Forecasting Economic Time Series with Locally Adaptive Signal Extraction

Paolo Giordani Mattias Villani

October 2008

Motivation

- Most macroeconomic series show strong evidence of non-Gaussian features: local mean and/or variance and/or persistence are not constant (Clements and Hendry 99, Stock and Watson 96 and 08).
- Yet point forecasts from simple Gaussian AR(MA) are hard to beat. Performance of more complex nonlinear models is often disappointing. Simple Gaussian TVP models better than Markov switching (Marcellino 08, Clements and Krolzig 98, Bessec and Bouabdallah 05).
- Project goal 1: understand why fixed parameter models are rejected in testing yet hard to beat in forecasting.

- Project goal 2: is conditional variance equally hard to forecast? Which models do best? In particular, is smooth better than sudden variation?
- Project goal 3: evaluate various versions of *LASER*, a univariate forecasting model recently developed at the Riksbank, which allows for a variety of extensions of a basic ARMA model.

Locally Adaptive Signal Extraction and Regression

- Similar univariate models in Koopman and Bos 04 and Giordani and Kohn 08, among others.
- *LASER* models y_t as the sum of 3 processes, all of which can have mixture of normals (MN) innovations:
    1. A local mean μ_t, any conditionally Gaussian process (Gaussian conditional on latent indicators).
    2. A finite-order, stationary, AR process x_t with (i) MN innovations in the residual log variance (ii) MN errors (iii) unknown lag length p.
    3. An independently distributed measurement error/AO process ϵ_t with (i) MN innovations in the log variance (ii) MN errors.

- General framework: Observation equation and transition equation for x_t:

$$
\begin{aligned}
y_t &= \mu_t + x_t + \epsilon_t && (1)\\
x_t &= \rho_1 x_{t-1} + ... + \rho_p x_{t-p} + u_t && (2)\\
\epsilon_t &\sim MoN(k_y, \boldsymbol{\pi}_y, \boldsymbol{\alpha}_y, \sigma^2_{y,t}\boldsymbol{\xi}^2_y) && (3)\\
u_t &\sim MoN(k_x, \boldsymbol{\pi}_x, \boldsymbol{\alpha}_x, \sigma^2_{x,t}\boldsymbol{\xi}^2_x). && (4)
\end{aligned}
$$

- Notice $\sigma^2_{y,t}$, $\sigma^2_{x,t}$. Requires several choices to be made operational.

Modelling shifts in local mean

- Simple 1: non-stationary μ_t, TVP or change-points

$$\begin{aligned} \mu_t &= \mu_{t-1} + \sigma_{\mu,1} u_{\mu,t} \text{ with prob. } \pi_1 \\ &= \mu_{t-1} \text{ with prob. } 1 - \pi_1. \end{aligned} \quad (5)$$

- Simple 2: stationary μ_t, change-points

$$\begin{aligned} \mu_t &= \overline{\mu} + \sigma_{\mu,2} u_{\mu,t} \text{ with prob. } \pi_1 \\ &= \mu_{t-1} \text{ with prob. } 1 - \pi_1. \end{aligned} \quad (6)$$

- Simple 3: stationary μ_t, cocktail

$$\begin{aligned} \mu_t &= \mu_{t-1} + \sigma_{\mu,1} u_{\mu,t} \text{ with prob. } \pi_1 && (7) \\ &= \overline{\mu} + \sigma_{\mu,2} u_{\mu,t} \text{ with prob. } \pi_2 && (8) \\ &= \mu_{t-1} \text{ with prob. } 1 - \pi_1 - \pi_2. \end{aligned}$$

- All 3 unappealing for most macro series with infrequent shifts: we don't expect y_t to jump to its new (distant) local mean. Poor forecasts when this assumption is made.
- Our choice when shifts infrequent: smoother transition to new local mean

$$\begin{aligned}
\Delta\mu_t &= (1-\rho_\mu)(\widetilde{\mu}_t - \mu_{t-1}) \\
\widetilde{\mu}_t &= \mu_{t-1} + \sigma_{\mu,1} u_{\mu,t} \text{ with prob. } \pi_1 \\
\widetilde{\mu}_t &= \overline{\mu} + \sigma_{\mu,2} u_{\mu,t} \text{ with prob. } \pi_2 \\
\widetilde{\mu}_t &= \widetilde{\mu}_{t-1} \text{ with prob. } 1-\pi_1-\pi_2,
\end{aligned} \tag{9}$$

$\rho = 0.8$ on quartely data.

