NBP Narodowy Bank Polski

NBP Working Paper No. 258

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

Siem Jan Koopman, André Lucas, Marcin Zamojski

NBP Working Paper No. 258

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

Siem Jan Koopman, André Lucas, Marcin Zamojski

Economic Institute
Warsaw, 2017

Siem Jan Koopman – Vrije Universiteit Amsterdam and Tinbergen Institute
André Lucas – Vrije Universiteit Amsterdam and Tinbergen Institute
Marcin Zamojski – University of Gothenburg

We are grateful to Narodowy Bank Polski for financial support. Koopman acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78) at Aarhus University, Denmark, and funded by the Danish National Research Foundation. Koopman and Lucas are grateful for support from the European Union Seventh Framework Programme (FP7-SSH/2007-2013, grant agreement 320270 – SYRTO). All numerical results presented in this paper were obtained with OxMetrics 7 or Julia at the Dutch National Cluster (LISA).

This research project was conducted under the NBP Economic Research Committee's open competition for research projects to be carried out by the NBP staff and economists from outside the NBP and was financed by the NBP.

Published by:
Narodowy Bank Polski
Education & Publishing Department
ul. Świętokrzyska 11/21
00-919 Warszawa, Poland
phone +48 22 185 23 35
www.nbp.pl

ISSN 2084-624X

Contents

Abstract

We consider score-driven time-varying parameters in dynamic yield curve models and investigate their in-sample and out-of-sample performance for two data sets. In a univariate setting, score-driven models were shown to offer competitive performance to parameter-driven models in terms of in-sample fit and quality of out-of-sample forecasts but at a lower computational cost. We investigate whether this performance and the related advantages extend to more general and higher-dimensional models. Based on an extensive Monte Carlo study, we show that in multivariate settings the advantages of score-driven models can even be more pronounced than in the univariate setting. We also show how the score-driven approach can be implemented in dynamic yield curve models and extend them to allow for the fat-tailed distributions of the disturbances and the time-variation of variances (heteroskedasticity) and covariances.

Keywords: term-structure, dynamic Nelson-Siegel models, non-Gaussian distributions, time-varying parameters, observation-driven models, parameter-driven models.

JEL classification codes: C15, C32, C33, C58, C63, E43, E52, E58.

1. Introduction

We study the in-sample and the out-of-sample performance of dynamic yield curve models for which the time-varying parameters are formulated either as observation-driven or as parameter-driven models. Our main focus is on the parsimonious yield curve model of Nelson and Siegel (1987) which is extended within a dynamic setting by Diebold and Li (2006) and Diebold, Rudebusch, and Aruoba (2006); it is called the dynamic Nelson-Siegel model (DNSM). The model describes the yield curve as a linear function of three latent factors referred to as level, slope, and curvature. These factors have an impact on the shape of the yield curve. For reasons of tractability, it is usually assumed that (i) the yield is a linear function of the factors, (ii) the disturbances for each interest rate equation are Gaussian with a constant variance, and (iii) that the disturbances are serially and mutually uncorrelated. Under these strict assumptions, we can consider a linear Gaussian state space formulation for the DNSM and we can estimate its static parameters by exact maximum likelihood methods based on the Kalman Filter; see Diebold, Rudebusch, and Aruoba (2006). The parsimony of the specification enables us to fit the DNSM using low-frequency, monthly data while maintaining an acceptable ratio of observations versus parameters. This is a useful feature in many empirical settings where good quality high-frequency data may be scarce.

The strict assumptions of normally distributed, uncorrelated, and homoskedastic residuals may, however, be harder to justify for data observed at a higher frequency, for example, daily data. Similarly, these assumptions are likely to be violated in times of economic unrest, such as during the financial crisis of 2007–2008. For instance, Koopman, Mallee, and Van der Wel (2010) and Mesters, Schwaab, and Koopman (2014) show that volatilities of pricing errors are indeed time-varying even at monthly or weekly frequencies and that the inclusion of only a common time-varying volatility factor improves the fit of the models considerably. Even though we focus on the daily data, our robustness analysis

confirms these results as we find evidence for both conditional heteroskedasticity and fat-tailedness of pricing errors for monthly and weekly data as well. When we leave the linear, homoskedastic Gaussian model assumptions, however, we typically face a substantially more complicated estimation challenge. For example, Koopman, Mallee, and Van der Wel (2010) and Mesters, Schwaab, and Koopman (2014) adopt simulation-based estimation methods, such as Monte Carlo importance sampling, to obtain the maximum likelihood estimates.

In our study, we consider a different class of models with the aim to describe the term structure dynamics more accurately in terms of the empirical, stylized properties of interest rate data observed at higher (daily) frequencies while keeping the estimation process feasible. In particular, we adopt the class of score-driven models as introduced by Creal, Koopman, and Lucas (2011, 2013); see also Harvey (2013).[1] An important advantage of score-driven models is that they are a sub-class of observation-driven models in the classification of Cox (1981). Therefore, an explicit, tractable expression for the likelihood function is available which facilitates the estimation of the static parameters in these models. As a result, estimation and inference by standard maximum likelihood procedures are relatively straightforward to implement and to execute. Moreover, the framework easily allows for further generalisations of the yield curve model, as long as the conditional observation density of the cross-section of yields remains tractable.

We consider a sequence of score-driven models that gradually relaxes the restrictive distributional assumptions discussed above. Specifically, we extend the model with time-varying volatility structures, possibly driven by common components. Furthermore, apart from having disturbances from a multivariate Gaussian distribution, we also allow disturbances to come from fat-tailed distributions, namely from the multivariate Student's t distribution. Overall, we consider 13 competing specifications from which ten specifications belong to the

[1] For a more complete compilation of recent work on score-driven models, we refer to the website gasmodel.com.

class of score-driven models and three specifications are within the parameter-driven framework. Koopman, Lucas, and Scharth (2016) show in a univariate context that the performance of score-driven models is competitive in terms of in-sample fit and quality of out-of-sample forecasts when compared to that of parameter-driven models, for a range of data-generating processes. Their results surprisingly continue to hold if the parameter-driven model is the true data-generating process. At the same time, however, all of their results are restricted to the univariate setting. Using our current set-up and framework, we are able to verify whether these results continue to hold in a multivariate context. In fact, we find that the advantages in the multivariate context may even be larger in certain cases than in the univariate setting.

In the empirical part of the paper, we consider all 13 models for the analysis of two daily interest rate datasets. The first is the typical U.S. data used in the literature. We also model interest rates in Poland which has a less developed financial market. The modelling of this data creates some interesting questions and challenges. For one, the market in Poland is less liquid so the time-varying volatility is more pronounced and periods of high pricing errors are prolonged. Furthermore, the sample length is much shorter than for U.S. or for other (developed) European economies which is part of the motivation to use higher frequency data. Our overall conclusion from the analysis is that by increasing the complexity of the model, we obtain considerable improvements in the fit. Of all extensions in our study, the largest improvements in fit are obtained by the extension that allows for the heteroskedasticity in pricing errors. Also, by allowing for multiple volatility factors, we provide new insight into what drives the dynamics of the yield curve in high-volatility regimes. This is in line with the parameter-driven results of Koopman, Mallee, and Van der Wel (2010). The benefits of allowing for fat-tailed conditional distributions are much smaller once heteroskedasticity is present in the model specification.

To quantify the potential gains from gradually increasing model complexity, we also conduct an extended Monte Carlo study. In a simulation study, the true values of the dynamic parameters are set a-priori. We can therefore evaluate all models in terms of the root-mean-square error (RMSE) and mean absolute percentage error (MAPE) for the whole sample. Based on such simulations, we find that the impact of the model extensions depends heavily on the time-to-maturity of the yield. Having more features in the model appears to benefit the short maturity yields mostly, with an improved in-sample and out-of-sample fit. For the long-term maturities, simple models that do not incorporate all relevant features appear to perform as accurately as their more complex counterparts. Differences in in-sample fit only occur during low-volatility and high-volatility regimes. For multiple-period-ahead forecasts we observe a gradual decay in relative performance of the simple models with the short-term maturities being especially affected. Despite their comparable performance, the extended models still have the advantage of offering valuable insights into what drives the changes in short-term interest rates and how likely these changes are to persist.

The remainder of this paper is organised as follows. In Section 2 we distinguish between parameter-driven and observation-driven models. We present the general specification of the dynamic Nelson-Siegel model and the generalisations used in Section 3. There, we also highlight the different estimation approaches used, i.e., for the parameter-driven and score-driven models. In Section 4, we present the results of a large-scale Monte Carlo study in which we assess the performance of our models. An empirical study of U.S. and Polish interest rates is presented in Section 5. Finally, Section 6 concludes.

2. Score-driven and parameter-driven methods

To introduce score-driven models and distinguish them from parameter-driven models, consider a time-series model of an observed time series $y_1, \ldots, y_T$ characterised by the conditional density

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

where y_t is an $N \times 1$ vector and $Y_{t-1} = \{y_1, \ldots, y_{t-1}\}$ is the information set available to the agent just before time t. The conditional observation density $p(\cdot \mid Y_{t-1}, \boldsymbol{f}_t; \boldsymbol{\theta})$ also depends on two parameter vectors: $\boldsymbol{\theta}$ and $\boldsymbol{f}_t$, which lie in the parameter spaces Θ and $\mathcal{F}$, respectively. The parameters grouped in $\boldsymbol{f}_t$ are assumed to be time-varying with possibly unknown dynamics that may also depend on the static parameters $\boldsymbol{\theta}$. At this level of generality, both the dynamic and static parameters can enter the model eq. (1) non-linearly. For instance, in Creal, Koopman, Lucas, and Zamojski (2015) I model time-varying relative risk aversion in the consumption capital asset pricing model with the standard constant relative risk aversion (CRRA) utility function.

Depending on how the dynamics of the time-varying parameter are specified, we distinguish parameter-driven and observation-driven models as in Cox (1981). In parameter-driven models we have $\boldsymbol{f}_{t+1} = \phi\left(\zeta_t, \boldsymbol{f}_t\right)$ for some function $\phi\left(\cdot\right)$, where ζ_t is a random innovation given the information set Y_{t-1}. In the parameter-driven setting, observations and the latent factors can evolve independently and the transition equations of the dynamic parameters are 'parameter-driven' since we need to impose distributional assumptions on $\boldsymbol{f}_t$ as well. A typical example of a parameter-driven model is the stochastic volatility model; see Shephard (2005) for an overview. Parameter-driven models are of special interest in structural modelling as they allow one to independently study the effects of shocks to different time-varying parameters; see Fernández-Villaverde and Rubio-Ramírez (2013) for a discussion on the benefits of incorporating stochas-

tic volatility in dynamic stochastic general equilibrium models. More generally, many parameter-driven models can be formulated in state space form with the dynamic parameter $\boldsymbol{f}_t$ treated as a latent state vector. For the purpose of modelling the term-structure of interest rates with the DNSMs, we consider a linear dynamic specification for $\boldsymbol{f}_t$,

$$\boldsymbol{f}_{t+1} = \delta + \Phi \boldsymbol{f}_t + \eta_t, \qquad \eta_t \sim N\left(0, \boldsymbol{\Sigma}\right). \tag{2}$$

Estimation of parameter-driven models may be cumbersome since the state variables need to be 'integrated out' from the joint density of observations and latent factors. If the necessary conditional distributions are not readily available in closed form, obtaining maximum likelihood estimators for $\boldsymbol{\theta}$ requires the use of computationally intensive simulation methods; see Durbin and Koopman (2012) for a general discussion.

By contrast, in observation-driven models in the sense proposed by Cox, the dynamic parameters $\boldsymbol{f}_t$ in eq. (1) vary over time as a function of lagged dependent variables, lagged exogenous variables, and the static parameters, $\boldsymbol{\theta}$. In other words, these models only have a single source of error which stems from the conditional distribution in eq. (1). Formally we write, $\boldsymbol{f}_{t+1} = \phi\left(Y_t, \boldsymbol{f}_1; \boldsymbol{\theta}\right)$. And in particular, we may have that $\boldsymbol{f}_{t+1} = \phi\left(y_t, \boldsymbol{f}_t; \boldsymbol{\theta}\right)$ if $\{y_t\}$ come from a Markov process. A typical example of an observation-driven model is the family of generalised autoregressive conditional heteroskedasticity (GARCH) models originated by Engle (1982) and Bollerslev (1986). In this class of models, values of dynamic parameters are perfectly predictable one-step-ahead given past information. This feature of observation-driven models makes evaluation of the likelihood straightforward through a direct prediction error decomposition.

For observation-driven models, the usual treatment imposes a linear, autoregressive updating function for the dynamic parameter,

$$\boldsymbol{f}_{t+1} = \boldsymbol{\omega} + \boldsymbol{B}\left(\boldsymbol{f}_t - \boldsymbol{\omega}\right) + \boldsymbol{A} s_t(y_t, \boldsymbol{f}_t; \boldsymbol{\theta}), \tag{3}$$

where $\boldsymbol{\omega}$, $\boldsymbol{A}$, and $\boldsymbol{B}$ are part of the static parameter vector $\boldsymbol{\theta}$. In this specification, $\boldsymbol{\omega}$ captures the unconditional mean, $\boldsymbol{B}$ determines the persistence, and $\boldsymbol{A}$ the learning rate. The function $s_t(\cdot)$ can be chosen in a flexible way. Depending on the choice of s_t we obtain different classes of observation-driven models. Creal, Koopman, and Lucas (2013) and Harvey (2013) propose a scaled first derivative of the local criterion function as a choice for s_t. This is also the approach that is adopted in Creal, Koopman, Lucas, and Zamojski (2015) who work within the method of moments framework. Due to this we refer to this class of models as score-driven models. In particular, the generalised autoregressive score (GAS) method of Creal, Koopman, and Lucas (2013) sets

$$s_t = S_t \nabla_t, \qquad \nabla_t = \frac{\partial p\left(y_t | \boldsymbol{f}_t, Y_t; \boldsymbol{\theta}\right)}{\partial \boldsymbol{f}_t^{\intercal}}, \qquad S_t = \mathcal{I}_{t|t-1}^{-k}, \tag{4}$$

where $\mathcal{I}_{t|t-1} = \mathrm{E}[\nabla_t \nabla_t^{\intercal}] = -\mathrm{E}[{}^{\partial^2 p\left(y_t | \boldsymbol{f}_t, Y_t; \boldsymbol{\theta}\right)}/_{\partial \boldsymbol{f}_t^2}]$ is the conditional Fisher information matrix and $k = \{0, {}^1/_2, 1\}$. This choice of the functional specification of s_t to drive the dynamic parameters is far from arbitrary. If $k = 1$ and $\boldsymbol{B} = \boldsymbol{I}$, eq. (3) reduces to a scaled Gauss-Newton update in the time dimension; if $k = 0$ (unit scaling) we take steepest ascent updates. Note that in practice, it is often preferable to avoid unit scaling as it may lead the filter to over-react to outliers. Furthermore, if $\boldsymbol{f}_t$ is indeed time-varying, it is intuitive to temporarily adjust the value of $\boldsymbol{f}_t$ if observations persistently signal that the model locally does not fit the data. By improving the fit in the neighbourhood of the most recent observation, the score-driven approach tries to improve the fit of the model in other regions of the sample space as well. This also means that score-driven models are more intuitively thought of as filters rather than as data-generating processes. In the case of modelling volatility with a GAS model, Nelson (1992) shows that these filters can produce consistent estimates of paths of time-varying parameters even if the filter (or model) is not the data-generating process.

Score-driven models encompass a wide range of well-known and popular observation-driven models; see Creal, Koopman, and Lucas (2013) for an overview.

Considerable work has been done in recent years to establish the theoretical properties of score-driven models. In particular, Blasques, Koopman, and Lucas (2015) show that generalised autoregressive score updates improve the local Kullback–Leibler divergence between the true data density and the model density. They further show that any observation-driven update with similar optimality properties needs to be 'score-equivalent'. The properties of the maximum likelihood estimator are considered in Blasques, Koopman, and Lucas (2014).

Koopman, Lucas, and Scharth (2016) show that for univariate models with time-varying parameters, score-driven models offer similar predictive accuracy as parameter-driven models. This is notable, as in many cases estimation of parameter-driven models is more involved and may be less stable.[2] Based on Koopman, Lucas, and Scharth (2016) and Blasques, Koopman, and Lucas (2015) there is now clear theoretical and empirical evidence that univariate score-driven models outperform other observation-driven models. It is thus interesting to see if score-driven models perform similarly well in a multivariate setting. To that end, modelling the term structure of interest rates offers a great challenge to the score-driven method. Although conceptually simple, the DNSMs we consider in this paper involve time-varying parameters both in means and variances, and consider both multivariate Gaussian and multivariate Student's t distributions.

[2]For instance, when estimation is done through importance sampling, stability may be affected by the choice of the importance density.

3. Analysing score-driven dynamic Nelson-Siegel models

We present a general formulation of the dynamic Nelson-Siegel model of interest rates within the setting of score-driven models. Various extensions of the baseline model that vary in complexity are considered. We start with the baseline model and gradually relax the assumptions, in particular the restrictive distributional assumptions. We treat ten model specifications within the setting of score-driven models and three specifications within the setting of parameter-driven models. These last three models are mainly adopted as benchmarks since our focus is on the score-driven models.

The dynamic Nelson-Siegel model is due to Diebold and Li (2006) and Diebold, Rudebusch, and Aruoba (2006) who modify the parsimonious model of Nelson and Siegel (1987) as

$$y_{it}(\tau_i) = f_{1,t} + f_{2,t}\left(\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i}\right) + f_{3,t}\left(\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i}\right) + \varepsilon_{it}, \qquad (5)$$

where $y_{it}(\tau_i)$ denotes the interest rate (yield) for maturity τ_i observed at time t . The ε_{it} are assumed to be drawn independently from a Gaussian distribution. The Nelson-Siegel model expresses the yield curve as a non-linear function of the parameter λ and an affine transformation of three other factors, $\boldsymbol{f}_t = (f_{L,t}, f_{S,t}, f_{C,t})$. Diebold and Li (2006) also interpret the three factors as the level, slope, and curvature of the yield curve. This is because the long-term interest rates are explained only by the level factor, $f_{L,t}$, since the loadings on other factors tend to 0 as the maturity τ_i increases. The other two factors correspond to the short-term and medium-term interest rates. The parameter λ defines what constitutes a 'medium-term' yield in this model as it specifies for which maturity the loading on $f_{C,t}$ is maximised. All four parameters can be estimated in a static setting to fit a single cross-section of interest rates. The static Nelson-Siegel model is still routinely employed by central banks and other financial market practitioners on a daily basis; see for instance Christensen, Diebold, and Rudebusch (2011).

Diebold, Rudebusch, and Aruoba (2006) further show that it is possible to model the evolution in time of the latent factors by assuming that they evolve stochastically and follow a vector autoregressive process of order one. By assuming normality and homoskedasticity of disturbances they recast the full model into a linear Gaussian state space model, where the latent factors follow a vector autoregressive process as in eq. (2). The statistical treatment for this model, including likelihood evaluation, can be based on the Kalman Filter and related algorithms.

3.1. Single volatility factor

Many extensions of the baseline model of Diebold, Rudebusch, and Aruoba (2006) are considered and treated in the literature. Each extension or generalisation typically aims to relax one or more assumptions underlying the baseline model. For example, Koopman, Mallee, and Van der Wel (2010) introduce a conditional heteroskedasticity specification for the measurement disturbances and specify a stochastically time-varying process for the rate of decay parameter λ. Mesters, Schwaab, and Koopman (2014) allow the disturbances to come from a fat-tailed distribution with stochastically time-varying variances. For both of these extensions, the volatility is supposed to be driven by a single, common factor.

We consider these cases and specifically follow Mesters, Schwaab, and Koopman in modelling log-volatility as a time-varying process using score-driven dynamics as in eq. (3). We define the following DNSM with a time-varying log-volatility factor

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim t\left(0, \boldsymbol{\Sigma}, \nu\right) \\ \boldsymbol{\Lambda} &= \begin{bmatrix} \mathbf{1} & \frac{1-e^{-\lambda\tau}}{\lambda\tau} & \frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \end{bmatrix}, & \boldsymbol{\Sigma} &= \operatorname{diag}\left(\sigma_1^2 e^{h_t}, \ldots, \sigma_p^2 e^{h_t}\right), \end{aligned} \tag{6}$$

where $\boldsymbol{y}_t$ is a $p \times 1$ vector of interest rates. For expositional purposes, we assume that the disturbances are drawn from a Student's t distribution with ν degrees of freedom and scale matrix $\boldsymbol{\Sigma}$. The results for the Gaussian distribution follow by considering the limiting case as ν tends to infinity. Depending on the distributional assumptions we label the two models as '`GAS t 1TV`' and '`GAS N 1TV`' to reflect the type of distribution (Normal, '`N`', or Student's t, '`t`') as well as the number of volatility factors ('`1TV`'; one time-varying). To identify the latent volatility factor, it is necessary to fix one of the loadings σ_i^2 at 1. The volatility factor is then scaled up or down for the remaining maturities. We impose the normalisation restriction on the one-year maturity. Hence we set $\sigma_{12m}^2 = 1$. This choice does not affect the results of our study.

The score and Fisher information matrix in this case are given by

$$\nabla_{\boldsymbol{f}_t} = \nu_p \left[\nu + k_t\right]^{-1} \boldsymbol{\Lambda}^\intercal \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t, \qquad \mathcal{I}_{ff} = \nu_p \left[\nu_p + 2\right]^{-1} \boldsymbol{\Lambda}^\intercal \boldsymbol{\Sigma}^{-\mathbf{1}} \boldsymbol{\Lambda}, \tag{7}$$

$$\nabla_{\boldsymbol{h}_t} = \frac{1}{2}\left(\nu_p \left[\nu + k_t\right]^{-1} k_t - p\right), \qquad \mathcal{I}_{hh} = \frac{1}{2}\left[\nu_p + 2\right]^{-1} \nu p, \tag{8}$$

where $k_t = \boldsymbol{\varepsilon}_t^\intercal \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t$, $\nu_p = \nu + p$, and $\mathcal{I}_{fh} = 0$. We provide details of the derivation in Appendix B. For the normal distribution, this simplifies to

$$\nabla_{\boldsymbol{f}_t} = \boldsymbol{\Lambda}^\intercal \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t, \qquad \mathcal{I}_{ff} = \boldsymbol{\Lambda}^\intercal \boldsymbol{\Sigma}^{-\mathbf{1}} \boldsymbol{\Lambda}, \tag{9}$$

$$\nabla_{\boldsymbol{h}_t} = \frac{1}{2}\left(k_t - p\right), \qquad \mathcal{I}_{hh} = \frac{p}{2}. \tag{10}$$

For the estimation process, we impose block diagonality on the static parameter matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ in the score transition eq. (3). Hence we do not capture spillovers from the volatility factors to the level equation, or vice versa. Although, as is shown in Creal, Koopman, Lucas, and Zamojski (2015), block diagonality of the $\boldsymbol{A}$ matrix can be relaxed to allow for leverage effects, in this study we have opted for sparsity to facilitate easier comparisons between the different models.

