Real-time estimation of unemployment with dynamic factor and time-varying state space models

Caterina Schiavoni [1,2] Franz Palm [2]
Stephan Smeekes [2] Jan van den Brakel [1,2]

[1]Statistics Netherlands, Heerlen

[2]Department of Quantitative Economics, Maastricht University

October 26, 2018

Outline

The Dutch Labour Force Survey (LFS)

Rotation panel design: each month a new sample enters the panel and is observed five times at quarterly intervals:

quarter

month

A B C	D E F	G H I	J K L	M N O	P Q R	} wave 1
	A B C	D E F	G H I	J K L	M N O	} wave 2
		A B C	D E F	G H I	J K L	} wave 3
			A B C	D E F	G H I	} wave 4
				A B C	D E F	} wave 5

- 5 monthly time series ($\boldsymbol{y}_t^k$) of the unemployed labour force.
- Aim: have a representative panel for the population.
- Induces Rotation Group Bias (RGB).

The labour force model

The measurement equation takes the form:

$$\boldsymbol{y}_t^k = \boldsymbol{\imath}_5 \underbrace{\theta_t^{k,y}}_{\substack{\text{common population parameter} \\ \text{i.e., the unemployment}}} + \overbrace{\begin{pmatrix} 0 \\ \boldsymbol{\lambda}_t^k \end{pmatrix}}^{\text{RGB}} + \overbrace{\boldsymbol{e}_t^k}^{\substack{\text{survey} \\ \text{errors}}}, \qquad (1)$$

where $\theta_t^{k,y} = L_t^{k,y} + S_t^{k,y}$, and k indicates that the variable is observed at the low (monthly) frequency.

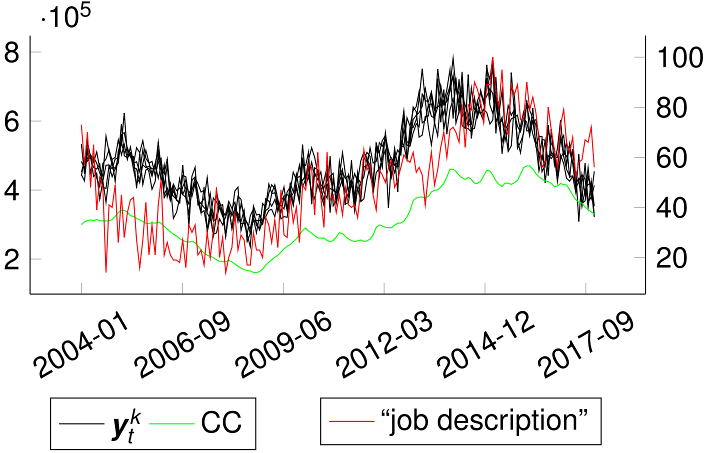
The transition equations are:

$$L_t^{k,y} = L_{t-1}^{k,y} + R_{t-1}^{k,y},$$
$$R_t^{k,y} = R_{t-1}^{k,y} + \eta_{R,t}^{k,y}, \quad \eta_{R,t}^{k,y} \sim N\left(0, \sigma_{R,y}^2\right).$$

$S_t^{k,y}$ is modeled as a trigonometric stochastic seasonal component.

Auxiliary series

We extend the model by including auxiliary series about job search behaviour from weekly Google Trends (GT) and monthly claimant counts (CC).



Two-step estimator

Problems to be addressed:

- High-dimensionality of the Google Trends $\left(\boldsymbol{x}_t^{GT}\right)$.
- Publication delays of the LFS and the CC can cause "ragged edge" datasets.

The two-step estimator by Doz et al. (2011) combines factor models with the Kalman filter and hence addresses both of these issues.

Two-step estimator (2)

State space representation of the dynamic factor model:

$$\begin{aligned} \boldsymbol{x}_t^{GT} &= \Lambda \boldsymbol{f}_t + \varepsilon_t, \qquad & \mathrm{E}(\varepsilon_t \varepsilon_t') &= \Psi, \\ \boldsymbol{f}_t &= \boldsymbol{f}_{t-1} + \boldsymbol{u}_t, \qquad & \mathrm{E}(\boldsymbol{u}_t \boldsymbol{u}_t') &= I_r, \end{aligned}$$

where $\boldsymbol{x}_t^{GT} \sim I(1)$ of dimension n, and $\boldsymbol{f}_t \sim I(1)$ of dimension r.

The I(1) setting is needed to exploit the correlation between the factors and the slope of the unobserved trend of the target variable $\boldsymbol{y}_t^k$.

