Advances in Machine Learning for Economic Forecasting

Eleni Kalamara
King's College London

George Kapetanios
King's College London

Motivation

- Recent surge in data collection *(Big Data era)*.
  - Different types of *Big* data: **textual data**, financial transactions, selected internet searches, surveys.
- Big Data may be able to aid in improving economic forecasts.
- Traditional forecasting tools cannot handle the size and complexity inherent in Big Data.
- Econometricians have refined numerous techniques from different disciplines to digest the ever-growing amount of data, avoid overfitting and improve forecast accuracy.
  $\longrightarrow$ e.g. Factor models.
- But many issues remain. An important issue is that macroeconomic forecasts need time series models that possibly allow for structural change and that side of big data methods remains relatively underdeveloped.

A Time Series Model for Unstructured Data

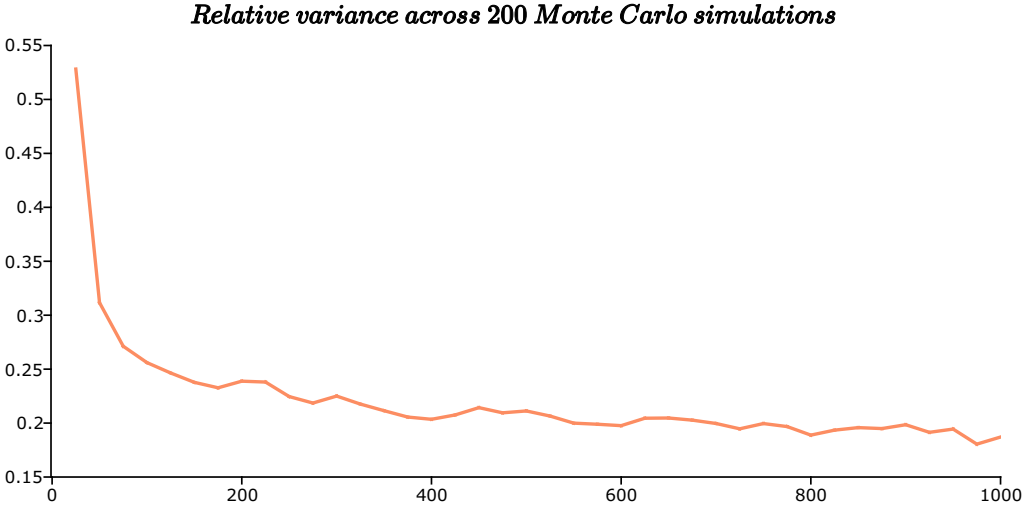
A simple model

- Recently researchers use big unstructured datasets to improve inference on unobserved variables and forecasting.
- Eg payroll data to improve unemployment analysis.
- But they construct summaries of the big dataset rather than use all of it.
- Let

$$X_i = F + \epsilon_i, i = 1, ..., N$$

  where X_i are observed, F and ϵ_i are unobserved, $F \sim niid(0, \sigma_f^2)$, and $\epsilon_i \sim niid(0, \sigma_i^2)$. We are interested in $Var(F|X_1, ..., X_N)$. Is $Var(F|\bar{X})$, $\bar{X} = \frac{1}{N}\sum_i X_i$ a good enough alternative?
- Yes but only under restrictive assumptions - $\sigma_i^2 = \sigma^2$ for all i. ▸ see details
- We suggest a state space model that uses the full big dataset.

