The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2018-3

Bayesian Analysis of Dynamic Transformed Exponential Family of Multivariate Models

Tevfik Aktekin
Peter T. Paul College of Business and Economics
University of New Hampshire, USA

Nick Polson
Booth School of Business
University of Chicago, USA

Refik Soyer
Department of Decision Sciences
The George Washington University, USA

Bayesian Analysis of Dynamic Transformed Exponential Family of Multivariate Models

Tevfik Aktekin
Decision Sciences
University of New Hampshire[*]

Nicholas G. Polson
Booth School of Business
University of Chicago

Refik Soyer
Decision Sciences
George Washington University

06/18/2018

Abstract

Dynamic transformed exponential family of models include the exponential, Weibull, Gamma, generalized Gamma, and Poisson families. Sequential on-line updating is an important aspect of fast filtering, inference and prediction. With this in mind, we develop particle filters tailored for our proposed family of dynamic models. The key for implementation is analytically tractable marginal likelihoods, a property not typically found in state space models outside of linear Gaussian models. To illustrate our methodology, we use simulated data examples and a real application for modeling the joint dynamics of stochastic volatility in financial indexes, the VIX and VXN.

Keywords: Transformed-exponential families, particle learning, state-space, dynamic time-series, stochastic volatility

[*]Aktekin is an Associate Professor of decision sciences at the Peter T. Paul College of Business and Economics in the University of New Hampshire. email: tevfik.aktekin@unh.edu. Polson is a Professor of Econometrics and Statistics at the Chicago Booth School of Business. email: ngp@chicagobooth.edu. Soyer is a Professor of decision sciences at the George Washington University School of Business. email: soyer@gwu.edu.

1 Introduction

Dynamic non-Gaussian models are of particular interest in modeling the underlying characteristics of spatio-temporal data. Most of these applications require fast on-line filtering, learning, and prediction as data arises sequentially. Non-Gaussian models can accurately represent observation features (e.g. skewness and heavy tails) and we present a rich class of multivariate dynamic transformed exponential family of models.

Many traditional stationary time series models can be viewed as state space models dating back to Kalman (1960) in engineering applications. Meinhold and Singpurwalla (1983) provide a clear treatment of the Kalman filter from a Bayesian perspective. State space models have been popular in many applications (e.g. finance, economics, engineering, environmental sciences, social sciences, among many others) due to their computational ease; see West and Harrison (1999) and Prado and West (2010) for reviews of state space models. The treatment of non-Gaussian state space models has received less attention in the literature due to computational constraints and lack of analytical tractability. Exponential family models with dynamic Gaussian state parameters are provided in West et al. (1985). Other notable classes of non-Gaussian state space models are discussed in Kitagawa (1987), Carlin et al. (1992), Frühwirth-Schnatter (1994), Kitagawa (1996), Durbin and Koopman (1997), Shephard and Pitt (1997), Durbin and Koopman (2000), Gamerman et al. (2013), among many others. However, almost none of this work focuses on the analysis of multivariate non-Gaussian time series data.

In this paper, we introduce a rich class of state space models for multivariate time series whose observations follow transformed exponential family of including the exponential, Weibull, Gamma, generalized Gamma, and Poisson. A key feature is the ability to introduce correlations across time as well as across series. We assume that the series are exposed to the same random common environment which is assumed to evolve dynamically. In a similar vein, Aktekin et al. (2018) consider only the case of count data. To provide filtering, smoothing, prediction, and learning we develop particle learning (PL) algorithms tailored for our proposed family of state space models that can address fast learning of both dynamic state and static parameters under certain conditions; see Lopes and Tsay (2011) and Singpurwalla et al. (2018) for reviews of particle based methods. Our PL algorithm is flexible and does not require a multi-dimensional proposal density for generating from the joint distribution of dynamic state and static parameter vectors, unlike traditional Markov chain Monte Carlo (MCMC) methods. To fully implement PL, we need to be able to sample from the propagation density and be able to evaluate the conditional predictive density, both of which are available for our class of models. We show that our PL algorithm does not require case-by-case considerations as the forms of the propagation and conditional predictive densities are readily available as functions of the model parameters and data.

One noteworthy advantage of our family of models is the availability of the marginal likelihood functions, a property not found outside of linear and Gaussian state space models (Kantas et al., 2015; Barra et al., 2017). The availability of the marginal likelihoods in our model obtained by integrating out the state variables, can be attributed to the imposed dynamics of our state variables. Recent work by Gamerman et al. (2013)

and Creal (2017) consider classes of non-Gaussian state space models with tractable marginal likelihoods for univariate settings. We build on this to obtain results for multivariate models. Another important feature is the closed form availability of the conditional distribution of a time series. Such a feature will be attractive in scenarios where information is available at each point in time about the rest of the series.

The rest of the paper is organized as follows. Section 2 presents the general form of our class of models, followed by examples of various observation equations in Section 3. Section 4 discusses estimation techniques using the PL algorithm and Section 5 covers numerical examples. Finally, Section 6 concludes with directions for future research.

2 Dynamic Transformed-Exponential Family (DTEF) Models

Consider a sequence of evenly spaced multivariate time series denoted by $\{(Y_{11},\dots,Y_{1T})$ $,\dots,$ $(Y_{J1},\dots,Y_{JT})\}$, where Y_{jt} represents the observations at time t of series j with $t=1,\dots,T$ and $j=1,\dots,J$. Our series are influenced by the same random common environment, namely θ_t. Lindley and Singpurwalla (1986) argue that using a common environment acting on all components of a system will induce correlations between the failure times of each component in reliability assessments. Following their approach, we propose the following general state space representation where a dynamically changing random common environment will induce correlations across the J time series.

The observation density from which the time series are sampled is assumed to have the general form given by

$$p(Y_{jt}|\theta_t,\lambda_j,\boldsymbol{\nu}) = f'(Y_{jt},\lambda_j,\boldsymbol{\nu})\theta_t^{g'(Y_{jt},\boldsymbol{\nu})}\exp\{-\theta_t h'(Y_{jt},\lambda_j,\boldsymbol{\nu})\}, \tag{1}$$

where θ_t's, λ_j's, and $\boldsymbol{\nu}$ are three sets of model parameters each representing a different aspect of time series data. θ_t's represent the effect of the random common environment on the J series as a function of time (we also refer to it as the state parameter in the general jargon of state space models). Both λ_j's and $\boldsymbol{\nu}$ represent random static effects that do not evolve over time. More specifically, $\boldsymbol{\nu}$ are a collection of static hyper-parameters which may be common or series specific and λ_j's represent static parameters that are specific to each series similar to random effects.

The above family of non-Gaussian distributions can also be referred to as the dynamic transformed-exponential family (DTEF) of models. Given θ_t's, λ_j's, and $\boldsymbol{\nu}$, Y_{jt}'s are assumed to be conditionally independent, therefore we can write the likelihood at time t as $p(\boldsymbol{Y}_t|\theta_t,\boldsymbol{\lambda},\boldsymbol{\nu}) = \prod_{j=1}^{J} p(Y_{jt}|\theta_t,\lambda_j,\boldsymbol{\nu})$ where $\boldsymbol{Y}_t = \{Y_{1t},\dots,Y_{Jt}\}$. Therefore, the general form can be written as

$$p(\boldsymbol{Y}_t|\theta_t,\boldsymbol{\lambda},\boldsymbol{\nu}) = f(\boldsymbol{Y}_t,\boldsymbol{\lambda},\boldsymbol{\nu})\theta_t^{g(\boldsymbol{Y}_t,\boldsymbol{\nu})}\exp\{-\theta_t h(\boldsymbol{Y}_t,\boldsymbol{\lambda},\boldsymbol{\nu})\}, \tag{2}$$

where $\boldsymbol{\lambda} = \{\lambda_1,\dots,\lambda_J\}$. In (2), the functions $f(.), g(.)$, and $h(.)$ will determine the form of the non-Gaussian family considered. We remark here that $f(.)$ and $g(.)$ can be functions of data as well as both static parameters, $g(.)$ is only allowed to be a function of data and the common static parameter, $\boldsymbol{\nu}$ and not $\boldsymbol{\lambda}$. The attractive feature of our proposed class of models is that the filtering densities and the marginal likelihoods conditional on the static parameters can be obtained in their general forms as functions of $f(.), g(.)$ and $h(.)$ if we assume a certain type of state evolution on the random common environment, θ_t's.

The functions $f(.), g(.)$ and $h(.)$ are chosen such that the function (2) is a proper probability density. Different forms of $f(.), g(.)$ and $h(.)$ leads to many well known non-Gaussian family of models such as the Poisson, exponential, Weibull, Gamma, and the generalized Gamma (that can also cover distributions such as the half-normal, Rayleigh, Maxwell-Boltzmann, chi-squared, and the log-normal).

We adopt the state evolution first introduced by Smith and Miller (1986) in modeling univariate dynamic time series. Various versions of this state evolution has been considered by Harvey and Fernandes (1989), Uhlig (1997), Aktekin and Soyer (2011), Aktekin et al. (2013), and Gamerman et al. (2013). Accordingly, we assume that θ_t evolves as

$$\theta_t = \frac{\theta_{t-1}}{\gamma}\epsilon_t, \tag{3}$$

where the error term conditional on the past and the static parameters (if needed) follows a Beta distribution of the form

$$(\epsilon_t|D^{t-1}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \sim Beta[\gamma\alpha_{t-1}, (1-\gamma)\alpha_{t-1}],$$

where $\alpha_{t-1} > 0, 0 < \gamma < 1$ and $D^{t-1} = \{D^{t-2}, \boldsymbol{Y}_{t-1}\}$.

In the above, γ is a discount parameter that may depend on the past data but not on θ_t. γ also influences the correlations of time series across time as well as across series whose implications we investigate in the sequel. In addition, the estimation of γ can be possible if the marginal likelihoods conditional on the static parameters were to be available analytically. We explore implementations for various classes of DTEF models with analytically tractable conditional marginal likelihoods.

To be able to sequentially update the dynamic states, we assume initially (prior to observing any data) that $(\theta_0|D^0) \sim Gamma(\alpha_0, \beta_0)$, then by induction we can show that

$$(\theta_{t-1}|D^{t-1}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \sim Gamma(\alpha_{t-1}, \beta_{t-1}),$$

where the updated conditional is given by

$$p(\theta_t|D^{t-1}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \int_0^{\theta_{t-1}/\gamma} p(\theta_t|\theta_{t-1}, D^{t-1}, \boldsymbol{\lambda}, \boldsymbol{\nu}) p(\theta_{t-1}|D^{t-1}, \boldsymbol{\lambda}, \boldsymbol{\nu}) d\theta_{t-1} \tag{4}$$

$$\sim Gamma(\gamma\alpha_{t-1}, \gamma\beta_{t-1}). \tag{5}$$

Here, $p(\theta_t|\theta_{t-1}, D^{t-1}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ is a scaled Beta density defined over $(0; \theta_{t-1}/\gamma)$ as implied by the state evolution (3).