As a benchmark we consider a parameter-driven specification of the model in eq. (6). Given that the model is both non-Gaussian and non-linear we estimate

it sequentially using importance sampling methods as in Mesters, Schwaab, and Koopman (2014). To that end, we follow Shephard and Pitt (1997) and Durbin and Koopman (1997) and apply Kalman filtering and smoothing methods to an approximating linear Gaussian state space model. The paths of time-varying mean and volatility factors are obtained through signal extraction, see Durbin and Koopman (2012) for details. We refer to this model as '`IS t 1TV`'.

3.2. Individual volatility factors and dense covariance matrices

In a parameter-driven framework, increasing the size of the state vector to have more volatility factors is computationally demanding. Even with a single volatility factor, the model is both non-Gaussian and non-linear in the state variables. It implies that the log-likelihood function—for a given parameter vector—needs to be evaluated using Monte Carlo methods such as importance sampling. Modelling multiple volatilities implies that we cannot easily perform the integration sequentially as in Mesters, Schwaab, and Koopman (2014), since we have a multivariate state (or signal) and typically analyse long time series. On the one hand, the numerically accelerated importance sampling method of Koopman, Lucas, and Scharth (2015), which is both computationally and numerically efficient in larger samples, requires a univariate signal for a feasible implementation. On the other hand, more feasible Monte Carlo estimation methods as developed by Shephard and Pitt (1997) and Durbin and Koopman (1997) for a multivariate signal, do not necessarily perform well in long samples.

In contrast, extensions to non-normal distributions, time-varying yield curve coefficients $\boldsymbol{f}_t$, and time-varying volatilities are relatively easy to achieve in score-driven models. We extend the multivariate heavy-tailed model for time-varying volatilities and correlations of Creal, Koopman, and Lucas (2011) to also account for the latent level, slope, and curvature in the mean. For this purpose we consider the model

$$\boldsymbol{y}_t = \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim t\left(0, \boldsymbol{\Sigma}, \nu\right) \tag{11}$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \mathbf{1} & \frac{1-e^{-\lambda\tau}}{\lambda\tau} & \frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t, \tag{12}$$

where $\boldsymbol{D}_t$ is the time-varying, diagonal standard deviation matrix and $\boldsymbol{R}$ is the static correlation matrix. The ith diagonal element of $\boldsymbol{D}_t^2$ is defined as d_{it}^2 and is specified by $d_{it}^2 = e^{h_{it}}$ where h_{it} is a time-varying process. Hence we have imposed a positive standard deviation by modelling log-volatility directly as in the previous section. To obtain further parsimony, we limit our attention to two settings. First, we assume that the disturbances are uncorrelated, that is $\boldsymbol{R} = \boldsymbol{I}$. The resulting model as indicated by '`GAS t pTV`' (Student's t-distributed errors and p volatility factors). Second, we relax the independence assumption and assume an equicorrelation structure for the correlation matrix, that is $\boldsymbol{R} = (1-\rho)\boldsymbol{I} + \rho\boldsymbol{\iota}\boldsymbol{\iota}^\intercal$. The off-diagonal parameter $-(p-1)^{-1} < \rho < 1$ can be estimated jointly with the other static parameters in $\boldsymbol{\theta}$. We refer to this model as '`GAS t pTV+c`'. We obtain two similar models ('`GAS N pTV`' and '`GAS N pTV+c`') by replacing the Student's t distribution in eq. (11) by a multivariate Normal distribution with covariance matrix $\boldsymbol{\Sigma}$.

We provide the score and Fisher information matrices for the scale matrix $\boldsymbol{\Sigma}$. These results can be adopted directly when the disturbances are uncorrelated. Only minor modifications are needed to account for the equicorrelation matrix. For details we refer to Creal, Koopman, and Lucas (2011). We have

$$\nabla_{\boldsymbol{f}_t} = \bar{\nu}\boldsymbol{\Lambda}^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t, \tag{13}$$

$$\mathcal{I}_{ff} = 2\tilde{\nu}\boldsymbol{\Lambda}^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}, \tag{14}$$

$$\nabla_{\mathrm{v}(\Sigma)} = \mathcal{K}\left(\frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{\Psi}^\intercal(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\,\mathrm{vec}(\bar{\nu}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\intercal - \boldsymbol{\Sigma})\right), \tag{15}$$

$$\mathcal{I}_{\mathrm{v}(\Sigma)\,\mathrm{v}(\Sigma)} = \mathcal{K}(\boldsymbol{\Sigma})\,\mathcal{K}(\boldsymbol{\Omega} + \boldsymbol{b}\boldsymbol{c}^\intercal)\,\mathcal{K}(\boldsymbol{\Sigma}), \tag{16}$$

where $\mathrm{v}(\Sigma)$ denotes the half vectorization of $\boldsymbol{\Sigma}$ and only contains entries from the lower triangle, $\otimes$ denotes the Kronecker product, $\boldsymbol{\Psi}$ is the duplication matrix so that $\boldsymbol{\Psi}\,\mathrm{v}(\Sigma) = \mathrm{vec}(\boldsymbol{\Sigma})$, see Abadir and Magnus (2005) for further details, $\boldsymbol{\Psi}_+ = (\boldsymbol{\Psi}^\intercal\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^\intercal$, $\nu_p = \nu + p$, $k_t = \boldsymbol{\varepsilon}_t^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t$, $\bar{\nu} = \nu_p[\nu + k_t]^{-1}$, $\tilde{\nu} = \nu_p[2(\nu_p +$

$2)]^{-1}$, $\boldsymbol{\Omega} = \tilde{\nu}\boldsymbol{\Psi}^{\intercal}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\boldsymbol{\Psi}$, $\boldsymbol{\Omega}^{-1} = \tilde{\nu}^{-1}\boldsymbol{\Psi}_{+}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\boldsymbol{\Psi}_{+}^{\intercal}$, $\boldsymbol{c} = \boldsymbol{\Psi}^{\intercal}\operatorname{vec}(\boldsymbol{\Sigma}^{-1})$, $\boldsymbol{b} = [2(\nu_p + 2)]^{-1}\boldsymbol{c}$. The matrix function $\mathcal{K}(\cdot)$ selects rows (and columns) that correspond to the position of diagonal entries of $\boldsymbol{\Sigma}$ in its half vectorization.[3] The three resulting matrices in eq. (16) are square and it is possible to compute the inverse Fisher information efficiently by applying the Sherman–Morrison formula. For $\bar{\nu} = 1$ and $\tilde{\nu} = \frac{1}{2}$, we obtain scores and the Fisher information for the Gaussian distribution.

Although we model the dynamics of the mean factors and the volatility factors jointly, we do not allow for interactions between them. Thus, and as in Section 3.1, we impose block diagonality on the matrix coefficients $\boldsymbol{A}$ and $\boldsymbol{B}$ in the transition eq. (3). See Appendix D for details on the full structure of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. We model the blocks corresponding to the dynamics of the regression parameter $\boldsymbol{f}_t$ in both matrices as dense. The blocks corresponding to the volatility dynamics (h_{it}) are not diagonal, as is typically done; see Creal, Koopman, and Lucas (2011). Instead, we allow spillovers from the volatility dynamics from one maturity into other maturities. To avoid an excessive number of parameters, we restrict the off-diagonal elements of the blocks in $\boldsymbol{A}$ and $\boldsymbol{B}$ corresponding to the h_{it}s to be the same, resulting in only 2 additional parameters compared to the diagonal case. Note, that these off-diagonal elements would be estimated close to 0 if there is no commonality in the volatilities.

These are the most complicated model specifications under consideration. Nevertheless, further extensions are still possible and easy to implement given that the Fisher information matrix for multivariate t distribution is well documented, see (Mitchell, 1989; Arellano-Valle, 2010; Creal, Koopman, and Lucas, 2011). In particular, it is possible to add time-varying skewness or to let the dependence parameter ν change over time. Furthermore, inclusion of macroeconomic variables to directly measure impact of the monetary policy is also feasible. In this case, the mean and/or variance factors would be affected by changes in

[3] Given that we assume no correlation between disturbances, all other rows and columns contain only zeros.

the monetary policy either directly or through the score and the matrix $\boldsymbol{A}$ in eq. (3), see Creal, Koopman, Lucas, and Zamojski (2015).[4] It is also possible to incorporate no-arbitrage into our score-driven Nelson-Siegel models by extending them in a similar fashion to Christensen, Diebold, and Rudebusch (2011). We leave such extensions for future work.

3.3. Homoskedasticity

The dynamic Nelson-Siegel models are not well grounded in economic theory or in finance. In particular, the most commonly used specifications do not exclude arbitrage opportunities. Note, however, that there exist extensions to the standard Nelson-Siegel model which incorporate the no-arbitrage assumption, see, e.g., Filipović (1999); Diebold, Piazzesi, and Rudebusch (2005) or Christensen, Diebold, and Rudebusch (2011). In the yield curve literature DNSMs are regarded as misspecified models. It is unlikely that the model provides an appropriate description of the data-generating process. Still, the DNSM is highly popular among practitioners. This is probably due to its empirical performance. Even the most simple version of the DNSM produces good empirical results despite the assumptions of normally distributed errors with static diagonal variance matrices.

Our most basic specification is obtained when errors for all maturities are normally independently distributed with a unit scale matrix, that is $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$. This elementary version of the DNSM is considered in Durbin and Koopman (2012) and we label it as '`1C`' (one, constant volatility factor). Depending on the distributional assumption and the estimation method we have '`GAS t 1C`', '`GAS N 1C`', and '`KF 1C`'; where model '`KF 1C`' is Gaussian and estimated with the Kalman Filter.

[4] In Creal, Koopman, Lucas, and Zamojski (2015), I propose to use this approach to easily incorporate the leverage effect in their score-driven GMM model with conditional heteroskedasticity. The leverage effect is captured by allowing the time-varying volatility to react to the scores of the static parameters.

Table 1

Models

This table presents a summary view of the models we consider in this paper. PD and SD denote parameter-driven and score-driven, respectively. Number of parameters is given under the assumption that 11 maturities are modelled.

Symbol	Type	Estimated with	Parameters	Fat tails	Heteroskedasticity	Correlation
'KF 1C'	PD	Kalman Filter	23			
'KF pC'	PD	Kalman Filter	33			
'IS t 1TV'	PD	Importance sampling	33	✓[5]	✓	
'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	✓	✓	✓

We also consider DNSMs with $\mathbf{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2)$ which we refer to as 'GAS t pC', 'GAS N pC', and 'KF pC'. In case of normal errors and parameter-driven dynamics for the latent mean factors (i.e. 'KF pC'), the model reduces to the specification of Diebold, Rudebusch, and Aruoba (2006). We use 'KF pC' as our primary benchmark model. Under the assumption of normal errors, the two models can be treated by the Kalman Filter without violating its optimality properties.

Table 1 presents a summary overview of all thirteen model-method combinations which we consider in our current study.

[5]When estimating, we set the degrees of freedom at 50 in order to ensure stability of the algorithm. Similar restrictions are also imposed by Mesters, Schwaab, and Koopman (2014).

4. Simulation study

In this section we present results of a considerable Monte Carlo study. We investigate how well the different model specifications perform depending on the underlying data-generating process.

4.1. Design of the study

We consider 12 different data-generating processes which reflect the diversity of the models in Section 3. In particular, we vary if (and how many) volatility factors are part of the specification. In the simulated world, yields are generated by one of the following models: (i) a Gaussian state-space model, (ii) a non-Gaussian state-space model, (iii) Gaussian errors in the observation equations combined with deterministic paths for the factors, and (iv) same as (iii) but with fat-tailed disturbances.

To limit our attention to relevant scenarios, we simulate data based on the empirical results we discuss later in Section 5. To that end, in the 'deterministic paths' DGPs we use the filtered paths that we obtained for the U.S. data between 2007 and 2012. We specifically choose this period to include both a period of relative calm and the financial crisis. Other parameters are also in line with what we see in the empirical applications. Through this design we can both compare the performance of different models and moreover investigate whether the empirical results in Section 5 are biased in any way. We label the set of DGPs (iii) and (iv) as '`Det`'.

We also consider a number of state-space data-generating processes. For this set of DGPs, parameter values are adopted from results obtained with the Kalman Filter ('`KF pC`') for the mean factors. In some cases, we introduce heteroskedasticity with a common volatility factor.

Note that in all cases, the GAS models are misspecified in the sense that, at the very least, the GAS transition equation is not correctly specified vis-à-vis

Table 2

Data-generating processes

The table contains a summary of all data-generating processes which we consider in the simulation study. Degrees of freedom and the value of the true correlation parameter are in parentheses when applicable. Other parameter values and paths of time-varying parameters are based on the results we obtain in Section 5 for the U.S. data with the 'reference model'.

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

the data-generating process. By contrast, we do consider cases for which the parameter-driven models are correctly specified and can be expected to perform well. A summary of the data-generating processes with their reference models is shown in Table 2.

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. The estimation of our models is based on the first 1,000 observations for which we compute measures of in-sample fit. We then use the remaining 50 observations to assess the models' out-of-sample and forecasting performance.

4.2. In-sample performance

We start our analysis by looking at the models' in-sample performance using model selection criteria. In Figure 1 and Table 3 we compare different method-

Table 3

AIC for different models

The table contains median values of AIC relative to values achieved by 'KF pC'. Higher is better. Cases where AIC is higher than for the benchmark model are in bold.

AIC	'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'	'IS t 1TV'	'KF pC'	'KF 1C'	'KF pC' value
'Det t pTV+c'	**1.23**	**1.20**	1.00	0.95	0.94	0.90	0.86	0.82	0.72	0.66	**1.25**	1.00	0.91	-23328.67
'Det N pTV+c'	**1.22**	**1.20**	0.99	0.95	0.93	0.90	0.85	0.81	0.72	0.66	**1.21**	1.00	0.91	-23899.32
'Det t pTV'	**1.26**	**1.23**	**1.15**	**1.13**	**1.09**	**1.08**	0.98	0.94	0.86	0.78	**1.25**	1.00	0.91	-20499.55
'Det N pTV'	**1.46**	**1.39**	**1.33**	**1.24**	**1.23**	**1.16**	**1.08**	**1.03**	0.82	0.07	**1.42**	1.00	0.21	-16977.44
'Det t 1TV'	**1.20**	**1.20**	**1.09**	**1.09**	**1.09**	**1.08**	0.95	0.90	0.86	0.80	**1.26**	1.00	0.93	-20381.89
'Det N 1TV'	**1.20**	**1.20**	**1.09**	**1.07**	**1.08**	**1.06**	0.94	0.89	0.84	0.78	**1.25**	1.00	0.93	-19799.62
'Sto N 1TV'	**1.13**	**1.12**	1.00	0.99	**1.01**	0.99	0.97	0.92	0.89	0.83	**1.23**	1.00	0.93	-21173.79
'Sto t30 1TV'	**1.14**	**1.13**	**1.02**	1.00	**1.02**	1.00	0.99	0.93	0.90	0.83	**1.24**	1.00	0.93	-20520.13
'Sto t50 1TV'	**1.13**	**1.13**	1.00	1.00	**1.01**	1.00	0.98	0.93	0.90	0.83	**1.24**	1.00	0.93	-20790.07
'Sto t30 1C'	0.97	0.96	0.87	0.86	0.87	0.87	0.87	0.87	0.87	0.86	**1.02**	1.00	1.00	-25552.36
'Sto t50 1C'	0.97	0.96	0.87	0.86	0.87	0.87	0.87	0.87	0.86	0.86	**1.02**	1.00	1.00	-25815.85
'Sto N 1C'	0.96	0.96	0.87	0.86	0.87	0.86	0.87	0.86	0.86	0.86	**1.01**	1.00	1.00	-26185.63

model combinations based on the achieved values of log-likelihood and the Akaike information criterion. For clarity of exposition we limit our attention in Figure 1 to a selection of representative data-generating processes. Table 3 contains all results. Note, that the choice of AIC over BIC is not material in this case as the results are qualitatively similar.

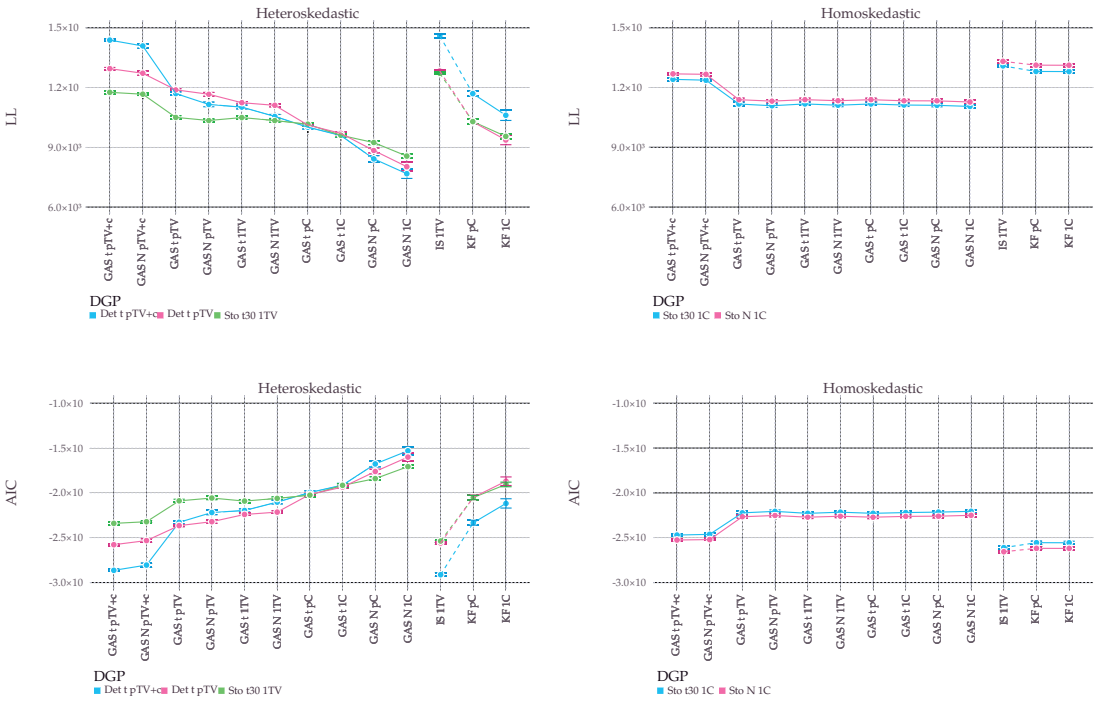
In each sub-figure of Figure 1, we show the 7 observation-driven models on the left and the 3 parameter-driven models on the right. Within each group we further sort the models based on their complexity. The more complicated and flexible models are presented on the left and the simple models on the right. The data-generating processes are divided into two groups depending on whether volatility is time-varying (panels on the left) or not (right). We present both the median values and the corresponding interquartile ranges. In Table 3, we show median values of the AIC for all cases normalised by the values achieved with the benchmark parameter-driven model.

We first consider the AIC performance of models for the DGPs with time-varying variances (left-hand panels). We find that similarly complex models have

Figure 1

Log-likelihood and information criteria

This figure compares values of log-likelihood (top) and AIC (bottom) for different methods and a selection of DGPs. We divide the DGPs into two groups depending on whether volatility is time-varying (left) or not (right). We sort method-model combinations according to their complexity with the most complex models on the left. Error-bars reflect interquartile ranges.



a similar AIC performance. For example, the AIC of '`KF 1C`' and '`GAS N 1C`' are very similar, even though these models are based on entirely different modelling paradigms. We also see that increasing the complexity of the models generally improves the model's in-sample fit while accounting for the number of parameters. Allowing for more heterogeneity across the variances and for fat-tailedness in a setting with constant variances results in substantially better performance in terms of lower AIC values despite the substantial increase in the number of parameters. For models with a deterministic factor path, the improvements continue all the way (adding fat-tails, adding more volatility components, adding cross-sectional equicorrelations). For the state space DGP with only one volatility component ('`Sto t30 1TV`'), the largest increases in performance are obtained

by relaxing the fixed volatility assumption. Beyond that, there is a fairly flat performance, irrespective of whether we allow for fat tails or more volatility components. The latter is intuitive as the DGP only has one volatility component. There is a further increase in performance (and decrease in AIC) if we allow for cross correlation.[6]

The performance of the parameter-driven models, which is visualised by points on the right-hand side of Figure 1, confirms that more complexity helps the model to better fit the data, even after accounting for the number of parameters. The parameter-driven models with fixed variances are clearly too restrictive. The fat-tailed state space model with a single time-varying variance component clearly outperforms the other state space specifications. Note, that estimation of this model is challenging and time-consuming.

For the homoskedastic DGPs ('`Sto t30 1C`', '`Sto N 1C`'), all GAS models perform roughly equally well, except the ones with non-zero cross correlation. In particular and as expected, there is hardly any difference in log-likelihood performance between models with constant and with time-varying volatilities, respectively. Similarly unsurprising is the better in-sample performance of the three state space models ('`IS t 1TV`', '`KF pC`', '`KF 1C`') which are relatively less mis-specified and reflect the parameter-driven nature of the DGPs.

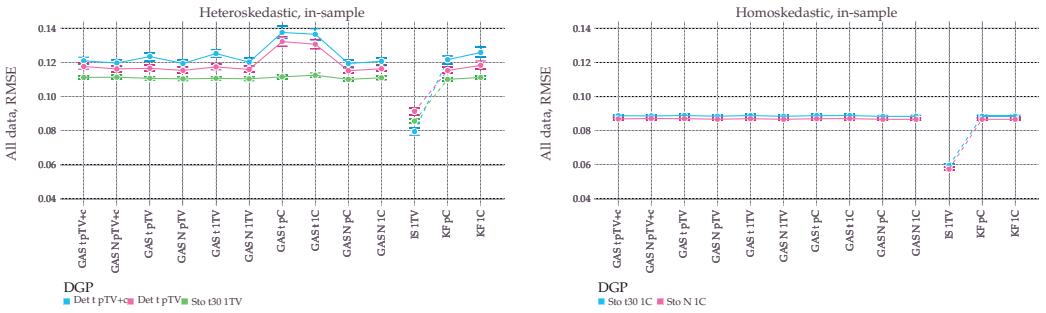
We further find that the largest differences in performance are obtained for short maturities and during spells of high volatility. We attribute these differences to a varying precision with which the unobserved components can be filtered from the data using the different methods. In high-volatility regimes,

[6] The decrease in AIC and the corresponding increase in log-likelihood for the models with an equicorrelation structure ('`GAS N pTV+c`', '`GAS t pTV+c`') is better understood by looking at the results for the homoskedastic DGPs (right-hand panels). Note that the DGPs generate contemporaneous cross-correlation between all yields through the common factor structure. As the GAS models are observation-driven, they only account for the time $t-1$ predicted part of the contemporaneous cross-correlation, and therefore miss part of it. Allowing the pricing errors in the GAS model to be cross-correlated using the equicorrelation structure alleviates this aspect of model misspecification and increases the likelihood. This is clearly seen in the right-hand panels by the drop in AICs from the models with zero correlation to the models with an equicorrelation structure.

