Journal of Econometrics 253 (2026) 106155

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Unobserved component models, approximate filters and dynamic adaptive mixture models

Leopoldo Catania [a,*], Enzo D'Innocenzo [b], Alessandra Luati [b,c]

[a] University of Aarhus and CoRE, Denmark
[b] University of Bologna, Italy
[c] Imperial College London, United Kingdom

ARTICLE INFO

Keywords:
Mixture models
Score driven models
Robust filters

ABSTRACT

State estimation in unobserved component models with parameter uncertainty is traditionally performed through approximate filters, where Gaussian distributions with given moments are employed to replace otherwise intractable conditional densities. This paper re-examines signal-plus-noise models where parameter uncertainty is induced by a latent variable that may assume a fixed number of states. First, it is shown that, for these models, the approximate filters commonly adopted in the literature can be obtained as linear combinations of minimum variance linear unbiased estimators. Second, it is observed that they coincide with filters implied by a novel class of dynamic adaptive mixture models, where the parameters of a mixture of distributions evolve over time following a recursion that is based on the score of the one-step-ahead predictive distribution. Focusing on a robust specification, where the mixture components are Student's t distributions, we prove existence, stationarity, and ergodicity of the data generating process as well as invertibility of the filter, and consistency and asymptotic normality of the maximum likelihood estimator of the static parameters. An application to energy spot prices is discussed, where the novel specification is compared with, and shown to outperform, robust score-driven filters and the related class of mixture autoregressive models.

1. Introduction

Filtering in the Gaussian signal-plus-noise models is optimally performed through the Kalman filter (Kalman, 1960; Kalman and Bucy, 1961), see also Harvey (1989), West and Harrison (1997) and Durbin and Koopman (2012). When the system parameters change over time and are not predictable without uncertainty, then Kalman filter updates are modified in ways that depend on the nature of parameter variation.

A natural form of parameter uncertainty may be induced by a latent variable that selects the value of the system parameters among a fixed number of alternatives. Harrison and Stevens (1976) emphasise that this case is likely to always represent the real situation in socio-economic applications and in many production processes. They also recognise that accounting for parameter uncertainty often induces non-linearities in the model, which lead to filtering recursions that are not operable.

A feasible solution is provided by Shumway and Stoffer (1991), who analysed the Gaussian signal-plus-noise model with parameter uncertainty, under the assumption that the latent variable is identically and independently distributed over time. The filter they derive relies on approximating the distribution of the observables, given the past history and the realisation of the latent variable, by

* Corresponding author.
E-mail addresses: leopoldo.catania@econ.au.dk (L. Catania), enzo.dinnocenzo2@unibo.it (E. D'Innocenzo), a.luati@imperial.ac.uk (A. Luati).

https://doi.org/10.1016/j.jeconom.2025.106155
Received 16 December 2024; Received in revised form 4 November 2025; Accepted 18 November 2025
Available online 28 November 2025
0304-4076/

a Gaussian. As such, the filter belongs to the family of approximate filters based on a "collapsing step", where collapsing is referred to the practice of replacing intractable densities in filtering recursions by Gaussian densities, as discussed by Kim (1994) and Kim and Nelson (1999). Analog filters, based on some approximations or collapsing steps, have been widely applied over the years, see Ackerson and Fu (1970), Bar-Shalom and Tse (1975), Gordon and Smith (1990), Akashi and Kumamoto (1977), Shephard (1994), Godsill (1997), Gerlach et al. (2000), Kalpakis et al. (2001), Frühwirth-Schnatter (2001), Xiong and Yeung (2002), and Yu (2012). The book of Kim and Nelson (1999) covers the case when the latent variable is a first order Markov chain, while Chen and Liu (2000) develop a generic sequential Monte Carlo method to perform approximate filtering in this class of models. In all cases, it is not obvious to figure out how the induced approximations affect the filtering procedure.

As a first contribution, this paper shows that the approximate estimator obtained by Shumway and Stoffer (1991) can be directly derived, with no underlying distributional assumption and no collapsing step, as a linear combination of minimum variance unbiased estimators, with weights determined to ensure unbiasedness. So far, this seems to be the sole exact result linking unobserved component models with parameter uncertainty and approximate filters.

Second, we observe that the same filter is implied by a dynamic adaptive mixture model, with Gaussian component densities and score-driven type updates. This allows us to embed it in a different framework that turns out to encompass a much wider class of estimators, designed either to approximate filters for non-linear non-Gaussian unobserved component models or to directly estimate time-varying parameters of conditional densities. The framework is provided by the dynamic adaptive mixture models introduced by Catania (2019), where the parameters of a mixture of distributions evolve over time following a recursion that is based on the score of the one-step-ahead predictive distribution. This way of updating dynamic parameters is based on the score-driven modelling methodology introduced by Creal et al. (2013) and Harvey (2013), and pioneered by Masreliez (1975). The novelty here lies in considering mixtures of Student's *t* distributions.

The use of the score as an updating mechanism for time-varying parameters of a conditional density in observation-driven models has received growing attention in recent years, due to its tractability in non-Gaussian models (see, among others, Blasques et al., 2023, 2022; Francq and Zakoian, 2023; Gorgi, 2020; Hansen and Schmidtblaicher, 2021; Holý and Tomanová, 2022). Within this literature, the properties of score-driven models when interpreted as approximating filters associated with non-Gaussian parameter-driven models have been investigated. The filters by Harvey and Luati (2014) and Gasperoni et al. (2023), for example, recover the innovation form of the Kalman filter in a linear Gaussian model as the limit case of a Student's *t* model, where the Kalman gain is treated as an unknown parameter to be estimated. Monache et al. (2021) specify a score-driven model for the system matrices of a linear Gaussian model, and develop a hybrid state-space score-driven model that retains conditionally Gaussian state and observation densities but requires a modified Kalman filter algorithm for filtering and smoothing. Buccheri et al. (2025) adopt a perturbation approach, where the observation density is expanded around a particular solution that corresponds to a degenerate prior for the unobserved state. On the applied side, a comparison on the predictive performance of parameter-driven and observation-driven models is considered by Koopman et al. (2016). Our contribution in this paper is to provide a formal link between unobserved component models with parameter uncertainty and score-driven filters, by showing that the former are approximated by Gaussian dynamic adaptive mixture score-driven filters that weight minimum variance unbiased estimators with filtered probabilities, to ensure unbiasedness.

If interpreting these processes either as approximated filters of parameter-driven models or as observation-driven models for time-varying parameters is a matter of context or background (Cox, 1981), embedding them in the framework discussed in the paper provides the analytic solution for the rigorous study of their stochastic and asymptotic properties. The key argument is provided by the score of a Student's *t* log-likelihood, a bounded function of the prediction error that allows us to prove existence, stationarity, and ergodicity of the data generating process as well as invertibility of the filter and thus consistency and asymptotic normality of the maximum likelihood estimator of the static parameters.

The theoretical properties in the limit case of Gaussian mixing densities remain a challenge, as the prediction error (the score), there, is linear in the observations and, thus, unbounded. The Gaussian mixture case is discussed throughout the paper and results are obtained as the limit case of infinite degrees of freedom, whenever these limits, within the methodology developed in this paper, exist.

In practice, the filter implied by a dynamic mixture model with Student's *t* densities is robust to extreme observations, in the sense of Calvet et al. (2015). Indeed, the model encompasses the robust filter by Harvey and Luati (2014) as a particular case. Note that this property is not valid in the Gaussian case, where the response to extreme observations becomes linear in the limit. The robustness of the score of a heavy-tailed distribution associated with the flexibility of the mixing design make the model suitable for signal extraction and prediction from datasets exhibiting instabilities, spikes, and extreme events.

To illustrate, we analyse an energy time series dataset, where it is crucial that predictions are not affected by abrupt changes with transitory effects. Estimates from Student's *t* and Gaussian dynamic mixture models are compared with estimates from a contaminated signal-plus-noise model computed with the Kalman filter, with the robust filter by Harvey and Luati (2014) and with the related class of mixtures of autoregressions of Wong and Li (2000) and Wong et al. (2009). In the latter, the parameter values of pure autoregressive processes depend on the realisation of a latent variable, assumed to be independent over time. The first specification of this kind is the Gaussian mixture transition distribution model of Le et al. (1996), based on previous results by Raftery (1985) and Martin and Raftery (1987), where the authors comment on the article by Kitagawa (1987), also related to the same class of models. Wong and Li (2000) have generalised the Gaussian mixture transition model and developed the mixture autoregressive model, see also the discussion in Berchtold and Raftery (2002), then further extended by Wong et al. (2009) to the case of Student's *t* mixtures of autoregressions. The results of the analysis will show that none of the aforementioned models outperform the Student's *t* dynamic adaptive mixture model developed in the paper.

The rest of the paper is organised as follows. Section 2 discusses the link between the Gaussian linear signal-plus-noise model with parameter uncertainty and the dynamic adaptive mixture model specification. Section 3 introduces the general specification of the model and then specialises to the Gaussian and Student's t mixture components. In Section 4, the stochastic properties of the Student's t specification are derived, while the asymptotic properties of the maximum likelihood estimator are studied in Section 5. The results of an extensive simulation study are summarised in Section 6. The data analysis is deferred to Section 7. Concluding remarks and directions for further research are outlined in Section 8. Proofs of theoretical results, technical details and additional results appear as supplementary materials available online.

