Evaluating Multilevel Predictions from Trading Data

⇒**Sven Banisch**
Robin Lamarche-Perrin
Eckehard Olbrich

Max-Planck-Institut für
Mathematik
in den **Naturwissenschaften**

DPG Frühjahrstagung 2016

Regensburg

mathe MACS
Mathematics for Multilevel Anticipatory Complex Systems

MAX-PLANCK-GESELLSCHAFT

Overview

- This work is an attempt to apply the theoretical ideas developed in
  - R. Lamarche-Perrin, S. Banisch and E. Olbrich, *The Information Bottleneck Method for Optimal Prediction of Multilevel Agent-based Systems*, accepted by ACS / MPIMIS preprint 55/2015
- in the context of trading data and studies on economic complexity
  - Hidalgo/Hausmann *The building blocks of economic complexity* PNAS **106** (2009) 10570–10575.
  - Tacchella et al. *A new metrics for countries' fitness and products' complexity* Scientific reports **2** (2012).
  - Cristelli et al. *The Heterogeneous Dynamics of Economic Complexity* PLoS ONE **10(2)** (2015), e0117174.
- Objectives:
  - evaluation of predictive power of proposed and alternative measures
  - refined understanding of different predictability regimes (Christelli et al. 2015)

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

- 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 $\psi : \Sigma \rightarrow \mathcal{S}_\psi$ 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 $T(X^{t+1}|X^t)$

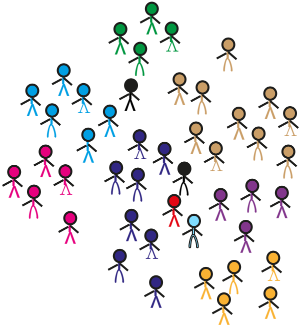
- Initial State $X^0 \in \Sigma$
- Current State $X^t \in \Sigma$ with Current Time $t \in \mathbb{N}$
- Future State $X^{t+\tau} \in \Sigma$ with Prediction Horizon $\tau \in \mathbb{N}$

- 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:
  - In most cases no complete microscopic model $(X^t \xrightarrow{T^\tau} X^{t+\tau})$ 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
- It might be useful to explore observables on different levels of aggregation!