Figure 2

In-sample RMSE

This figure compares values of average (both in time and in cross-section) root-mean-square errors for different methods and a selection of DGPs. We divide the DGPs into two groups depending on whether volatility is time-varying (left) or not (right). We sort method-model combinations according to their complexity with the most complex models on the left. Error-bars reflect interquartile ranges.



the GAS models yield more precise estimates of particularly the slope and the curvature of the yield curve compared to the state space specifications with constant variances and thin-tailed distributions. For details on the differences in cross-sectional and time-series performance, see Appendix C.

To conclude the analysis of in-sample performance we look at the root-mean-square error and the mean absolute percentage error. The errors are computed based on the fit of the full yield curve and averaged both in the cross-section and in the time dimension. The results for the two measures of fit are qualitatively similar and, for brevity, we focus on the RMSEs in Figure 2. The figure contains median values of RMSEs across all Monte Carlo simulations along with interquartile ranges for a selection of DGPs. See Table A.1 in Appendix A for full results.

At this level of aggregation it appears that—for the most part—all considered models offer similar in-sample fit as measured by the RMSE. This is despite the vast differences in AIC and log-likelihood as discussed earlier based on Figure 1 and Table 3. It is, however, not surprising given that the largest differences in tracking performance are obtained for (i) the short maturities and (ii) in high-

volatility regimes, both of which are under-represented in the aggregation. There are two exceptions. First, the two GAS models with fat-tailed disturbances and constant variances offer a considerably worse fit if the underlying DGP includes time-varying variances. This is due to the fact, that adjusting the static variances to better match the brief crisis period would decrease the tracking performance in the majority of the sample. Instead the optimiser opts, given the provided flexibility, to grossly underestimate the degrees of freedom at below 10.0 when the true value is set at 30.0. This effectively classifies the whole crisis period where volatility clustering manifests itself as a prolonged series of outliers.

By contrast, the parameter-driven model with a single common volatility factor and fat-tailed disturbances ('`IS t 1TV`') produces significantly lower RMSE regardless of whether the DGP includes time-varying volatilities or not. This is due to the fact that we obtain the paths of the factors through signal extraction which involves some smoothing and improves in-sample fit. Alternatively, it is also possible to filter out factor paths by means of particle filtering in which case we expect the performance to deteriorate to the levels exhibited by the other filters.[7]

Particularly interesting is the comparable performance of the very simple parameter-driven models and the most complex model we consider, i.e., '`GAS t pTV+c`',[8] for the '`Det t pTV+c`' data-generating process. In this case, the GAS model is much closer to the true DGP than the '`KF`' parameter-driven models, but they all perform similarly well. This is evidence of how much flexibility even the simple models have to fit the data by adjusting the decay parameter, λ, and decreasing the persistence of the slope and curvature factors.[9] As we discuss next, this is due to over-fitting as the out-of-sample performance is considerably weaker.

[7] Note, that estimation of this model with importance sampling is cumbersome and time-consuming. The use of particle filters given the large scale nature of this Monte Carlo study was not feasible.

[8] Student's t distributed errors with time-varying variances and equicorrelation.

[9] Evidence of the latter can be seen in Figure Figure C.2 of Appendix C.

Table 4

Out-of-sample RMSE and MAPE

We present median values of root-mean-square errors and mean absolute percentage errors relative to values achieved by 'KF pC'. Lower is better. We mark outperforming cases in bold.

Panel A: Root-mean-square error for the yields

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

Panel B: Mean absolute percentage error for the yields

MAPE	'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'	**0.86**	**0.87**	**0.87**	**0.86**	**0.92**	**0.89**	1.12	1.09	**0.92**	**0.94**	1.00	**0.97**	0.04
'Det N pTV+c'	**0.86**	**0.88**	**0.88**	**0.87**	**0.92**	**0.90**	1.12	1.09	**0.93**	**0.95**	1.00	**0.97**	0.04
'Det t pTV'	**0.90**	**0.91**	**0.90**	**0.91**	**0.94**	**0.93**	1.15	1.11	**0.97**	**0.98**	1.00	**0.96**	0.04
'Det N pTV'	**0.86**	**0.93**	**0.87**	**0.92**	**0.95**	**0.95**	1.18	1.13	**0.99**	1.26	1.00	2.22	0.04
'Det t 1TV'	1.00	1.00	**0.99**	**0.99**	1.00	**0.99**	1.16	1.18	1.01	1.03	1.00	1.02	0.05
'Det N 1TV'	1.00	1.00	1.00	1.00	1.00	1.00	1.15	1.18	1.01	1.03	1.00	1.02	0.06
'Sto N 1TV'	1.01	1.00	1.01	1.00	1.01	**0.99**	1.03	1.04	1.00	1.01	1.00	1.01	0.02
'Sto t30 1TV'	1.00	**0.98**	1.01	**0.99**	1.01	**0.99**	1.02	1.03	1.00	1.01	1.00	1.01	0.02
'Sto t50 1TV'	1.01	1.00	1.02	**0.99**	1.01	1.00	1.03	1.04	1.00	1.01	1.00	1.01	0.02
'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	0.02
'Sto t50 1C'	1.01	1.04	1.01	1.00	1.01	1.00	1.01	1.01	1.00	1.00	1.00	1.00	0.02
'Sto N 1C'	1.01	**0.98**	1.01	1.00	1.01	1.00	1.01	1.01	1.00	1.00	1.00	1.00	0.02

4.3. Out-of-sample performance (one-step-ahead predictions)

We now consider the out-of-sample performance of the considered models. For this purpose we take the estimated values of parameters based on the first 1,000 observations in each replication and compute recursively 50 one-step-ahead

Table 4

Out-of-sample RMSE and MAPE (continued)

Panel C: Root-mean-square error for the curvature factor

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

predictions in the remainder of the sample. [10] For each specification we obtain 825,000 one-period-ahead forecasts out-of-sample (1,500 replications × 50 out-of-sample observations × 11 yields). We compute the RMSEs and MAPEs and similarly to our in-sample analysis we average the results both across time and in the cross-section. Table 4 contains the results.

Even at this level of aggregation substantial differences in performance are clearly visible among the competing specifications. We do not observe differences in out-of-sample performance for DGPs with constant variances. However, the farther away the data-generating process is from satisfying the homoskedasticity, normality, and independence assumptions, the better is the performance of score-driven models as compared to the traditional parameter-driven dynamic Nelson-Siegel model of Diebold, Rudebusch, and Aruoba (2006). The most com-

[10] This choice of sample length is arbitrary, but we have verified that slightly increasing or decreasing the number of observations does not change the results qualitatively. However, a significant increase in the number of observations in the out-of-sample analysis might affect the results. This is because all of the score-driven models are always mis-specified in the Monte Carlo simulations and, as Nelson (1992) points out, mis-specified filters may generate worse long-term forecasts even if they make consistent one-step-ahead predictions. We find our sample length choice to be a reasonable compromise between the short-term consistency and real-world requirements.

plex score-driven model which we consider, '`GAS t pTV+c`', generates 14% more accurate forecasts. Note that these models offered comparable in-sample fit.

We find that the score-driven models are better equipped to track the slope and—especially—the curvature factors if the traditional assumptions are violated. Unlike the differences in in-sample performance these improvements are present for all but the longest maturities. To give a better insight into the scale of out-performance, we present the RMSEs and the MAPEs for short-term interest rates in Table A.2 (appendix) where the improvement reaches 25%. The improvement in fit stems here from much higher precision with which the GAS models track the slope and the curvature out-of-sample, see Panel C of Table 4. The results in Table 4 and Table A.2 further reinforce our conclusion that allowing for any form of heteroskedasticity produces the largest improvements in fit and forecasting performance. Whether to relax the independence assumption and the choice of the number of volatility factors depends on which maturities on the yield curve are of most interest. Relaxing the normality assumption seems to produce only slight improvements while substantially increasing the complexity of the model.

4.4. Forecasting performance (n-period-ahead predictions)

We conclude the simulation study by looking at performance of the different method-model specifications in producing n-period-ahead forecasts. In the case of the benchmark parameter-driven models, generating such forecasts is as easy as simply supplying the Kalman Filter with data augmented by missing values to match the desired forecast horizon. The Kalman Filter produces minimum MSE forecasts and their associated standard errors in a routine fashion, see Durbin and Koopman (2012, Ch. 4.11) for details.

In the observation-driven framework, one-step-ahead forecasts are obtained as a byproduct of estimation. This is because values of dynamic parameters in this class of models are perfectly predictable one-step-ahead given past information.

However, this also implies that in the standard approach it is not possible to obtain confidence intervals for these short-term predictions, see e.g. Pascual, Romo, and Ruiz (2006). Confidence intervals can be, however, obtained for misspecified models as proposed in Zamojski (2016). The performance of one-period-ahead forecasts is discussed earlier in the previous subsection. Here, we focus on longer-horizon predictions.

Longer-horizon forecasts in the observation-driven framework can be obtained through simulations by treating the adopted score-driven model as if it was the data-generating process. For instance, given the parameter estimates at time $t = T$, one can simulate M sample paths for $y_{i,T+1}, \ldots, y_{i,T+n}$ and compute statistics of interests: mean and standard deviation or median and appropriate quantiles. The quality of forecasts depends not only on how well the model matches the true DGP but also on the number of performed simulations. This is problematic in our setting for two reasons. First, the DNSMs themselves are clearly misspecified both in the simulation setting and in the empirical applications which may decrease their forecasting performance over longer periods; though, according to Nelson (1992), such misspecification does not affect consistency of the one-period-ahead forecast. Secondly, performing such simulations for the more complex models we consider is computationally expensive to the point of being prohibitive for our Monte Carlo study.

To address this issue we propose a less computationally intensive procedure for obtaining n-period-ahead forecasts which we call bootcasting. The bootcasting algorithm takes the following five steps:

Step 1: Estimate the scores, $\{\nabla_t\}$, with the chosen score-driven filter and normalise them by pre-multiplying them by the appropriate square-root-inverse Fisher information matrix to create a series of raw updating steps.

Step 2: Generate M series of forecasting raw steps by sampling blocks of size n with replacement from among the raw steps from Step 1.

Step 3: Re-scale the full block of forecasting raw steps by pre-multiplying them with the time, $t = T$, square-root of the Fisher information matrix.

Step 4: For each bootcasting sample apply the score-driven transition equation to obtain n-period 'bootcasts' of the time-varying parameters.

Step 5: Forecasts are obtained as the median (or mean) of the empirical distribution function of the time-varying parameters. The confidence intervals are obtained by taking appropriate quantiles of the EDF and can be computed to arbitrary precision by increasing M.

The motivation for this procedure stems from the statistical properties of the GAS steps, s_t, which are zero in expectation and have a constant unit variance if obtained with a square-root inverse of the Fisher information matrix (Creal, Koopman, and Lucas, 2013). This allows us to resample blocks of such steps. It follows that, if the square-root inverse was used in estimation, Step 1 and Step 3 can be omitted. If the scaling matrix is chosen to be the inverse Fisher information matrix, the procedure can be expressed as:

$$
\begin{aligned}
s_{T+k} &= \underbrace{\mathcal{I}_{T|T-1}^{-1}\underbrace{\mathcal{I}_{T|T-1}^{\frac{1}{2}}\underbrace{\mathcal{I}_{j|j-1}^{-\frac{1}{2}}\nabla_j}_{\text{Normalise}}}_{\text{Re-scale}}}_{\text{GAS update as in eq. (3)}} \\
&= \mathcal{I}_{T|T-1}^{-\frac{1}{2}}\mathcal{I}_{j|j-1}^{-\frac{1}{2}}\nabla_j, \qquad (17) \\
s_{T+k+1} &= \mathcal{I}_{T|T-1}^{-\frac{1}{2}}\mathcal{I}_{j+1|j}^{-\frac{1}{2}}\nabla_{j+1}, \qquad (18)
\end{aligned}
$$

where j is a randomly chosen offset such that $T - n \geq j \geq 1$.

The use of quantiles to obtain the final forecasts and confidence intervals around them is motivated by the fact that such measures are invariant under monotone transformations. Thus, bootcasted time-varying parameters can be used to produce forecasts for the observations as well. In the context of term-structure modelling with DNSMs and GAS it is sufficient to only bootcast the mean factors, $\boldsymbol{f}_t$.

There are benefits of this procedure. First, we do not sacrifice quality of forecasts which we verify for a selection of DGPs.[11] At the same time, we are able to avoid both inverting large, dense matrices in, e.g., eq. (16) at every step of the recursion as well as sampling from multivariate distributions. For $M = 2000$ simulations and depending on the model complexity, bootcasting is 2–10 times faster than the traditional procedure. These gains are present even for the simplest models we consider. A further benefit is that we are able to account for the misfit due to misspecification. For instance, if we observe a structural break in the data, the bootcasting procedure will account for the possibility of such an event occurring in the future at the observed, realised frequency.

We now proceed to compare forecasting performance between the competing specifications. In Figure 3, we present the results for two score-driven models, '`GAS t pTV+c`' and '`GAS t 1TV`', which allow for time-varying volatilities and fat-tailed disturbances. Forecasts of up to ten periods ahead are obtained with the aforementioned 'bootcasting' procedure. As a benchmark, we use the classical dynamic Nelson-Siegel model of Diebold, Rudebusch, and Aruoba (2006, '`KF pC`') for which the forecasts are produced by the Kalman Filter. In the figure we focus on a single heteroskedastic DGP (panels on the left hand side) and one with constant variances (on the right). We report average values of bias, RMSEs, and MAPEs of the forecasts as well as span of their confidence intervals.

First, we consider the DGP with time-varying variances and cross-correlation. The Kalman Filter appears to have an advantage for short-term forecasts with smaller bias, as well as lower RMSE and MAPE. This is likely due to over-fitting as discussed earlier. The relative performance of this model deteriorates quickly as the forecast horizon is extended. By the end of the 10-period horizon, forecasts based on the '`KF pC`' model clearly have the highest MAPEs. The RMSEs and values of bias show similar, albeit, weaker deterioration.

We find that '`GAS t pTV+c`' performs considerably better than the simpler

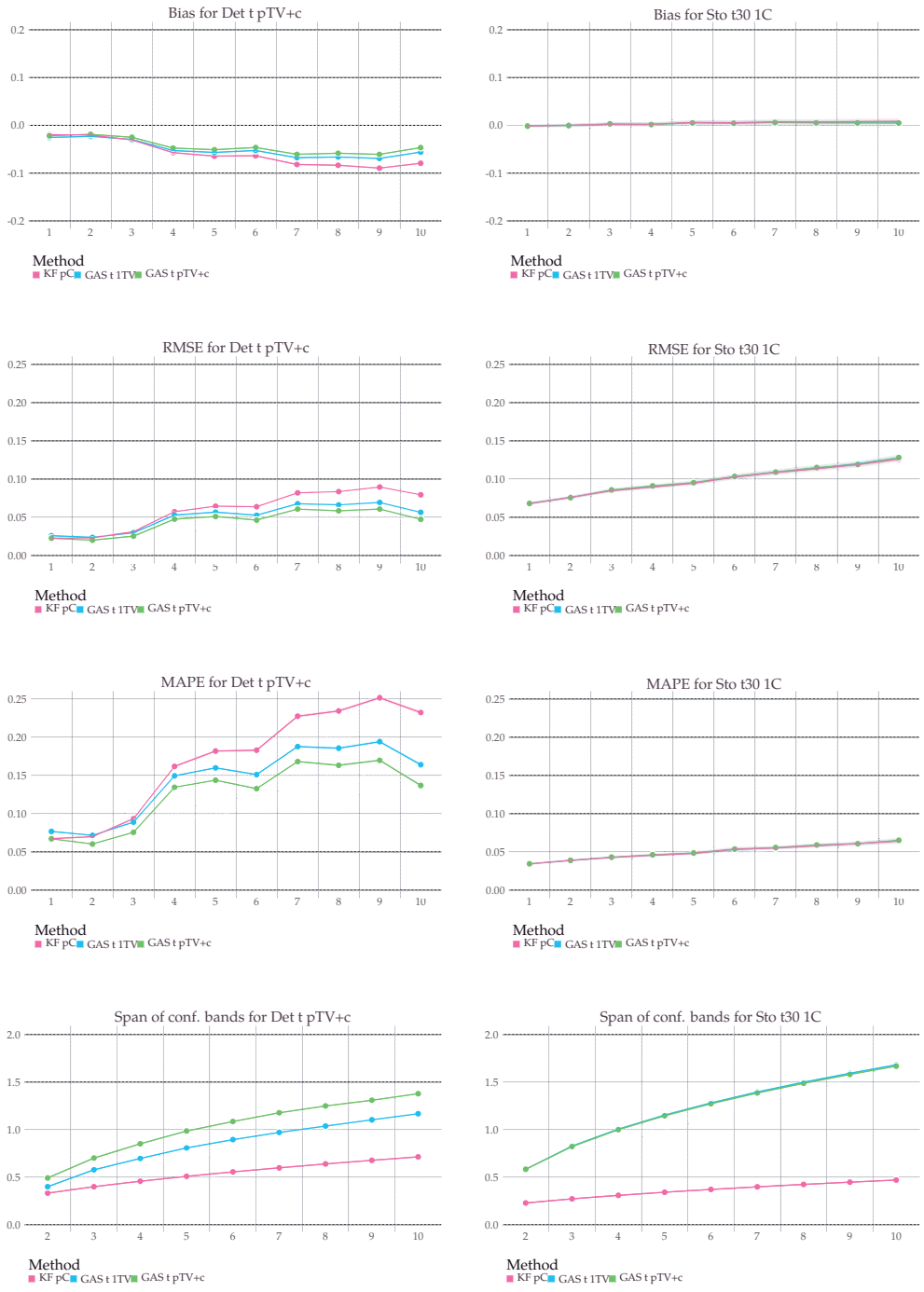
[11] The untabulated results are available upon request.

Figure 3

Forecasting performance: all maturities

Figure 4

Forecasting performance: short-term maturities



'`GAS t 1TV`' model and produces forecasts with smaller bias, RMSEs, and MAPEs for all horizons. This is due to the fact that in this case, '`GAS t pTV+c`' fits (feature-wise) much better with the analysed DGP.

The results in Figure 3 are averaged across all maturities. In Figure 4, we instead focus on the short-term maturities. The patterns which we saw for all data are even more pronounced here. This is not surprising, given our previous results on in-sample and out-of-sample fit. In particular, the difference in forecast accuracy obtained with the score-driven models and the benchmark parameter-driven model rises to 5 percentage points by the end of the 10-period horizon.

Finally and as expected, we do not observe any significant differences in forecasting performance between models with constant and with time-varying volatilities for homoskedastic DGPs.

Given the good performance of the GAS model in a controlled setting, we can now proceed to analyse the empirical data. This is done in the next section.

5. Empirical application

In this section we bring the discussed models to real-world, daily data from two countries. As is common in the literature, we look at the the U.S. market. Thus, our results can be directly compared to those of, e.g., Diebold and Li (2006) or Diebold, Rudebusch, and Aruoba (2006). We contribute to this literature by studying model performance at the high, daily frequency and by relaxing distributional assumptions. We also look at Poland which is a less economically and financially developed country. The choice of this dataset creates some interesting questions and challenges. For one, the Polish market is less liquid which is reflected by longer and more pronounced periods of high volatility. Moreover, note that the least restrictive model which we consider requires estimation of around 60 static (hyper-)parameters. The size of the model is not an important issue for the Western economies, especially the U.S., where monthly data is available for 50 years or more. However, low availability of data is one of the major obstacles in using recent advances in methods in developing countries, including in Poland. Given that even in the most parsimonious DNSM there are more than 30 parameters to be estimated, using short time series may result in the methods over-fitting the data.

To further motivate our choice of daily datasets, we note that with high frequency data it is possible to identify the channels through which the monetary policy affects the yield curve and thus agents' expectations. This is because we observe the potentially affected yields without much delay after policy-induced shocks. As Mesters, Schwaab, and Koopman (2014) argue, this is especially important after the collapse of Lehman Brothers in 2008, as policy interventions were being undertaken on a weekly (or even day-by-day) basis. The impact of the monetary policy can be established through an event study or by extending a chosen model to include macroeconomic variables as discussed in Section 3.[12]

[12]Nonetheless, we also look at monthly and weekly data in Section 5.1.

We use U.S. and Polish zero-coupon yields as provided by ICAP (Datastream) for 11 maturities: 3, 6, 9, 12, 18, 24, 36, 60, 84, 108, and 120 months. The data is reported at daily frequency. Due to differences in availability of the data, the sample period is 1997–2015 (4815 observations) for U.S. and 2007–2015 (2262 observations) for Poland.[13] The yield curves are derived based on a number of financial instruments. In particular, the short-term rates are based on LIBOR (WIBOR for Poland), while the mid- and long-term zeros are constructed based on forward rate agreements and swaps respectively.[14] For clarity of exposition we focus on four maturities in most of the tables and figures, i.e., on 6-, 12-, 60-, and 120-month yields.

Table 5 contains summary statistics for the two datasets. For each maturity we report mean, standard deviation, minimum, 25-th percentile, median, 75-th percentile, maximum, skewness, and kurtosis. We also show autocorrelation coefficients for a range of lags (from one day to two years). From the table we see that the yield curves in both markets are on average upward sloping, although the pattern is less pronounced in Poland. Additionally, the volatilities are decreasing with maturity. This is why we need more factors to explain dynamics on the short-term-end of the yield curve. Given the high frequency of the data it is not surprising that the yields are highly persistent. The persistence decreases in time, although the slower the longer the maturity. The table reveals that the data proxy for the curvature factor is the least persistent.

In Panel A of Figure 5, we present the dataset around the time of the 2008

[13] Based on the information from Thomson Reuters, ICAP started reporting the data for Poland in 2007 and provided a two-year 'instant' history. The data in the period 2005–2007 is of a noticeably lower quality and we thus remove it from the sample.

[14] We note that commercially available data on Polish yields starts, at the earliest, in 2001, and even then the liquidity of many (even benchmark) bonds is uneven at the beginning of the period. For this reason, other studies of the term structure of interest rates in Poland also construct the yield curves using interbank interest rates, see, e.g., Marciniak (2006); Kliber (2009) and Dziwok (2013). As a robustness check in Section 5.2, we look at the Fama-Bliss zero-coupon yields that are based on U.S. sovereign bonds. We find that the resulting filtered paths of the latent factors are very similar to those estimated using ICAP data. Given the lower availability of high quality data based on government bonds for Poland and the lack of substantial differences in this robustness analysis, we use the ICAP data for both markets for the sake of consistency.

