Combining Factor Forecasts and Regularization Methods

Marcel Stechert, Marcelo Avila and Maik Haberer | Free University Berlin

Motivation 1

- We examine inflation forecasts by using estimated factors from a large number of predictors to reduce the dimension of the data (Stock and Watson 2002, JASA; Stock and Watson 2002, JBES). The large macro data is available from McCracken and Ng (2006, JBES) click to follow link

Drawbacks with many macro time series:

1. Macro variables are highly cross-correlated, implying that more data are not always better (Boivin and Ng 2006, JOE)
2. Only small fraction of macro variables have explanatory power for inflation

- Reduce noise through elastic net (EN) regressions to eliminate irrelevant predictors (Bai and Ng 2008, JOE; Ng 2013, HOEF) → sharpen factors!
- We expect factors, estimated from pre-selected (targeted) data, to perform better in pseudo-out-of-sample forecasting exercises.
- Main research questions:
  1. How to use EN for variable selection in combination with factor model forecasts?
  2. Do factor-augmented models (targeted vs. untargeted predictors) perform better in forecasting?

Factor Model 2

- Following Stock and Watson (2002, JASA) we assume a **factor model structure** for the N-dimensional data matrix X_t ($N \times 1$):

$$X_t = \Lambda \mathcal{F}_t + e_t \,,$$

- where $\mathcal{F}_t$ ($r \times 1$) are the latent factors, Λ ($N \times r$) contains the factor loadings, e_t ($N \times 1$) are idiosyncratic disturbances and X has size $T \times N$.
- The **factor-augmented prediction equation** then is:

$$y_{t+h} = \beta_F' \mathcal{F}_t + \beta_\omega' \omega_t + \epsilon_{t+h} \implies \mathbb{E}_t\{y_{t+h}\} = \hat{y}_{t+h|t} = \hat{\beta}'_F \hat{\mathcal{F}}_t + \hat{\beta}'_\omega \omega_t \,,$$

- where ω is a $m \times 1$ vector of observed variables, h represents the forecast horizon and ϵ_{t+h} is the forecast error. $\widehat{F}_t$, $\widehat{\beta}_F$ and $\widehat{\beta}_\omega$ are the estimated factors and coefficients. Additionally: $r << N$.

Two-Step Procedure:

- As the latent factors are unobserved, we follow a **two-step procedure**:
  1. Estimate factors consistently by (asymptotic) PCA (Stock and Watson 2002, JASA; Stock and Watson 2002, JBES; Bai and Ng 2013, JOE)
  2. Plug estimated factors, $\hat{\mathcal{F}}_t$, into the forecasting equation and predict $\hat{y}_{t+h|t}$ direct to desired horizon (see Boivin and Ng 2005, IJCB for detailed comparison)

Regularized Regressions 3

- **Lasso** (Tibshirani 1996, JRSS) minimizes the SSR plus a penalty term, where λ governs the degree of penalization.

$$\sum_{i=1}^{n}\left(y_i + \beta_0 + \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\left[\|\beta\|_1\right]$$

- For sufficiently high λ, (some) coefficients are shrunk completely to zero, but in groups of highly correlated variables, Lasso tends to select only one and disregard the others.
- **Ridge** regression (Hoerl and Kennard 1970, TEM):