Ratio $Var(F|X_1, ..., X_N)$ / $Var(F|\bar{X})$



Factor Analysis: a (very) general overview

$$\underset{N\times 1}{X_t} = \Lambda \underset{r\times 1}{F_t} + \epsilon_t \qquad t = 1 \dots T$$

$$\begin{bmatrix} u_t* \\ \epsilon_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_r & 0 \\ 0 & \psi \end{bmatrix} \right)$$

- Static: X_t depends only on F_t (Stock and Watson, 2002)
- Dynamic:* Further include: $F_t = A_1 F_{t-1} + A_p F_{t-p} + u_t$. (Doz, Giannone and Reichlin, 2011, 2012; Forni et al., 2000, 2005)
- Exact: ψ is diagonal and ϵ_t time invariant
- Approximate (Some) cross-sectional & temporal dependence in ϵ_t

The model

The N dimensional balanced dataset:

$$X_t = \Lambda F_t + \xi_t \tag{1}$$

The unstructured data set:

$$Z_t = B_t F_t + \epsilon_t \tag{2}$$

- Z_t: $k_t \times 1$, $z_t = (z_{1t}, \dots z_{k_t t})'$, where $k_t >> T$ ‣ see example
- **Importantly,** k_t has a time-varying dimension. There can be a different number of events at every period and each event can be represented by a vector of different dimension. (*e.g number of newspaper articles or employees in a payroll, at each t*)

$$F_t = CF_{t-1} + \eta_t \tag{3}$$

State Space Form

Define:

$$Y_t = \begin{pmatrix} X_t \\ Z_t \end{pmatrix} = \begin{pmatrix} \Lambda \\ B_t \end{pmatrix} F_t + \begin{pmatrix} \zeta_t \\ \epsilon_t \end{pmatrix} \tag{1}$$

And re-write:

$$Y_t = \boldsymbol{\Lambda_{0,t}} F_t + \boldsymbol{\zeta_t} \qquad \text{Measurement eq}$$

$$F_t = CF_{t-1} + \eta_t \qquad \text{Transition eq.}$$

where $\boldsymbol{\Lambda_{0,t}} = (\Lambda, B_t)\prime$.

Model characteristics - extensions

- Deal with missing values in non-synchronous way.
- Model Estimation: Kalman Filter and Maximum Likelihood.
- The model can represent a variety of features of the unstructured data: squares and other higher moments.
- X_t with ragged edges.
- The model can accommodate mixed frequencies; X_t can follow a lower frequency and Z_t a higher one.
- Enable nowcasting and forecasting at both high and low frequency, extracting a high frequency factor.

A Monte Carlo Simulation

Different Specifications for DGP for: $Z_t = B_t F_t + \epsilon_t$:

- indiosyncratic components ξ_t, ϵ_t: both cross-sectionally and temporally independent. **(exact model)**
- $\epsilon_t \sim N(0, \Sigma_{\epsilon_t})$, Σ_{ϵ_t} = $\sigma_{it}^2 * I_{max(k_t)}$, $\sigma_{it}^2 \in \mathbf{U}(1, 3)$.
- $u_t \sim N(0, I_n)$.
- Assume: $r = 1$, $n = 1$, $\Lambda_0 = \{1, \ldots 1\}$
- factor DGP: $C = \beta * I_r$, $\beta = \{0.5\}$
- $T = 50, 100, 200,$
- $\max k_t = 10, 50, 100, 500, 1000$

Comparators

Model 1: Not include z_t (standard factor model), i.e:

$$X_t = \Lambda F_t + \xi_t$$

Model 2:

$$Y_t^* = \begin{pmatrix} X_t \\ Z_t^* \end{pmatrix} = \begin{pmatrix} \Lambda \\ B_t^* \end{pmatrix} F_t + \begin{pmatrix} \xi_t \\ \epsilon_t^* \end{pmatrix}$$

where:

- $Z_t^* = \sum_{k=1}^{k_t} Z_t / k_t$ - average of the unstructured data set at each point in time t.
- $Var(\epsilon_t^*)$ = $\frac{\bar{\sigma}_{i,t}^2}{\max k_t}$

Keep the same factor structure

Average of Relative RMSEs over 200 Monte Carlo simulations

True Parameters : $\beta = 0.5, \sigma_i^2 \in \mathbf{U}(1,3)$

		Model 1					Model 2				
		max(k_t)									
		10	50	100	500	1000	10	50	100	500	1000
T	50	0.666	0.215	0.190	0.076	0.096	0.339	0.307	0.306	0.193	0.103
	100	0.662	0.222	0.266	0.168	0.123	0.995	0.362	0.421	0.229	0.167
	200	0.280	0.243	0.261	0.181	0.102	0.461	0.386	0.409	0.246	0.139

Table: Average of relative RMSE of the HSS over Model 1 and Model 2 respectively.
Model 1: does not include unstructured dataset (Z_t), **Model 2**: includes the average of Z_t
.