Table 5

Summary statistics

Our data consists of daily zero-coupon yields in Poland and the U.S. at maturities of 3, 6, 9, 12, 18, 24, 36, 60, 84, 108, and 120 months. For each maturity we report mean, standard deviation, minimum, 25-th percentile, median, 75-th percentile, maximum, skewness, and kurtosis (Pearson). We also show autocorrelation coefficients at four lags: one day (1d), a month (1m), a year (1y), and two years (2y). The table contains results for data proxies of the latent mean factors. The longest maturity, 120-month, is adopted as a proxy for the level. Slope is approximated by the spread between the 120-month and 3-month yields. Proxy of the curvature factor is computed as the difference between the spread associated with 3-month and 24-month yields and the spread associated with the 24-month and 120-month yields.

Panel A: U.S., sample period is 1997–2015, 4815 observations

Maturity	Mean	Std. dev.	Min.	Q25	Median	Q75	Max.	Skewness	Kurtosis	1d ACF	1m ACF	1y ACF	2y ACF
3m	2.78	2.37	0.22	0.34	1.89	5.44	7.15	0.35	1.49	1.000	0.99	0.62	0.11
6m	2.81	2.37	0.22	0.40	2.03	5.41	7.44	0.36	1.53	1.000	0.99	0.63	0.12
9m	2.86	2.36	0.23	0.46	2.19	5.35	7.62	0.36	1.56	1.000	0.99	0.64	0.14
12m	2.91	2.35	0.25	0.52	2.38	5.32	7.72	0.35	1.59	0.999	0.99	0.65	0.16
18m	3.05	2.30	0.30	0.68	2.70	5.25	7.82	0.31	1.63	1.000	0.99	0.68	0.22
24m	3.19	2.25	0.34	0.87	2.98	5.20	7.87	0.27	1.67	1.000	0.99	0.70	0.28
36m	3.46	2.14	0.42	1.27	3.43	5.24	7.96	0.18	1.74	0.999	0.98	0.73	0.36
60m	3.91	1.92	0.73	1.96	4.00	5.42	7.95	0.04	1.83	0.999	0.98	0.75	0.45
84m	4.26	1.76	1.14	2.54	4.40	5.62	8.00	-0.02	1.92	0.999	0.97	0.74	0.50
108m	4.50	1.66	1.44	2.97	4.67	5.78	8.02	-0.06	1.98	0.999	0.97	0.74	0.54
120m (Level proxy)	4.60	1.62	1.57	3.12	4.78	5.84	8.07	-0.08	2.00	0.999	0.97	0.73	0.56
Slope proxy	1.82	1.29	-1.12	0.73	1.79	2.88	4.55	0.04	1.93	0.999	0.95	0.27	-0.42
Curvature proxy	-1.01	0.85	-4.05	-1.69	-1.03	-0.29	1.31	0.11	2.27	0.996	0.90	0.27	0.01

Panel B: Poland, sample period is 2007–2015, 2262 observation

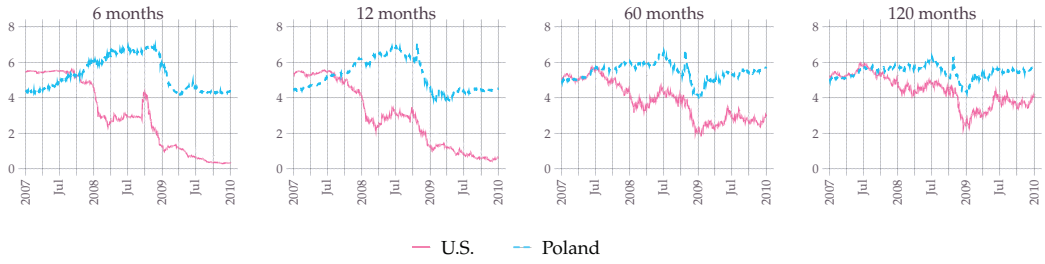
Maturity	Mean	Std. dev.	Min.	Q25	Median	Q75	Max.	Skewness	Kurtosis	1d ACF	1m ACF	1y ACF	2y ACF
3m	4.14	1.31	1.51	2.87	4.21	4.95	6.93	0.02	2.43	0.997	0.97	0.20	-0.14
6m	4.21	1.33	1.48	2.86	4.33	4.95	6.92	-0.02	2.40	0.997	0.97	0.24	-0.12
9m	4.18	1.39	1.58	2.72	4.35	4.95	6.90	-0.12	2.34	0.999	0.97	0.29	-0.10
12m	4.15	1.32	1.48	2.81	4.42	4.88	7.10	-0.20	2.37	0.999	0.97	0.31	-0.11
18m	4.18	1.31	1.46	2.96	4.56	4.90	7.05	-0.33	2.38	0.999	0.97	0.38	-0.06
24m	4.23	1.32	1.51	3.07	4.64	4.97	6.98	-0.46	2.37	0.999	0.96	0.44	-0.02
36m	4.33	1.29	1.50	3.29	4.73	5.23	6.88	-0.60	2.37	0.999	0.96	0.48	0.01
60m	4.48	1.23	1.54	3.60	4.87	5.47	6.74	-0.74	2.48	0.999	0.95	0.48	0.03
84m	4.57	1.16	1.63	3.82	4.94	5.52	6.58	-0.84	2.65	0.999	0.95	0.46	0.03
108m	4.63	1.10	1.73	3.95	4.98	5.52	6.39	-0.93	2.82	0.999	0.95	0.44	0.03
120m (Level proxy)	4.65	1.07	1.78	4.00	4.99	5.51	6.31	-0.95	2.87	0.998	0.95	0.44	0.04
Slope proxy	0.51	0.85	-2.23	-0.06	0.64	1.27	2.14	-0.56	2.66	0.992	0.90	-0.53	-0.10
Curvature proxy	-0.33	0.60	-2.20	-0.77	-0.41	0.16	1.08	-0.10	2.49	0.982	0.82	0.15	0.22

Figure 5

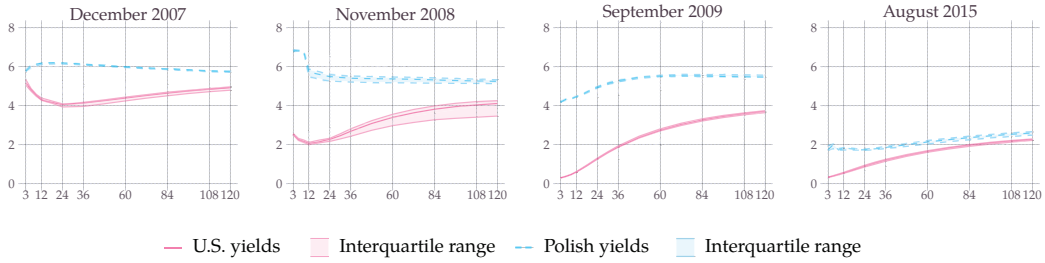
Zero-coupon yields

Our data consists of daily zero-coupon yields in Poland and the U.S. at maturities of 3, 6, 9, 12, 18, 24, 36, 60, 84, 108, and 120 months. We present a sub-sample of the daily data between 2007 and 2010 for a limited number of maturities in Panel A. Panel B contains full yield curves for a selection of months. In Panel B, we aggregate the data and report median values alongside interquartile ranges.

Panel A: Evolution in time of yields for chosen maturities



Panel B: Cross-section of yields for a selection of months



financial crisis. There are notable similarities in evolution of interest rates both in the U.S. and in Poland. We again see that the short-term yields are more volatile and less persistent than the longer maturities. For either country, it is visually apparent that the yields are driven by unobserved components as yields tend to move in the same direction. Although we do not model the two data sets jointly, co-movement of maturities between countries suggests that the factor structure is global in nature. Having said that, there are notable differences between the more economically developed U.S. and Poland. Short-term interest rates in the U.S. are considerably lower and the U.S. yield curve tends to be more steeply upward sloping not only on average but also during the financial crisis. By comparison, the Polish yield curve is much flatter.

Table 6

In-sample and out-of-sample fit for empirical data

The table contain estimation results for the dynamic Nelson-Siegel models considered in this paper. We fit the models to U.S. and Polish daily interest rates. We report the average (per observation) contributions to the log-likelihood (LL), AIC, and BIC measures. For clarity of exposition, we report RMSE and MAPE values multiplied by 100.0.

Panel A: U.S., sample period is 1997–2015, 4815 observations

				All maturities				Short-term maturities			
				In-sample		Out-of-sample		In-sample		Out-of-sample	
	LL	AIC	BIC	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
'GAS t pTV+c'	21.79	-43.57	-43.49	35.35	14.06	4.20	2.91	28.79	11.59	1.58	2.60
'GAS N pTV+c'	20.99	-41.96	-41.88	35.21	14.70	3.35	2.35	28.19	12.63	1.47	2.41
'GAS t pTV'	16.36	-32.70	-32.62	9.16	3.74	3.15	2.29	7.81	2.93	1.46	2.56
'GAS N pTV'	15.84	-31.65	-31.57	8.29	3.44	3.10	2.39	6.89	2.60	1.47	2.46
'GAS t 1TV'	15.25	-30.48	-30.43	8.20	3.46	3.28	2.67	6.72	3.10	1.66	2.88
'GAS N 1TV'	14.85	-29.68	-29.63	7.91	3.43	3.17	2.61	6.56	3.25	1.61	2.72
'GAS t pC'	13.68	-27.34	-27.29	9.61	3.73	3.40	2.76	9.13	3.68	1.96	3.34
'GAS t 1C'	12.98	-25.96	-25.93	9.30	3.91	3.41	2.75	9.12	4.38	1.98	3.50
'GAS N pC'	12.81	-25.60	-25.55	7.78	3.72	3.11	2.42	6.71	3.74	1.63	2.84
'GAS N 1C'	12.13	-24.25	-24.22	7.84	3.85	3.09	2.37	7.26	4.27	1.64	2.89
'IS t 1TV'	18.35	-36.68	-36.63	5.29	3.29			4.05	3.06		
'KF pC'	15.47	-30.93	-30.89	7.89	3.71	3.18	2.55	6.72	3.65	1.73	3.05
'KF 1C'	12.94	-25.88	-25.85	7.99	3.90	3.28	2.62	7.47	4.45	1.78	3.18

Across time there has been more variation in the level of short-term interest rates in the U.S. The country has experienced many periods in which its yield curve was inverted or inverted with a hump (1998, 2000/2001, 2006, 2007/2008, 2008/2009). Apart from the financial crisis, in which there was a clear distinction between short-term and other interest rates in Poland, the yield curve was inverted for about a year starting in 2012.

Table 6 summarises the results of the empirical study.[15] We fit 13 models to the interest rate data from both countries. The results are, for the most part, in line with those in the Monte Carlo study. In particular, we see that

[15] Parameter estimates along with standard errors are reported in Appendix D.

Table 6

In-sample and out-of-sample fit for empirical data (continued)

Panel B: Poland, sample period is 2007–2015, 2262 observations

				All maturities				Short-term maturities			
				In-sample		Out-of-sample		In-sample		Out-of-sample	
	LL	AIC	BIC	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
'GAS t pTV+c'	18.41	-36.76	-36.61	18.66	3.19	9.73	4.02	24.22	4.50	17.26	7.65
'GAS N pTV+c'	17.20	-34.35	-34.21	18.01	3.08	10.60	4.37	26.65	4.69	18.28	8.04
'GAS t pTV'	15.61	-31.16	-31.01	9.41	1.51	8.48	3.55	13.18	2.15	13.77	6.18
'GAS N pTV'	15.08	-30.10	-29.96	8.72	1.46	7.90	3.34	12.95	2.16	13.60	6.16
'GAS t 1TV'	14.66	-29.29	-29.20	8.95	1.51	8.38	3.56	12.42	2.14	13.56	6.13
'GAS N 1TV'	14.27	-28.51	-28.42	8.53	1.47	7.90	3.35	12.41	2.13	13.43	6.10
'GAS t pC'	13.17	-26.30	-26.22	10.54	1.59	9.01	3.83	12.84	2.13	13.96	6.40
'GAS t 1C'	12.69	-25.35	-25.29	12.45	1.74	10.03	4.17	13.87	2.19	15.87	6.99
'GAS N pC'	12.27	-24.52	-24.43	8.42	1.47	8.34	3.52	12.34	2.13	14.41	6.38
'GAS N 1C'	10.99	-21.95	-21.89	8.44	1.47	8.54	3.62	11.76	2.00	13.99	6.22
'IS t 1TV'	18.62	-37.22	-37.14	8.91	1.41			16.48	2.75		
'KF pC'	15.44	-30.85	-30.77	11.64	1.87	7.93	3.43	17.82	3.03	13.21	6.08
'KF 1C'	11.11	-22.20	-22.15	8.71	1.51	8.98	3.79	11.83	2.02	14.38	6.38

relaxing the assumptions of normality, homoskedasticity, and independence leads to huge improvements in in-sample fit as measured by the AIC and BIC. Between the simplest score-driven model and the most complex we are able to almost double the log-likelihood and even after accounting for the presence of additional parameters it is clear that the more complex models are preferred.

For the U.S. data the improvements are also clearly visible in the overall reduction of RMSEs and MAPEs, both in-sample and out-of-sample. This is with exception of the two models which allow for cross-sectional correlations in pricing errors for which the improvement in AIC/BIC is accompanied by a substantial deterioration of in-sample RMSEs and MAPEs. As discussed before, the equicorrelation structure allows these models to capture the part of contemporaneous shifts in the yield curve which are not predictable one period ahead. The surprising results might be due to the fact that the volatility of the latent factors,

in particular of the level factor, is dominating the variance of pricing errors and contributes to the very high estimates of correlation we obtain in these models.

If we do not relax the independence assumption, these results are fully in line with our conclusions from the Monte Carlo study. Overall, the '`GAS N pTV`' and '`GAS t pTV`' models offer the best out-of-sample performance (based on one-period-ahead predictions) for the U.S. data. In the case of Poland, the results are inconclusive in the short-horizon setup we consider here.

In Figure 6, we show the filtered paths of the latent factors while Figure 7 contains a selection of fitted interest rates for the U.S. data. Analogous figures for Poland can be found in Appendix A. We present the results for two sub-periods. The plots on the left-hand side contain the factors and yields around the 2008 financial crisis. On the right are the most recent results at the end of our sample period. Overall, we see that allowing for time-varying volatilities and fat-tailed disturbances has a considerable impact on the three mean factors. Especially affected is the curvature factor which becomes more persistent. The bottom panels of Figure 6 further suggest that the Kalman Filter estimated variances are too high at the end of the sample.

Finally, we find that allowing for multiple time-varying volatility factors leads to new insights about the dynamics of pricing errors in the DNSMs. If we allow for only a single common volatility factor, the long-term maturities are estimated to have higher volatilities than the short-term ones. This is also true for the models with multiple volatility factors, but only in non-crisis periods. In high-volatility regimes, these models suggest a reversal of this ranking with the short-term maturities having considerably higher variances of pricing errors. With the benefit of hindsight, we can attribute this phenomenon to the short-term increase of counter-party risk during the crisis period which would predominantly affect the short-term maturities. During the financial crisis regulators had to step in to provide liquidity to the interbank market. A relatively high frequency DNSM with time-varying volatilities of pricing errors may then be a useful tool for

Figure 6

Filtered latent factors, U.S.

This figure contains filtered paths of level, slope, curvature, as well as volatility paths for a selection of maturities. The column on the left contains results around the 2008 financial crisis while the right column presents end of our sample period. We use daily data and the sample period is 1997–2015.

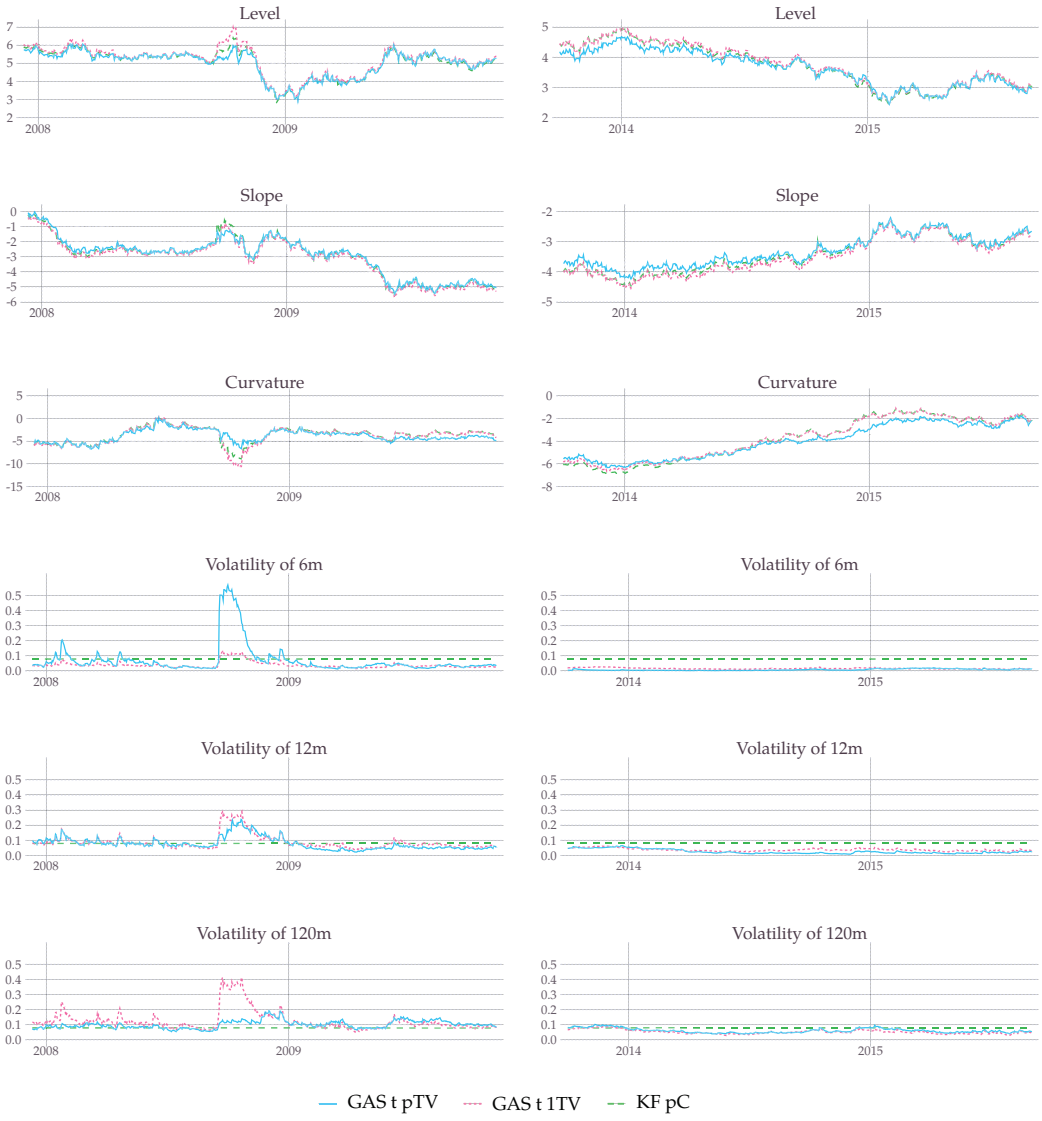


Figure 7

Estimated yields, U.S.

We plot fitted data for a range of maturities as estimated with three of our models. The column on the left contains results around the 2008 financial crisis while the right column presents end of our sample period. We use daily data and the sample period is 1997–2015.

a central bank in gauging the effectiveness of such interventions.

5.1. Robustness to changes in sampling frequency

In this section, we compare estimation results obtained for three different sampling frequencies: monthly, weekly, and daily. The lower frequency series are obtained by resampling the daily data.

We first focus on the '`GAS t pTV`' model. In Figure 8, we compare estimation results obtained for the three different sampling frequencies. Here, we use the daily U.S. data from ICAP. Setting aside the different degree of path smoothness, it is evident from the figure that there are no qualitative differences between results obtained for the three datasets. Additionally, filtered paths of the mean factors are also similar quantitatively, i.e., the paths for monthly data are only a heavily smoothed version of the paths we obtain at the daily frequency. We also find evidence for conditional heteroskedasticity and fat-tailedness of pricing errors at lower frequencies. In particular, the estimated fat-tailedness is significantly higher for the monthly data. The variances of monthly pricing errors increase in the same periods as for the daily data, but the periods of high volatility are estimated as considerably longer. Our results are in line with Koopman, Mallee, and Van der Wel (2010) and Mesters, Schwaab, and Koopman (2014) who also show that volatilities of pricing errors are time-varying at lower sampling frequencies.

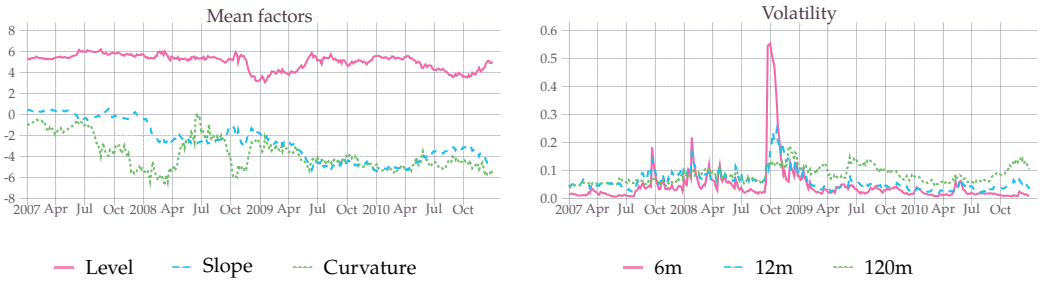
In Figure 9, we study the impact of sampling frequency on the log-likelihood and the AIC for all the score-driven models we consider in the paper. For clarity, we present average (per observation) values of these measures of in-sample fit. The general pattern is similar regardless of the sampling frequency: allowing for more features in the model leads to improvements in fit. However, the relative importance of the features depends on the sampling frequency. At the daily frequency, relaxing the normality assumption leads to only small benefits. The largest improvements are obtained by allowing for heteroskedasticity in pricing

Figure 8

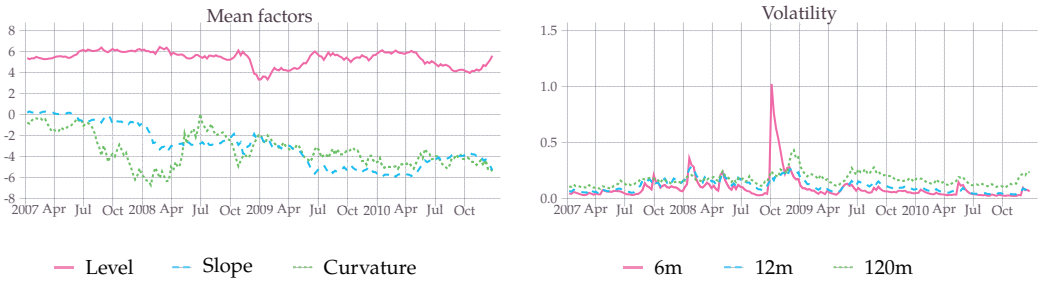
Impact of sampling frequency on the filtered factors

This figure contains filtered paths of the level, slope, and curvature, as well as volatility paths for a selection of maturities estimated with the '`GAS t pTV`' model. In Panel A, we present results for the daily U.S. data which we use in the remainder of the paper. Panels B and C contain results for weekly and monthly data. In all cases, we use the full sample period from 1997 to 2015 to estimate the model. We report the results for 2007–2010.