2. Unobserved component models with parameter uncertainty

Let $\{y_t\}_{t\in\mathbb{Z}}$ be a sequence of random variables whose realisations at time $t = 1, 2, \ldots, T$ are interpreted as noisy observations of an unobservable process $\{\mu_t\}_{t\in\mathbb{Z}}$, often referred to as the underlying signal of $\{y_t\}_{t\in\mathbb{Z}}$. Let $\{S_t\}_{t\in\mathbb{Z}}$ be a sequence of independent and identically distributed (iid) unobserved Categorical random variables taking values in the state space $\mathcal{S} = \{1, \ldots, J\}$. To illustrate the problem, consider a simple signal-plus-noise model with parameter uncertainty

$$\begin{array}{lcll} y_t & = & \tau(S_t)\mu_t + \varepsilon_t, & \varepsilon_t \sim \mathrm{iid}(0, \sigma^2) \\ \mu_{t+1} & = & \phi\mu_t + \eta_t, & \eta_t \sim \mathrm{iid}(0, \varpi) \end{array} \tag{1}$$

where $\tau(S_t) = \tau_j$ with probability $\alpha_j = P(S_t = j)$, $j = 1, \ldots, J$, $\sum_{j=1}^{J} \alpha_j = 1$; the error terms ε_s and η_r are mutually independent and independent of S_t, for all $t, r, s \in \mathbb{Z}$. At each time point, the latent variable S_t determines the system parameter τ_j by selecting its value among a fixed number of alternatives, each one occurring with probability α_j. For simplicity of exposition, we assume that only τ depends on S_t. Nonetheless, the subsequent analysis remains valid in the more general case where σ^2 and ϖ also depend on S_t.

Let $\mathcal{F}_t = \sigma(y_{t-s}, s \geq 0)$ denote the sigma-algebra generated by the random variables $y_t, y_{t-1}, \ldots$ and define the one-step-ahead predictor $\mu_{t|t-1} = \mathbb{E}(\mu_t|\mathcal{F}_{t-1})$ with variance $P_{t|t-1} = \mathrm{Var}(\mu_t|\mathcal{F}_{t-1})$. The optimal filtering problem requires deriving the filtered estimator $\mu_{t|t} = \mathbb{E}(\mu_t|\mathcal{F}_t)$ as the mean of the conditional distribution of the unobserved state μ_t given the past and the current observation, then used to update the predicted values $\mu_{t+1|t}$. (Durbin and Koopman, 2012) show that, with no underlying distributional assumptions, the minimum variance linear unbiased estimator of μ_{t+1} conditional on $\mathcal{F}_t$ is equivalent to $\mu_{t+1|t}$ in a linear Gaussian model. In the following section, we extend the result to the non-linear non-Gaussian model in Eq. (1).

2.1. Unbiased filters

Conditional on $\mathcal{F}_{t-1}$ and $S_t = j$, the minimum variance linear unbiased estimator (MVLUE) of μ_{t+1} is $\tilde{\mu}_{j,t+1} = \tilde{\beta}_{j,t} + \tilde{\kappa}_{j,t} y_t$, where $\tilde{\beta}_{j,t} = \mu_{t|t-1}(\phi - \tilde{\kappa}_{j,t}\tau_j)$ and $\tilde{\kappa}_{j,t} = \phi P_{t|t-1}\tau_j/(\tau_j^2 P_{t|t-1} + \sigma^2)$ are obtained by equating $\mathbb{E}(\tilde{\mu}_{j,t+1} - \mu_{t+1}|\mathcal{F}_{t-1}, S_t = j) = 0$ and then minimising $\mathrm{Var}(\tilde{\mu}_{j,t+1} - \mu_{t+1}|\mathcal{F}_{t-1}, S_t = j)$, so that $\tilde{\mu}_{j,t+1} = \phi\mu_{t|t-1} + \tilde{\kappa}_{j,t}v_{j,t}$, with $v_{j,t} = y_t - \tau_j\mu_{t|t-1}$.

A weighted average of conditional MVLUE, $\tilde{\mu}_{t+1} = \sum_{j=1}^{J} \tilde{\xi}_{j,t}\tilde{\mu}_{j,t+1}$, with weights given by the filtered probabilities $\tilde{\xi}_{j,t} = P(S_t = j|\mathcal{F}_t)$, such that $\tilde{\xi}_{j,t} > 0$ for $j = 1, \ldots, J$ and $\sum_{j=1}^{J} \tilde{\xi}_{j,t} = 1$, gives the unbiased estimator

$$\tilde{\mu}_{t+1} = \phi\mu_{t|t-1} + \sum_{j=1}^{J} \tilde{\xi}_{j,t}\tilde{\kappa}_{j,t}v_{j,t}. \tag{2}$$

The latter satisfies $\mathbb{E}(\tilde{\mu}_{t+1} - \mu_{t+1}|\mathcal{F}_{t-1}) = 0$, since $\mathbb{E}(\tilde{\xi}_{j,t}v_{j,t}|\mathcal{F}_{t-1}) = \alpha_j\mathbb{E}(v_{j,t}|\mathcal{F}_{t-1}, S_t = j) = 0$ for all j, as the filtered probabilities can be written as

$$\tilde{\xi}_{j,t} = \frac{\alpha_j p(y_t|\mathcal{F}_{t-1}, S_t = j)}{\sum_{k=1}^{J} \alpha_k p(y_t|\mathcal{F}_{t-1}, S_t = k)} \tag{3}$$

where $p(y_t|\mathcal{F}_{t-1}, S_t = j)$ is the density of the random variable y_t conditional on $\mathcal{F}_{t-1}$ and $S_t = j$, i.e. to its past and to the system being at time t in the state j.

The unbiased estimator derived in Eq. (2) assigns weights to J conditional Kalman gains, $\tilde{\kappa}_{j,t}$, and prediction errors, $v_{j,t}$, according to the filtered probabilities $\tilde{\xi}_{j,t}$, and reduces to the Kalman filter in the case when $S_t = j$ with probability 1. If conditioning is restricted to $\mathcal{F}_{t-1}$, $\tilde{\beta}_t = \mu_{t|t-1}(\phi - \tilde{\kappa}_t\bar{\tau})$ and $\tilde{\kappa}_t = (\phi P_{t|t-1}\bar{\tau})/[\mu_{t|t-1}^2(v - \bar{\tau}^2) + vP_{t|t-1} + \sigma^2]$, respectively, where $\bar{\tau} = \sum_{j=1}^{J} \alpha_j\tau_j$ and $v = \sum_{j=1}^{J} \alpha_j\tau_j^2$. The MVLUE becomes $\tilde{\mu}_{t+1} = \phi\mu_{t|t-1} + \tilde{\kappa}_t v_t$, with $v_t = y_t - \tau\mu_{t|t-1}$. If $J = 1$, i.e. there is no uncertainty in the system parameter, then $v = \bar{\tau}^2$. The MVLUE thus coincides with the Kalman filter and $\tilde{\kappa}_t$ is the Kalman gain, see Durbin and Koopman (2012). Let $\mathcal{S}_t = \sigma(S_{t-s}, s \geq 0)$, then if it is assumed that S_t is observed with a lag, then $p(y_t|\mathcal{F}_{t-1}, S_{t-2} = j_2, S_{t-3} = j_3, \ldots)$, where $j_i \in \mathcal{S}$ for $i \in \mathbb{N}$, is the Gaussian density, and the estimator in Eq. (2) is the uniformly minimum variance linear unbiased estimator (UMVLUE) since, in this case, $\mathbb{E}(\mu_{t+1}|\mathcal{F}_t, \mathcal{S}_{t-1}) = \tilde{\mu}_{t+1}$.

2.2. Approximate filters

The estimator (2) coincides with the filter obtained by Shumway and Stoffer (1991, equations 11, 14, 16), by assuming that $p(y_t|\mathcal{F}_{t-1}, S_t = j)$ is the Gaussian density and then applying standard results related to the multivariate Normal distribution. Unbiasedness is verified a posteriori.

For the model in Eq. (1), $p(y_t|\mathcal{F}_{t-1}, S_t = j)$ is not the Gaussian density, even if ε_t and η_t are jointly Gaussian, because $p(y_t|\mathcal{F}_{t-1}, S_t = j, S_{t-1} = j_1, \dots)$, which in this case is Gaussian, is not equal to $p(y_t|\mathcal{F}_{t-1}, S_t = j)$, which is a mixture of Gaussian densities even if S_t is iid, in that

$$p(y_t|\mathcal{F}_{t-1}, S_t = j) = \lim_{s\to\infty} \sum_{j_1=1}^{J} \cdots \sum_{j_s=1}^{J} p(y_t|\mathcal{F}_{t-1}, S_t = j, S_{t-1} = j_1, \dots S_{t-s} = j_s) \prod_{m=1}^{s} \alpha_{j_m}, \tag{4}$$

see the detailed discussion in Chapter 6 of Shumway and Stoffer (2017).

A similar scheme has been explicitly adopted by Bar-Shalom and Tse (1975) under the name of probabilistic data association filters, by Harrison and Stevens (1976) in their multi-process models, and by Gordon and Smith (1990) in their extension of the linear dynamic model. These methods rely on a "collapsing step", as described in Kim (1994) and Kim and Nelson (1999), where an intractable density is replaced by the Gaussian density, which is required to make the filter operable. What characterises all the aforementioned approximate filters is the inclusion of the filtered probabilities, $\tilde{\xi}_{j,t}$, that play a predominant role also in the Hamilton (1989) filter, as well as in several approximate filters for non-linear models, such as the filters of Gray (1996) and Klassen (2002) for Markov-Switching models with generalised autoregressive conditional heteroskedasticity (GARCH), and the mixture Kalman filters by Chen and Liu (2000).