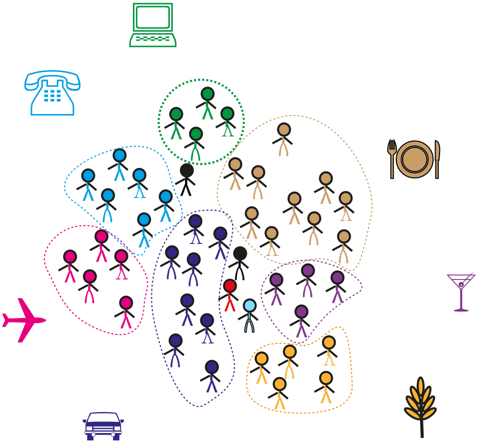
Economy as a Multilevel System

Individuals/Households

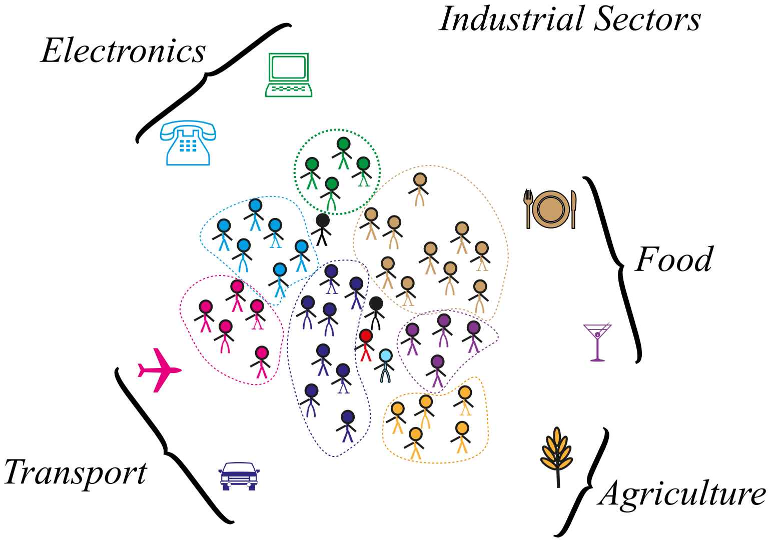


Economy as a Multilevel System

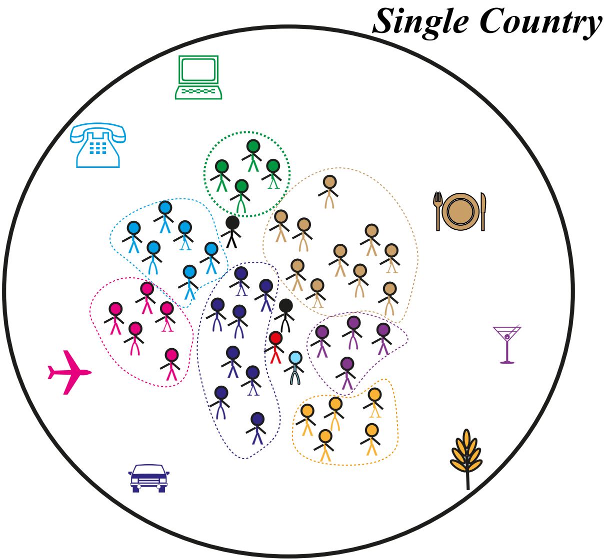
Firms/Production



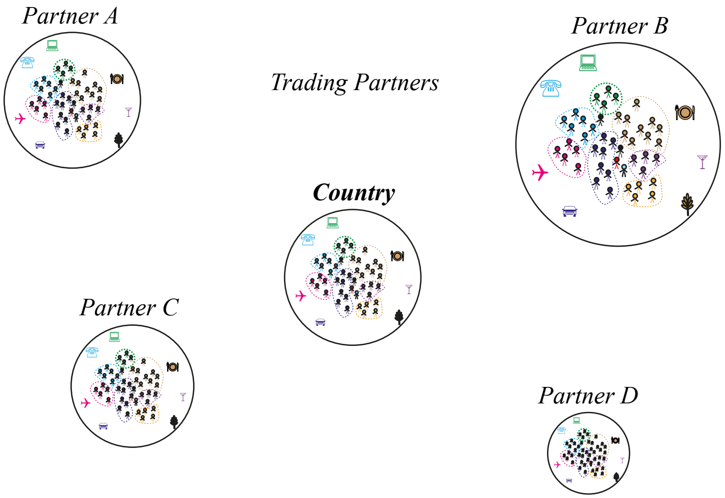
Economy as a Multilevel System



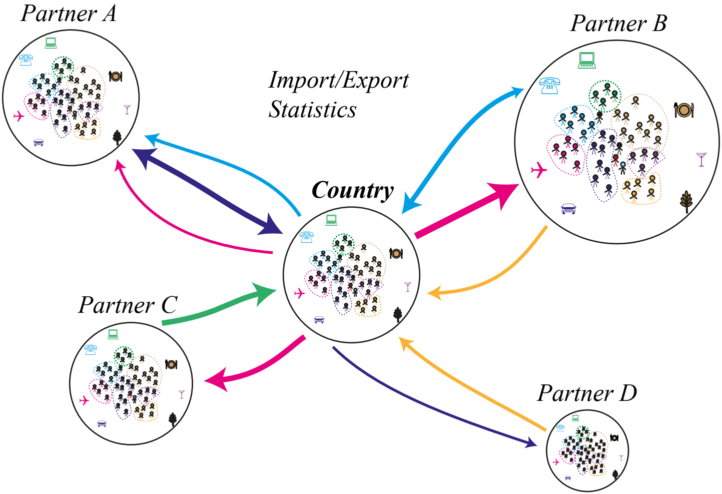
Economy as a Multilevel System



Economy as a Multilevel System



Economy as a Multilevel System



- 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	95	71 (147 ISIC)	since 1967

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

- 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)

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

Application | International Trade Data

2016

What should we measure...

Macroscopic Information:

GDP(2016)

ΔGDP(2015-2016)

Economic Complexity(2016)

Fitness(2016)

Mesoscopic Information:

Production(2016)

pre-measurement φ

Inference

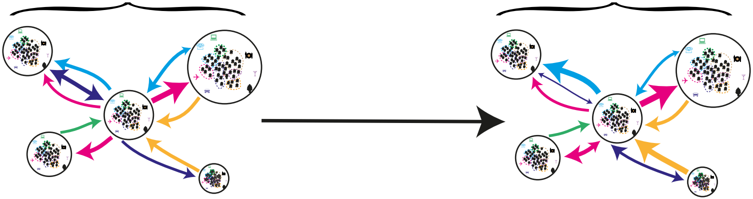
Evaluation

2021

... to most efficiently predict growth?

