Evaluating Multilevel Predictions from Data - The Case of Trading Data to Predict GDP Growth

Sven Banisch

Robin Lamarche-Perrin

⇒ Eckehard Olbrich

Max-Planck-Institut für Mathematik in den Naturwissenschaften

Conference on Complex Systems 2015

Tempe, Arizona / September 28 – October 2

mathe MACS

Mathematics for Multilevel Anticipatory Complex Systems

MAX-PLANCK-GESELLSCHAFT

$$X^0 \xrightarrow{\qquad T^t \qquad} X^t \xrightarrow{\qquad T^\tau \qquad} X^{t+\tau}$$

- Markovian Kernel $T(X^{t+1}|X^t)$

- Initial State $X^0 \in \Sigma$
- Current State $X^t \in \Sigma$ with Current Time $t \in \mathbb{N}$
- Future State $X^{t+\tau} \in \Sigma$ with Prediction Horizon $\tau \in \mathbb{N}$

$$\begin{tikzcd} & & \psi(X^{t+\tau}) \\ X^0 \arrow[r, "T^t"] & X^t \arrow[r, "T^\tau"] & X^{t+\tau} \arrow[u, "\psi"] \end{tikzcd}$$

- Markovian Kernel $T(X^{t+1}|X^t)$
- Initial State $X^0 \in \Sigma$
- Current State $X^t \in \Sigma$ with Current Time $t \in \mathbb{N}$
- Future State $X^{t+\tau} \in \Sigma$ with Prediction Horizon $\tau \in \mathbb{N}$
- Post-measurement $\psi : \Sigma \to \mathcal{S}_\psi$ defined by $\Pr(\psi(X)|X)$

$$
\begin{tikzcd}
 & \phi(X^t) \arrow[r, dashed, "\text{prediction}"] & \psi(X^{t+\tau}) \\
X^0 \arrow[r, "T^t"] & X^t \arrow[u, "\phi"] \arrow[r, "T^\tau"] & X^{t+\tau} \arrow[u, "\psi"']
\end{tikzcd}
$$

- Markovian Kernel $T(X^{t+1}|X^t)$

- Initial State $X^0 \in \Sigma$
- Current State $X^t \in \Sigma$ with Current Time $t \in \mathbb{N}$
- Future State $X^{t+\tau} \in \Sigma$ with Prediction Horizon $\tau \in \mathbb{N}$

- Post-measurement $\psi : \Sigma \to \mathcal{S}_\psi$ defined by $\Pr(\psi(X)|X)$
- Pre-measurement $\phi : \Sigma \to \mathcal{S}_\phi$ defined by $\Pr(\phi(X)|X)$

- Naively one might think that aggregation always means losing information and therefore the microscopic description would be the best
- However:
  1. In most cases no complete microscopic model is available, thus the predictor has to be inferred from the data
  2. Even if models are available their computation might need a longer time than the prediction horizon
  3. observations might be costly which effectively restricts the number of observables available for prediction

- It might be useful to explore observables on different levels of aggregation!

- Naively one might think that aggregation always means losing information and therefore the microscopic description would be the best
- However:
  1. In most cases no complete microscopic model is available, thus the predictor has to be inferred from the data
  2. Even if models are available their computation might need a longer time than the prediction horizon
  3. observations might be costly which effectively restricts the number of observables available for prediction

- It might be useful to explore observables on different levels of aggregation!

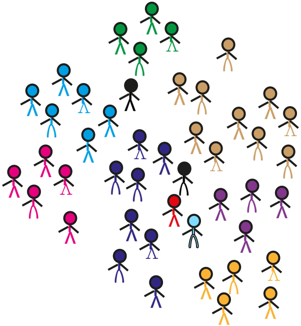
- Naively one might think that aggregation always means losing information and therefore the microscopic description would be the best
- However:
  - In most cases no complete microscopic model is available, thus the predictor has to be inferred from the data

  ⇒ The microscopic state space is high-dimensional which leads to exponentially increasing data requirements and makes inference at this level often infeasible in practice

  ⇒ consider observables on different levels of aggregation!

