Nowcasting with Dynamic Data Masking and Regularized Regression

Erik Alpkvist [1] Jonas Hallgren [2†] Timo Koski [2]
Johannes Siven [1] Ard den Reijer [4]

[1]GSA Capital
London, UK

[†]Corresponding and presenting author
[2]KTH, Royal Institute of Technology
Stockholm, Sweden

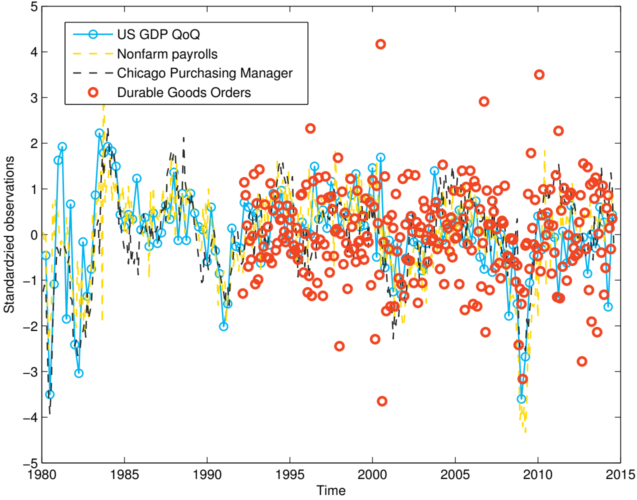
[4]Sveriges Riksbank (Central Bank of Sweden)
Stockholm, Sweden

CFE 2015
Dec 12, 2015, University of London

In *nowcasting*, one wants to forecast a low frequency time series, y, using observations of several higher frequency time series, Z. A typical setting is:

y — quarterly real GDP growth.
Z — a broad cross section of higher frequency economic indicators (e.g. weekly and monthly.)

Task: Find $\mathbb{E}[y_N \mid Z_t]$.


US GDP QoQ
Nonfarm payrolls
Chicago Purchasing Manager
Durable Goods Orders
Standardzied observations
Time
5
4
3
2
1
0
−1
−2
−3
−4
−5
1980
1985
1990
1995
2000
2005
2010
2015

DATA

- Release time stamps are essential.
- ALFRED is used in the study.

MODELS

- Models are on the form

$$y_N = w^T F(Z_t) + \epsilon = w^T X + \epsilon.$$

  We seek the projections F.
- Two popular approaches: MIDAS-type and Dynamic Linear Factor models (DLM).
- The suggested approach is related to MIDAS.

MIDAS

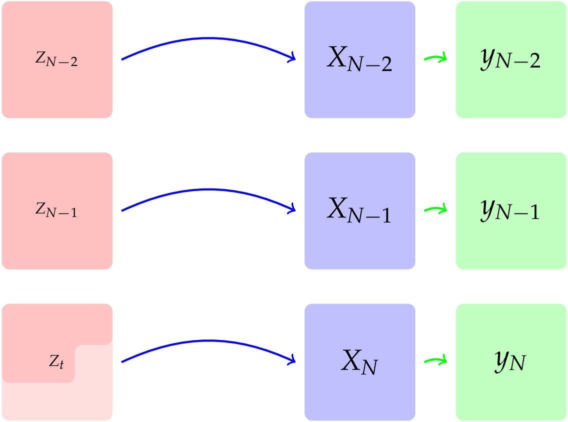
- Mixed-data sampling (MIDAS)[1] projects the time series onto each quarter

$$F^{\text{MIDAS}}(Z_t) = X_t^{\text{MIDAS}}$$

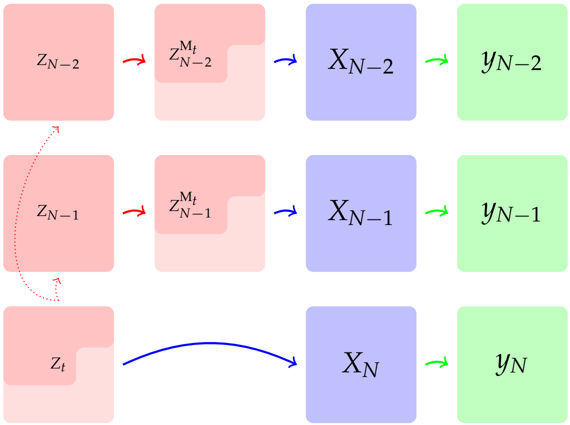
  by e.g. averaging the values. The prediction is $\hat{y}_{N|t} = w^T X_t^{\text{MIDAS}}$.
- Fine for historical regression.
- Forecasting is more difficult due to the "ragged edge" problem.

