wrong?
  - Long Term Capital Management (LTCM):
    "Theoretically, the odds against a loss such as August's [1998] had been prohibitive; an event so freakish as to be unlikely to occur even once over the entire life of the Universe and even over numerous repitions of the Universe." [5] [6]
  - Financial Crisis
    "We were seeing things that were 25-standard deviation moves, several days in a row." [7] [8]
    25-σ event: ≈ 1 in 10^{137} years (age of univ.: $\approx 1.3 \cdot 10^{10}$ years)

---

[5] Roger Lowenstein, (2001). *When Genius Failed*, Fourth Estate. (p. 159)

[6] Kolman, (1999). *LTCM speaks*, Derivatives Strategy, April 1999.

[7] Financial Times, (2007). *Goldman pays the price of being big*
https://www.ft.com/content/d2121cb6-49cb-11dc-9ffe-0000779fd2ac

[8] Kevin Dowd, John Cotter, Chris Humphrey and Margaret Woods *How Unlucky is 25-Sigma?* available on arXiv:
https://arxiv.org/ftp/arxiv/papers/1103/1103.5672.pdf

Introduction & Motivation

Question: What is a more reasonable weaker assumption that still provides us with enough structure to answer some interesting questions?

Introduction & Motivation

Question: What is a more reasonable weaker assumption that still provides us with enough structure to answer some interesting questions?

- Types of dynamical systems: objective - inter subjective - subjective

Introduction & Motivation

Question: What is a more reasonable weaker assumption that still provides us with enough structure to answer some interesting questions?

- Types of dynamical systems: objective - inter subjective - subjective
  - objective (Physics)
  - inter subjective (Things we agree upon as a society/institute/research community/ ... : Social Sciences (Economics, Finance, ...))
  - subjective

Introduction & Motivation

Question: What is a more reasonable weaker assumption that still provides us with enough structure to answer some interesting questions?

- Types of dynamical systems: objective - inter subjective - subjective
  - objective (Physics)
  - inter subjective (Things we agree upon as a society/institute/research community/ ... : Social Sciences (Economics, Finance, ...))
  - subjective
- Inter subjective things can change.

Introduction & Motivation

Question: What is a more reasonable weaker assumption that still provides us with enough structure to answer some interesting questions?

- Types of dynamical systems: objective - inter subjective - subjective
  - objective (Physics)
  - inter subjective (Things we agree upon as a society/institute/research community/ ... : Social Sciences (Economics, Finance, ...))
  - subjective
- Inter subjective things can change.
- The severity of assumptions should correspond to the robustness of the system they describe.

Introduction & Motivation

Question: What is a more reasonable weaker assumption that still provides us with enough structure to answer some interesting questions?

- Types of dynamical systems: objective - inter subjective - subjective
  - objective (Physics)
  - inter subjective (Things we agree upon as a society/institute/research community/ ... : Social Sciences (Economics, Finance, ...))
  - subjective
- Inter subjective things can change.
- The severity of assumptions should correspond to the robustness of the system they describe.
- **Main Assumption:** Inter subjective things change gradually. (The state parameter of the system is a C^1-function of time.)

Setup

What do we need?

- A convenient class of non-stationary processes.
  - Brownian Motion with time varying volatility

$$dX(t) = \sigma(t)dB(t), \;\; \sigma(\cdot) \in C^1$$

  - more general:
    Semimartingales with 'continuously differentiable' characteristics [9]

[9] Jean Jacod, Albert N. Shiryaev, (2003). *Limit Theorems for Stochastic Processes* Springer-Verlag Berlin Heidelberg, Ch.2.

Setup

What do we need?

- A convenient class of non-stationary processes.
  - Brownian Motion with time varying volatility

$$dX(t) = \sigma(t)dB(t), \quad \sigma(\cdot) \in C^1$$

  - more general:
    Semimartingales with 'continuously differentiable' characteristics [9]
- A formal object that is able to track the developments of a system (Representativeness).
  - Representativeness Metric
    (What we learn and how we change our behavior and correspondingly the rules of the system because of it)

[9] Jean Jacod, Albert N. Shiryaev, (2003). *Limit Theorems for Stochastic Processes* Springer-Verlag Berlin Heidelberg, Ch.2.

