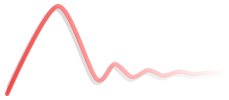

Dynare Working Papers Series
https://www.dynare.org/wp/

Pruned Skewed Kalman Filter and Smoother: With Application to the Yield Curve

Gaygysyz Guljanov
Willi Mutschler
Mark Trede

Working Paper no. 78

December 2022

CEPREMAP
CENTRE POUR LA RECHERCHE ECONOMIQUE ET SES APPLICATIONS
48, boulevard Jourdan — 75014 Paris — France
https://www.cepremap.fr

Pruned Skewed Kalman Filter and Smoother: With Application to the Yield Curve

Gaygysyz Guljanov[a], Willi Mutschler[b,*], Mark Trede[a]

[a] *Center for Quantitative Economics, University of Münster.*
[b] *School of Business and Economics, University of Tübingen.*

Abstract

The *Skewed Kalman Filter* is a powerful tool for statistical inference of asymmetrically distributed time series data. However, the need to evaluate Gaussian cumulative distribution functions (cdf) of increasing dimensions, creates a numerical barrier such that the filter is usually applicable for univariate models and under simplifying conditions only. Based on the intuition of how skewness propagates through the state-space system, a computationally efficient algorithm is proposed to *prune* the overall skewness dimension by discarding elements in the cdfs that do not distort the symmetry up to a pre-specified numerical threshold. Accuracy and efficiency of this *Pruned Skewed Kalman Filter* for general multivariate state-space models are illustrated through an extensive simulation study. The *Skewed Kalman Smoother* and its pruned implementation are also derived. Applicability is demonstrated by estimating a multivariate dynamic Nelson-Siegel term structure model of the US yield curve with Maximum Likelihood methods. We find that the data clearly favors a skewed distribution for the innovations to the latent level, slope and curvature factors.

Keywords: state-space models, skewed Kalman filter, skewed Kalman smoother, closed skew-normal, dimension reduction, yield curve, term structure, dynamic Nelson-Siegel

*The authors thank Dietmar Bauer for both critical and helpful comments as well as sharing his codes on the Mendell-Elston method with us.

**The second author acknowledges financial support from the Deutsche Forschungsgemeinschaft (DFG) through Grant No. 411754673. Declarations of interest: none.

Replication codes are available at `https://github.com/wmutschl/pruned-skewed-kalman-paper`.

*Corresponding author.

Email addresses: `gaygysyz.guljanov@wiwi.uni-muenster.de` (Gaygysyz Guljanov), `willi@mutschler.eu` (Willi Mutschler), `mark.trede@uni-muenster.de` (Mark Trede)

1. Introduction

State-space models and the Kalman filter are at the heart of modern signal processing and robust control, but they also play an inherent role in statistics and econometrics, particularly for the estimation of structural models such as vector autoregressive or dynamic stochastic general equilibrium models, see e.g. Kilian & Lütkepohl (2017) and Fernández-Villaverde et al. (2016) for textbook introductions. The reduced-forms of these structural models are special cases of the class of linear state-space models developed by engineers and statisticians to describe physical and dynamic systems. Furthermore, by assuming Gaussianity for both the structural innovations and measurement errors the state and control variables of the system become normally distributed. In this case, the Kalman filter is an efficient recursive procedure for inference about the state vector and can be exploited to compute the exact Gaussian likelihood function. The filter is optimal in the sense that it minimizes the one-step ahead prediction errors covariance matrix. More importantly, from an applied and computational perspective, Kalman filter operations can be performed rapidly and efficiently.