[1] Ghysels et al. (2004)

MIDAS



Dynamic Masking



Dynamic masking

- Simple idea: Construct regression features dynamically. Similar to MIDAS but the ragged edge problem disappears!
- Features are constructed for immediate use.
- The model is updated when data is updated.
- In both DLM and MIDAS the model is constant throughout the quarter and only data is updated.

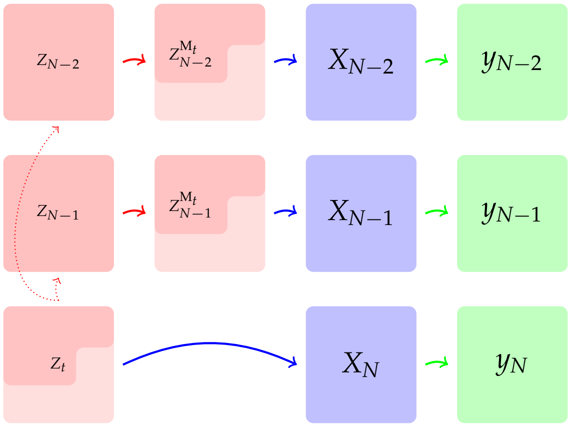
ALGORITHM

1. Find shape of the data available today.
2. Mask out unavailable data from old data-vintages.
3. Use MIDAS-type projection of masked data onto each quarter.
4. Regress on the masked data to get w.
5. Predict: $\hat{y}_{N|t} = w^T F_t^{\text{Mask}}$

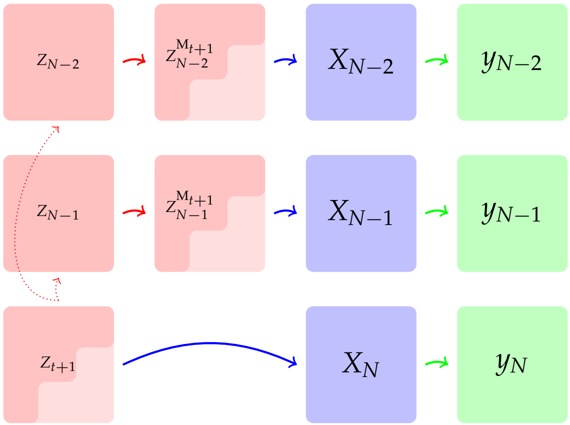
ALGORITHM

1. Find shape of the data available today.
2. Mask out unavailable data from old data-vintages.
3. Use MIDAS-type projection of masked data onto each quarter.
4. Regress on the masked data to get w.
5. Predict: $\hat{y}_{N|t} = w^T F_t^{\text{Mask}} = w^T X_t^{\text{Mask}}$.

Time t



Time $t+1$

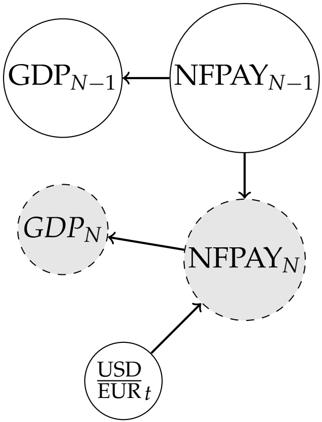


Why Masking?

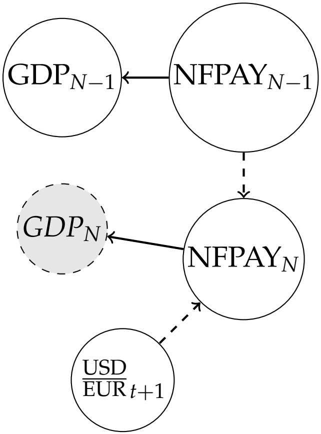
Why Masking?

- Theoretical motivation
- Flexible and easy in implementation
- Good results

Time t



Time $t+1$



$p_\theta(\mathrm{GDP}_N \mid \mathrm{NFPAY}_{N-1}, \frac{\mathrm{USD}}{\mathrm{EUR}}_t)$

$p_{\tilde{\theta}}(\mathrm{GDP}_N \mid \mathrm{NFPAY}_N)$

In general: $\tilde{\theta} \neq \theta$!