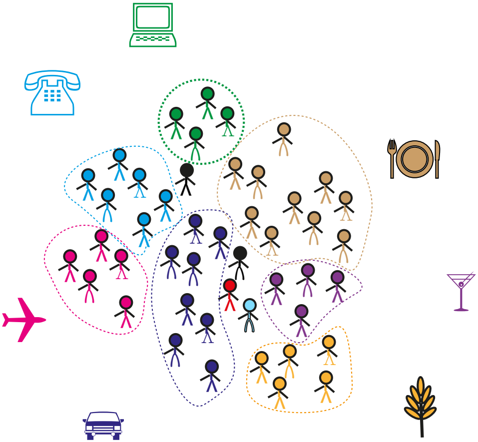
Economy as a Multilevel System

Individuals/Households

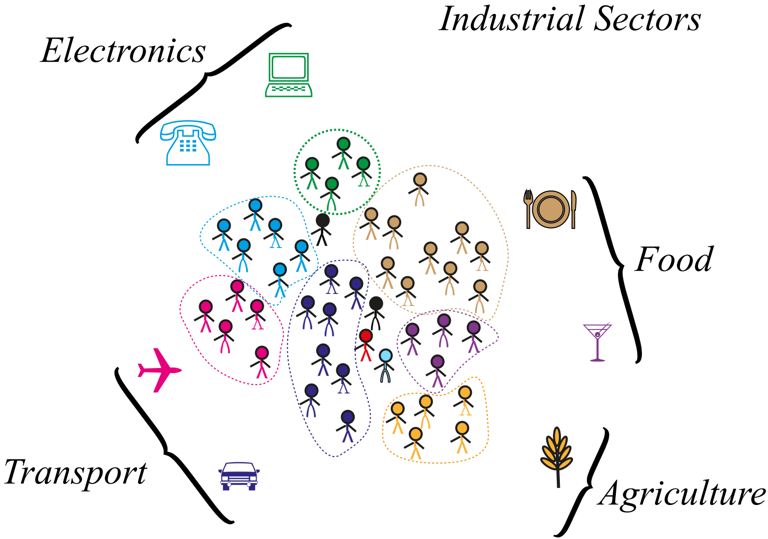


Economy as a Multilevel System

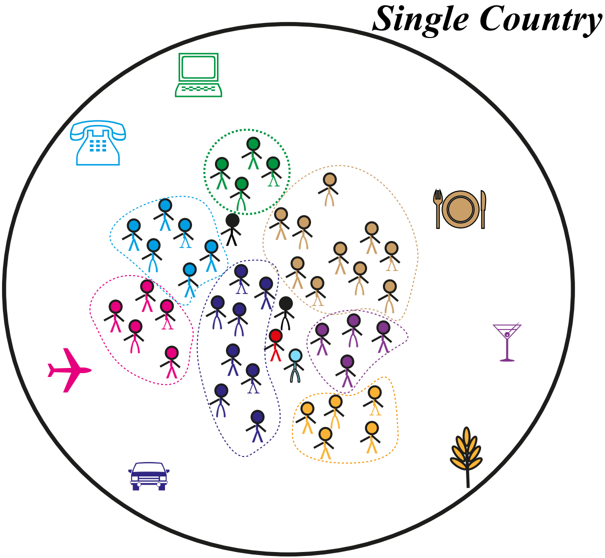
Firms/Production



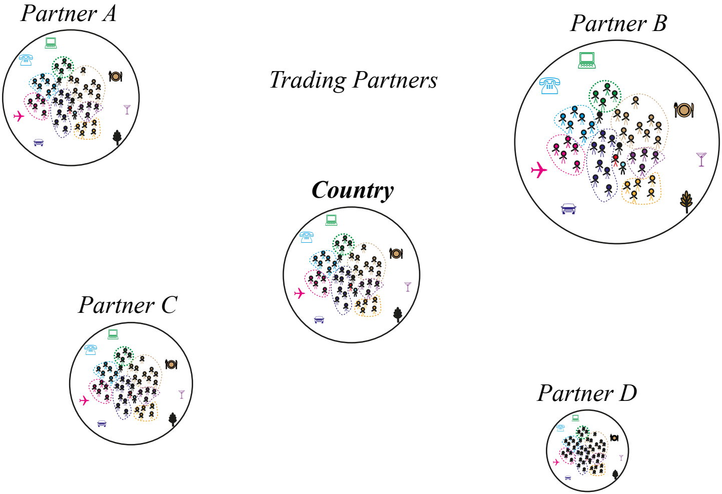
Economy as a Multilevel System



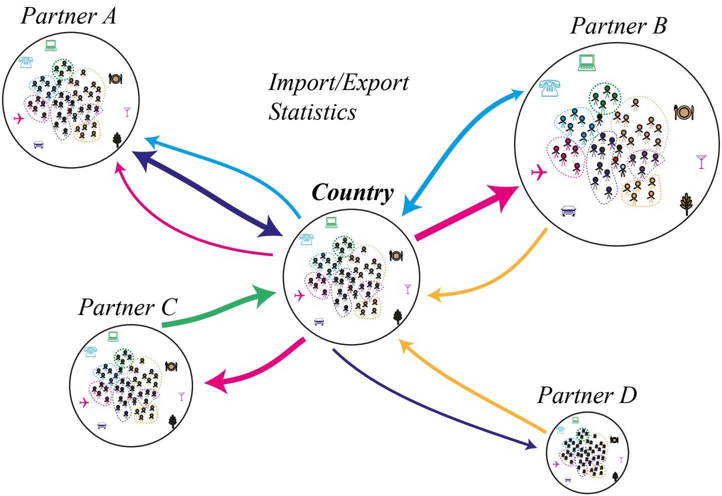
Economy as a Multilevel System



Economy as a Multilevel System



Economy as a Multilevel System



Application | International Trade Data

- In recent years large amounts of data on international trade have been made available
  - export/import volumes between countries for different products (based on UN Comtrade)

data set	countries (regions)	product classes	time
BACI	>200	$\approx$ 5000	since 1994*
TradeMap	>200	5300	since 2001
CHELEM	94	71 (147 ISIC)	since 1967

*data dating back to 1980 is available at lower resolution level

Application | International Trade Data

- In recent years large amounts of data on international trade have been made available
  - export/import volumes between countries for different products (based on UN Comtrade)

data set	countries (regions)	product classes	time