Setup

A map

$$I : \mathbb{R} \times \mathbb{R} \to [0, \infty), \quad (t_1, t_2) \mapsto I(t_1, t_2) \tag{1}$$

is called a **representativeness metric** when it fulfills

i) $I(t, t) = 0$ for all $t \in \mathbb{R}$.

ii) $I(t_1, t_2) = I(t_2, t_1)$ for all $t_1, t_2 \in \mathbb{R}$.

iii) $I(t_1, t_3) \leq I(t_1, t_2) + I(t_2, t_3)$ for all $t_1, t_2, t_3 \in \mathbb{R}$.

iv) $\frac{\partial I(t_1, t_2)}{\partial t_2} \geq 0$ for a fixed t_1 and $t_1 < t_2$,

where $t_1, t_2 \in \mathbb{R}$ refer to points in time.

Setup

A map

$$I : \mathbb{R} \times \mathbb{R} \to [0, \infty), \quad (t_1, t_2) \mapsto I(t_1, t_2) \tag{1}$$

is called a **representativeness metric** when it fulfills

i) $I(t, t) = 0$ for all $t \in \mathbb{R}$.

ii) $I(t_1, t_2) = I(t_2, t_1)$ for all $t_1, t_2 \in \mathbb{R}$.

iii) $I(t_1, t_3) \leq I(t_1, t_2) + I(t_2, t_3)$ for all $t_1, t_2, t_3 \in \mathbb{R}$.

iv) $\frac{\partial I(t_1, t_2)}{\partial t_2} \geq 0$ for a fixed t_1 and $t_1 < t_2$,

where $t_1, t_2 \in \mathbb{R}$ refer to points in time.

- This metric measures how much inter subjective properties change over time.
- For our brownian motion:

$$I(t_1, t_2) = \int_{t_1}^{t_2} |\frac{\partial \sigma(s)}{\partial s}| ds \ . \tag{2}$$

Error 1: Nonrepresentativeness Error

Assume we are located at some point in time, T.
Question: Given the representativeness metric. Can one quantify the bias of an estimator when using a specific data set?

- Assume: Perfect knowledge
- Assume: Infinite Data Density
- Objective: Non-stationary data, apply ordinary estimators (e.g. sample variance) and quantify the bias.

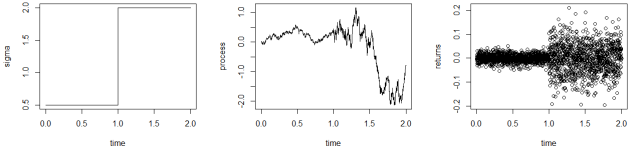
Error 1: Nonrepresentativeness Error

Assume we are located at some point in time, T.

Question: Given the representativeness metric. Can one quantify the bias of an estimator when using a specific data set?

- Assume: Perfect knowledge
- Assume: Infinite Data Density
- Objective: Non-stationary data, apply ordinary estimators (e.g. sample variance) and quantify the bias.

$$\sigma(t) = \begin{cases} \sigma_1 & \text{for } t \in [0,1] \\ \sigma_2 & \text{for } t \in (1,2] \end{cases} . \tag{3}$$



Error 1: Nonrepresentativeness Error

via typical 'algebraic induction'

$$\sigma^2_{[0,2]} = \frac{1}{2}\sigma^2_1 + \frac{1}{2}\sigma^2_2 \to \frac{1}{n}\sum_{i=1}^{n}\sigma^2_i \to \frac{1}{2}\int_0^2 \sigma^2_s ds \tag{4}$$

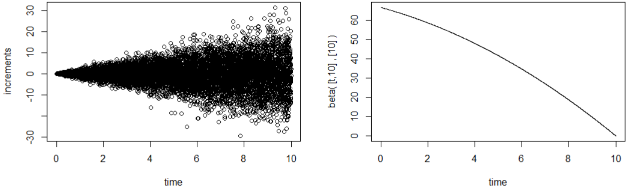
and analogously for any $\sigma^2_{[t,T]}$.