Panel A: Daily observations, $\nu = 26.38$



Panel B: Weekly observations, $\nu = 24.42$



Panel C: Monthly observations, $\nu = 18.04$

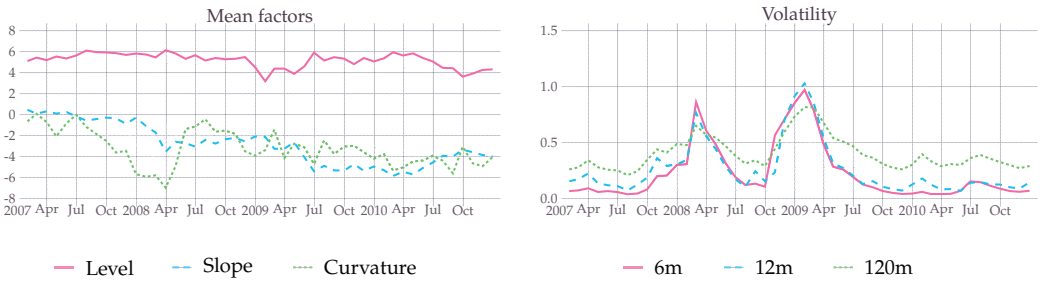
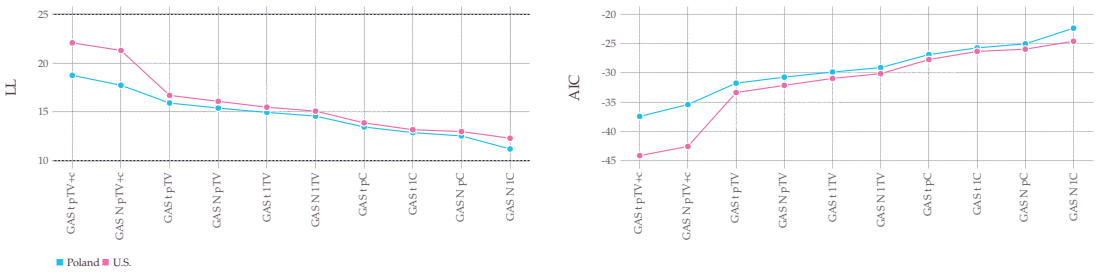


Figure 9

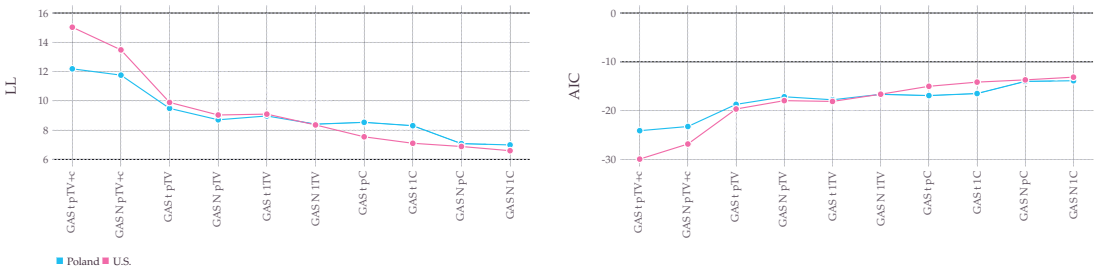
Impact of sampling frequency on in-sample fit

This figure compares values of log-likelihood (left) and AIC (right) for different score-driven models fitted to empirical data. We present results for the U.S. and Polish data. Due to different sample lengths, we report the average contributions to the log-likelihood and the AIC of each observation. Panel A, B, and C show results for daily, weekly, and monthly sampling frequencies. We sort models according to their complexity within the most complex models on the left.

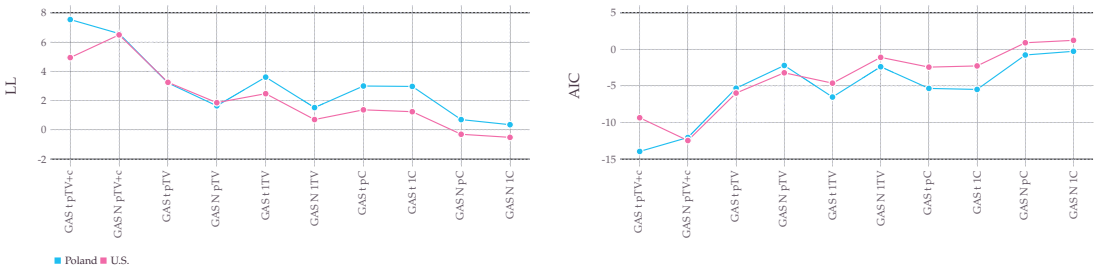
Panel A: Daily data. Sample length is 4815 and 2262 for for U.S. and Poland.



Panel B: Weekly data. Sample length is 964 and 453 for for U.S. and Poland.



Panel C: Monthly data. Sample length is 223 and 105 for for U.S. and Poland.



errors. Based on our results, the lower the sampling frequency, the more important it is to allow for non-Gaussian errors. This is despite the fact, that even at the low, monthly frequency we see improvement in both the AIC and the log-likelihood after allowing for some form of conditional heteroscedasticity. The importance of allowing for fat-tailedness errors in the monthly case may be partly due to the lagging nature of score-driven models, in that the the filter has too little time to adjust the conditional variance leaving many under- or over-scaled residuals. The results point to an interesting trade-off where working with relatively high-frequency data may allow us to keep the normality assumption if we allow for heteroskedasticity; while for the low-frequency data we need to relax both assumptions.

Ultimately, our ability to filter out the time-varying variance of pricing errors depends on the magnitude of the changes and, indeed, on how frequently we sample the data. At some very low frequency there might not be enough persistence in the time series for the method to pick up the time-variation in volatility even if the time-varying model is the closest to the true DGP. This is to say; the simpler models may be able to fit the data equally well by increasing unconditional variance and/or increasing the fat-tailedness of the distribution of pricing errors. A question then remains, which of these models have the potential to bring the most insight about the data and, in particular, what are the economic channels through which the yield curve is affected. Under the assumption, that the true, high-frequency DGP includes time-varying volatility (e.g., in the form of stochastic volatility), modelling low frequency data as homoskedastic would lead to overestimating the variance of the pricing errors in good states, i.e., most of the time historically. Allowing for fat-tailed pricing errors would mean admitting defeat in modelling the term structure in crises with just some improvement in the good states. As interest rates are observable at higher frequencies, in both cases, we then leave valuable information on the table. Furthermore, there are reasons to believe that mis-specified filters of the type we discuss in the paper are better able to track the true underlying processes as sampling frequency is increased, see, e.g., Nelson (1992) or Nelson and Foster (1994).

Naturally, various market participants may want to model more than just the term structure of interest rates as we do in this paper. For instance, central banks may be interested in including in the model some time-series of macroeconomics variables that are observable at lower frequencies. In this case, it is still possible to use 'high-frequency' Nelson-Siegel models and we refer to the rapidly growing literature on modelling mixed-frequency datasets, see e.g., Ghysels, Santa-Clara, and Valkanov (2004); Andreou, Ghysels, and Kourtellos (2010); Ghysels (2016), or Blasques, Koopman, Mallee, and Zhang (2016).

5.2. Impact of dataset choice

In this section we investigate whether sample selection impacts the empirical results. Our main empirical analysis uses daily data from ICAP which relies on interbank interest rates to construct the short-end of the yield curve. This dataset is interesting as it contains high-frequency data for a range of countries. As a robustness check, we now consider the standard unsmoothed Fama-Bliss zero-coupon yields obtained based on the CRSP unsmoothed Fama and Bliss (1987) forward rates. Unlike the ICAP data, the Fama-Bliss dataset is available only for U.S. and is based on sovereign bonds. It is also widely used in the literature, e.g., by Diebold and Li (2006) or Diebold, Rudebusch, and Aruoba (2006). For consistency, we study yields with the same maturities as in the remainder of Section 5. We fit the '`GAS t pTV`' model to data from 1997–2009 and from both datasets. For the sake of brevity, we focus on the '`GAS t pTV`' model with (possibly) fat-tailed pricing errors and individual latent volatility factors. Figure 10 contains filtered paths of level, slope, curvature, as well as volatility paths for the 6-, 12-, and 120-month yields.

The results are very similar overall. The only exception is the level factor which is consistently estimated higher for the ICAP data. This is unsurprising given that interbank interest rates include a premium to account for counter-

Figure 10

ICAP vs. Fama-Bliss

This figure contains filtered paths of the level, slope, and curvature, as well as volatility paths for a selection of maturities estimated with the '`GAS t pTV`' model for two datasets. We consider both monthly data obtained from ICAP and Fama-Bliss zero-coupon yields. Sample period for estimation is 1997-2009. We report the results for 2006–2009.

party risk. Notably, however, even for this factor the extracted dynamics are similar. The choice of dataset here does not significantly affect estimates of hyper-parameters. For instance, the degrees of freedom (ν) are estimated at 10.64 and 11.54 for the ICAP and Fama-Bliss datasets respectively. Similarly, the difference between estimates of the decay parameters, λ, are small: 0.081 for ICAP and 0.075 for Fama-Bliss.

6. Conclusions

In this paper, we have studied in-sample, out-of-sample, and forecasting performance of dynamic yield curve models with score-driven time-varying parameters. We considered a sequence of score-driven models that gradually relax the restrictive distributional assumptions which are usually imposed on dynamic yield curve models and which are cumbersome to relax in the parameter-driven framework. Specifically, we extend the model with time-varying volatility structures, possibly driven by common components. Furthermore, apart from having disturbances from a multivariate Gaussian distribution, we also allow disturbances to come from fat-tailed distributions, namely from the multivariate Student's t distribution. Finally we allow the pricing errors to be correlated in the cross-sectional dimension.

Based on an extensive Monte Carlo study, we find that the generalised autoregressive score models of Creal, Koopman, and Lucas (2013) produce similar in-sample fit as correctly specified state-space models while offering superior out-of-sample performance when the aforementioned distributional assumptions are violated.

Having more features in the model appears to benefit the time-series (yields) which load on multiple latent mean factors or on latent factors about which there is little information in the cross-section. In the context of interest rates, this means that the complex score-driven models offer an improved in-sample and considerably better out-of-sample fit for short-term maturities. Similarly the latent curvature factor which determines the yields in the short- and medium-run is tracked with higher precision. We find this to be especially true in periods when the time-varying volatilities are considerably above or below their unconditional means.

We have also proposed a novel forecasting procedure which can be used to quickly and efficiently construct multiple-period-ahead predictions and corresponding out-of-sample confidence intervals for score-driven models.

Finally, we have studied the dynamics of yield curves in the U.S. and in Poland. Our overall conclusion from this analysis is that by increasing the complexity of the model, we obtain considerable improvements in the fit. The largest improvements are obtained by allowing for heteroskedasticity in pricing errors. Also, by allowing for multiple volatility factors, we provide new insights as to what drives the dynamics of the yield curve in high-volatility regimes. Relaxing the normality assumption does not lead to significant benefits.

References

Abadir, Karim M, and Jan R Magnus, 2005, *Matrix algebra* . vol. 1 (Cambridge University Press).

Andreou, Elena, Eric Ghysels, and Andros Kourtellos, 2010, Regression models with mixed sampling frequencies, *Journal of Econometrics* 158, 246–261.

Arellano-Valle, Reinaldo B, 2010, On the information matrix of the multivariate skew-t model, *Metron* 68, 371–386.

Blasques, F, SJ Koopman, M Mallee, and Z Zhang, 2016, Weighted maximum likelihood for dynamic factor analysis and forecasting with mixed frequency data, *Journal of Econometrics*.

Blasques, Francisco, Siem Jan Koopman, and Andre Lucas, 2014, Maximum likelihood estimation for generalized autoregressive score models, .

Blasques, Francisco, Siem Jan Koopman, and André Lucas, 2015, Information-theoretic optimality of observation-driven time series models for continuous responses, *Biometrika* 102, 325–343.

Bollerslev, Timothy, 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307–327.

Christensen, Jens HE, Francis X Diebold, and Glenn D Rudebusch, 2011, The affine arbitrage-free class of Nelson-Siegel term structure models, *Journal of Econometrics* 164, 4–20.

Cox, David R., 1981, Statistical analysis of time series: some recent developments, *Scandinavian Journal of Statistics* 8, 93–115.

Creal, Drew, Siem Jan Koopman, and André Lucas, 2011, A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations, *Journal of Business & Economic Statistics* 29, 552–563.

Creal, Drew, Siem Jan Koopman, and André Lucas, 2013, Generalized autoregressive score models with applications, *Journal of Applied Econometrics* 28, 777–795.

——— , and Marcin Zamojski, 2015, Generalized autoregressive method of moments, Discussion paper, Tinbergen Institute Discussion Paper 15-138/III.

Diebold, Francis X, and Canlin Li, 2006, Forecasting the term structure of government bond yields, *Journal of Econometrics* 130, 337–364.

Diebold, Francis X, Monika Piazzesi, and Glenn Rudebusch, 2005, Modeling bond yields in finance and macroeconomics, Discussion paper, National Bureau of Economic Research.

Diebold, Francis X, Glenn D Rudebusch, and S Boragan Aruoba, 2006, The macroeconomy and the yield curve: a dynamic latent factor approach, *Journal of Econometrics* 131, 309–338.

Durbin, James, and Siem Jan Koopman, 1997, Monte Carlo maximum likelihood estimation for non-Gaussian state space models, *Biometrika* 84, 669–684.

——— , 2012, *Time Series Analysis by State Space Methods* (Oxford University Press: Oxford, UK) 2 edn.

Dziwok, Ewa, 2013, Weryfikacja modeli krzywej dochodowości na potrzeby polityki pieniężnej NBP, *Materiały i Studia* 296.

Engle, Robert F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica* 50, 987–1007.

Fama, Eugene F, and Robert R Bliss, 1987, The information in long-maturity forward rates, *The American Economic Review* pp. 680–692.

Fernández-Villaverde, Jesús, and Juan Rubio-Ramírez, 2013, Macroeconomics and volatility: Data, models, and methods, in Daron Acemoglu, Manuel Arel-

lano, and Eddie Dekel, ed.: *Advances in Economics and Econometrics: Tenth World Congress* . pp. 137–183 (Cambridge University Press: Cambridge, UK).

Filipović, Damir, 1999, A note on the nelson–siegel family, *Mathematical finance* 9, 349–359.

Ghysels, Eric, 2016, Macroeconomics and the reality of mixed frequency data, *Journal of Econometrics*.

——— , Pedro Santa-Clara, and Rossen Valkanov, 2004, The midas touch: Mixed data sampling regression models, Discussion paper, CIRANO.

Harvey, Andrew C, 2013, *Dynamic models for volatility and heavy tails: with applications to financial and economic time series* . No. 52 (Cambridge University Press).

Kliber, Paweł, 2009, Estymacja struktury terminowej stóp procentowych w polsce, *Bank i Kredyt* pp. 107–122.

Koopman, Siem Jan, Rutger Lit, and André Lucas, 2015, Intraday stock price dependence using dynamic discrete copula distributions, Discussion paper, Tinbergen Institute Discussion Paper 15-037/III/DSF90.

Koopman, Siem Jan, André Lucas, and Marcel Scharth, 2015, Numerically accelerated importance sampling for nonlinear non-Gaussian state-space models, *Journal of Business & Economic Statistics* 33, 114–127.

Koopman, Siem Jan, Andre Lucas, and Marcel Scharth, 2016, Predicting time-varying parameters with parameter-driven and observation-driven models, *Review of Economics and Statistics*.

Koopman, Siem Jan, Max IP Mallee, and Michel Van der Wel, 2010, Analyzing the term structure of interest rates using the dynamic Nelson–Siegel model with time-varying parameters, *Journal of Business & Economic Statistics* 28, 329–343.

Marciniak, Marek, 2006, Yield curve estimation at the National Bank of Poland, *Bank i Kredyt* 10, 52–74.

Mesters, Geert, Bernd Schwaab, and Siem Jan Koopman, 2014, A dynamic yield curve model with stochastic volatility and non-Gaussian interactions: an empirical study of non-standard monetary policy in the euro area, Discussion paper, Tinbergen Institute Discussion Paper 14-071/III.

Mitchell, Ann ES, 1989, The information matrix, skewness tensor and a-connections for the general multivariate elliptic distribution, *Annals of the Institute of Statistical Mathematics* 41, 289–304.

Nelson, Charles R, and Andrew F Siegel, 1987, Parsimonious modeling of yield curves, *Journal of Business* pp. 473–489.

Nelson, Daniel B, 1992, Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model, *Journal of Econometrics* 52, 61–90.

Nelson, Daniel B., and Dean P. Foster, 1994, Asymptotic filtering theory for univariate arch models, *Econometrica* 62, pp. 1–41.

Pascual, Lorenzo, Juan Romo, and Esther Ruiz, 2006, Bootstrap prediction for returns and volatilities in GARCH models, *Computational Statistics & Data Analysis* 50, 2293–2312.

Shephard, Neil, 2005, *Stochastic Volatility: Selected Readings* (Oxford University Press: Oxford, UK).

——— , and Michael K Pitt, 1997, Likelihood analysis of non-Gaussian measurement time series, *Biometrika* 84, 653–667.

Zamojski, Marcin, 2016, Filtering with confidence: In-sample confidence bands for GARCH filters, Discussion paper, SYRTO (Systemic Risk Tomography) Working Paper 4/2016.

A. Supplementary tables and figures

Table A.1

In-sample RMSE and MAPE in the simulation study

We present median values of root-mean-square errors and mean absolute percentage errors relative to values achieved by 'KF pC'. Lower is better. We mark outperforming cases in bold.

Panel A: Root-mean-square error

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'	'IS t 1TV'	'KF pC'	'KF 1C'	'KF pC' value
'Det t pTV+c'	**0.99**	**0.98**	1.02	**0.98**	1.03	**0.99**	1.13	1.12	**0.98**	**0.99**	**0.65**	1.00	1.03	0.12
'Det N pTV+c'	**0.99**	**0.98**	1.01	**0.98**	1.03	**0.99**	1.14	1.13	**0.98**	**0.99**	**0.65**	1.00	1.03	0.12
'Det t pTV'	1.02	1.01	1.01	1.00	1.02	1.01	1.15	1.13	1.00	1.01	**0.79**	1.00	1.03	0.12
'Det N pTV'	1.03	1.03	1.03	1.01	1.05	1.01	1.19	1.19	1.00	1.04	**0.87**	1.00	1.16	0.17
'Det t 1TV'	1.03	1.01	1.02	1.00	1.02	1.00	1.33	1.31	1.00	1.01	**0.77**	1.00	1.01	0.12
'Det N 1TV'	1.03	1.01	1.04	1.00	1.04	1.00	1.33	1.31	1.00	1.01	**0.77**	1.00	1.01	0.12
'Sto N 1TV'	1.01	1.01	1.01	1.00	1.01	1.00	1.01	1.02	1.00	1.01	**0.77**	1.00	1.01	0.11
'Sto t30 1TV'	1.01	1.01	1.01	1.00	1.01	1.00	1.01	1.02	1.00	1.01	**0.78**	1.00	1.01	0.11
'Sto t50 1TV'	1.01	1.01	1.01	1.00	1.01	1.00	1.01	1.02	1.00	1.01	**0.77**	1.00	1.01	0.11
'Sto t30 1C'	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.67**	1.00	1.00	0.09
'Sto t50 1C'	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.67**	1.00	1.00	0.09
'Sto N 1C'	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.66**	1.00	1.00	0.09

Panel B: Mean absolute percentage error

MAPE	'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'	'IS t 1TV'	'KF pC'	'KF 1C'	'KF pC' value
'Det t pTV+c'	**0.97**	**0.96**	**0.98**	**0.96**	1.00	**0.97**	1.07	1.07	**0.97**	**0.98**	**0.57**	1.00	1.02	0.03
'Det N pTV+c'	**0.97**	**0.96**	**0.98**	**0.96**	1.00	**0.97**	1.07	1.07	**0.97**	**0.99**	**0.58**	1.00	1.01	0.03
'Det t pTV'	**0.99**	**0.99**	**0.99**	**0.98**	1.00	**0.99**	1.08	1.07	1.00	1.01	**0.73**	1.00	1.01	0.03
'Det N pTV'	1.00	1.02	1.00	**0.99**	1.02	1.01	1.11	1.11	1.00	1.13	**0.77**	1.00	1.43	0.03
'Det t 1TV'	1.01	1.01	1.01	1.00	1.01	1.00	1.13	1.13	1.00	1.02	**0.75**	1.00	1.01	0.03
'Det N 1TV'	1.01	1.01	1.02	1.00	1.02	1.00	1.14	1.15	1.00	1.02	**0.75**	1.00	1.01	0.03
'Sto N 1TV'	1.01	1.00	1.01	1.00	1.01	1.00	1.02	1.03	1.00	1.01	**0.72**	1.00	1.01	0.02
'Sto t30 1TV'	1.01	1.02	1.01	1.00	1.01	1.00	1.02	1.03	1.00	1.01	**0.77**	1.00	1.01	0.02
'Sto t50 1TV'	1.01	1.02	1.01	1.00	1.01	1.00	1.02	1.03	1.00	1.01	**0.76**	1.00	1.01	0.02
'Sto t30 1C'	1.01	**0.98**	1.01	1.00	1.01	1.00	1.01	1.01	1.00	1.00	**0.70**	1.00	1.00	0.02
'Sto t50 1C'	1.00	**0.96**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.68**	1.00	1.00	0.02
'Sto N 1C'	1.00	**0.97**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.67**	1.00	1.00	0.02

Table A.2

Out-of-sample RMSE and MAPE in the simulation study: short-term maturities

We present median values of root-mean-square errors and mean absolute percentage errors relative to values achieved by 'KF pC'. Lower is better. We mark outperforming cases in bold.