As a consequence, we can show that the filtering density conditional on the static parameters will be

$$p(\theta_t|D^t, \boldsymbol{\lambda}, \boldsymbol{\nu}) \propto p(\boldsymbol{Y}_t|\theta_t, \boldsymbol{\lambda}, \boldsymbol{\nu})p(\theta_t|D^{t-1}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \tag{6}$$

$$\sim Gamma(\alpha_t, \beta_t), \tag{7}$$

where $p(\boldsymbol{Y}_t|\theta_t, \boldsymbol{\lambda}, \boldsymbol{\nu})$ is the observation equation from (2) and $(\theta_t|D^{t-1}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ is given by (5). Thus, the recursive parameter updates can be written as

$$\alpha_t = \gamma\alpha_{t-1} + g(\boldsymbol{Y}_t, \boldsymbol{\nu}),\ \beta_t = \gamma\beta_{t-1} + h(\boldsymbol{Y}_t, \boldsymbol{\lambda}, \boldsymbol{\nu}) \tag{8}$$

Given that the observation equation of a given model follows the form in (2), the above updating schemes can be used where all the multivariate data is pooled together in learning about the common random environment. We remark here that the updating of α_t and β_t are only available conditional on both sets of static parameters and their sequential estimation requires further attention. However, we also note that most multivariate state space models outside of Gaussianity typically will not lead to tractable state filtering densities even conditionally. Thus, the above updating scheme is an attractive feature of our proposed class of models. A similar updating scheme is considered in Gamerman et al. (2013) who only consider univariate state space models.

An important by-product of our proposed classes of DTEF models is the availability of the joint marginal distribution (marginal likelihood) of $\boldsymbol{Y}_t$ given the past conditional on the static parameters. This density is free of θ_t's. This joint marginal likelihood density rarely is analytically tractable outside of Gaussian state space models. In its general form, the density can be written as

$$p(\boldsymbol{Y}_t|D^{t-1}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \frac{\Gamma[\gamma\alpha_{t-1} + g(\boldsymbol{Y}_t, \boldsymbol{\nu})] f(\boldsymbol{Y}_t, \boldsymbol{\lambda}, \boldsymbol{\nu})(\gamma\beta_{t-1})^{\gamma\alpha_{t-1}}}{\Gamma(\gamma\alpha_{t-1})[\gamma\beta_{t-1} + h(\boldsymbol{Y}_t, \boldsymbol{\lambda}, \boldsymbol{\nu})]^{\gamma\alpha_{t-1} + g(\boldsymbol{Y}_t, \boldsymbol{\nu})}}. \tag{9}$$

There are mainly two uses for the above: 1) for one step-ahead online forecasting and 2) for marginal likelihood calculations in the estimation of static parameters such as $\boldsymbol{\nu}$ and γ. The fact that the form is free of the dynamic state parameters makes its availability an important feature for computational purposes. The form (9) leads to dynamic multivariate generalizations of known distributions. For instance, when the likelihood is Poisson, (9) is a dynamic multivariate negative binomial distribution (Aktekin et al., 2018). The availability of these known multivariate distributions allows us to obtain dynamic correlation estimates in a sequential manner whose practicality is discussed in our numerical examples section. In what follows, we present details of various DTEF models and discuss their implementation.

3 Examples of DTEF Models

3.1 Exponential State Space Model

In the case where Y_{jt}'s are exponentially distributed with rate parameter $\lambda_j\theta_t$, we can write the observation equation as

$$p(Y_{jt}|\theta_t, \lambda_j) = \lambda_j\theta_t \exp\{-(\theta_t\lambda_j Y_{jt})\}, \tag{10}$$

and

$$p(\boldsymbol{Y}_t|\theta_t, \boldsymbol{\lambda}) = \Big(\prod_{j=1}^{J} \lambda_j\Big)\theta_t^J \exp\{-\theta_t \sum_{j=1}^{J} \lambda_j Y_{jt}\}. \tag{11}$$

Using the general form in (2), we can decompose $f(.), g(.)$, and $h(.)$ as

$$f(\boldsymbol{Y}_t, \boldsymbol{\lambda}) = \prod_{j=1}^{J} \lambda_j,\ g(\boldsymbol{Y}_t) = J, \text{ and } h(\boldsymbol{Y}_t, \boldsymbol{\lambda}) = \sum_{j=1}^{J} \lambda_j Y_{jt},$$

Here the dependence on the common static parameter $\boldsymbol{\nu}$ is omitted as it will be an empty set. The filtering density conditional on λ_j's can be obtained via

$$p(\theta_t|D^t, \boldsymbol{\lambda}) \sim Gamma(\alpha_t, \beta_t), \tag{12}$$

with recursive updates of

$$\alpha_t = \gamma\alpha_{t-1} + J = \gamma\alpha_{t-1} + g(\boldsymbol{Y}_t), \text{ and } \beta_t = \gamma\beta_{t-1} + \sum_{j=1}^{J} \lambda_j Y_{jt} = \gamma\beta_{t-1} + h(\boldsymbol{Y}_t, \boldsymbol{\lambda}).$$

Using (9), the joint marginal distribution of $\boldsymbol{Y}_t$ given the past can be written as

$$p(\boldsymbol{Y}_t|D^{t-1}, \boldsymbol{\lambda}, \gamma) = \frac{\Gamma(\gamma\alpha_{t-1} + J)(\prod_{j=1}^{J} \lambda_j)(\gamma\beta_{t-1})^{\gamma\alpha_{t-1}}}{\Gamma(\gamma\alpha_{t-1})(\gamma\beta_{t-1} + \sum_{j=1}^{J} \lambda_j Y_{jt})^{\gamma\alpha_{t-1}+J}} \tag{13}$$

$$= \frac{(\prod_{j=1}^{J} \lambda_j/\gamma\beta_{t-1})\gamma\alpha_{t-1}(\gamma\alpha_{t-1}+1)\dots(\gamma\alpha_{t-1}+J-1)}{(1 + \sum_{j=1}^{J} \lambda_j Y_{jt}/\gamma\beta_{t-1})^{\gamma\alpha_{t-1}+J}}. \tag{14}$$

The above is a dynamic version of a J dimensional multivariate Lomax distribution of Nayak (1987) with parameters $\lambda_j/\gamma\beta_{t-1}$ and $\gamma\alpha_{t-1}$. When $\gamma\beta_{t-1}$ and $\gamma\alpha_{t-1}$ are constant, the above can be tied to the static version considered in Nayak (1987) and is typically used for modeling heavy tailed events in reliability analysis in assessing component lifetimes.

3.2 Weibull State Space Model

Another commonly used distribution in the analysis of non-Gaussian data is the Weibull which can exhibit many different density shapes. The Weibull distribution is widely used in modeling time to event type phenomenon and has found widespread application areas in reliability, survival, and risk analyses. It is possible to present the Weibull model via our general form using the following parameterization

$$p(Y_{jt}|\theta_t\lambda_j,\delta) = \theta_t\lambda_j bY_{jt}^{\delta-1}\exp\{-\theta_t\lambda_j Y_{jt}^{\delta}\}, \tag{15}$$

$$p(\boldsymbol{Y}_t|\theta_t,\boldsymbol{\lambda},\delta) = \theta_t^J\delta^J\prod_{j=1}^{J}\lambda_j\prod_{j=1}^{J}\left(Y_{jt}^{\delta-1}\right)\exp\{-\theta_t(\sum_{j=1}^{J}\lambda_j Y_{jt}^{\delta})\}, \tag{16}$$

where the scale parameter is $\lambda_j\theta_t$ and the shape parameter is δ which can be common to all J series or can be series specific. We remark here that it would be straightforward to assume the shape parameter to vary for each series by simply adding an index j. In this case where $\boldsymbol{\nu} = \delta$, we can write

$$f(\boldsymbol{Y}_t,\boldsymbol{\lambda},\boldsymbol{\nu}) = \delta^J\prod_{j=1}^{J}\lambda_j\prod_{j=1}^{J}\left(Y_{jt}^{\delta-1}\right),\ g(\boldsymbol{Y}_t,\boldsymbol{\nu}) = J,\ \text{and}\ h(\boldsymbol{Y}_t,\boldsymbol{\lambda},\boldsymbol{\nu}) = (\sum_{j=1}^{J}\lambda_j Y_{jt}^{\delta}).$$

The joint marginal distribution of $\boldsymbol{Y}_t$ given the past can be written as

$$p(\boldsymbol{Y}_t|D^{t-1},\boldsymbol{\lambda},\delta,\gamma) = \frac{\Gamma(\gamma\alpha_{t-1}+J)\delta^J\prod_{j=1}^{J}\lambda_j\left(\prod_{j=1}^{J}Y_{jt}\right)^{\delta-1}(\gamma\beta_{t-1})^{\gamma\alpha_{t-1}}}{\Gamma(\gamma\alpha_{t-1})(\gamma\beta_{t-1}+\sum_{j=1}^{J}\lambda_j Y_{jt}^{\delta})^{\gamma\alpha_{t-1}+J}}. \tag{17}$$

This is related to the multivariate generalized Burr distribution; see Tadikamalla (1980). When $\delta = 1$, it reduces to the dynamic multivariate Lomax distribution given by (14). It is sometimes referred to as the compound Weibull distribution (Dubey, 1968). We refer to the above as the dynamic multivariate Burr distribution.

