Nowcasting using regression on signatures

Samuel N. Cohen[1,2], Giulia Mantoan[3] Lars Nesheim[2,5]
Áureo de Paula[2,4] **Arthur Turrell**[3, 6] Lingyi Yang[1,2]

[1]University of Oxford [2]The Alan Turing Institute [3]Bank of England [4]University College London [5]University of Bristol [6]No.10 Downing Street

With thanks to former colleagues at the Office for National Statistics.

The views are those of the authors and do not necessarily reflect those of the UK government or Bank of England.

Motivation: policymakers need (the best possible) nowcasts

Examples:

- Monetary policy acts with a lag: want up-to-date info on where economy is now to chart right course for future.
- Financial stability-related debt measures are lagged due to reporting requirements: want full picture of current risks.
- National accounts take time to assemble: want to fill in slow-to-collect data with good early estimates.
- ...

Commonly occurring challenges in nowcasting

- **Ragged-edge:** official data released with significant delay; wide range of other predictors available more frequently.
  Solution: Bridge, MIDAS, discrete-time Kalman filter
- **High dimensional information**: many variables used to predict a single target.
  Solution: Dynamic factor models, Bayesian vector auto-regressions (with shrinkage), penalised estimation methods (L2, L2)
- **Time-varying parameters and nonlinearities**: time series methods often assume stationarity, linearity, constant model parameters; in nowcasting these assumptions can be restrictive.
  Solution: state-space (eg Kalman filters) and other models that are flexible, allow for non-stationarity, nonlinearity, and time-varying parameters.

New challenges in nowcasting

Want to incorporate highly informative alternative and big data sources. But how? Issues include:

1. time series that are too short in the time dimension;
2. missing data points (because collection interrupted or impacted by, eg, covid); and
3. the need to adopt a non-trivial non-linear model.

New challenges in nowcasting

Want to incorporate highly informative alternative and big data sources. But how? Issues include:

1. time series that are too short in the time dimension;
2. missing data points (because collection interrupted or impacted by, eg, covid); and
3. the need to adopt a non-trivial non-linear model.

In this paper, **we propose a nowcasting model (regression on signatures) that addresses issues (2) and (3) and subsumes the linear Kalman filter as a special case.**

Our nowcasting method

- is compatible with standard dimensional reduction techniques

Our nowcasting method

- is compatible with standard dimensional reduction techniques
- handles missing observations and mixed frequencies

Our nowcasting method

- is compatible with standard dimensional reduction techniques
- handles missing observations and mixed frequencies
- requires minimal additional effort when inferring non-linear relationships

Our nowcasting method

- is compatible with standard dimensional reduction techniques
- handles missing observations and mixed frequencies
- requires minimal additional effort when inferring non-linear relationships
- is transparent in its estimation (unlike, say, ML)

Our nowcasting method

- is compatible with standard dimensional reduction techniques
- handles missing observations and mixed frequencies
- requires minimal additional effort when inferring non-linear relationships
- is transparent in its estimation (unlike, say, ML)
- because of the above, is more robust in stressed times

How our nowcasting method works

- **Uses path signatures**
  $\psi_{k,t} \equiv S^k(t, X^1, X^2, \ldots, X^n)$; we use these mathematical objects to capture time series information. Signatures naturally allow for missing data, mixed frequency, irregular sampling, and complex non-linearities.

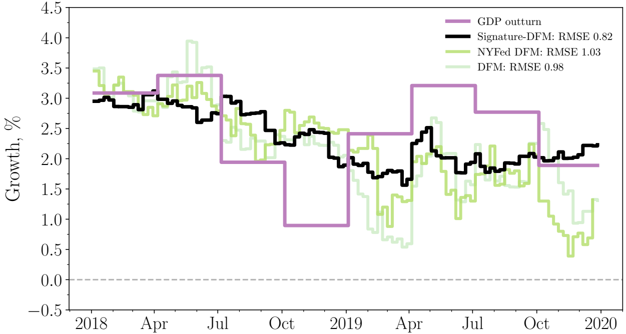


Figure: Out-of-sample NY Fed GDP nowcast

How our nowcasting method works

- **Uses path signatures**

  $\psi_{k,t} \equiv S^k(t, X^1, X^2, \ldots, X^n)$; we use these mathematical objects to capture time series information. Signatures naturally allow for missing data, mixed frequency, irregular sampling, and complex non-linearities.