Empirical application - Forecasting Inflation using newspaper articles scores

- Really many papers on forecasting with factor models.
- **The unstructured data set** Z_t: Let M be the maximum number of articles that appeared monthly. That said, let z_t^k, be a $k_t \times 1$ vector of sentiment scores where k is the number of articles for each point period t and $k_t = 1 \ldots M$. This implies that for s, where $k < s < M$ the observations of z_t^k are missing. ▸ see example
- We estimate the high dimensional state space model to extract a factor using sentiment/uncertainty scores extracted from newspaper articles[1] ▸ see text methods .
- **Benchmark:** an FADL-type model of the form:

$$\hat{x_{t+h}} = \hat{\alpha} + \hat{\beta} x_t + \sum_j \hat{\gamma}_j \cdot \chi_{jt}$$

where χ_{jt} : macro/fin factors (Redl, 2017).

[1] The *sentiment* on each article is measured using a dictionary based method

A selection of Forecast Results for Inflation

Text Model	h = 3	h= 6	h= 9
Loughran_sentiment	0.828 **	0.823 ***	0.850***
Harvard_sentiment	0.831 ***	0.813 ***	0.850 ***
vader_sentiment	0.853***	0.856***	0.864 ***
stability_sentiment	0.874	0.824***	0.845 ***
opinion_sentiment	0.885***	0.932	0.930
tf_idf_econom	0.889 *	0.865 ***	0.906 *
Nyman_sentiment	0.933 ***	0.964***	0.972
economcounts	0.938**	0.933**	0.965 **
tf_idf_uncert	0.939	0.951	0.934
alexopoulos_09	0.964	0.953	0.973 ***
uncertaincounts	0.973	0.951	0.971
Afinn_sentiment	0.985	1.004	1.001
baker_bloom_davis	1.001	0.967	0.973
husted	1.001	0.979	0.983

Table: relative RMSEs, based on the estimated factor using a text method.* Denotes rejection at the 10% level, ** at the 5% level and *** 1 (D-M test)

A selection of Forecast Results for monthly GDP growth

Text Model	h = 3	h= 6	h= 9
Harvard_sentiment	0.861**	0.745	0.688
Loughran_sentiment	0.881	0.803	0.754
economcounts	0.905	0.852	0.829
opinion_sentiment	0.910	0.855	0.824
stability_sentiment	0.922	0.865	0.835
uncertaincounts	0.925	0.893	0.883
tf_idf_econom	0.926	0.871	0.845
vader_sentiment	0.929*	0.863	0.809
alexopoulos_09	0.943	0.919	0.913
tf_idf_uncert	0.957	0.922	0.919
Nyman_sentiment	0.960	0.923	0.893
Afinn_sentiment	0.975	0.970	0.967
husted	0.979	0.972	0.986
baker_bloom_davis	0.983	0.978	0.980

Table: relative RMSEs, based on the estimated factor using a text method.* Denotes rejection at the 10% level, ** at the 5% level and *** 1 (D-M test)

Time Variation in Machine Learning Models

Idea

- Extend Machine Learning models to fit to the applied time series setting and account for structural breaks using the kernel based approach of Giraitis, Kapetanios and Yates (2014).
- Focus on the regression-like tools because they are the most natural for macroeconomic applications. In particular, examine the **support vector regressor (SVR)** (Vapnik, 1998) and **neural nets** (Friedman, Hastie and Tibshirani, 2001) and propose a ground theoretical framework to allow for structural breaks.

