Dynamic term structure models with score-driven time-varying parameters: estimation and forecasting

Siem Jan Koopman [1] André Lucas [1] Marcin Zamojski [2]

[1] Vrije Universiteit Amsterdam and Tinbergen Institute [2] University of Gothenburg

IAAE Conference
Nicosia, Cyprus
June 25–28, 2019

Paper/slides @ zamojski.net More @ gasmodel.com

Motivation

- Modelling cross-sections of interest rates in time is challenging.
- Successful models/methods are not well grounded in theory and rely on strong assumptions.
- Although short-term forecasting performance is usually good, the results are hard to explain and hard to tie back to economics.
- We look at the dynamic Nelson-Siegel model (Diebold, Rudebusch, and Aruoba, 2006).
  - Parameter-driven approach that requires assumptions about Gaussianity, homoskedasticity, and independence.
- We want to know:
  - Which assumptions should be relaxed at low/higher frequencies?
  - What is the relative value-added of relaxing these distributional assumptions?
  - What are relative performance gains depending on whether we use monthly, weekly, or daily data.

Motivation

- We propose to extract latent factors using score-driven models (SDMs) which are easy to implement and relatively quick to estimate.
  - We can easily relax the normality, homoskedasticity, and independence assumptions. One at a time and all at once.
  - Even though score-driven models tend to be misspecified, they have been shown to offer similar fit to correctly specified models in the univariate setting (Koopman, Lucas, and Scharth, 2016).
  - We conduct a large scale Monte Carlo study to verify whether this holds in the multivariate setting as well.

Findings

- It is hard to beat the simple model for very short-term forecasts.
- We can expect out-performance after few periods.
- We can gain new insights about the nature of changes in the yield curve if we try.
- Allowing for time-varying variances improves fit the most at higher frequencies. At lower frequencies, non-normality dominates.
- Performance improvement is visible mostly for shorter maturities.
- We find that SDM are able to fit the data very well (up to 30% improvement in likelihoods and informational content).

Outline

- SDMs/ODMs vs. PDMs
- Nelson-Siegel
- Extensions
- Simulations
- Empirics
- Conclusions

Score Driven Models (Creal et al., 2013)

A general class of likelihood-based observation-driven models with an observation equation:

$$y_t \sim p\left(y_t | Y_{t-1}, \boldsymbol{f}_t; \boldsymbol{\theta}\right),$$

$$Y_t = \{y_1, \ldots, y_t\}.$$

The transition equation / updating rule is autoregressive:

$$\boldsymbol{f}_{t+1} = \boldsymbol{\omega} + \boldsymbol{B}\left(\boldsymbol{f}_t - \boldsymbol{\omega}\right) + \boldsymbol{A} s_t(y_t, \boldsymbol{f}_t).$$

where the innovation or 'driving' mechanism s_t is given as:

$$s_t = S_t \nabla_t, \qquad \nabla_t = \frac{\partial p\left(y_t | \boldsymbol{f}_t, Y_t; \boldsymbol{\theta}\right)}{\partial \boldsymbol{f}_t^\intercal},$$

$$S_t = \mathcal{I}_{t|t-1}^{-k}, \qquad \mathcal{I}_{t|t-1} = \mathrm{E}\left[\nabla_t \nabla_t^\intercal\right] = -\mathrm{E}\left[\frac{\partial^2 p\left(y_t | \boldsymbol{f}_t, Y_t; \boldsymbol{\theta}\right)}{\partial \boldsymbol{f}_t^2}\right]$$

Creal et al. (2018) extend the method to other loss functions, e.g., the GMM criterion.

Score Driven Models (Creal et al., 2013)

- Updates, s_t, can be linear, non-linear in $\boldsymbol{f}_t$.
- Scaling, S_t, can be linear, non-linear, or invariant in $\boldsymbol{f}_t$.
- Gradients and Hessian are familiar entities in estimation and optimization.
- Typically computationally simple to use.
- Provides an internally consistent and general updating rule.
- More of a method than a model (widely applicable, but also misspecified?).
- Arguably, the only difficulty... deriving $\mathcal{I}_{t|t-1}$ (approximate methods are a work in progress, papers often take shortcuts).

