JMLR: Workshop and Conference Proceedings 80:1–10, 2017 ACML 2017

Supplement : Learning Predictive Leading Indicators for Forecasting Time Series Systems with Unknown Clusters of Forecast Tasks

Magda Gregorová[1,2] MAGDA.GREGOROVA@HESGE.CH
Alexandros Kalousis[1,2] ALEXANDROS.KALOUSIS@HESGE.CH
Stéphane Marchand-Maillet[2] STEPHANE.MARCHAND-MAILLET@UNIGE.CH
[1] *Geneva School of Business Administration, HES-SO University of Applied Sciences of Western Switzerland;* [2] *University of Geneva, Switzerland*

Editors: Yung-Kyun Noh and Min-Ling Zhang

Appendix A. Experimental data and transformations

Table 1 lists the measurement sites of the Water Service of the US Geological Survey (http://www.usgs.gov/) whose data we use in the USGS experiments in section 5.2 of the main text. The original data are the daily averages of the physical discharge in cubic feet per second (parameter code 00060) downloaded from the USGS database on 9/9/2016. We have used the data up to 31/12/2014 and before modelling transformed them by taking the year-on-year log-differences.

Table 1: Measurement sites for the river-flow data

Code	Description
06191500	Yellowstone River at Corwin Springs MT
06192500	Yellowstone River near Livingston MT
06214500	Yellowstone River at Billings MT
06295000	Yellowstone River at Forsyth MT
06309000	Yellowstone River at Miles City MT
06327500	Yellowstone River at Glendive MT
06329500	Yellowstone River near Sidney MT
01129200	CONNECTICUT R BELOW INDIAN STREAM NR PITTSBURG, NH
01129500	CONNECTICUT RIVER AT NORTH STRATFORD, NH
01131500	CONNECTICUT RIVER NEAR DALTON, NH
01138500	CONNECTICUT RIVER AT WELLS RIVER, VT
01144500	CONNECTICUT RIVER AT WEST LEBANON, NH
01154500	CONNECTICUT RIVER AT NORTH WALPOLE, NH
01170500	CONNECTICUT RIVER AT MONTAGUE CITY, MA
01184000	CONNECTICUT RIVER AT THOMPSONVILLE, CT

Table 3 lists the macro-economic indicators of Stock and Watson (2012) used in our economic experiment in section 5.2 in the main text. Before using for modelling we have applied the same pre-processing steps as in Stock and Watson (2012).

We have

- transformed the monthly data to quarterly by taking the quarterly averages (column Q in table 3);
- applied the stationarizing transformations described in table 2 (column T in table 3);
- cleaned the data from outliers by replacing observations with absolute deviations from median larger than 6 times the interquartile range by the median of the 5 preceding values.

Table 2: Stationarizing transformations

T	Transformation
1	$y_t = z_t$
2	$y_t = z_t - z_{t-1}$
3	$y_t = (z_t - z_{t-1}) - (z_{t-1} - z_{t-2})$
4	$y_t = \log(z_t)$
5	$y_t = \ln(z_t / z_{t-1})$
6	$y_t = \ln(z_t / z_{t-1}) - \ln(z_{t-1} / z_{t-2})$

z_t is the original data, y_t is the transformed series

Appendix B. Experimental results

This section provides further details on experimental results not included in the main text due to space limitation.

B.1. Synthetic experiments

Fig. 1 shows the synthesis of the model parameter matrices $\mathbf{W}$ for the six synthetic experimental designs. The displayed structures correspond to the schema of the $\mathbf{W}$ matrix presented in Fig. 1 of the main text. For the figure, the matrices were binarised to simply indicate the existence (1) or non-existence (0) of a G-causal link. The white-to-black shading reflects the number of experimental replications in which this binary indicator is active (equal to 1). So, a black element in the matrix means that this G-causal link was learned in all the 20 re-samples of the generating process. White means no G-causality in any of the re-samples. Though none of the sparse method was able to clearly and systematically recover the true structures, VARL1 and VARLG clearly suffer from more numerous and more frequent over-selections than CLVAR which matches the true structures more closely and with higher selection stability (fewer light-shaded elements). The 4th experimental set-up is included in the main text as Fig. 4.