3. Dynamic adaptive mixture models

In its more general formulation, the dynamic adaptive mixture model by Catania (2019) for time-varying conditional locations assumes that, for $j = 1, \dots, J$,

$$p(y_t|\mathcal{F}_{t-1}) = \sum_{j=1}^{J} \alpha_j p_j(y_t; \mu_{j,t|t-1}, \boldsymbol{\theta}_j) \tag{5}$$

with

$$\mu_{j,t+1|t} = \omega_j(1-\phi_j) + \phi_j \mu_{j,t|t-1} + \kappa_j \xi_{j,t} s_j(y_t, \mu_{j,t|t-1}) \tag{6}$$

where

$$s_j(y_t, \mu_{j,t|t-1}) = \frac{\partial \log p_j(y_t; \mu_{j,t|t-1}, \boldsymbol{\theta}_j)}{\partial \mu_{j,t|t-1}}$$

is the score function of $\mu_{j,t|t-1}$ associated with the mixture component density $p_j(y_t; \mu_{j,t|t-1}, \boldsymbol{\theta}_j)$, depending on the static parameter $\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j \subset \mathbb{R}^m$, and

$$\xi_{j,t} = \frac{\alpha_j p_j(y_t; \mu_{j,t|t-1}, \boldsymbol{\theta}_j)}{\sum_{j=1}^{J} \alpha_j p_j(y_t; \mu_{j,t|t-1}, \boldsymbol{\theta}_j)} \tag{7}$$

with $\alpha_j > 0$, $\sum_{j=1}^{J} \alpha_j = 1$ and $\omega_j, \phi_j, \kappa_j$ real coefficients. Given that

$$\xi_{j,t} s_j(y_t, \mu_{j,t|t-1}) = \frac{\partial \log p(y_t|\mathcal{F}_{t-1})}{\partial \mu_{j,t|t-1}}, \tag{8}$$

the resulting estimator for μ_{t+1} based on $\mathcal{F}_{t-1}$ is

$$\mu_{t+1|t} = \sum_{j=1}^{J} \alpha_j \omega_j (1-\phi_j) + \sum_{j=1}^{J} \alpha_j \phi_j \mu_{j,t|t-1} + \sum_{j=1}^{J} \alpha_j \kappa_j \xi_{j,t} s_j(y_t, \mu_{j,t|t-1}). \tag{9}$$

The properties of the score function in tracking time-varying parameters of conditional densities have been studied based on information-theoretic criteria (Blasques et al., 2015) and explicit and implicit stochastic gradient descent methods (Gorgi et al., 2023; Lange et al., 2025). The results typically hold under the hypothesis of a misspecified true (unknown) density. For both explicit and implicit stochastic gradient descent methods, Heel et al. (2025) have recently derived a mean square error bound for misspecified score-driven filters, with respect to the pseudo-true parameter, that requires mild regularity conditions.

3.1. Gaussian components and approximate filters

If the mixing densities in (5) are Gaussian, i.e. $\log p_j(y_t; \mu_{j,t|t-1}, \boldsymbol{\theta}_j) \propto \left(y_t - \mu_{j,t|t-1}\right)^2$, then $s_j(y_t, \mu_{j,t|t-1}) \propto y_t - \mu_{j,t|t-1}$, i.e. the score is proportional to the one-step-ahead conditional prediction error and, as such, it is a linear (unbounded) function of y_t.

If $p_j(y_t; \mu_{j,t|t-1}, \boldsymbol{\theta}_j)$ is used to approximate $p(y_t|\mathcal{F}_{t-1}, S_t = j)$ (a collapsing step), then the $\xi_{j,t}$ in (7) are equal to the filtered probabilities $\tilde{\xi}_{j,t}$ in Eq. (3), and, with $\omega_j = 0$ and $\phi_j = \phi$, it is immediate to recognise that (9) is the estimator obtained by Shumway and Stoffer (1991), up to the scaling term $\bar{\tau}$ which, in this case, is directly included in $\mu_{t|t-1}$.

Interpreting each mixture weight α_j as the prior probability of sampling from the j-th component and the ratio $p_j(y_t; \mu_{j,t|t-1}, \boldsymbol{\theta}_j)/p(y_t|\mathcal{F}_{t-1})$ as the likelihood at time t for component j given information up to time $t-1$, then $\xi_{j,t}$ is the posterior probability of sampling from the j-th mixture component. Thus, dynamic adaptive mixture models with Gaussian mixing densities provide a rationale for approximate filters based on collapsing steps.

Note that specifying a dynamic adaptive mixture model like (5) does not entail assuming an underlying unobserved component model like (1). In the terminology of Cox (1981), the latter is a parameter-driven model, where the dynamics of the unobserved component are driven by an error term characterised by its own distribution, while equation (5) specifies an observation-driven model, where the dynamics of the time-varying signal are a function of the random variables $y_{t-1}, y_{t-2}, \dots$ only. In general, for the first class of models it is relatively straightforward to establish stationarity and ergodicity and often privative to have closed-form expressions for the filtered densities. Conversely, in observation driven models, conditional densities are available in closed form by construction, but establishing the probabilistic properties is often complicated due to the non-linearity of the filter.

3.2. Student's *t* components and robust filters

Let the mixing densities in (5) be Student's t with ν_j degrees of freedom, location parameters $\mu_{j,t|t-1}$ as in (6) and scale equal to φ_j,

$$p_j(y_t;\mu_{j,t|t-1},\nu_j,\varphi_j) = \frac{\Gamma\left(\frac{\nu_j+1}{2}\right)}{\Gamma\left(\frac{\nu_j}{2}\right)\sqrt{\pi\nu_j\varphi_j^2}}\left(1+\frac{(y_t-\mu_{j,t|t-1})^2}{\nu_j\varphi_j^2}\right)^{-\frac{\nu_j+1}{2}}. \tag{10}$$

One has

$$s_j(y_t,\mu_{j,t|t-1}) = \frac{\nu_j+1}{\nu_j\varphi_j^2}u_{j,t} \tag{11}$$

where

$$u_{j,t} = \frac{\upsilon_{j,t}}{1+\upsilon_{j,t}^2/\left(\nu_j\varphi_j^2\right)} \tag{12}$$

and

$$\upsilon_{j,t} = y_t - \mu_{j,t|t-1} \tag{13}$$

is the j-th one-step-ahead prediction error.

Up to a constant factor, $s_j(y_t,\mu_{j,t|t-1})$ and $u_{j,t}$ in Eqs. (11) and (12) are informationally equivalent, thus the filter we shall consider in the following will replace $s_j(y_t,\mu_{j,t|t-1})$ by $u_{j,t}$ in Eqs. (6) and (9). With this notation, when $\nu_j\to\infty$ for all j, then $u_{j,t}=\upsilon_{j,t}$ for all j and the model will become the mixture of Gaussian densities discussed in the previous section and we shall recover the estimator obtained by Shumway and Stoffer (1991), and, for $J=1$, the innovation form of the Kalman filter in the steady state.

If $J=1$ and $\nu<\infty$, then the robust filter developed by Harvey and Luati (2014) is recovered. As discussed by Harvey and Luati (2014), robustness of the filter comes from the redescending shape of the score $u_{j,t}$ in (12), that is a bounded transformation of the linear innovation error $\upsilon_{j,t}$ in (13). The redescending shape of the score implies that the influence of extreme observations on the filtered signal is reduced. This is coherent with assuming a heavy-tailed observation density and requiring that the filter is robust to extreme observations, in the sense of Calvet et al. (2015).

4. Probabilistic properties

Let us consider the process $\{y_t\}_{t\in\mathbb{Z}}$ defined by equations (5)–(8), with Student's t mixture components, i.e. the specification of Section 3.2. The random variable y_t admits the following stochastic representation

$$y_t = \sum_{j=1}^{J} z_{j,t}\left(\mu_{j,t|t-1}+\varphi_j\epsilon_{j,t}\right), \tag{14}$$

where $\{z_{j,t}\}_{t\in\mathbb{Z}}$ are time-independent random variables equal to one with $\mathbb{P}(z_{j,t}=1)=\alpha_j$ and zero otherwise, and such that $\sum_{j=1}^J z_{j,t}=1$, while $\{\epsilon_{j,t}\}_{t\in\mathbb{Z}}$ are iid Student's t distributed random variables with zero mean and unit scale, independent of $\{z_{j,t}\}_{t\in\mathbb{Z}}$. As $\mu_{t|t-1}$ is a function of y_{t-1} (and its past values), equation (14) is a non-linear stochastic difference equation in y_t. In the following, for a generic random variable $w_{i,t}$, we define $\tilde{w}_{i,t}^2 = w_{i,t}^2/(\nu_i\varphi_i^2)$, for $i=1,\dots,J$, and with $\mathcal{M}=\otimes_{j=1}^J\mathcal{M}_j$, where $\mathcal{M}_j=\left[\omega_j-\varpi_j,\omega_j+\varpi_j\right]$, with $\varpi_j=\varphi_j^2\sqrt{\nu_j}/2$, and $\otimes$ denotes the usual Cartesian product. The next theorem gives sufficient conditions for the existence of a stationary solution of Eq. (14).