Score-driven Dynamic Nelson-Siegel Model

$$\boldsymbol{y}_t = \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathrm{N}\left(0, \boldsymbol{\Sigma}\right),$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \mathbf{1} & \frac{1-e^{-\lambda \boldsymbol{\tau}}}{\lambda \boldsymbol{\tau}} & \frac{1-e^{-\lambda \boldsymbol{\tau}}}{\lambda \boldsymbol{\tau}} - e^{-\lambda \boldsymbol{\tau}} \end{bmatrix},$$

$$\boldsymbol{f}_{t+1} = \boldsymbol{\Phi} \boldsymbol{f}_t + \eta_t, \qquad \boldsymbol{\eta}_t \sim \mathrm{N}\left(0, \boldsymbol{\Omega}\right).$$

Score-driven Dynamic Nelson-Siegel Model

$$
\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim t\left(0, \boldsymbol{\Sigma}, \nu\right), \\
\boldsymbol{\Lambda} &= \begin{bmatrix} \mathbf{1} & \frac{1-e^{-\lambda \boldsymbol{\tau}}}{\lambda \boldsymbol{\tau}} & \frac{1-e^{-\lambda \boldsymbol{\tau}}}{\lambda \boldsymbol{\tau}} - e^{-\lambda \boldsymbol{\tau}} \end{bmatrix}, \\
\boldsymbol{\Sigma} &= \boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t, \\
\boldsymbol{D}_t &= \mathrm{diag}\left(\omega_1^2 e^{h_t}, \ldots, \omega_p^2 e^{h_t}\right), \\
\boldsymbol{f}_{t+1} &= \boldsymbol{\omega} + \boldsymbol{B}(\boldsymbol{f}_t - \boldsymbol{\omega}) + \boldsymbol{A} \boldsymbol{s}_t.
\end{aligned}
$$

Score-driven Dynamic Nelson-Siegel Model

Symbol	Type	Estimated with	Parameters	Fat tails	Heteroskedasticity	Correlation
'KF 1C'	PD	Kalman Filter	23			
'KF pC'	PD	Kalman Filter	33			
'GAS N 1C'	SD	GAS	23			
'GAS N pC'	SD	GAS	33			
'GAS t 1C'	SD	GAS	24	✓		
'GAS t pC'	SD	GAS	34	✓		
'GAS N 1TV'	SD	GAS	35		✓	
'GAS t 1TV'	SD	GAS	36	✓	✓	
'GAS N pTV'	SD	GAS	57		✓	
'GAS t pTV'	SD	GAS	58	✓	✓	
'GAS N pTV+c'	SD	GAS	58		✓	✓
'GAS t pTV+c'	SD	GAS	59	✓	✓	✓

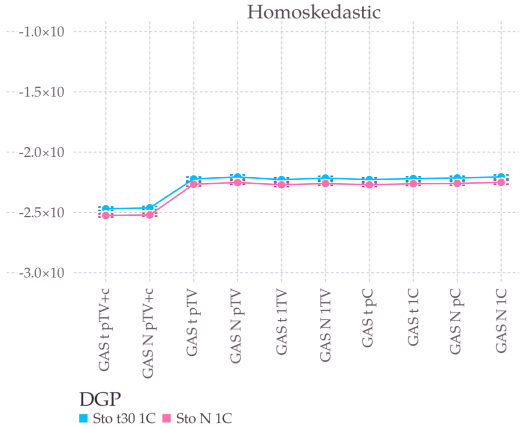
- 2 parameter-driven benchmarks.
- 10 score-driven filters:
  - 5 filters with fat-tails (robust to outliers),
  - 6 filters allowing for heteroskedasticity,
  - 2 filters allowing for contemporaneous correlations in pricing errors.

