Score-Driven Models: Methods and Applications*

Mariia Artemova[1,2], Francisco Blasques[1,2],
Janneke van Brummelen[1,2], Siem Jan Koopman[1,2,3]

[1]*School of Business and Economics, Vrije Universiteit Amsterdam*

[2]*Tinbergen Institute Amsterdam*

[3]*CREATES, Aarhus University*

January 2022

Summary

The flexibility, generality and feasibility of score-driven models have contributed much to the impact of score-driven models in both research and policy. Score-driven models provide a unified framework for modeling time variation in parametric models. Unlike many other parametric models, given predictive likelihood score-driven models provide a flexible and intuitive way of modeling the dynamics while keeping estimation procedure and inference relatively simple. The developments in the theory and methodology of score-driven models made this class of models even more appealing. This led to new formulations of empirical dynamic models which are of relevance in economics and finance. In the context of macroeconomic studies, the key examples are nonlinear autoregressive, dynamic factor, dynamic spatial, and Markov switching models. In the context of finance studies, the major examples are models for integer-valued time series, multivariate scale models, and dynamic copula models. In finance applications, score-driven models are especially important since they provide updating mechanisms for time-varying parameters that limit the effect of the influential observations and outliers that are often present in financial time series.

Keywords: Score-driven time-varying parameters, Nonlinear autoregressive model, Integer-valued time series, Dynamic Markov switching model, Dynamic factor model, Dynamic spatial model, Multivariate volatility, Dynamic copula model.

Subjects: Dynamic Econometrics, Financial Econometrics, Economics, Finance

*This is a draft of an article that has been published by Oxford University Press in the Oxford Research Encyclopedia of Economics and Finance on 15 August 2022.

Score-Driven Models: Why, Where, and When to Use

Score-driven models have been introduced by Creal et al. (2013) and Harvey (2013) where a unified framework for modeling time-varying parameters is provided within an observation-driven approach. A review and a treatment of the methodological and theoretical aspects of score-driven models are provided by Artemova et al. (2022). The key principle of using the score function of the predictive likelihood as the driving mechanism for time-varying parameters in a model can be exploited for many different and important purposes in empirical studies related to the fields of economics and finance.

An extensive and up-to-date list of score-driven models proposed in the literature can be found online at `http://www.gasmodel.com/`. It includes contributions on linear and nonlinear regression models with time-varying coefficients for location and scale, but also dynamic models with time-varying higher order moments, and dynamic mixture distributions models. During the 2008–2022 period, many researchers have contributed to this field of research. This chapter aims to present an overview of these developments together with illustrations which are relevant for empirical studies in economics and finance.

The score-driven model belongs to the class of observation-driven models as defined by Cox (1981). The dynamic features of an observation-driven model rely on time-varying parameters which are defined as functions of past observations. Hence, the time-varying parameter is perfectly one-step-ahead predictable given the past information. It also implies that the likelihood function of the model is available in closed form, whether or not the model is non-Gaussian, is based on nonlinear relations, and/or relies on complex updating equations. As a result, the computational cost of parameter estimation using the method of maximum likelihood is relatively low when compared to the class of parameter-driven models. In the parameter-driven class, the time-varying parameter is a dynamic function of stochastic error terms which need to be integrated out from the joint density of observations and dynamic parameters to obtain the likelihood function. This integration task requires typically intensive computational simulation-based methods when the model departs from linear and Gaussian structures; see Gouriéroux and Monfort (1997) and Durbin and Koopman (2000). It is an important reason why observation-driven models, rather than parameter-driven models, are increasingly used in applied econometric studies where the time-variation of parameters need to be considered. The sub-class of score-driven models in the family of observation-driven models makes their use even more convenient as it offers sound updating mechanisms for time-varying parameters while from the outset it is not always clear how to choose or design an updating function. The score function provides a practical solution to this problem while it also has convenient optimality properties as argued in Blasques et al. (2015).

Time-varying parameters are commonly associated with the mean (location) or the variance (scale) of the observation density. However, score-driven models are not limited to their handling of time-varying regression coefficients or time-varying conditional volatility. It is also recognized in the literature that other features of the conditional observation density can be subject to time-variation. For example, Diebold et al. (1994) and Filardo (1994) emphasize that for Markov switching models it can be restrictive to assume a constant transition probability matrix. Also, Patton (2006) argues for the importance of the copula parameter to be time-varying. In the model specifications discussed in these contributions, the time-varying parameters are modeled as some function of the exogenous variables and/or lagged dependent variable. The adopted model specification is usually motivated to obtain an economic interpretation. However, in many situations the choice of updating equation, and its functional form, are not immediately obvious. In such circumstances, the score-driven approach is appealing as it provides a flexible and intuitive way of modeling time-varying parameters where the updating equation exploits information from the whole density and is not based on a particular choice from the set of first and higher order moments.

In general, the specification of the time-varying parameter update depends on the context of the model but also on the empirical research question. The main purpose of the paper is to show that the score-driven modeling approach can be easily incorporated into a range of models which play an important role in empirical applications. However, the review is by no means exhaustive. In the remainder of the Chapter, first, the general statistical formulation of the score-driven model is introduced. Next, reviews of three groups of score-driven models are provided: (i) univariate models, (ii) multivariate models, and (iii) models relevant for financial econometrics. In these three reviews, the focus is on the relevant features and aspects of the score-driven modeling approach.

General Framework of Score-Driven Models

We present a general framework for the modeling of time-varying parameters based on the score function of the predictive likelihood function which is used as a driver for the updating of the time-varying parameters. The score function can be with respect to the parameter directly or to a monotonic transformation of the parameter through a link function. The score-driven modeling approach originates from Creal et al. (2013) and Harvey (2013). A theoretical and methodological review of score-driven models can be found in Artemova et al. (2022).

The time-varying parameter is denoted by $\boldsymbol{f}_t$ and generally represents a $p \times 1$ vector. The time-varying parameter can also be a scalar then $p = 1$ and $\boldsymbol{f}_t \equiv f_t$. The updating equation for $\boldsymbol{f}_t$ is given by

$$\boldsymbol{f}_{t+1} = \boldsymbol{\omega} + \boldsymbol{A}\boldsymbol{s}_t + \boldsymbol{B}\boldsymbol{f}_t, \qquad \boldsymbol{s}_t = \boldsymbol{S}_t \nabla_t, \qquad \nabla_t = \frac{\partial}{\partial \boldsymbol{f}_t} \log p(\mathbf{y}_t | \boldsymbol{f}_t, \mathcal{F}_{t-1}; \boldsymbol{\psi}), \tag{1}$$

where $\mathcal{F}_{t-1}$ is the information set available up to time $t-1$, $\boldsymbol{\omega}$ is a $p \times 1$ vector of constants, $\boldsymbol{A}$ and $\boldsymbol{B}$ are $p \times p$ coefficient matrices, $\boldsymbol{s}_t$ is the scaled score of the predictive logdensity $\log p(\mathbf{y}_t | \boldsymbol{f}_t, \mathcal{F}_{t-1}; \boldsymbol{\psi})$ with respect to vector $\boldsymbol{f}_t$, $\boldsymbol{S}_t$ is the $p \times p$ scaling matrix, and $\boldsymbol{\psi}$ is a vector of other static parameters. In practical terms, the updating equation uses the score function of the predictive likelihood contribution at time t as its driving mechanism. The time-varying parameter update simply uses a *local* score measure as the innovation term. From the perspective of optimization, the parameter $\boldsymbol{f}_{t+1}$ is updated in the direction of the "steepest" increase of the predictive likelihood at time t. The specification of the linear updating function can be extended by including lags of $\boldsymbol{s}_t$ and $\boldsymbol{f}_t$ in the updating equation (1). This will also lead to more coefficient matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.

The typical specification for the scaling matrix $\boldsymbol{S}_t$ is given by

$$\boldsymbol{S}_t = \mathcal{I}_{t|t-1}^{-\zeta}, \qquad \zeta = \left\{0,\ \frac{1}{2},\ 1\right\},$$

where $\mathcal{I}_{t|t-1}$ is a Fisher information matrix and power ζ is a scalar. In case $\zeta \neq 0$, the score is corrected for the local curvature of the predictive density and it ensures that $\boldsymbol{s}_t$ has a finite covariance matrix. However, in many applications of empirical relevance, the Fisher information matrix is analytically intractable. In such cases, it is advantageous to use the identity matrix as the scaling matrix, that is $\zeta = 0$, or to rely on numerical integration methods for obtaining an estimate of $\mathcal{I}_{t|t-1}$. An alternative to overcome numerical challenges is to obtain an estimate of $\mathcal{I}_{t|t-1}$ on the basis of a feasible (approximate) information matrix; see Zhang et al. (2011) and Patton et al. (2019)). The choice of the scaling matrix will be discussed case by case for the presented models.

Univariate Score-Driven Models

The discussion starts with score-driven models in application to univariate time series. First, nonlinear autoregressive models and models for integer-valued time series are considered. This is a natural starting point given that these models are closely related to time-varying location models discussed in Artemova et al. (2022). Next, score-driven dynamic Markov switching and mixture models are presented. The section ends with an empirical illustration of the score-driven dynamic Markov switching model.

Nonlinear Autoregressive Models

Linear regression is probably the most widely applied statistical technique. Despite its simplicity, this method can offer a surprisingly accurate fit and insight in multiple statistical and econometric problems. This is the case also in time series analysis, where linear autoregressive (AR) models can offer an exceptionally simple yet accurate way of modeling dynamics in the data. However, in many applications it is important to account for time variation in regression coefficients which may occur for multiple reasons, ranging from nonlinear relationships to structural changes in the data generating process.

Consider the following linear AR(1) model with a time-varying autoregressive parameter,

$$y_t = c + \phi_t y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim p_\varepsilon(\varepsilon_t, \boldsymbol{\psi}),$$

where $\{y_t\}_{t=1}^T$ is the observed data, c is an intercept parameter, ϕ_t is a time-varying autoregressive parameter, and ε_t is a mean-zero disturbance that is distributed with density $p_\varepsilon(\varepsilon_t, \boldsymbol{\psi})$ which is parametrized by the vector $\boldsymbol{\psi}$. For simplicity, demeaned series y_t are considered $(c = 0)$ which can be extended straightforwardly to allow for a non-zero mean.

The score-driven modeling framework provides an intuitive way for modeling the time-varying parameter ϕ_t. Assume that ε_t is Gaussian with variance σ^2 then the score-driven updating equation for $\phi_t \equiv f_t$ is as follows:

$$f_{t+1} = \omega + \alpha \left(y_t - f_t y_{t-1}\right) \frac{y_{t-1}}{\sigma^2} + \beta f_t,$$

where the scaling $S_t = 1$ is used. The score expression results in an update for f_t which is governed by the term $(y_t - f_t y_{t-1}) y_{t-1}$, with the magnitude of the adjustments being determined both by α and

$1/\sigma^2$. The term $(y_t - f_t y_{t-1})y_{t-1}$ can be seen as an estimate of the covariance between y_{t-1} and the error term $(y_t - f_t y_{t-1})$, based on a single observation. The time-varying parameter f_t is thus adjusted to ensure that the error term is orthogonal to y_{t-1}.