Theorem 4.1. (Strong stationarity) *Let us consider the model specified by Eqs. (5)–(8) with densities as in (10). Under (14), for fixed $\boldsymbol{\mu}=(\mu_1,\dots,\mu_J)'\in\mathcal{M}\subset\mathbb{R}^J$, y_t is a function of $z_{j,t}$ and $\epsilon_{j,t}$. If*

$$\sum_{l=1}^{J}\alpha_l\mathbb{E}\left[\ln\sup_{\boldsymbol{\mu}\in\mathcal{M}}\max_j\sum_{i=1}^{J}\left|\phi_j e_{i,j}+\kappa_i g_{ij,t}(\boldsymbol{\mu})\right|\,|z_{l,t}=1\right]<0 \tag{15}$$

with $e_{i,j}=1$ if $i=j$ and 0 otherwise, where
$g_{ij,t}(\boldsymbol{\mu})=\xi_{i,t}(\boldsymbol{\mu})u_{i,t}(\boldsymbol{\mu})[u_{i,t}(\boldsymbol{\mu})(e_{i,j}-z_{j,t})n_i-\sum_{k=1}^J\xi_{k,t}(\boldsymbol{\mu})u_{k,t}(\boldsymbol{\mu})(e_{k,j}-z_{k,t})n_k]+\xi_{i,t}(\boldsymbol{\mu})\dot{u}_{i,t}(\boldsymbol{\mu})(e_{i,j}-z_{j,t})n_i$,
with $\dot{u}_{i,t}(\boldsymbol{\mu})=2\tilde{u}_{i,t}^2(\boldsymbol{\mu})-1/(1+\tilde{\upsilon}_{i,t}^2(\boldsymbol{\mu}))$, and $n_i=(\nu_i+1)/(\nu_i\varphi_i)$,
then a stationary and ergodic solution of Eq. (14) exists and is unique.

The notation $\xi_{i,t}(\boldsymbol{\mu})$, $v_{i,t}(\boldsymbol{\mu})$, $u_{i,t}(\boldsymbol{\mu})$ emphasises that $g_{ij,t}(\boldsymbol{\mu})$ in Theorem 4.1 depends on the variable $\boldsymbol{\mu}$ through the weights, the innovation errors, and the scores, respectively. Note, however, that, as functions of y_t, they depend on the stochastic processes ϵ_t and $z_{j,t}$ as well, i.e., for example, $v_{j,t}(\boldsymbol{\mu}) = \sum_{i=1}^{J} z_{i,t}\left(\mu_i + \varphi_i \epsilon_{i,t}\right) - \mu_j$.

The proof is based on Theorem 3.1 in Bougerol (1993) and consists in showing that the random function that map the J dimensional vector of $\mu_{j,t|t-1}$ into the vector of $\mu_{j,t+1|t}$, evaluated at $\boldsymbol{\mu}$, forms a stationary and ergodic sequence of Lipschitz maps, with Lipschitz coefficient satisfying a contraction condition that follows by (15). The latter is quite abstract, so sufficient conditions are given below, typically more restrictive than actually required. An alternative strategy consists in constructing a feasible invertibility region along the lines traced in Blasques et al. (2018). In the current setting, this approach is complicated by the large number of parameters, already in the case $J = 2$. We may observe, however, that the amplitude of the region decreases as ν_j increases and that, in principle, $|\phi_j| > 1$ for some j is admitted.

A sufficient non-stochastic condition implying the high level condition (15) is

$$\sum_{l=1}^{J} \alpha_l \ln\left(\max_j [|\phi_j| + \sum_{i=1}^{J} |\kappa_i| m_{i,j,l}]\right) < 0$$

where $m_{i,j,l} = \varphi_i\sqrt{\nu_i}(n_i\varphi_i\sqrt{\nu_i}r_{i,j,l} + \sum_{k=1}^{J} n_k\varphi_k\sqrt{\nu_k}r_{k,j,l})/4 + n_i r_{i,j,l}/\varphi_i$ with n_i as in Theorem 4.1 and $r_{i,j,l} = |\mathbb{1}(i=j) - \mathbb{1}(i=l)|$. This follows by iterating expectations, by using Jensen's and triangular inequalities and by observing that $\mathbb{E}[\sup_{\boldsymbol{\mu}\in\mathcal{M}} |g_{i,j,t}(\boldsymbol{\mu})| \, |z_{tl} = 1] \le m_{i,j,l}$. If $J = 2$ a sharper bound is obtained, as in this case $\mathbb{E}[|g_{i,j,t}(\boldsymbol{\mu})||z_{tl} = 1] = 0$ for $i, j = 1, 2$ implying that $m_{1,1,1} = m_{1,2,1} = m_{2,1,2} = m_{2,2,2} = 0$.

Weak stationarity of the process is established in the following theorem.

Theorem 4.2. (Weak stationarity) *If* $\min_{j=1,\dots,J}\{\nu_j\} > 1$, *and* $\max_{j=1,\dots,J} |\phi_j| < 1$, *then*

$$\mathbb{E}[y_t] = \sum_{j=1}^{J} \alpha_j \omega_j.$$

If, in addition, $\min_{j=1,\dots,J}\{\nu_j\} > 2$ *and Theorem 4.1 holds, then,*

$$\mathbb{E}[y_t^2] = \sum_{j=1}^{J} \alpha_j \left(\omega_j + \frac{\kappa_j^2 \sigma_{s_j}^2}{1-\phi_j^2} + \varphi_j^2 \mathbb{E}[\epsilon_{j,t}^2]\right)$$

where $\sigma_{s_j}^2 = \mathbb{E}[(\xi_{j,t}u_{j,t})^2] < \infty$ *and* $\{y_t\}_{t\in\mathbb{Z}}$ *is weakly stationary.*

While the first moment is available in closed form, the second moment depends on the quantity $\sigma_{s_j}^2$, that is finite as $u_{j,t}$ is bounded, and constant under strong stationarity.

Finally, under the same assumptions as in Theorem 4.2, the h-step ahead conditional expectations $\mathbb{E}[y_{T+h}|\mathcal{F}_T]$ and $\mathbb{E}[y_{T+h}^2|\mathcal{F}_T]$ are obtained by repeated applications of the law of iterated expectations,

$$\mathbb{E}[y_{T+h}|\mathcal{F}_T] = \sum_{j=1}^{J} \alpha_j \left(\sum_{l=2}^{h} \phi_j^{l-2}\right)\omega_j(1-\phi_j) + \sum_{j=1}^{J} \alpha_j \phi_j^{h-1} \mu_{j,T+1|T}$$

and

$$\begin{aligned}\mathbb{E}[y_{T+h}^2|\mathcal{F}_T] = &\sum_{j=1}^{J} \gamma_j \left[\sum_{l=2}^{h} \left(\phi_j^2\right)^{l-2}\right]\omega_j(1-\phi_j) + \sum_{j=1}^{J} \alpha_j \left[\sum_{l=2}^{h} \left(\phi_j^2\right)^{l-2}\right]\kappa_j^2\sigma_{s_j}^2 \\ &+ \sum_{j=1}^{J} \alpha_j \left(\phi_j^2\right)^{h-1} \mu_{j,T+1|T}^2 + 2\sum_{j=1}^{J} \gamma_j \phi_j \left[\sum_{l=3}^{h} \phi_j^{l-3}\frac{\gamma_j}{\alpha_j} + \phi_j^{h-2}\mu_{j,T+1|T}\right] \\ &+ 2\sum_{j=1}^{J} \gamma_j \left(\phi_j^2\right)^{h-2} \phi_j \mu_{j,T+1|T} + \sum_{j=1}^{J} \alpha_j \varphi_j^2 \mathbb{E}\left[\epsilon_{j,t}^2\right],\end{aligned}$$

where γ_j and $\sigma_{s_j}^2$ are as in Theorem 4.2, $\mathbb{E}[\epsilon_{j,t}^2] = \nu_j/(\nu_j - 2)$, and $\sum_{l=i}^{h} \cdot = 0$ for $h < i$.

Remark 1. When $\nu_j \to \infty$ for $j = 1, \dots, J$, the Student's t dynamic adaptive mixture model converges to the Gaussian model discussed in Section 3.1, and the score $u_{j,t}$ coincides with the prediction error $v_{j,t}$. It follows that, of the properties derived in this section, only the first moment remains valid in the Gaussian case. The second moment requires $\mathbb{E}[\xi_{j,t}v_{j,t}^2] < \infty$ to exist, and strong stationarity requires bounds on the Lipschitz coefficient associated with the recursion, see Sections S1.1 and S1.2. With the methods developed and used in this paper, both conditions are not straightforward to prove. Our conjecture, also validated by simulations, is that the growth of the score is controlled by the mixture weights, which make the Gaussian mixture work in practice. However, due to the high non-linearity and complexity of the mixing weights, bounds on their product with the score are non-trivial to derive.