- Formal argument: via characteristic functions, i.e. Levy continuity. Tightness of the sequence of distributions is evident for L^1 random variables (additional assumption).
- Define the **representativeness error** as the 'distance' between σ_T and $\sigma_{[t,T]}$, i.e.
$$\beta(T;[t,T]) = d(\sigma_T, \sigma_{[t,T]})$$
- 'Representativeness' is now an data set/inverval dependent property!

Error 1: Nonrepresentativeness Error

$$\beta([t,T],[T]) = |\sigma_T^2 - \frac{1}{T-t}\int_t^T \sigma^2(u)du| \,. \tag{5}$$

Example: $T = 10, t \in [0,10], \sigma(t) = t$, then



- Write $\sigma^2_{[t,T]}$ for the variance that one would get when applying stationary techniques on the data from the interval $[t,T]$.
- Then:

$$\beta([t,T],[T]) = |\sigma_T^2 - \sigma^2_{[t,T]}| = |\sigma_T^2 - f(I(t,T))| \tag{6}$$

Error 2: Convergence Error

'But in a number of applications we have at our disposal only a discrete-observation of the [data generating process]' [10]

- For i.i.d. normal observations and $d = |\cdot|$:
  convergence rate $E\left[|S_n^2 - \sigma^2|\right] \approx \sqrt{1.129^2 \frac{\sigma^2}{n}}$ [11]

[10] Gejza Dohnal, *On Estimating the Diffusion Coefficient* Journal of Applied Probability, Vol. 24, No. 1 (Mar., 1987), pp. 105-114

[11] ? - There are some interesting Berry-Esseen type questions here

Error 2: Convergence Error

'But in a number of applications we have at our disposal only a discrete-observation of the [data generating process]' [10]

- For i.i.d. normal observations and $d = |\cdot|$:
  convergence rate $E\left[|S_n^2 - \sigma^2|\right] \approx \sqrt{1.129^2 \frac{\sigma^2}{n}}$ [11]
- For i.i.d. normal observations and $d = (\cdot)^2$:
  convergence rate $E\left[(S_n^2 - \sigma^2)^2\right] = \sqrt{2\frac{\sigma^4}{n}}$

[10] Gejza Dohnal, *On Estimating the Diffusion Coefficient* Journal of Applied Probability, Vol. 24, No. 1 (Mar., 1987), pp. 105-114

[11] ? - There are some interesting Berry-Esseen type questions here

Error 2: Convergence Error

'But in a number of applications we have at our disposal only a discrete-observation of the [data generating process]' [10]

- For i.i.d. normal observations and $d = |\cdot|$: convergence rate $E\left[|S_n^2 - \sigma^2|\right] \approx \sqrt{1.129^2 \frac{\sigma^2}{n}}$ [11]
- For i.i.d. normal observations and $d = (\cdot)^2$: convergence rate $E\left[(S_n^2 - \sigma^2)^2\right] = \sqrt{2\frac{\sigma^4}{n}}$
- Natural Emulation ($d = |\cdot|$): $\sqrt{c_{[a,b]}\frac{\sigma^2_{[a,b]}}{\#[a,b]}}$ (Simulations - check)

$$\alpha_{[a,b]} = |c_{[a,b]}\sigma^2_{[a,b]} - s^2_{[a,b]}| = \sqrt{\frac{c_{[a,b]}\sigma^2_{[a,b]}}{\gamma \cdot (b-a)}}, \tag{7}$$

where γ denotes frequency of observations (assuming equidistant observations).

[10] Gejza Dohnal, *On Estimating the Diffusion Coefficient* Journal of Applied Probability, Vol. 24, No. 1 (Mar., 1987), pp. 105-114

[11] ? - There are some interesting Berry-Esseen type questions here

Minimal Bias Estimates

Assuming that the functional form of the combined error is given by ordinary error propagation:

$$t^* = argmin_{s\in[t,T]}\sqrt{\alpha^2_{[s,T]} + \beta^2_{[s,T]}} \tag{8}$$

$$= argmin_{s\in[t,T]}\sqrt{\left(\sqrt{\frac{c_{[s,T]}\sigma^2_{[s,T]}}{\gamma\cdot(T-s)}}\right)^2 + \left(\sigma^2_T - \sigma^2_{[s,T]}\right)^2}. \tag{9}$$