Modelling shifts in variances

Same approach as in Giordani and Kohn 08, but on two variances

$$\begin{aligned} \ln \sigma_{y,t} &= \ln \sigma_{y,t-1} + \delta_y e_{y,t} \text{ with prob. } \pi_y^\sigma \\ \ln \sigma_{y,t} &= \ln \sigma_{y,t-1} \text{ with prob. } 1 - \pi_y^\sigma, \end{aligned} \tag{10}$$

and

$$\begin{aligned} \ln \sigma_{x,t} &= \ln \sigma_{x,t-1} + \delta_x e_{x,t} \text{ with prob. } \pi_x^\sigma \\ \ln \sigma_{x,t} &= \ln \sigma_{x,t-1} \text{ with prob. } 1 - \pi_x^\sigma, \end{aligned} \tag{11}$$

Allows for infrequent, large shifts as well as for continuos, small shifts.

Four forecasting models

Four versions of LASER:

1. ARMA. $\mu_t, \sigma_{x,t}$ and $\sigma_{\epsilon,t}$ constant and all innovations normal. Benchmark.
2. Shifts. Infrequent shifts in μ_t, constant $\sigma_{x,t}$ and $\sigma_{y,t}$ and normal innovations.
3. Robust TVP. Normal innovations in μ_t, $\ln(\sigma_{x,t})$ and $\ln(\sigma_{y,t})$ and MN innovations elsewhere.
4. Robust Shifts. Infrequent shifts in μ_t, $\ln(\sigma_{x,t})$ and $\ln(\sigma_{y,t})$ and MN innovations elsewhere.

Priors

- Common where possible. In actual applications we would use domain knowledge.
- Lag length 1-4: $prob(p=1, p=2, p=3, p=4) = (0.4, 0.3, 0.2, 0.1)$.

$$\rho_1, ..., \rho_p | p \sim N\left(\begin{bmatrix} \underline{\rho} \\ 0 \\ ... \\ 0 \end{bmatrix}, \underline{\sigma}^2_\rho \begin{bmatrix} 1 & \underline{\rho} & \underline{\rho}^2 & \underline{\rho}^3 \\ & 1 & \underline{\rho} & \underline{\rho}^2 \\ & & 1 & \underline{\rho} \\ & & & 1 \end{bmatrix} \right)$$

if $\rho_1, ..., \rho_p$ stationary, 0 otherwise. $\underline{\rho} = 0.8$ and $\underline{\sigma}_\rho = 0.2$.

- $\sigma^2_{\mu,1}$ inverse gamma

$$\xi^2_y \sim IG(10 \times \lambda^2 \widehat{\sigma}^2_y, 10),$$

where $\widehat{\sigma}^2_y$ is the sample variance of $y^s_t = 0.8 y^s_{t-1} + 0.2 y_t . \lambda = 0.25$ in *Robust Shifts* and *Shifts* and $\lambda = 0.25\sqrt{0.02}$ in *Robust TVP*. Same one step-ahead variance.

Priors

- For *(Robust) Shifts* $\pi_1 = 0.02$. Average interval between shifts 12 years. Implied distribution for the number of shifts binomial, with most mass in the 0-8 interval in the period 1959-2007 and probability of a shift equally distributed across all time periods (see Koop and Potter 07 for a discussion).
- $$\begin{aligned} \delta_y^2 &\sim IG(10 \times \lambda_y^2, 10). \\ \delta_x^2 &\sim IG(10 \times \lambda_x^2, 10). \end{aligned}$$
- For *Robust Shifts*, $\lambda_y = \lambda_x = 0.7$, $\pi_y = \pi_x = 0.02$. Interpretation: a ± 0.7 shock to $\log(\sigma_y)$ increases (decreases) σ_y by 50% (25%).
- For *Robust TVP* $\pi_y = \pi_x = 1$ and $\lambda_y = \lambda_x = 0.7\sqrt{0.02}$.

Inference

- MCMC. Standard Gibbs very inefficient. We use the algorithms of Gerlach et al. 00 and Giordani and Kohn 08.
- Not conditionally Gaussian, but can be made conditionally Gaussian in blocks.