Simulations, DGPs

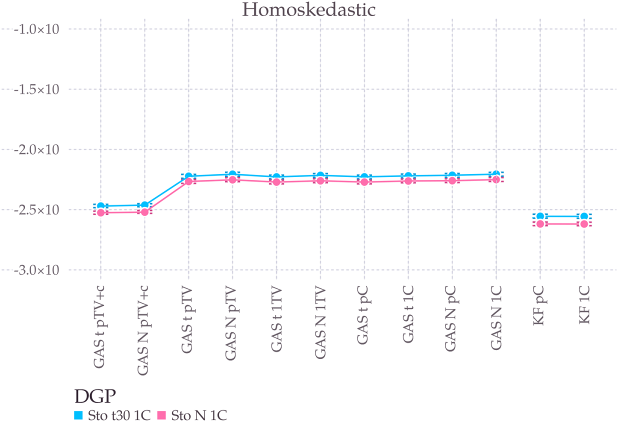
Symbol	DGP	Distribution	Fat tails (df)	Heteroskedasticity	Correlation (ρ)
'Det t pTV'	Deterministic paths	Student's t	✓(30)	✓	✓(0.35)
'Det N pTV'	Deterministic paths	Normal		✓	✓(0.35)
'Det t pTV'	Deterministic paths	Student's t	✓(30)	✓	
'Det N pTV'	Deterministic paths	Normal		✓	
'Det t 1TV'	Deterministic paths	Student's t	✓(30)	✓	
'Det N 1TV'	Deterministic paths	Normal		✓	
'Sto t30 1TV'	Stochastic paths	Student's t	✓(30)	✓	
'Sto t50 1TV'	Stochastic paths	Student's t	✓(50)	✓	
'Sto N 1TV'	Stochastic paths	Normal		✓	
'Sto t30 1C'	Stochastic paths	Student's t	✓(30)		
'Sto t50 1C'	Stochastic paths	Student's t	✓(50)		
'Sto N 1C'	Stochastic paths	Normal			

- For each data-generating process we construct 1,500 samples of length T=1,050.
- This corresponds to around 4 years of daily data.
- We simulate data for 11 maturities: 3, 6, 9, 12, 18, 24, 36, 60, 84, 108, and 120 months.

Homoskedastic simulations, in-sample fit, AIC



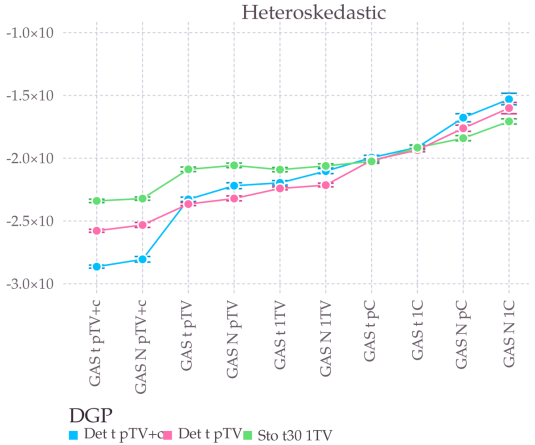
Homoskedastic simulations, in-sample fit, AIC



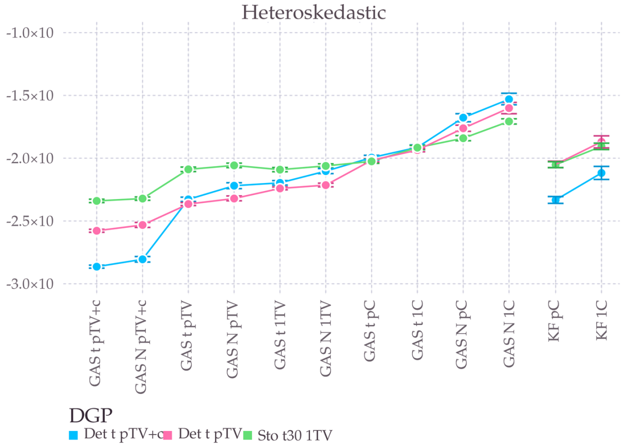
Homoskedastic simulations, out-of-sample RMSE, 1-ahead

RMSE	‘GAS t pTV+c’	‘GAS N pTV+c’	‘GAS t pTV’	‘GAS N pTV’	‘GAS t 1TV’	‘GAS N 1TV’	‘GAS t pC’	‘GAS t 1C’	‘GAS N pC’	‘GAS N 1C’	‘KF pC’	‘KF 1C’	‘KF pC’ value
‘Sto t30 1C’	1.01	1.01	1.01	1.00	1.01	1.00	1.01	1.01	1.00	1.00	1.00	1.00	0.09
‘Sto N 1C’	1.01	1.01	1.01	1.00	1.01	1.00	1.01	1.01	1.00	1.00	1.00	1.00	0.09