- **Selects signature terms through standard dimensional reduction**

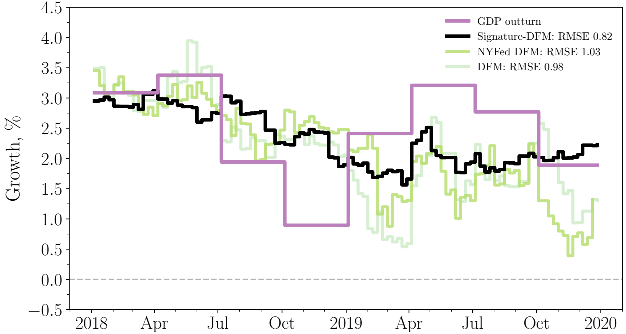


Figure: Out-of-sample NY Fed GDP nowcast

How our nowcasting method works

- **Uses path signatures**

  $\psi_{k,t} \equiv S^k(t, X^1, X^2, \ldots, X^n)$; we use these mathematical objects to capture time series information. Signatures naturally allow for missing data, mixed frequency, irregular sampling, and complex non-linearities.

- **Selects signature terms through standard dimensional reduction**

- **Regresses on signatures**

  $Y_t = \sum_{k=0}^{K}(\alpha_k + \beta_k Y_{t-})\psi_{k,t} + \epsilon_t,$

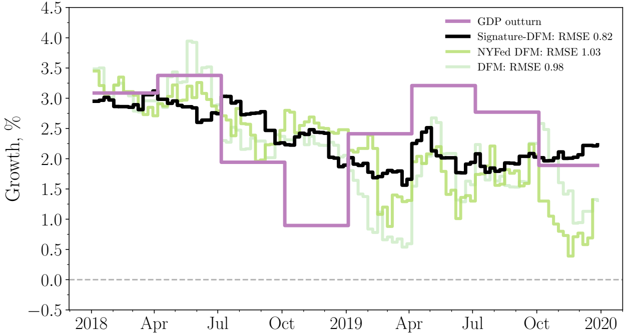


Figure: Out-of-sample NY Fed GDP nowcast

How our nowcasting method works

- **Uses path signatures**
  $\psi_{k,t} \equiv S^k(t, X^1, X^2, \ldots, X^n)$; we use these mathematical objects to capture time series information. Signatures naturally allow for missing data, mixed frequency, irregular sampling, and complex non-linearities.
- **Selects signature terms through standard dimensional reduction**
- **Regresses on signatures**
  $Y_t = \sum_{k=0}^{K}(\alpha_k + \beta_k Y_{t-})\psi_{k,t} + \epsilon_t,$
- ***Outperforms other methods***

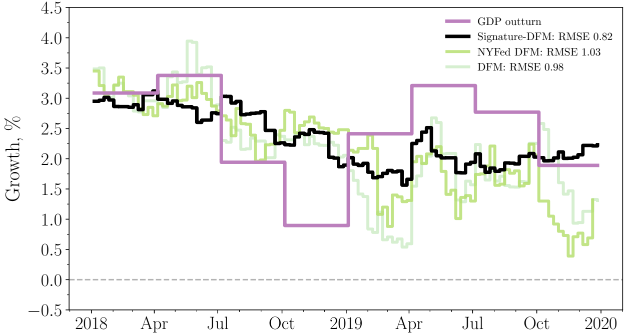


Figure: Out-of-sample NY Fed GDP nowcast

Background: Path Theory

Paths and path integrals

A path in $\mathbb{R}^d$ is $X : [a, b] \to \mathbb{R}^d$, where each component is a 1-D map $X^{(k)} : [a, b] \to \mathbb{R}$.

A path integral is given by

$$\int_a^b f(t) dX_t.$$

with the normal definition of Riemann integration: $\lim_{\pi \to 0} \sum_i f(X_{\xi_i})(X_{t_i} - X_{t_{i-1}})$ for mesh-size π and $t_{i-1} \leq \xi_i \leq t_i$.

Path integrals are:

- invariant to translations in X,
- invariant to reparameterisation in t.

Path signatures

Let

$$S(X)^{(k)}_{a,t} = \int_{a<s<t} dX^{(k)}_s = X^{(k)}_t - X^{(k)}_a.$$

Next let us define the double iterated integral as

$$S(X)^{(k,l)}_{a,t} = \int_{a<s<t} S(X)^{(k)}_{a,t} dX^{(l)}_s = \int_{a<r<s<t} dX^{(k)}_r dX^{(l)}_s.$$

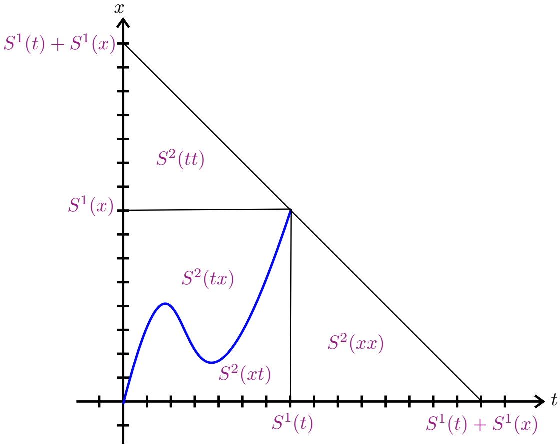
Similarly, we define all the higher order index terms as iterated integrals, then the signature of the path is the infinite collection of all such terms

$$S(X)_{a,b} = \Big(1, S(X)^{(k)}_{a,b}, \dots, S(X)^{(d)}_{a,b}, S(X)^{(1,1)}_{a,b}, S(X)^{(1,2)}_{a,b}, \dots\Big).$$

Visualisation of signature terms

The path signature captures geometric information, such as the order of events.