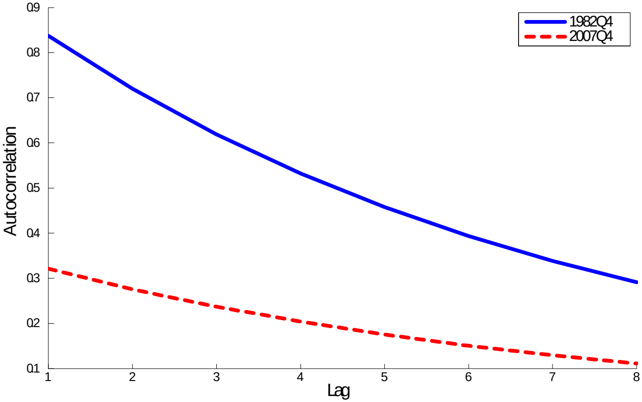
Time variation in persistence

- The AR parameters $\rho_1, ..., \rho_p$ are fixed.
- But if $\sigma_{y,t}$ and $\sigma_{x,t}$ are not, the local autocorrelation structure can change substantially (*local signal extraction*).
- Example for $p = 1$

$$\begin{aligned} corr(y_t, y_{t-1}) &\rightarrow \rho \text{ for } \sigma_{y,t}/\sigma_{x,t} \rightarrow 0 \\ corr(y_t, y_{t-1}) &\rightarrow 0 \text{ for } \sigma_{y,t}/\sigma_{x,t} \rightarrow \infty \end{aligned}$$

- Parsimonious approach to shifts in persistence, particularly for large p. Probably sufficient on macro data.
- Example: on both US (figure) and Swedish inflation we find local autocorrelations in 2006q4 much lower than in 1982q4.

Autocorrelation functions for US inflation



Forecasting exercise

- 3 US series, 3 corresponding Swedish. Quarterly 1959q2-2006q4 US and 1980q2-2007q4 Sweden.
- real GDP growth, CPI inflation, three month treasury bill.
- Start forecasting after 10 years.
- Caveat. Forecasts average uncertainty over parameters, errors, shifts: The forecast distributions can be asymmetric and RMSFE may not be representative.

Point forecasts: Results

1. *Robust Shifts* performs substantially better than *Shifts*. Somewhat better than *Robust TVP* one quarter ahead, about as well at four.
2. *Robust Shifts* sizably outperforms *Robust TVP* for interest rate forecasting.
3. *Robust Shifts* performs slightly better than *ARMA* on US data and substantially better on Swedish data.
4. The process for the local mean μ_t assumed for *Shifts, Robust TVP* and *Robust Shifts* is arguably poor for inflation and interest rates.

U.S. 1969Q2-2006Q4

	1Q			4Q		
	GDP	Infl.	Rate	GDP	Infl	Rate
ARMA	3.274	1.753	0.851	**3.430**	2.542	1.830
Shift	0.985	0.993	1.011	1.033	1.005	0.997
Robust TVP	1.002	1.023	1.133	1.014	0.972	0.974
Robust Shift	**0.974**	1.029	**0.866**	1.052	1.001	0.942
Random walk	1.269	**0.989**	0.912	1.363	**0.948**	**0.931**

Sweden 1990Q2-2007Q4

	$1Q$			$4Q$		
	GDP	Infl.	Rate	GDP	Infl.	Rate
ARMA	2.095	3.117	1.071	2.805	3.337	2.404
Shift	1.010	0.989	0.962	0.996	0.966	0.935
Robust TVP	0.986	0.946	0.971	**0.983**	0.951	0.938
Robust Shift	**0.985**	**0.909**	**0.873**	0.989	**0.935**	**0.837**
Random walk	1.115	1.261	1.161	1.184	1.051	0.918

U.S. 1990Q2-2006Q4

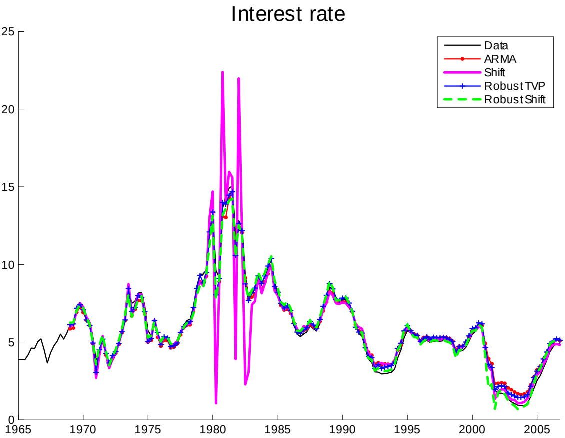
	1Q			4Q		
	GDP	Infl.	Rate	GDP	Infl.	Rate
ARMA	1.991	1.289	0.409	**1.950**	1.410	1.419
Shift	0.985	**0.993**	1.011	1.033	1.005	0.997
Robust TVP	1.002	1.023	1.133	1.014	**0.972**	0.974
Robust Shift	**0.974**	1.029	**0.866**	1.052	1.001	**0.942**
Random walk	2.086	1.345	1.900	2.402	1.710	1.201

Large shifts and normal errors: a poor modelling assumption

- Intuition: if normality is assumed and a shock is larger than 2*std*, it will be interpreted as a break.
- US interest rates extreme case: outliers/increased variance confused for breaks by *Shifts*.
- Swedish inflation: *Shifts* takes longer to adapt to a mean shift.