Fig. 2 summarises the scaling properties of the CLVAR method with increasing increasing sample size T and the number of time series K. In each experiment, we selected a single hyper-parameter combination (near the optimal) and measured the time in seconds (on a

Table 3: Macro-economic data and transformations

Code	Q	T	Description
GDP251	Q	5	real gross domestic product, quantity index (2000=100) , saar
CPIAUCSL	M	6	cpi all items (sa) fred
FYFF	M	2	interest rate: federal funds (effective) (% per annum,nsa)
PSCCOMR	M	5	real spot mrkt price idx:bls & crb: all commod(1967=100)
FMRNBA	M	3	depository inst reserves:nonborrowed,adj res req chgs(mil$,sa)
FMRRA	M	6	depository inst reserves:total,adj for reserve req chgs(mil$,sa)
FM2	M	6	money stock:m2 (bil$,sa)
GDP252	Q	5	real personal consumpt expend, quantity idx (2000=100) , saar
IPS10	M	5	industrial production index - total index
UTL11	M	1	capacity utilization - manufacturing (sic)
LHUR	M	2	unemployment rate: all workers, 16 years & over (% ,sa)
HSFR	M	4	housing starts:nonfarm(1947-58),total farm& nonfarm(1959-)
PWFSA	M	6	producer price index: finished goods (82=100,sa)
GDP273	Q	6	personal consumption expenditures, price idx (2000=100) , saar
CES275R	M	5	real avg hrly earnings, prod wrkrs, nonfarm - goods-producing
FM1	M	6	money stock: m1(bil$,sa)
FSPIN	M	5	s& p's common stock price index: industrials (1941-43=10)
FYGT10	M	2	interest rate: u.s.treasury const matur,10-yr.(% per ann,nsa)
EXRUS	M	5	united states,effective exchange rate(merm)(index no.)
CES002	M	5	employees, nonfarm - total private

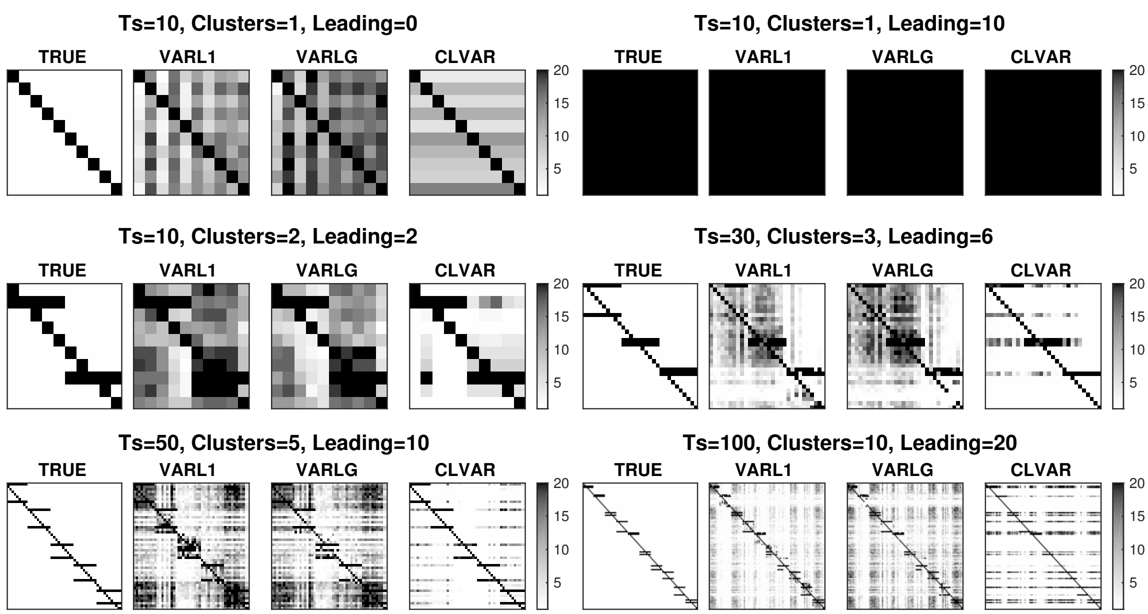

Figure 1: Synthesis of model parameters $\mathbf{W}$

single Intel(R) Xeon(R) CPU E5-2680 v2 @ 2.80GHz) and the number of iterations needed till the convergence of the objective (with 10^{-5} tolerance) for the 20 data re-samples. We used the ℓ_2 regularised solution as a warm start. The empirical results correspond to the theoretical complexity analysis of section 3.2 in the main text. For an experimental set-up with fixed number of series K (and G-causal structure), the run-time typically grows fairly slowly with the sample sizes T. However, the increases are much more important when moving to larger experiments, with higher K and more complicated structures. Here the growth in run-time is accompanied by higher number of iterations. From our experimental set-up it is difficult to separate the effect of enlarging the time-series systems in terms of higher K from the effect of more complicated structures in terms of higher number of clusters and leading indicators. In reality, we expect these to go hand-in-hand so in this sense our empirical analysis complements the theoretical asymptotic complexity analysis of section 3.2 of the main text