5. Maximum likelihood estimation

Let $\boldsymbol{\theta} = (\alpha_j, \boldsymbol{\theta}_j')' \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta} \subset \mathbb{R}^{6J}$, with $\boldsymbol{\theta}_j = (\omega_j, \kappa_j, \phi_j, \varphi_j, \nu_j)'$, $j = 1, 2, \dots, J$. Denote as $\{\hat{\mu}_{t|t-1}(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ the sequence of dynamic parameters computed according to $\boldsymbol{\theta}$ and initialised at time $t = 1$ at the arbitrary values $\hat{\mu}_{j,1|0}(\boldsymbol{\theta}) = \hat{\mu}_{j,1}$, according to (6), for $j = 1, \dots, J$. The empirical log-likelihood function is

$$\hat{\mathcal{L}}_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \hat{\ell}_t(\boldsymbol{\theta}), \tag{16}$$

where $\hat{\ell}_t(\boldsymbol{\theta}) = \ln \sum_{j=1}^{J} \alpha_j p_j(y_t|\hat{\mu}_{j,t|t-1}(\boldsymbol{\theta}), \varphi_j, \nu_j)$, with the associated MLE

$$\hat{\boldsymbol{\theta}}_T = \text{argmax}_{\theta\in\boldsymbol{\Theta}} \hat{\mathcal{L}}_T(\boldsymbol{\theta}). \tag{17}$$

Let $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ be the true vector of parameters according to which the observed process is assumed to be generated. The following Theorem establishes the strong consistency of $\hat{\boldsymbol{\theta}}_T$ for $\boldsymbol{\theta}_0$.

Theorem 5.1. *Let the assumptions of Theorem 4.1 hold at $\boldsymbol{\theta}_0$ and suppose that $\alpha_1 > \alpha_2 > \cdots > \alpha_J > 0$ with $\sum_{j=1}^{J} \alpha_j = 1$, $\boldsymbol{\theta}_k = \boldsymbol{\theta}_j$ if and only if $k = j$, and $\kappa_j \neq 0$ for $j = 1, \dots, J$. If*

$$\mathbb{E}\left[\ln \sup_{\boldsymbol{\mu}\in\mathcal{M}} \max_{j} \sum_{i=1}^{J} |\phi_j e_{i,j} + \kappa_i \xi_{i,t}(\boldsymbol{\mu}) h_{i,j,t}(\boldsymbol{\mu})|\right] < 0 \tag{18}$$

for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta} \subset \mathbb{R}^{6J}$ is a compact parameter space, and

$$h_{i,j,t}(\boldsymbol{\mu}) = (e_{i,j} - \xi_{i,t}(\boldsymbol{\mu}))n_i u_{i,t}^2(\boldsymbol{\mu}) + \left[2\, u_{i,t}^2(\boldsymbol{\mu})/(\nu_j \varphi_j^2) - \left(1 + v_{i,t}^2(\boldsymbol{\mu})/(\nu_j \varphi_j^2)\right)^{-1}\right] e_{i,j},$$

with n_i as in Theorem 4.1,

then $\hat{\theta}_T \xrightarrow{a.s.} \theta_0$, as $T \to \infty$.

Note that, differently from Theorem 4.1, in this setting $\xi_{j,t}(\boldsymbol{\mu}), v_{j,t}(\boldsymbol{\mu}), u_{j,t}(\boldsymbol{\mu})$ (and their transforms) are treated as functions of y_t, i.e., for example $v_{j,t}(\boldsymbol{\mu}) = y_t - \mu_j$.

The proof, in Supplementary S1.3, is a classical proof of consistency in non-linear dynamic models. The identification assumption on the mixture weights is not restrictive, as it only avoids permutation of the J mixture components, see Aitkin and Rubin (1985). The condition $\kappa_j \neq 0$ ensures identification of ϕ_j for $j = 1, \dots, J$. The contraction condition (18) can be checked via simulations and is required for the almost sure convergence of the empirical log-likelihood function (16) to the population log-likelihood, $\mathcal{L}(\theta) = \mathbb{E}[\ell_t(\boldsymbol{\theta})]$, where $\ell_t(\boldsymbol{\theta})$ is defined as $\hat{\ell}_t(\boldsymbol{\theta})$ with $\hat{\mu}_{j,t|t-1}(\boldsymbol{\theta})$ replaced by $\mu_{j,t|t-1}(\boldsymbol{\theta})$, for all $j = 1, \dots, J$.

The next theorem, based on Theorem 6.2 of White (1994) and proved in Section S1.4, establishes the asymptotic normality of $\hat{\boldsymbol{\theta}}_T$.

Theorem 5.2. *Let the assumptions of Theorem 4.1 hold and suppose that: i) $\theta_0 \in int(\boldsymbol{\Theta})$, ii) $\mathbb{E}[\partial\ell_t(\theta_0)/\partial\theta\partial\theta^\top]$ is invertible, and iii) it holds that*

$$\rho_j := |\phi_j| + \max_i |\kappa_i| \bar{h}_{i,j} < 1, \tag{19}$$

for all $j = 1, \dots, J$ where

$$\bar{h}_{i,j} = e_{i,j}(2n_i\nu_i + 1)/8 + (1 - e_{i,j})n_i\nu_i/4,$$

then $\sqrt{T}(\hat{\theta}_T - \theta_0) \Rightarrow \mathcal{N}(\mathbf{0}, \mathcal{I}^{-1}(\theta_0))$, where $\mathcal{I}(\theta_0)$ is the Fisher Information matrix evaluated at the true parameter vector θ_0.

We note that the sufficient non-stochastic condition in (19) also implies (18) (details are reported in the supplementary materials).

6. Simulation results

6.1. Finite sample properties of the MLE

Computing the MLE in practice can be challenging due to the possible presence of local maxima in the empirical log-likelihood function $\hat{\mathcal{L}}_T(\theta)$. An initial guess for the optimiser can be constructed following the discussion in Catania (2019). Section S6.1 in the Supplementary materials reports the details of a Monte Carlo study to assess the finite sample properties of the MLE for varying sizes $T = 500$, $T = 1000$, and $T = 5000$. Overall, results indicate that a sample size of $T = 1000$ is appropriate to estimate a specification with two regimes, i.e. $J = 2$. Results also show that when $T = 500$, sizable biases affect the estimate of the degrees of freedom parameters. This result is largely expected and suggests that, when the sample size is small, it is reasonable to impose the constraint $\nu_j = \nu$ for all j and estimate a single parameter ν. Alternatively, to design a robust filter, ν can be set to some fixed value, e.g. $\nu = 6$.

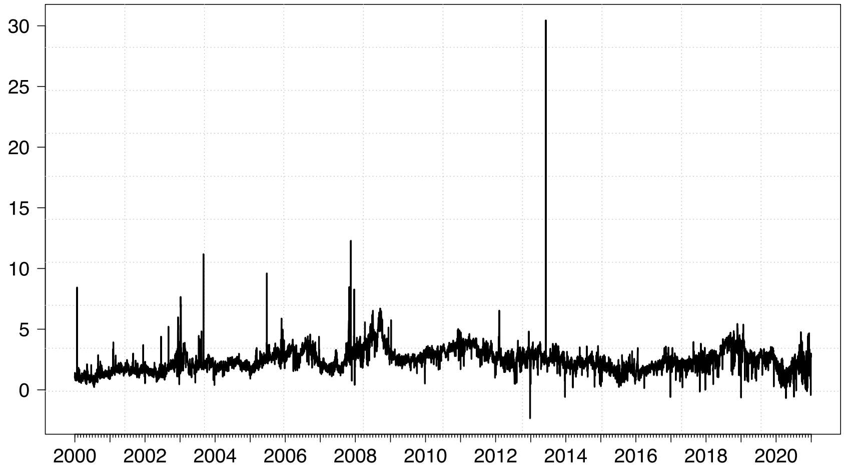

Fig. 1. Average daily NordPool energy spot price from Energinet.

6.2. *Filtering in a signal plus heavy-tailed noise and parameter uncertainty*

It is interesting to assess how well a Student's t DAMM performs when applied to data generated from the state space model in (1) with fat-tailed measurement noise. The simulation analysis reported in Section S6.2 of the supplementary materials shows that the Student's t DAMM performs remarkably well when noise follows a Student's t distribution with ν degrees of freedom, especially when $\nu = 1$ and $\nu = 2$. For comparison, we consider estimates obtained by the Gaussian DAMM, the Kalman Filter, and a particle filter at the true values of the parameters. Results show that the Gaussian DAMM performs well for $\nu \geq 2$, whereas the KF, as expected, does not deliver sensible estimates when $\nu \leq 2$. For larger ν results from the Student's t and Gaussian DAMMs closely align.

7. Empirical illustration

We analyse the daily average of the hourly "day ahead NordPool energy spot price" in Danish Krone per megawatt-hour (MWh) over the period 31 December 1999 to 31 December 2020, for a total of $T = 7672$ observations. Data can be freely downloaded from Energinet,[1] which is the Danish national transmission system operator for electricity and natural gas. The dataset is provided at the hourly frequency: we aggregate the series from the DK1 "price area" to daily frequency using the UTC time zone and taking the mean of the hourly prices. Finally, we standardise the data to have unit variance and remove the seasonality in it using weekly dummy variables.

The graph of the time series is reported in Fig. 1 and shows a number of outlying observations associated with temporary imbalances in the energy market. The Augmented Dickey-Fuller test statistics (Dickey and Fuller, 1979) is equal to −6.10, thus leading us to reject the null hypothesis of a unit root. The empirical skewness and kurtosis coefficients are 3.64 and 86.54, respectively, indicating strong departure from a Gaussian data-generating process.