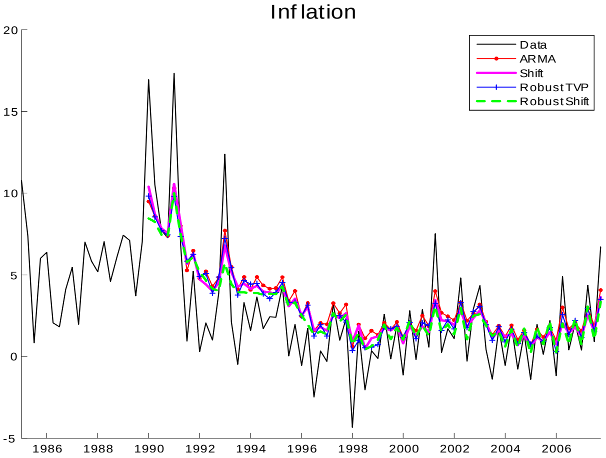
4 quarter ahead US interest rate forecasts

rate



2.*pdf*

4 quarter ahead Swedish inflation forecasts



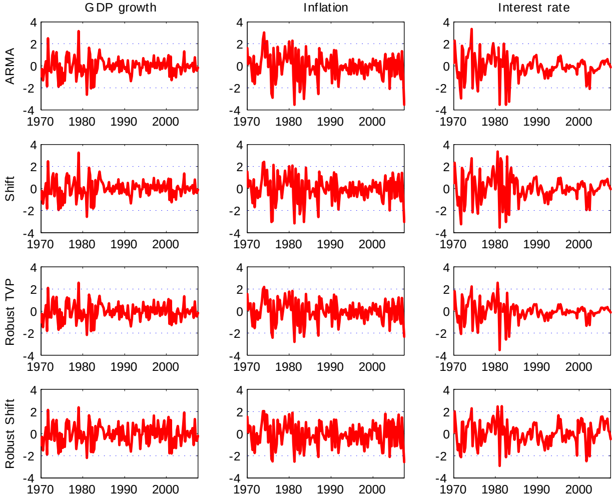
Interval forecasts

- Much sharper conclusions than for point forecasts:
  1. Robust Shifts is the only model to do well on all six series and sub-samples.
  2. Robust Shifts performs drammatically better than ARMA and Shifts: much easier to beat fixed parameter specifications.
  3. Robust Shifts largely and consistently outperforms Robust TVP.
- *Tests of correct interval coverage* (see paper): independence of forecast errors and correct coverage.
- *Normalized forecast errors*:

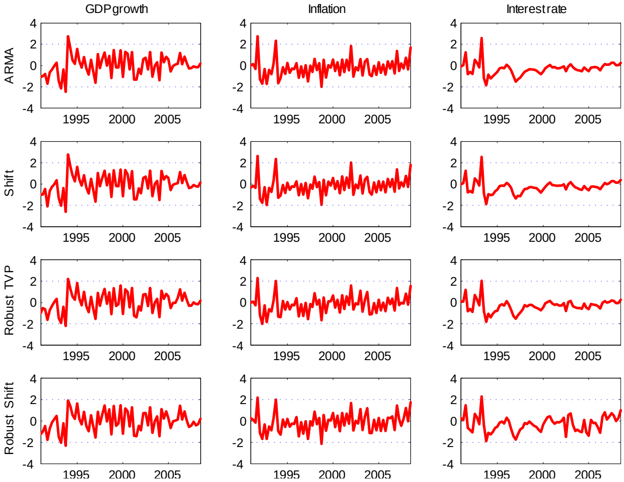
$$u_{t+1} = \Phi^{-1}[F_t(y_{t+1})],$$

  $F_t(y_{t+1})$ one-step-ahead predictive CDF evaluated at the realization y_{t+1}, and $\Phi^{-1}(\cdot)$ inverse Gaussion CDF.
- If the model is correct, u_{t+1} should be and $NID(0, 1)$.

Normalized forecast errors, US



Normalized forecast errors, Sweden



Conclusions

- We should be extra cautious: only six series, only one history. This said...
- If large parameter shifts are allowed, at least fat tailed errors strongly recommended on quarterly data, a must at higher frequencies.
- If this is done, Robust Shifts type models seem a promising forecasting tool, and very promising for interval forecasting.