The steps

$$\begin{aligned} \boldsymbol{x}_{t|\nu}^{GT} &= \hat{\Lambda}\hat{\boldsymbol{f}}_{t|\nu} + \hat{\varepsilon}_{t|\nu}, \quad \widehat{\text{cov}}(\varepsilon_{t|\nu}) = \text{diag}(\hat{\Psi}), \\ \Delta\hat{\boldsymbol{f}}_{t|\nu} &= \hat{\boldsymbol{u}}_{t|\nu}. \end{aligned} \tag{2}$$

1. Λ, $\boldsymbol{f}_t$ and Ψ are estimated by Principal Component Analysis applied to the differenced data, $\Delta\boldsymbol{x}_t^{GT}$, in a balanced dataset;

2. $\boldsymbol{f}_t$ are re-estimated with the Kalman filter applied to the approximated model on the entire dataset $\Rightarrow$ we need to condition on the information set Ω_ν: $\hat{\boldsymbol{f}}_{t|\nu} = \text{E}[\boldsymbol{f}_t|\Omega_\nu; \mathcal{M}_{(\hat{\Lambda},\hat{\Psi})}]$. ν is the time of a particular data release.

Nowcasting in a high-dimensional state space model

We combine (1) and (2) to get our final model:

$$\begin{pmatrix} \boldsymbol{y}^k_{t|\nu} \\ x^{k,CC}_{t|\nu} \\ \boldsymbol{x}^{k,GT}_{t|\nu} \end{pmatrix} = \begin{pmatrix} \imath_5 \theta^{k,y}_{t|\nu} \\ \theta^{k,CC}_{t|\nu} \\ \hat{\Lambda} \boldsymbol{f}^k_{t|\nu} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\lambda}^k_t \\ 0 \\ \boldsymbol{0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{e}^k_{t|\nu} \\ \varepsilon^{k,CC}_{t|\nu} \\ \boldsymbol{\varepsilon}^{k,GT}_{t|\nu} \end{pmatrix},$$

$$\begin{pmatrix} R^{k,y}_{t|\nu} \\ R^{k,CC}_{t|\nu} \\ \boldsymbol{f}^k_{t|\nu} \end{pmatrix} = \begin{pmatrix} R^{k,y}_{t-1|\nu} \\ R^{k,CC}_{t-1|\nu} \\ \boldsymbol{f}^k_{t-1|\nu} \end{pmatrix} + \begin{pmatrix} \eta^{k,y}_{R,t|\nu} \\ \eta^{k,CC}_{R,t|\nu} \\ \boldsymbol{u}^k_{t|\nu} \end{pmatrix}.$$

After the first step, $\boldsymbol{x}^{GT}_t$ is aggregated to the lower frequency of $\boldsymbol{y}^k_t$ and the observation equations for $\boldsymbol{y}^k_t$, $x^{k,CC}_t$ and $\boldsymbol{x}^{k,GT}_t$ are stacked together.

Nowcasting in a high-dimensional state space model

Where we get the gain from

$$\text{cov}\left(\eta^{k,y}_{R,t|\nu},\quad \eta^{k,CC}_{R,t|\nu},\quad \boldsymbol{u}'^{k}_{t|\nu}\right)' =$$

$$= \begin{bmatrix} \sigma^2_{R,y} & \rho_{CC}\sigma_{R,y}\sigma_{R,CC} & \rho_{1,GT}\sigma_{R,y} & \cdots & \rho_{r,GT}\sigma_{R,y} \\ \rho_{CC}\sigma_{R,y}\sigma_{R,CC} & \sigma^2_{R,CC} & 0 & \dots & 0 \\ \rho_{1,GT}\sigma_{R,y} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{r,GT}\sigma_{R,y} & 0 & 0 & \dots & 1 \end{bmatrix}.$$

The innovations of the trends' slopes and of the factors are allowed to be correlated. Potential gains in precision $\left(\text{MSE}\left(\hat{\theta}^{k,y}_t\right)\right)$ if $|\rho|$ is large (Harvey and Chung, 2000).

Nowcasting in a high-dimensional state space model

How we get the nowcast

$$\begin{pmatrix} \boldsymbol{y}^k_{t|\nu} \\ x^{k,CC}_{t|\nu} \\ \boldsymbol{x}^{k,GT}_{t|\nu} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\imath}_5 \theta^{k,y}_{t|\nu} \\ \theta^{k,CC}_{t|\nu} \\ \hat{\Lambda} \boldsymbol{f}^k_{t|\nu} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\lambda}^k_t \\ 0 \\ \boldsymbol{0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{e}^k_{t|\nu} \\ \varepsilon^{k,CC}_{t|\nu} \\ \boldsymbol{\varepsilon}^{k,GT}_{t|\nu} \end{pmatrix},$$

$$\begin{pmatrix} R^{k,y}_{t|\nu} \\ R^{k,CC}_{t|\nu} \\ \boldsymbol{f}^k_{t|\nu} \end{pmatrix} = \begin{pmatrix} R^{k,y}_{t-1|\nu} \\ R^{k,CC}_{t-1|\nu} \\ \boldsymbol{f}^k_{t-1|\nu} \end{pmatrix} + \begin{pmatrix} \eta^{k,y}_{R,t|\nu} \\ \eta^{k,CC}_{R,t|\nu} \\ \boldsymbol{u}^k_{t|\nu} \end{pmatrix}.$$