3.3 Generalized Gamma State Space Model

Another rich non-Gaussian family of models arises from the generalized gamma (GG) distribution first considered by Stacy (1962). The GG distribution entails several well known distributions such as the half-normal, Rayleigh, Maxwell-Boltzmann, Chi-squared, and the log-normal and can therefore exhibit various density shapes. An important feature of the GG family is the flexibility of its hazard rate function that can be used to represent nonmonotonic unimodal or bathtub-shaped hazard functions. It has found applications in many fields such as reliability (Stacy, 1962), call center operations (Aktekin and Soyer, 2014), and marketing (Allenby et al., 1999). We consider the following parametrization of the GG distribution

$$p(Y_{jt}|\theta_t\lambda_j,\phi,\delta)=\frac{\delta(\theta_t\lambda_j)^\phi}{\Gamma(\phi)}Y_{jt}^{\phi\delta-1}\exp\{-\theta_t\lambda_jY_{jt}^\delta\}, \tag{18}$$

which leads to

$$p(\boldsymbol{Y}_t|\theta_t,\boldsymbol{\lambda},\phi,\delta)=\theta_t^{J\phi}\Big(\frac{\delta}{\Gamma(\phi)}\Big)^J\Big(\prod_{j=1}^J\lambda_j\Big)^\phi\Big(\prod_{j=1}^JY_{jt}\Big)^{\phi\delta-1}\exp\{-\theta_t\sum_{j=1}^J\lambda_jY_{jt}^\delta\}, \tag{19}$$

In this setting with $\boldsymbol{\nu}=(\phi,\delta)$. $f(.),g(.)$, and $h(.)$ are given by

$$f(\boldsymbol{Y}_t,\boldsymbol{\lambda},\boldsymbol{\nu})=\Big(\frac{\delta}{\Gamma(\phi)}\Big)^J\Big(\prod_{j=1}^J\lambda_j\Big)^\phi\Big(\prod_{j=1}^JY_{jt}\Big)^{\phi\delta-1},g(\boldsymbol{Y}_t,\boldsymbol{\nu})=J\phi,\text{and}h(\boldsymbol{Y}_t,\boldsymbol{\lambda},\boldsymbol{\nu})=\sum_{j=1}^J\lambda_jY_{jt}^\delta$$

Next, we focus our attention on the analysis of a special case of the GG distribution when $\delta=1$ in the above which yields the gamma distribution. Thus, we can write

$$p(Y_{jt}|\phi,\theta_t\lambda_j)=\frac{(\theta_t\lambda_j)^\phi}{\Gamma(\phi)}Y_{jt}^{\phi-1}\exp\{-\theta_t\lambda_jY_{jt}\}, \tag{20}$$

$$p(\boldsymbol{Y}_t|\theta_t,\boldsymbol{\lambda},\phi)=\frac{\theta_t^{J\phi}(\prod_{j=1}^J\lambda_j)^\phi}{\Gamma(\phi)^J}\Big(\prod_{j=1}^JY_{jt}\Big)^{\phi-1}\exp\{-\theta_t\sum_{j=1}^J(\lambda_jY_{jt})\}, \tag{21}$$

where the scale parameter $(\lambda_j\theta_t)$ varies as a function of time and series and the shape parameter (ϕ) can either be assumed to be common to all j series or can be series specific. The marginal likelihood can be obtained as

$$p(\boldsymbol{Y}_t|D^{t-1},\boldsymbol{\lambda},\phi,\gamma)=\frac{\Gamma(\gamma\alpha_{t-1}+J\phi)(\prod_{j=1}^J\lambda_j)^\phi\Big(\prod_{j=1}^JY_{jt}\Big)^{\phi-1}(\gamma\beta_{t-1})^{\gamma\alpha_{t-1}}}{\Gamma(\gamma\alpha_{t-1})\Gamma(\phi)^J(\gamma\beta_{t-1}+\sum_{j=1}^J\lambda_jY_{jt})^{\gamma\alpha_{t-1}+J\phi}}, \tag{22}$$

which is a dynamic version of the generalized multivariate Lomax distribution of Nayak (1987) (a.k.a. the multivariate Beta prime).

For settings where regression type estimation is of interest (prediction in the same time period as opposed to one step ahead predictions), the joint predictive densities (marginal likelihoods) (9) are available for every class of DTEF model considered here. Not only these marginal likelihoods can assist with the estimation of hyper parameters and model selection but they can also lead to the conditional mean (regression) estimates as well. This is one of the notable strengths of our DTEF family of models that is typically not found outside of Gaussian state space models. For the Gamma model, the conditional expectation given the static parameters can obtained as

$$E(Y_{jt}|Y_{1t},\dots,Y_{j-1,t},D^{(t-1)},\boldsymbol{\lambda},\boldsymbol{\phi},\gamma)=\frac{\phi_j\Big(\gamma\beta_{t-1}+(\sum_{i=1}^{j-1}\lambda_i Y_{it})\Big)}{\lambda_j\Big(\gamma\alpha_{t-1}+(\sum_{i=1}^{j-1}\phi_i)-1\Big)}, \tag{23}$$

where $\boldsymbol{\phi}=(\phi_1,\dots,\phi_J)$, ϕ_j's are component specific shape parameters and Y_{jt} is a linear function of $Y_{1t},\dots,Y_{j-1,t}$. Such a linear relationship can be better observed in a bivariate Gamma model, whose conditional mean estimates would be as

$$E(Y_{1t}|Y_{2t},D^{(t-1)},\boldsymbol{\lambda},\boldsymbol{\phi},\gamma)=\frac{\phi_1\Big(\gamma\beta_{t-1}+\lambda_2 Y_{2t}\Big)}{\lambda_1\Big(\gamma\alpha_{t-1}+\phi_2-1\Big)}. \tag{24}$$

Given the estimates of both the dynamic and static parameters, the above can be computed as a Monte Carlo average. Even though, we do not make use of these conditional expectations (regressions) in our numerical example, they may be used viable alternatives for prediction in various other settings. For instance in reliability analysis, we may have information about the failure time of a component during a certain time period which may lead to better estimates of the failure time of a second component given the first one.

Table 1 summarizes our DTEF family of models where the different combinations of $f(.), g(.)$ and $h(.)$ for each model is shown. We note here that the details of the Poisson model presented in the table can be found in Aktekin et al. (2018).

4 Algorithmic Issues: Particle Learning (PL)

We now turn our attention to the estimation of the dynamic state, θ_t's as well as the static parameters, λ_j's. It is possible to consider two options: MCMC or particle filtering (PF) methods. The main goal is to sequentially sample from, $p(\theta_1,\dots,\theta_t,\lambda_1,\dots,\lambda_j|D^t)$, the joint distribution of all model parameters. In order to implement an MCMC algorithm, we need to generate θ_t's via $p(\theta_1,\dots,\theta_t|\lambda_1,\dots,\lambda_j,D^t)$ and λ_j's via $p(\lambda_1,\dots,\lambda_j|\theta_1,\dots,\theta_t,D^t)$ iteratively in the form of a Gibbs sampler. However, generating the θ_t's would require a significant amount of effort as the conditional filtering densities cannot be obtained in closed form, which in turn suggests the use of a Metropolis-Hastings step to generate the θ_t's as a single block or a forward filtering backward sampling algorthm (FFBS). Moreover, Storvik (2002) points out that the MCMC methods require the restarting of each chain as new observations are observed which increases the computational burden significantly. The same argument is re-iterated by Gramacy and Polson (2011) in the context of sequential design and optimization of Gaussian processes.

In this paper, we consider PF methods, more specifically the particle learning (PL) technique whose requirements for implementation can be met given the conditional conjugate nature of our model. The main idea of particle filtering is based on the re-

Model	$\boldsymbol{f}(.)$	$\boldsymbol{g}(.)$	$\boldsymbol{h}(.)$
Exponential	$\prod_{j=1}^{J} \lambda_j$	J	$\sum_{j=1}^{J}(\lambda_j Y_{jt})$
Weibull	$\delta^J(\prod_{j=1}^{J} \lambda_j)(\prod_{j=1}^{J} Y_{jt}^{\delta-1})$	J	$\sum_{j=1}^{J}(\lambda_j Y_{jt}^{\delta})$
Gamma	$\left(\frac{1}{\Gamma(\phi)}\right)^J \left(\prod_{j=1}^{J} \lambda_j\right)^{\phi} \left(\prod_{j=1}^{J} Y_{jt}\right)^{\phi-1}$	$J\phi$	$\sum_{j=1}^{J}(\lambda_j Y_{jt})$
Gen. Gamma	$\left(\frac{\delta}{\Gamma(\phi)}\right)^J \left(\prod_{j=1}^{J} \lambda_j\right)^{\phi} \left(\prod_{j=1}^{J} Y_{jt}\right)^{\phi\delta-1}$	$J\phi$	$\sum_{j=1}^{J}(\lambda_j Y_{jt}^{\delta})$
Poisson	$\prod_{j=1}^{J} \left(\frac{\lambda_j^{Y_{jt}}}{Y_{jt}!}\right)$	$\sum_{j=1}^{J} Y_{jt}$	$\sum_{j=1}^{J} \lambda_j$

Table 1: *Summary of the $f(.), g(.)$ and $h(.)$ functions which lead to different classes of DTEF models. For Weibull, Gamma, and generalized Gamma models, the shape and scale parameters are shown as common to each series for notational convencience and they can be extended to individually represent each series.*

balancing of a finite number of particles of the posterior states given the next data point proportional to the likelihood. Our PL approach is based on the work discussed in Carvalho et al. (2010). We also note that our algorithm can be classified as an optimal particle filter as it does not require any MCMC steps when the discount parameter, γ, is treated as a known quantity which we suggest for settings where fast learning and prediction is of interest such as online environments.

In what follows, we first summarize the general steps of the PL algorithm followed by a more detailed version tailored for our class of DTEF models. The main goal in PL is to generate joint sequential samples from the density $p(\theta_{t+1}, \boldsymbol{\lambda}|D^{t+1}, \boldsymbol{\nu})$. To achieve this, we first resample joint particles from $p(\theta_t, \boldsymbol{\lambda}|D^{t+1}, \boldsymbol{\nu})$ (the key idea here is that the jointly sampled particles are coupled) and propagate θ_{t+1} from the density $p(\theta_{t+1}|\theta_t, D^{t+1}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ (note that this propagation density is defined conditional on D^{t+1} and not D^t unlike the state evolution distribution). In other words, we eventually end up obtaining samples from $p(\theta_{t+1}|D^{t+1}, \boldsymbol{\nu})$ via

$$p(\theta_{t+1}|D^{t+1}, \boldsymbol{\nu}) = \int p(\theta_{t+1}|\theta_t, D^{t+1}, \boldsymbol{\lambda}, \boldsymbol{\nu}) p(\theta_t, \boldsymbol{\lambda}|D^{t+1}, \boldsymbol{\nu}) d\theta_t d\boldsymbol{\lambda}.$$