For time series that contain influential incidental points and outliers, it can be beneficial to model the innovations by the Student's t distribution, $\varepsilon_t \sim t_\lambda(0, \sigma^2)$. The score-driven modeling framework automatically ensures that the impact of these observations on the update of f_t is bounded since the score function takes the following form,

$$\nabla_t = \frac{\lambda + 1}{\lambda \sigma^2} w_t (y_t - f_t y_{t-1}) y_{t-1},$$
$$w_t = \left(1 + \frac{(y_t - f_t y_{t-1})^2}{\lambda \sigma^2}\right)^{-1},$$

where the magnitude of the adjustment also depends on the weight w_t so that the effect of the score on the update is limited. The weight w_t plays a similar role as in location score-driven models with Student's t innovations discussed in Artemova et al. (2022).

In the discussion, the dynamics of ϕ_t are modeled "directly". Alternatively, the dynamics of ϕ_t can be modeled using a link function, $\phi_t = h(f_t)$, as proposed by Blasques et al. (2020). For example, in practice it is often desirable to rule out negative temporal dependence. For this, the exponential link function $h(f_t) = \exp(f_t)$ can be used. To rule out both negative temporal dependence and explosive behaviour while allowing for the near unit-root dynamics in y_t, the logistic function $h(f_t) = [1 + \exp(-f_t)]^{-1} \in (0, 1)$ can be considered. The use of the transformation function should be properly accounted in the updating equation, see Blasques et al. (2020) for the details. In addition, extra flexibility can be achieved by allowing time variation in the conditional volatility of the innovations, see Delle Monache and Petrella (2017). Modeling interaction between time-varying autoregressive coefficient and volatility can be of great importance in macroeconomic applications.

Models for Integer-Valued Time Series

Discrete-valued time series are often encountered in empirical applications. For instance, they can play important role when considering crime and transaction data. Observation-driven models have been successfully applied in these situations. For example, Fokianos et al. (2009) and Zhu (2011) model time series of counts using conditional Poisson and negative binomial (NB) distributions, respectively, and the dynamics of the conditional mean parameter is modeled using the lags of the dependent variable. This approach can be further generalized by means of the score-driven modeling methodology.

Specifically, Blasques et al. (2018) and Harvey and Kattuman (2020) model transaction data as well as data on coronavirus deaths using a score-driven model with a negative binomial distribution,

$$y_t|\mathcal{F}_{t-1} \sim \mathcal{NB}(\lambda_t, r),$$

where $\{y_t\}_{t=1}^T$ is a time series, $r > 0$ is the dispersion parameter, and λ_t is a time-varying location parameter. The updating equation for the reparameterized time varying parameter f_t, where $\lambda_t = h(f_t)$, directly follows from the score of the predictive likelihood, which is given by,

$$\nabla_t = \lambda_t^{-1}(y_t - \lambda_t)(r\lambda_t + 1)^{-1}\frac{\partial h(f_t)}{\partial f_t}, \tag{2}$$

where the link function $h(f_t)$ is used to ensure that $\lambda_t > 0$. Multiple link functions $h(\cdot)$ can be employed, such as an exponential function $\lambda_t = \exp(f_t)$. The score scaled by the inverse Fisher information matrix is then given by,

$$s_t = \frac{y_t - \exp(f_t)}{\exp(f_t)}, \quad \text{since} \quad \mathcal{I}_{t|t-1} = \lambda_t(r\lambda_t + 1)^{-1}.$$

Score-driven model can also accommodate an updating equation which can account for the presence of many zeros in the data. Blasques et al. (2018) consider a zero-inflated negative binomial distribution which properly treats zero values without any information loss. Particularly,

$$\begin{aligned} &y_t \sim 0 \quad \text{with probability } \pi, \\ &y_t|\mathcal{F}_{t-1} \sim \mathcal{NB}(\lambda_t, r) \quad \text{with probability } 1-\pi, \end{aligned}$$

where $\pi \in [0, 1)$ denotes the probability of zero values. The score update naturally incorporates information from the density (a mixture of densities in this case) into the update for λ_t. When $y_t \neq 0$ the score expression is the same as in (2). However, zero observations are treated differently:

$$\nabla_t = \frac{\pi - 1}{(r\lambda_t + 1)(1 + \pi(r\lambda_t + 1)^{1/r} - \pi)}\frac{\partial h(f_t)}{\partial f_t} \quad \text{for } y_t = 0,$$

meaning that the score expression also depends on the probability π of a zero value. When $\pi = 1$ it implies that $\nabla_t = 0$, hence the score contribution for the update is equal to 0. In turn, when $\pi = 0$ the score simplifies to (2) with $y_t = 0$.

For count data with extreme observations, Gorgi (2020) introduced a new general class of observation-driven models. The idea is to model the data based on a heavy-tailed mixture of negative binomial distributions, known as the beta negative binomial (BNB) distribution,

$$y_t|\mathcal{F}_{t-1} \sim \mathcal{BNB}(\lambda_t, r, \phi),$$

where $r > 0$ is the dispersion parameter and $\phi > 1$ is the tail parameter. The dynamics of λ_t are in turn modeled using an observation-driven updating equation. Specifically, the parameter can be updated based on the score of the predictive likelihood. The score-driven update ensures that the dynamics for λ_t is *itself* robust to outliers in y_t since it exploits relevant information from the BNB density. The BNB score-driven model is general enough since in the limit, when the tail parameter $\phi \to \infty$ and the dispersion parameter $r \to \infty$, the model approximates the Poisson autoregressive model and the score-driven-INAR model with NB distribution, introduced in Gorgi (2018).

Score-Driven Dynamic Markov Switching and Mixture Models

Markov switching models, introduced by Hamilton (1989), have proven to be useful in capturing changes in the level, volatility, or dependence structure of time series. In economics and finance, changes may occur due to policy changes or other influential and disruptive events. In particular, this class of models allows for switching between regimes according to transition probabilities that are estimated from the data. Typically, such transition probabilities are, however, assumed to be static. This can be restrictive in many applications, as highlighted by Diebold et al. (1994), Filardo (1994) and Haas et al. (2013). To relax these assumptions Bazzi et al. (2017) and Catania (2019) introduced dynamic Markov switching model and dynamic adaptive mixture model, respectively. In these models, time-varying parameters are modeled using a score-driven framework which provides an intuitive functional form for the time-varying parameter update. In the model of Catania (2019), the mixture components can be any parametric distribution with the possibility of allowing parameters of the distributions to be time-varying. Parameters of the mixture component and the weights are modeled using the score-driven framework. In turn, Bazzi et al. (2017) introduced a dynamic Markov regime switching model where the transition probability from one state to another changes over time.

Further discussion is based on the model of Bazzi et al. (2017). For simplicity, the focus is on the case with two regimes/states which can easily be generalized to a case with more states. The idea is to model the dynamics of the transition matrix Π_t,

$$\Pi_t = \begin{bmatrix} \pi_{00,t} & \pi_{01,t} \\ \pi_{10,t} & \pi_{11,t} \end{bmatrix} = \begin{bmatrix} \pi_{00,t} & 1 - \pi_{00,t} \\ 1 - \pi_{11,t} & \pi_{11,t} \end{bmatrix},$$

where $0 < \pi_{00,t}, \pi_{11,t} < 1$ denote the probabilities of staying at the same state. The transition probabilities are parameterized using a logistic link function to ensure that they lie in the $(0, 1)$ interval,

$$\pi_{ii,t} = \frac{\exp(-f_{i,t})}{1 + \exp(-f_{i,t})}, \quad i = 0, 1,$$

and then $\boldsymbol{f}_t = (f_{0,t}, f_{1,t})'$ is modeled using the score of the predictive likelihood and is updated according to equation (1). The two other probabilities $\pi_{01,t}$ and $\pi_{10,t}$ for the state switches follow straightforwardly since probability of switching from state 0 to 1 is equal to $1 - \pi_{00,t}$ and from state 1 to 0 is equal to $1 - \pi_{11,t}$.

To obtain the score, first the expression for the predictive likelihood needs to be derived. The conditional density of y_t is essentially a mixture of two distributions,

$$p(y_t|\mathcal{F}_{t-1};\boldsymbol{\psi}) = p(y_t|\theta_0,\mathcal{F}_{t-1};\boldsymbol{\psi})P(z_t = 0|\mathcal{F}_{t-1};\boldsymbol{\psi}) + p(y_t|\theta_1,\mathcal{F}_{t-1};\boldsymbol{\psi})\left(1 - P(z_t = 0|\mathcal{F}_{t-1};\boldsymbol{\psi})\right),$$

where y_t is a time series, z_t is an unobserved discrete Markov chain stochastic process, θ_i are regime-specific parameters ($i = 0, 1$) and $\boldsymbol{\psi}$ contains static parameters.

Naturally, the predictive probability $P(z_t = 0|\mathcal{F}_{t-1};\boldsymbol{\psi})$ can be obtained using a Hamilton filter,

$$P(z_t = 0|\mathcal{F}_{t-1};\boldsymbol{\psi}) = \pi_{00,t}P(z_{t-1} = 0|\mathcal{F}_{t-1};\boldsymbol{\psi}) + (1 - \pi_{11,t})P(z_{t-1} = 1|\mathcal{F}_{t-1};\boldsymbol{\psi}),$$

where $\pi_{00,t}$ and $\pi_{11,t}$ are functions of the time-varying parameters $f_{0,t}$ and $f_{1,t}$, respectively. The filtered probability $P(z_{t-1} = 0|\boldsymbol{\psi},\mathcal{F}_{t-1})$ can be obtained using law of conditional probability and noticing that $\mathcal{F}_{t-1} = \mathcal{F}_{t-2} \cup y_{t-1}$.