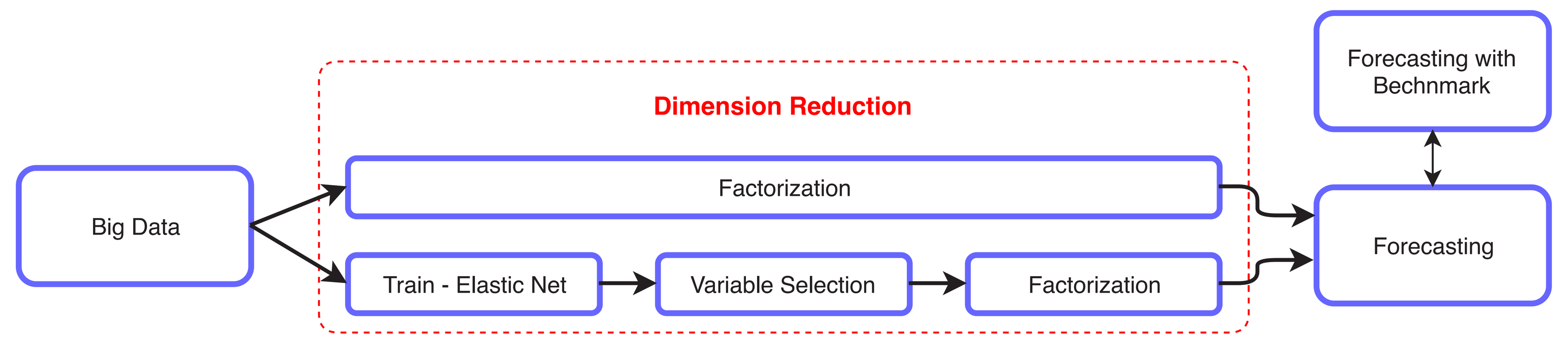
$$\sum_{i=1}^{n}\left(y_i + \beta_0 + \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\left[\tfrac{1}{2}\|\beta\|_2^2\right]$$

- Ridge approach cannot produce sparse estimates but regularization path is smooth and more stable than that of Lasso.
- Mixture of both, by controlling α ($0 \le \alpha \le 1$), gives the **Elastic Net** (Zhou and Hastie, 2005, JRSS) (Lasso: $\alpha = 1$; Ridge: $\alpha = 0$):

$$\sum_{i=1}^{n}\left(y_i + \beta_0 + \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \boldsymbol{\lambda}\left[\tfrac{1-\alpha}{2} \times \|\beta\|_2^2 + \boldsymbol{\alpha} \times \|\beta\|_1\right]$$