Time-varying neural nets

A general definition of a multi-layer (deep) neural network follows: Let $\mathbf{x} = (x_1, \ldots, x_p)\prime$ be the input vector.
Let $h_1 \ldots h_L$ be vectors of activation functions ‣ see types for each of the L (hidden) layers of the network representing non-linear transformations of the data. Denote by g_l the l-layer which is a vector of functions of length equal to the number of J_l nodes in that layer, such that $g_0 = \mathbf{x}$. The overall structure of the network is equal to:

$$G = g_L(g_{L-1}(\ldots(g_0(.))$$

where:

$$g_l(\mathbf{x}) = W_{1,l} h_l \left(W_{2,l}\mathbf{x} + \mathbf{b}_l\right) \qquad \forall 1 \leq l \leq L$$

and $W_{1,l}$, $W_{2,l}$ and $\mathbf{b}_l$ are matrices and vectors of weight parameters.

Time-varying neural nets

The model can then be written as (Friedman, Hastie and Tibshirani, 2001):

$$y_t = G\left(x_t, \beta^0\right) + \varepsilon_t, t = 1, ..., T \tag{2}$$

where x_t is $p \times 1$, β^0 is $k \times 1$ and contains all model parameters and G denotes the overall nonlinear mapping. We estimate this model by penalised least squares, i.e.

$$\hat{\beta} = \arg\min_{\beta} \frac{\|\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{X}, \beta)\|_2^2}{T} + \lambda \|\boldsymbol{\beta}\|_1 .$$

where $\boldsymbol{y} = (y_1, ..., y_T)'$ and $\boldsymbol{G}(\boldsymbol{X}, \beta) = (G(x_1, \beta), ..., G(x_T, \beta))'$

Time-varying neural nets

Let the model now be extended to the case

$$y_t = G\left(\boldsymbol{x}_t, \beta_t^0\right) + \varepsilon_t \tag{3}$$

where β_t^0 is a persistent, bounded, possibly stochastic process and:

$$\left\|\beta_t^0 - \beta_s^0\right\| \leq C \frac{|t-s|}{\min(t,s)} \sup_{s \leq h \leq t} \left\|\beta_h^0\right\|, \text{ for some } C > 0. \tag{4}$$

We estimate this model by time varying penalised least squares, i.e.

$$\hat{\beta}_t = \arg\min_{\beta} \frac{\|\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{X}, \beta)\|_{w_t,2}^2}{T} + \lambda \|\beta\|_1 .$$

where

$$\|\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{X}, \beta)\|_{w_t,2}^2 = \sum_{j=1}^{T} w_{t,j} \left(y_j - G\left(\boldsymbol{x}_j, \beta\right)\right)^2 ,$$

and $w_{t,j} = K\left(\frac{t-j}{H}\right)$, for some kernel function K and bandwidth $H = o(T)$, $H \to \infty$.

Time-varying neural nets

- There is considerable work in the above setting where β_t^0 is allowed to be stochastic.
- In a series of papers, Giraitis, Kapetanios et al (2014 (JoE), 2015 (JAE), 2018 (JTSA)) show that kernel based estimation of β_t^0, in many contexts (regression, ML) is consistent and asymptotically normal, even if β_t^0 is stochastic and satisfies a smoothness condition.
- Recently Dendramis, Giraitis and Kapetanios (2019) extend this to a large dimensional setting, providing sharp probability exponential inequalities for weighted, randomly scaled sums of mixing and possibly fat-tailed data that allows time varying estimation of large covariance matrices.

Time-varying neural nets

Theorem

Let model (3) with condition (4) hold and let (i) ε_t be a martingale difference process that is independent of $\boldsymbol{x}_t$ and (ii) G be a function with bounded first derivatives. Then, for all t,

$$\frac{\left\| \boldsymbol{G}\left(\boldsymbol{X}, \hat{\beta}_t\right) - \boldsymbol{G}\left(\boldsymbol{X}, \beta^{(T),0}\right) \right\|_{w_t,2}^2}{T} = O_p\left(\left(\frac{\log k}{H}\right)^{1/2} \sup_t \left\| \beta_t^0 \right\|_1\right). \tag{5}$$

where $\beta^{(T),0} = (\beta_1^{0\prime}, \ldots, \beta_T^{0\prime})'$ *and* $\boldsymbol{G}\left(\boldsymbol{X}, \beta^{(T),0}\right) = (G\left(x_1, \beta_1^0\right), \ldots, G\left(x_T, \beta_T^0\right))'$