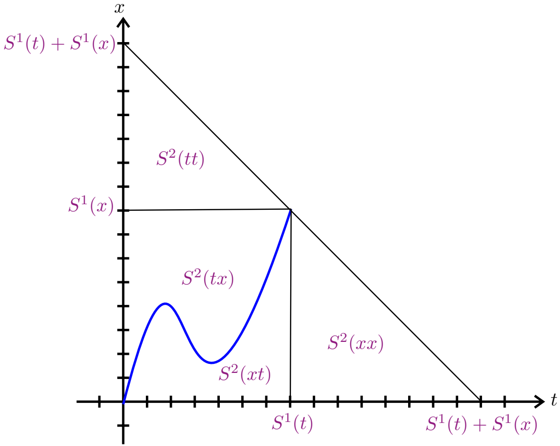
1. First level terms give the change/increment in each dimension between start and end of path



Visualisation of signature terms

The path signature captures geometric information, such as the order of events.

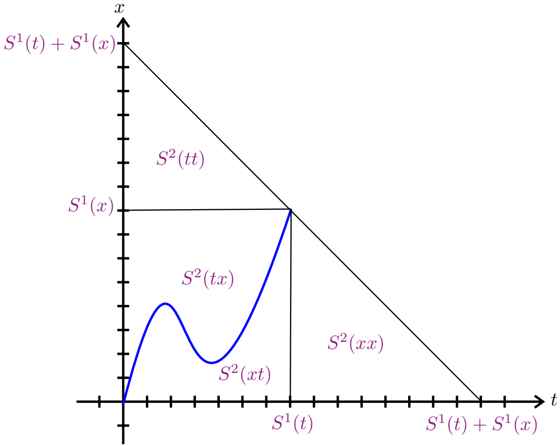
1. First level terms give the change/increment in each dimension between start and end of path
2. Second level related to areas bounded by the path



Visualisation of signature terms

The path signature captures geometric information, such as the order of events.

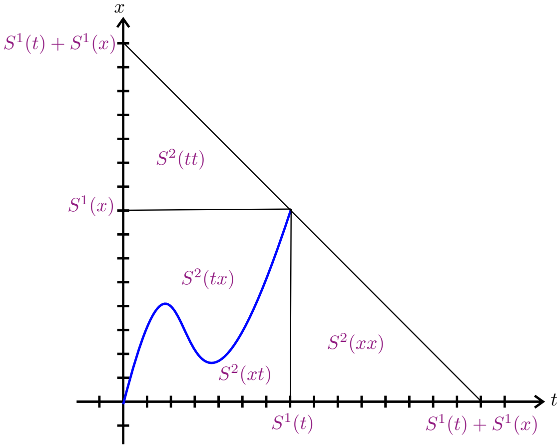
1. First level terms give the change/increment in each dimension between start and end of path
2. Second level related to areas bounded by the path
3. Cross terms related to area



Visualisation of signature terms

The path signature captures geometric information, such as the order of events.

1. First level terms give the change/increment in each dimension between start and end of path
2. Second level related to areas bounded by the path
3. Cross terms related to area
4. Infinite number of terms to capture all information—cf other universal function approximations



Model theory: nowcasting with signatures

Nowcasting with Signatures

Let Y_t be the low-frequency variable we want to nowcast and $X_t \in \mathbb{R}^K$, be the K high-frequency explanatory variables.

We compute a set of truncated signatures $S^{(n)}(X_t)$. As the signatures have a universal approximation property for function on streams, we suppose that

$$Y_t = \beta^\intercal S^{(n)}(X_t) + \epsilon_t,$$

where we can fit the coefficients β through (regularised) regression. Note that the first term of the truncated signature object is 1, allowing an intercept to be fitted.

What about the infinite number of terms!?

There is factorial decay of the magnitude of signature terms with increasing levels (**Lyons (2007)**). Let $X : [0, T] \to \mathbb{R}^d$ be a path with finite one-variation (total variation), that is

$$||X||_{1,[0,T]} \equiv \sup_{P \in \mathcal{P}} \sum_{i=0}^{n_P - 1} |X_{i+1} - X_i| < \infty,$$

where $\mathcal{P}$ is the set of all partitions of $[0, T]$ and n_P is the number of points in partition P. Then, for each level $k \geq 1$,

$$S^k(X \dots X)_{[0,T]} = \int_0^T \cdots \int_0^{t_1} dX_{t_0} \dots dX_{t_{k-1}} \leq \frac{||X||^k_{1,[0,T]}}{k!}. \quad (1)$$

If the signature is truncated at a sufficiently high level, it still captures most of the information

What about the curse of dimensionality?