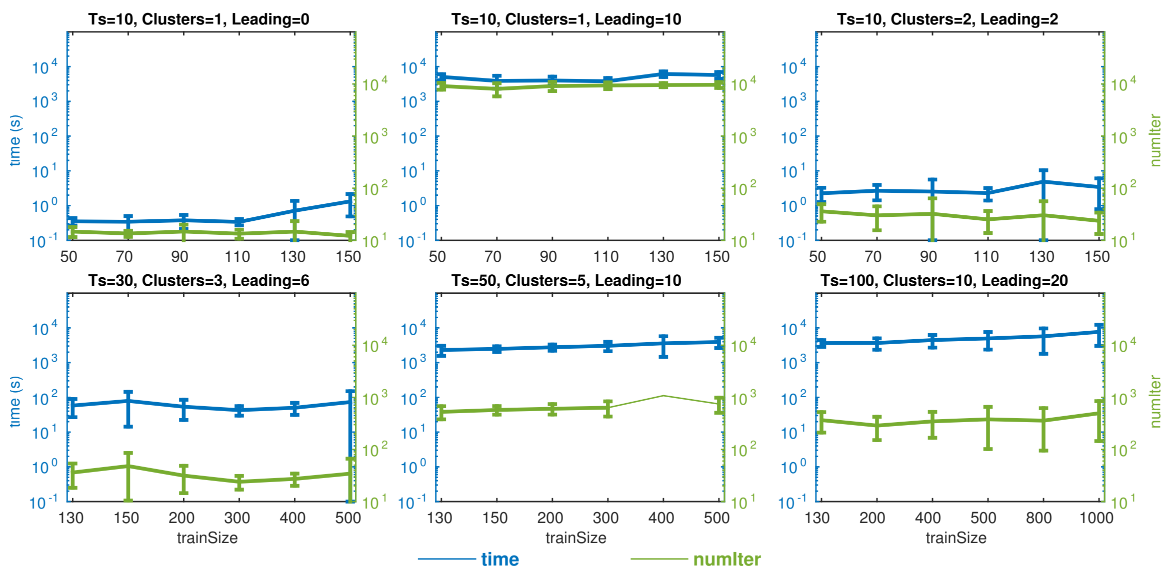

Figure 2: Runtime and number of iterations

Table 4 provides the numerical data behind the plots of Fig. 3(a) in the main text. The predictive accuracy is measured by mean squared error (MSE) of 1-step-ahead forecasts relative to the forecasts produced by the VAR with the true generative coefficients (the irreducible error). The relative MSE is averaged over the 500 hold-out points (the models are fixed and the forecasts are produced by sliding forward over the dataset). The *avg* and *std* are the average and standard deviation calculated over the 20 re-samples of the data for each experimental design.