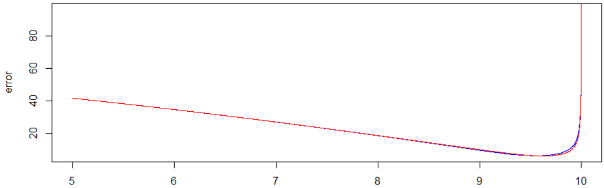
Minimal Bias Estimates

Assuming that the functional form of the combined error is given by ordinary error propagation:

$$t^* = argmin_{s\in[t,T]}\sqrt{\alpha^2_{[s,T]} + \beta^2_{[s,T]}} \tag{8}$$

$$= argmin_{s\in[t,T]}\sqrt{\left(\sqrt{\frac{c_{[s,T]}\sigma^2_{[s,T]}}{\gamma\cdot(T-s)}}\right)^2 + \left(\sigma^2_T - \sigma^2_{[s,T]}\right)^2}. \tag{9}$$

Example: $([5,10], \sigma(t) = t)$ Simulation (functional form above(blue - c assumed constant at 1.129^2) vs empirical error(red)):

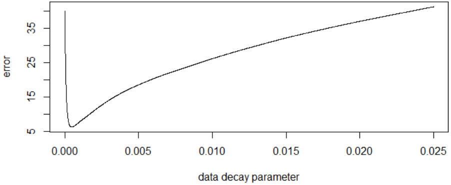


Minimal Bias Estimates - Exponential Weighting

$$p_{T-j} \propto e^{-\frac{\ln 2}{\tau}(T-j)}\,, \tag{10}$$

where τ is a data decay parameter.

- low data decay parameter $\rightarrow$ 'more data'
- high data decay parameter $\rightarrow$ 'less data'
- formal description via 'effective number of scenarios'[12].
- Simulation $([5, 10], \sigma(t) = t)$:

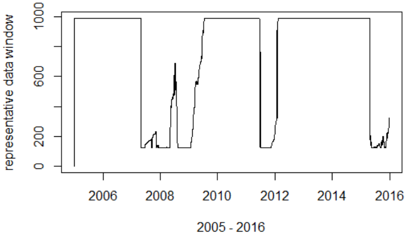


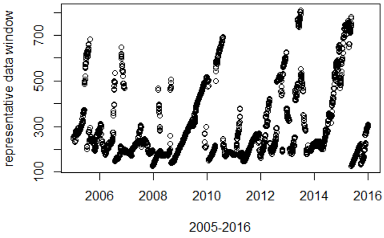
[12] Attilio Meucci, (2012). *Effective Number of Scenarios in Fully Flexible Probabilities* SSRN (1971808)

Algorithms

There are two ways to get to a minimum:

- from right: Start with small data set and increase it until marginal increase in rep error exceeds marginal decrease in convergence error. (right)
- from left: Start with a large data window and shorten it until the difference in the representativeness metric becomes 'tolerable'. (left)[13]





[13] Dax Data, 2001-2016

Non-omniscience

- In Reality we obviously do not know the state of the world.
- We need a proxy that can be an indication of when the rules of the game change.
- There are a variety of plausible options for a choice of the representativeness metric.
- People do not act out of nothing: liquidity metrics (weighted volatility (Amihud), trading volume, spread dynamics, position sizes, ...), fear indices, entropy-based, ...
- **For applications the main structural assumption will be the choice of the representativeness metric.**
- Since here I am only introducing the methodology, I want to sidestep that problem.

Application: Banking Regulation

'The supervisory authority may also require a bank to calculate its Expected Shortfall using a shorter observation period if, in the supervisor's judgement; this is justified by a significant upsurge in price volatility. In this case, however, the period should be no shorter than 6 months.' [14]

Questions/Observations:

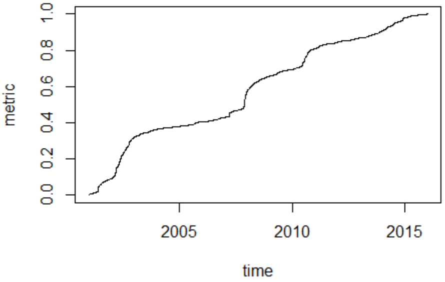
1) Can we use the method developed in this paper to give the regulator a tool that yields the 'optimal amount of data' which should not be shorter than 6 months?
2) Is that a good idea?
3) 'representativeness metric' = 'price volatility'.