BACI	>200	≈ 5000	since 1994*
TradeMap	>200	5300	since 2001
CHELEM	94	71 (147 ISIC)	since 1967**

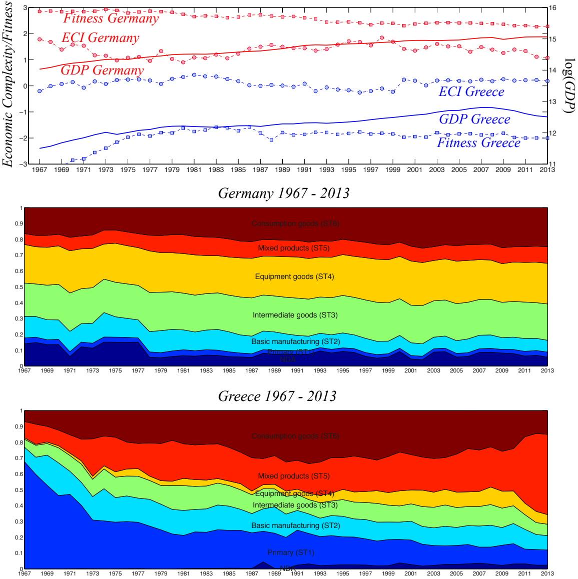
*data dating back to 1980 is available at lower resolution level

** Thanks to CEPII (`http://www.cepii.fr`) for providing us access to the CHELEM database.

Application | International Trade Data

- Measures of economic complexity (Hidalgo/Hausmann 2009)and fitness (Tacchella et al. 2012) proposed on the basis of trade data
  - Compute performance of countries based on their embeddedness in the trade network in the spirit of PageRank
  - Aggregate information from the structure of exports of countries into a single observable
  - Predictive power for growth potential of countries
- Aim here: evaluation of predictive power and comparison to less–aggregated observables
  - CHELEM database provides various product aggregations (production chains, stages, sectors, technological levels)
  - Expect that proportion of exports within the different aggregates is also informative about future
  - »Simple« and easy to interpret; does not take network structure into account