However, in real data applications non-Gaussianity, and particularly skewness, is a feature of many time series commonly used to estimate linear state-space models (de Roon & Karehnke, 2017; Fagiolo et al., 2008). For example in the economics literature the estimated innovations systematically exhibit strong asymmetry as recently shown by Lindé et al. (2016) for monetary policy shocks and by Ludvigson et al. (2021) for both financial and macroeconomic uncertainty shocks. The pronounced negative skewness of stock returns and the implications for asset pricing and investment management have been extensively documented, see e.g. Neuberger (2012). Likewise, Ruge-Murcia (2017) finds that skewed nominal shocks (e.g. on inflation) are important to account for slope changes in the yield curve, while asymmetric real shocks (e.g. on consumption) are important to account for level changes. This fits seamlessly into the Macro-Finance literature, which advocates the use of Kalman filtering techniques to model and forecast yield curves. In fact, it has long been recognized that the first three principal components of yields, commonly labeled level, slope, and curvature, provide a sufficient empirical summary of the entire yield curve.[1] Following Diebold & Li (2006), it has become standard practice to analyze the term structure of interest rates by estimating dynamic Nelson & Siegel (1987)-type models with the conventional *Gaussian Kalman filter*. Recently, Bauer & Chernov (2021) argue that conditional yield skewness is an important summary statistic about the state of the economy, (not only but) especially in the face of unprecedented low interest rates. Consequently, this calls into question the validity of the Gaussian assumption when estimating yield curves with Kalman filtering techniques.

Therefore, one needs to adapt the state-space modelling framework and algorithms to account for skewness in the error term distribution. To this end, we consider the closed skew-normal (CSN) distribution of González-Farías et al. (2004b) as a suitable alternative, because it generalizes the Gaussian distribution

[1] Granted there is also some consensus that additional factors are important in determining the shape of the yield curve, see e.g. Gürkaynak & Wright (2012) for a review.

by adding skewness while preserving some important properties of the normal distribution; namely, being closed under (i) marginalization, (ii) conditional distributions, (iii) linear transformations (full column or row rank), (iv) summation of independent random variables from this family, and (v) joint distributions of independent random variables in this family, see e.g. Azzalini & Capitanio (2014) and Genton (2004) for excellent textbook introductions. Moreover, it includes both the normal distribution as well as the popular skew-normal distribution proposed by Azzalini (1985) and Azzalini & Dalla Valle (1996) as special cases. Since the three basic tools in the implementation of the Kalman filter are closure under linear transformation, summation, and conditioning, the use of this distribution allows the development of closed-form recursions that are almost identical to the *Gaussian Kalman Filter*. Accordingly, Naveau et al. (2005) and Cabral et al. (2014) formulate *Skewed Kalman Filters* based on the CSN distribution for linear state-space systems, whereas Rezaie & Eidsvik (2014, 2016) develop *Skewed Unscented Kalman Filters* for nonlinear state-space systems and also discuss the computational aspects.

The (closed) skew-normal distribution is applied in all kinds of disciplines, to list a few: insurance claims in property-liability (Eling, 2012), growth-at-risk analysis (Adrian et al., 2019; Wei et al., 2021; Wolf, 2022), mental well-being studies (Pescheny et al., 2021), modelling psychiatric measures (Counsell et al., 2011), risk management (Vernic, 2006), stochastic frontier models (Chen et al., 2014), stock returns (Chen et al., 2003), and time series econometrics (Karlsson et al., 2021; Manouchehri & Nematollahi, 2019). The *Skewed Kalman Filter*, however, is rarely used despite its great potential and familiarity of implementation. Particularly, in economics and econometrics the literature is very sparse, Cabral et al. (2014) being a notable exception for the analysis of UK gas consumption and Emvalomatis et al. (2011) for estimating dynamic efficiency measurements in agricultural economics. Other applications in statistics and signal processing are usually limited to simplified conditions and univariate settings. We suspect that this is mainly due to a numerical problem known as "increasing skewness dimensions" for which we present a solution in this paper. Intuitively, the problem stems from the fact that the probability density function (pdf) of the CSN distribution has two dimensions as we multiply a Gaussian pdf by the ratio of two Gaussian cumulative distribution functions (cdf). While the Gaussian pdf is the resemblance to the normal distribution, the skewness dimension stems from the Gaussian cdfs. Even though evaluating Gaussian cdfs is a well understood task, it can become numerically difficult if the dimension of the cdfs becomes very large. In more detail, if one assumes the CSN distribution for the error terms (shocks) in the state transition equation, we can establish the closure under summation property required for the Kalman filtering steps based on the fact that the sum of two independent CSN distributed variables is also CSN. However, the skewness dimension grows over time, as the dimensions of the resulting Gaussian cdfs that need to be evaluated increase by the sum of the cdf dimensions in both variables. We therefore need to deal with high-dimensional Gaussian cdfs, the computation of which is not difficult in the univariate case, but becomes cumbersome if not infeasible for high-dimensional distributions, a point echoed by Amsler et al. (2021) for the skew-normal distribution.