Together with the dynamic adaptive mixture model (DAMM) with Student's t components discussed in Section 3.2 (TDAMM), we consider the Gaussian mixture case (GDAMM), obtained by letting $\nu_j \to \infty$ for all j, which coincides with the filter discussed in Section 3.1 and implemented by Harrison and Stevens (1976) and Shumway and Stoffer (1991), among others. We also include the MAR model of Wong and Li (2000) with Gaussian mixture components (GMAR), and with Student's t components (TMAR), see Wong et al. (2009). Specifically, in the GMAR(K) model with J components, $p(y_t|\mathcal{F}_{t-1}, S_t = j) \sim N(\mu_{j,t|t-1}, \varphi_j^2)$, where $\mu_{j,t|t-1} = \phi_{j,0} + \sum_{k=1}^{K} \phi_{j,k} y_{t-i}$ for $j = 1, \ldots, J$ such that $\mu_{t|t-1} = \mathbb{E}[y_{t+1}|\mathcal{F}_t] = \tilde{\phi}_0 + \sum_{k=1}^{K} \tilde{\phi}_k y_{t-k}$, where $\tilde{\phi}_k = \sum_{j=1}^{J} \alpha_j \phi_{j,k}$ for $k = 1, \ldots, K$, and the innovation form is $y_t = \mu_{t|t-1} + v_t$ with $\mu_{t+1|t} = \tilde{\phi}_0 + \sum_{k=1}^{K} \tilde{\phi}_k \mu_{t-k|t-k-1} + \sum_{k=1}^{K} \tilde{\phi}_k v_{t-k}$. The TMAR($K$) is analogous but with $p(y_t|\mathcal{F}_{t-1}, S_t = j) \sim t_{\nu_j}(\mu_{j,t|t-1}, \varphi_j)$. We also consider the dynamic conditional score (DCS) model of Harvey and Luati (2014), which coincides with TDAMM with $J = 1$, and the AR(1)-signal plus noise model $y_t = \omega + \mu_t + \varepsilon_t$, with $\mu_{t+1} = \phi\mu_t + \eta_t$, where ε_t and η_t are mutually independent and independent over time, with $\mathbb{E}[\varepsilon_t] = \mathbb{E}[\eta_t] = 0$, and $\mathbb{E}[\varepsilon_t^2] = \sigma^2$, $\mathbb{E}[\eta_t^2] = \tau^2$.

7.1. *Estimation*

The AR(1)-signal plus noise model is estimated via the quasi maximum likelihood estimator using the Kalman filter (KF), which gives $\hat{\omega} = 2.401(0.003)$, $\hat{\sigma} = 0.512(0.812)$, $\hat{\phi} = 0.970(0.285)$, and $\hat{\tau} = 0.208(0.574)$. All the other models are estimated via the MLE. For the GMAR and TMAR models, we compute the MLE using the Expectation-Maximization algorithm of Dempster et al. (1977), thereby

[1] `https://www.energidataservice.dk/tso-electricity/NordpoolMarket`

Table 1
GDAMM, TDAMM, and DCS estimated parameters (standard errors).

	GDAMM				TDAMM		DCS
	$j=1$	$j=2$	$j=3$	$j=4$	$j=1$	$j=2$	–
ω_j	2.891 (0.094)	2.549 (0.034)	2.514 (0.086)	3.450 (1.913)	2.452 (0.041)	2.436 (0.018)	2.465 (0.057)
ϕ_j	0.993 (0.001)	0.959 (0.002)	0.956 (0.006)	0.983 (0.032)	0.983 (0.000)	0.930 (0.001)	0.978 (0.002)
κ_j	0.215 (0.006)	0.516 (0.016)	1.191 (0.068)	9.223 (8.926)	0.187 (0.003)	1.476 (0.008)	0.238 (0.005)
φ_j	0.285 (0.003)	0.396 (0.004)	0.880 (0.023)	4.387 (0.400)	0.306 (0.002)	0.484 (0.004)	0.318 (0.004)
α_j	0.561 (0.010)	0.334 (0.010)	0.096 (0.005)	0.010 (0.002)	0.780 (0.005)	0.220 (0.005)	–
ν_j	–	–	–	–	16.695 (1.179)	3.268 (0.047)	3.513 (0.097)

Table 2
GMAR and TMAR estimated parameters (standard errors).

	GMAR									TMAR				
	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$	$j=9$	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
ϕ_{0j}	−0.268 (0.038)	−0.016 (0.033)	−0.005 (0.050)	0.118 (0.030)	0.157 (0.032)	0.127 (0.152)	0.306 (0.076)	0.383 (0.244)	1.420 (0.870)	−0.035 (0.031)	0.101 (0.029)	0.135 (0.028)	0.283 (0.076)	−0.052 (0.042)
ϕ_{1j}	0.357 (0.023)	0.424 (0.015)	0.090 (0.025)	1.280 (0.018)	0.853 (0.016)	1.800 (0.077)	−0.013 (0.032)	0.128 (0.241)	15.620 (0.994)	0.347 (0.013)	0.747 (0.018)	1.246 (0.017)	−0.006 (0.024)	0.067 (0.017)
ϕ_{2j}	0.168 (0.033)	0.009 (0.015)	−0.033 (0.024)	−0.107 (0.021)	−0.100 (0.022)	−0.595 (0.066)	0.354 (0.046)	0.178 (0.157)	−17.210 (1.708)	0.008 (0.016)	−0.038 (0.020)	−0.107 (0.019)	0.295 (0.035)	0.015 (0.021)
ϕ_{3j}	0.223 (0.030)	−0.014 (0.022)	0.005 (0.028)	−0.122 (0.016)	−0.007 (0.020)	0.012 (0.077)	0.110 (0.057)	0.765 (0.215)	13.602 (2.923)	−0.020 (0.020)	−0.009 (0.016)	−0.110 (0.016)	0.663 (0.052)	0.020 (0.028)
ϕ_{4j}	−0.015 (0.015)	0.006 (0.019)	0.107 (0.036)	0.007 (0.018)	0.076 (0.020)	−0.240 (0.109)	0.640 (0.061)	−0.486 (0.256)	−2.127 (4.062)	0.007 (0.021)	0.123 (0.018)	0.006 (0.016)	0.013 (0.052)	0.139 (0.034)
ϕ_{5j}	0.451 (0.026)	0.003 (0.020)	−0.039 (0.033)	−0.097 (0.019)	0.025 (0.015)	0.141 (0.109)	−0.234 (0.052)	0.037 (0.222)	−0.944 (3.952)	0.003 (0.022)	0.064 (0.018)	−0.066 (0.017)	−0.002 (0.043)	−0.038 (0.030)
ϕ_{6j}	−0.008 (0.019)	−0.111 (0.025)	0.830 (0.032)	0.013 (0.019)	0.101 (0.018)	0.012 (0.099)	−0.007 (0.043)	0.009 (0.219)	−6.689 (4.743)	−0.052 (0.021)	0.080 (0.017)	−0.012 (0.010)	−0.080 (0.054)	0.782 (0.031)
ϕ_{7j}	−0.097 (0.022)	0.705 (0.018)	0.032 (0.029)	−0.002 (0.016)	−0.010 (0.015)	−0.244 (0.087)	0.019 (0.042)	0.226 (0.124)	0.066 (5.040)	0.733 (0.018)	−0.015 (0.013)	0.001 (0.012)	−0.019 (0.049)	0.033 (0.026)
φ_j	0.120 (0.013)	0.165 (0.010)	0.220 (0.014)	0.125 (0.010)	0.222 (0.010)	0.476 (0.041)	0.350 (0.023)	0.924 (0.058)	0.757 (0.350)	0.158 (0.012)	0.233 (0.011)	0.160 (0.012)	0.366 (0.031)	0.185 (0.017)
α_j	0.084 (0.011)	0.156 (0.013)	0.123 (0.011)	0.122 (0.013)	0.276 (0.020)	0.063 (0.009)	0.129 (0.014)	0.045 (0.009)	0.002 (0.001)	0.153 (0.013)	0.327 (0.024)	0.238 (0.018)	0.134 (0.013)	0.148 (0.013)
ν_j	–	–	–	–	–	–	–	–	–	61.911 (389.075)	15.556 (15.616)	1.529 (0.139)	5.287 (2.722)	2.668 (0.899)

avoiding direct numerical maximization of the likelihood function.[2] For the DAMM specifications, the likelihood is maximized numerically using the differential evolution algorithm implemented in the DEoptim package of R (Mullen et al., 2011). Additional computational details are provided in Section S7.1 of the supplementary materials.

We estimate the GDAMM and TDAMM for all $J \in (1, \ldots, 5)$ and the GMAR and TMAR for all $J \in (1, \ldots, 10)$ and $K \in (0, \ldots, 15)$ and select the best specification using the BIC.[3] The selected models are: i) GDAMM with $J = 4$, ii) TDAMM with $J = 2$, iii) GMAR with $J = 9$ and $K = 7$, and iv) TMAR with $J = 5$ and $K = 7$.

Estimates for the GDAMM and TDAMM are reported in Table 1. If the autoregressive parameters ϕ_j of GDAMM and TDAMM are stable across all models and specifications, the score coefficients κ_j and the mixture weights α_j are the parameters that change the most across mixture components. In the TDAMM, the tails of the mixing densities are substantially different, as indicated by the dissimilar estimates of degrees of freedom parameters ν_j. Indeed, TDAMM allows for the full flexibility to model mixtures of densities with heavy tails ($\hat{\nu}_2 = 3.268$) and closer to Gaussian tails ($\hat{\nu}_1 = 16.695$), in this example with associated probabilities $\alpha_2 = 0.220$ and $\alpha_1 = 0.780$. The scale parameter remains nearly the same in the $J = 2$ TDAMM components while varies substantially in the Gaussian case, where $J = 4$ mixture components with different variances are selected for the model to recover a signal that is similar to the signal recovered by the TDAMM (see Fig. 2). DCS or TDAMM with $J = 1$ is a good compromise as the estimates lie in between those of the two components of TDAMM, except for the degrees of freedom parameter, associated with the heaviest tailed density. We assess the validity of the invertibility condition in (18) based on the estimated parameters via Monte Carlo simulation. For the TDAMM model, however, the condition is not satisfied due to the large estimated values of ϕ_j – the computed value is 0.55. Nevertheless, it is well documented in the literature that such sufficient conditions are most of the time too restrictive to fulfil in practice, especially in models with multiple dynamic parameters.