Time - varying SVR

- Let $x_t = [x_{1t}, \dots x_{Nt}]'$ be the vector of covariates and y_t be the target variable.

$$y_t = \beta^{0\prime} \boldsymbol{x}_t + \varepsilon_t. \tag{6}$$

- Recall that SVR operates by solving (Vapnik et al. (1997), pg. 156-158) :

$$\hat{\beta} = \min_{\beta} \|\beta\|^2 \text{, s.t.}$$

$$y_t - \beta' \boldsymbol{x}_t \leq \epsilon + \xi_t, \quad \beta' \boldsymbol{x}_t - y_t \leq \epsilon + \xi_t^*, \quad \xi_t^*, \xi_t \geq 0$$

where ϵ denotes a preselected error margin parameter (tuning parameter) and ξ_t^*, ξ_t are slack variables.

Time - varying SVR

Dual Formulation of the problem:

$$\max_{\boldsymbol{\alpha},\boldsymbol{\alpha}^*}\left[-\sum_{i,j=1}^{T}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)\boldsymbol{x}_i'\boldsymbol{x}_j-\epsilon\sum_{i=1}^{T}\left(\alpha_i+\alpha_i^*\right)+\sum_{i=1}^{T}\left(\alpha_i+\alpha_i^*\right)y_i\right]$$

$$\text{s.t.}\sum_{i=1}^{T}\left(\alpha_i-\alpha_i^*\right)=0 \text{ and } \alpha_i,\alpha_i^*\geq 0.$$

Then,

$$\hat{\boldsymbol{\beta}}_t=\sum_{i=1}^{T}\left(\alpha_{ti}-\alpha_{ti}^*\right)\boldsymbol{x}_i$$

The value of parameter ϵ defines a margin of tolerance where no penalty is given to the errors. Thus, the formulation of the problem can be viewed as a penalised optimisation procedure where there is a positive value which controls the penalty imposed on observations that lies outside the ϵ and helps to prevent overfitting (Steinwart and Christmann, 2008).

Time - varying SVR

Following Giraitis, Kapetanios and Yates (2014); Kapetanios and Zikes (2018), we incorporate weights $w_{t,j}$ where $w_{t,j} = K\left(\frac{t-j}{H}\right)$, for some kernel function K which is centered on the time-point of interest and decays for more distant observations and bandwidth $H = o(T)$, $H \to \infty$, the optimization problem becomes:

$$\max_{\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_t^*} \Bigg[-\sum_{i,j=1}^{T} w_{t,j} w_{t,i} \left(\alpha_i - \alpha_i^*\right) \left(\alpha_j - \alpha_j^*\right) \boldsymbol{x}_i' \boldsymbol{x}_j - \epsilon \sum_{i=1}^{T} w_{t,i} \left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{T} w_{t,i} \left(\alpha_i + \alpha_i^*\right) y_i \Bigg] \tag{7}$$

$$\text{s.t.} \sum_{i=1}^{T} \left(\alpha_i - \alpha_i^*\right) = 0 \text{ and } \alpha_i, \alpha_i^* \geq 0, \sum_{i=1}^{T} w_{t,i} = T$$

Then,

$$\hat{\beta}_t = \sum_{i=1}^{T} w_{t,i} \left(\alpha_{ti} - \alpha_{ti}^*\right) \boldsymbol{x}_i.$$

Empirics - Forecasting US Industry portfolios

Dataset

- **Targets:** Three Standard industry portfolios of U.S. equities. [2]
  - Cnsm: Consumer Durables, NonDurables, Wholesale, Retail.
  - Manuf: Manufacturing, Energy, Utilities.
  - HiTech: Business Equipment, Telephone, Television Transmission.
- **Predictors:** "Zoo factors" of Feng, Giglio and Xiu (2019) [3].
- **Full sample length**: 1976m1 - 2017m10: Evaluation period starts in 2002.