Aggretated and less aggregated observables



Aggregated

ECI: Economic complexity HIDALGO/HAUSMANN 2009

Fitness: Weighted fitness TACCHELLA ET AL. 2012

Less aggregated

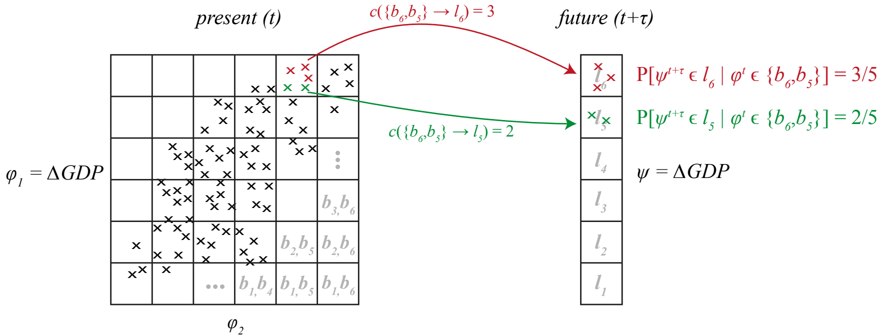
Production stages

Sectors

Prodcution chains

Prediction Method

- Using observables at time t $(\phi_1(X^t), \phi_2(X^t))$ to predict the GDP at time $t+\tau$ $(\psi(X^{t+\tau}))$ or the respective growth rate
  Similar to Cristelli et al. 2015
  - Binning the data and count the number of transitions
    $c(\phi_1 \in b_i \wedge \phi_2 \in b_j \to \psi \in l_k) = c(\{b_i, b_j\} \to l_k)$
  - Predictor: (empirical) conditional probability
    $P(l_k | \{b_i, b_j\}) = \frac{c(\{b_i, b_j\} \to l_k))}{c(\{b_i, b_j\}))}$



Evaluating Probabilistic Forcasts

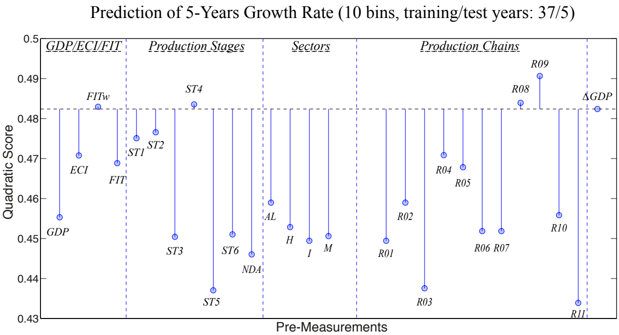
- Split data into training and test set (5 years for testing) and train the predictor $S(l_k|\{b_i, b_j\})$ on the training data

Evaluating Probabilistic Forcasts

- Split data into training and test set (5 years for testing) and train the predictor $S(l_k|\{b_i, b_j\})$ on the training data
- Probabilistic forecasts can be evaluated by *scoring rules*. A scoring rule evaluates an observed data point (i, j, k) on the test data by assigning a score $S(P, k)$.
- For *proper* scoring rules the expected score is maximized if P is the *true* distribution. Proper scores are:
  - Ignorance score: $S(l_k|\{b_i, b_j\}) = \log(P(l_k|\{b_i, b_j\}))$
    - Information-theoretic interpretation
    - Problem with unobserved transitions: $S(l_k|\{b_i, b_j\}) = -\infty$ if $P(l_k|\{b_i, b_j\}) = 0$
  - Quadratic score (used in the following): $S(l_k|\{b_i, b_j\}) = 2P(l_k|\{b_i, b_j\}) - \sum_{k'} P(l_{k'}|\{b_i, b_j\})^2$