And this is the core numerical problem of the *Skewed Kalman Filter*, as in typical applications of state-space models this skewness dimension quickly increases and may even explode as the recursion proceeds over many time steps. Therefore, Rezaie & Eidsvik (2014) argue that for practical purposes one must either assume simplified conditions or re-fit the updated distribution in some way. Accordingly, Naveau et al. (2005) and Cabral et al. (2014) propose to only assume the CSN distribution for the initial state vector. As the sum of a CSN distributed variable with a normally distributed variable is also CSN, the Kalman filtering steps can be easily derived, but the skewness dimension stays constant. However, as we will show, the impact of the initial distribution – and the amount of skewness – vanishes rather quickly over time, which is typically not a feature we find in real data applications. Alternatively, Naveau et al. (2005) tailor an extended univariate state-space model by splitting up the state vector in a linear and a skewed part, enabling them to do the filtering without letting the skewness dimension explode. However, general state-space models like the reduced-form representations of structural economic models cannot be put into this extended form. Recently, Arellano-Valle et al. (2019) include the CSN distribution in the measurement equation, while the state transition shocks are still modelled as normally distributed. But there is abundant evidence that skewness actually stems from the innovations and not from the measurement errors; thus, their approach is not viable in a more general setting. To sum up, the numerical problems lead to the fact that real data applications in the literature are limited to univariate models only, although the underlying research questions would arguably favor a multivariate approach.

To overcome all of these barriers, our main contribution is to propose a computationally efficient way to approximate the updated distribution by reducing the skewness dimension. Our algorithm is based on the fact that a CSN distributed random variable can be defined via a conditional distribution of two normally distributed variables. Intuitively, in this representation, it is the correlation between the two random variables that introduces asymmetry and skewness. If this correlation is large, then the asymmetry of the conditional random variable, which is CSN distributed, is also large. However, if the correlation is small, then the symmetry is only slightly affected and the CSN distribution is very similar to the Gaussian distribution. In the extreme case, when there is no correlation at all, the conditional random variable will be exactly the same as a normally distributed one. In other words, the *Skewed Kalman Filter* morphs into the *Gaussian Kalman Filter*. Our approach is therefore based on a threshold, say 0.001 in absolute value, at which we discard weakly correlated elements in the skewed Kalman filtering steps as they do not distort the symmetry measurably. By doing so, we effectively reduce the overall skewness dimension by the number of discarded variables, and make the *Skewed Kalman Filter* applicable for multivariate state-space models without any restrictions. We call this algorithm the *Pruned Skewed Kalman Filter*. To motivate and derive the algorithm, we analytically show how skewness propagates through the system, which is a second contribution of this paper. Our third contribution is to derive the *Skewed Kalman Smoother* in closed-form. As far as we are aware of, we are the first to provide the closed-form expressions and, more importantly, to

implement the smoothing steps using our pruning algorithm.

We find that our algorithm works well in practice in terms of accuracy, speed, and applicability. To this end, we provide ample Monte Carlo simulation evidence in both univariate and multivariate settings to show that when the data is skewed our *Pruned Skewed Kalman* algorithm (i) filters and smooths the unobserved state vector more accurately than the conventional Kalman algorithm, (ii) is only slightly more time-consuming than the Gaussian Kalman filter for evaluating the likelihood function, and (iii) provides accurate quasi-maximum likelihood estimators for the shock parameters in finite samples. As a real data exercise, we demonstrate the usefulness of the filter by revisiting the multivariate dynamic Nelson & Siegel (1987) term structure model and estimate the latent factors for the US yield curve using Kalman filtering techniques. Estimates with the conventional Kalman filter, as in e.g. Diebold et al. (2006), yield significant skewness in the smoothed error term distribution, which propagates through the state-space system and makes the estimated latent factors skewed as well. This result contradicts the premise of linearity and Gaussianity; consequently, we show that by utilizing the CSN distribution we can not only match the skewed pattern but also our Maximum Likelihood estimation with the *Pruned Skewed Kalman Filter* is clearly favored by the data.