Heteroskedastic simulations, in-sample fit, AIC



Heteroskedastic simulations, in-sample fit, AIC



Out-of-sample RMSE, yields, 1-ahead

RMSE	'GAS t pTV+c'	'GAS N pTV+c'	'GAS t pTV'	'GAS N pTV'	'GAS t 1TV'	'GAS N 1TV'	'GAS t pC'	'GAS t 1C'	'GAS N pC'	'GAS N 1C'	'KF pC'	'KF 1C'	'KF pC' value
'Det t pTV+c'	**0.97**	**0.96**	**0.98**	**0.96**	1.00	**0.96**	1.08	1.08	**0.97**	**0.98**	1.00	1.00	0.05
'Det N pTV+c'	**0.97**	**0.97**	**0.98**	**0.96**	1.00	**0.97**	1.09	1.08	**0.97**	**0.98**	1.00	1.00	0.05
'Det t pTV'	**0.99**	**0.99**	**0.98**	**0.98**	**0.99**	**0.98**	1.08	1.08	**0.99**	1.00	1.00	**0.99**	0.05
'Det N pTV'	**0.97**	1.00	**0.97**	**0.99**	1.00	**0.99**	1.10	1.10	1.00	1.10	1.00	1.67	0.05
'Det t 1TV'	1.01	1.00	1.00	1.00	1.00	1.00	1.09	1.10	1.01	1.02	1.00	1.01	0.05
'Det N 1TV'	1.01	1.01	1.00	1.00	1.00	1.00	1.09	1.10	1.01	1.01	1.00	1.01	0.06
'Sto N 1TV'	1.01	1.00	1.01	**0.99**	1.01	**0.99**	1.03	1.03	1.00	1.01	1.00	1.01	0.07
'Sto t30 1TV'	1.00	**0.99**	1.01	**0.99**	1.01	**0.99**	1.02	1.03	1.00	1.01	1.00	1.01	0.07
'Sto t50 1TV'	1.01	1.00	1.02	**0.99**	1.01	**0.99**	1.02	1.03	1.00	1.01	1.00	1.01	0.07
'Sto t30 1C'	1.01	1.01	1.01	1.00	1.01	1.00	1.01	1.01	1.00	1.00	1.00	1.00	0.09
'Sto t50 1C'	1.01	1.01	1.01	1.00	1.01	1.00	1.01	1.01	1.00	1.00	1.00	1.00	0.09
'Sto N 1C'	1.01	1.01	1.01	1.00	1.01	1.00	1.01	1.01	1.00	1.00	1.00	1.00	0.09

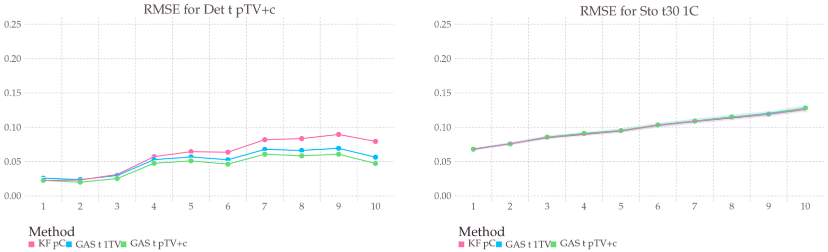
Out-of-sample RMSE, short-term yields, 1-ahead