Then the score vector ∇_t of the predictive loglikelihood is given by,

$$\begin{aligned}\nabla_t &= \frac{p(y_t|\theta_0, \mathcal{F}_{t-1}; \boldsymbol{\psi}) - p(y_t|\theta_1, \mathcal{F}_{t-1}; \boldsymbol{\psi})}{p(y_t|\mathcal{F}_{t-1}; \boldsymbol{\psi})} \times g(\boldsymbol{f}_t, \boldsymbol{\psi}, \mathcal{F}_{t-1}),\\ g(\boldsymbol{f}_t, \boldsymbol{\psi}, \mathcal{F}_{t-1}) &= \begin{pmatrix} P(z_{t-1} = 0|\mathcal{F}_{t-1}; \boldsymbol{\psi})\pi_{00,t}(1 - \pi_{00,t}) \\ -P(z_{t-1} = 1|\mathcal{F}_{t-1}; \boldsymbol{\psi})\pi_{11,t}(1 - \pi_{11,t}) \end{pmatrix}.\end{aligned} \tag{3}$$

The score expression has a very intuitive interpretation. If the scaled difference between $p(y_t|\theta_0, \mathcal{F}_{t-1}; \boldsymbol{\psi})$ and $p(y_t|\theta_1, \mathcal{F}_{t-1}; \boldsymbol{\psi})$ is positive, then in the next period one should expect an increase in $f_{0,t}$ and a decrease in $f_{1,t}$. In other words, the probability $\pi_{00,t}$ is updated positively if it is more likely that in the previous period observations were sampled from $p(y_t|\theta_0, \mathcal{F}_{t-1}; \boldsymbol{\psi})$ rather than from $p(y_t|\theta_1, \mathcal{F}_{t-1}; \boldsymbol{\psi})$, and vice versa. In turn, the magnitude of the updates is controlled by the term $g(\boldsymbol{f}_t, \boldsymbol{\psi}, \mathcal{F}_{t-1})$. To be more precise, it is controlled by the filtered conditional probabilities of being in regime 0 or 1. This means that if state 0 was not visited by the chain at time $t-1$, that is, $P(z_{t-1} = 0|\mathcal{F}_{t-1}; \boldsymbol{\psi}) = 0$, there is no relevant information for $\pi_{00,t}$ update, while a large step is taken to update $\pi_{11,t}$. Hence, if there is certainty about being in regime $z_{t-1} = 1$, then the filter can "learn" from the data about $\pi_{11,t}$ but not much about $\pi_{00,t}$.

Interestingly, the score expression in score-driven Markov switching models is very similar to the one obtained in the mixture model of Catania (2019). Specifically, for the case with two regimes, Harvey and Palumbo (2021) show that the score expression is of the following form,

$$\nabla_t = \frac{p(y_t|\theta_0, \mathcal{F}_{t-1}; \boldsymbol{\psi}) - p(y_t|\theta_1, \mathcal{F}_{t-1}; \boldsymbol{\psi})}{p(y_t|\mathcal{F}_{t-1}; \boldsymbol{\psi})} \times \pi_t(1 - \pi_t), \tag{4}$$

where in the case of the mixture model π_t is the probability of being in state one at time t. The score expressions (3) and (4) are almost the same except that in Markov switching model the filtered probabilities $P(z_{t-1} = i|\mathcal{F}_{t-1}; \boldsymbol{\psi})$ also enter the score expression. This is not surprising since, in contrast to the mixture model, in the Markov switching model the latent process is assumed to be a Markov chain. Apart from this, these two score-driven models are closely related to each other.

Illustration: Unemployment Rate

To demonstrate the flexibility of the score-driven modeling approach, an empirical illustration in application to the changes in the US unemployment rate during the period 1960-2019 is considered[1].

To illustrate the advantages of the score-driven model, the comparison is made between the classic static Markov switching model and the score-driven dynamic Markov switching model. For both models, a model specification with four states was found to capture the dynamics of the series well. Particularly, two regimes for the time-varying mean (low $\mu_{1,t}$ and high $\mu_{2,t}$ mean regimes) and two regimes for the volatility (low σ_1^2 and high σ_2^2 volatility regimes) are considered. The mean in regime i is modeled as an autoregressive process with regime-specific coefficients: $\mu_{i,t} = \phi_{0i} + \phi_{1i}y_{t-1} + \phi_{2i}y_{t-2}$. For the volatility, the transition probabilities Π_t^σ are allowed to be time-varying and are modeled using score-driven dynamics, whereas in the classic Markov switching model the transition probabilities Π^σ are assumed to be static. Hence, the transition matrix is of the following form:

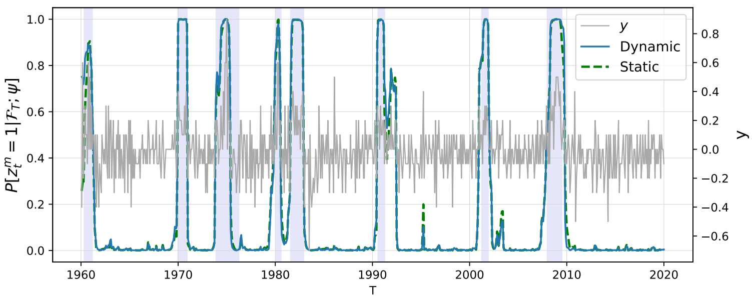
$$\Pi_t = \Pi^\mu \otimes \Pi_t^\sigma, \quad \Pi^\mu = \begin{bmatrix} \pi_{11}^\mu & (1-\pi_{11}^\mu) \\ (1-\pi_{22}^\mu) & \pi_{22}^\mu \end{bmatrix}, \quad \Pi_t^\sigma = \begin{bmatrix} \pi_{11,t}^\sigma & (1-\pi_{11,t}^\sigma) \\ (1-\pi_{22,t}^\sigma) & \pi_{22,t}^\sigma \end{bmatrix},$$

The results are demonstrated in Figure 1. The smoothed probabilities of being in the high mean regime are almost the same for the two models. Indeed, both models capture well the increases in the unemployment rate, which coincide with the periods of the US recessions. The similarity of the results is not surprising since the transition probabilities for the mean are not time-varying in either model. In contrast, the results for the smoothed high variance regime are different between the two models. Specifically, around the period 1975-1990 there were several consecutive recessions and there are several spikes in the squared time series. The score-driven dynamic Markov switching model is able to distinguish between these four recession episodes, while the static Markov switching model does not adapt. This difference occurs due to the fact that in the score-driven dynamic model the filtered high variance probability decreases during the period 1975-1990, meaning that probability of staying in the high volatility regime becomes lower. In contrast, in the static model, the probability cannot change over time. Similar pattern appears around 2010. Therefore, the dynamic score-driven Markov switching model provides an extra layer of flexibility in capturing regimes, which is important in economic applications.

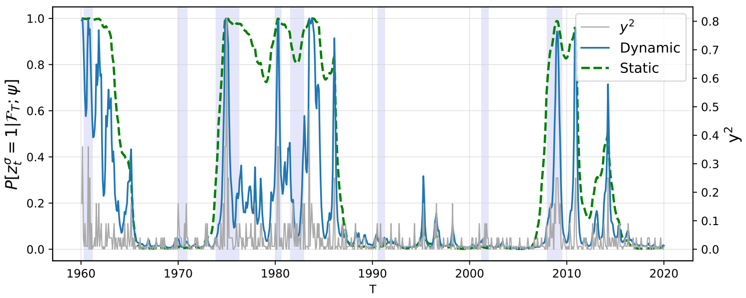
Multivariate Score-Driven Models

The functionality of the score-driven modeling approach is not limited to univariate time series and can also be considered in the multivariate time series applications. While the data is multivariate, the

(a) Smoothed High (Recession) Mean Regime



(b) Smoothed High Variance Regime



(c) Filtered High Variance probability

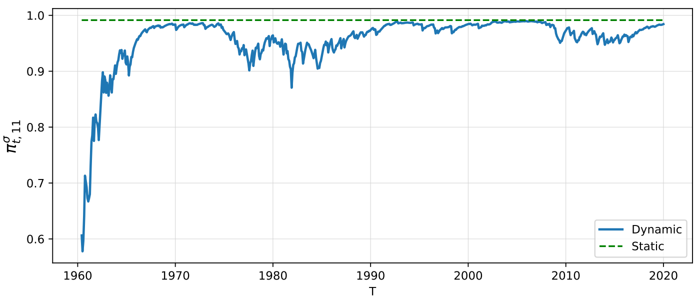


Figure 1: Empirical illustration: score-driven dynamic Markov switching model.
Shaded areas correspond to the US recessions according to the NBER.

time-varying parameter can still be a scalar. The models under consideration are popular in economic and finance-related applications.

Score-Driven Dynamic Factor Models

Dynamic factor models are widely used in econometrics as a convenient way to capture common dynamics and cross-sectional dependence across multiple time series. The use of common factors is especially beneficial when the cross-sectional dimension of the dataset is large. In the literature, factors are usually modeled using a parameter-driven approach as in Doz et al. (2012) where factors have their own source of uncertainty. However, when innovations are non-Gaussian or the factors' updating equation is non-linear, the likelihood is not available in a closed form. This results in a high computational burden, especially when the dimension of the model is large. Alternatively, one can estimate the factors using principal component analysis as in Stock and Watson (2002). However, since the factors are static, forecasts cannot be produced directly from the model and auxilliary regressions are required to make predictions. In contrast, the score-driven dynamic factor model, proposed in Creal et al. (2014), provides a unified framework for estimation and forecasting while keeping the estimation procedure relatively simple.

Assume that the dynamics of the time series $\{\mathbf{y}_t\}_{t=1}^T$ can be summarized by several common factors $\boldsymbol{f}_t$ which leads to the following observation equation:

$$\mathbf{y}_t = \boldsymbol{z} + \boldsymbol{Z}\boldsymbol{f}_t + \varepsilon_t, \quad \varepsilon_t \sim \boldsymbol{p}_{\boldsymbol{\varepsilon}}(\varepsilon_t; \boldsymbol{\Sigma}, \boldsymbol{\lambda}),$$

where $\mathbf{y}_t$ is an $N \times 1$ random vector, $\boldsymbol{f}_t$ is a $p \times 1$ vector of latent factors, $\boldsymbol{Z}$ is an $N \times p$ matrix of factor loadings, and $\boldsymbol{z}$ is an $N \times 1$ vector of intercepts. The innovations $\boldsymbol{\varepsilon}_t$ are drawn from a multivariate density $\boldsymbol{p}_{\boldsymbol{\varepsilon}}(\varepsilon_t; \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ with an $N \times N$ scale matrix $\boldsymbol{\Sigma}$, and parameterized by the vector $\boldsymbol{\lambda}$.