Panel A: Root-mean-square error

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

Panel B: Mean absolute percentage error

MAPE	'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'	**0.76**	**0.79**	**0.78**	**0.78**	**0.83**	**0.82**	1.17	1.10	**0.85**	**0.91**	1.00	**0.97**	0.06
'Det N pTV+c'	**0.76**	**0.79**	**0.78**	**0.78**	**0.83**	**0.81**	1.18	1.11	**0.86**	**0.91**	1.00	**0.96**	0.05
'Det t pTV'	**0.82**	**0.85**	**0.83**	**0.84**	**0.88**	**0.87**	1.26	1.18	**0.94**	**0.97**	1.00	**0.91**	0.05
'Det N pTV'	**0.81**	**0.89**	**0.82**	**0.88**	**0.91**	**0.92**	1.31	1.24	**0.97**	1.31	1.00	2.06	0.05
'Det t 1TV'	**0.95**	**0.95**	**0.95**	**0.95**	**0.97**	**0.96**	1.33	1.38	1.00	1.07	1.00	1.07	0.06
'Det N 1TV'	**0.96**	**0.96**	**0.96**	**0.96**	**0.98**	**0.96**	1.31	1.37	1.01	1.07	1.00	1.06	0.07
'Sto N 1TV'	1.01	1.00	1.02	1.00	1.02	1.00	1.05	1.06	1.00	1.04	1.00	1.05	0.02
'Sto t30 1TV'	1.01	**0.98**	1.01	1.00	1.01	1.00	1.04	1.06	1.01	1.04	1.00	1.06	0.02
'Sto t50 1TV'	1.02	1.01	1.03	1.00	1.02	1.00	1.03	1.05	1.01	1.04	1.00	1.06	0.02
'Sto t30 1C'	1.02	1.01	1.01	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.03
'Sto t50 1C'	1.01	1.02	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.00	1.00	1.00	0.03
'Sto N 1C'	1.01	**0.95**	1.01	1.01	1.01	1.00	1.01	1.01	1.00	1.00	1.00	1.00	0.03

Figure Figure A.1

Filtered latent factors, Poland

This figure contains filtered paths of level, slope, curvature, as well as volatility paths for a selection of maturities. The column on the left contains results around the 2008 financial crisis while the right column presents end of our sample period. We use daily data and the sample period is 2007–2015.

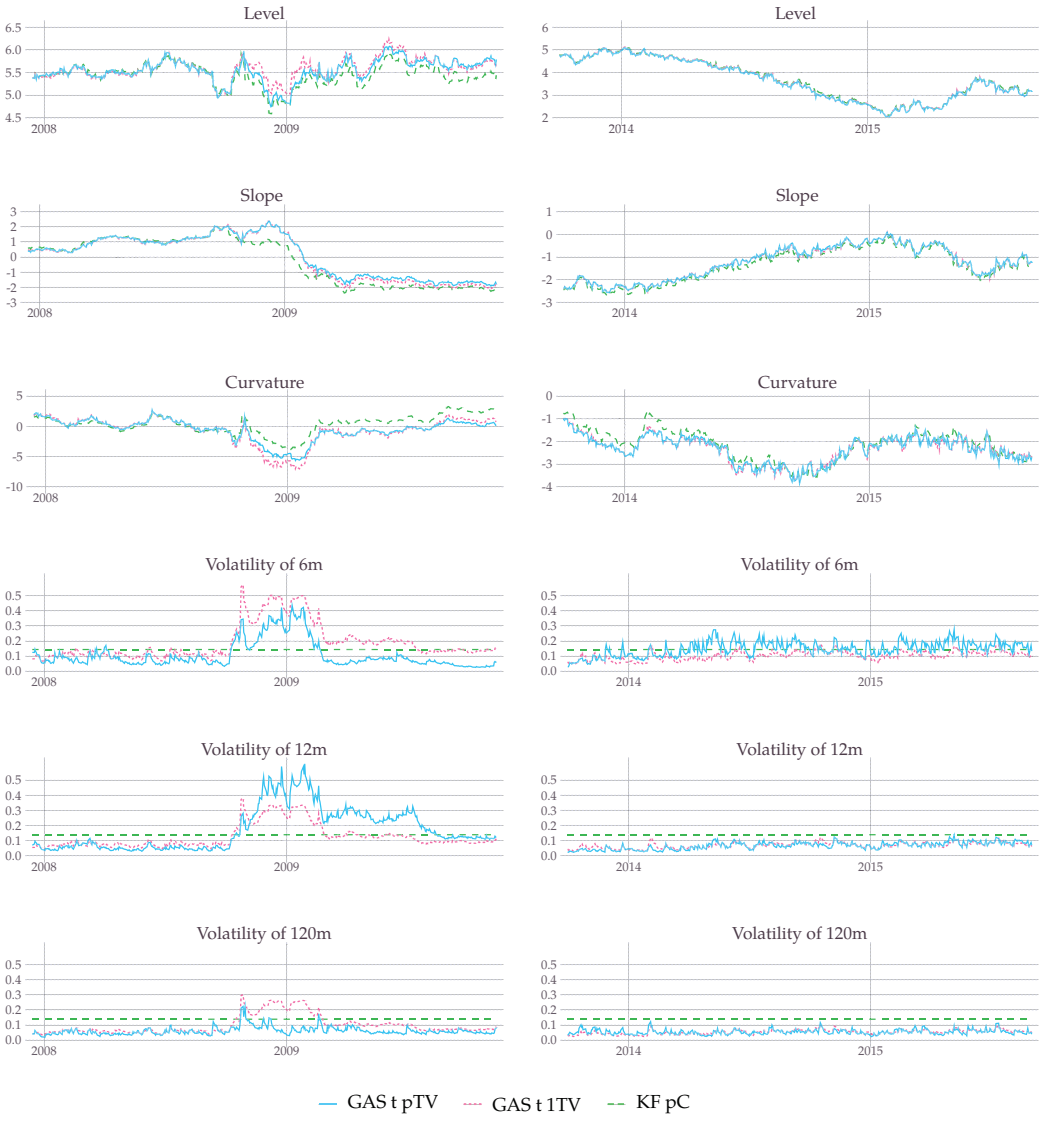
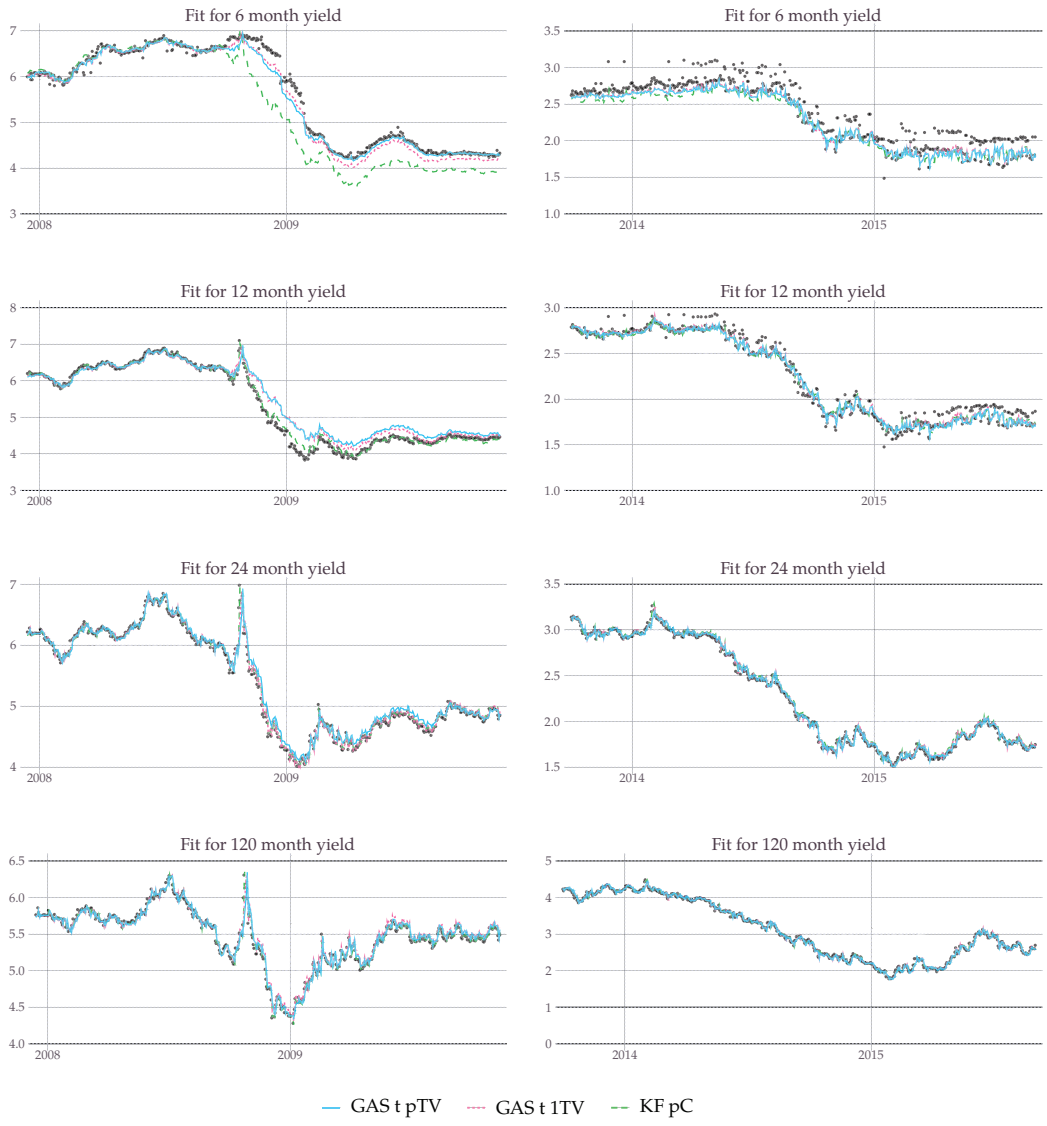


Figure Figure A.2

Estimated yields, Poland

We plot fitted data for a range of maturities as estimated with three of our models. The column on the left contains results around the 2008 financial crisis while the right column presents end of our sample period. We use daily data and the sample period is 2007–2015.



B. Scores and Fisher information

Here we provide derivations for the score vectors and Fisher information matrices in a multivariate Student's t distribution. The Fisher information matrix for multivariate t distribution is well studied (Mitchell, 1989; Arellano-Valle, 2010; Creal, Koopman, and Lucas, 2011). In what follows, we limit our attention to the single volatility factor model. We first derive the Fisher information for the latent mean factors and then for the time-varying volatility.

A variable y is said to be drawn from a multivariate $t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if it has the following density function:

$$\frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}(\nu\pi)^{-\frac{p}{2}}(|\boldsymbol{\Sigma}|)^{-\frac{1}{2}}\left[1+\frac{1}{\nu}(\boldsymbol{y}-\boldsymbol{\mu})^{\intercal}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right]^{-\frac{\nu+p}{2}}, \tag{B.19}$$

where $\boldsymbol{\mu}$ is the location vector, $\boldsymbol{\Sigma}$ is the scale matrix, ν denotes degrees of freedom, and p is the dimensionality. We can model the term-structure of interest rates by setting $\boldsymbol{\mu} = \boldsymbol{\Lambda}\boldsymbol{f}_t$. In this case, the log-likelihood function for a single observation takes the form:

$$\begin{aligned} l_t = &\log\left[\Gamma\left(\frac{\nu+p}{2}\right)\right] - \log\left[\Gamma\left(\frac{\nu}{2}\right)\right] - \frac{p}{2}\log(\nu\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}|) \\ &- \frac{\nu+p}{2}\log\left[1+\frac{1}{\nu}(\boldsymbol{y}_t-\boldsymbol{\Lambda}\boldsymbol{f}_t)^{\intercal}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t-\boldsymbol{\Lambda}\boldsymbol{f}_t)\right] \end{aligned} \tag{B.20}$$

Latent factors

We can filter the latent level, slope, and curvature of the yield curve using the score of the log-likelihood w.r.t. $\boldsymbol{f}_t$. Let $\boldsymbol{\varepsilon}_t = \boldsymbol{y}_t - \boldsymbol{\Lambda}\boldsymbol{f}$ and $\boldsymbol{K}_t = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\varepsilon}_t$,[16] then the score and the Hessian are:

$$\frac{\partial l}{\partial \boldsymbol{f}} = -\frac{\nu+p}{2}\left[1+\frac{1}{\nu}\boldsymbol{\varepsilon}_t^{\intercal}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\right]^{-1}\left(\frac{1}{\nu}\right)(-2)\,\boldsymbol{\Lambda}^{\intercal}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t$$

[16] $\boldsymbol{K}_t \sim t_\nu(\boldsymbol{0}, \boldsymbol{I})$, which means that K_t are no longer correlated but they maintain their tail-dependence.

$$= (\nu + p)\left[\nu + \boldsymbol{\varepsilon}_t^\intercal \boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\right]^{-1}\boldsymbol{\Lambda}^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t$$

$$= (\nu + p)\left[\nu + \boldsymbol{K}_t^\intercal\boldsymbol{K}_t\right]^{-1}\boldsymbol{\Lambda}^\intercal\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{K}_t; \tag{B.21}$$

$$\begin{aligned}\frac{\partial^2 l}{\partial \boldsymbol{f}^2} &= (\nu + p)\left[\nu + \boldsymbol{\varepsilon}_t^\intercal \boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\right]^{-1}\frac{\partial}{\partial \boldsymbol{f}}\left\{\boldsymbol{\Lambda}^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\right\} \\ &\quad + (\nu + p)\frac{\partial}{\partial \boldsymbol{f}}\left\{\left[\nu + \boldsymbol{\varepsilon}_t^\intercal \boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\right]^{-1}\right\}\boldsymbol{\varepsilon}_t^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda} \\ &= (\nu + p)\left\{2\frac{\boldsymbol{\Lambda}^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}}{\left[\nu + \boldsymbol{\varepsilon}_t^\intercal \boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\right]^2} - \frac{\boldsymbol{\Lambda}^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}}{\nu + \boldsymbol{\varepsilon}_t^\intercal \boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t}\right\} \\ &= \frac{\nu + p}{\nu}\left\{2\frac{\frac{1}{\nu}\boldsymbol{\Lambda}^\intercal\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{K}_t\boldsymbol{K}_t^\intercal\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Lambda}}{\left[1 + \frac{1}{\nu}\boldsymbol{K}_t^\intercal\boldsymbol{K}_t\right]^2} - \frac{\boldsymbol{\Lambda}^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}}{1 + \frac{1}{\nu}\boldsymbol{K}_t^\intercal\boldsymbol{K}_t}\right\}.\end{aligned} \tag{B.22}$$

The $f_t - f_t$ block of the Fisher information matrix, $\mathcal{I}$, can be derived by noting that $\mathcal{I}_{ff} = -\mathrm{E}\left[\frac{\partial^2 l}{\partial \boldsymbol{f}^2}\right]$. Note that at this stage we maintain full generality. The results we obtain are applicable even if the scale matrix, $\boldsymbol{\Sigma}$ is dense. For brevity of exposition, we denote $x_i = {}^{\star}_{i}D(\boldsymbol{\theta})$ if x_i is both an i-th independent (from all other draws considered) draw and a draw from a distribution D with parameters θ.[17] Additionally, $y = {}^{\star}_{i..j}D(\boldsymbol{\theta})$ if y is a transformation of the independent draws $\{x_{i..j}\}$ and is thus not independent of them. Given that $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$ (possibly $\boldsymbol{\Sigma}_t$) are known at time t, we can limit the analysis to derivation of:

$$\mathrm{E}\left[\frac{\frac{1}{\nu^2}\boldsymbol{K}_t\boldsymbol{K}_t^\intercal}{\left(1 + \frac{1}{\nu}\boldsymbol{K}_t^\intercal\boldsymbol{K}_t\right)^2}\right] \tag{B.23}$$

$$\mathrm{E}\left[\frac{1}{1 + \frac{1}{\nu}\boldsymbol{K}_t^\intercal\boldsymbol{K}_t}\right] \tag{B.24}$$

$\frac{1}{\nu}\boldsymbol{K}_t^\intercal\boldsymbol{K}_t$ expands to a convolution of F-distributed variables. In a general case it is hard to derive the distributional properties of this quantity, but we note that even though $\boldsymbol{K}_t$ was standardised and is thus uncorrelated it maintains tail

[17] For instance, if $x_i = {}^{\star}_{i}B(\alpha, \beta)$, $y_j = {}^{\star}_{j}B(\alpha, \beta)$, and $z_i = {}^{\star}_{i}B(\alpha, \beta)$ then x_i and y_j are independent draws from Beta distribution with shape parameters α and β while x_i and z_i refer to the same draw.

dependence of ε_t via a common draw from a $\Gamma\left(\frac{v}{2}, \frac{2}{v}\right)$ distribution, then:

$$\left[1+\frac{1}{\nu}\boldsymbol{K}_t^{\intercal}\boldsymbol{K}_t\right]^{-1} = \left[1+\frac{1}{\nu}\sum_{i=1}^{p}{}^{\star}_{i}F\left(1,\nu\right)\right]^{-1} = \left[1+\left({}^{\star}_{p+1}\Gamma\left(\frac{v}{2},2\right)\right)^{-1}\sum_{i=1}^{p}{}^{\star}_{i}\Gamma\left(\frac{1}{2},2\right)\right]^{-1}$$

$$\overset{18}{=} \frac{{}^{\star}_{p+1}\Gamma\left(\frac{v}{2},2\right)}{{}^{\star}_{p+1}\Gamma\left(\frac{v}{2},2\right)+{}^{\star}_{1..p}\Gamma\left(\frac{p}{2},2\right)} \tag{B.25}$$

$$= {}^{\star}_{1..p}B\left(\frac{v}{2},\frac{p}{2}\right) \tag{B.26}$$

$$\overset{E()}{=} \frac{\nu}{\nu+p}, \tag{B.27}$$

where in (B.25) we use the property that a convolution of independently drawn random variables from $\Gamma\left(\alpha_i,\beta\right)$ is distributed as $\Gamma\left(\sum_i \alpha,\beta\right)$; in (B.26) we use the property that $\frac{{}^{\star}_{i}\Gamma(\alpha_i,\beta)}{{}^{\star}_{i}\Gamma(\alpha_i,\beta)+{}^{\star}_{j}\Gamma(\alpha_j,\beta)}$ is distributed as $B\left(\alpha_i,\alpha_j\right)$; and in (B.27) we apply the expectation operator.

By similar arguments, the diagonal elements of the matrix in (B.23) are a product of ${}^{\star}_{1..p+1}B\left(\frac{1}{2},\frac{v+p-1}{2}\right)$ and ${}^{\star}_{1..p+1}B\left(\frac{v}{2},\frac{p}{2}\right)$ which in turn lie on a hyperplane of $Dir\left(\frac{1}{2},\frac{v}{2},\frac{p-1}{2}\right)$ and we denote them as ${}^{\star}_{p+1}Dir\left(\cdot\right)_{\ \frac{1}{2}}$ and ${}^{\star}_{p+1}Dir\left(\cdot\right)_{\ \frac{v}{2}}$ respectively. The relationship with Dirichlet distribution is easily seen by rewriting (B.23) into:

$$\frac{{}^{\star}_{1}\Gamma\left(\frac{1}{2},2\right)}{{}^{\star}_{1}\Gamma\left(\frac{1}{2},2\right)+{}^{\star}_{p+1}\Gamma\left(\frac{v}{2},2\right)+{}^{\star}_{2..p}\Gamma\left(\frac{p-1}{2},2\right)}\frac{{}^{\star}_{p+1}\Gamma\left(\frac{v}{2},2\right)}{{}^{\star}_{1}\Gamma\left(\frac{1}{2},2\right)+{}^{\star}_{p+1}\Gamma\left(\frac{v}{2},2\right)+{}^{\star}_{2..p}\Gamma\left(\frac{p-1}{2},2\right)}. \tag{B.28}$$

The expectation of diagonal elements in (B.23) is thus:

$$\mathrm{E}\left[{}^{\star}_{p+1}Dir\left(\cdot\right)_{\ \frac{1}{2}}{}^{\star}_{p+1}Dir\left(\cdot\right)_{\ \frac{v}{2}}\right] = \mathrm{E}\left[{}^{\star}_{p+1}Dir\left(\cdot\right)_{\ \frac{1}{2}}\right]\mathrm{E}\left[{}^{\star}_{p+1}Dir\left(\cdot\right)_{\ \frac{v}{2}}\right] + \mathrm{cov}\left[{}^{\star}_{p+1}Dir\left(\cdot\right)_{\ \frac{1}{2}},{}^{\star}_{p+1}Dir\left(\cdot\right)_{\ \frac{v}{2}}\right]$$

[18] Given that all constituent draws from the Gamma distributions in the denominator are independent, we also have ${}^{\star}_{p+1}\Gamma\left(\frac{v}{2},2\right)+{}^{\star}_{1..p}\Gamma\left(\frac{p}{2},2\right) \equiv {}^{\star}_{p+1}\Gamma\left(\frac{v}{2},2\right)+{}^{\star}_{1}\Gamma\left(\frac{1}{2},2\right)+{}^{\star}_{2..p}\Gamma\left(\frac{p-1}{2},2\right) \equiv {}^{\star}_{1..p+1}\Gamma\left(\frac{\nu+p}{2},2\right)$, which will be used at later stage.

$$
\begin{aligned}
&= \frac{\nu}{(\nu+p)^2} - \frac{2\nu}{(\nu+p)^2(\nu+p+2)} \\
&= \frac{v}{(v+p)(v+p+2)}
\end{aligned} \tag{B.29}
$$

It is possible to show that the off-diagonal elements are 0 in expectation. Combining (B.29) and (B.27) we get that the $\boldsymbol{f}-\boldsymbol{f}$ block of the Fisher information matrix is:

$$
\begin{aligned}
\mathcal{I}_{ff} &= -\frac{\nu+p}{\nu}\left[\frac{2v}{(v+p)(v+p+2)} - \frac{v}{v+p}\right]\boldsymbol{\Lambda}^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda} \\
&= \left(1 - \frac{2}{v+p+2}\right)\boldsymbol{\Lambda}^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda} \\
&= c_\nu\boldsymbol{\Lambda}^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}
\end{aligned} \tag{B.30}
$$

Since the score for $\boldsymbol{f}$ is antisymmetric w.r.t. $\boldsymbol{y}-\boldsymbol{\Lambda f}$ while score for $\boldsymbol{\Sigma}$ is symmetric, the off-diagonal blocks of the Fisher information matrix are 0 (contraction).