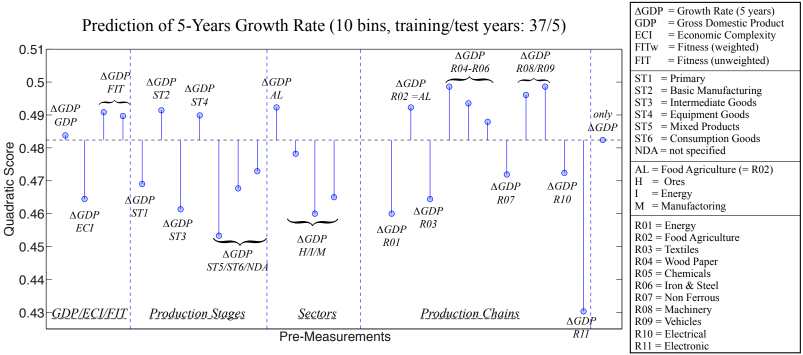
Evaluating Probabilistic Forcasts

- Split data into training and test set (5 years for testing) and train the predictor $S(l_k|\{b_i, b_j\})$ on the training data
- Probabilistic forecasts can be evaluated by *scoring rules*. A scoring rule evaluates an observed data point (i, j, k) on the test data by assigning a score $S(P, k)$.
- For *proper* scoring rules the expected score is maximized if P is the *true* distribution. Proper scores are:
  - Ignorance score: $S(l_k|\{b_i, b_j\}) = \log(P(l_k|\{b_i, b_j\}))$
    - Information-theoretic interpretation
    - Problem with unobserved transitions: $S(l_k|\{b_i, b_j\}) = -\infty$ if $P(l_k|\{b_i, b_j\}) = 0$
  - Quadratic score (used in the following): $S(l_k|\{b_i, b_j\}) = 2P(l_k|\{b_i, b_j\}) - \sum_{k'} P(l_{k'}|\{b_i, b_j\})^2$
- We compare predictors using their average score on the test data.



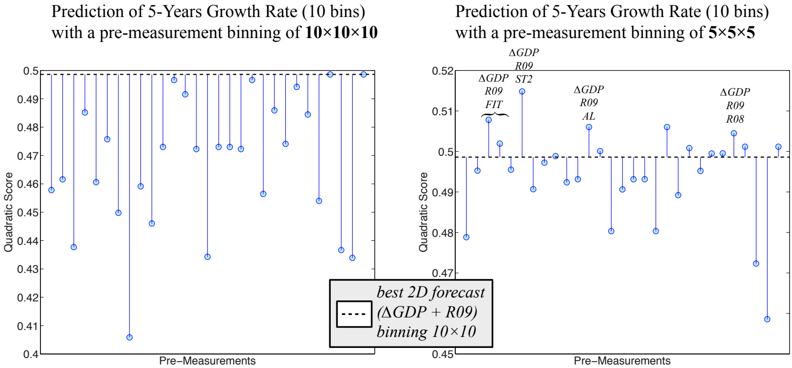
ΔGDP	=	Growth Rate (5 years)
GDP	=	Gross Domestic Product
ECI	=	Economic Complexity
FITw	=	Fitness (weighted)
FIT	=	Fitness (unweighted)
ST1	=	Primary
ST2	=	Basic Manufacturing
ST3	=	Intermediate Goods
ST4	=	Equipment Goods
ST5	=	Mixed Products
ST6	=	Consumption Goods
NDA	=	not specified
AL	=	Food Agriculture (= R02)
H	=	Ores
I	=	Energy
M	=	Manufactoring
R01	=	Energy
R02	=	Food Agriculture
R03	=	Textiles
R04	=	Wood Paper
R05	=	Chemicals
R06	=	Iron & Steel
R07	=	Non Ferrous
R08	=	Machinery
R09	=	Vehicles
R10	=	Electrical
R11	=	Electronic

- Predicting the 5-years growth rate by a selection of single pre-measurements ($\phi = \phi_1$)



- Predicting the 5-years growth rate by a combination of current growth rate and a selection of pre-measurements ($\phi = (\phi_1, \phi_2)$)

Results 〉 Three pre-measurements



- No improvement of forecast if three measures are combined ($\phi = (\phi_1, \phi_2, \phi_3)$) due to overfitting
- But: decreasing the number of bins for the ϕ (of course, not for ψ!) increases scores for particular measurement combinations.
  - Raises questions related to optimal binning

Summary | Outlook

- Data from multilevel systems, such as international trade, can be observed on different levels of aggegation
  - We study the trade-off between the higher information content of less aggregated descriptions and the better inferrability of predictors using higher-level aggregates
  - Trade data: aggregations over meaningful groups of products may outperform higher aggregated measures such as economic complexity while still allowing proper inference of the predictor from the limited amount of data.