At higher truncation levels, the number of signature terms can still be numerous. We solve this with **standard regularisation techniques**. For example elastic net (**Zou (2005)**), which includes both L1 penalty and L2 penalty terms, can be used to perform (signature) feature selection. The optimisation problem being solved is given by

$$\min_{\alpha_k,\beta_k} \frac{1}{n}\left|\left|Y_t - \sum_{k=1}^{K}(\alpha_k + \beta_k Y_{t-})\psi_{k,t}\right|\right|_2^2 + \gamma\lambda||\beta||_1 + \gamma(1-\lambda)||\beta||_2^2,$$

where n is the number of time points that data were collected from, γ is the regularisation strength parameter, and $\lambda \in (0,1)$ is the L1 ratio.

Useful properties of signatures for nowcasting

- The signature terms describe a continuous path rather than discrete points: so can be used with irregular sampling patterns
- Path signatures capture geometric information, such as the order of events
- Any continuous function on paths can be approximated arbitrarily well through a linear combination of signature terms (Stone-Wierstrass theorem)
- Can capture linear relationships with at most one signature integral that is wrt to data; non-linear relationships can be captured with higher level truncation. Means do not need to pre-process/transform the data (eg differencing) if using high level of truncation.

Subsuming the Kalman filter: proof

Definitions: continuous time state-space models

We assume a hidden process Y and observed process X.

$$dY_t = (F_t Y_t + f_t)dt + \sigma_t dV_t$$
$$dX_t = (H_t Y_t + h_t)dt + dW_t,$$

where $F \in \mathbb{R}^{d\times d}$, $\sigma \in \mathbb{R}^{d\times p}$, $f \in \mathbb{R}^d$, $H \in \mathbb{R}^{m\times d}$, $h \in \mathbb{R}^m$, so Y is d-dimensional, X is m-dimensional. The processes V, W are Brownian motions, and W is independent of Y.

Regression on signatures generalises the Kalman filter. I

Theorem

The continuous time filter (Kalman-Bucy):

$$d\hat{Y}_t = (F_t\hat{Y}_t + f_t)dt + R_t H_t^\top (dX_t - (H_t\hat{Y}_t + h_t)dt) \tag{2}$$

can be equivalently written as a linear function of the initial state $d\hat{Y}_0$ and the signature of the augmented observation process (t, X_t). In particular, this representation uses only the signature terms where the observation X appears once in the iterated integral.

Regression on signatures generalises the Kalman filter. II

$A_t \equiv F_t - R_t H_t^\top H_t,$

$\xi_t \equiv \int_0^t (f_s - R_s H_s^\top h_s) ds + \int_0^t R_s H_s^\top dX_s$

which gives us the simplified expression for the filter

$$d\hat{Y}_t = A_t \hat{Y}_t dt + d\xi_t.$$

Denote the iterated integrals of A and ξ by

$\mathbb{A}_t^n = \int_0^t \int_0^{t_1} \cdots \int_0^{t_{n-1}} \Big(A_{t_1} A_{t_2} \cdots A_{t_n} \Big) dt_n \cdots dt_1$

$\Xi_t^n = \int_0^t \int_0^{t_1} \cdots \int_0^{t_{n-1}} \Big(A_{t_1} A_{t_2} \cdots A_{t_n} \xi_{t_n} \Big) dt_n \cdots dt_1,$

Regression on signatures generalises the Kalman filter. III

Writing $d\hat{Y}_t$ and $\hat{Y}$ in integral from, we have:

$$\hat{Y}_t = \hat{Y}_0 + \int_0^t A_s \hat{Y}_s ds + \xi_t.$$

Can then show that the value of $\hat{Y}$ can be expressed through the series solution

$$\hat{Y}_t = \sum_{n \geq 0} \left(\mathbb{A}_t^n \hat{Y}_0 + \Xi_t^n \right).$$

This is the initial state and the signatures!

Simulations: performance vs Kalman filter

The setup of the simulations

How will the signature method perform when the sampling is a) regular but mixed frequency, b) irregular, c) non-linearly transformed, and d) irregular **and** non-linearly transformed? Suppose each simulation/path starts at $Y_0 = 0.1$, and the simulation follows the dynamics