In a similar fashion we can model the single volatility factor, h_t. The score and the Hessian for h_t are:

$$
\frac{\partial l}{\partial h} = -\frac{1}{2}p + \frac{\nu+p}{2}\frac{\boldsymbol{K}_t^\intercal\boldsymbol{K}_t}{1+\frac{1}{\nu}\boldsymbol{K}_t^\intercal\boldsymbol{K}_t} \tag{B.31}
$$

$$
\frac{\partial^2 l}{\partial \boldsymbol{h}^2} = -\frac{\nu+p}{2}\frac{\boldsymbol{K}_t^\intercal\boldsymbol{K}_t}{\left(1+\frac{1}{\nu}\boldsymbol{K}_t^\intercal\boldsymbol{K}_t\right)^2} \tag{B.32}
$$

The derivation of the $h-h$ block of the Fisher information matrix is straightforward given (B.25). We restate the Hessian as:

$$
\begin{aligned}
\frac{\partial^2 l}{\partial \boldsymbol{h}^2} &= \left(\frac{\overset{\star}{}_{p+1}\Gamma\left(\frac{v}{2},2\right)}{\overset{\star}{}_{p+1}\Gamma\left(\frac{v}{2},2\right) + \overset{\star}{}_{1..p}\Gamma\left(\frac{p}{2},2\right)}\right)^2 \frac{\overset{\star}{}_{1..p}\Gamma\left(\frac{p}{2},2\right)}{\overset{\star}{}_{p+1}\Gamma\left(\frac{v}{2},2\right)} \\
&= \frac{\overset{\star}{}_{p+1}\Gamma\left(\frac{v}{2},2\right)}{\overset{\star}{}_{p+1}\Gamma\left(\frac{v}{2},2\right) + \overset{\star}{}_{1..p}\Gamma\left(\frac{p}{2},2\right)} \frac{\overset{\star}{}_{1..p}\Gamma\left(\frac{p}{2},2\right)}{\overset{\star}{}_{p+1}\Gamma\left(\frac{v}{2},2\right) + \overset{\star}{}_{1..p}\Gamma\left(\frac{p}{2},2\right)}
\end{aligned} \tag{B.33}
$$

Similarly to (B.28), the relationship with a two dimensional Dirichlet is appraent. The Hessian w.r.t. the single volatility factor is distributed as $Dir\left(\frac{v}{2},\frac{p}{2}\right)$. The

$h - h$ block of the Fisher information matrix then follows as:

$$\begin{aligned}\mathcal{I}_{hh} &= \frac{\nu + p}{2}\left(\frac{p\nu}{(p+\nu)^2} - \frac{2p\nu}{(p+\nu)^2(p+\nu+2)}\right) \\ &= \frac{p\nu}{2(p+\nu+2)}\end{aligned} \tag{B.34}$$

C. Time-series and cross-sectional performance in the simulation study

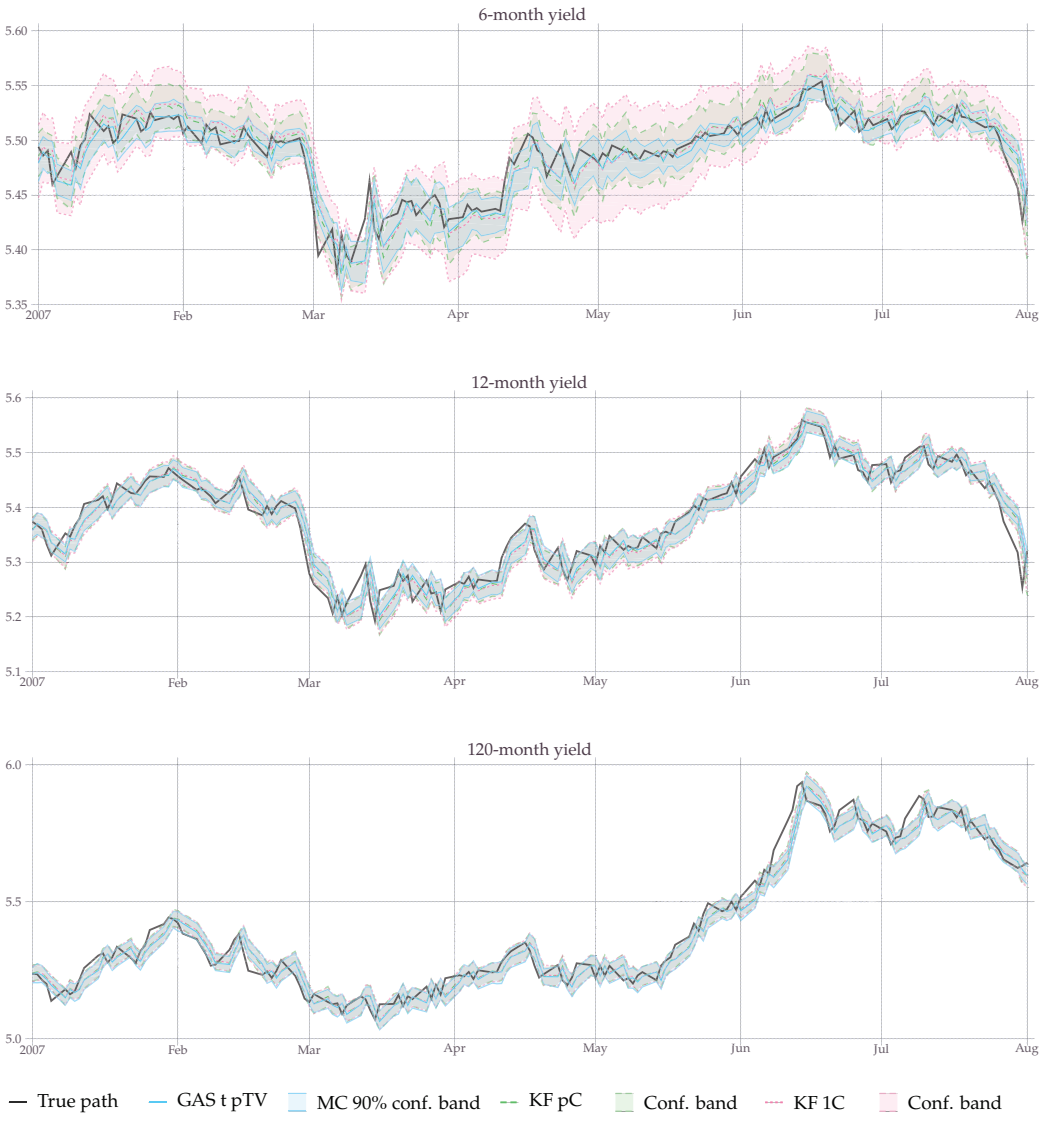
Figure Figure C.1 presents how well we are able to fit a selection of maturities over time. In this figure, we look at the '`Det t pTV`' DGP. As the factor paths are deterministic in this case, we are able to plot the results with Monte Carlo confidence bands. Here, we compare the '`GAS t pTV'` specification with the standard parameter-driven models which can be estimated with the Kalman Filter. It is clear from the figure that the largest differences are obtained for the short maturities. Consider the first panel with the 6-month interest rate. Although all three models track the true values of the yield correctly on average, the simple models produce much noisier results. This can be inferred from the much wider Monte Carlo confidence bands obtained for the simple models. These differences are due to a varying precision with which the unobserved components can be filtered from the data using the different methods. For longer maturities, all methods perform similarly. This is in line with the structure of the DNSM: the long term yield corresponds to the first factor ($f_{L,t}$), which is the dominant factor in the yield curve dynamics. All methods are capable of filtering this dominant signal from the data. Filtering the 6-month yield signal from the data is harder for two reasons. First, the 6-month yield is typically much more volatile, particularly so over the sample period, which includes the financial crisis. Second, given that the long-term yield captures the level of the yield curve, filtering the short yield requires us to also have accurate descriptions of the more subtle dynamic aspects of the yield curve, in particular the slope and also the curvature. The difference in factor tracking across the three models is also clearly visible in the top panels of Figures Figure C.2 and Figure C.3.

To shed more light on the performance of the models in different periods of time, in Figures Figure C.2 and Figure C.3, we plot filtered factor patterns and parameter precision for a high-volatility and a low-volatility period, respectively. For the high-volatility period, we see that the GAS model yields more precise

Figure Figure C.1

Time-series performance

This figure shows how well the competing methods are able to fit the evolution of yields in the '`Det t pTV`' data-generating process. In this DGP, all replications share the factor and interest rate paths in common. We show paths of the median estimates along with their 90% Monte Carlo confidence intervals. The dates on the horizontal axis refer to the time period from which the deterministic paths in the DGP were sourced.



estimates of particularly the slope and the curvature of the yield curve compared to the state space specifications with constant variances and thin-tailed distributions. In particular, we see that it is important to allow for time-varying

Figure Figure C.2

Factors and cross-sections in high volatility regimes

This figure shows how well the competing methods are able to track the factors and fit the cross-section of interest rates for the '`Det t pTV`' data-generating process in high volatility regimes. In this DGP, all replications share the factor and interest rate paths in common. We show paths of the median estimates along with their 90% Monte Carlo confidence intervals. The top two rows contain results for the latent factors: level, slope, curvature, and two (out of 11) volatilities. We also report the distribution of the decay parameter, λ. In the bottom two rows we look at differences in the fit in the cross-section of interest rates. The third row considers three synthetic days at the beginning, in the middle, and towards the end of a 'crisis' with a sudden increase in volatility (second row). The fourth row decomposes one such synthetic day and shows the fit for arbitrarily chosen replications.

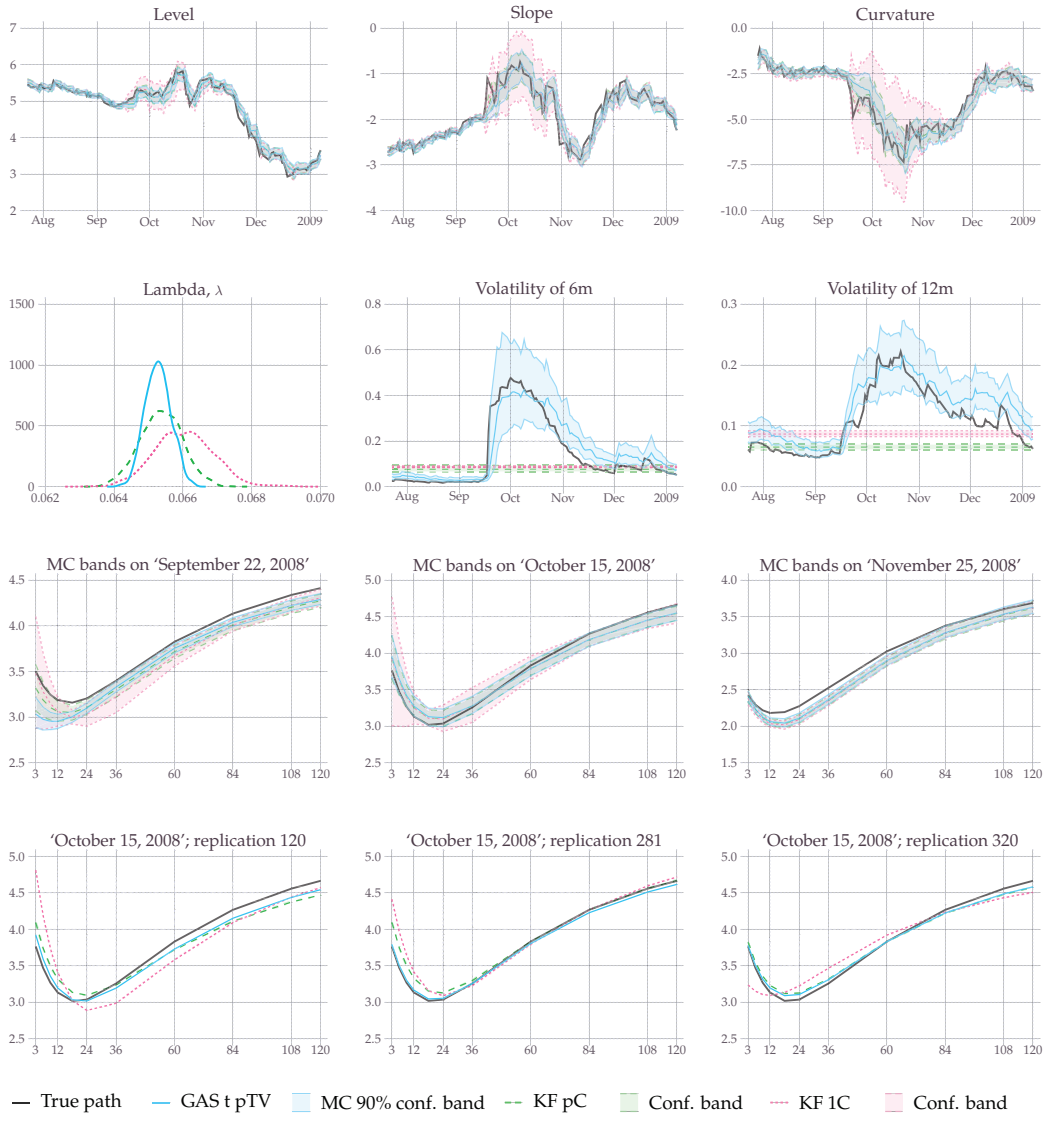
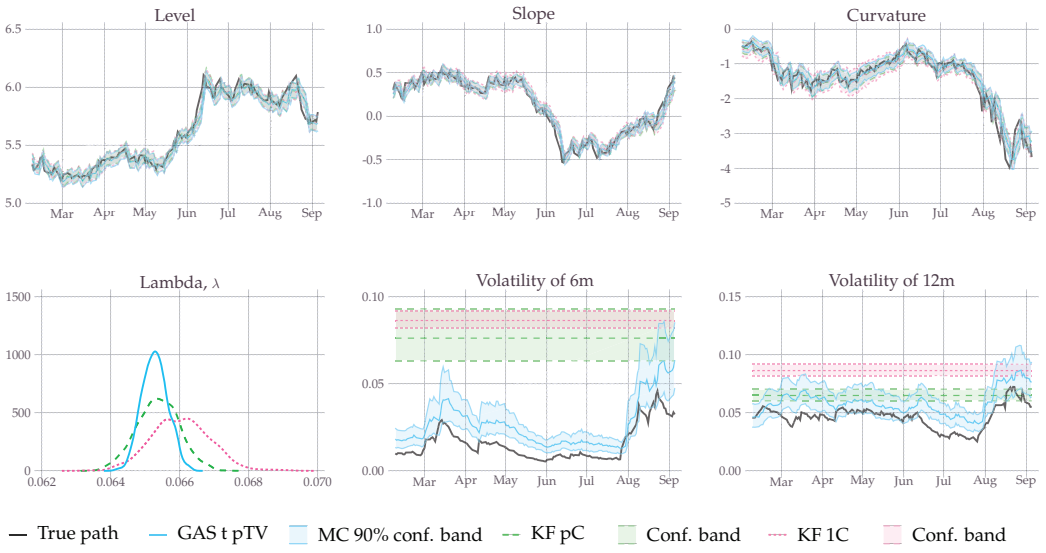


Figure Figure C.3

Factor tracking in low volatility regimes

This figure shows how well the competing methods are able to track the factors and fit the cross-section of interest rates for the '`Det t pTV`' data-generating process in low volatility regimes. See Figure Figure C.2 for detailed description.



variances in the short and medium rates over the heat of the financial crisis in 2008. The constant variances of the plain vanilla state space models appear substantially misspecified during this period.

The effect is further underlined by considering the fitted yield curves on the simulated 'October 15, 2008', for the three models across 3 arbitrarily chosen simulation replications. This is done in the bottom row of panels in Figure Figure C.2. The strongly inverted U-shaped yield curve is rather uncommon given the overall historical pattern of the data. For all three simulations, however, it is clear that the GAS specification is much better able to adapt to this non-standard yield curve pattern than the state space models with fixed variances, even though also these models allow for time-variation in the coefficients of the DNSMs.

The good tracking performance of the GAS model in this multivariate context is further corroborated by looking at the low-volatility regime in Figure Figure C.3. The filtered dynamic patterns of the three latent mean factors are quite

comparable across models. Also in this case, the curvature factor shows the biggest differences. The state space models, however, estimate the volatility of the short and medium yields far too high. The GAS models, by contrast, turn out to be much closer to the low volatility patterns used to generate the data. In addition, the GAS model appears to be able to capture the dynamics of volatility well, even though the level is still slightly too high, though the error is much smaller than for the state space models with constant variances. This is not surprising given the misspecified nature of the GAS model. In fact, Nelson and Foster (1994) attribute this outcome to the inability of the filter to match the true 'variance of the variance' everywhere in the state space. Other examples of observation-driven filters producing 'bias-at-extremes' can be seen in Koopman, Lit, and Lucas (2015), Creal, Koopman, Lucas, and Zamojski (2015), or Zamojski (2016).

D. Estimation details and results

We are able to clearly show benefits of relaxing assumptions individually because the score-driven models which we consider in the paper are nested. In this section we present a general score-driven Nelson-Siegel model and lay down all the restriction which we impose for parsimony. All ten GAS models can be derived from the following specification by imposing further parameter restrictions. Note that the speed and stability of estimation can be considerably improved for simpler models (e.g., '`GAS t 1TV`') using the analytical results which we provide in Section 3. We refer you to Section 3 for description of the notation.

$$\boldsymbol{y}_t = \begin{bmatrix} \mathbf{1} & \frac{1-e^{-\lambda\tau}}{\lambda\tau} & \frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \end{bmatrix} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim t\left(0, \boldsymbol{\Sigma}, \nu\right),$$

$$\boldsymbol{\Sigma} = \boldsymbol{D}_t \left[(1-\rho)\, \boldsymbol{I} + \rho \boldsymbol{\iota}\boldsymbol{\iota}^\intercal \right] \boldsymbol{D}_t, \qquad \boldsymbol{D}_t^2 = \operatorname{diag}\left(\sigma_{3m}^2 e^{h_{3m,t}}, \ldots, \sigma_{120m}^2 e^{h_{120m,t}} \right),$$

$$\begin{bmatrix} \boldsymbol{f}_{t+1} \\ \boldsymbol{h}_{t+1} \end{bmatrix} = \begin{bmatrix} f_{L,t+1} \\ f_{S,t+1} \\ f_{C,t+1} \\ h_{3m,t+1} \\ \vdots \\ h_{120m,t+1} \end{bmatrix} = \boldsymbol{\omega} + \boldsymbol{B} \left(\begin{bmatrix} \boldsymbol{f}_t \\ \boldsymbol{h}_t \end{bmatrix} - \boldsymbol{\omega} \right) + \boldsymbol{A} s_t, \qquad \boldsymbol{\omega} = \begin{bmatrix} \omega_L \\ \omega_S \\ \omega_C \\ \omega_{3m} \\ \vdots \\ \omega_{120m} \end{bmatrix},$$

$$\boldsymbol{A} = \begin{bmatrix} A_{LL} & A_{LS} & A_{LC} & & 0 & \\ A_{SL} & A_{SS} & A_{SC} & & & \\ A_{CL} & A_{CS} & A_{CC} & & & \\ & & & A_{3m} & A_{OD} & \ldots & A_{OD} \\ & 0 & & A_{OD} & \ddots & \ddots & \vdots \\ & & & \vdots & \ddots & \ddots & A_{OD} \\ & & & A_{OD} & \ldots & A_{OD} & A_{120m} \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} B_{LL} & B_{LS} & B_{LC} & & 0 & \\ B_{SL} & B_{SS} & B_{SC} & & & \\ B_{CL} & B_{CS} & B_{CC} & & & \\ & & & B_{3m} & B_{OD} & \ldots & B_{OD} \\ & 0 & & B_{OD} & \ddots & \ddots & \vdots \\ & & & \vdots & \ddots & \ddots & B_{OD} \\ & & & B_{OD} & \ldots & B_{OD} & B_{120m} \end{bmatrix},$$

where s_t is the score-driven step as defined in Section 2. Given this specification, Tables D.1 and D.2 contain estimation results for the U.S. and the Polish data respectively. We report parameter estimates with robust standard errors.

Table D.1

Parameter estimates, U.S.

The table contains estimation results for 10 observation-driven models which we consider in the paper. Here, we look at the U.S. market. The sample period is 1997–2015 and contains 4,815 daily observations. We report robust standard errors in parentheses. The table also includes implicit and non-zero parameter restrictions where applicable.

	'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'
ν	19.451	∞	26.382	∞	34.573	∞	17.321	2.686	∞	∞
	(3.590)		(3.714)		(5.930)		(1.098)	(0.159)		
λ	0.061	0.058	0.059	0.053	0.051	0.049	0.052	0.051	0.053	0.054
	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
ω_L	6.851	6.526	8.095	7.437	7.327	7.410	7.123	7.204	7.376	7.371
	(0.213)	(0.206)	(0.151)	(0.216)	(0.096)	(0.124)	(0.085)	(0.085)	(0.094)	(0.094)
ω_S	-1.225	-0.981	-1.944	-1.572	-1.554	-1.619	-1.520	-1.556	-1.619	-1.635
	(0.167)	(0.157)	(0.080)	(0.110)	(0.060)	(0.069)	(0.046)	(0.042)	(0.045)	(0.049)
ω_C	0.248	-1.194	-2.684	0.384	0.965	0.748	1.825	1.465	0.773	0.749
	(0.527)	(0.159)	(0.777)	(0.954)	(0.342)	(0.447)	(0.372)	(0.383)	(0.364)	(0.359)
A_{LL}	0.471	0.753	1.171	1.145	1.140	1.109	1.080	1.120	1.186	1.242
	(0.078)	(0.030)	(0.062)	(0.062)	(0.076)	(0.067)	(0.166)	(0.138)	(0.053)	(0.059)
A_{LS}	-0.294	-0.088	0.099	0.134	0.081	0.086	0.062	0.082	0.163	0.229
	(0.081)	(0.032)	(0.054)	(0.053)	(0.064)	(0.059)	(0.134)	(0.111)	(0.045)	(0.051)
A_{LC}	-0.039	0.000	0.048	0.038	0.075	0.055	0.038	0.041	0.042	0.055
	(0.012)	(0.011)	(0.010)	(0.009)	(0.010)	(0.009)	(0.026)	(0.021)	(0.009)	(0.010)
A_{SL}	0.348	0.104	-0.186	0.052	-0.015	-0.099	-0.032	-0.063	-0.186	-0.209
	(0.078)	(0.023)	(0.069)	(0.071)	(0.076)	(0.082)	(0.103)	(0.105)	(0.091)	(0.097)
A_{SS}	1.105	0.932	0.894	1.042	1.010	0.893	0.994	0.982	0.791	0.749
	(0.087)	(0.026)	(0.064)	(0.060)	(0.068)	(0.075)	(0.087)	(0.094)	(0.082)	(0.090)
A_{SC}	0.043	0.010	-0.052	-0.015	-0.042	-0.035	-0.028	-0.034	-0.025	-0.042
	(0.011)	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)	(0.014)	(0.013)	(0.013)	(0.014)
A_{CL}	0.577	0.106	0.527	-0.588	0.246	-0.163	1.237	0.851	-0.189	-0.399
	(0.111)	(0.069)	(0.198)	(0.239)	(0.211)	(0.192)	(0.604)	(0.595)	(0.252)	(0.268)
A_{CS}	0.577	0.147	0.395	-0.553	0.154	-0.145	0.926	0.610	-0.109	-0.355
	(0.105)	(0.070)	(0.174)	(0.212)	(0.184)	(0.177)	(0.475)	(0.469)	(0.230)	(0.244)
A_{CC}	0.845	0.816	1.088	0.925	0.984	0.912	1.149	1.092	0.884	0.839
	(0.038)	(0.020)	(0.034)	(0.034)	(0.031)	(0.028)	(0.080)	(0.063)	(0.036)	(0.043)
B_{LL}	0.999	1.000	1.000	0.997	0.998	0.999	1.000	1.000	0.998	0.998
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
B_{LS}	0.004	0.001	0.001	-0.001	0.001	0.001	-0.000	-0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
B_{LC}	-0.002	-0.003	0.001	0.003	0.002	0.001	-0.000	0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
B_{SL}	0.001	-0.000	-0.000	0.002	0.001	-0.000	-0.007	-0.006	-0.001	-0.001
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)