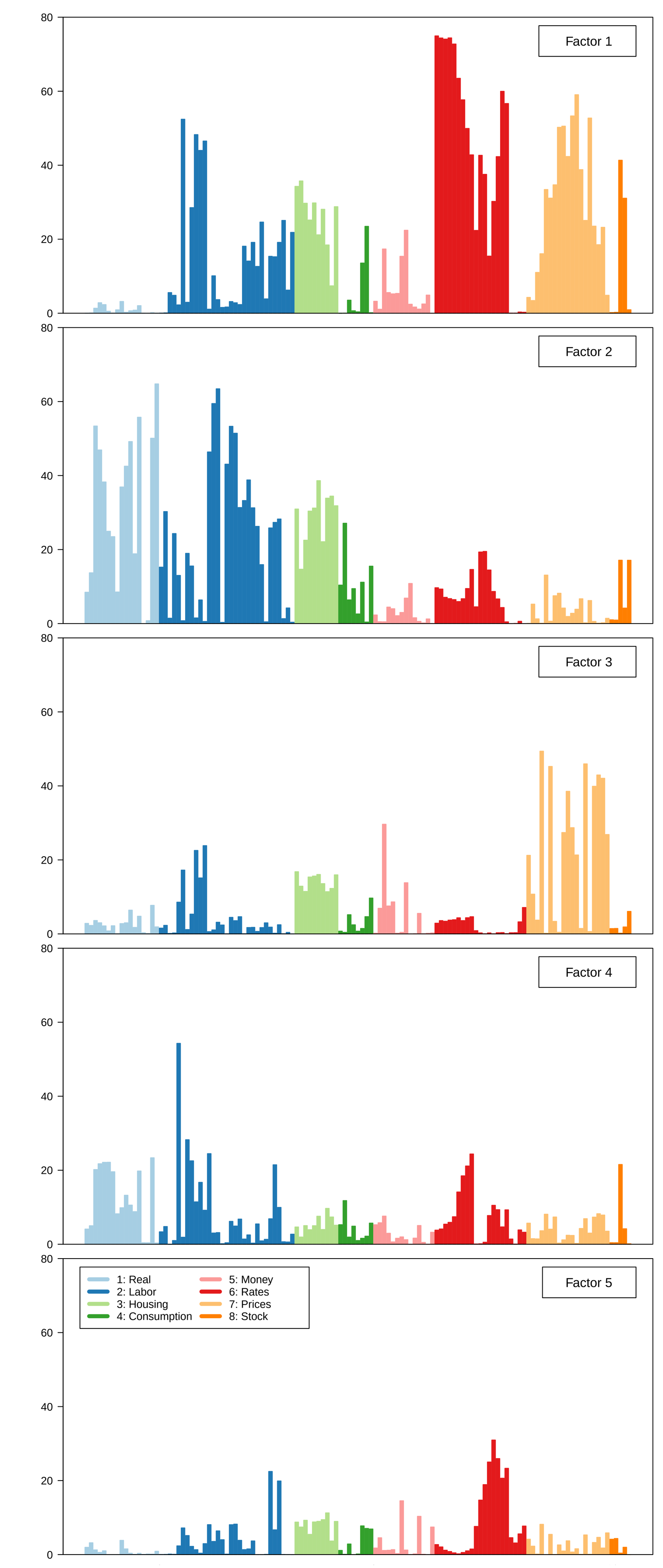
Friedman et al. 2010, JSS

Factors from Targeted vs. Untargeted Predictors: Step-by-Step 4



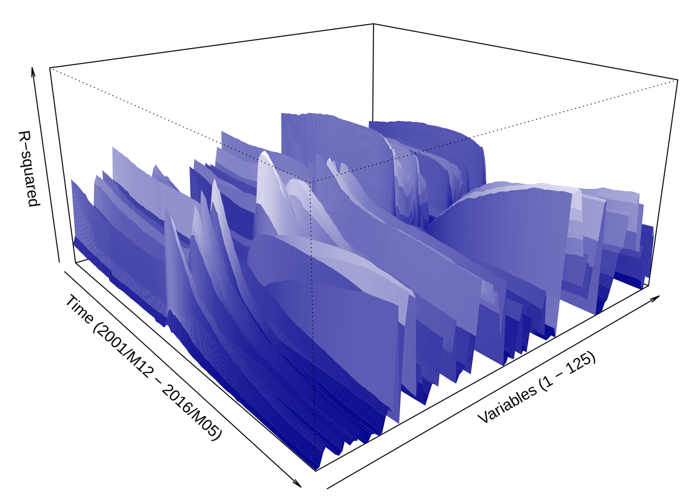
Visualised Factors 5

Marginal R-squares of $\hat{F}_{t,1}$ to $\hat{F}_{t,5}$ click to enlarge

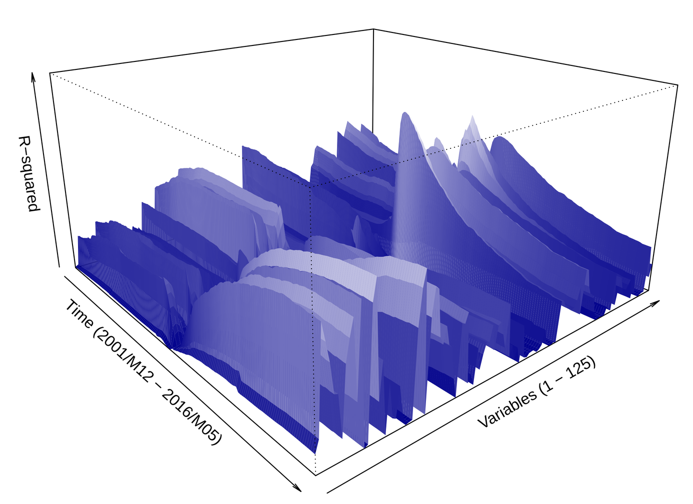


Notes: Chart shows the R-square from regressing the individual series given on the x-axis onto $\hat{F}_{t,k}$, where $k = 1, ..., 5$, seperatly. Full sample ranges from from 1971/M01 to 2016/M05.