Estimates for GMAR and TMAR are reported in Table 2. Many autoregressive coefficients are not statistically different from zero and, overall, estimates are heterogeneous between Gaussian and Student's t cases and across mixtures' components, due to the large number of lags selected by the BIC. The high order of MAR models indicated by the information criterion is required to face the complexity of the dataset.

[2] For the TMAR model, the M-step requires solving for ν_j, $j = 1, \ldots, J$. We implement this step using a simple bisection root-finding method.

[3] We impose the constraint $\min_j \alpha_j > 0.01$ to avoid estimating some mixture components associated with very low probabilities. All BIC and log-likelihood values are reported in Section S7.2 of the supplementary materials.

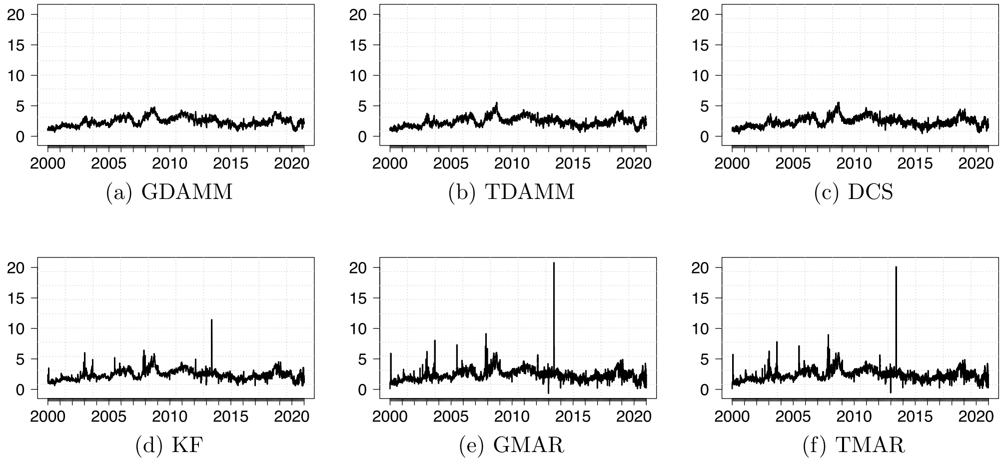

Fig. 2. Signal extraction for the NordPool spot price over the period 1 January 2000 to 31 December 2020.

Table 3
Mean squared forecast error computed over the out of sample period from 2 July 2010 to 31 December 2020 for $h = 1, \dots, 7$ steps ahead. Results are reported relative to the KF which acts as a benchmark. Grey cells emphasise specifications that belongs to the superior set of models according to the model confidence set procedure at the 75 % confidence level.

$h =$	1	2	3	4	5	6	7
GDAMM	**0.989**	0.923	0.910	0.910	**0.911**	0.910	0.909
TDAMM	**0.975**	**0.914**	**0.902**	**0.903**	**0.904**	**0.903**	**0.902**
GMAR	**1.012**	0.992	0.982	0.991	0.999	1.012	1.019
TMAR	**1.023**	0.997	0.989	0.999	1.006	1.024	1.032
KF	**1.000**	1.000	1.000	1.000	1.000	1.000	1.000
DCS	**0.986**	0.923	0.910	**0.910**	**0.908**	**0.908**	**0.909**

7.2. Signal extraction

Fig. 2 reports the filtered signal over the full sample, $\hat{\mu}_{t|t-1}(\hat{\boldsymbol{\theta}}_T)$ for $t = 1, \dots, T$, for all specifications. We note that GMAR, TMAR, and, to a slightly lesser extent, the KF recover a noisy signal, where all the spikes are present, which essentially mimics the original series. This is not surprising since, for these specifications, $\hat{\mu}_{t|t-1}(\hat{\boldsymbol{\theta}}_T)$ is a linear function of past observations. Indeed, even if in the MAR(1) model there are J conditional means and the conditional distribution of y_t given $\mathcal{F}_{t-1}$ is mixed Gaussian, the *filter* for its conditional mean is still linear, and coincides with that of an AR(1) model, i.e. the steady state of the KF for a linear Gaussian signal-plus-noise model with variance of the measurement error equal to zero. Results for GDAMM, TDAMM, and DCS are remarkably good. Signals estimated by GDAMM and TDAMM are very similar, empirically confirming the conjecture in Remark 1, at the cost for the Gaussian mixture of having $J = 4$ components compared to the parsimonious representation of the Student's t mixture of $J = 2$ components.

7.3. Out of sample analysis

The different specifications are compared in an out-of-sample analysis. We divide the series into two equal parts composed of $T = 3836$ observations each. The values of J and K are re-estimated over the in-sample period. We observe that TDAMM and GDAMM select the same J over the full sample and in-sample period, while orders selection for GMAR and TMAR changes to the pairs $J = 6$, $K = 10$ and $J = 5$, $K = 8$, respectively. We perform a rolling forecast analysis using an expanding window and parameters estimated over the in-sample period. We compute h-step-ahead point predictions with $h = 1, \dots, 7$, i.e. up to a week. Note that predictions are available in closed form for all specifications, see Section 4.

Mean squared forecast errors (MSFE) computed over the forecast period are reported in Table 3, relatively to the KF specification that acts as a benchmark. Values lower than one indicate outperformance of the benchmark and vice versa. Grey cells emphasise those specifications that belong to the superior set of models according to the model confidence set procedure of Hansen et al. (2011) at the 75 % confidence level. Results indicate that the MSFE of the GDAMM and TDAMM is generally smaller than the one obtained by the KF and the other specifications. TDAMM is often the top performer and provides the lowest MSFE at each horizon. Furthermore, TDAMM is the sole specification that belongs to the superior set of models for all forecast horizons.

In summary, the analysis confirms that a dynamic mixture of Student's t densities, some of which may be approximating a Gaussian density, is a flexible, parsimonious and analytically tractable specification for robust estimation and prediction of datasets characterised by complex features.

8. Conclusions

The contribution of this paper to the literature on dynamic models for time-varying parameters of predictive densities is twofold. First, we revise signal-plus-noise models with parameter uncertainty and relate them to the class of dynamic adaptive mixture models, implying the same filtering recursions traditionally obtained, in the former case, based on collapsing steps. This is a progress toward reconciling parameter-driven models and observation-driven models, or nonlinear unobserved component models and approximate filters. Second, we specify a class of robust dynamic mixture models whose stochastic and asymptotic properties can be made explicit. The methods developed in the paper are based on the properties of the score of the Student's t log-likelihood, a bounded function of the data, and can be extended straightforwardly to other similar cases, or to mixtures of different distributions, as discussed in Harvey (2013).

Unicted reference

Berkes et al. (2003), Billingsley (1961), Ghalanos and Theussl (2015), Gordon et al. (1993), Holzmann et al. (2006), Kalliovirta et al. (2015), Krengel (1985), Meitz et al. (2018), Price et al. (2006), Rao (1962), Storn and Price (1997), Straumann (2005), Straumann and Mikosch (2006), Ye (1987).

Funding

We are grateful to the Editor, Viktor Todorov, the Associate Editor, and two anonymous reviewers for their useful remarks that greatly improved the quality of the current version of the manuscript. We also thank the participants at the NBER-NFS Time Series Conference in Houston (2021), the 7th RCEA Time Series Workshop conference in Milan (2021), the Rome-Waseda Time Series Symposium in Rome (2022), the SNDE Annual Symposium in Padua (2024), the CFE-CMStat conference in London (2024), and seminar participants in VU Amsterdam, CREST Paris, Goethe University Frankfurt, GREATS Birmingham, and ESSEC Paris. Leopoldo Catania acknowledges the research support of the Project 2 grant of the Danish Council for Independent Research (IRFD), Social Sciences, number 3099-00089B, and of the Aarhus Universitets Forskningsfond (AUFF).

Supplementary material

Supplementary material associated with this article can be found in the online version at 10.1016/j.jeconom.2025.106155.