In case of Student's t innovations with λ degrees of freedom, the likelihood contribution at time t is as follows:

$$l_t = \log\Gamma\left(\frac{\lambda+N}{2}\right) - \log\Gamma\left(\frac{\lambda}{2}\right) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{N}{2}\log(\lambda\pi) - \frac{\lambda+N}{2}\log\left(1 + \frac{(\mathbf{y}_t - \boldsymbol{z} - \boldsymbol{Z}\boldsymbol{f}_t)'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \boldsymbol{z} - \boldsymbol{Z}\boldsymbol{f}_t)}{\lambda}\right).$$

Hence, the score expression takes the form,

$$\begin{aligned} \nabla_t &= \frac{\lambda+N}{\lambda} w_t \boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\left(\mathbf{y}_t - \boldsymbol{z} - \boldsymbol{Z}\boldsymbol{f}_t\right), \\ w_t &= \left(1 + \frac{(\mathbf{y}_t - \boldsymbol{z} - \boldsymbol{Z}\boldsymbol{f}_t)'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \boldsymbol{z} - \boldsymbol{Z}\boldsymbol{f}_t)}{\lambda}\right)^{-1}. \end{aligned} \tag{5}$$

An important feature of the score-driven factor model is that the presence of the time-varying weight w_t makes the dynamics of $\boldsymbol{f}_t$ more robust against outliers and influential observations. Specifically, "large" prediction errors $\boldsymbol{\varepsilon}_t$ lead to a small weight w_t, consequently, down-weighting large positive or negative observations in the update for $\boldsymbol{f}_t$. In turn, when $\lambda \to \infty$ the score simplifies to that of a

Gaussian model, and the factor is a linear function of the prediction error $\mathbf{y}_t - \boldsymbol{z} - \boldsymbol{Z}\boldsymbol{f}_t$. Specifically, when $p = 1$ and the scaling matrix $\boldsymbol{S}_t = \mathcal{I}_{t|t-1}^{-1}$ the updating equation simplifies to,

$$f_{t+1} = \omega + \alpha \left(\boldsymbol{Z}'\Sigma^{-1}\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}\Sigma^{-1} \left(\mathbf{y}_t - \boldsymbol{z} - \boldsymbol{Z}f_t\right) + \beta f_t,$$

where, intuitively, the update is based on the generalized least squares estimator. The equation can also be interpreted as a scaled location model.

In factor models, both the factors and the loadings are unobserved and identified up to a rotation. In score-driven factor models, the loadings also enter into the factors' updating equation which complicates the identification. Hence, to identify the model parameters several restrictions need to be imposed. First, the data is usually standardized meaning that all series have sample mean equal to 0 and sample variance equal to 1. This implies that one can set $\boldsymbol{z} = \mathbf{0}$ and $\boldsymbol{\omega} = \mathbf{0}$. Next, matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are assumed to be diagonal. As for the loadings, the first p rows of the matrix of factor loadings, $\boldsymbol{Z}$, are restricted to have a form of lower-diagonal matrix with 1s on the diagonal.

In the paper, Creal et al. (2014) provide even a more general and flexible setup by allowing observations to be coming from different distributions which are linked to each other by a small set of latent dynamic factors. They also consider a problem of missing observations which appear either regularly due to different data sampling frequency or irregularly. Intuitively, when observations are absent, their contribution to the likelihood is equal to 0 which also leads to a score contribution of zero. Particularly, when all cross-sectional observations are missing at time t this implies that $\boldsymbol{s}_t = \mathbf{0}_p$.

Score-driven factor models are very attractive when datasets are large. For this reason, many extensions have been developed in the literature. For example, when modeling the time-varying covariance/scale matrix of a multivariate time series the number of parameters grows fast once more series are added. To resolve this issue Creal et al. (2011) impose a factor structure on the time-varying volatilities and/or time-varying correlations of the covariance/scale matrix of the multivariate Student's t distribution. The factors are driven by the score which bounds the effect of the extreme observations and outliers. This is especially relevant in finance applications. An extension of this model was further considered in Liu (2016).

Factor models can also be used to model time-varying default probabilities. For instance, Babii et al. (2019) model default probabilities as a function of time-varying latent geographic and credit score factors which are updated based on the score. In the paper, they demonstrate the computational benefits of using a score-driven approach in big data applications. Moreover, factor-augmented regressions have proven to be successful in the forecasting literature. Gorgi et al. (2019b) and Blasques et al. (2021) consider factor-augmented regressions models with score-driven factors when observations are of mixed frequency and the factors' loadings are clustered, respectively. Finally, a score-driven factor structure was exploited in copula models; see for example Oh and Patton (2017), Oh and Patton (2018), and Opschoor et al. (2020). Since it is not computationally feasible to estimate copula models when the cross-sectional dimension is large, these models are of great importance in applied work.

Illustration: Global Commodity Price Indices

The empirical illustration presents an application of the score-driven dynamic factor model to the analysis of the co-movements between global commodity price indices. The dataset under consideration is similar to the one considered in Zhang and Broadstock (2020). It consists of seven major commodity classes: Energy, Beverage, Fertilizers, Food, Metal, Precious metals, and Raw materials. Similarly to Zhang and Broadstock (2020), monthly crude oil price data is used as a proxy for the energy sector and monthly commodity price indices from the World Bank database are used for the other six classes. The data spans the period from April 1983 untill February 2021[2.].

Zhang and Broadstock (2020) document rising connectedness between the commodity prices. This finding highlights the importance of studying the co-movements between the series. The score-driven factor model with Gaussian innovations can be used to extract a factor that is common to all the classes. The standardized series and the extracted common factor are presented in Figure 2. Clearly, the estimated common factor captures well the co-movements between the series. It reveals a big downturn during the 2008 recession and a somewhat smaller one during the COVID-19 recession. This indicates that major co-movements between the commodity prices arise during the moments of the economic downturns. Furthermore, all price indices seem to comove which is indicated by the positive loadings on the common factor. Food, Raw materials and Metal are exposed the most to the common factor as they have the largest loadings, while the Precious Metals are the least exposed.

Score-Driven Dynamic Spatial Models

In panel data applications, cross-sectional observations are often dependent. A spatial dependence structure may emerge, for example, due to geographical proximity. Spatial models take into account

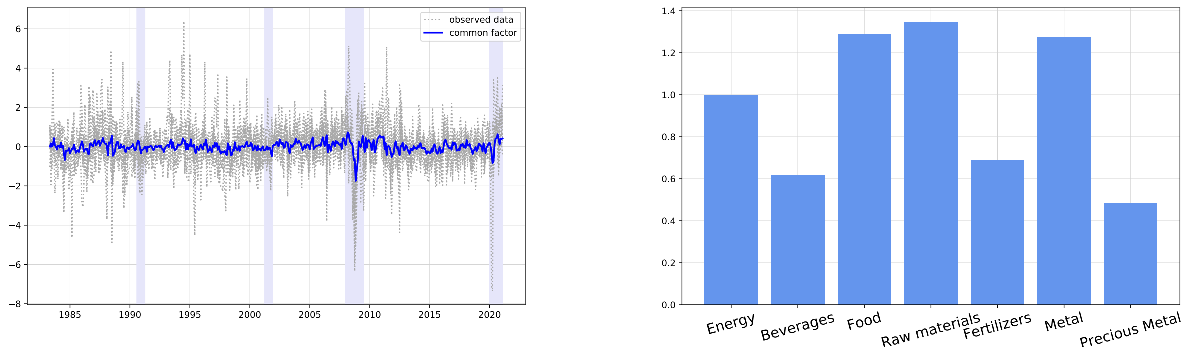


Figure 2: Empirical illustration: score-driven dynamic factor model.
Estimated common factor for seven major commodity classes (left) and estimated loadings on the common factor (right). Shaded areas in the left figure correspond to the US recessions according to the NBER.

the fact that "neighbouring" units might be more dependent and subject to spatial spillover effects. To model spatial dependencies, the Spatial Durbin Model (Anselin, 1988), and multiple well known extensions, are often considered. In these models, spatial dependencies are determined by a spatial weighting matrix which is calibrated or estimated from relevant data, and assumed to be fixed over time. This can, however, be restrictive in some applications.

Blasques et al. (2016) generalize the spatial Durbin model by allowing the spatial dependence parameter ρ to be time-varying. In particular, Blasques et al. (2016) consider a spatial model given by,

$$\mathbf{y}_t = \boldsymbol{W}_t \mathbf{y}_t + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \boldsymbol{p}_{\varepsilon}(\boldsymbol{\varepsilon}_t; \boldsymbol{\Sigma}, \boldsymbol{\lambda}), \quad t = 1, ..., T,$$

where $\boldsymbol{W}_t \equiv \rho_t \boldsymbol{W}$, $\mathbf{y}_t$ is an $N \times 1$ vector that consists of cross-sectional units that are subject to spatial dependence, $\boldsymbol{W}$ is an $N \times N$ row-normalized matrix of exogenous spatial weights which contains information about the "distance" between the units, $\boldsymbol{X}_t \equiv (\boldsymbol{\iota}_N : \boldsymbol{K}_t : \boldsymbol{W}\boldsymbol{K}_t)$ consists of an $N \times 1$ vector of ones, $N \times k$ matrix of exogenous regressors $\boldsymbol{K}_t$ and their spatial lags $\boldsymbol{W}\boldsymbol{K}_t$, $\boldsymbol{\beta} \equiv (\beta_1, \boldsymbol{\beta}_2', \boldsymbol{\beta}_3')'$ contains unknown parameters that correspond to $\boldsymbol{X}_t$, and $\boldsymbol{\varepsilon}_t$ is an $N \times 1$ vector of innovations with multivariate density $\boldsymbol{p}_{\varepsilon}(\boldsymbol{\varepsilon}_t; \boldsymbol{\Sigma}, \boldsymbol{\lambda})$.

The idea is to model the dynamics of ρ_t using a score-driven model. Since the matrix $\boldsymbol{W}$ is row-normalized, the spatial dependence parameter ρ_t should be smaller than one in absolute terms, i.e. $\rho_t \in (-1, 1)$. Specifically, ρ_t can be modeled as a bounded monotonic transformation of a time varying parameter f_t where the dynamics of f_t are defined using the score-driven model (1).

To obtain the score updating equation, consider the time t contribution to the log-likelihood function,

$$l_t = \log \boldsymbol{p}_{\boldsymbol{\varepsilon}}(\mathbf{y}_t - h(f_t)\boldsymbol{W}\mathbf{y}_t - \boldsymbol{X}_t\boldsymbol{\beta}, \boldsymbol{\Sigma}; \boldsymbol{\lambda}) + \log |(\boldsymbol{I}_N - h(f_t)\boldsymbol{W})|,$$

where the last term captures the non-linearity of the model in $\rho_t = h(f_t)$ and $h(\cdot)$ is a monotonic transformation. A typical choice of the transformation function is the hyperbolic tangent function.

The expression for the score depends on the density $\boldsymbol{p}_{\boldsymbol{\varepsilon}}$ under consideration. Possible choices are Gaussian and Student's t distributions. Particularly, when the disturbance vector $\boldsymbol{\varepsilon}_t$ is multivariate Student's t distributed with scale matrix $\boldsymbol{\Sigma}$ and degrees of freedom parameter λ, the score expression is as follows:

$$\nabla_t = \left(\frac{\lambda + N}{\lambda}\right)\left(w_t \mathbf{y}_t' \boldsymbol{W}' \boldsymbol{\Sigma}^{-1}\left(\mathbf{y}_t - h(f_t)\boldsymbol{W}\mathbf{y}_t - \boldsymbol{X}_t\boldsymbol{\beta}\right) - \operatorname{tr}\left(\boldsymbol{Z}(f_t)\boldsymbol{W}\right)\right) \cdot \frac{\partial h(f_t)}{\partial f_t}, \tag{6}$$

$$w_t = \left(1 + \frac{1}{\lambda}(\mathbf{y}_t - h(f_t)\boldsymbol{W}\mathbf{y}_t - \boldsymbol{X}_t\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - h(f_t)\boldsymbol{W}\mathbf{y}_t - \boldsymbol{X}_t\boldsymbol{\beta})\right)^{-1}.$$

where $\boldsymbol{Z}(f_t) = (\boldsymbol{I}_N - h(f_t)\boldsymbol{W})$. Therefore, the time-varying parameter f_t is driven by past observations and it integrates both direct effects arising from changes in the explanatory variables and indirect effects arising from the neighboring units. Intuitively, the score expression (6) is similar to the score (5) in the factor model. However, there is an additional term $\operatorname{tr}((\boldsymbol{Z}(f_t)\boldsymbol{W})$ that accounts for the effects of feedback loops where unit i affects unit j and unit j also affects unit i. Similarly to the score-driven factor model with Student's t innovations, the presence of weight w_t bounds the effect of extreme observations on the update.