- Kalman filter applied to the whole state-space model.
- We can nowcast the variable of interest, providing unemployment estimates in real time before LFS data become available: $\hat{\theta}^{k,y}_{t|\nu} = \mathrm{E}[\theta^{k,y}_t | \Omega_\nu; \mathcal{M}_{(\hat{\Lambda},\hat{\Psi})}]$ is the nowcast for $\theta^{k,y}_t$.

Set-up: smooth trend model

$$\begin{pmatrix} y_t^k \\ \boldsymbol{x}_t^k \end{pmatrix} = \begin{pmatrix} L_t^k \\ \Lambda f_t^k \end{pmatrix} + \begin{pmatrix} \varepsilon_t^{k,y} \\ \boldsymbol{\varepsilon}_t^{k,x} \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_t^{k,y} \\ \boldsymbol{\varepsilon}_t^{k,x} \end{pmatrix} \sim N\left(\mathbf{0}, \text{diag}(0.5)\right),$$

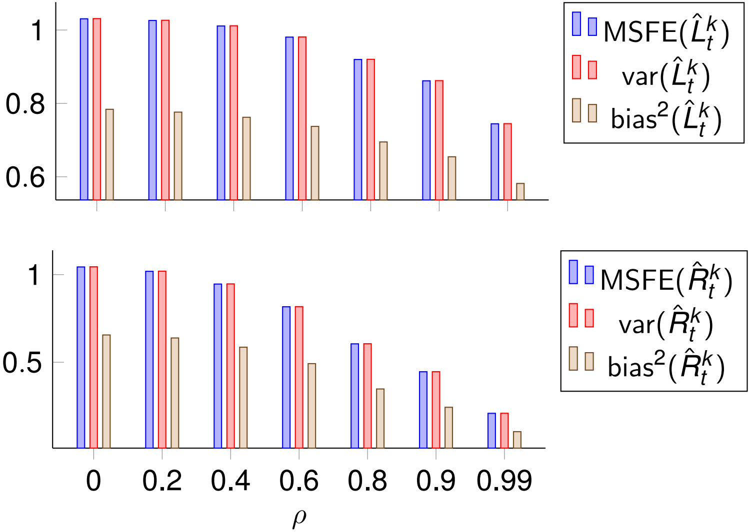
$$L_t^k = L_{t-1}^k + R_{t-1}^k,$$

$$\begin{pmatrix} R_t^k \\ f_t^k \end{pmatrix} = \begin{pmatrix} R_{t-1}^k \\ f_{t-1}^k \end{pmatrix} + \begin{pmatrix} \eta_{R,t}^k \\ u_t^k \end{pmatrix}, \quad \begin{pmatrix} \eta_{R,t}^k \\ u_t^k \end{pmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right).$$

- 1 factor f_t^k. $\Lambda = \left(\Lambda_0', \Lambda_1'\right)'$, $\underset{50\times 1}{\Lambda_0} = \mathbf{0}$, $\underset{50\times 1}{\Lambda_1} \sim U(0,1)$.
- $\boldsymbol{x}_t^k$ released at the same time without publication delays.
- y_t^k not observed at time t.
- $T = 150$ (number of months), $n = 100$, $n_{\text{sim}} = 500$.
- Recursive nowcast in the third part of the sample.

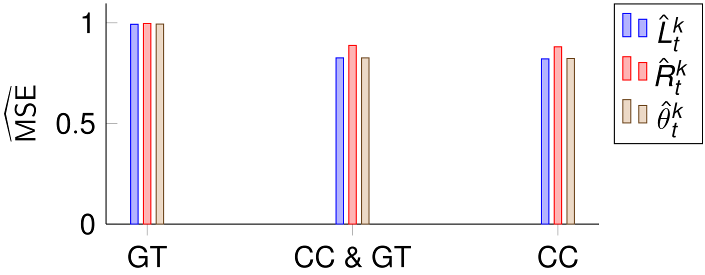
Results

Results relative to the model without auxiliary series.