RMSE	'GAS t pTV+c'	'GAS N pTV+c'	'GAS t pTV'	'GAS N pTV'	'GAS t 1TV'	'GAS N 1TV'	'GAS t pC'	'GAS t 1C'	'GAS N pC'	'GAS N 1C'	'KF pC'	'KF 1C'	'KF pC' value
'Det t pTV+c'	**0.78**	**0.82**	**0.80**	**0.80**	**0.85**	**0.83**	1.14	1.08	**0.87**	**0.93**	1.00	**0.96**	0.02
'Det N pTV+c'	**0.79**	**0.81**	**0.80**	**0.81**	**0.85**	**0.83**	1.15	1.09	**0.87**	**0.93**	1.00	**0.96**	0.02
'Det t pTV'	**0.84**	**0.87**	**0.85**	**0.86**	**0.90**	**0.90**	1.23	1.16	**0.96**	**0.99**	1.00	**0.92**	0.02
'Det N pTV'	**0.82**	**0.90**	**0.83**	**0.89**	**0.91**	**0.93**	1.28	1.20	**0.98**	1.33	1.00	1.86	0.02
'Det t 1TV'	**0.95**	**0.95**	**0.96**	**0.95**	**0.97**	**0.96**	1.29	1.35	1.01	1.07	1.00	1.07	0.02
'Det N 1TV'	**0.96**	**0.97**	**0.96**	**0.95**	**0.98**	**0.96**	1.28	1.34	1.02	1.07	1.00	1.06	0.03
'Sto N 1TV'	1.01	1.00	1.02	1.00	1.02	1.00	1.04	1.06	1.01	1.04	1.00	1.05	0.05
'Sto t30 1TV'	1.01	1.00	1.02	1.00	1.02	1.00	1.03	1.06	1.00	1.04	1.00	1.05	0.05
'Sto t50 1TV'	1.02	1.00	1.03	1.00	1.02	1.00	1.04	1.06	1.01	1.04	1.00	1.06	0.05
'Sto t30 1C'	1.01	1.02	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.08
'Sto t50 1C'	1.01	1.02	1.01	1.00	1.01	1.00	1.01	1.01	1.00	1.00	1.00	1.00	0.08
'Sto N 1C'	1.01	1.01	1.01	1.00	1.01	1.00	1.01	1.01	1.00	1.00	1.00	1.00	0.08

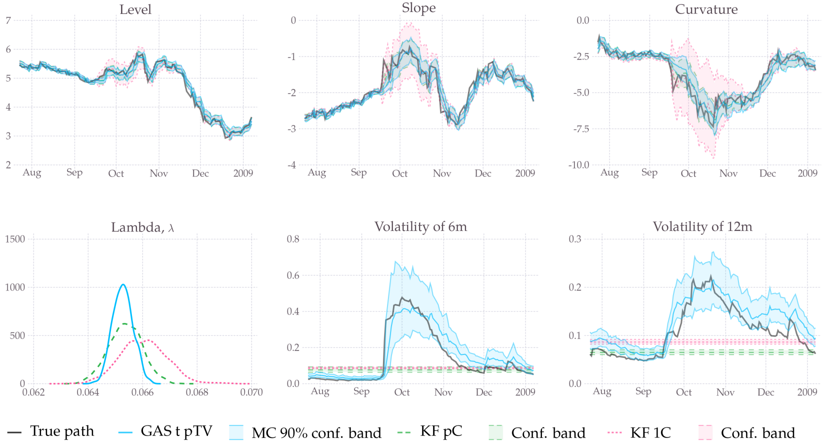
Out-of-sample RMSE, curvature factor, 1-ahead

RMSE	'GAS t pTV+c'	'GAS N pTV+c'	'GAS t pTV'	'GAS N pTV'	'GAS t 1TV'	'GAS N 1TV'	'GAS t pC'	'GAS t 1C'	'GAS N pC'	'GAS N 1C'	'KF pC'	'KF 1C'	'KF pC' value
'Det t pTV+c'	**0.84**	**0.86**	**0.87**	**0.86**	**0.93**	**0.91**	1.14	1.12	**0.97**	**0.94**	1.00	**0.99**	0.16
'Det N pTV+c'	**0.84**	**0.86**	**0.86**	**0.86**	**0.93**	**0.91**	1.13	1.12	**0.96**	**0.94**	1.00	**0.99**	0.16
'Det t pTV'	**0.84**	**0.86**	**0.84**	**0.85**	**0.91**	**0.91**	1.12	1.11	**0.98**	**0.95**	1.00	1.00	0.17
'Det N pTV'	**0.81**	**0.87**	**0.81**	**0.86**	**0.91**	**0.93**	1.18	1.16	1.00	1.87	1.00	4.81	0.17
'Det t 1TV'	**0.93**	**0.94**	**0.94**	**0.95**	**0.93**	**0.94**	1.11	1.13	1.00	1.03	1.00	1.05	0.19
'Det N 1TV'	**0.95**	**0.96**	**0.95**	**0.97**	**0.95**	**0.96**	1.15	1.17	1.01	1.04	1.00	1.06	0.20
'Sto N 1TV'	**0.97**	**0.97**	**0.97**	**0.97**	**0.97**	**0.97**	**0.97**	**0.99**	1.01	1.05	1.00	1.02	0.24
'Sto t30 1TV'	**0.96**	**0.96**	**0.97**	**0.96**	**0.97**	**0.97**	**0.96**	**0.99**	1.00	1.04	1.00	1.02	0.24
'Sto t50 1TV'	**0.97**	**0.96**	**0.98**	**0.96**	**0.97**	**0.97**	**0.96**	**0.99**	1.00	1.04	1.00	1.02	0.24
'Sto t30 1C'	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.26
'Sto t50 1C'	**0.99**	**0.99**	1.00	1.00	1.00	**0.99**	1.00	1.00	**0.99**	**0.99**	1.00	1.00	0.26
'Sto N 1C'	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.26