[2] Focus on portfolios rather than individual assets because they have more stable betas, higher signal-to-noise ratios, and are less prone to missing data issues. Data from Kenneth French's website.

[3] 150 risk factors at the monthly frequency for the period from July 1976 to December 2017

Forecasting US industry portfolios

Table: Out of sample relative RMSEs using time-varying ML and standard ML model: Benchmark: AR(1)

	Steps Ahead	(1)	(3)	(6)	(9)	(12)
Cnsm	TVSVM	0.750	0.735	0.734	0.733	0.829
	SVM	0.868	0.844	0.830	0.839	0.831
	TVNN	0.940	1.000	1.000	0.890	0.983
	NN	1.106	1.030	1.090	0.925	1.050
Manuf	TVSVM	0.684	0.864	0.800	0.605	0.798
	SVM	0.892	0.871	0.809	0.811	0.793
	TVNN	1.035	1.068	0.947	0.984	0.887
	NN	0.968	1.299	0.991	0.944	0.936

Forecasting US industry portfolios

Table: Out of sample relative RMSEs using time-varying ML and standard ML model: Benchmark: AR(1)

	Steps Ahead	(1)	(3)	(6)	(9)	(12)
HiTech	TVSVM	0.608	0.604	0.805	0.809	0.813
	SVM	0.829	0.811	0.803	0.822	0.821
	TVNN	0.982	0.987	0.997	0.964	0.908
	NN	0.926	0.927	0.926	0.944	0.983

Empirics 2 - UK GDP growth

- We use the time-varying ML models to forecast monthly UK GDP at $h = 1, 3, 6, 9, 12$ step ahead using a large panel of survey indicators.
- Full sample length: 2000m1 - 2018m8
- Reserve 92 months for forecast evaluation (post crisis)
- Derive direct forecasts generated by the time-varying neural nets and time-varying support vector regressions and compare with standard AR(1). For comparison, forecasts are also derived using the standard neural nets and support vector regressors under the same specifications.

Empirical Application - GDP growth

In-sample period: 2000m1-2009m12					
Model	h = 1	h = 3	h= 6	h= 9	h= 12
TVNN	0.750***	0.698*	0.742*	0.840	0.825*
NN	0.843	0.721	0.778	0.843	0.900***
TVSVR	0.945	0.925	0.887	0.915	0.987
SVR	0.716	0.646***	0.664***	0.696*	0.764

Table: Average RMSEs at h=1,3,6,9,12 for the time-varying and the standard ML models relative to the AR(1). *, **, *** are the results from Diebold and Mariano (1995) test with Harvey's (1997) adjustment for predictive accuracy.* Denotes rejection at the 10% level, ** at the 5% level and *** 1% level.

Thank you

Thank you

Appendix

Example of spreadsheet of an unstructured dataset

-1.04	-9.75	-5.78	0.89	NaN	NaN	NaN	NaN	NaN
-0.28	-7.79	-1.54	-0.70	NaN	NaN	NaN	NaN	NaN
0.10	-0.05	0.57	0.47	0.54	-0.96	-0.10	0.32	-0.08
3.12	4.22	-0.93	1.45	-0.15	0.19	NaN	NaN	NaN
-0.43	0.43	-2.42	2.45	-0.22	0.80	NaN	NaN	NaN
0.06	-13.71	0.33	3.13	0.57	0.09	1.20	0.57	0.96
0.09	2.34	0.51	3.05	0.32	0.04	NaN	NaN	NaN
0.12	-13.15	0.64	1.70	-0.07	-1.47	NaN	NaN	NaN
-0.53	-8.77	-2.92	1.03	0.28	-1.10	0.30	0.87	0.24
-1.27	-8.21	-7.04	0.38	0.25	-0.75	NaN	NaN	NaN
-1.53	-7.94	-8.52	1.36	0.29	-0.74	NaN	NaN	NaN
-1.10	-9.46	-6.12	2.13	-0.06	-1.23	0.50	0.70	0.4
-0.69	2.48	-3.82	NaN	-0.24	4.70	NaN	NaN	NaN
-0.93	NaN	-5.17	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