In practice, using this model one can account for time-variation in shock spillovers. In financial applications, the parameter ρ_t can also be interpreted as a systemic risk measure or the market's perception of contagion. For example, there is evidence that sovereign and bank credit spreads in Europe are strongly cross-sectionally dependent. Using dynamic spatial models applied to the Europe sovereign and bank credit default swap (CDS) spreads, Blasques et al. (2016) and Foglia and Angelini (2019) reveal that the intensity of the spatial dependence is time-varying and seems to respond to the business cycle and policy measures implemented by the European Central Bank during the sovereign debt crisis.

Furthermore, using a spatial score-driven model, Böhm et al. (2022) analyze the strength of the business cycle synchronization in Europe. They find that the dynamics of the spatial dependence parameter is cyclical and the business cycle synchronization in Europe is stronger during periods of financial distress. Moreover, to account for the global diffusion of shocks, Catania and Billé (2017) extended the dynamic spatial model by introducing, in addition to global spillovers, spatial dependence into the error term $\boldsymbol{\varepsilon}_t$. Therefore, a change in the disturbance of a single location can produce impacts on nearby disturbances.

Alternatively, one can allow for the spatial weight matrix $\boldsymbol{W}$ *itself* to be time-varying as in Billé et al. (2019). Intuitively, while the geographical distance between units might not be changing over time, the "spatial radius" within which neighbourhoods should be considered might be time-varying. The weights of the $N \times N$ symmetric spatial weight matrix, $\boldsymbol{W}(\gamma_t, \boldsymbol{d})$, at time t can be modeled as a monotonic decreasing transformation of a time-varying distance decay parameter γ_t and a static $N \times N$ distance matrix $\boldsymbol{d}$, i.e. $W_{ij,t} \equiv h(\gamma_t, d_{ij})$. The parameter γ_t then characterizes the importance of higher-order neighbours: the lower γ_t the more important higher-order neighbours are. Possible choices of transformation functions are $h(\gamma_t, d_{ij}) = d_{ij}^{-\gamma_t}$ and $h(\gamma_t, d_{ij}) = \exp(-\gamma_t d_{ij})$ which ensure that the larger the distance between the units the smaller the weight is.

This approach provides the "spatial radius" within which higher-order neighborhoods play an important role and hence defines the sparsity of the weight matrix $\boldsymbol{W}$. Specifically, when the parameter γ_t is low it means that higher order neighbours are important, which results in a non-sparse weight matrix and vice versa. Since the choice of the weight matrix is not trivial, this approach allows flexible specification of the weight matrix, and therefore it helps to avoid model misspecification due to an incorrectly specified $\boldsymbol{W}$.

Illustration: Credit Default Swap Spreads

The empirical illustration demonstrates the evolution of the systemic risk measure and the importance of modeling time variation for parameter ρ_t. The dataset under consideration is similar to the one used in Blasques et al. (2016) and it contains the sovereign CDS spreads data for six European countries: Belgium, France, Germany, Italy, Portugal, and Spain. It spans the period from 1 January 2009 untill 31 December 2020, meaning that it covers both the period during the sovereign debt crisis in Europe, as in Blasques et al. (2016), as well as the COVID-19 recession (Figure 3). The comparison is made between four models: the static spatial Durbin model with Gaussian and Student's t innovations, and the score-driven dynamic spatial Durbin model with Gaussian and Student's t innovations. From Figure 3, it is apparent that the sovereign CDS spreads have several extreme observations. Therefore, one would expect that the robust filter performs better.

The advantage of using a Student's t distribution for the innovations becomes immediately evident, since these models lead to substantially higher loglikelihood values and lower AICc values (Table 1). Incorporating dynamics into the spatial parameter leads to even further improvements. The estimated spatial parameter is demonstrated in Figure 3. Overall, the dynamics implied by the Student's t score-driven and Gaussian score-driven models are comparable. The results reveal a decrease in the spatial dependence after the European sovereign debt crisis, which is in line with the findings in Blasques et al. (2016). The major differences between the models implied spatial dependences occur during the COVID-19 pandemic. The Gaussian score-driven model reacts immediately with a substantial increase in the dependence and after that it takes some time for the spatial dependence parameter to decrease, while the Student's t score-driven model bounds the influence of the observations, which leads to a moderate increase in the spatial dependence.

	static Gaussian	static Student's t	score-driven Gaussian	score-driven Student's t
logL	-54236.62	-53820.17	-46882.65	-46608.12
AICc	108493.2	107664.3	93787.29	93242.25

Table 1: Comparison of the likelihood and AICc values for static and dynamic spatial models.

Score-Driven Models in Financial Econometrics

The score-driven modeling approach can be incorporated in multivariate scale models and in dynamic copula models. Such models are used heavily in methodological and empirical research related to financial econometrics. Furthermore, many financial time series are subject to heavy noise and "outlying" observations which may require robust methods for their analysis. It is shown that score-driven models are rather effective in the introduction of robustness features in models and methods which are used regularly in financial econometrics.

Score-Driven Multivariate Scale Models

It is often important to jointly model the conditional volatilities and conditional correlations of a group of assets. For example for effective risk management and asset pricing. A popular way to jointly model volatilities are multivariate GARCH models, see Bauwens et al. (2006) for an overview. Among this class of models, a highly successful model is the dynamic conditional correlation (DCC) model of Engle (2002) which models the variances separately from the correlations. An alternative

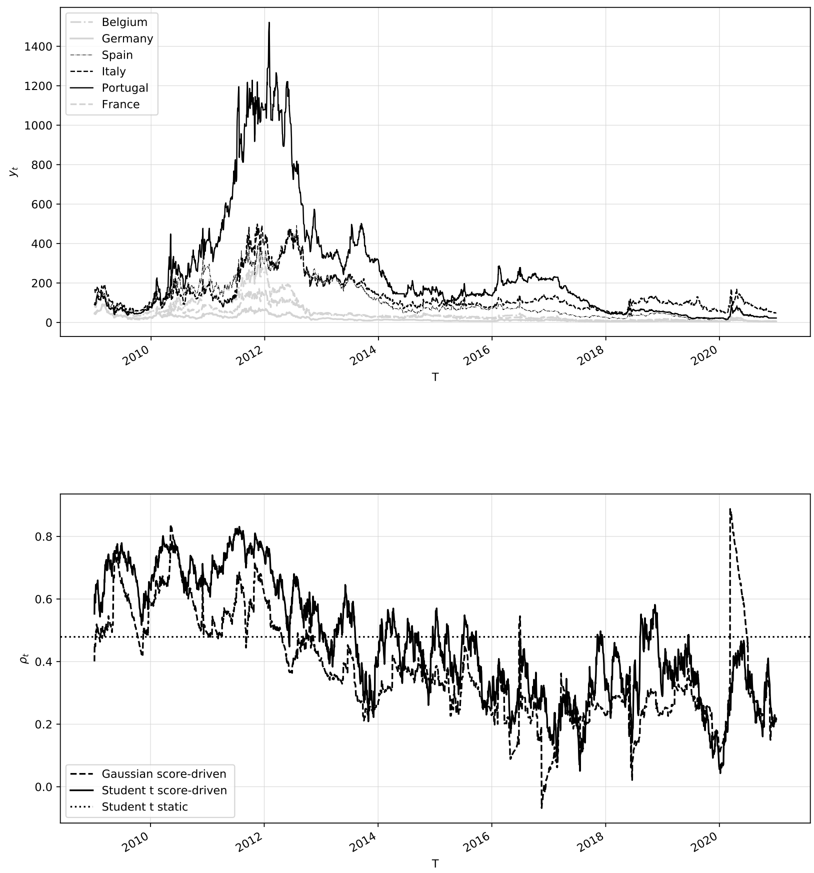


Figure 3: Empirical illustration: score-driven dynamic spatial Durbin model.
CDS spreads data for six countries (top) and estimated spatial parameter (bottom).

option is to use multivariate score-driven volatility models, which have the same advantages of the univariate score-driven scale models extensively discussed in Artemova et al. (2022).

Let $\mathbf{y}_t$ be an N-dimensional observation vector with mean zero and observation equation

$$\mathbf{y}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t$ is i.i.d. and $\boldsymbol{H}_t$ is the conditional covariance matrix of $\mathbf{y}_t$. $\boldsymbol{H}_t$ is driven by the time-varying parameter vector $\boldsymbol{f}_t$, i.e. $\boldsymbol{H}_t = \boldsymbol{H}(\boldsymbol{f}_t)$. In turn, the dynamics of $\boldsymbol{f}_t$ are modeled using equation (1).