[14] *Minimum capital requirements for market risk*, Bank for International Settlements, Jan 2016, section 3, par. 181 (e)

Application: Banking Regulation

Representativeness Metric: Price Volatility, i.e.

$$I(t_1, t_2) = \sum_{t_1}^{t_2} R_t^2 \tag{11}$$



DAX returns, 2001-2016, renormarlised

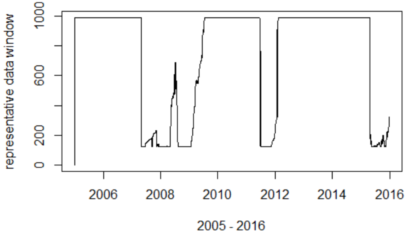
Application: Banking Regulation

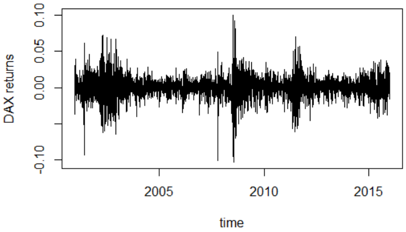
Simple implementation:

- $i = 125$, $D = 1000$
- While
$$\frac{\frac{1}{i} d_I(X_{t-i}, X_t)}{\frac{1}{D} d_I(X_{t-D}, X_t)} \notin [1 - \varepsilon_D, 1 + \varepsilon_D]$$
and $D > i$ reduce D, the optimal amount of data and $\varepsilon_D = \frac{1}{\sqrt{D}}$.

Translation: If short-term and long-term velocity do not fit well together: Reduce the amount of data until it is tolerable with respect to something that depends on the convergence error.

Application: Banking Regulation

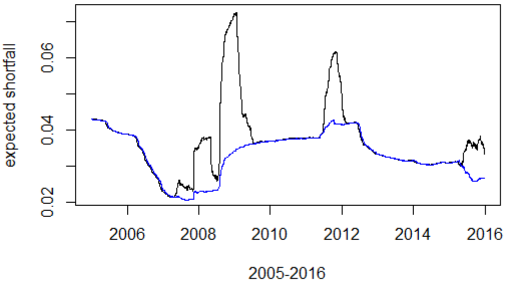




Application: Banking Regulation

Objective: Crude(!) estimate of the effect on risk measure dynamics.

- Estimate historic variance (blue - 1000 days rolling, black - algo)
- Calculate 97.5% quantile (normal distribution)
- multiply by appropriate constant to get Expected Shortfall (normal).



Application: Banking Regulation

Observations:

- Capital Requirements would have increased massively at the start of the financial crisis.
- If the regulator would have done this: 'Second round effects'.
- The analysis is crude but more realistic scenarios would probably be worse (heavy tails? second rounds effects? actual Basel 3 methodology? liquidity risk?)
- Counterperspective: So, the data keeping a bank alive (from a QRM perspective) in the crisis is the calm period before the crisis that is - arguably - not representative for the current situation?

Further Research

- Bourbaki this
  - locally stationary processes
  - estimator convergence for more general distributions + formalisations
  - Differential Geometry ($\kappa =$ nonrep, distance to horizon)
- 'The Role of Data Reduction Techniques in Portfolio Management' (with Francesco Cesarone, Fabrizio Lillo, Fabio Tardella) [15]
- 'Optimal Data Windows in Time Series Estimation' (with Steffen Günther, Angelo Racalbuto)
- Model Risk or Time Dynamics?
- ...

[15] Partly sponsored through ACRI Research Prize (QFW2018)

Conclusion

- I presented a quantitative theory of representativeness and identified optimal data sets in a non-stationary setup.
- Tradeoff: Representativeness vs Estimator Convergence.
- Impact Study

Thank you for your attention!
Questions?
Suggestions?
Remarks?

Information & Contact

`https://ssrn.com/abstract=3063182`

jakob.krause@wiwi.uni-halle.de

`https://jakobkrause.wordpress.com/`