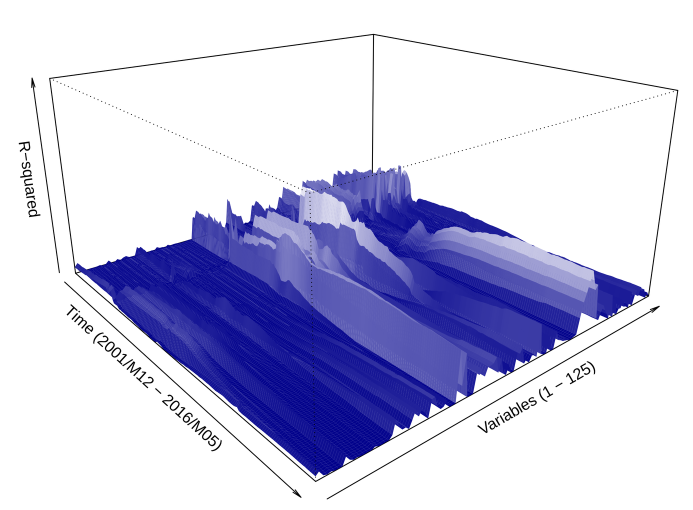
Marginal R-squares of $\hat{F}_{t,1}$ click to enlarge



Marginal R-squares of $\hat{F}_{t,2}$ click to enlarge

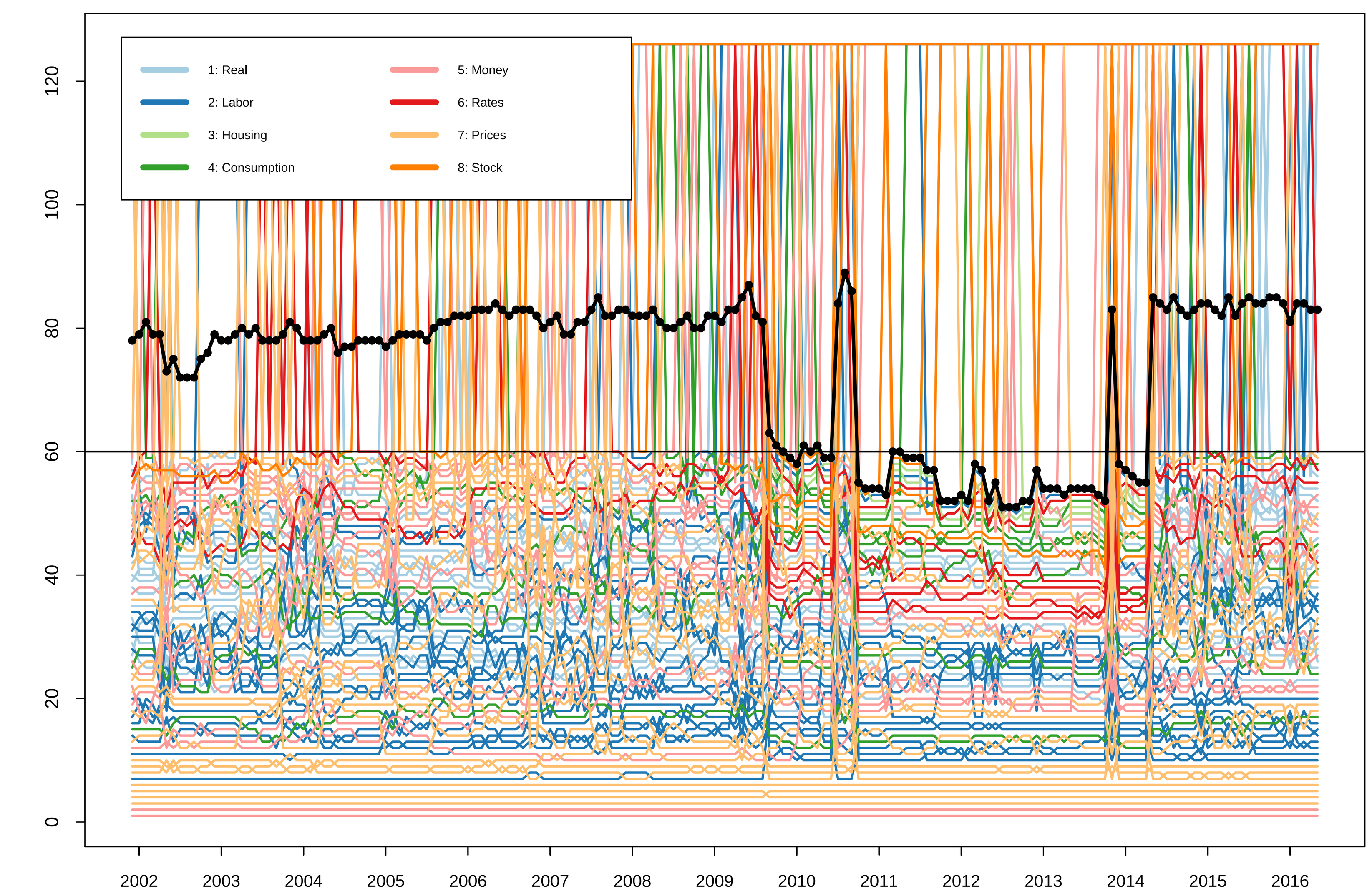


Marginal R-squares of $\hat{F}_{t,5}$ click to enlarge



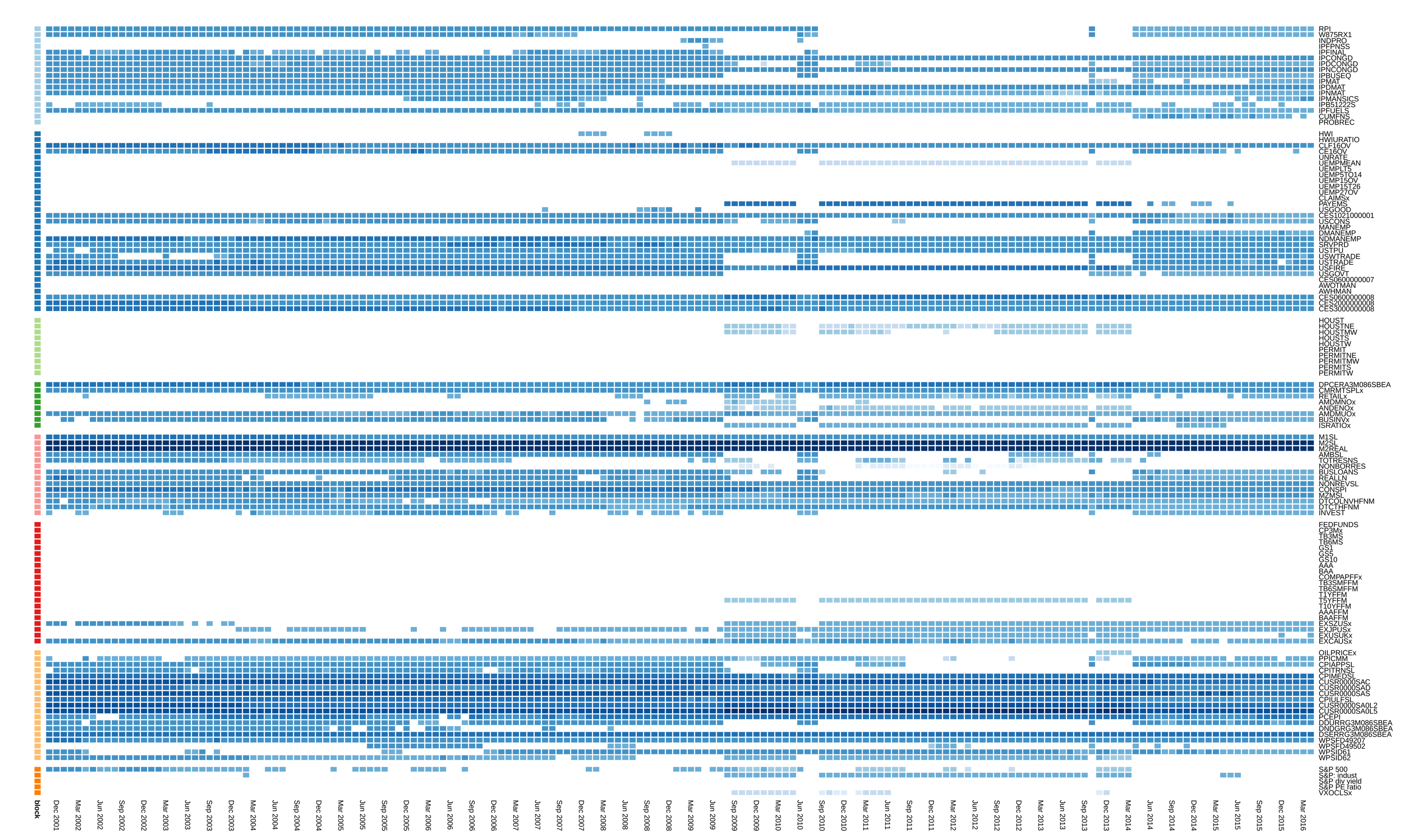
Visualised Elastic Net 6

Rank Plot of Selected Variables click to enlarge



Notes: Y-axis depicts the rank and x-axis time. The black dotted line shows the original number of variables selected (non-zero coefficients) by the EN. Colored lines represent the variables associated to the rank selected by the EN. The black horizontal line shows the ordering-threshold (here: 60 variables).

Heatmap of Selected Variables click to enlarge