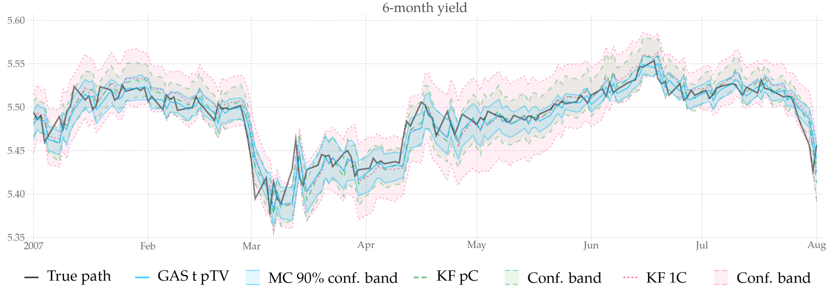
Out-of-sample RMSE, short-term yields, n-ahead



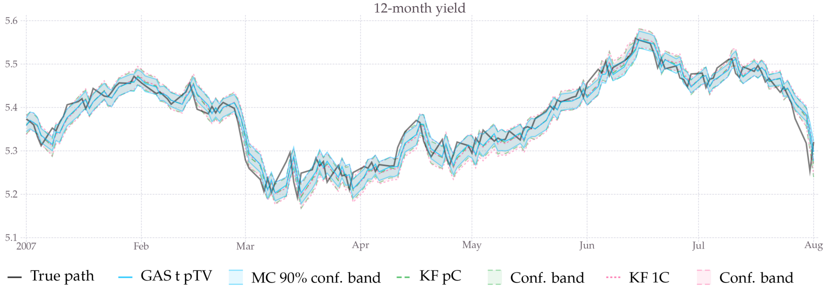
In-sample factor fit



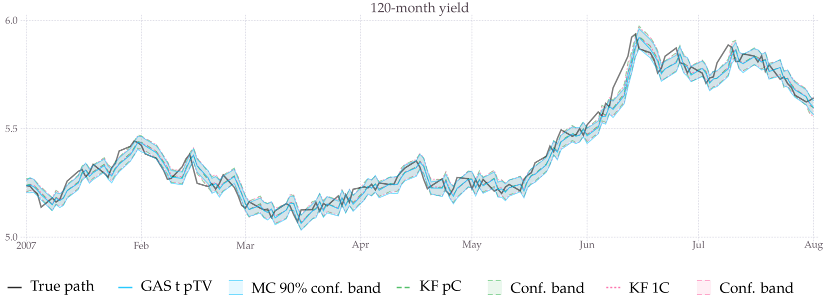
In-sample, short-term yield fit



In-sample, medium-term yield fit



In-sample, long-term yield fit



Empirical application

- Zero-coupon yields from ICAP for 14 countries (sample periods and number of observations in brackets):
  - Australia (1999–2017, 4790)
  - Canada (1999–2017, 4790)
  - Czech Rep. (2005–2017, 3265)
  - Denmark (1998–2017, 5130)
  - Hungary (2005–2017, 3265)
  - Japan (1998–2012, 3805)
  - Norway (2000–2017, 4571)
  - New Zealand (2005–2017, 3265)
  - Poland (2006–2017, 2915)
  - South Africa (2005–2017, 3265)
  - Sweden (1998–2017, 5131)
  - Switzerland (1997–2017, 5368)
  - U.K. (1997–2017, 5368)
  - U.S. (1997–2017, 5338)
- For most, short term rates are based on LIBOR. Medium and long term rates are based on FRAs and SWAPs.
- 11 maturities: 3, 6, 9, 12, 18, 24, 36, 60, 84, 108, and 120 months.