Table 4: Synthetic experiments: Relative MSE over true model

trainSize	stat	AR	VARL2	VARL1	VARLG	CLVAR
Ts=10, Clusters=1, Leading=0						
50	avg	1.225	3.113	1.974	2.513	1.772
	std	0.102	0.637	0.310	0.515	0.643
70	avg	1.143	2.570	1.503	1.868	1.395
	std	0.063	0.474	0.162	0.304	0.328
90	avg	1.098	2.186	1.324	1.545	1.219
	std	0.043	0.353	0.125	0.203	0.082
110	avg	1.071	1.926	1.232	1.368	1.160
	std	0.030	0.268	0.086	0.133	0.055
130	avg	1.052	1.745	1.172	1.267	1.121
	std	0.020	0.230	0.062	0.096	0.038
150	avg	1.043	1.617	1.143	1.214	1.102
	std	0.017	0.178	0.050	0.076	0.034
Ts=10, Clusters=1, Leading=10						
50	avg	12.564	23.449	42.347	42.055	22.319
	std	5.258	18.273	32.893	32.115	15.458
70	avg	12.649	12.712	9.610	9.634	11.692
	std	4.507	4.565	3.681	3.472	4.769
90	avg	12.281	4.905	4.943	4.897	4.934
	std	4.697	1.545	1.513	1.547	1.584
110	avg	12.583	3.569	3.565	3.562	3.509
	std	4.316	0.940	0.939	0.940	0.917
130	avg	12.028	2.527	2.523	2.519	2.477
	std	3.762	0.480	0.479	0.477	0.486
150	avg	11.637	2.195	2.194	2.189	2.158
	std	3.157	0.417	0.417	0.414	0.433
Ts=10, Clusters=2, Leading=2						
50	avg	1.429	2.730	1.743	2.124	1.674
	std	0.167	0.565	0.246	0.357	0.308
70	avg	1.336	2.296	1.522	1.689	1.405
	std	0.096	0.346	0.206	0.199	0.139
90	avg	1.295	1.998	1.398	1.499	1.289
	std	0.100	0.253	0.139	0.146	0.116
110	avg	1.258	1.824	1.320	1.396	1.199
	std	0.064	0.214	0.082	0.103	0.062
130	avg	1.237	1.686	1.283	1.347	1.156
	std	0.057	0.181	0.078	0.085	0.050
150	avg	1.229	1.566	1.255	1.300	1.133
	std	0.054	0.171	0.076	0.081	0.043

continues in next page ...

... continues from previous page

trainSize	stat	AR	VARL2	VARL1	VARLG	CLVAR
Ts=30, Clusters=3, Leading=6						
130	avg	1.532	2.424	1.355	1.490	1.261
	std	0.142	0.361	0.091	0.131	0.081
150	avg	1.512	2.245	1.319	1.414	1.225
	std	0.129	0.308	0.081	0.111	0.062
200	avg	1.471	1.861	1.257	1.289	1.172
	std	0.113	0.205	0.062	0.069	0.048
300	avg	1.451	1.586	1.164	1.224	1.100
	std	0.103	0.135	0.039	0.051	0.031
400	avg	1.438	1.375	1.119	1.151	1.073
	std	0.100	0.086	0.028	0.036	0.024
500	avg	1.434	1.298	1.097	1.116	1.059
	std	0.099	0.068	0.023	0.027	0.018
Ts=50, Clusters=5, Leading=10						
130	avg	3.327	4.376	2.024	2.160	1.746
	std	0.557	0.792	0.306	0.277	0.223
150	avg	3.254	3.920	1.801	1.915	1.577
	std	0.544	0.685	0.205	0.215	0.196
200	avg	3.200	3.203	1.583	1.620	1.324
	std	0.506	0.523	0.146	0.145	0.109
300	avg	3.161	2.280	1.347	1.392	1.128
	std	0.491	0.301	0.083	0.092	0.034
400	avg	3.123	1.844	1.228	1.299	1.091
	std	0.478	0.192	0.053	0.070	0.025
500	avg	3.097	1.622	1.172	1.214	1.081
	std	0.468	0.144	0.039	0.050	0.025
Ts=100, Clusters=10, Leading=20						
130	avg	1.890	3.362	1.518	1.807	1.734
	std	0.202	0.536	0.118	0.184	0.178
200	avg	1.845	2.855	1.350	1.478	1.415
	std	0.189	0.420	0.078	0.106	0.109
400	avg	1.801	2.196	1.225	1.268	1.180
	std	0.177	0.264	0.050	0.059	0.049
600	avg	1.783	1.820	1.137	1.191	1.108
	std	0.172	0.182	0.031	0.043	0.026
800	avg	1.777	1.644	1.104	1.129	1.084
	std	0.170	0.142	0.023	0.029	0.021
1000	avg	1.774	1.501	1.090	1.100	1.065
	std	0.170	0.115	0.020	0.022	0.019

Table 5 provides the numerical data behind the plots of Fig. 3(b) in the main text. The selection accuracy of the true G-causal links is measured by the average between the false negative and false positive rates. The *avg* and *std* are the average and standard deviation calculated over the 20 re-samples of the data for each experimental design.