Notes: This figure depicts whether a variable is selected and a give time and is relevance, measured by the the absolute value of the estimated coefficient. Light blue shows a small absolute value, dark blue a high absolute value and white implicates that the variable has no relevance. At any given time at most 60 variables are selected.

Forecasting 7

Pseudo Out-of-Sample Forecast Simulation Results (2001/M12 - 2016/M05): Relative Mean Squared Error compared to Benchmark Model

	full sample						*pre − Lehman*						*post − Lehman*					
h =	1	2	3	6	12	24	1	2	3	6	12	24	1	2	3	6	12	24
$AR(p)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$VAR(p)$	0.97	0.95	0.92	0.88	0.96	0.86	1.08	1.11	1.09	1.08	0.97	1.01	0.90	0.86	0.83	0.78	0.95	0.74
$SW_{full}(1)$	0.94	0.91	0.89	0.87	1.03	0.65	0.69	0.63	0.66	0.70	0.67	0.75	1.11	0.91	0.89	0.87	1.03	0.65
$SW_{full}(3)$	0.94	0.80	0.79	0.79	1.28	0.93	0.61	0.55	0.57	0.61	0.45	0.61	0.94	0.80	0.79	0.79	1.28	0.93
$SW_{full}(5)$	0.34	0.27	0.27	0.28	0.41	0.30	0.31	0.28	0.29	0.30	0.22	0.26	0.34	0.27	0.27	0.28	0.41	0.30
$SW_{full}(10)$	0.34	0.26	0.26	0.27	0.44	0.33	0.27	0.24	0.25	0.26	0.21	0.28	0.34	0.26	0.26	0.27	0.44	0.33
$SW_{full}(15)$	0.37	0.30	0.28	0.31	0.46	0.33	0.22	0.19	0.20	0.21	0.16	0.20	0.37	0.30	0.28	0.31	0.46	0.33
$SW_{EN}(1)$	0.92	0.89	0.86	0.85	1.00	0.63	0.67	0.61	0.64	0.68	0.65	0.73	1.08	0.89	0.86	0.85	1.00	0.63
$SW_{EN}(3)$	0.92	0.78	0.77	0.77	1.25	0.91	0.59	0.53	0.55	0.59	0.44	0.60	0.92	0.78	0.77	0.77	1.25	0.91
$SW_{EN}(5)$	0.33	0.26	0.26	0.27	0.40	0.29	0.30	0.27	0.29	0.29	0.21	0.25	0.33	0.26	0.26	0.27	0.40	0.29
$SW_{EN}(10)$	0.33	0.26	0.25	0.26	0.43	0.32	0.26	0.24	0.25	0.25	0.20	0.28	0.33	0.26	0.25	0.26	0.43	0.32
$SW_{EN}(15)$	0.36	0.29	0.28	0.30	0.45	0.33	0.21	0.19	0.20	0.20	0.15	0.19	0.36	0.29	0.28	0.30	0.45	0.33

