Unobserved Component Models with Parameter Uncertainty, Approximated Filters, and Dynamic Adaptive Mixture Models

Leopoldo Catania[1] Enzo D'Innocenzo[2] Alessandra Luati[3]

[1]Aarhus BSS and CREATES

[2]Vrije Universiteit Amsterdam

[3]University of Bologna

2021 NBER-NSF Time Series Conference
Duncan Hall, Rice University
15-16 October 2021

Linear signal plus noise

Consider the following linear signal plus noise specification:

$$y_t = \tau\mu_t + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$$
$$\mu_{t+1} = \phi\mu_t + \eta_t, \quad \eta_t \overset{iid}{\sim} (0, \varpi).$$

We compute $\mu_{t+1|t} = \mathbb{E}[\mu_{t+1}|\mathcal{F}_t]$ via the Kalman filter (KF), which, in its innovation form, is:

$$\mu_{t+1|t} = \phi\mu_{t|t-1} + k_t v_t,$$

where $v_t = y_t - \tau\mu_{t|t-1}$ is the prediction error and $k_t = (\phi P_{t|t-1}\tau)/(\tau^2 P_{t|t-1} + \sigma^2)$ is the Kalman gain.

- If ε_t and η_t are Gaussian, the KF is the minimum prediction error variance unbiased estimator (MVUE)
- If ε_t and η_t are not Gaussian, the KF is the minimum prediction error variance linear unbiased estimator (MVLUE)

Parameter uncertainty

Assume that there is uncertainty (for example) in the τ parameter, i.e.:

$$y_t = \tau_{S_t}\mu_t + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$$
$$\mu_{t+1} = \phi\mu_t + \eta_t, \quad \eta_t \overset{iid}{\sim} (0, \varpi),$$

where S_t selects, independently over time, the τ parameter between a number of alternative values, $\tau_1, \dots, \tau_J$, with probabilities $P(S_t = j) = \alpha_j$.

In this case, computing the MVUE is unfeasible because its computation requires (discrete) integration of all past possible realizations of S_t, which are J^t.

Note that the independence of the S_t variables is of no help here.

Parameter uncertainty: the MVLUE

We define the linear estimator $\mu^*_{t+1} = \beta^*_t + \kappa^*_t y_t$ for β^*_t and κ^*_t that solves:

$$\begin{cases} \mathbb{E}[\mu^*_{t+1} - \mu_{t+1} | \mathcal{F}_{t-1}] = 0 \\ \mathbb{V}[\mu^*_{t+1} - \mu_{t+1} | \mathcal{F}_{t-1}] = !\min \end{cases},$$

and find that

$$\mu^*_{t+1} = \phi \mu_{t|t-1} + \kappa^*_t v_t,$$

where $v_t = y_t - \bar{\tau}\mu_{t|t-1}$ and

$$\kappa^*_t = \frac{\phi P_{t|t-1} \bar{\tau}}{\mu^2_{t|t-1}(\upsilon - \bar{\tau}^2) + \upsilon P_{t|t-1} + \sigma^2}$$

with $\bar{\tau} = \sum_{j=1}^{J} \alpha_j \tau_j$ and $\upsilon = \sum_{j=1}^{J} \alpha_j \tau_j^2$.

When $J = 1 \Rightarrow$ no uncertainty $\Rightarrow$ the KF.

Parameter uncertainty: the MCVLUE

We introduce the minimum conditional predictive variance linear unbiased estimator (MCVLUE). Let $\tilde{\mu}_{j,t+1} = \tilde{\beta}_{j,t} + \tilde{\kappa}_{j,t} y_t$ for $j = 1, \ldots, J$ with $\tilde{\beta}_{j,t}$ and $\tilde{\kappa}_{j,t}$ defined such that

$$\begin{cases} \mathbb{E}[\mu^*_{j,t+1} - \mu_{t+1} | \mathcal{F}_{t-1}, S_t = j] = 0 \\ \mathbb{V}[\mu^*_{j,t+1} - \mu_{t+1} | \mathcal{F}_{t-1}, S_t = j] = !\min \end{cases},$$

and find

$$\tilde{\mu}_{j,t+1} = \phi \mu_{t|t-1} + \tilde{\kappa}_{j,t} v_{j,t},$$

where

$$\tilde{\kappa}_{j,t} = \frac{\phi P_{t|t-1} \tau_j}{\tau_j^2 P_{t|t-1} + \sigma^2},$$

and $v_{j,t} = y_t - \tau_j \mu_{t|t-1}$.