The best way to represent the decomposition of the PL algorithm is through the writing of the Bayes' rule conditional on the past dynamic state parameter and the static parameters via

$$p(\mathbf{Y_{t+1}}|\theta_{t+1}, \boldsymbol{\lambda}, \boldsymbol{\nu}) p(\theta_{t+1}|\theta_t, D^t, \boldsymbol{\lambda}, \boldsymbol{\nu}) = p(\mathbf{Y_{t+1}}|\theta_t, \boldsymbol{\lambda}, \boldsymbol{\nu}) p(\theta_{t+1}|\theta_t, D^{t+1}, \boldsymbol{\lambda}, \boldsymbol{\nu}), \quad (25)$$

where on the left hand side, the first term is the likelihood from (2) and the second term is implied by the state transition equation from (3) which will be a scaled Beta density as $Beta[\gamma\alpha_{t-1}, (1-\gamma)\alpha_{t-1}]$ defined over $(0; \theta_t/\gamma)$. Thus, to implement the PL algorithm, we need $p(\mathbf{Y_{t+1}}|\theta_t, \boldsymbol{\lambda}, \boldsymbol{\nu})$ for computing the weights for resampling and $p(\theta_{t+1}|\theta_t, D^{t+1}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ for propagation of the states. It is possible to write the forms of the propagation density and the resampling weights in a general form for the proposed class of models. The propagation density for all the models can be represented via

$$p(\theta_{t+1}|\theta_t, D^{t+1}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \propto p(\boldsymbol{Y}_{t+1}|\theta_{t+1}, \boldsymbol{\lambda}, \boldsymbol{\nu}) p(\theta_{t+1}|\theta_t, \boldsymbol{\lambda}, \boldsymbol{\nu}, D^{t+1}) \quad (26)$$

$$\propto \theta_{t+1}^{\gamma\alpha_t + g(\boldsymbol{Y}_t, \boldsymbol{\nu}) - 1} \Big(1 - \frac{\gamma}{\theta_t}\theta_{t+1}\Big)^{(1-\gamma)\alpha_t} \exp\{-\theta_{t+1} h(\boldsymbol{Y}_t, \boldsymbol{\lambda}, \boldsymbol{\nu})\}, \quad (27)$$

which is the density form of the scaled hyper-geometric Beta distribution (Gordy, 1998) with parameters $\gamma\alpha_t + g(\boldsymbol{Y}_t, \boldsymbol{\nu}), (1-\gamma)\alpha_t$ and $(\theta_t/\gamma)h(\boldsymbol{Y}_t, \boldsymbol{\lambda}, \boldsymbol{\nu})$ defined over the range, $(0, \theta_t/\gamma)$. Typically, even though we can analytically obtain the propagation densities, they do not represent a class of models that we can easily generate samples from. Instead, we use an alternative step based on the importance sampling approach considered in Pitt and Shephard (1999) for particle filters.

Next step is to obtain the conditional predictive density, $p(\mathbf{Y_{t+1}}|\theta_t, \boldsymbol{\lambda}, \boldsymbol{\nu})$, which is used to compute the weights of the particles for resampling. For all of the DTEF classes of models considered previously, we can obtain

$$p(\mathbf{Y_{t+1}}|\theta_t,\boldsymbol{\lambda},\boldsymbol{\nu}) = \frac{p(\mathbf{Y_{t+1}}|\theta_{t+1},\boldsymbol{\lambda},\boldsymbol{\nu})p(\theta_{t+1}|\theta_t,D^t,\boldsymbol{\lambda},\boldsymbol{\nu})}{p(\theta_{t+1}|\theta_t,D^{t+1},\boldsymbol{\lambda},\boldsymbol{\nu})} \tag{28}$$

$$= f(\mathbf{Y}_t,\boldsymbol{\lambda},\boldsymbol{\nu})\Big(\frac{\theta_t}{\gamma}\Big)^{g(\mathbf{Y}_t,\boldsymbol{\nu})}\frac{B[\gamma\alpha_t+g(\mathbf{Y}_t,\boldsymbol{\nu}),(1-\gamma)\alpha_t]}{B[\gamma\alpha_t,(1-\gamma)\alpha_t]}{}_1F_1(a^\star,b^\star,c^\star), \tag{29}$$

where $a^\star = \gamma\alpha_t + g(\mathbf{Y}_t,\boldsymbol{\nu}), b^\star = \alpha_t + g(\mathbf{Y}_t,\boldsymbol{\nu})$ and $c^\star = -(\theta_t/\gamma)h(\mathbf{Y}_t,\boldsymbol{\lambda},\boldsymbol{\nu})$. In the above, ${}_1F_1(a^\star,b^\star,c^\star)$ represents confluent hyper-geometric function (CHF). As we only need to evaluate the resampling weights, we do not need to recognize their forms as known densities and the CHF can be evaluated in a straightforward manner in most computing environments .

Given the availability of the propagation and the conditional predictive densities, we can now make use of the PL algorithm of Carvalho et al. (2010) to update both the dynamic and the static parameters. In general, we can summarize the PL algorithm as follows. First, we start by jointly resampling the state and static particles at time t using weights proportional to the conditional marginal density (or the conditional predictive likelihood). This way only particle sets that are more likely are moved forward in the filter. After the resampling step, the propagation of the current state at time t to the future state at $t+1$ occurs. An attractive feature of approach is that in both steps, future observations are used, a property not shared with other PF methods. In addition, resampling the particles first has many advantages as pointed out by Liu and Chen (1998). Due to the order of operations, the PL method is typically referred to as a resample-propagate method. In the last step, the static parameters are generated using their conditional sufficient statistics that are available in our model. This is another noteworthy property of the algorithm which makes estimating static parameters a straightforward exercise (given that the conditional sufficient statistics are available). The idea of tracking sufficient statistics in particle methods dates back to the works of Storvik (2002) and Fearnhead (2002).

To implement the algorithm, we first define z_t as the essential vector of parameters which will be stored at each point in time. We define $z_t = \{\theta_t,\boldsymbol{\lambda},s_t\}$ where s_t is the conditional sufficient statistics for λ_j's. As a consequence, z_t needs to be updated sequentially. Therefore, the algorithm can be summarized as

1. (Resample) $\{z_t\}_{i=1}^N$ from $z_t^{(i)} = \{\theta_t,\boldsymbol{\lambda},s_t\}^{(i)}$ using weights $w_t^{(i)} \propto p(\mathbf{Y_{t+1}}|z_t^{(i)})$
2. (Propagate) $\{\theta_t^{(i)}\}$ to $\{\theta_{t+1}^{(i)}\}$ via $p(\theta_{t+1}|z_t^{(i)},\mathbf{Y_{t+1}})$
3. (Update) $s_{t+1}^{(i)} = f(s_t^{(i)},\theta_{t+1}^{(i)},\mathbf{Y_{t+1}})$
4. (Sample) $(\lambda)^{(i)}$ from the conditional posterior, $p(\lambda|s_{t+1}^{(i)})$

As $p(\theta_{t+1}|z_t^{(i)},\mathbf{Y_{t+1}})$ is not a well known density that can be easily sampled from, we propagate samples for step 2 using an importance sampling step in our numerical examples.

In step 3 of the PL algorithm, the conditional sufficient statistics are updated according to a deterministic function f. If we assume Gamma distributed priors for the λ_j's as $Gamma(a_{j,0}, b_{j,0})$, we can obtain such updating for each family of DTEF model considered in this study. As an example, when the likelihood is exponential, we can write

$$p(\lambda_j|\theta_1, \ldots, \theta_{t+1}, D^{t+1}) \propto p(Y_{j,1}, \ldots, Y_{j,t+1}|\theta_{t+1}, \lambda_j)p(\lambda_j) \tag{30}$$

$$\propto \left(\lambda_j^{t+1} \exp\{-\lambda_j \sum_{i=1}^{t+1} \theta_i Y_{ji}\}\right)\left(\lambda_j^{a_{j,0}-1} e^{-b_{j,0}\lambda_j}\right), \tag{31}$$

which is a Gamma density as $(\lambda_j|\theta_1, \ldots, \theta_{t+1}, D^{t+1}) \sim Gamma(a_{j,t+1}, b_{j,t+1})$ where the deterministic mapping of f required for the implementation of the PL algorithm can be summarized as

$$a_{j,t+1} = a_{j,0} + (t+1) \text{ and } b_{j,t+1} = b_{j,0} + \Big(\sum_{i=1}^{t+1} \theta_i Y_{ji}\Big), \tag{32}$$

or as desired using the conditional sufficient statistics updating form

$$a_{j,t+1} = a_{j,t} + 1 \text{ and } b_{j,t+1} = b_{j,t} + (\theta_{t+1} Y_{j,t+1}), \tag{33}$$

where$a_{j,t+1}$ and $b_{j,t+1}$ are defined as conditional sufficient statistics (conditional on the previous time period's sufficient statistics). In summary, our goal in sequential inference is to obtain $(\lambda_j|D^{t+1})$ for each j which can be thought as a mixture approximation as

$$\approx p^N(\lambda_j|D^{t+1}) = \frac{1}{N}\sum_{i=1}^{N} p(\lambda_j|s_{t+1}^{(i)}), \tag{34}$$

given samples from $s_{t+1}^{(i)}$ from $(s_{t+1}|D^{t+1})$. Note that in the above, the conditioning in $p(\lambda_j|s_{t+1}^{(i)})$ will include the dependence on the state, θ_{t+1}. Given (33), it is possible to fully implement the PL algorithm by starting sampling from the initial prior for θ_0. Table 2 provides a summary of conditional sufficient statistics updating for each DTEF model.