Table D.1

Parameter estimates, U.S. (continued)

	'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'
B_{SS}	0.997	0.996	0.998	1.000	0.998	0.997	0.997	0.996	0.997	0.997
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
B_{SC}	0.001	0.007	0.001	-0.002	-0.000	0.001	0.005	0.005	0.002	0.002
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
B_{CL}	0.001	0.003	0.001	0.006	0.004	0.003	0.003	0.003	0.005	0.004
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
B_{CS}	-0.001	-0.001	0.002	0.004	-0.001	0.000	-0.001	-0.003	0.001	0.000
	(0.001)	(0.002)	(0.003)	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
B_{CC}	1.000	1.000	0.994	0.993	0.997	0.997	1.000	1.000	0.997	0.997
	(0.001)	(0.001)	(0.003)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
ω_{3m}	-4.592	-2.936	-6.398	-5.084	-5.448	-5.665				
	(0.126)	(0.283)	(0.350)	(0.153)	(0.127)	(0.182)				
ω_{6m}	-4.240	-3.049	-6.849	-8.173	ω_{3m}	ω_{3m}				
	(0.146)	(0.382)	(0.403)	(0.405)						
ω_{9m}	-3.139	-2.661	-6.094	-6.484	ω_{3m}	ω_{3m}				
	(0.166)	(0.397)	(0.198)	(0.277)						
ω_{12m}	-2.770	-2.340	-5.944	-6.075	ω_{3m}	ω_{3m}				
	(0.203)	(0.322)	(0.231)	(0.283)						
ω_{18m}	-2.499	-2.042	-4.263	-6.113	ω_{3m}	ω_{3m}				
	(0.189)	(0.299)	(0.376)	(0.461)						
ω_{24m}	-1.961	-1.979	-5.003	-5.773	ω_{3m}	ω_{3m}				
	(0.152)	(0.309)	(0.194)	(0.137)						
ω_{36m}	-1.611	-1.735	-4.046	-5.134	ω_{3m}	ω_{3m}				
	(0.088)	(0.229)	(0.165)	(0.403)						
ω_{60m}	-2.899	-1.375	-3.964	-4.621	ω_{3m}	ω_{3m}				
	(0.139)	(0.254)	(0.138)	(0.147)						
ω_{84m}	-1.612	-1.400	-4.657	-5.198	ω_{3m}	ω_{3m}				
	(0.083)	(0.256)	(0.166)	(0.140)						
ω_{108m}	-1.703	-1.504	-5.155	-5.667	ω_{3m}	ω_{3m}				
	(0.117)	(0.253)	(0.217)	(0.149)						
ω_{120m}	-1.710	-1.644	-5.121	-5.425	ω_{3m}	ω_{3m}				
	(0.113)	(0.271)	(0.136)	(0.115)						
A_{3m}	0.235	0.155	0.197	0.213	0.115	0.028				
	(0.011)	(0.007)	(0.020)	(0.032)	(0.029)	(0.003)				
A_{6m}	0.240	0.140	0.164	0.054	A_{3m}	A_{3m}				
	(0.013)	(0.006)	(0.013)	(0.016)						
A_{9m}	0.199	0.116	0.153	0.051	A_{3m}	A_{3m}				
	(0.009)	(0.005)	(0.017)	(0.009)						
A_{12m}	0.196	0.107	0.106	0.048	A_{3m}	A_{3m}				
	(0.008)	(0.005)	(0.017)	(0.006)						

Table D.1

Parameter estimates, U.S. (continued)

	'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'
A_{18m}	0.186	0.104	0.100	0.031	A_{3m}	A_{3m}				
	(0.013)	(0.005)	(0.013)	(0.004)						
A_{24m}	0.168	0.101	0.100	0.035	A_{3m}	A_{3m}				
	(0.015)	(0.006)	(0.012)	(0.007)						
A_{36m}	0.165	0.099	0.104	0.045	A_{3m}	A_{3m}				
	(0.018)	(0.006)	(0.011)	(0.008)						
A_{60m}	0.138	0.103	0.090	0.057	A_{3m}	A_{3m}				
	(0.015)	(0.006)	(0.006)	(0.008)						
A_{84m}	0.156	0.096	0.046	0.030	A_{3m}	A_{3m}				
	(0.015)	(0.006)	(0.004)	(0.005)						
A_{108m}	0.169	0.088	0.038	0.022	A_{3m}	A_{3m}				
	(0.016)	(0.006)	(0.006)	(0.005)						
A_{120m}	0.174	0.084	0.047	0.030	A_{3m}	A_{3m}				
	(0.016)	(0.006)	(0.007)	(0.007)						
A_{OD}	0.009	-0.005	0.009	0.002						
	(0.007)	(0.001)	(0.002)	(0.001)						
B_{3m}	0.971	0.993	0.981	0.918	0.967	0.990				
	(0.005)	(0.001)	(0.003)	(0.015)	(0.011)	(0.002)				
B_{6m}	0.978	0.994	0.982	0.990	B_{3m}	B_{3m}				
	(0.004)	(0.001)	(0.003)	(0.004)						
B_{9m}	0.988	0.997	0.980	0.992	B_{3m}	B_{3m}				
	(0.003)	(0.001)	(0.004)	(0.002)						
B_{12m}	0.988	0.998	0.987	0.994	B_{3m}	B_{3m}				
	(0.003)	(0.001)	(0.004)	(0.001)						
B_{18m}	0.986	0.996	0.994	0.998	B_{3m}	B_{3m}				
	(0.005)	(0.001)	(0.002)	(0.001)						
B_{24m}	0.988	0.994	0.982	0.990	B_{3m}	B_{3m}				
	(0.004)	(0.001)	(0.005)	(0.003)						
B_{36m}	0.987	0.989	0.987	0.996	B_{3m}	B_{3m}				
	(0.004)	(0.002)	(0.003)	(0.003)						
B_{60m}	1.000	0.993	0.988	0.992	B_{3m}	B_{3m}				
	(0.002)	(0.001)	(0.002)	(0.002)						
B_{84m}	0.990	0.995	0.993	0.994	B_{3m}	B_{3m}				
	(0.004)	(0.001)	(0.001)	(0.003)						
B_{108m}	0.985	0.996	0.994	0.994	B_{3m}	B_{3m}				
	(0.005)	(0.001)	(0.002)	(0.003)						
B_{120m}	0.983	0.996	0.990	0.991	B_{3m}	B_{3m}				
	(0.005)	(0.001)	(0.003)	(0.004)						
B_{OD}	-0.005	0.000	-0.001	-0.000						
	(0.003)	(0.000)	(0.000)	(0.000)						

Table D.1

Parameter estimates, U.S. (continued)

	'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'
ρ	0.987 (0.001)	0.986 (0.001)								
σ^2_{3m}	1.000	1.000	1.000	1.000	2.018 (0.074)	1.696 (0.086)	0.009 (0.000)	0.005 (0.000)	0.012 (0.001)	0.006 (0.000)
σ^2_{6m}	1.000	1.000	1.000	1.000	0.195 (0.007)	0.222 (0.015)	0.001 (0.000)	σ^2_{3m}	0.002 (0.000)	σ^2_{3m}
σ^2_{9m}	1.000	1.000	1.000	1.000	0.577 (0.007)	0.580 (0.007)	0.002 (0.000)	σ^2_{3m}	0.005 (0.000)	σ^2_{3m}
σ^2_{12m}	1.000	1.000	1.000	1.000	1.000	1.000	0.004 (0.000)	σ^2_{3m}	0.009 (0.001)	σ^2_{3m}
σ^2_{18m}	1.000	1.000	1.000	1.000	0.691 (0.012)	0.689 (0.013)	0.003 (0.000)	σ^2_{3m}	0.004 (0.000)	σ^2_{3m}
σ^2_{24m}	1.000	1.000	1.000	1.000	0.973 (0.043)	0.732 (0.043)	0.003 (0.000)	σ^2_{3m}	0.003 (0.000)	σ^2_{3m}
σ^2_{36m}	1.000	1.000	1.000	1.000	2.796 (0.121)	1.964 (0.138)	0.007 (0.000)	σ^2_{3m}	0.007 (0.000)	σ^2_{3m}
σ^2_{60m}	1.000	1.000	1.000	1.000	3.842 (0.152)	2.869 (0.180)	0.011 (0.000)	σ^2_{3m}	0.012 (0.000)	σ^2_{3m}
σ^2_{84m}	1.000	1.000	1.000	1.000	1.825 (0.067)	1.429 (0.078)	0.005 (0.000)	σ^2_{3m}	0.006 (0.000)	σ^2_{3m}
σ^2_{108m}	1.000	1.000	1.000	1.000	1.519 (0.065)	1.184 (0.068)	0.004 (0.000)	σ^2_{3m}	0.004 (0.000)	σ^2_{3m}
σ^2_{120m}	1.000	1.000	1.000	1.000	2.048 (0.088)	1.567 (0.093)	0.006 (0.000)	σ^2_{3m}	0.006 (0.000)	σ^2_{3m}

Table D.2

Parameter estimates, Poland

The table contains estimation results for 10 observation-driven models which we consider in the paper. Here, we look at the Polish market. The sample period is 2007–2015 and contains 2,262 daily observations. We report robust standard errors in parentheses. The table also includes implicit and non-zero parameter restrictions where applicable.

	'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'
ν	10.608	∞	19.219	∞	24.895	∞	15.767	1.227	∞	∞
	(2.560)		(3.702)		(3.714)		(3.608)	(0.126)		
λ	0.046	0.045	0.055	0.055	0.054	0.053	0.057	0.057	0.061	0.062
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ω_L	5.234	5.212	5.385	5.233	5.385	5.303	5.327	5.380	5.310	5.377
	(0.068)	(0.163)	(0.060)	(0.059)	(0.061)	(0.064)	(0.352)	(0.021)	(0.043)	(0.037)
ω_S	-0.970	-0.921	-1.111	-1.068	-1.056	-1.031	-0.860	-1.116	-1.046	-1.133
	(0.022)	(0.042)	(0.060)	(0.033)	(0.051)	(0.037)	(0.665)	(0.049)	(0.043)	(0.044)
ω_C	0.482	0.503	-0.300	0.276	-0.337	-0.096	-0.551	-0.366	-0.228	-0.444
	(0.211)	(0.412)	(0.193)	(0.217)	(0.230)	(0.243)	(0.234)	(0.198)	(0.162)	(0.111)
A_{LL}	0.639	0.665	1.061	0.921	0.982	0.875	1.098	0.839	0.930	0.985
	(0.051)	(0.071)	(0.056)	(0.049)	(0.049)	(0.045)	(0.071)	(0.061)	(0.051)	(0.049)
A_{LS}	-0.166	-0.241	-0.009	-0.143	-0.102	-0.155	-0.039	0.021	-0.115	-0.060
	(0.044)	(0.056)	(0.032)	(0.034)	(0.028)	(0.031)	(0.058)	(0.044)	(0.032)	(0.028)
A_{LC}	-0.020	-0.021	0.057	-0.002	0.044	0.004	0.064	0.055	0.005	0.031
	(0.015)	(0.018)	(0.014)	(0.015)	(0.012)	(0.013)	(0.041)	(0.012)	(0.016)	(0.013)
A_{SL}	-0.121	-0.122	-0.652	-0.647	-0.719	-0.616	-0.770	-0.289	-0.773	-0.827
	(0.041)	(0.062)	(0.072)	(0.060)	(0.066)	(0.061)	(0.087)	(0.115)	(0.073)	(0.075)
A_{SS}	0.772	0.852	0.514	0.503	0.486	0.528	0.444	0.562	0.429	0.373
	(0.051)	(0.044)	(0.041)	(0.043)	(0.042)	(0.043)	(0.054)	(0.075)	(0.050)	(0.048)
A_{SC}	0.008	0.021	-0.074	-0.079	-0.071	-0.031	-0.112	-0.035	-0.067	-0.093
	(0.010)	(0.015)	(0.017)	(0.020)	(0.017)	(0.017)	(0.024)	(0.025)	(0.024)	(0.021)
A_{CL}	0.704	1.198	1.099	1.570	1.830	1.763	1.342	0.688	2.003	1.743
	(0.155)	(0.182)	(0.192)	(0.158)	(0.178)	(0.156)	(0.432)	(0.318)	(0.247)	(0.237)
A_{CS}	0.522	1.146	0.907	1.465	1.504	1.519	1.293	0.463	1.602	1.278
	(0.138)	(0.205)	(0.159)	(0.120)	(0.134)	(0.123)	(0.265)	(0.196)	(0.176)	(0.151)
A_{CC}	0.807	1.024	0.960	1.226	1.098	1.139	1.105	0.725	1.265	1.113
	(0.059)	(0.047)	(0.069)	(0.057)	(0.054)	(0.049)	(0.204)	(0.076)	(0.080)	(0.069)
B_{LL}	1.000	1.000	0.996	0.996	0.996	0.995	0.998	0.999	0.997	0.998
	(0.003)	(0.009)	(0.002)	(0.002)	(0.002)	(0.003)	(0.005)	(0.001)	(0.002)	(0.002)
B_{LS}	-0.002	0.002	0.000	0.003	0.000	0.001	-0.002	0.003	0.001	0.000
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)
B_{LC}	0.002	-0.000	0.002	0.002	0.002	0.003	-0.004	-0.001	0.002	0.001
	(0.002)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
B_{SL}	0.001	-0.002	0.001	0.003	0.002	0.004	-0.001	-0.002	-0.001	-0.002
	(0.003)	(0.004)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)

Table D.2

Parameter estimates, Poland (continued)

	'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'
B_{SS}	0.998	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	0.999
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
B_{SC}	0.001	0.004	0.002	-0.001	0.001	-0.000	0.010	0.010	0.004	0.005
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
B_{CL}	-0.007	-0.000	0.012	0.002	0.005	0.009	-0.003	0.007	0.010	0.009
	(0.007)	(0.018)	(0.005)	(0.006)	(0.005)	(0.005)	(0.021)	(0.004)	(0.006)	(0.006)
B_{CS}	0.000	-0.010	-0.005	-0.015	-0.006	-0.003	-0.010	-0.022	-0.006	-0.006
	(0.003)	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.025)	(0.003)	(0.005)	(0.005)
B_{CC}	1.000	1.000	0.995	1.000	0.998	0.994	1.000	0.990	0.989	0.990
	(0.003)	(0.008)	(0.004)	(0.003)	(0.003)	(0.003)	(0.012)	(0.002)	(0.004)	(0.004)
ω_{3m}	-2.811	-2.971	-4.281	-3.888	-5.032	-5.272				
	(0.501)	(0.439)	(0.198)	(0.324)	(0.135)	(0.189)				
ω_{6m}	-3.363	-3.521	-4.596	-4.901	ω_{3m}	ω_{3m}				
	(0.243)	(0.396)	(0.243)	(0.571)						
ω_{9m}	-2.746	-3.636	-5.899	-5.985	ω_{3m}	ω_{3m}				
	(0.554)	(1.128)	(0.184)	(0.141)						
ω_{12m}	-4.405	-4.745	-5.504	-5.448	ω_{3m}	ω_{3m}				
	(0.190)	(0.540)	(0.263)	(0.214)						
ω_{18m}	-4.779	-4.678	-5.235	-5.593	ω_{3m}	ω_{3m}				
	(0.141)	(0.258)	(0.165)	(0.178)						
ω_{24m}	-4.891	-5.754	-5.968	-6.718	ω_{3m}	ω_{3m}				
	(0.150)	(0.290)	(0.176)	(0.247)						
ω_{36m}	-4.431	-5.150	-6.004	-6.127	ω_{3m}	ω_{3m}				
	(0.154)	(0.637)	(0.121)	(0.190)						
ω_{60m}	-4.430	-5.027	-5.484	-5.802	ω_{3m}	ω_{3m}				
	(0.146)	(2.607)	(0.148)	(0.203)						
ω_{84m}	-4.626	-4.909	-5.661	-5.940	ω_{3m}	ω_{3m}				
	(0.150)	(0.451)	(0.146)	(0.143)						
ω_{108m}	-4.676	-4.964	-5.977	-6.103	ω_{3m}	ω_{3m}				
	(0.146)	(1.062)	(0.124)	(0.126)						
ω_{120m}	-4.741	-5.132	-5.893	-5.993	ω_{3m}	ω_{3m}				
	(0.135)	(0.276)	(0.117)	(0.141)						
A_{3m}	0.179	0.058	0.174	0.137	0.213	0.033				
	(0.016)	(0.010)	(0.017)	(0.075)	(0.040)	(0.005)				
A_{6m}	0.172	0.073	0.136	0.061	A_{3m}	A_{3m}				
	(0.018)	(0.012)	(0.014)	(0.019)						
A_{9m}	0.170	0.087	0.141	0.115	A_{3m}	A_{3m}				
	(0.013)	(0.030)	(0.016)	(0.016)						
A_{12m}	0.200	0.086	0.114	0.073	A_{3m}	A_{3m}				
	(0.022)	(0.010)	(0.017)	(0.014)						

Table D.2

Parameter estimates, Poland (continued)

	'GAS t pIV+c'	'GAS N pIV+c'	'GAS t pIV'	'GAS N pIV'	'GAS t 1IV'	'GAS N 1IV'	'GAS t pC'	'GAS t 1C'	'GAS N pC'	'GAS N 1C'
A_{18m}	0.194	0.081	0.111	0.049	A_{3m}	A_{3m}				
	(0.021)	(0.015)	(0.027)	(0.013)						
A_{24m}	0.223	0.095	0.190	0.039	A_{3m}	A_{3m}				
	(0.017)	(0.034)	(0.038)	(0.006)						
A_{36m}	0.201	0.079	0.180	0.060	A_{3m}	A_{3m}				
	(0.021)	(0.047)	(0.023)	(0.019)						
A_{60m}	0.218	0.078	0.185	0.067	A_{3m}	A_{3m}				
	(0.021)	(0.066)	(0.026)	(0.013)						
A_{84m}	0.202	0.098	0.175	0.068	A_{3m}	A_{3m}				
	(0.022)	(0.070)	(0.029)	(0.017)						
A_{108m}	0.181	0.088	0.178	0.076	A_{3m}	A_{3m}				
	(0.022)	(0.046)	(0.027)	(0.022)						
A_{120m}	0.182	0.096	0.162	0.059	A_{3m}	A_{3m}				
	(0.020)	(0.049)	(0.030)	(0.015)						
A_{OD}	0.028	-0.000	0.021	0.001						
	(0.007)	(0.002)	(0.005)	(0.001)						
B_{3m}	0.982	1.000	0.956	0.961	0.931	0.985				
	(0.005)	(0.002)	(0.009)	(0.013)	(0.017)	(0.005)				
B_{6m}	0.976	0.992	0.977	0.993	B_{3m}	B_{3m}				
	(0.006)	(0.004)	(0.005)	(0.008)						
B_{9m}	0.985	0.995	0.962	0.948	B_{3m}	B_{3m}				
	(0.004)	(0.012)	(0.008)	(0.011)						
B_{12m}	0.968	0.993	0.980	0.973	B_{3m}	B_{3m}				
	(0.007)	(0.005)	(0.006)	(0.007)						
B_{18m}	0.949	0.990	0.974	0.976	B_{3m}	B_{3m}				
	(0.010)	(0.004)	(0.008)	(0.006)						
B_{24m}	0.952	0.990	0.945	0.987	B_{3m}	B_{3m}				
	(0.009)	(0.008)	(0.022)	(0.008)						
B_{36m}	0.956	0.999	0.936	0.974	B_{3m}	B_{3m}				
	(0.008)	(0.007)	(0.015)	(0.016)						
B_{60m}	0.950	1.000	0.947	0.981	B_{3m}	B_{3m}				
	(0.008)	(0.022)	(0.013)	(0.010)						
B_{84m}	0.955	0.985	0.936	0.969	B_{3m}	B_{3m}				
	(0.008)	(0.044)	(0.018)	(0.016)						
B_{108m}	0.949	1.000	0.915	0.956	B_{3m}	B_{3m}				
	(0.013)	(0.012)	(0.024)	(0.026)						
B_{120m}	0.950	0.991	0.924	0.974	B_{3m}	B_{3m}				
	(0.013)	(0.024)	(0.026)	(0.017)						
B_{OD}	-0.014	-0.001	-0.006	0.000						
	(0.004)	(0.003)	(0.003)	(0.001)						

Table D.2

Parameter estimates, Poland (continued)

	'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'
ρ	0.897 (0.011)	0.845 (0.017)								
σ^2_{3m}	1.000	1.000	1.000	1.000	3.303 (0.275)	2.730 (0.226)	0.013 (0.001)	0.004 (0.000)	0.011 (0.001)	0.008 (0.000)
σ^2_{6m}	1.000	1.000	1.000	1.000	2.359 (0.157)	2.145 (0.148)	0.010 (0.001)	σ^2_{3m}	0.014 (0.001)	σ^2_{3m}
σ^2_{9m}	1.000	1.000	1.000	1.000	1.115 (0.072)	0.938 (0.078)	0.009 (0.001)	σ^2_{3m}	0.019 (0.002)	σ^2_{3m}
σ^2_{12m}	1.000	1.000	1.000	1.000	1.000	1.000	0.008 (0.001)	σ^2_{3m}	0.016 (0.001)	σ^2_{3m}
σ^2_{18m}	1.000	1.000	1.000	1.000	0.633 (0.018)	0.639 (0.020)	0.004 (0.000)	σ^2_{3m}	0.006 (0.000)	σ^2_{3m}
σ^2_{24m}	1.000	1.000	1.000	1.000	0.384 (0.021)	0.351 (0.021)	0.002 (0.000)	σ^2_{3m}	0.002 (0.000)	σ^2_{3m}
σ^2_{36m}	1.000	1.000	1.000	1.000	0.468 (0.026)	0.422 (0.026)	0.003 (0.000)	σ^2_{3m}	0.003 (0.000)	σ^2_{3m}
σ^2_{60m}	1.000	1.000	1.000	1.000	0.694 (0.039)	0.612 (0.037)	0.004 (0.000)	σ^2_{3m}	0.005 (0.000)	σ^2_{3m}
σ^2_{84m}	1.000	1.000	1.000	1.000	0.547 (0.030)	0.483 (0.030)	0.003 (0.000)	σ^2_{3m}	0.003 (0.000)	σ^2_{3m}
σ^2_{108m}	1.000	1.000	1.000	1.000	0.516 (0.030)	0.457 (0.029)	0.003 (0.000)	σ^2_{3m}	0.003 (0.000)	σ^2_{3m}
σ^2_{120m}	1.000	1.000	1.000	1.000	0.617 (0.036)	0.536 (0.034)	0.003 (0.000)	σ^2_{3m}	0.004 (0.000)	σ^2_{3m}

www.nbp.pl