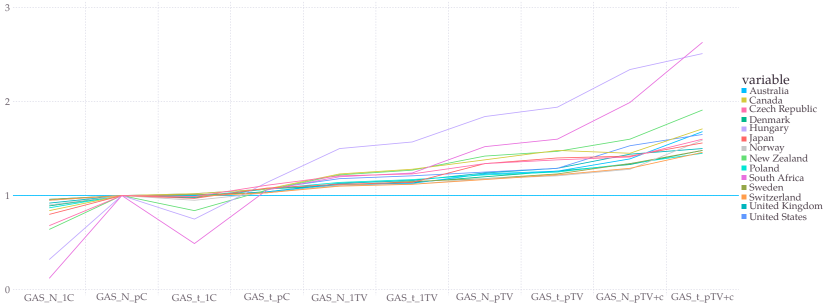
Empirical application

AIC

Model	Australia	Canada	Czech Republic	Denmark	Hungary	Japan	Norway	New Zealand	Poland	South Africa	Sweden	Switzerland	United Kingdom	United States
'KF 1C'	-24.18	-18.92	-16.44	-24.38	-2.5	-27.7	-24.38	-11.97	-23.13	-0.3	-29.92	-33.56	-25.76	-27.25
'KF pC'	-29.73	-28.47	-27.84	-33.11	-19.33	-40.08	-31.55	-26.95	-33.88	-13.17	-35.57	-40.01	-34.33	-32.81
'GAS N 1C'	-23.79	-19.72	-16.9	-24.47	-2.96	-28.82	-25.15	-13.09	-22.38	-1.18	-29.64	-33.55	-26.09	-25.49
'GAS N pC'	-24.79	-23.5	-24.72	-27.35	-9.37	-35.98	-27.81	-20.45	-25.59	-10.22	-30.97	-35.19	-28.42	-26.84
'GAS t 1C'	-24.73	-23.92	-24.48	-26.88	-7	-34.95	-26.38	-17.1	-25.43	-4.98	-31.2	-34.7	-28.3	-27.34
'GAS t pC'	-25.69	-25.21	-27.56	-29.16	-10.72	-39	-28.62	-21.73	-27.5	-10.85	-32.04	-36.13	-29.56	-28.74
'GAS N 1TV'	-27.66	-28.99	-29.88	-30.4	-14.07	-40.45	-30.78	-24.99	-29.29	-12.23	-34.45	-38.73	-32.22	-31.63
'GAS t 1TV'	-27.92	-30.02	-30.34	-31.14	-14.71	-41.18	-31.1	-25.93	-29.95	-12.72	-35.59	-39.52	-32.88	-32.48
'GAS N pTV'	-30.5	-32.41	-33.14	-32.8	-17.24	-48.19	-32.85	-28.97	-31.08	-15.52	-36.5	-41.17	-35.26	-33.58
'GAS t pTV'	-31.22	-34.87	-34.11	-34.4	-18.18	-50.45	-33.72	-30.13	-32.08	-16.34	-38.21	-42.81	-36.68	-34.69
'GAS N pTV+c'	-34.47	-34.01	-34.92	-36.28	-21.97	-51.07	-35.5	-32.65	-34.33	-20.36	-41.53	-45.23	-41.01	-41.08
'GAS t pTV+c'	-41.76	-40.17	-39.45	-40.36	-23.52	-56.28	-44.12	-39.01	-37.17	-26.89	-45.88	-51.23	-42.52	-44.22

Empirical application

Log-likelihood



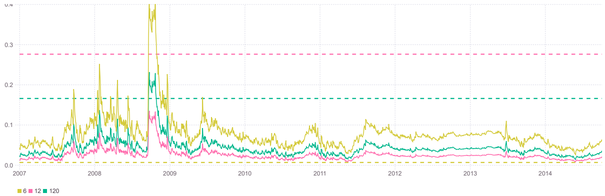
Empirical application, U.S.