ΔGDP(2016-2021)

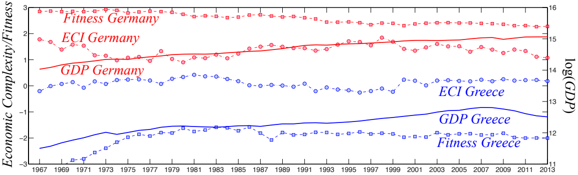
post-measurement ψ



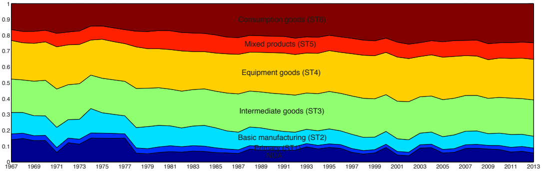
Application | International Trade Data

- 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

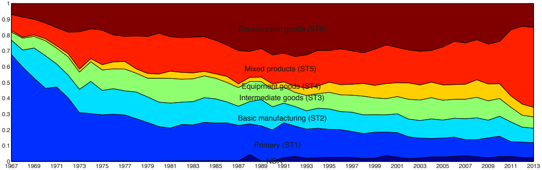
Aggretated and less aggregated observables



Germany 1967 - 2013



Greece 1967 - 2013



Aggregated

ECI: Economic complexity HIDALGO/HAUSMANN 2009

Fitness: Weighted fitness TACCHELLA ET AL. 2012

Less aggregated

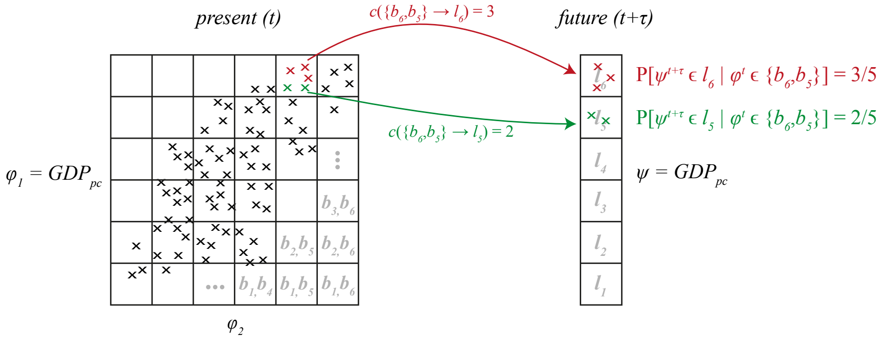
Production stages

Sectors

Prodcution chains

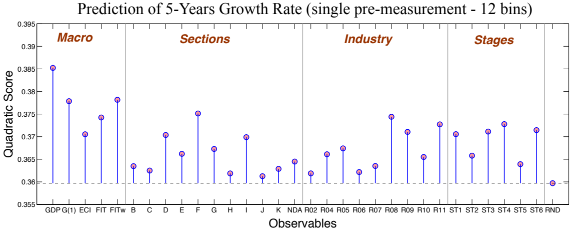
Prediction Method

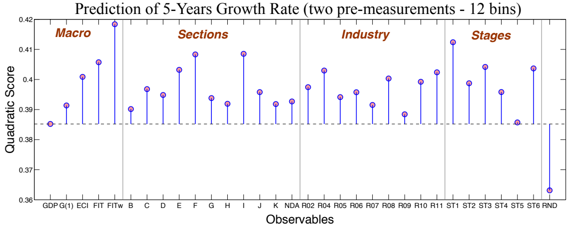
- Using observables at time t $(\phi_1(X^t), \phi_2(X^t))$ to predict the GDP at time $t+\tau$ $(\psi(X^{t+\tau}))$ or the respective growth rate
  Similar to Cristelli et al. 2015
  - Binning the data and count the number of transitions
    $c(\phi_1 \in b_i \wedge \phi_2 \in b_j \to \psi \in l_k) =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

- Leaving-one out cross-validation: for each observation o, train predictor $P(l_k|\{b_i, b_j\})$ using all data except o
- Probabilistic forecasts can be evaluated by *scoring rules*. A scoring rule evaluates an observed data point $o = (i, j, k)$ 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 over all data points o.





GDP = Gross Domestic Product
G(1) = Previous Growth Rate (1 year)
ECI = Economic Complexity
FIT = Fitness (unweighted)
FITw = Fitness (weighted)

B = Construction Products
C = Basic Metals
D = Textiles
E = Wood/Paper
F = Metal Products
G = Chemicals
H = Mining
I = Energy
J = Agriculture
K = Food Products

NDA = not specified

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

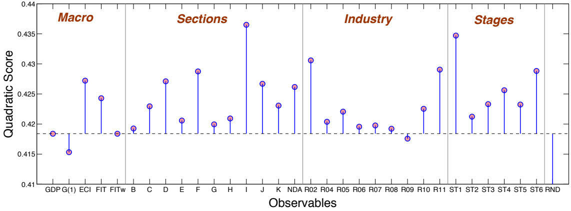
ST1 = Primary
ST2 = Basic Manufacturing
ST3 = Intermediate Goods
ST4 = Equipment Goods
ST5 = Mixed Products
ST6 = Consumption Goods

RND = Surrogate Data

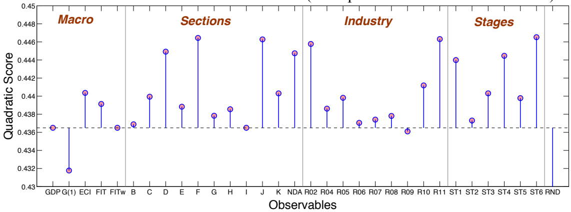
- The best single pre-measurement is current GDP
- Augmented with weighted fitness the score increases considerably, but also most product aggregations provide additional information.