METHODS

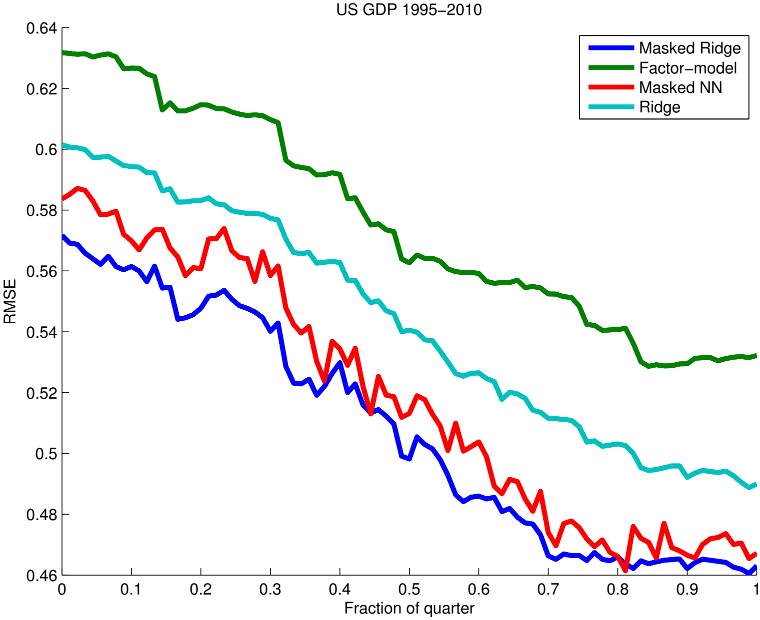
Given X^{Mask} we predict $\hat{y} = w^T \Phi(X^{\text{Mask}})$.

- Regression: Φ is the identity and the estimates are
  - $w^{\text{OLS}} \leftarrow \arg\min_w \|y - w^T F^{\text{Mask}}\|_2$,
  - $w^{\text{Ridge}} \leftarrow \arg\min_w \|y - w^T F^{\text{Mask}}\|_2 + \lambda \|w\|_2$.
- Neural Network

$$\hat{y} = w^T \Phi^{\text{NN}}(X^{\text{Mask}}) = \sum_{k=1}^{M} w_k h(\theta_k^T X^{\text{Mask}}),$$

  where h is the activation function.
- Kernel methods predict using training data:

$$\Phi^{\text{Kernel}}(X^{\text{Mask}}) = k(X^{\text{Mask}}, X^{\text{Mask}}_{\text{Train}}).$$


US GDP 1995–2010
0.64
0.62
0.6
0.58
0.56
0.54
0.52
0.5
0.48
0.46
RMSE
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Fraction of quarter
Masked Ridge
Factor–model
Masked NN
Ridge

Why Masking?

- Theoretical motivation
- Flexible and easy in implementation
- Good results

Any remark, question or suggestion is welcomed!

jonas@math.kth.se
http://people.kth.se/~jhallg/

REFERENCES

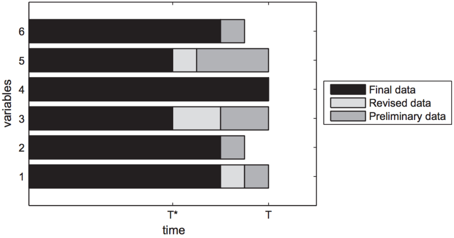
Doz, Catherine and Giannone, Domenico and Reichlin, Lucrezia A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164(1):,188–205, 2011.

Kees E Bouwman and Jan PAM Jacobs. Forecasting with real-time macroeconomic data: the ragged-edge problem and revisions. *Journal of Macroeconomics*, 33(4):784–792, 2011.

Eric Ghysels, Pedro Santa-Clara, and Rossen Valkanov. The MIDAS touch: Mixed data sampling regression models. *Finance*, 2004.

Kenneth F Wallis. Forecasting with an econometric model: The ragged edge problem. *Journal of Forecasting*, 5(1):1–13, 1986.

Ragged Edge[3]



1,6 are published with lags. 2,3,5 are preliminary. 4 is given: interest rate.
Revision can take time:
In 2004 US revised the money supply (M2) series from January 1959 onwards[2]!

[2] Bouwman and Jacobs (2011)
[3] Wallis (1986)