Estimating the static parameters, $\boldsymbol{\nu}$ and γ

One of the requirements of the PL framework for updating the static parameters is the analytically tractability of the conditional sufficient statistics, which was available for the λ_j's. As these conditional sufficient statistics will not be tractable for the common static parameters, $\boldsymbol{\nu}$ and γ, we can instead use the conditional marginal likelihoods that are available for the proposed class of DTEF models and are free of the dynamic state parameters. In its general form, the posterior distributions of all common static parameters

Model	Sufficient Statistics Updating
Exponential	$a_{j,t+1} = a_{j,t} + 1$
	$b_{j,t+1} = b_{j,t} + (\theta_{t+1} Y_{j,t+1})$
Weibull	$a_{j,t+1} = a_{j,t} + 1$
	$b_{j,t+1} = b_{j,t} + (\theta_{t+1} Y^{\delta}_{j,t+1})$
Gamma	$a_{j,t+1} = a_{j,t} + \phi$
	$b_{j,t+1} = b_{j,t} + (\theta_{t+1} Y_{j,t+1})$
Gen. Gamma	$a_{j,t+1} = a_{j,t} + \phi$
	$b_{j,t+1} = b_{j,t} + (\theta_{t+1} Y^{\delta}_{j,t+1})$
Poisson	$a_{j,t+1} = a_{j,t} + Y_{jt}$
	$b_{j,t+1} = b_{j,t} + \theta_{t+1}$

Table 2: *Summary of conditional sufficient statistics updating for the familty of DTEF models*

including the discount factor, γ, can be written as

$$p(\boldsymbol{\nu}, \gamma | D^t, \boldsymbol{\lambda}) \propto \prod_{t=1}^{T} p(\boldsymbol{Y}_t | D^{t-1}, \boldsymbol{\nu}, \gamma, \boldsymbol{\lambda}) p(\boldsymbol{\nu}) p(\gamma), \tag{35}$$

where the marginal likelihood term, $p(\boldsymbol{Y}_t | D^{t-1}, \boldsymbol{\nu}, \gamma)$, can be obtained via (2). Depending on the form of the marginal likelihood, we can use a Metropolis-Hastings algorithm to generate the samples from (35). Alternatively, in the case where only a single parameter such as γ needs to be learned, we can assume a discrete prior and obtain a discrete posterior distribution proportional to the product of the marginal likelihood and the prior. For instance, in the exponential model, we can write (35) recursively and free of $\boldsymbol{\nu}$ as

$$p(\gamma | D^t, \boldsymbol{\lambda}) \propto p(\boldsymbol{Y}_t | D^{t-1}, \gamma, \boldsymbol{\lambda}) p(\gamma | D^{t-1}, \boldsymbol{\lambda}),$$
$$p(\gamma = i | D^t, \boldsymbol{\lambda}) = \frac{p(\boldsymbol{Y}_t | D^{t-1}, \gamma = i, \boldsymbol{\lambda}) p(\gamma = i | D^{t-1}, \boldsymbol{\lambda})}{\sum_{i=1}^{K} \Big(p(\boldsymbol{Y}_t | D^{t-1}, \gamma = i, \boldsymbol{\lambda}) p(\gamma = i | D^{t-1}, \boldsymbol{\lambda}) \Big)}, \text{ for } i = 1, \ldots, K,$$

where the marginal likelihood is the multivariate Lomax distribution as in (14). The above can be added to the end of each iteration of the PL algorithm after the updating of λ_j's.

5 Numerical Examples

In order to demonstrate the implementation of our DTEF models, we present a simulated example using data generated from the Weibull model and an actual example on stochastic volatility for the SP500 and the Nasdaq indexes that are drawn from the publicly available Yahoo finance online database. The data as well as the R code are available upon request via email from the authors.

5.1 Weibull Example: Simulated Sets

We constructed 8 simulated sets from the data generating process of the Weibull model. For each one of the simulations, the γ parameter was set between 0.2 and 0.9 with increments of 0.1. Each set consists of 250 observations for five series (50 for each series). Initially, we set $\theta_0 \sim G(\alpha_0 = 1, \beta_0 = 1)$ which represents the starting status of the random common environment. Even though this is not a requirement of our model, it is not unreasonable to assume that the random common environment is initialized around the unit scale (with mean $\alpha_0/\beta_0 = 1$). This way the scale of the static parameters, λ_j's are more interpretable with respect to the scale of the data generated. In all 8 simulations, λ_j's, were set to 1,2, 3, 4, and 5, respectively. In addition, the shape parameter (denoted by δ) was set at 3. We remark here that the sample correlations across the five series for each of the simulated sets ranged between 0.01 and 0.91 with most cases between 0.25 and 0.50. A quick summary of our simulated data generation is as follows,

- Generate θ_0 via $p(\theta_0) \sim Gamma(\alpha_0, \beta_0)$.

For $t = 2, \ldots, T$,

- Generate θ_t via $p(\theta_t|\theta_{t-1}) \sim ScaledBeta(\gamma\alpha_{t-1}, (1-\gamma)\alpha_{t-1})$ where $\theta_t \in (0, \theta_{t-1}/\gamma)$.
- Generate Y_{jt}'s via $p(Y_{jt}|\theta_t, \lambda_j, \delta) \sim Weibull(\delta, \theta_t\lambda_j)$ for $j = 1, \ldots, J$.
- Update $\alpha_t = \gamma\alpha_{t-1} + J$.

In order to implement the PL algorithm, we generated 1,000 particles for each scenario. We remark here that we also investigated the effects of generating larger sets of particles and found that the results were identical. To preserve space, we omit the inference results of estimation of our experiments with particle sizes larger than 1,000. For priors, we set $\theta_0 \sim Gamma(10, 10)$ so that the random common environment is roughly centered around the unit scale. The static parameter priors were assumed to be uninformative as $\lambda_j \sim Gamma(0.001, 0.001)$ for $j = 1, \ldots, 5$. We also investigated the implications of imposing tighter priors centered around their true means and found that the results were very similar. For real data applications, we suggest the use of vague priors on the static parameters as their true values will be unknown unlike the simulated examples. The discount parameter γ was assumed to be discrete uniform between 0.01 and 0.99 with 100 atoms. We assumed that the shape parameter was fixed at its true value at 3.

In step 2 of the PL algorithm detailed in Section 3, we need to be able to generate samples from the propagation density $p(\theta_{t+1}|z_t, \boldsymbol{Y}_{t+1})$ which has the form of a scaled hyper-geometric Beta (HGB) distribution. Depending on $f(.), g(.),$ and, $h(.)$, generating samples from this HGB density may become difficult. Alternatively, in our numerical examples we considered an importance sampling step similar to the approach proposed by Pitt and Shephard (1999). This introduces a new propagate-resample step in lieu of the propagation step where θ_{t+1}'s are first propagated from the state transition equation (3) then are resampled using weights proportional to the likelihood. More specifically, we replace step 2 of our PL algorithm from Section 3 with the following propagate-resample steps

	Series 1	Series 2	Series 3	Series 4	Series 5
γ = 0.2	0.20	0.24	0.22	0.19	0.22
γ = 0.3	0.22	0.23	0.19	0.21	0.20
γ = 0.4	0.23	0.22	0.22	0.19	0.17
γ = 0.5	0.24	0.20	0.22	0.19	0.16
γ = 0.6	0.20	0.23	0.19	0.19	0.18
γ = 0.7	0.19	0.24	0.17	0.24	0.18
γ = 0.8	0.16	0.23	0.23	0.21	0.18
γ = 0.9	0.16	0.23	0.23	0.21	0.17

Table 3: *Weibull Example: MAPE estimates for several simulations for various values of the discount factor.*

- (Propagate) $\{\tilde{\theta}_{t+1}\}^{(i)}$ from $p(\theta_{t+1}|\theta_t^{(i)}, \boldsymbol{\lambda}^{(i)}, D^t)$
- (Resample) $\{\theta_{t+1}\}^{(i)}$ using weights $w_{t+1} \propto p(\mathbf{Y_t}|\tilde{\theta}_{t+1}^{(i)}, \boldsymbol{\lambda}^{(i)})$

Figure 1 shows the time series plot of the simulated data (in light gray) versus the filtered mean estimates (in red) for one of the simulated examples of the above mentioned Weibull model. Overall, at first glance, the model seems to provide a reasonable fit to the generated multivariate data unless there is a drastic downward or upward jump in the time series.

To demonstrate the in-sample fit performance of our proposed model, we computed the median absolute percentage errors (MAPE) as (actual data - posterior mean estimate divided by actual data). Table 3 shows a summary of the MAPEs for each series for all 8 simulated sets. The MAPEs range between 16 % and 24 % and do not exhibit any noticeable patterns as we increase or decrease the magnitude of the discount factor γ.

5.2 Gamma Example: Stochastic Volatility Data

Stochastic volatility models are widely used in finance and econometrics for assessing the time dependent fluctuations of financial assets or financial indexes. In our numerical illustration, we consider two volatility market indexes, VIX and VXN, that are known to be highly correlated. VIX is the ticker symbol for the Chicago Board Options Exchange (CBOE) volatility index, which shows the market's expectation of a 30-day volatility. It is constructed using the implied volatilities of a wide range of SP500 index options. VXN is a measure of market expectations of 30-day volatility for the Nasdaq-100 market index, as implied by the price of near-term options on this index. The monthly time series data is obtained from the publicly available Yahoo finance online database for a span of five years between October 2012 and October 2017. A time series plot of VIX and VXN are given in Figure 2 where we can observe co-movement of the series which makes it a viable candidate for our DTEF for modeling correlated multivariate data.

Moreover, stochastic volatility data is known to exhibit non-Guassian features such as longer tails and non-symmetricies, properties that can be accounted by our family of DTEF models. In this numerical example, we illustrate the implementation of our