Table 5: Synthetic experiments: Selection errors of true G-causal links

trainSize	stat	AR	VARL2	VARL1	VARLG	CLVAR
Ts=10, Clusters=1, Leading=0						
50	avg	0.000	0.500	0.268	0.316	0.151
	std	0.000	0.000	0.062	0.072	0.153
70	avg	0.000	0.500	0.245	0.318	0.119
	std	0.000	0.000	0.027	0.056	0.094
90	avg	0.000	0.500	0.207	0.301	0.118
	std	0.000	0.000	0.027	0.026	0.043
110	avg	0.000	0.500	0.172	0.285	0.122
	std	0.000	0.000	0.026	0.027	0.034
130	avg	0.000	0.500	0.141	0.268	0.138
	std	0.000	0.000	0.017	0.029	0.036
150	avg	0.000	0.500	0.120	0.245	0.143
	std	0.000	0.000	0.019	0.027	0.041
Ts=10, Clusters=1, Leading=10						
50	avg	0.450	0.000	0.027	0.031	0.043
	std	0.000	0.000	0.084	0.095	0.112
70	avg	0.450	0.000	0.013	0.013	0.010
	std	0.000	0.000	0.054	0.058	0.026
90	avg	0.450	0.000	0.001	0.000	0.017
	std	0.000	0.000	0.005	0.000	0.042
110	avg	0.450	0.000	0.000	0.000	0.000
	std	0.000	0.000	0.000	0.000	0.000
130	avg	0.450	0.000	0.000	0.000	0.000
	std	0.000	0.000	0.000	0.000	0.000
150	avg	0.450	0.000	0.000	0.000	0.000
	std	0.000	0.000	0.000	0.000	0.000
Ts=10, Clusters=2, Leading=2						
50	avg	0.222	0.500	0.190	0.268	0.168
	std	0.000	0.000	0.041	0.041	0.103
70	avg	0.222	0.500	0.179	0.248	0.096
	std	0.000	0.000	0.072	0.039	0.061
90	avg	0.222	0.500	0.159	0.222	0.094
	std	0.000	0.000	0.067	0.030	0.051
110	avg	0.222	0.500	0.155	0.200	0.058
	std	0.000	0.000	0.086	0.031	0.044

continues in next page ...

... continues from previous page

trainSize	stat	AR	VARL2	VARL1	VARLG	CLVAR
130	avg	0.222	0.500	0.192	0.205	0.067
	std	0.000	0.000	0.107	0.076	0.054
150	avg	0.222	0.500	0.210	0.196	0.054
	std	0.000	0.000	0.118	0.064	0.034
Ts=30, Clusters=3, Leading=6						
130	avg	0.321	0.500	0.175	0.204	0.173
	std	0.000	0.000	0.015	0.017	0.037
150	avg	0.321	0.500	0.165	0.192	0.176
	std	0.000	0.000	0.016	0.018	0.027
200	avg	0.321	0.500	0.178	0.183	0.165
	std	0.000	0.000	0.046	0.014	0.029
300	avg	0.321	0.500	0.225	0.166	0.107
	std	0.000	0.000	0.012	0.027	0.034
400	avg	0.321	0.500	0.202	0.242	0.098
	std	0.000	0.000	0.013	0.012	0.029
500	avg	0.321	0.500	0.184	0.220	0.095
	std	0.000	0.000	0.009	0.013	0.022
Ts=50, Clusters=5, Leading=10						
130	avg	0.321	0.500	0.226	0.226	0.123
	std	0.000	0.000	0.040	0.010	0.032
150	avg	0.321	0.500	0.211	0.217	0.110
	std	0.000	0.000	0.039	0.011	0.035
200	avg	0.321	0.500	0.212	0.196	0.077
	std	0.000	0.000	0.047	0.009	0.027
300	avg	0.321	0.500	0.257	0.169	0.038
	std	0.000	0.000	0.011	0.021	0.008
400	avg	0.321	0.500	0.236	0.183	0.034
	std	0.000	0.000	0.012	0.047	0.011
500	avg	0.321	0.500	0.224	0.218	0.033
	std	0.000	0.000	0.011	0.033	0.011
Ts=100, Clusters=10, Leading=20						
130	avg	0.321	0.500	0.120	0.164	0.260
	std	0.000	0.000	0.006	0.009	0.054
200	avg	0.321	0.500	0.098	0.132	0.159
	std	0.000	0.000	0.006	0.006	0.028
400	avg	0.321	0.500	0.158	0.094	0.121
	std	0.000	0.000	0.006	0.006	0.016
600	avg	0.321	0.500	0.129	0.175	0.104
	std	0.000	0.000	0.004	0.005	0.011
800	avg	0.321	0.500	0.110	0.154	0.100
	std	0.000	0.000	0.004	0.005	0.011
1000	avg	0.321	0.500	0.096	0.137	0.090

continues in next page ...