Predicting the 5-years growth rate (3D & 4D)

Prediction of 5-Years Growth Rate (three pre-measurement - 12×12×3)



Prediction of 5-Years Growth Rate (three pre-measurement - 12×12×3×3)



GDP = Gross Domestic Product
G(1) = Previous Growth Rate (1 year)
ECI = Economic Complexity
FIT = Fitness (unweighted)
FITw = Fitness (weighted)

B = Construction Products
C = Basic Metals
D = Textiles
E = Wood/Paper
F = Metal Products
G = Chemicals
H = Mining
I = Energy
J = Agriculture
K = Food Products

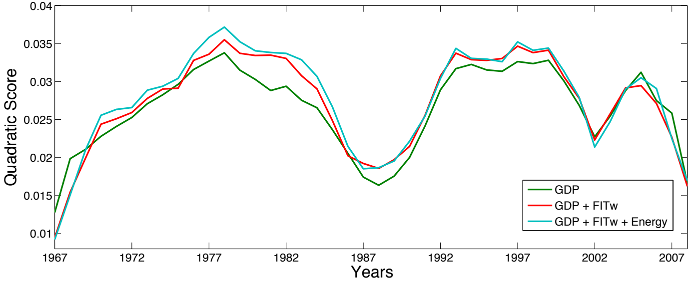
NDA = not specified

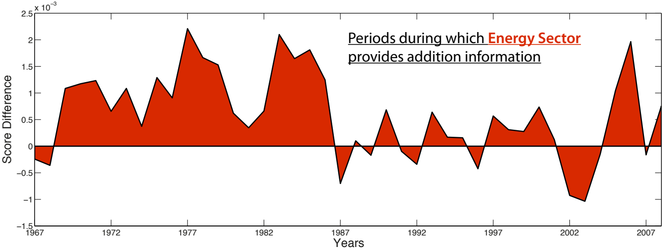
R01 = Energy
R02 = Food Agriculture
R03 = Textiles
R04 = Wood/Paper
R05 = Chemicals
R06 = Iron & Steel
R07 = Non Ferrous
R08 = Machinery
R09 = Vehicles
R10 = Electrical
R11 = Electronic

ST1 = Primary
ST2 = Basic Manufacturing
ST3 = Intermediate Goods
ST4 = Equipment Goods
ST5 = Mixed Products
ST6 = Consumption Goods

RND = Surrogate Data

- Coarse information on energy production (Section I) and several other aggregates provide useful additional information.
- i.p.: consumption goods, electronics, agriculture, metal products and textiles.


Quadratic Score
Years
GDP
GDP + FITw
GDP + FITw + Energy
0.04
0.035
0.03
0.025
0.02
0.015
0.01
1967
1972
1977
1982
1987
1992
1997
2002
2007


x 10^-3
Score Difference
Years
Periods during which Energy Sector provides addition information
2.5
2
1.5
1
0.5
0
−0.5
−1
−1.5
1967
1972
1977
1982
1987
1992
1997
2002
2007

Overview

- This work is an attempt to apply the theoretical ideas developed in
  - R. Lamarche-Perrin, S. Banisch and E. Olbrich, *The Information Bottleneck Method for Optimal Prediction of Multilevel Agent-based Systems*, accepted by ACS / MPIMIS preprint 55/2015
- in the context of trading data and studies on economic complexity
  - Hidalgo/Hausmann *The building blocks of economic complexity* PNAS **106** (2009) 10570–10575.
  - Tacchella et al. *A new metrics for countries' fitness and products' complexity* Scientific reports **2** (2012).
  - Cristelli et al. *The Heterogeneous Dynamics of Economic Complexity* PLoS ONE **10(2)** (2015), e0117174.
- We show:
  - Among the economic complexity measures, weighted fitness (Tacchella et al. 2015) performs best.
  - Product aggregates provide interpretable complementary information.
  - Evaluation method captures heterogeneous predictability in different time periods (crisis as non-stationarity?).
- Next: information-theoretic understanding & non-homogeneous predictors & optimal binning & etcetera