There are numerous options for the functional link between the (multivariate) time-varying parameter $\boldsymbol{f}_t$ and the covariance matrix $\boldsymbol{H}_t = \boldsymbol{H}(\boldsymbol{f}_t)$. For instance, one can take $\boldsymbol{f}_t = \mathrm{vech}(\boldsymbol{H}_t)$, where vech vectorizes the lower-triangular part of a matrix. In many cases, however, it is preferred to disentangle the volatilities and the correlations, to be able to model the correlations explicitly. For example, Creal et al. (2011) consider the following parameterization of $\boldsymbol{H}_t$, which is also used for the DCC model:

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t \,, \tag{7}$$

where $\boldsymbol{D}_t$ is a diagonal matrix containing the standard deviations and $\boldsymbol{R}_t$ is a positive definite correlation matrix. Either $\boldsymbol{D}_t$ or $\boldsymbol{R}_t$, or both can be chosen to be time-varying. The correlation matrix $\boldsymbol{R}_t$ can be decomposed as follows: $\boldsymbol{R}_t = \boldsymbol{\Delta}_t^{-1} \boldsymbol{Q}_t \boldsymbol{\Delta}_t^{-1}$, where $\boldsymbol{Q}_t$ is symmetric positive definite and $\boldsymbol{\Delta}_t$ is diagonal with the square root of the diagonal elements of $\boldsymbol{Q}_t$ on the diagonal. The time-varying parameter $\boldsymbol{f}_t$ can then be specified as

$$\boldsymbol{f}_t = \begin{pmatrix} \mathrm{diag}(\boldsymbol{D}_t^2) \\ \mathrm{vech}(\boldsymbol{Q}_t) \end{pmatrix} . \tag{8}$$

If the innovations $\boldsymbol{\varepsilon}_t$ are multivariate standard normal, then the corresponding score is given by

$$\nabla_t = \frac{\partial \log p(\mathbf{y}_t | \boldsymbol{f}_t, \mathcal{F}_{t-1}; \boldsymbol{\psi})}{\partial \boldsymbol{f}_t} = \frac{1}{2} \Psi_t' \mathcal{D}_N' \boldsymbol{H}_t^{-1} \otimes \boldsymbol{H}_t^{-1} [w_t \mathrm{vec}(\mathbf{y}_t \mathbf{y}_t') - \mathrm{vec}(\boldsymbol{H}_t)] \,, \tag{9}$$

$$\Psi_t = \Psi(\boldsymbol{f}_t) = \frac{\partial \mathrm{vech}(\boldsymbol{H}_t)}{\partial \boldsymbol{f}_t'} \quad \text{for } \boldsymbol{H}_t = \boldsymbol{H}(\boldsymbol{f}_t) \,, \tag{10}$$

where $\mathcal{D}_N$ is the duplication marix, i.e. $\mathcal{D}_N \mathrm{vech}(\boldsymbol{A}) = \mathrm{vec}(\boldsymbol{A})$, $\otimes$ is the Kronecker product and $w_t = 1$. The choice of parameterization of $\boldsymbol{H}_t$ is fully accounted for by the matrix Ψ_t. The value of Ψ_t for the specific parameterization in (7) and (8) is provided by Creal et al. (2011, Equation (16)). Interestingly, according to (9), regardless of the precise parameterization, the driving mechanism of $\boldsymbol{f}_t$ depends on the deviations of the outer product $\mathbf{y}_t \mathbf{y}_t'$ from the covariance matrix $\boldsymbol{H}_t$.

Creal et al. (2011) instead consider a standardized multivariate Student's t with mean zero, variance $\boldsymbol{I}_N$ and $\nu > 2$ degrees of freedom, for which the score is also given by (9), but then for $w_t = (\nu + N)/(\nu - 2 + \mathbf{y}_t' \boldsymbol{H}_t^{-1} \mathbf{y}_t)$. It is also possible to use a regular Student's t distribution and allow for $\nu \leq 2$, but then $\boldsymbol{H}_t$ should be interpreted as a scale matrix instead of a covariance matrix. For the Student's t distribution, the updating of $\boldsymbol{f}_t$ depends on the deviations of the weighted outer product $w_t \mathrm{vec}(\mathbf{y}_t \mathbf{y}_t')$ from the covariance matrix $\mathrm{vec}(\boldsymbol{H}_t)$. The weight w_t decreases for large values of $\mathbf{y}_t' \boldsymbol{H}_t^{-1} \mathbf{y}_t$, so it downweights large observations, much like the univariate Student's t scale models discussed in Artemova et al. (2022). It is also possible to consider other multivariate distributions for the innovations, such as the multivariate generalized t-distribution.

If $\boldsymbol{H}_t$ is parametrized as in (7) and (8), parameter restrictions on the coefficient matrices are needed to ensure that the volatilities are positive and $\boldsymbol{Q}_t$ is positive definite. Alternatively, $\mathrm{diag}(\boldsymbol{D}_t^2)$ in (8) can be replaced by $\log(\mathrm{diag}(\boldsymbol{D}_t^2))$ to automatically ensure the positivity of the volatilities.

A disadvantage of the parameterization (7) and (8) is that the number of the time-varying factors that drive the correlation matrix $\boldsymbol{R}_t$ is greater than the number of free correlations in $\boldsymbol{R}_t$. This causes the information matrix to be singular, so instead of the inverse, a pseudo inverse has to be used to scale the score. An alternative way to parametrize the model, which avoids this problem, is to use hyperspherical coordinates to model the time-varying correlation matrix $\boldsymbol{R}_t$, as discussed by Creal et al. (2011). Another possibility to limit the number of factors is to impose a common dynamic factor structure on the volatilities and correlations, as discussed in section "Score-Driven Dynamic Factor Models". To further limit the number of parameters, the coefficient matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ can be restricted to be diagonal or scaled identity matrices, like for the DCC model.

Score-Driven Volatility Models with Realized Measures

Due to the increasing availability of high-frequency data over the years, there have been many developments in the literature concerning volatility models that include variables that take into account intraday data. More specifically, it is common to use realized measures, which are estimates of the ex-post daily return variation. They can be constructed as the sum of squared intraday returns (realized variance), or using more advanced techniques like kernel estimation (realized kernels), see Barndorff-Nielsen et al. (2008). It is also possible to construct a realized covariance matrix for a group of asset returns, using, for example, the multivariate realized kernel of Barndorff-Nielsen et al. (2011).

In the so-called High-Frequency-Based Volatility (HEAVY) models, originally proposed for univariate settings by Shephard and Sheppard (2010) and extended to multivariate settings by Noureldin et al. (2012), the returns and realized measures are modeled jointly. In these models, the returns and the realized measures are each driven by their own latent time-varying parameter. This contrasts the Realized GARCH model of Hansen et al. (2012), where the daily returns and the realized measures of the variance are both driven by the *same* underlying volatility process. Gorgi et al. (2019a) extend this model to the multivariate setting and introduce score-driven dynamics to the underlying volatility process, which is based on the predictive joint likelihood function of the returns and the realized covariance matrix.

Suppose that for every time point t, there is an N-dimensional observation vector $\mathbf{y}_t$ of returns and that an $N \times N$ realized covariance matrix $\boldsymbol{X}_t$ is available. Then let

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{H}_t^{1/2} \boldsymbol{\varepsilon}_t \,, \\ \boldsymbol{X}_t &= \boldsymbol{V}_t^{1/2} \boldsymbol{\eta}_t \boldsymbol{V}_t^{1/2} \,, \end{aligned} \tag{11}$$

where $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ are serially and mutually independent, and the covariance matrices $\boldsymbol{H}_t$ and $\boldsymbol{V}_t$ are assumed to be $\mathcal{F}_{t-1}$ measurable and are linked via the following dynamic equation

$$\mathbf{H}_t = \boldsymbol{\Lambda} \mathbf{V}_t \boldsymbol{\Lambda}' \,,$$

where $\boldsymbol{\Lambda} = (\lambda_{ij})$ is an $N \times N$ invertible coefficient matrix with $\lambda_{11} > 0$ for identification. The functional link between the time-varying parameter vector $\boldsymbol{f}_t$ and $\boldsymbol{V}_t$ can be set to $\boldsymbol{f}_t = \mathrm{vech}(\boldsymbol{V}_t)$, such that $\boldsymbol{f}_t$ is a $N(N+1)/2$ dimensional vector. It is also possible to consider different parameterizations, as discussed in section "Score-Driven Multivariate Scale Models", but for ease of exposition only this simple parameterization is considered here.

If in (11) $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \boldsymbol{I}_N)$ is multivariate standard normally distributed and $\boldsymbol{\eta}_t \sim \mathcal{W}(\boldsymbol{I}_N/\nu, \nu)$ is Wishart distributed with expectation $\boldsymbol{I}_N$ and degrees of freedom $\nu \geq N$, as in Gorgi et al. (2019a), the scaled score is equal to:

$$\boldsymbol{s}_t = \frac{1}{\nu+1} \left(\nu[\mathrm{vech}(\boldsymbol{X}_t) - \boldsymbol{f}_t] + [\mathrm{vech}(\boldsymbol{\Lambda}^{-1} \mathbf{y}_t \mathbf{y}_t' (\boldsymbol{\Lambda}')^{-1}) - \boldsymbol{f}_t] \right) \,. \tag{12}$$

It stands out that the magnitude of $\boldsymbol{s}_t$ depends on the difference between $\boldsymbol{X}_t$ and $\boldsymbol{V}_t$ weighted by the degrees of freedom of the Wishart distribution ν, whereas the contribution of the deviation of $\mathbf{y}_t \mathbf{y}_t'$ from $\boldsymbol{V}_t$ only receives a weight of one, reflecting the fact that the intraday information is more informative about the underlying covariance structures than the daily returns. This model does use thin-tailed distributions, and therefore there is no downweighting mechanism in the score for larger values of $\mathbf{y}_t \mathbf{y}_t'$ (or $\boldsymbol{X}_t$). Although the model lacks robust updating, in their empirical application to fifteen assets from the NYSE, Gorgi et al. (2019a) find that the model performs well in-sample, and outperforms competing models in out-of-sample density forecasts of the daily returns.

Because realized measures are constructed as sums of intraday returns, which can be noisy, realized covariance matrices typically contain many outliers. The Wishart distribution is not flexible enough to capture this behaviour, as is shown by Opschoor et al. (2018). Hence, the model might improve in case a heavy-tailed distribution is used for modeling the covariance matrix $\boldsymbol{X}_t$, such that the score-driven filter is robust against outliers. For example, consider (11), but let $\boldsymbol{\varepsilon}_t$ be standardized multivariate Student's t distributed with $\nu_0 > 2$ degrees of freedom and let $\boldsymbol{\eta}_t$ be matrix-F distributed with $\nu_1 > N$ and $\nu_2 > N$ degrees of freedom, leading to a model similar to that of Opschoor et al. (2018). The matrix-F distribution is the multivariate counterpart of the regular F distribution and is obtained as a product of a Wishart and an inverse-Wishart distributed random matrices. If $\nu_2 \to \infty$, its distribution becomes Wishart with ν_1 degrees of freedom. For simplicity, consider $\boldsymbol{\Lambda} = \boldsymbol{I}_N$, i.e. $\boldsymbol{V}_t = \boldsymbol{H}_t$. The computation of the information matrix is cumbersome, so as proposed by Opschoor et al. (2018), it is convenient to use the inverse Fischer information matrix of the Normal/Wishart model as a scaling matrix, that is $\boldsymbol{S}_t = \frac{2}{\nu_1+1}\mathcal{B}_N(\boldsymbol{H}_t \otimes \boldsymbol{H}_t)\mathcal{D}_N$ where $\mathcal{B}_N = (\mathcal{D}_N'\mathcal{D}_N)^{-1}\mathcal{D}_N'$. The scaled score $\boldsymbol{s}_t$ then becomes

$$\boldsymbol{s}_t = \frac{1}{\nu_1+1}\Bigg(\nu_1\left[\frac{\nu_1+\nu_2}{\nu_2-N-1}\text{vech}\left(\boldsymbol{X}_t\left(\boldsymbol{I}_N + \frac{\nu_1\boldsymbol{H}_t^{-1}\boldsymbol{X}_t}{\nu_2-N-1}\right)^{-1}\right) - \boldsymbol{f}_t\right] \\ + \left[w_t\text{vech}(\mathbf{y}_t\mathbf{y}_t') - \boldsymbol{f}_t\right]\Bigg),$$

where $w_t = (\nu_0 + N)/(\nu_0 - 2 + \mathbf{y}_t'\boldsymbol{H}_t^{-1}\mathbf{y}_t)$, and where $\boldsymbol{H}_t = \text{devech}(\boldsymbol{f}_t)$. The weighting term w_t downweights large values of $\mathbf{y}_t\mathbf{y}_t'$, and is identical to the weighting term of the plain multivariate Student's t scale model in (9). Furthermore, large values of the realized covariance matrix $\boldsymbol{X}_t$ are also downweighted in a similar manner, which reflects the fact that the matrix-F distribution is fat-tailed. Namely, instead of taking the difference between $\text{vech}(\boldsymbol{X}_t)$ and $\boldsymbol{f}_t$ as in (12), $\boldsymbol{X}_t$ is weighted by a matrix that is inversely related to the matrix $\boldsymbol{X}_t$. So again, the use of heavy-tailed distributions induces a robust updating of $\boldsymbol{f}_t$ because of the score-driven dynamics. If $\nu_0 \to \infty$ and $\nu_2 \to \infty$, the score becomes equal to (12), for $\nu = \nu_1$, because then the innovations $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ are standard normal and Wishart with ν_1 degrees of freedom, respectively. It is also possible to consider another combination of distributions for the innovations, which will automatically lead to other score-driven dynamics.