▸ Go back

Text methods: Algorithm-based metrics

Positive and negative dictionary	Boolean	Computer science-based
Financial stability (Correa et al., 2017)	Economic Uncertainty (Alexopoulos, Cohen et al., 2009)	VADER sentiment (Gilbert, 2014)
Finance oriented (Loughran and McDonald, 2013)	Monetary policy uncertainty (Husted, Rogers and Sun, 2017)	'Opinion' sentiment (Hu et al., 2017; Hu and Liu, 2004)
Afinn sentiment (Nielsen, 2011)	Economic Policy Uncertainty (Baker, Bloom and Davis, 2016)	Punctuation economy (this paper: See details)
Harvard IV (used in Tetlock (2007))		
Anxiety-excitement (Nyman et al., 2018)		
Single word counts of "uncertain" and "econom"		
tf-idf applied to "uncertain" and "econom"		

Go back

Average of Relative RMSEs over 500 Monte Carlo simulations

True Parameters : $\sigma = 1, q = 0.9$						
	Model 1			**Model 2**		
T \ **max**k_t	100	500	1000	100	500	1000
50	0.928	0.750	0.640	0.928	0.757	0.635
100	0.920	0.741	0.642	0.914	0.742	0.640
400	0.921	0.742	0.635	0.920	0.755	0.640
1000	0.914	0.742	0.6357	0.921	0.748	0.647

Table: Average of RMSEs of the High Dimensional state space relative to the comparator models. Model1: not include unstructured dataset (Z_t), Model2: includes the average of Z_t

Popular activation functions

- Logistic function: $h(x) = \frac{1}{1+\exp -x}$
- hyperbolic tangent (tanh) function: $h(x) = \tanh(\mathrm{x})$
- Rectified Linear Unit (RELU) function: $h(x) = \max(0, x)$

go back

Conditional Variances

$Var(F|X_1, \dots X_N) = \Sigma_{FF} - \Sigma_{FX}\Sigma_{XX}^{-1}\Sigma_{XF}$. Given that:

$Var(X_i) = 1 + \sigma_i^2$, $\quad Var(\bar{X}) = 1 + \frac{1}{N^2}\sum_{i=1}^{N}\sigma_i^2$, $\quad Cov(X_i, F) = 1$, $\quad Cov(\bar{X}, F) = 1$, we have

$$Var(F|\bar{X}) = 1 - \frac{1}{1 + \frac{1}{N^2}\sum_{i=1}^{N}\sigma_i^2},$$

And applying the Sherman-Morrison formula :

$$Var(F|X_1, ..., X_N) = 1 - \sum_{i=1}^{N}\frac{1}{\sigma_i^2} + \frac{\sum_{i=1}^{N}\sum_{j=1}^{n}\frac{1}{\sigma_i^2\sigma_j^2}}{1 + \sum_{i=1}^{N}\frac{1}{\sigma_i^2}}.$$

go back

Conditional Variances

It holds that: $Var(F|\bar{X}) \geq Var(f|X_1, X_N)$,i.e:,

$$\sum_{i=1}^{N} \frac{1}{\sigma_i^2} - \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\sigma_i^2 \sigma_j^2}}{1 + \sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \geq \frac{1}{1 + \frac{1}{N^2} \sum_{i=1}^{N} \sigma_i^2}$$

If $\sigma_i^2 = \sigma^2$, for all $i \in N$,

$$\frac{N}{\sigma^2} - \frac{\frac{N^2}{\sigma^4}}{1 + \frac{N}{\sigma^2}} - \frac{1}{1 + \frac{\sigma^2}{N}} \geq 0 \tag{8}$$

