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

Mathematics for Multilevel Anticipatory Complex Systems
Predicting Multilevel Systems

General Setting

\[ X^0 \xrightarrow{T^t} X^t \xrightarrow{T^\tau} X^{t+\tau} \]

- 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 \( \tau \in \mathbb{N} \)
**Predicting Multilevel Systems**

**General Setting**

\[
\begin{align*}
X^0 & \xrightarrow{T^t} X^t & \xrightarrow{T^\tau} X^{t+\tau} \\
\psi(X^{t+\tau}) & \uparrow \psi \\
\end{align*}
\]

- **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** \( \psi : \Sigma \rightarrow S_\psi \) defined by \( \Pr(\psi(X)|X) \)
Predicting Multilevel Systems

General Setting

\[ \phi(X^t) \xrightarrow{\text{prediction}} \psi(X^{t+\tau}) \]

- 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 \( \tau \in \mathbb{N} \)
- Post-measurement \( \psi : \Sigma \rightarrow S_\psi \) defined by \( \Pr(\psi(X)|X) \)
- Pre-measurement \( \phi : \Sigma \rightarrow S_\phi \) 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!
Predicting Multilevel Systems

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

Electronics

Industrial Sectors

Food

Agriculture

Transport
Economy as a Multilevel System

Single Country
Economy as a Multilevel System

Partner A

Partner B

Trading Partners

Country

Partner C

Partner D

Evaluating Multilevel Predictions from Data

Banisch, Lamarche-Perrin, Olbrich
Economy as a Multilevel System

Partner A

Partner B

Partner C

Partner D

Import/Export Statistics

Country

Evaluating Multilevel Predictions from Data
Banisch, Lamarche-Perrin, Olbrich
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)

<table>
<thead>
<tr>
<th>data set</th>
<th>countries (regions)</th>
<th>product classes</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>BACI</td>
<td>&gt;200</td>
<td>≈ 5000</td>
<td>since 1994*</td>
</tr>
<tr>
<td>TradeMap</td>
<td>&gt;200</td>
<td>5300</td>
<td>since 2001</td>
</tr>
<tr>
<td>CHELEM</td>
<td>94</td>
<td>71 (147 ISIC)</td>
<td>since 1967</td>
</tr>
</tbody>
</table>

*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)

<table>
<thead>
<tr>
<th>data set</th>
<th>countries (regions)</th>
<th>product classes</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>BACI</td>
<td>&gt;200</td>
<td>(\approx 5000)</td>
<td>since 1994*</td>
</tr>
<tr>
<td>TradeMap</td>
<td>&gt;200</td>
<td>5300</td>
<td>since 2001</td>
</tr>
<tr>
<td>CHELEM</td>
<td>94</td>
<td>71 (147 ISIC)</td>
<td>since 1967**</td>
</tr>
</tbody>
</table>

*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
Aggregated and less aggregated observables

**Germany 1967 - 2013**

- **Fitness Germany**
- **ECI Germany**
- **GDP Germany**

**Greece 1967 - 2013**

- **Fitness Greece**
- **ECI Greece**
- **GDP Greece**

---

**Aggregated**

**ECI:** Economic complexity  
**Hidalgo/Hausmann**  
2009

**Fitness:** Weighted fitness  
**Tacchella et al.** 2012

---

**Less aggregated**

Production stages  
Sectors  
Production chains

---

*Evaluating Multilevel Predictions from Data*  
Banisch, Lamarche-Perrin, Olbrich
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
  \[
c(\phi_1 \in b_i \land \phi_2 \in b_j \rightarrow \psi \in l_k) = c\left(\{b_i, b_j\} \rightarrow l_k\right)
  \]

- Predictor: (empirical) conditional probability
  \[
P(l_k|\{b_i, b_j\}) = \frac{c(\{b_i, b_j\} \rightarrow l_k))}{c(\{b_i, b_j\}))}
  \]

---

**Diagram:**

- \( \phi_1 = \Delta GDP \)
- \( \phi_2 \)
- \( \psi = \Delta GDP \)
- Present (\( t \))
- Future (\( t+\tau \))

- Transition counts:
  - \( c(\{b_6, b_5\} \rightarrow l_6) = 3 \)
  - \( c(\{b_6, b_5\} \rightarrow l_5) = 2 \)
  - \( P[\psi^{t+\tau} \in l_6 | \phi^t \in \{b_6, b_5\}] = 3/5 \)
  - \( P[\psi^{t+\tau} \in l_5 | \phi^t \in \{b_6, b_5\}] = 2/5 \)
• 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 Forecasts

- 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): $S(l_k|\{b_i, b_j\}) = 2P(l_k|\{b_i, b_j\}) - \sum_{k'} P(l_{k'}|\{b_i, b_j\})^2$
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.
- Predicting the 5-years growth rate by a selection of single pre-measurements ($\hat{\phi} = \phi_1$)
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 $\phi$ (of course, not for $\psi$!) 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 aggregation
  - 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.
Data from multilevel systems, such as international trade, can be observed on different levels of aggregation.

- 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)
Data from multilevel systems, such as international trade, can be observed on different levels of aggregation

- 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
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*

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*
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?
Given a micro-resolution of $N$ micro-bins, there are:

- $\frac{N(N-1)}{2}$ possible bins
- $2^{N-1}$ possible binnings $\rightarrow$ 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 $O(N^2)$
- In this context, finding a binning that minimises the sum of the scores can also be done in quadratic time $O(N^2)$

$\rightarrow$ See dynamic algorithms for the Ordered Set Partitioning Problem

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