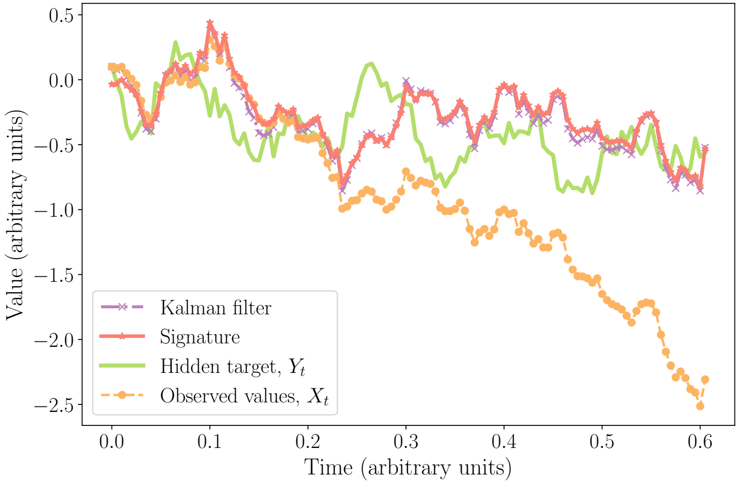
$$dY_t = -Y_t dt + \sqrt{2} dV_t$$
$$dX_t = 10 Y_t dt + dW_t,$$

where V_t, W_t are independent, normally distributed r.v.s in $\mathbb{R}$.

The end time of each simulation, T, is randomly sampled as Uniform on $[0.1, 1]$. We take the mesh size in time to be $\Delta t = 0.005$.

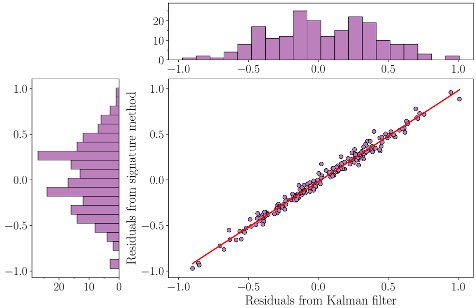
Signature similar to Kalman filter in tracking true values...

Mixed frequency, regular sampling



...this is confirmed by a comparison of residuals

Mixed frequency, regular sampling

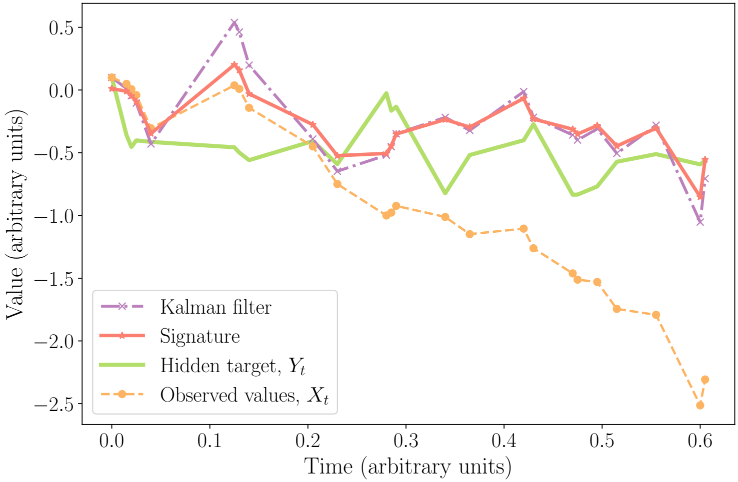


Intercept: -0.02

Gradient: 1.00

Irregular sampling: Kalman filtering and Signatures comparable

Downsampling used to get irregular series



Nonlinear transformation of input data

The signature method can be applied almost off-the-shelf to transformed problems. We demonstrate by applying a sigmoid transform to the input data of our simulation example, i.e. for input X we take

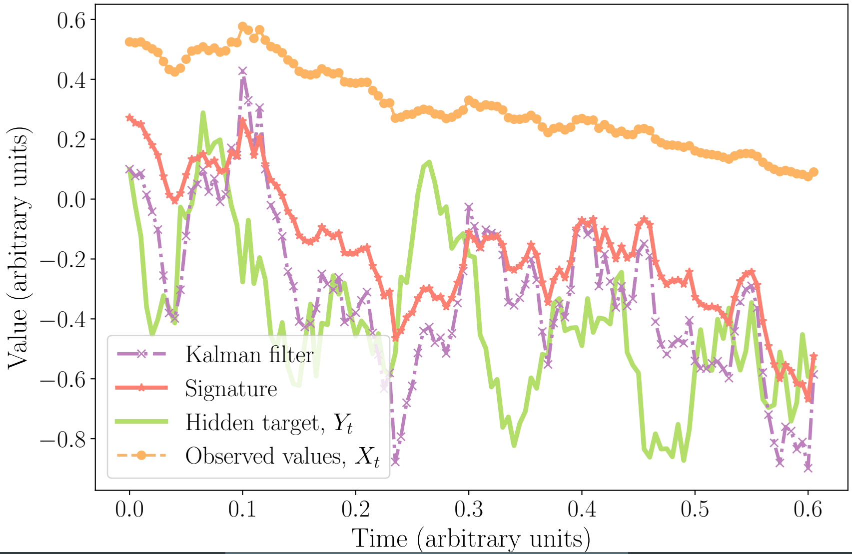
$$\bar{X} = g(X) = \frac{1}{1 + e^{-X}},$$