We set $\alpha = \frac{N}{\sigma^2}$. Then:

$$\alpha - \frac{\alpha^2}{1 + \alpha} - \frac{1}{1 + \frac{1}{\alpha}} \geq 0$$

or

$$\alpha + \alpha^2 + 1 + \alpha - \alpha^2 - \alpha - 1 - \alpha \geq 0$$

But $\alpha + \alpha^2 + 1 + \alpha - \alpha^2 - \alpha - 1 - \alpha = 0$ ▸ go back

References I

Alexopoulos, Michelle, Jon Cohen, et al. 2009. "Uncertain times, uncertain measures." *University of Toronto Department of Economics Working Paper*, 352.

Baker, Scott R, Nicholas Bloom, and Steven J Davis. 2016. "Measuring economic policy uncertainty." *The Quarterly Journal of Economics*, 131(4): 1593–1636.

Correa, Ricardo, Keshav Garud, Juan M Londono, Nathan Mislang, et al. 2017. "Constructing a Dictionary for Financial Stability." Board of Governors of the Federal Reserve System (US).

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

Doz, Catherine, Domenico Giannone, and Lucrezia Reichlin. 2012. "A quasi–maximum likelihood approach for large, approximate dynamic factor models." *Review of economics and statistics*, 94(4): 1014–1024.

References II

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu. 2019. "Taming the factor zoo: A test of new factors." National Bureau of Economic Research.

Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. 2000. "The generalized dynamic-factor model: Identification and estimation." *Review of Economics and statistics*, 82(4): 540–554.

Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. 2005. "The generalized dynamic factor model: one-sided estimation and forecasting." *Journal of the American Statistical Association*, 100(471): 830–840.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. *The elements of statistical learning*. Vol. 1, Springer series in statistics New York.

Gilbert, CJ Hutto Eric. 2014. "Vader: A parsimonious rule-based model for sentiment analysis of social media text."

Giraitis, Liudas, George Kapetanios, and Tony Yates. 2014. "Inference on stochastic time-varying coefficient models." *Journal of Econometrics*, 179(1): 46–65.

References III

Hu, Guoning, Preeti Bhargava, Saul Fuhrmann, Sarah Ellinger, and Nemanja Spasojevic. 2017. "Analyzing Users' Sentiment Towards Popular Consumer Industries and Brands on Twitter." 381–388, IEEE.

Hu, Minqing, and Bing Liu. 2004. "Mining and summarizing customer reviews." 168–177, ACM.

Husted, Lucas F., John Rogers, and Bo Sun. 2017. "Monetary Policy Uncertainty." Board of Governors of the Federal Reserve System (U.S.) International Finance Discussion Papers 1215.

Kapetanios, George, and Filip Zikes. 2018. "Time-varying Lasso." *Economics Letters*, 169: 1–6.

Loughran, Tim, and Bill McDonald. 2013. "IPO first-day returns, offer price revisions, volatility, and form S-1 language." *Journal of Financial Economics*, 109(2): 307–326.

Nielsen, Finn Årup. 2011. "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs." *arXiv preprint arXiv:1103.2903*.

References IV

Nyman, Rickard, Sujit Kapadia, David Tuckett, David Gregory, Paul Ormerod, and Robert Smith. 2018. "News and narratives in financial systems: exploiting big data for systemic risk assessment." *Bank of England Staff Working Papers*, 704.

Steinwart, Ingo, and Andreas Christmann. 2008. *Support vector machines*. Springer Science & Business Media.

Stock, James H, and Mark W Watson. 2002. "Forecasting using principal components from a large number of predictors." *Journal of the American statistical association*, 97(460): 1167–1179.

Tetlock, Paul C. 2007. "Giving content to investor sentiment: The role of media in the stock market." *The Journal of finance*, 62(3): 1139–1168.

Vapnik, Vladimir, Christopher JC Burges, Linda Kaufman, Alex J Smola, and Harris Drucker. 1997. "Support vector regression machines." 155–161.