Parameter uncertainty: the MCVLUE

The MCVLUE is then defined as $\tilde{\mu}_{t+1} = \sum_{j=1}^{J} \xi_{j,t} \tilde{\mu}_{j,t+1}$, such that

$$\tilde{\mu}_{t+1} = \phi \mu_{t|t-1} + \sum_{j=1}^{J} \xi_{j,t} \tilde{\kappa}_{j,t} v_{j,t},$$

where $\xi_{j,t} = P(S_t = j|\mathcal{F}_t)$, for $j = 1, \ldots, J$, are filtered probabilities given by:

$$\xi_{j,t} = \frac{\alpha_j p_j(y_t|\mathcal{F}_{t-1}, S_t = j)}{\sum_{k=1}^{J} \alpha_k p_k(y_t|\mathcal{F}_{t-1}, S_t = k)},$$

where $p_j(\cdot|\mathcal{F}_{t-1}, S_t = j)$ is the density of the random variable $y_t|(\mathcal{F}_{t-1}, S_t = j)$.

I) $\mathbb{E}[\tilde{\mu}_{t+1} - \mu_{t+1}|\mathcal{F}_{t-1}] = 0$, i.e. the MCVLUE is unbiased.

II) Preliminary results suggest that when there is parameter uncertainty: $\mathbb{V}[\tilde{\mu}_{t+1} - \mu_{t+1}|\mathcal{F}_{t-1}] < \mathbb{V}[\mu^*_{t+1} - \mu_{t+1}|\mathcal{F}_{t-1}]$, i.e. MCVLUE is more efficient than MVLUE.

Parameter uncertainty: the MCVLUE

It turns out that the innovation form $\tilde{\mu}_{t+1} = \phi\mu_{t|t-1} + \sum_{j=1}^{J} \xi_{j,t}\tilde{\kappa}_{j,t}\upsilon_{j,t}$ has been implemented extensively as a result of a "collapsing step", i.e. an approximation required to avoid the marginalization of $S_{t-1}, S_{t-2}, \ldots,$ without realising that it is the MCVLUE! For example by:

- Bar-Shalom and Tse (1975) in their "probabilistic data association filters".
- Harrison and Stevens (1976) in their multi-process models.
- Gordon and Smith (1990) in their extension of the linear dynamic model.
- Shumway and Stoffer (1991) due to the wrong conclusion about $p_j(y_t|\mathcal{F}_{t-1}, S_t = j)$ being Gaussian in the model with parameter uncertainty.
- Kim (1994) and Kim and Nelson (1999), in order to make their filter operable.

MCVLUE is a Dyanmic Adaptive Mixture Model (DAMM)

The DAMM model of Catania (2019) is an observation driven model that postulates an innovation form for the dynamic parameters in the first place and a distributional assumption of the kind $y_t|(\mathcal{F}_{t-1}, S_t) \sim D(\theta_{S_t,t})$. The recursion for $\theta_{S_t,t}$ is a score driven one, Creal et al. (2013) and Harvey (2013):

$$\theta_{j,t+1} = \omega_j + \kappa_j \xi_{j,t} u_{j,t} + \phi_j \theta_{j,t},$$

where $u_{j,t} = \frac{\partial \log p_j(y_t|\mathcal{F}_{t-1}, S_t = j)}{\partial \theta_{j,t}}$.

When y_t is conditionally Gaussian ($D \equiv \mathcal{N}$) and the dynamic parameters are the conditional means ($\theta_{S_t,t} \equiv \mu_{S_t,t}$), the filter implied by the DAMM for $E[y_{t+1}|\mathcal{F}_t]$ is equivalent to the one implied by the MCVLUE in the steady state ($\tilde{\kappa}_{j,t} = \tilde{\kappa}_j$).

So the ad hoc "collapsing step" actually leads to the MCVLUE which coincides with the DAMM $\Rightarrow$ let's study it!

What we do in the paper

1) We study the statistical properties of a DAMM model with conditionally Student's t shocks for the location parameters.
   i) Conditions for strong and weak stationarity.
   ii) Closed form solutions for $\mathbb{E}[y_{t+h}|\mathcal{F}_t]$ and $\mathbb{V}[y_{t+h}|\mathcal{F}_t]$
   iii) Sufficient conditions for the continuous invertibility of the filtered sequences $\hat{\theta}_{j,t+1}$.
   iv) Consistency and asymptotic Normality of the maximum likelihood estimator.
2) Comparison with the mixture autoregressive models of Wong and Li (2000) and Wong et al. (2009).
3) Monte Carlo simulation analysis.
4) Empirical comparison between the DAMM, KF, MAR, and the robust KF (RobKF) of Calvet et al. (2015).