with the derivative of the inverse of g given by

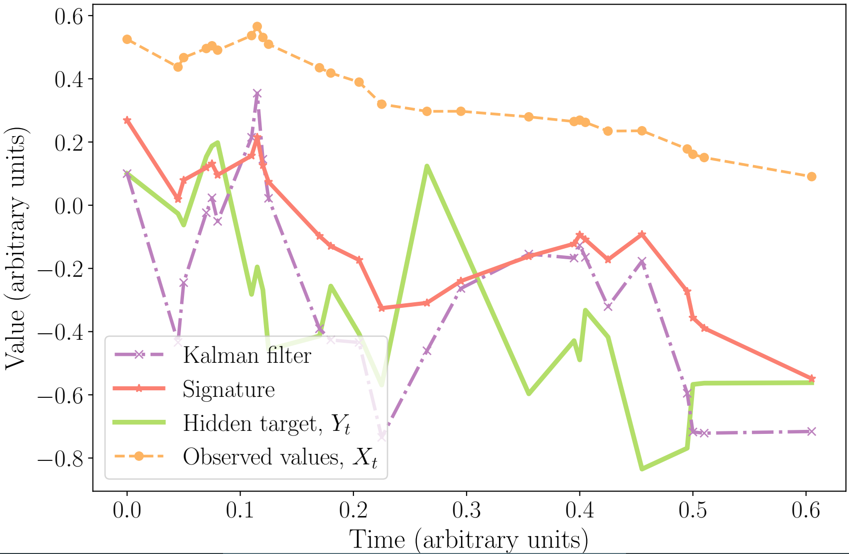
$$\frac{dg^{-1}}{d\bar{X}} =: \bar{g}(\bar{X}) = \frac{1}{\bar{X}(1 - \bar{X})}.$$

The optimal filter is obtained by using $\bar{g}(\bar{X})d\bar{X}$ in the place of dX in $\hat{Y}_t$.

In a nonlinear context, signatures perform well



Signatures perform strongly with nonlinearity *and* irregular sampling



Application: Nowcasting US GDP growth

US GDP Nowcasting model

Question: How does the signature method perform alongside existing nowcasting models?

- Nowcast household income with a limited set of indicators
- Training: Jan 2000 to Dec 2015; validation: Jan 2016 to Dec 2017; Test: Jan 2018 to Dec 2019
- Nowcast horizons: t+0 day, t+15 days and t+45 days; where t denotes a reference quarter
- Models: DFM, Fed DFM, signature with feature selection via PCA, signature with dynamic factors as features

US GDP Nowcasting model

New York Fed Staff Nowcast reports weekly estimates of the US GDP utilising a range of macroeconomic indicators.

We benchmark the signature method against a dynamic factor model which is based on details provided in Bok et al. (2017) but with some differences in the data.

US GDP: Data

- In order to nowcast real US GDP growth, we use 33 of the 37 variables proposed in Bok et al. as the ISM data are not publicly available.
- The indicators are monthly variables that covers housing, income, manufacturing, labor, surveys, trade, and consumption.
- We apply the same factor loading to get four factors: a "global", "soft", "real", and "labour" factor.

More on the models

- Dynamic factor models assume a lower dimensional subspace where the factors evolve like some autoregressive process:

$$
\begin{aligned}
y_t &= \lambda_1 f_{1,t} + \cdots + \lambda_r f_{r,t} + e_t \\
f_{j,t} &= a_j f_{j,t-1} + u_{j,t}
\end{aligned}
$$

- We run two types of signature model, each using a different method of reducing the dimensionality of the feature space. **In the first**, we use **PCA** to reduce the dimensions to 4 using the same factor groupings as in the DFM (with the addition of time). Then we compute signatures of the time series of principal components. **In the second**, we compute signatures of the **factors** from the DFM model.

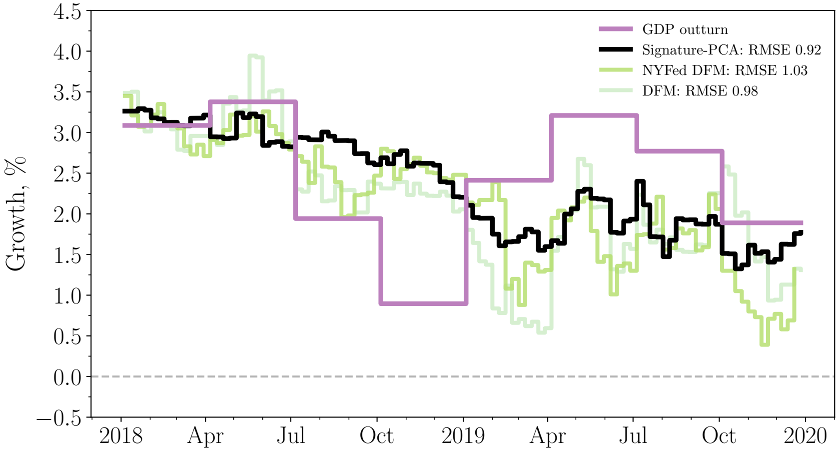
Results

	sig (PCA)	sig (DFM)	DFM	NY Fed DFM
validation period	0.57	**0.46**	0.85	0.86
test period	0.92	**0.82**	0.98	1.03