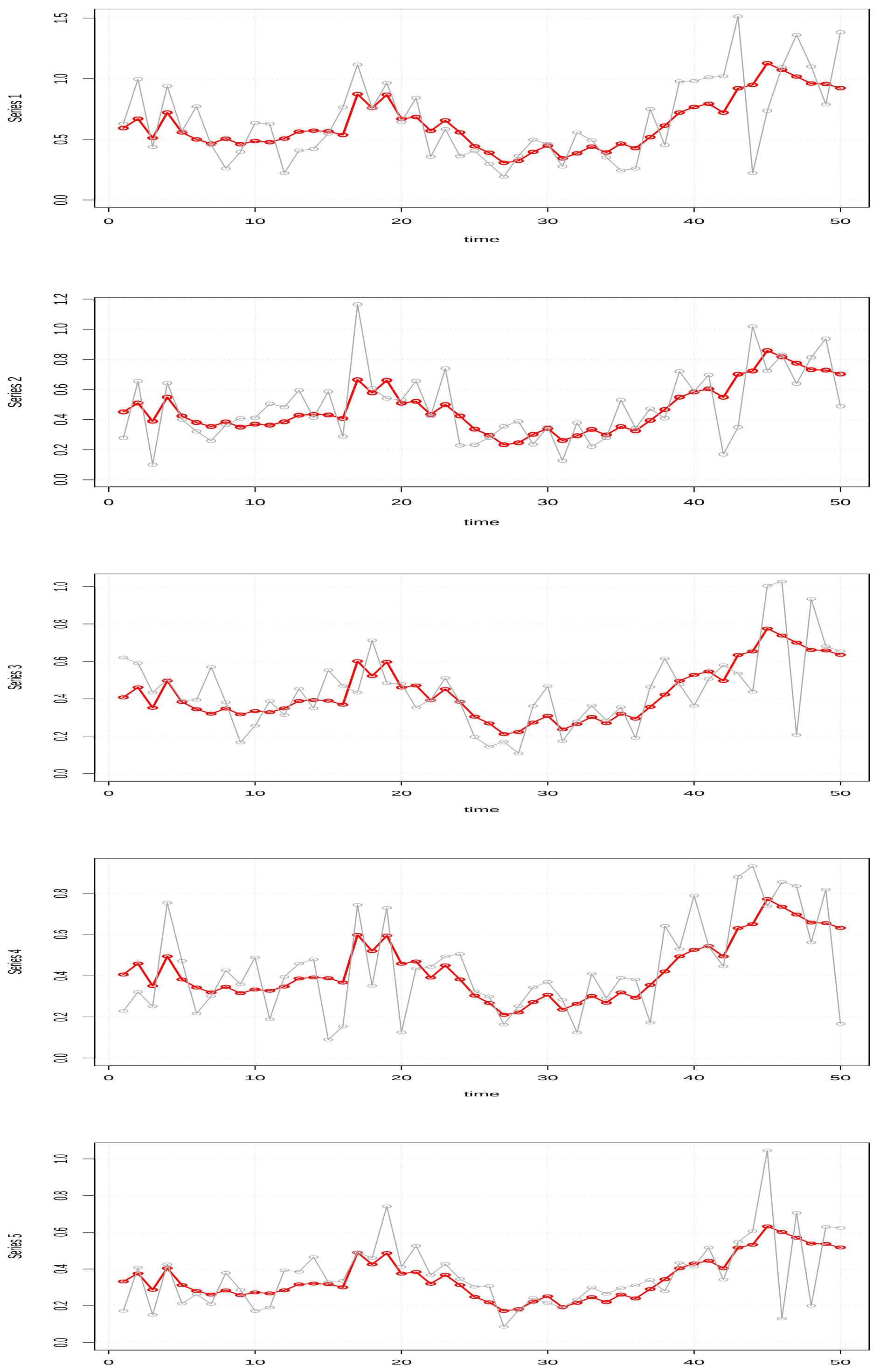

Figure 1: *Simulated Weibull Example: The gray lines represent the simulated data for five series and red lines are the respective filtered mean estimates.*

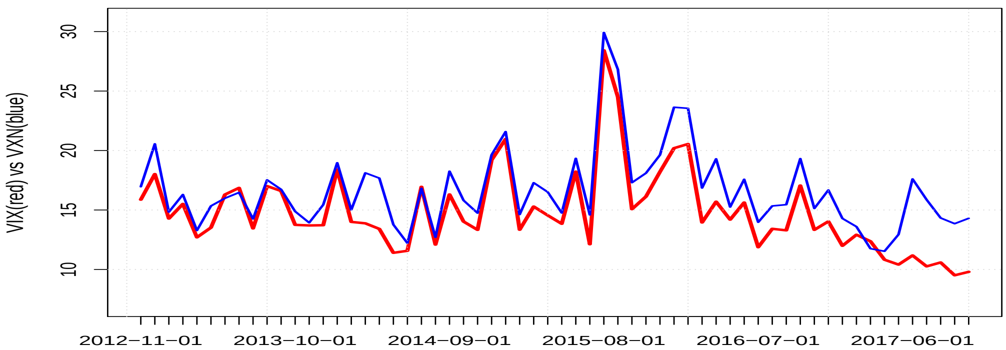

Figure 2: *VIX (red) versus VXN (blue) monthly time series plot for five years between 10/12 and 10/17*

Gamma model as stochastic volatility is typically modeled as a Gamma variable in the finance and econometrics literature (Creal, 2017). As per our notation introduced in (21), the shape parameter ϕ can be assumed to be common to all series or can be allowed to vary across series. This can informally be investigated by plotting the histogram of both series and by fitting respective Gamma densities. The histograms for both VIX and VXN data are shown in Figure 3 based on which we can infer non-Gaussianity with longer tails for both series. At first glance, it looks like the fitted Gamma densities may have similar shape parameters but it is tough to reach a definitive conclusion. Therefore, we used the method of moments to estimate series specific shape parameters, denoted as ϕ_1 (VIX) and ϕ_2 (VXN) in the rest of this section. The estimated values were $\phi_1 = 1.23$ and $\phi_2 = 1.44$. Thus, we use these estimates as fixed inputs in our PL algorithm in estimating the rest of the model parameters. Alternatively, it is also possible to treat the shape parameters as uncertain and use the approach discussed in (35) for estimation along with γ.

As before, we are interested in estimating the dynamic state variables, series specific static parameters, and the the discount factor, γ. To implement the PL algorithm, we found that generating 1,000 particles were more than enough for a problem this size (we experimented with various larger particle sizes and found that they all lead to almost identical results). The static parameter priors were assumed to follow $\lambda_j \sim Gamma(0.001, 0.001)$ for $j = 1, \ldots, 2$. γ was defined between 0.01 and 0.99 with 100 atoms as a discrete prior. Setting up an initial prior for θ_0 has proven to be a more difficult task than expected. Unlike the simulated examples, as we are not informed about the true distribution of the θ_t's, we first assumed an uninformative prior such as $Gamma(0.001, 0.001)$. However, depending on the initial sampled values of θ_0 (when it is very small or very large), our estimation procedure leads to unidentifiability issues between the dynamic state and the static parameters. To mitigate this issue, we instead considered more informative priors such as $Gamma(1, 1), Gamma(10, 10)$

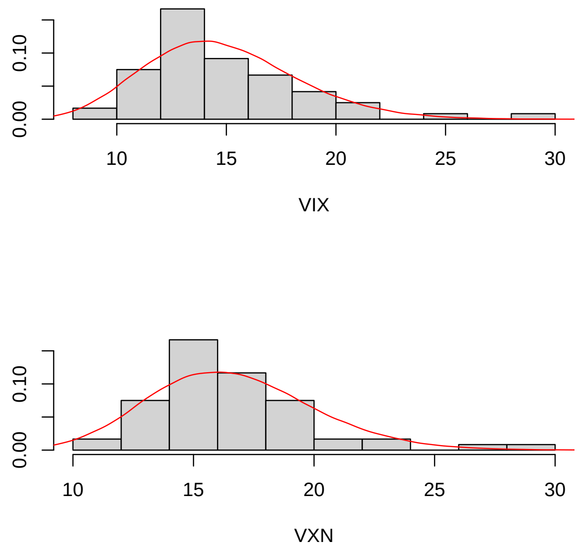

Figure 3: *The histograms for VIX and VXN (gray) versus the fitted Gamma densities (red)*

	Mean APE	Median APE	Pred. Mean APE	Pred. Median APE
VIX	0.102	0.095	0.14	0.13
VXN	0.093	0.093	0.13	0.14

Table 4: *Absolute percentage error summary for the multivariate Gamma model fit and predictive performance for the stochastic volatility example*

or $Gamma(100, 100)$, neither of which caused any further issues. When the prior is $Gamma(0.001, 0.001)$, the initial blind sampling of θ_0 may be from regions that are extremely small or large in which case the algorithm may take a long time to attain the underlying level. Therefore, we recommend the use of tighter priors for θ_0 that are centered around unity when fitting our DTEF models.

A time series plot of the actual VIX and VXN (gray) data versus the mean filtered estimates (red) are shown in Figure 4. The model provides a reasonable fit except for time periods when there are unexpected large jumps in the actual data (for instance between observations 30 and 40, there is a huge jump in the expected volatilities for both markets). To quantify the model fit, we calculated the mean (10.2% and 9.3%) and the median absolute percentage errors (9.3% and 9.5%) that are shown in Table 4. Both estimates show evidence in favour of our model providing a reasonable fit to actual data. In addition, we also obtained sequential one step ahead predictions of VIX and VXN jointly. The one step ahead predictive mean and median APE's were in the region of 13% and 14%.

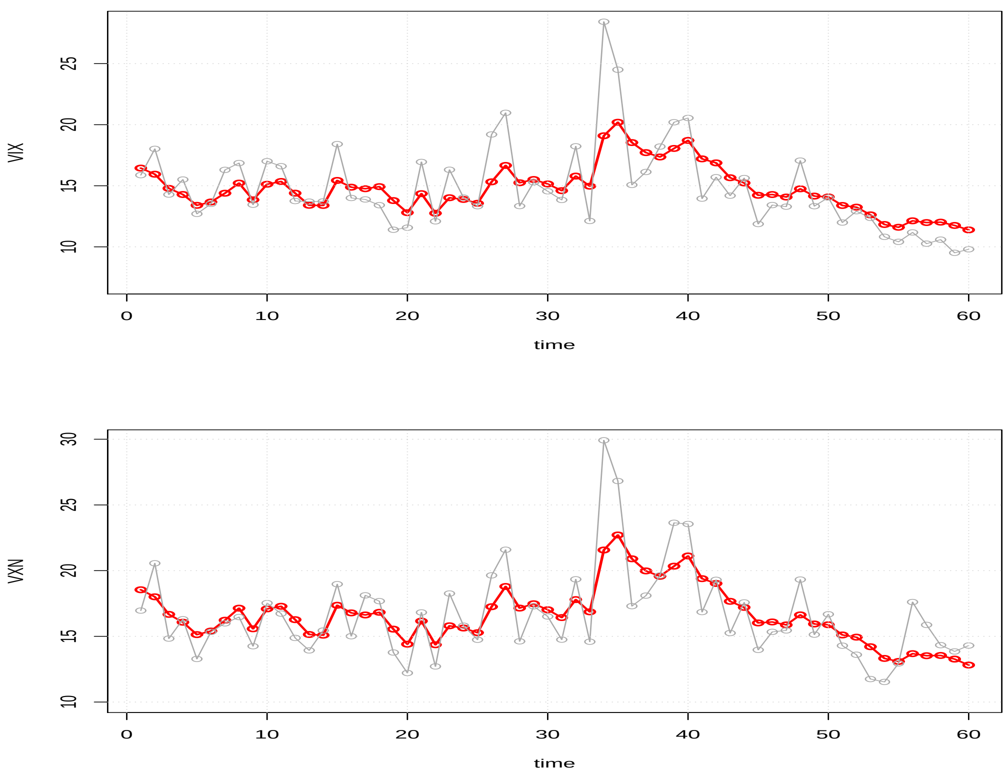

Figure 4: *VIX and VXN time series plots (gray) versus mean filtered estimates (red)*

	Pred. Mean APE	Pred. Median APE
VIX	0.320	0.276
VXN	0.297	0.256

Table 5: *Predictive performance of univariate Gamma models for the stochastic volatility example*

To further investigate the predictive performance of our Gamma DTEF model, we estimated a univariate state space model for each series individually where the likelihoods (observation equations) are Gamma densities and the state transitions are assumed to follow scaled Beta densities as before. A quick summary of the univariate Gamma model is as follows:

- $Y_t \sim Gamma(\phi, \theta_t)$ (observation equation)
- $(\theta_t|\theta_{t-1}) \sim ScaledBeta[\gamma\alpha_{t-1}, (1-\gamma)\alpha_{t-1}]$ with $\theta_t \in (0; \theta_{t-1}/\gamma)$ (transition equation)
- $(\theta_t|D^t) \sim Gamma(\alpha_t, \beta_t)$ (filtering density)
- $\alpha_t = \gamma\alpha_{t-1} + \phi$ and $\beta_t = \gamma\beta_{t-1} + Y_t$ (sequential updates)

As before, we used the method of moments to estimate the respective Gamma shape parameters and employed the same initial state prior, $\theta_0 \sim G(0.001, 0.001)$. To estimate γ, we used the marginal likelihood implied by the above observation and state equations. The posterior distribution of γ can be obtained proportional to the following

$$p(\gamma|D^t) \propto \prod_{i=1}^{t} \Big(\frac{\Gamma(\gamma\alpha_{i-1} + \phi)(\gamma\beta_{i-1})^{\gamma\alpha_{i-1}}}{\Gamma(\gamma\alpha_{i-1})(\gamma\beta_{i-1} + Y_i)^{\gamma\alpha_{i-1}+\phi}} \Big) p(\gamma)$$

The above posterior distribution is estimated sequentially with a discrete prior defined between 0.01 and 0.99 and is used in computing the one step-ahead predictions, $E(Y_t|D^{t-1}) = a(\gamma\beta_{t-1})/(\gamma\alpha_{t-1} - 1)$. The median and mean APE estimates are given in Table 5 for each series (we note here that the first two observation predictions were omitted from the APE calculations to provide a fair comparison to the multivariate DTEF model as their inaccuracy significantly skews the APE summary measures). The APE estimates are significantly above those obtained for the DTEF model which further supports the need to model these series as dependent and multivariate.

The posterior distributions of the static parameters, λ_1 (VIX) and λ_2 (VXN) are given in Figure 5. The estimation track paths λ_1, λ_2 and γ are shown in Figure 6. As expected, initially our uncertainty for λ_1 and λ_2 is relatively larger as we do not observe a lot of data. After roughly observing ten data points, the static parameter estimates seem to converge with diminishing credibility intervals as we become more informed. A similar pattern can be observed regarding γ, where initially it is distributed closer to the discrete prior centered around 0.5. After observing 15 data points, it also converges to the region of its final posterior distribution. Overall, the track plots gives us an informal and visual representation of the PL algorithm estimates over time.

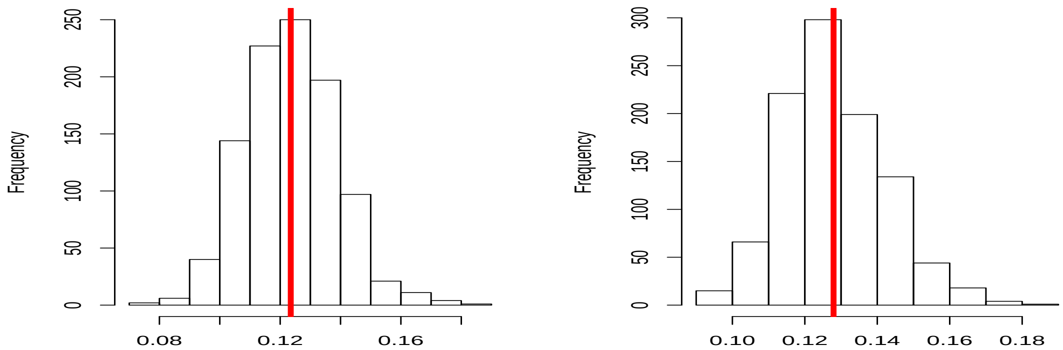

Figure 5: *Posterior distributions of λ_1 (VIX) and λ_2 (VXN) for the stochastic volatility example*

Another practical by-product of our DTEF models is the availability of the dynamically changing random common environment fluctuations and patterns as provided by the joint state distribution, $p(\theta_1, \ldots, \theta_t|D^t)$. In this example, the common environment can be interpreted as the general financial market conditions which influence both indexes in a similar manner. Figure 7 shows the boxplot of the dynamic state parameters representing how the financial market conditions evolve over time. The red line represents a pseudo average market expectation over the last five years (simply the average of the expected value of all the θ_ts over the last five years) which can be used as a visual benchmark to assess the overall health of the financial markets. This may allow financial analysts to infer the relative performance of the index with respect to the average market health over the last five years. For instance, if the estimate for a given month falls below or above this benchmark, this may reveal patterns about market patterns which would not be readily available just by observing the time series plots of each series individually. The additional insights provided by these dynamic state parameters can be attributed to the pooling of information across several series and the modeling of the uncertainty in the latent common random environment.

One-step Ahead Dynamic Correlation Estimates

The availability of the joint marginal likelihoods (9) for our class of DTEF models also leads to analytically tractable one-step ahead dynamic correlations, denoted by $cor(Y_{jt}, Y_{it}|D^{t-1})$ where the uncertainty of the state variable θ_t has been integrated out. Our view on the interpretation of this is that $cor(Y_{jt}, Y_{it}|D^{t-1})$ represents the correlation between two series when the uncertainty of the environment has been accounted for (a concept akin to the autocorrelation of residual terms in least squares regression models). Even though, $cor(Y_{jt}, Y_{it}|D^{t-1})$ can be obtained for all of the DTEF models presented in this paper, we only consider the case that is relevant to the analysis of the stochastic volatility data.

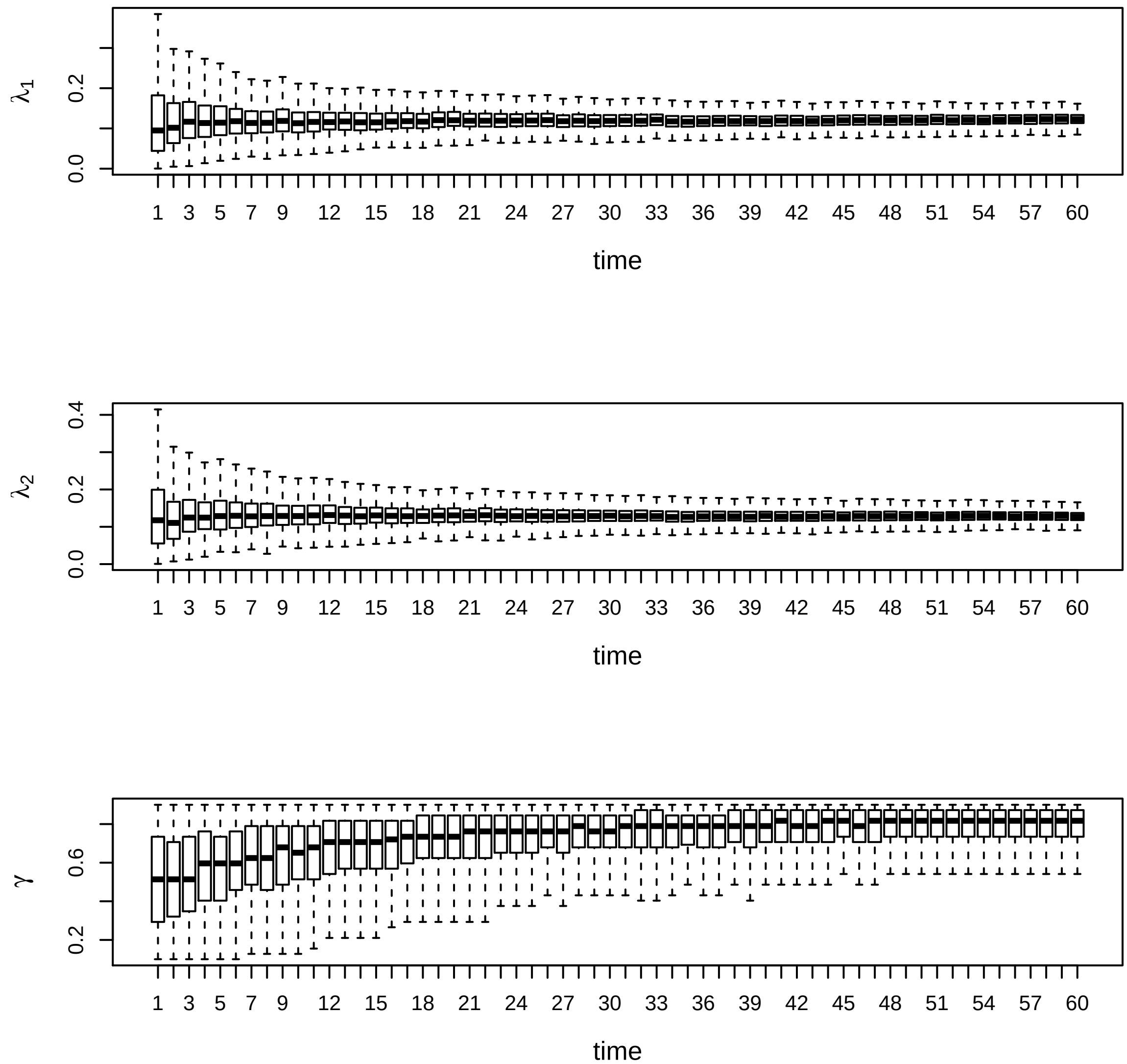

Figure 6: *The estimation track paths for the posterior distributions of λ_1 (VIX), λ_2 (VXN) and γ for the stochastic volatility example*

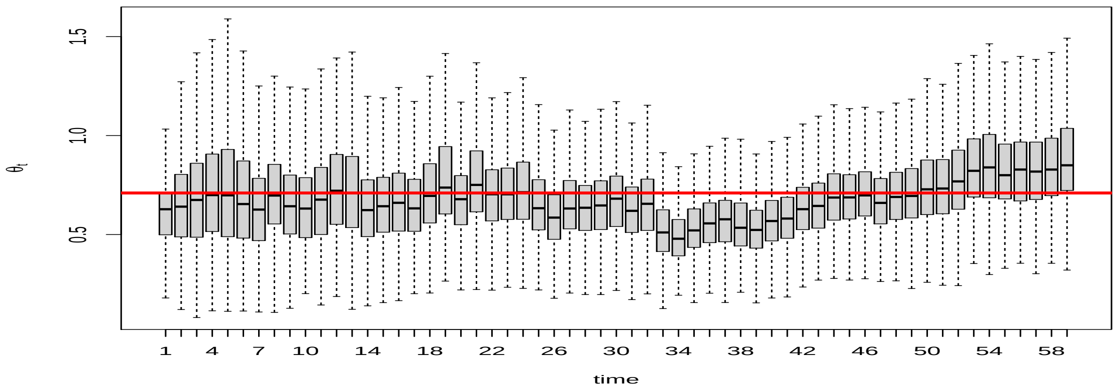

Figure 7: *Boxplot of the dynamic state parameters representing how the financial market conditions evolve over time. The red line represents a pseudo average market expectation over the last five years (simply the average of the expected value of all the θ_ts over the last five years) which can be used as a visual benchmark to assess the overall health of the financial markets.*

Using the joint marginal likelihoods (9), we can show that the one-step ahead dynamic correlations can be computed via

$$cor(Y_{jt}, Y_{it}|D^{t-1}) = \sqrt{\frac{\phi_i \phi_j}{(\gamma\alpha_{t-1} + \phi_i - 1)(\gamma\alpha_{t-1} + \phi_j - 1)}}, \tag{36}$$

where ϕ_i and ϕ_j are the respective shape parameters of the ith and jth series. The effect of time is embedded in α_t's as well as the posterior distribution of γ, $p(\gamma|D^t)$. Not every class of DTEF model allows for the one-step ahead dynamic correlations to vary across series when there are more than two series. For instance in the exponential model, we can show that the dynamic correlation estimate for any pair of i and j is computed as $1/(\gamma\alpha_{t-1})$ which is neither a function of i nor j.