- Sample period is 1997–2017 (5338 observations at daily frequency).
- 11 maturities: 3, 6, 9, 12, 18, 24, 36, 60, 84, 108, and 120 months.
- If volatility is assumed to be constant, the average loglikelihood contribution is 12.8 and 13.7 for the Normal and t (with 17.8 degrees of freedom) densities respectively.
- `GAS t pTV' achieves average loglikelihood contribution of 17.36 with 26.6 degrees of freedom.

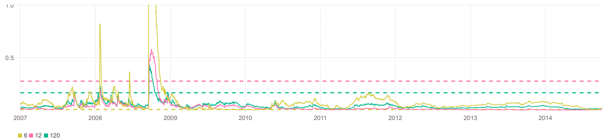
	'GAS t pTV+c'	'GAS N pTV+c'	'GAS t pTV'	'GAS N pTV'	'GAS t 1TV'	'GAS N 1TV'	'GAS t pC'	'GAS t 1C'	'GAS N pC'	'GAS N 1C'	'KF pC'	'KF 1C'	'KF pC' value
LL	22.12	20.55	17.36	16.8	16.25	15.82	14.37	13.68	13.43	12.75	16.41	13.63	
AIC	-44.22	-41.08	-34.69	-33.58	-32.48	-31.63	-28.74	-27.34	-26.84	-25.49	-32.81	-27.25	
OOS RMSE (all)	1.32	1.05	**0.99**	**0.97**	1.03	1.00	1.07	1.07	**0.98**	**0.97**	1.00	1.03	3.18
OOS MAPE (all)	1.14	**0.92**	**0.90**	**0.94**	1.05	1.02	1.08	1.08	**0.95**	**0.93**	1.00	1.03	2.55
OOS RMSE (short)	**0.91**	**0.85**	**0.84**	**0.85**	**0.96**	**0.93**	1.13	1.14	**0.94**	**0.95**	1.00	1.03	1.73
OOS MAPE (short)	**0.85**	**0.79**	**0.84**	**0.81**	**0.94**	**0.89**	1.10	1.15	**0.93**	**0.95**	1.00	1.04	3.05

Empirical application, U.S., pricing errors volatility

`GAS t 1TV': fixed weights/ordering, higher persistence and smaller update average needs across the cross-section

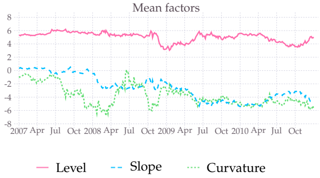


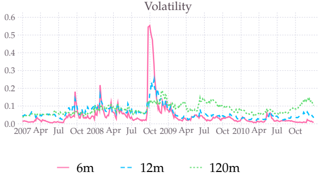
`GAS t pTV': switching ordering, lower persistence and larger updates



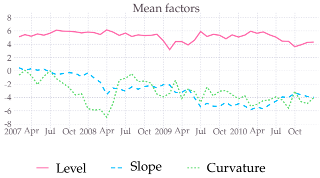
Empirical application, U.S., data frequency

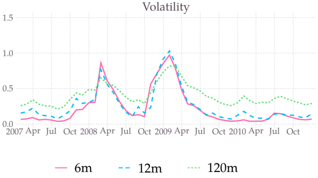
Daily observations, $\nu = 26.64$





Monthly observations, $\nu = 18.04$





Empirical application, U.S., data frequency

Empirical application, U.S., datasets

Conclusions

- It is hard to beat the simple model for short-term forecasting.
- But we can gain new insights about the nature of changes in the yield curve from relaxing some of the assumptions that would be very hard to relax with a parameter-driven model.
- At daily frequency, allowing for time-varying variances improves fit the most and uncovers additional short-term risk factors.
- At lower frequencies, misfit due to the filters' lagging requires fat tailed distributions.
- Performance improvement is visible mostly for shorter maturities.
- We find that SDM are able to fit the data very well (up to 30% improvement in likelihoods and informational content).
- For complicated models implementation of SDMs offers considerable speed and implementation benefits.
- Conjecture: factors extracted with some SDMs are less likely to suffer from over-fitting and can be used in subsequent macro/finance analyses.