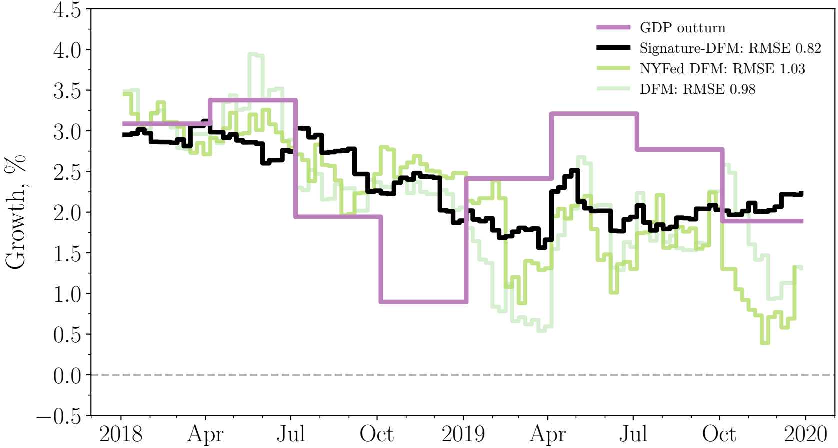
RMSE for the signature method on PCA and DFM factors, as well as our DFM and the NY Fed published values.

- The signature method outperforms the DFM and original NY Fed DFM
- Signature on dynamic factors has best performance: demonstrates that signature can extract richer features than standard DFM provides

Signature + PCA, test period. RMSE: sig (0.92), DFM (0.98), NY Fed Staff (1.03).



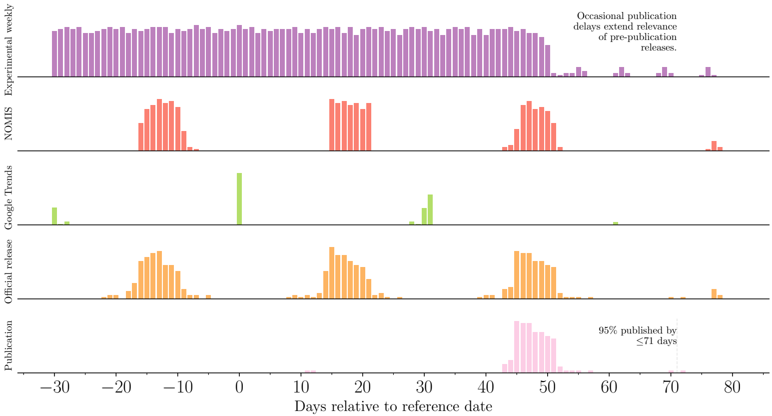
Signature on factors, test period. RMSE: sig (0.82), DFM (0.98), NY Fed Staff (1.03).