Plot 8 shows the one-step ahead dynamic correlations estimates between VIX and VXN for the stochastic volatility example. We can infer that the initial uncertainty in the correlation estimates is higher with respect to the end of the series. Such behavior may be attributed to the fact that posterior distribution of γ has larger variance initially as observed in the estimation trackplot of γ from Figure 6. As we learn more about the relationship between the two series, uncertainty about the discount factor diminishes as also reflected to the dynamic correlation estimates. Furthermore, the relative magnitude of the dynamic correlation estimates are also higher initially. Namely, when making initial predictions, learning about the two series jointly is relatively more important as opposed to learning about them individually. Whereas when the predictions are not highly correlated, considering both series jointly may not help us reduce the future uncertainty about these predictions. These correlations in turn may help us identify inflation points for decoupling multivariate time series models so that they can be treated

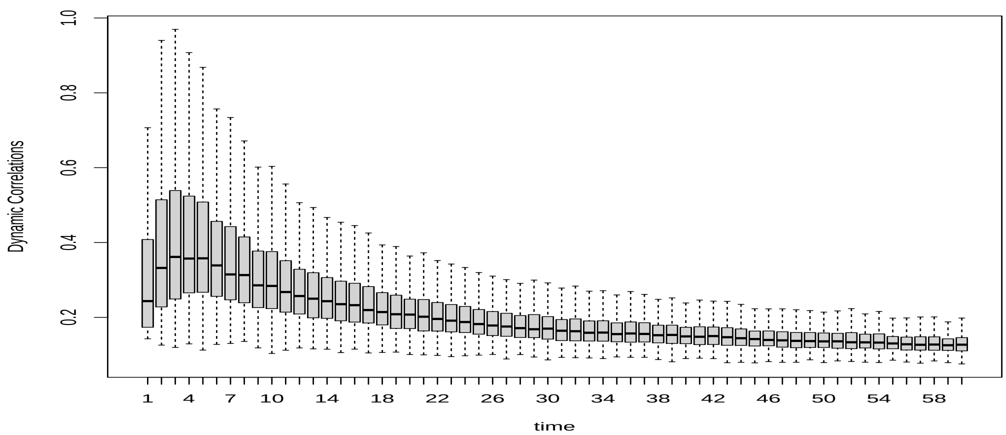

Figure 8: *Monthly one-step ahead dynamic correlation estimates over time for VIX and VXN indexes*

as univariate models once the state parameter and the discount parameters have been estimated. This is potentially a desirable feature in large networks of time series models where estimation speed is of utmost importance. For instance, in our example, the sample correlation estimate between the two series (VIX and VXN) is 0.91. However, dynamic correlation estimates from Figure 6 indicate are somewhere between 0.2 and 0.1. This implies that most of the correlation between the two series can be accounted for via the state parameters, θ_t's that represent the financial market conditions these two series are exposed to.

6 Concluding Remarks

In this paper, we developed a class of non-Gaussian state space models for multivariate time series such as the exponential, Weibull, Gamma, generalized Gamma, and the Poisson. The series are assumed to be exposed to the same common random environment which in term induces dependence across time as well as across series. By imposing a scaled Beta state transition density, our class of models leads to analytically tractable filtering and one-step ahead densities conditional on static (series specific) parameters. To estimate the dynamic state, static and hyper parameters, we developed fast particle learning algorithms tailored for our proposed family of models We argued that our PL algorithm is computationally less expensive with respect to MCMC methods and does not require case-by-case considerations as the forms of the propagation and conditional predictive densities are readily available as functions of the model parameters and data for all the classes of models considered here.

A useful property of our class of DTEF models is the availability of the marginal likelihood functions (a property not found in state space models outside of linear and Gaussian models). The availability of the marginal likelihood allows us to estimate hyper-parameters such as the discount factor (γ) with less relative computational burden due to the fact that the dynamic state vector can be integrated out. These marginal likelihoods also lead to straightforward estimations of the conditional means (regressions) and dynamic correlations. Furthermore, they can also be employed with online model selection as a model fit exercise. The multivariate marginal likelihoods for each one of our class of models can be tied back to known multivariate distributions such as the Lomax, generalized Lomax, and the multivariate Burr distributions, to name a few. To show the implementation of our models, we designed a simulated data example followed by a real application of stochastic volatility of financial indexes.

A noteworthy extension of the proposed DTEF model would be to treat the discount factor, γ as dynamic by imposing an evolution over time similar to that of the state parameter, θ_t. This approach may lead to a more flexible class of DTEF models but will increase the computational burden significantly as the marginal likelihoods cannot be used directly for estimation unlike our current approach.

References

Aktekin, T., Polson, N. G., and Soyer, R. (2018). Sequential Bayesian analysis of multivariate count data. *Bayesian Analysis*, 13(2):385–409.

Aktekin, T. and Soyer, R. (2011). Call center arrival modeling: A Bayesian state-space approach. *Naval Research Logistics*, 58(1):28–42.

Aktekin, T. and Soyer, R. (2014). Bayesian analysis of abandonment in call center operations. *Applied Stochastic Models in Business and Industry*, 30(2):141–156.

Aktekin, T., Soyer, R., and Xu, F. (2013). Assessment of mortgage default risk via Bayesian state space models. *Annals of Applied Statistics*, 7(3):1450–1473.

Allenby, G. M., Leone, R. P., and Jen, L. (1999). A dynamic model of purchase timing with application to direct marketing. *Journal of the American Statistical Association*, 94(446):365–374.

Barra, I., Hoogerheide, L., Koopman, S. J., and Lucas, A. (2017). Joint Bayesian analysis of parameters and states in nonlinear non-Gaussian state space models. *Journal of Applied Econometrics*, 32(5):1003–1026.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992). A monte carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, 87(418):493–500.

Carvalho, C., Johannes, M. S., Lopes, H. F., and Polson, N. (2010). Particle learning and smoothing. *Statistical Science*, 25(1):88–106.

Creal, D. D. (2017). A class of non-Gaussian state space models with exact likelihood inference. *Journal of Business & Economic Statistics*, 35(4):585–597.

Dubey, S. D. (1968). A compound Weibull distribution. *Naval Research Logistics*, 15(2):179–188.

Durbin, J. and Koopman, S. (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*, 62(1):3–56.

Durbin, J. and Koopman, S. J. (1997). Monte carlo maximum likelihood estimation for non-gaussian state space models. *Biometrika*, 84(3):669–684.

Fearnhead, P. (2002). Markov chain monte carlo, sufficient statistics, and particle filters. *Journal of Computational and Graphical Statistics*, 11(4):848–862.

Frühwirth-Schnatter, S. (1994). Applied state space modelling of non-gaussian time series using integration-based kalman filtering. *Statistics and Computing*, 4(4):259–269.

Gamerman, D., Dos-Santos, T. R., and Franco, G. C. (2013). A non-Gaussian family of state-space models with exact marginal likelihood. *Journal of Time Series Analysis*, 34(6):625–645.

Gordy, M. B. (1998). A generalization of generalized beta distributions. Technical report.

Gramacy, R. B. and Polson, N. G. (2011). Particle learning of Gaussian process models for sequential design and optimization. *Journal of Computational and Graphical Statistics*, 20(1):102–118.

Harvey, A. C. and Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business and Economic Statistics*, 7(4):407–417.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45.

Kantas, N., Doucet, A., Singh, S. S., Maciejowski, J., Chopin, N., et al. (2015). On particle methods for parameter estimation in state-space models. *Statistical Science*, 30(3):328–351.

Kitagawa, G. (1987). Non–gaussian state space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82(400):1032–1041.

Kitagawa, G. (1996). Monte carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25.

Lindley, D. V. and Singpurwalla, N. D. (1986). Multivariate distributions for the life lengths of components of a system sharing a common environment. *Journal of Applied Probability*, 23:418–431.

Liu, J. S. and Chen, R. (1998). Sequential monte carlo methods for dynamic systems. *Journal of the American Statistical Sssociation*, 93(443):1032–1044.

Lopes, H. F. and Tsay, R. S. (2011). Particle filters and bayesian inference in financial econometrics. *Journal of Forecasting*, 30(1):168–209.

Meinhold, R. J. and Singpurwalla, N. D. (1983). Understanding the kalman filter. *The American Statistician*, 37(2):123–127.

Nayak, T. K. (1987). Multivariate lomax distribution: properties and usefulness in reliability theory. *Journal of Applied Probability*, 24(01):170–177.

Pitt, M. L. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599.

Prado, R. and West, M. (2010). *Time series: modeling, computation, and inference*. CRC Press.

Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika*, 84(3):653–667.

Singpurwalla, N. D., Polson, N. G., and Soyer, R. (2018). From least squares to signal processing and particle filtering. *Technometrics*, 60(2):146–160.

Smith, R. and Miller, J. E. (1986). A non-Gaussian state space model and application to prediction of records. *Journal of the Royal Statistical Society, Series B*, 48(1):79–88.

Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, pages 1187–1192.

Storvik, G. (2002). Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing*, 50(2):281–289.

Tadikamalla, P. R. (1980). A look at the Burr and related distributions. *International Statistical Review*, pages 337–344.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica*, pages 59–73.

West, M. and Harrison, J. (1999). *Bayesian Forecasting and Dynamic Models, Second Edition*. Springer.

West, M., Harrison, J., and Migon, H. S. (1985). Dynamic generalized linear models and bayesian forecasting. *Journal of the American Statistical Association*, 80(389):73–83.