References

Ackerson, G., Fu, K., 1970. On state estimation in switching environments. IEEE Trans. Automat. Contr 15 (1), 10–17.
Aitkin, M., Rubin, D.B., 1985. Estimation and hypothesis testing in finite mixture models. J. R. Stat. Soci. Series B (Methodol.) 47 (1), 67–75.
Akashi, H., Kumamoto, H., 1977. Random sampling approach to state estimation in switching environments. Automatica 13 (4), 429–434.
Bar-Shalom, Y., Tse, E., 1975. Tracking in a cluttered environment with probabilistic data association. Automatica 11 (5), 451–460.
Berchtold, A., Raftery, A.E., 2002. The mixture transition distribution model for high-order markov chains and non-gaussian time series. pp. 328–356.
Berkes, I., Horváth, L., Kokoszka, P., 2003. Garch processes: structure and estimation. Bernoulli 9 (2), 201–227.
Billingsley, P., 1961. The lindeberg-levy theorem for martingales. Proc. Am. Math. Soci. 12 (5), 788–792.
Blasques, F., Brummelen, J.V., Koopman, S.J., Lucas, A., 2022. Maximum likelihood estimation for score-driven models. J. Econom. 227 (2), 325–346.
Blasques, F., Francq, C., Laurent, S., 2023. Quasi score-driven models. J. Econom. 234 (1), 251–275.
Blasques, F., Gorgi, P., Koopman, S.J., Wintenberger, O., 2018. Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. Electron. J. Stat. 12 (1), 1019–1052.
Blasques, F., Koopman, S.J., Lucas, A., 2015. Information-theoretic optimality of observation-driven time series models for continuous responses. Biometrika 102 (2), 325–343.
Bougerol, P., 1993. Kalman filtering with random coefficients and contractions. SIAM J. Control Optim. 31 (4), 942–959.
Buccheri, G., Bormetti, G., Corsi, F., Lillo, F., 2025. Beyond the Normal Approximation: Robust Filtering and Smoothing Via Perturbation Methods. Working paper.
Calvet, L.E., Czellar, V., Ronchetti, E., 2015. Robust filtering. J. Am. Stat. Assoc. 110 (512), 1591–1606.
Catania, L., 2019. Dynamic adaptive mixture models with an application to volatility and risk.
Chen, R., Liu, J.S., 2000. Mixture kalman filters. J. R. Stat. Soci. Series B (Stat. Methodo.) 62 (3), 493–508.
Cox, D.R., 1981. Statistical analysis of time series: some recent developments. Scand. J. Stat. 8 (2), 93–115.
Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. J. Appl. Econom. 28 (5), 777–795.
Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the em algorithm. J. R. Stat. Soci. Series B (Stat. Methodo.) 39 (1), 1–22.
Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. J. Am. Stat. Assoc. 74 (366), 427–431.
Durbin, J., Koopman, S.J., 2012. Time Series Analysis by State Space Methods. Oxford university press.
Francq, C., Zakoian, J.-M., 2023. Local asymptotic normality of general conditionally heteroskedastic and score-driven time-series models. Econ. Theory 39 (5), 1067–1092.
Frühwirth-Schnatter, S., 2001. Fully bayesian analysis of switching gaussian state space models. Ann. Inst. Stat. Math. 53 (1), 31–49.
Gasperoni, F., Luati, A., Paci, L., Innocenzo, E., 2023. Score-driven modeling of spatio-temporal data. J. Am. Stat. Assoc. 118 (542), 1066–1077.
Gerlach, R., Carter, C., Kohn, R., 2000. Efficient bayesian inference for dynamic mixture models. J. Am. Stat. Assoc. 95 (451), 819–828.
Ghalanos, A., Theussl, S., 2015. RSOLNP: General Non-Linear Optimization Using Augmented Lagrange Multiplier Method. R Package Version 1.
Godsill, S.J., 1997. Bayesian enhancement of speech and audio signals which can be modelled as arma processes. Int. Stat. Rev. 65 (1), 1–21.

Gordon, K., Smith, A., 1990. Modeling and monitoring biomedical time series. J. Am. Stat. Assoc. 85 (410), 328–337.
Gordon, N.J., Salmond, D.J., Smith, A.F., 1993. Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation. IEE Proceedings F (Radar And Signal Processing. Vol. 140.
Gorgi, P., 2020. Beta-negative binomial auto-regressions for modelling integer-valued time series with extreme observations. JJ. R. Stat. Soci. Series B 82 (5), 1325–1347.
Gorgi, P., Lauria, C.S.A., Luati, A., 2023. On the optimality of score-driven models.
Gray, S.F., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. J. Financ. Econ. 42 (1), 27–62.
Hamilton, J.D., 1989. Analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357–384.
Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79 (2), 453–497.
Hansen, P.R., Schmidtblaicher, M., 2021. A dynamic model of vaccine compliance: how fake news undermined the danish hpv vaccine program. J. Business Econ. Stat. 39 (1), 259–271.
Harrison, P.J., Stevens, C.F., 1976. Bayesian forecasting. JJ. R. Stat. Soci. Series B (Stat. Methodo.) 38 (3), 205–228.
Harvey, A., Luati, A., 2014. Filtering with heavy tails. J. Am. Stat. Assoc. 109 (507), 1112–1122.
Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge university press.
Harvey, A.C., 2013. Dynamic Models for Volatility and Heavy Tails: with Applications to Financial and Economic Time Series. Vol. 52. Cambridge University Press.
Heel, S. D.V., Dijk, D.V., Os, B.V., 2025. Stability and Performance Guarantees for Misspecified Multivariate Score-Driven Filters. Working paper.
Holý, V., Tomanová, P., 2022. Modeling price clustering in high-frequency prices. Quantitative Finance 22 (9), 1649–1663.
Holzmann, H., Munk, A., Gneiting, T., 2006. Identifiability of finite mixtures of elliptical distributions. Scand. J. Stat. 33 (4), 753–763.
Kalliovirta, L., Meitz, M., Saikkonen, P., 2015. A Gaussian mixture autoregressive model for univariate time series. J. Time Series Anal. 36 (2), 247–266.
Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. J. Basic Eng. 82 (1), 35–45.
Kalman, R.E., Bucy, R.S., 1961. New results in linear filtering and prediction theory. p. 109.
Kalpakis, K., Gada, D., Puttagunta, V., 2001. Distance measures for effective clustering of arima time-series. In: Proceedings 2001 IEEE International Conference on Data Mining. IEEE, pp. 273–280.
Kim, C.-J., 1994. Dynamic linear models with markov-switching. J. Econom. 60 (1–2), 1–22.
Kim, C.-J., Nelson, C.R., 1999. State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications. MIT Press Books.
Kitagawa, G., 1987. Non-Gaussian state-space modeling of nonstationary time series. J. Am. Stat. Assoc. 82 (400), 1032–1041.
Klassen, F., 2002. Improving garch volatility forecasts. Empir. Econ. 27 (2), 363–394.
Koopman, S.J., Lucas, A., Scharth, M., 2016. Predicting time-varying parameters with parameter-driven and observation-driven models. Rev. Econ. Stat. 98 (1), 97–110.
Krengel, U., 1985. Ergodic Theorems.
Lange, R.-J., Os, B.V., Dijk, D.V., 2025. Implicit Score-Driven Filters for Time-Varying Parameter Models. TI discussion paper.
Le, N.D., Martin, R.D., Raftery, A.E., 1996. Modeling flat stretches, bursts outliers in time series using mixture transition distribution models. J. Am. Stat. Assoc. 91 (436), 1504–1515.
Martin, R.D., Raftery, A.E., 1987. Comment: robustness, computation, and non-euclidean models. J. Am. Stat. Assoc. 82 (400), 1044–1050.
Masreliez, C., 1975. Approximate non-gaussian filtering with linear state and observation relations. IEEE Trans. Automat. Contr 20 (1), 107–110.
Meitz, M., Preve, D., Saikkonen, P., 2018. A Mixture Autoregressive Model Based on Student's t-Distribution.
Monache, D.D., Petrella, I., Venditti, F., 2021. Price dividend ratio and long-run stock returns: a score-driven state space model. pp. 1054–1065.
Mullen, K.M., Ardia, D., Gil, D.L., Windover, D., Cline, J., 2011. Deoptim: an r package for global optimization by differential evolution. J. Stat. Softw. 40, 1–26.
Price, K., Storn, R.M., Lampinen, J.A., 2006. Differential Evolution: a Practical Approach to Global Optimization. Springer Science & Business Media.
Raftery, A.E., 1985. A model for high-order markov chains. JJ. R. Stat. Soci. Series B (Stat. Methodo.) 47 (3), 528–539.
Rao, R.R., 1962. Relations between weak and uniform convergence of measures with applications. Ann. Math. Statist. 33 (2), 659–680.
Shephard, N., 1994. Partial non-Gaussian state space. Biometrika 81 (1), 115–131.
Shumway, R., Stoffer, D., 2017. In: Time series analysis and its applications. Springer, New York.
Shumway, R.H., Stoffer, D.S., 1991. Dynamic linear models with switching. J. Am. Stat. Assoc. 86 (415), 913. Addendum: JASA.
Storn, R., Price, K., 1997. Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces. J. Global Optim. 11 (4), 341–359.
Straumann, D., 2005. Estimation in Conditionally Heteroscedastic Time Series Models. Lecture Notes in Statistics. Berlin Heidelberg, Springer.
Straumann, D., Mikosch, T., 2006. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. Ann. Stat. 34 (5), 2449–2495.
West, M., Harrison, J., 1997. Bayesian Forecasting and Dynamic Models. Springer.
White, H., 1994. Estimation, Inference and Specification Analysis. Cambridge University Press.
Wong, C., Chan, W., Kam, P., 2009. A student t-mixture autoregressive model with applications to heavy-tailed financial data. Biometrika 96 (3), 751–760.
Wong, C.S., Li, W.K., 2000. On a mixture autoregressive model. J. R. Stat. Soci. Series B (Stat. Methodo.) 62 (1), 95–115.
Xiong, Y., Yeung, D.-Y., 2002. Mixtures of arma models for model-based time series clustering. In: 2002 IEEE International Conference on Data Mining. IEEE, pp. 717–720. Proceedings.
Ye, Y., 1987. Interior Algorithms for Linear, Quadratic, and Linearly Constrained Non-Linear Programming. Ph.d. thesis, department of ees, stanford university.
Yu, J., 2012. A particle filter driven dynamic gaussian mixture model approach for complex process monitoring and fault diagnosis. J. Process Control 22 (4), 778–788.