An example with US industrial production

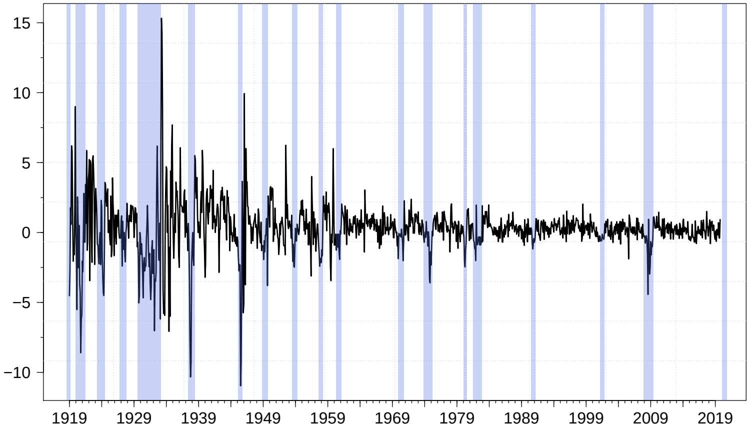
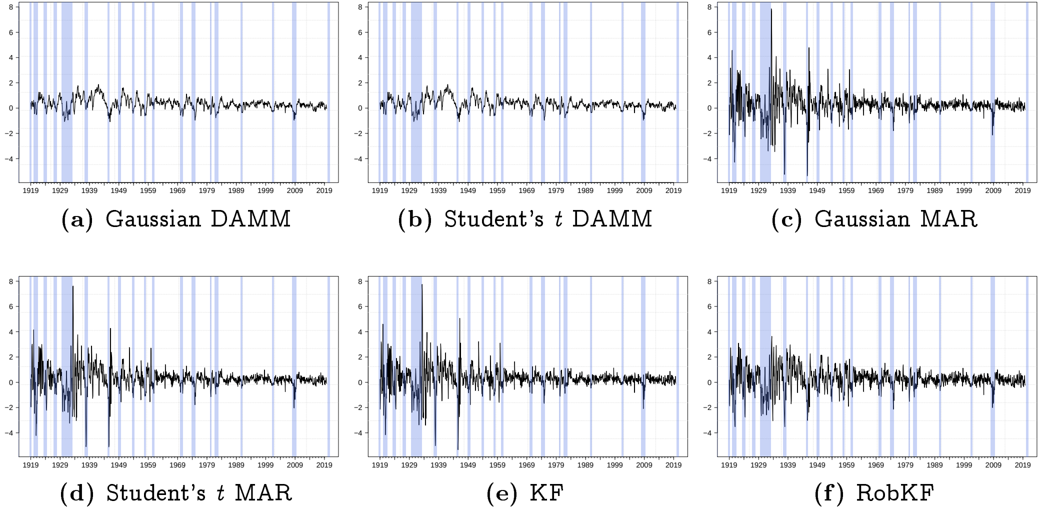


Figure: US industrial production at the monthly frequency from February 1919 to November 2019. Blue bands indicate period of recession according to the NBER recession indicators..

Filtered estimates



(a) Gaussian DAMM (b) Student's t DAMM (c) Gaussian MAR

(d) Student's t MAR (e) KF (f) RobKF

Figure: Signal extraction for the US industrial production from February 1919 to November 2019.

Conclusion

- When the conditional mean of a mixture of Gaussian distributions is estimated over time using the score (DAMM) we obtain a filter that resembles the ones discussed by Harrison and Stevens (1976) and Shumway and Stoffer (1991), among others, which we show is the MCVLUE.
- We study the statistical properties of these filters and the associated maximum likelihood estimator.
- An empirical illustration demonstrates the better performance of the DAMM with respect to the mixture of autoregressions of Wong and Li (2000) and Wong et al. (2009), and estimates from a contaminated signal plus noise model computed with the robust KF of Calvet et al. (2015).

Future research

- DAMM can also include time varying mixture probabilities and mixture components scales, see Catania (2019). This paper can be extended by including those features.
- Future research should focus on other links provided by score driven models (like the DAMM), and unobserved component models.

References I

Bar-Shalom, Y. and Tse, E. (1975). Tracking in a cluttered environment with probabilistic data association. *Automatica*, 11(5):451–460.

Calvet, L. E., Czellar, V., and Ronchetti, E. (2015). Robust filtering. *Journal of the American Statistical Association*, 110(512):1591–1606.

Catania, L. (2019). Dynamic adaptive mixture models with an application to volatility and risk. *Journal of Financial Econometrics*.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Gordon, K. and Smith, A. (1990). Modeling and monitoring biomedical time series. *Journal of the American Statistical Association*, 85(410):328–337.

Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society: Series B (Methodological)*, 38(3):205–228.

References II

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, volume 52. Cambridge University Press.

Kim, C.-J. (1994). Dynamic linear models with markov-switching. *Journal of Econometrics*, 60(1-2):1–22.

Kim, C.-J. and Nelson, C. R. (1999). State-space models with regime switching: classical and gibbs-sampling approaches with applications. *MIT Press Books*, 1.

Shumway, R. H. and Stoffer, D. S. (1991). Dynamic linear models with switching. *Journal of the American Statistical Association*, 86(415):763–769.

Wong, C., Chan, W., and Kam, P. (2009). A student t-mixture autoregressive model with applications to heavy-tailed financial data. *Biometrika*, 96(3):751–760.

References III

Wong, C. S. and Li, W. K. (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):95–115.