Notes: Expanding window forecasts were used. We use the average of the forecast errors and we will refer to the ratio of the MSE for a given method to the MSE of an $AR(p)$ as RMSE (relative mean-squared error). An entry less than one indicates that the specified method is superior to the simple AR(p) forecast. For the factor model forecasts, $SW(k)$, k indicates the number of factors used. In addition to the full sample analysis, we also consider two forecast subsamples: the pre- (2002/M1 - 2007/M12) and post-Lehman (2008/M1 - 2016/M05) period. The choice of the variables for the VAR(p) was based on the idea of a reduced-form three-variable New Keynesian model and, therefore, considers inflation, industrial production and the federal funds rate. All series for all forecasting exercises were appropriately transformed prior to analysis. Cumulative variance explained by the factors over the full sample: $\hat{\mathcal{F}}_{1:1}$ (18.6%), $\hat{\mathcal{F}}_{1:2}$ (31.4%), $\hat{\mathcal{F}}_{1:3}$ (39.2%), $\hat{\mathcal{F}}_{1:4}$ (44.5%), $\hat{\mathcal{F}}_{1:5}$ (49.2%), $\hat{\mathcal{F}}_{1:6}$ (53.4%), $\hat{\mathcal{F}}_{1:7}$ (56.6%), $\hat{\mathcal{F}}_{1:8}$ (59.1%), $\hat{\mathcal{F}}_{1:9}$ (61.4%), $\hat{\mathcal{F}}_{1:10}$ (63.4%), $\hat{\mathcal{F}}_{1:11}$ (65.4%), $\hat{\mathcal{F}}_{1:12}$ (67.2%), $\hat{\mathcal{F}}_{1:13}$ (68.8%), $\hat{\mathcal{F}}_{1:14}$ (70.4%), $\hat{\mathcal{F}}_{1:15}$ (71.8%). Number of factors selected by the eigenvalue ratio test, proposed by Ahn and Hohrenstein (2013, ECTR), which suggested three factors. However, as pointed out by Ludvigson and Ng (2009, RFS), regardless of the variance-explained this does not imply that such factors beyond criteria-selected numbers do not have any significant predictive power for the variable of interest.

Discussion and Outlook 8

Short Summary:

- Allowing the number, and the composition, of possible predictor series to change with the sample through an Elastic Net variable-selection procedure seems to provide an additional flexibility, which allows the forecast model to adapt to parameter instability in the data.

- Outlook and Extensions:
  - **Forecasting**: Use real-time data (handle missing data due to ragged edges with Kalman-filter or EM-algorithm), state-space model and dynamic factor estimates and/or time-varying factor loadings, forecast different series other than inflation, think about other types of forecast evaluation (e.g. density forecast instead of focusing on point estimates), consider non-linear combinations of the factors and predictors
  - **Variable pre-selection**: Cross-validation (maybe random blockwise forecasting but still backward-looking) and training of EN, sensitivity of EN with regards to hyper-parameters and selection criteria, Ordering of variables, select predictors with respect to forecast horizon