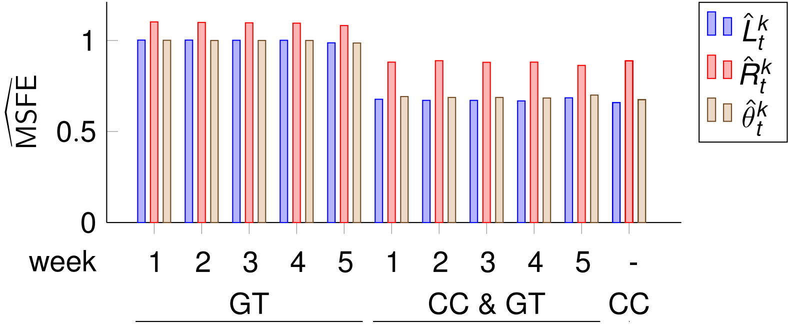
In-sample perfromance

- One factor is estimated from the Google Trends.
- $\hat{\rho}_{GT} = 0.421$, $(\hat{\rho}_{GT}, \hat{\rho}_{CC}) = (0.020, 0.905)$, $\hat{\rho}_{CC} = 0.908$.
- LR test rejects $H_0 : \rho_{CC} = 0$, but not $H_0 : \rho_{GT} = 0$.
- Results relative to the model that does not use any auxiliary series (Baseline).

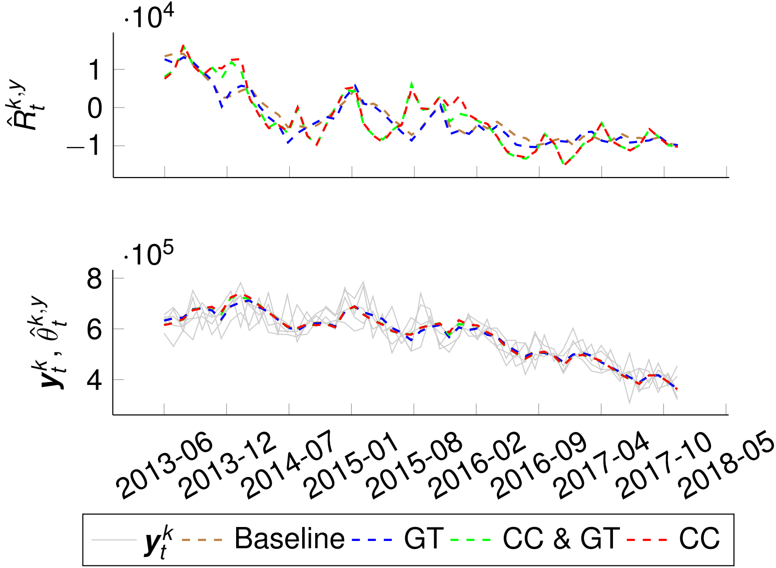


Out-of-sample performance

- Recursive nowcast in the third part of the sample.
- Results relative to the model that does not use any auxiliary series (Baseline).



Nowcast



$\cdot 10^4$
1
0
−1
$\hat{R}_t^{k,y}$
$\cdot 10^5$
8
6
4
$\boldsymbol{y}_t^k$, $\hat{\theta}_t^{k,y}$
2013-06
2013-12
2014-07
2015-01
2015-08
2016-02
2016-09
2017-04
2017-10
2018-05
$\boldsymbol{y}_t^k$
Baseline
GT
CC & GT
CC

How to improve the out-of-sample results?

- ✗ Modelling the seasonality of the factor with an ARIMA model (only better in-sample performance).
- ✗ Targeting the Google Trends first, with the Elastic Net.
- ✗ Including two factors of the Google Trends.
- ? Including the lags of the Google Trends' factor.

Time-varying correlation

The correlation between the target and the auxiliary series can be time dependent (van den Brakel and Krieg, 2016).

Two methods to estimate ρ_t:

- Generalized autoregressive score (GAS) approach.
- Kernel density-based approach.

Conclusions

- Our proposed method can yield large improvements, in terms of MSFE, in the nowcast of unobserved components, in a smooth trend model when a high-dimensional auxiliary series is included.
- The model is robust to the inclusion of auxiliary information that does not have predictive power, as it does not deteriorate the nowcast of the unobserved components.
- Google Trends of job search terms do not significantly improve the state space model used to estimate and nowcast the Dutch unemployment and its change. Yet, there is no risk in including them as auxiliary information.

Thank you!

References

Doz, C., Giannone, D., and Reichlin, L. (2011). A Two-step Estimator for Large Approximate Dynamic Factor Models Based on Kalman filtering. *Journal of Econometrics*, 164(1):188–205.

Harvey, A. and Chung, C.-H. (2000). Estimating the Underlying Change in Unemployment in the UK. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 163(3):303–309.

van den Brakel, J. A. and Krieg, S. (2016). Small Area Estimation with State Space Common Factor Models for Rotating Panels. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(3):763–791.