In the literature numerous other univariate and multivariate score-driven models that use realized measures have been considered. See for instance, Buccheri and Corsi (2019), Opschoor and Lucas (2019), and Opschoor and Lucas (2021).

Score-Driven Dynamic Copula Models

It is well known that copula models are used to characterize the dependence between variables and are highly relevant in financial risk management applications. Sklar's theorem in Sklar (1959) shows that multivariate distributions can be described by a set of marginal distributions and a copula function. Specifically, for the bivariate stochastic process $\{\mathbf{y}_t\}_{t=1}^T$ with $\mathbf{y}_t = (y_{1t}, y_{2t})'$ with conditional joint distribution $\boldsymbol{F}_t$ and conditional marginal distributions F_{1t} and F_{2t}, by the extended Sklar's theorem (Patton, 2006),

$$\mathbf{y}_t|\mathcal{F}_{t-1} \sim \boldsymbol{F}_t = \boldsymbol{C}_t(F_{1t}, F_{2t}),$$

where $\mathcal{F}_{t-1}$ is some information set and $\boldsymbol{C}_t$ is the conditional copula of $\mathbf{y}_t$ containing all information about the dependence between y_{1t} and y_{2t}. In other words, this theorem tells that it is possible to construct $\boldsymbol{F}_t$ by linking together any two marginal distributions F_{1t} and F_{2t} with any copula function using the same information set $\mathcal{F}_{t-1}$. This is a remarkable result since it provides great flexibility in modeling joint distributions. In general, copula models capture a wide range of dependence structures, such as asymmetric, nonlinear, and tail dependence in extreme events.

The copula $\boldsymbol{C}_t$ can be represented using the joint distribution of probability integral transforms of y_{1t} and y_{2t}:

$$\boldsymbol{U}_t|\mathcal{F}_{t-1} \sim \boldsymbol{C}_t, \;\; U_{1t} = F_{1t}(y_{1t}), \;\; U_{2t} = F_{2t}(y_{2t}),$$

where $\boldsymbol{U}_t$ is the probability integral transform. However, more flexibility can be achieved by allowing the copula parameter, δ_t, to be time-varying,

$$\boldsymbol{U}_t|\mathcal{F}_{t-1} \sim \boldsymbol{C}_t(\delta_t).$$

To ensure that the copula parameter, δ_t, lies in a particular range it is usually modeled as a monotonic transformation of f_t, i.e. $\delta_t = h(f_t)$. For example, in bivariate models, the parameter δ_t often corresponds to a correlation parameter. To ensure that $\delta_t \in (-1, 1)$, the hyperbolic tangent transformation function can be used.

Patton (2006) introduced a time-varying copula model and proposed to model the dynamics of f_t. However, the proposed approach does not take into account the copula specification and, in general, the choice of the updating equation is not obvious in many specifications. Score-driven models overcome these issues by defining an updating equation which makes use of the score of the predictive likelihood,

$$\nabla_t = \frac{\partial \log \boldsymbol{c}(\boldsymbol{u_t}; \delta_t)}{\partial \delta_t} \frac{\partial h(f_t)}{\partial f_t},$$

where $\boldsymbol{c}(\boldsymbol{u_t}; \delta_t)$ denotes the copula density.

The unknown model parameters $\boldsymbol{\psi}$, including parameters of the marginal distributions and the parameters of the copula, can be estimated by the method of maximum likelihood. In particular, the joint likelihood of the model is given by,

$$\begin{aligned} \mathcal{L}(\boldsymbol{\psi}) = \sum_{t=1}^{T} \log \boldsymbol{p}(\mathbf{y_t} | f_t, \mathcal{F}_{t-1}; \boldsymbol{\psi}) &= \sum_{t=1}^{T} \log p_1(y_{1t} | \mathcal{F}_{t-1}; \boldsymbol{\psi}_1) + \sum_{t=1}^{T} \log p_2(y_{2t} | \mathcal{F}_{t-1}; \boldsymbol{\psi}_2) \\ &+ \sum_{t=1}^{T} \log \boldsymbol{c}_t(F_{1t}(y_{1t}; \boldsymbol{\psi}_1), F_{2t}(y_{2t}; \boldsymbol{\psi}_2); \boldsymbol{\psi}_c), \end{aligned}$$

where $\boldsymbol{\psi} = (\boldsymbol{\psi}_1', \boldsymbol{\psi}_2', \boldsymbol{\psi}_c')'$ are the parameters to be estimated, including the parameters of the marginal distributions and the parameters of the copula. Theoretically, the parameters can be estimated from the joint likelihood. However, to reduce the computational burden the estimation is usually carried out in two steps. First, the parameters of the marginal distribution are estimated, next, the copula parameters are obtained conditional on the estimated marginal distribution parameters.

In the literature, different score-driven copula models have been considered. For example, Creal et al. (2013) introduced a bivariate score-driven Gaussian copula model and found that it generalizes the approach proposed by Patton (2006). In Figure 4, a comparison between the news impact curves (NICs) for these two models is provided. One can notice that the NIC of Patton (2006) model is independent of the value of the current correlation coefficient δ. In turn, the NIC for the score-driven model resembles the NIC of Patton (2006) model for low values of the correlation coefficient δ. However, when $\delta > 0$ the NIC the score-driven update is not linear anymore. In the score-driven model, "news" have a more pronounced "negative" effect on the transformed correlation coefficient for distant values of y_1 and y_2, with the difference between the models increasing with the increase in δ (in absolute values) and increase in the distance between y_1 and y_2. Specifically, if the correlation coefficient is large and positive, the "news" will have a large negative effect on the update when the difference between y_1 and y_2 is large.

In contrast to the Gaussian copula, the score-driven Student's t copula model limits the impact

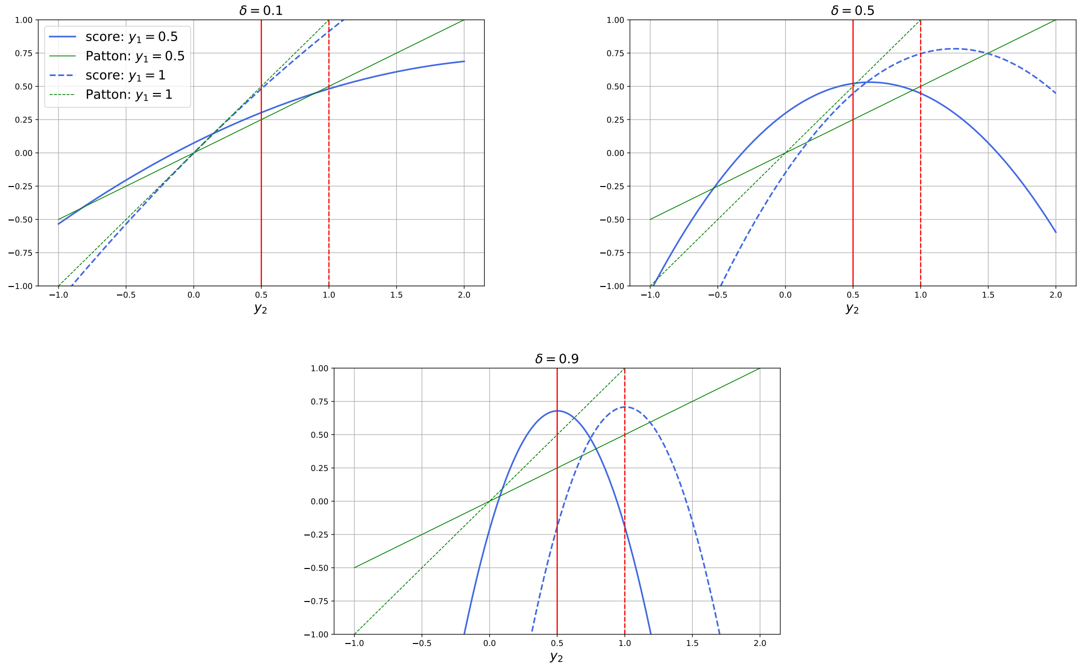


Figure 4: News impact curves of the Patton's and score-driven Gaussian copula models
News impact curves of the Patton's (green lines) and score-driven (blue lines) Gaussian copula models when y_2 varies, while y_1 is fixed at 0.5 (solid line) or 1 (dashed line) for different values of the correlation parameter $\delta = \{0.1, 0.5, 0.9\}$. The score is scaled by the Fisher information matrix with $\zeta = 1/2$ where ζ was defined in section "General Framework of Score-Driven Models".

of the possible outliers by allowing dependence in the tails. Janus et al. (2014) use a score-driven model with Student's t marginals and Student's t copula functions to model volatility and correlation in financial time series with long memory dynamics. They highlight that the robustness features of the Student's t copula are especially important for long memory models since the impact of the score on the update vanishes at a slow rate. Moreover, to capture heavy tailed properties as well as skewness, Lucas et al. (2014) use a score-driven model with a skewed-t distribution. Copula models that allow for asymmetry in the tails can also be considered. In particular, the Clayton copula provides support to the lower tail dependence, while the Gumbel copula supports upper tail dependence.

Several further extensions have been proposed in the literature. For example, Salvatierra and Patton (2015) adapt score-driven dynamic copula models by incorporating high-frequency information. They propose a model where the updating equation is equipped with specific realized measures of correlation. Bernardi and Catania (2019) extend the model with copula dependence parameters that are regime-specific and are updated according to a score-driven Markov switching model.