Application: Nowcasting UK unemployment through Covid

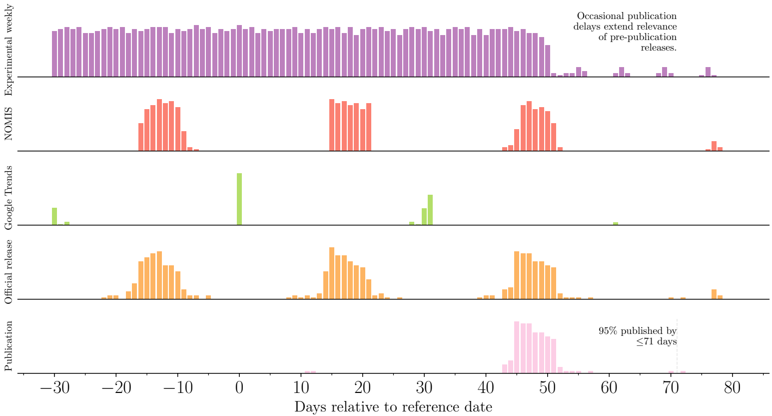
Nowcasting monthly UK unemployment

- **irregular publications**:
  delays, months vs weeks



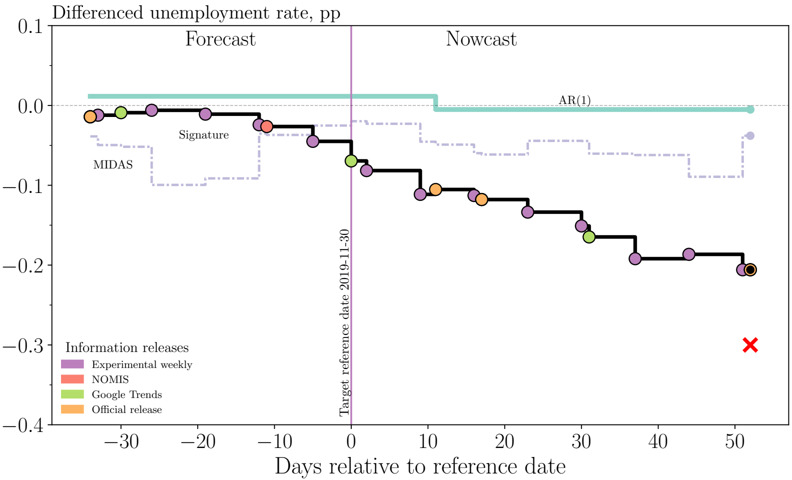
Nowcasting monthly UK unemployment

- **irregular publications**: delays, months vs weeks
- **mixed frequency data**:
  - monthly: Google Trends "unemployment benefit" and "seekers allowance", NOMIS claimant count
  - weekly: highly experimental labour market statistics on activity, inactivity, employment, and unemployment



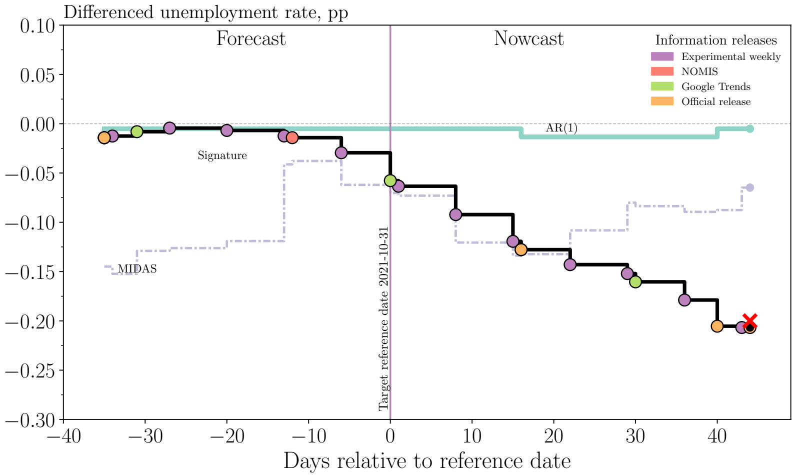
Out-of-sample results in “normal” times

- November 2019
- somewhat challenging: significant movement in unemployment

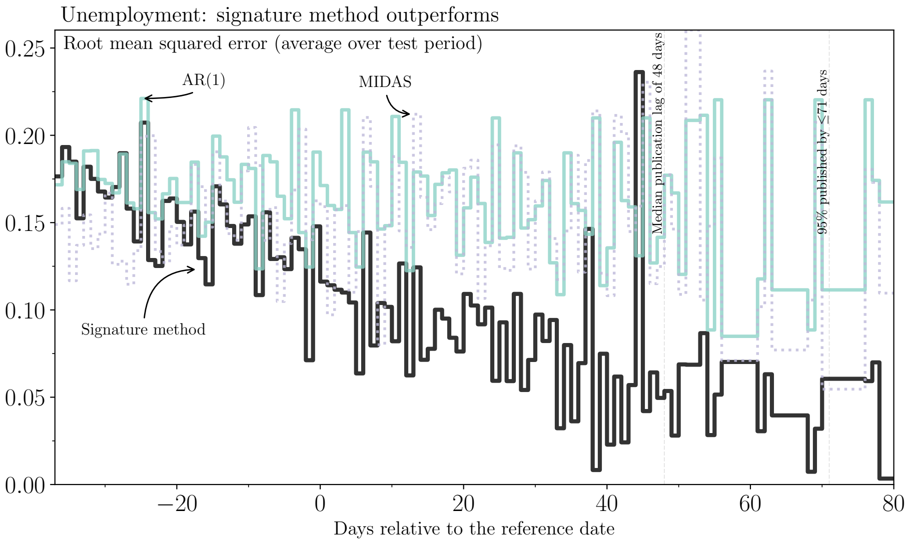


Out-of-sample results in Covid

- October 2021
- challenging: UK government's furlough (aka job retention) scheme ended the previous month



On average and out-of-sample, signatures outperform



Conclusions

Conclusions

- **Regression on signatures is a new class of nowcasting model**
- comfortable with noisy and difficult data: mixed frequencies, irregularity, missing data, non-linearities
- Regression on signatures subsumes the linear Kalman filter
- Nowcasting with signatures achieves **near optimal performance in simulations**
- Signatures have **strong performance in real-world applications**