... continues from previous page

trainSize	stat	AR	VARL2	VARL1	VARLG	CLVAR
	std	0.000	0.000	0.004	0.004	0.013

B.2. Real-data experiments

Table 6 provides the numerical data behind the plots of Fig. 5(a) in the main text. The predictive accuracy is measured by mean squared error (MSE) of 1-step-ahead forecasts relative to the forecasts produced random walk model (= uses the last observed value as the 1-step-ahead forecast). The relative MSE is averaged over the 30 and 300 hold-out points for the Economic and the USGS dataset respectively (the models are fixed and the forecasts are produced by sliding forward over the dataset). The *avg* and *std* are the average and standard deviation calculated over the 20 re-samples of the data for each experimental dataset.

Table 6: Real-data experiments: Relative MSE over random walk

trainSize	stat	AR	VARL2	VARL1	VARLG	CLVAR
Economic Ts=20						
50	avg	0.413	0.573	0.498	0.498	0.436
	std	0.035	0.026	0.016	0.018	0.036
70	avg	0.419	0.489	0.456	0.470	0.409
	std	0.028	0.025	0.041	0.027	0.025
90	avg	0.400	0.455	0.420	0.442	0.392
	std	0.025	0.024	0.023	0.021	0.025
110	avg	0.384	0.424	0.379	0.395	0.370
	std	0.022	0.024	0.022	0.029	0.025
130	avg	0.380	0.406	0.368	0.368	0.367
	std	0.023	0.026	0.022	0.026	0.024
USGS Ts=17						
200	avg	0.912	1.857	1.222	1.222	0.980
	std	0.013	0.098	0.058	0.046	0.069
300	avg	0.881	1.320	0.876	0.922	0.746
	std	0.003	0.040	0.012	0.019	0.035
400	avg	0.858	0.934	0.760	0.756	0.708
	std	0.001	0.030	0.026	0.012	0.019
500	avg	0.855	0.855	0.754	0.714	0.675
	std	0.003	0.021	0.011	0.007	0.015
600	avg	0.862	0.858	0.747	0.729	0.679
	std	0.002	0.004	0.004	0.020	0.024

Table 7 provides the numerical data behind the plots of Fig. 5(b) in the main text. The sparsity of the learned models is measured by the proportion of active edges in the learned G-causality graph. The *avg* and *std* are the average and standard deviation calculated over the 20 re-samples of the data for each experimental dataset.

Table 7: Real-data experiments: proportion of G-causal graph edges

trainSize	stat	AR	VARL2	VARL1	VARLG	CLVAR
Economic Ts=20						
50	avg	0.050	1.000	0.127	0.207	0.115
	std	0.000	0.000	0.092	0.012	0.024
70	avg	0.050	1.000	0.394	0.172	0.110
	std	0.000	0.000	0.177	0.012	0.026
90	avg	0.050	1.000	0.505	0.166	0.138
	std	0.000	0.000	0.100	0.011	0.053
110	avg	0.050	1.000	0.502	0.412	0.182
	std	0.000	0.000	0.009	0.204	0.040
130	avg	0.050	1.000	0.507	0.570	0.199
	std	0.000	0.000	0.007	0.096	0.042
USGS Ts=17						
200	avg	0.059	1.000	0.618	0.465	0.363
	std	0.000	0.000	0.117	0.061	0.057
300	avg	0.059	1.000	0.581	0.641	0.369
	std	0.000	0.000	0.018	0.010	0.081
400	avg	0.059	1.000	0.629	0.625	0.412
	std	0.000	0.000	0.143	0.005	0.060
500	avg	0.059	1.000	0.823	0.640	0.431
	std	0.000	0.000	0.061	0.009	0.057
600	avg	0.059	1.000	0.801	0.632	0.450
	std	0.000	0.000	0.014	0.090	0.020

References

James H. Stock and Mark W. Watson. Generalized Shrinkage Methods for Forecasting Using Many Predictors. *Journal of Business & Economic Statistics*, 2012.