Discussion

Many other extensions of score-driven models have been proposed in addition to the classes of models discussed in the previous sections. The list is too long for an exhaustive review of all score-driven models available in the literature. For completeness, a brief mention of some additional examples is provided.

In forecasting applications, the choice of the loss function/M-estimator depends on the application in use. As a result, there are situations when it is desirable not to use a likelihood but some other criterion function. For example, during the COVID-19 pandemic, one could argue that the cost of underpedicting COVID related deaths outweighs the cost of overprediction. In such scenarios, asymmetric loss functions can play an important role. This naturally leads to the idea of the time-varying parameter update being based on the score of the loss function under consideration. In contrast to score-driven models, this approach eliminates the need of specifying any parametric distribution which can be beneficial when the model is potentially misspecified.

One of the first papers that goes beyond the parametric framework is the paper by Creal et al. (2018). Economic models usually do not provide information about the parametric ditribution of the data while economic theory can often provide moment conditions. Therefore, it can be desirable to use Generalized method of moments (GMM) estimator instead of the likelihood. Exploiting this idea, Creal et al. (2018) propose to model the dynamics of time-varying parameters using a score of the local GMM criterion function. Furthermore, Patton et al. (2019) introduce a new semiparametric model for Value-at-Risk (VaR) and Expected Shortfall (ES) dynamics. VaR and ES can be consistently estimated using, so called, "FZ0 loss function", introduced by Fissler and Ziegel (2016). Hence, it is intuitive to base the updating equation on the score of this loss function. A similar idea was exploited in Catania and Luati (2019), who propose to model dynamics of the quantiles of a time series based on the score of the quantile loss function.

Blasques et al. (2022) introduce a new general class of quasi score-driven models. This class of models can incorporate many popular loss functions, which can, for example, make the update of the time-varying parameter robust to outliers as in score-driven models. The estimation can be carried out using the method of quasi-likelihood. Therefore, in contrast to score-driven models, there is no direct link between the functional form of the updating equation and the density of the innovations or loss function. Similar to the score-driven models, this class gives rise to a wide range of models that can be used in a variety of empirical applications, making the observation-driven modeling approach even more appealing.

Acknowledgement

The authors are grateful to Ekaterina Ugulava for her excellent research assistance, Julia Schaumburg for providing the code for score-driven dynamic spatial model illustration, and Marco Bazzi for providing the code for score-driven dynamic Markov switching illustration.

Further reading

Artemova, M., Blasques, F., van Brummelen, J., and Koopman, S. J. (2022). Score-Driven Models: Methodology and Theory. In *Oxford Research Encyclopedia of Economics and Finance*, page tba. Oxford University Press.

Creal, D., Koopman, S. J., and Lucas, A. (2008). A general framework for observation driven time-varying parameter models. Technical report, Tinbergen Institute Discussion paper.

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

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

Harvey, A. C. (2022). Score-driven time series models. *Annual Review of Statistics and Its Application*, 9:321–342.

References

Anselin, L. (1988). *Spatial econometrics: methods and models*, volume 4. Springer Science & Business Media.

Artemova, M., Blasques, F., van Brummelen, J., and Koopman, S. J. (2022). Score-Driven Models: Methodology and Theory. In *Oxford Research Encyclopedia of Economics and Finance*, page tba. Oxford University Press.

Babii, A., Chen, X., and Ghysels, E. (2019). Commercial and residential mortgage defaults: Spatial dependence with frailty. *Journal of Econometrics*, 212(1):47–77.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162(2):149–169.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109.

Bazzi, M., Blasques, F., Koopman, S. J., and Lucas, A. (2017). Time-varying transition probabilities for Markov regime switching models. *Journal of Time Series Analysis*, 38(3):458–478.

Bernardi, M. and Catania, L. (2019). Switching generalized autoregressive score copula models with application to systemic risk. *Journal of Applied Econometrics*, 34(1):43–65.

Billé, A. G., Blasques, F., and Catania, L. (2019). Dynamic spatial autoregressive models with time-varying spatial weighting matrices. *Available at SSRN 3241470*.

Blasques, F., Francq, C., and Laurent, S. (2022). Quasi score-driven models. *Journal of Econometrics*.

Blasques, F., Holý, V., and Tomanová, P. (2018). Zero-inflated autoregressive conditional duration model for discrete trade durations with excessive zeros. *arXiv preprint arXiv:1812.07318*.

Blasques, F., Hoogerkamp, M. H., Koopman, S. J., and van de Werve, I. (2021). Dynamic factor models with clustered loadings: Forecasting education flows using unemployment data. *International Journal of Forecasting*, 37:forthcoming.

Blasques, F., Koopman, S. J., and Lucas, A. (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika*, 102(2):325–343.

Blasques, F., Koopman, S. J., and Lucas, A. (2020). Nonlinear autoregressive models with optimality properties. *Econometric Reviews*, 39(6):559–578.

Blasques, F., Koopman, S. J., Lucas, A., and Schaumburg, J. (2016). Spillover dynamics for systemic risk measurement using spatial financial time series models. *Journal of Econometrics*, 195(2):211–223.

Böhm, H., Schaumburg, J., and Tonzer, L. (2022). Financial linkages and sectoral business cycle synchronization: Evidence from europe. *IMF Economic Review*, pages 1–37.

Buccheri, G. and Corsi, F. (2019). HARK the SHARK: Realized Volatility Modeling with Measurement Errors and Nonlinear Dependencies*. *Journal of Financial Econometrics*. nbz025.

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

Catania, L. and Billé, A. G. (2017). Dynamic spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Applied Econometrics*, 32(6):1178–1196.

Catania, L. and Luati, A. (2019). Semiparametric modeling of multiple quantiles. *Available at SSRN 3494995*.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, 8(2):93–115.

Creal, D., Koopman, S. J., and Lucas, A. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics*, 29(4):552–563.

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

Creal, D., Koopman, S. J., Lucas, A., and Zamojski, M. (2018). Generalized autoregressive method of moments.

Creal, D., Schwaab, B., Koopman, S. J., and Lucas, A. (2014). Observation-driven mixed-measurement dynamic factor models with an application to credit risk. *Review of Economics and Statistics*, 96(5):898–915.

Delle Monache, D. and Petrella, I. (2017). Adaptive models and heavy tails with an application to inflation forecasting. *International Journal of Forecasting*, 33(2):482–501.

Diebold, F. X., Lee, J.-H., and Weinbach, G. C. (1994). Regime switching with time-varying transition probabilities. In Granger, C. W. J. and Mizon, G., editors, *Nonstationary Time Series Analysis and Cointegration*, pages 283–302. Advanced Texts in Econometrics, Oxford University Press.

Doz, C., Giannone, D., and Reichlin, L. (2012). A quasi–maximum likelihood approach for large, approximate dynamic factor models. *Review of Economics and Statistics*, 94(4):1014–1024.

Durbin, J. and Koopman, S. J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):3–56.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics*, 12(3):299–308.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *Annals of Statistics*, 44(4):1680–1707.

Foglia, M. and Angelini, E. (2019). The time-spatial dimension of eurozone banking systemic risk. *Risks*, 7(3):75.

Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, 104(488):1430–1439.

Gorgi, P. (2018). Integer-valued autoregressive models with survival probability driven by a stochastic recurrence equation. *Journal of Time Series Analysis*, 39(2):150–171.

Gorgi, P. (2020). Beta-negative binomial auto-regressions for modelling integer-valued time series with extreme observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82:1325–1347.

Gorgi, P., Hansen, P., Janus, P., and Koopman, S. J. (2019a). Realized Wishart-GARCH: A score-driven multi-asset volatility model. *Journal of Financial Econometrics*, 17(1):1–32.

Gorgi, P., Koopman, S. J., and Li, M. (2019b). Forecasting economic time series using score-driven dynamic models with mixed-data sampling. *International Journal of Forecasting*, 35(4):1735–1747.

Gouriéroux, C. and Monfort, A. (1997). *Simulation-based Econometric Methods*. Oxford University Press.

Haas, M., Krause, J., Paolella, M. S., and Steude, S. C. (2013). Time-varying mixture GARCH models and asymmetric volatility. *The North American Journal of Economics and Finance*, 26:602–623.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

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

Harvey, A. C. and Kattuman, P. (2020). Time series models based on growth curves with applications to forecasting coronavirus. *Harvard Data Science Review*.

Harvey, A. C. and Palumbo, D. (2021). Regime switching models for directional and linear observations. *Discussion Paper*.

Janus, P., Koopman, S. J., and Lucas, A. (2014). Long memory dynamics for multivariate dependence under heavy tails. *Journal of Empirical Finance*, 29:187–206.

Liu, Y. (2016). *Time-varying correlation and common structures in volatility*. PhD Thesis: University of Amsterdam.

Lucas, A., Schwaab, B., and Zhang, X. (2014). Conditional euro area sovereign default risk. *Journal of Business & Economic Statistics*, 32(2):271–284.

Noureldin, D., Shephard, N., and Sheppard, K. (2012). Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, 27(6):907–933.

Oh, D. H. and Patton, A. J. (2017). Modeling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics*, 35(1):139–154.

Oh, D. H. and Patton, A. J. (2018). Time-varying systemic risk: Evidence from a dynamic copula model of CDS spreads. *Journal of Business & Economic Statistics*, 36(2):181–195.

Opschoor, A., Janus, P., Lucas, A., and Van Dijk, D. (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics*, 36(4):643–657.

Opschoor, A. and Lucas, A. (2019). Fractional integration and fat tails for realized covariance kernels. *Journal of Financial Econometrics*, 17(1):66–90.

Opschoor, A. and Lucas, A. (2021). Observation-driven models for realized variances and overnight returns applied to Value-at-Risk and expected shortfall forecasting. *International Journal of Forecasting*, 37(2):622–633.

Opschoor, A., Lucas, A., Barra, I., and Van Dijk, D. (2020). Closed-form multi-factor copula models with observation-driven dynamic factor loadings. *Journal of Business & Economic Statistics*, 38(1):1–14.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, 211(2):388–413.

Salvatierra, I. D. L. and Patton, A. J. (2015). Dynamic copula models and high frequency data. *Journal of Empirical Finance*, 30:120–135.

Shephard, N. and Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, 25(2):197–231.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Zhang, D. and Broadstock, D. C. (2020). Global financial crisis and rising connectedness in the international commodity markets. *International Review of Financial Analysis*, 68:101239.

Zhang, X., Creal, D., Koopman, S. J., and Lucas, A. (2011). Modeling dynamic volatilities and correlations under skewness and fat tails. *Discussion Paper*.

Zhu, F. (2011). A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis*, 32(1):54–67.

Notes

1. Monthly seasonally adjusted unemployment rate series has been retrieved from `https://fred.stlouisfed.org`.
2. The average crude oil spot prices series has been retrieved from the `investing.com` website.