Summary | Outlook

- Data from multilevel systems, such as international trade, can be observed on different levels of aggegation
  - We study the trade-off between the higher information content of less aggregated descriptions and the better inferrability of predictors using higher-level aggregates
  - Trade data: aggregations over meaningful groups of products may outperform higher aggregated measures such as economic complexity while still allowing proper inference of the predictor from the limited amount of data.
- Work in progress:
  - Optimal binning problem
  - Heterogeneous predictability (Christelli et al. 2015): Find predictors for different regimes of economic performance
  - Other forecast schemes (e.g. nearest-neighbor-based)

Summary | Outlook

- Data from multilevel systems, such as international trade, can be observed on different levels of aggegation
  - We study the trade-off between the higher information content of less aggregated descriptions and the better inferrability of predictors using higher-level aggregates
  - Trade data: aggregations over meaningful groups of products may outperform higher aggregated measures such as economic complexity while still allowing proper inference of the predictor from the limited amount of data.
- Work in progress:
  - Optimal binning problem
  - Heterogeneous predictability (CHRISTELLI ET AL. 2015): Find predictors for different regimes of economic performance
  - Other forecast schemes (e.g. nearest-neighbor-based)
- Talk addressing the theoretical framework on Friday October 2nd, 11:00 AM - 11:15 AM, Foundations of Complex Systems 1
  The Information Bottleneck Method for Optimal Prediction of the Voter Model

Acknowledgements

Thanks to my collaborators

Robin Lamarche-Perrin

Sven Banisch

mathe ACS Mathematics for Multilevel Anticipatory Complex Systems

SEVENTH FRAMEWORK PROGRAMME

KTS Klaus Tschira Stiftung gemeinnützige GmbH

References

R. Lamarche-Perrin, S. Banisch and E. Olbrich
The Information Bottleneck Method for Optimal Prediction of Multilevel Agent-based Systems
submitted to Advances in Complex Systems, online available as MPIMIS preprint 55/2015

C. A. Hidalgo, R. Hausmann
The building blocks of economic complexity
PNAS **106** (2009) 10570–10575.

A. Tacchella, W. Cristelli, G. Gabrielli and L. Pietronero
A new metrics for countries' fitness and products' complexity
Scientific reports **2** (2012).

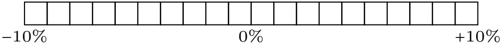
R. Hausmann, C. A. Hidalgo, S. Bustos, M. Coscia, S. Chung, J. Jimenez, A. Simoes, M. A. Yıldırım
The Atlas of Economic Complexity
http://atlas.cid.harvard.edu/

W. Cristelli, A. Tacchella and L. Pietronero
The Heterogeneous Dynamics of Economic Complexity
PLoS ONE **10(2)** (2015), e0117174.

Optimal Binning of Pre-measurements (1)

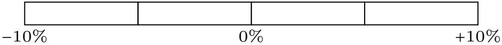
- The efficiency of prediction depends on the resolution that one uses to represent the observed values

  - High Resolution



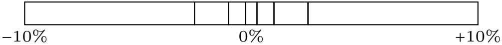
    ⟶ **Highly informative** in theory, but might lead to data **overfitting**

  - Low Resolution



    ⟶ **Less informative**, but allows **generalisation** from limited data

  - Adaptive Resolution



    ⟶ An interesting **trade-off** between information and generalisation

- Problem: How to find the optimal resolution for a given data set? In other words, which binning of the pre-measurement space minimises the score function?

Optimal Binning of Pre-measurements (2)

- Given a micro-resolution of N micro-bins, there are:
  - $\frac{N(N-1)}{2}$ possible bins
  - 2^{N-1} possible binnings $\longrightarrow$ **intractable** by brute-force algorithms

- However:
  - The logarithmic score is **additively decomposable**, that is the score of a binning is the sum of the scores of its bins
  - The scores of all of the $\frac{N(N-1)}{2}$ possible bins can be computed in **quadratic time** $\mathcal{O}(N^2)$
  - In this context, finding a binning that minimises the sum of the scores can also be done in **quadratic time** $\mathcal{O}(N^2)$

$\longrightarrow$ See dynamic algorithms for the *Ordered Set Partitioning Problem*

[Lamarche-Perrin *et al.*, MPI MIS preprint, 2014]