There are, of course, various other methodologies and algorithms for statistical inference of time series with asymmetric distributions. For instance, sequential Monte Carlo (SMC) methods can be easily adapted to skewed distributions, but the computational complexity and runtime of these filters grows rapidly as the state dimension increases. Skewness can also be modeled by a mixture of normal distributions for which there are many filtering algorithms. As recently pointed out by Nurminen et al. (2018), however, Gaussian mixtures (GM) have exponentially decaying tails and can thus be too sensitive to outlier measurements and the computational cost of a mixture reduction algorithm is expensive. Bayesian methods are frequently tailored to very specific modeling frameworks and assumptions, allowing for the fine-tuning of certain sampling algorithms, such as combining a Gibbs sampler with Metropolis-Hastings stages, as e.g. in Karlsson et al. (2021) for a VAR model. We do not claim that the *Pruned Skewed Kalman Filter* outperforms these approaches in and of itself, but we believe that the filter's ease of use, as well as its ability to be integrated into current toolboxes and estimating methodologies, will encourage and accelerate its adoption across a wide range of disciplines. Our exposition and implementation of the *Pruned Skewed Kalman Filter and Smoother* are very general and resemble the simplicity of the normal Kalman filtering and smoothing routines. In terms of modeling, empirical researchers can keep their linear state-space system and merely allow for additional flexibility by assuming the CSN distribution for the innovations rather than the normal one. In terms of computation, any estimating approach that incorporates Kalman filtering techniques, whether Bayesian or Frequentist, may be readily and simply changed by just switching out the Kalman filtering routine. For example, we already have a preliminary implementation and interface for incorporating the *Pruned Skewed Kalman Filter* into Dynare, which (among many other things) provides a toolbox to esti-

mate structural dynamic stochastic general equilibrium models with both Maximum Likelihood as well as Bayesian MCMC methods (Adjemian et al., 2022).[2] To emphasize this, we offer model-independent implementations of the *Pruned Skewed Kalman Filter and Smoother* in Julia, MATLAB, Python and R at `https://github.com/gguljanov/pruned-skewed-kalman`. Overall, we intend to provide an easy-to-use and intuitive tool for advancing empirical research in a wide range of fields where skewness is an important and integral element of the research agenda.

The remainder of the paper is organized as follows. In section 2, we review representations and properties of the CSN distribution that are needed for filtering and smoothing. Section 3 summarizes the closed-form expressions and the forward and backward recursion steps for the *Skewed Kalman Filter and Smoother*. In section 4, we first show how skewness propagates through the state-space system over time and then derive our *pruning algorithm*. Section 5 contains the Monte-Carlo evidence and section 6 our empirical application. Section 7 concludes.

2. Closed skew-normal distribution

In this section, we summarize the definition and properties of the CSN distribution. The exposition and notation follow closely González-Farías et al. (2004a), González-Farías et al. (2004b), Grabek et al. (2011) and Rezaie & Eidsvik (2014). Let $E_1 \sim N_p(0, \Sigma)$ and $E_2 \sim N_q(0, \Delta)$ be independent multivariate normally distributed random vectors. The $p \times p$ covariance matrix Σ is positive semi-definite, the $q \times q$ covariance matrix Δ is positive definite. Let μ and ν be real vectors of length p and q, respectively, and Γ a real $q \times p$ matrix. Define

$$\begin{aligned} W &= \mu + E_1 \\ Z &= -\nu + \Gamma E_1 + E_2. \end{aligned}$$

Then

$$\begin{pmatrix} W \\ Z \end{pmatrix} \sim N_{p+q}\left(\begin{bmatrix} \mu \\ -\nu \end{bmatrix}, \begin{bmatrix} \Sigma & \Sigma\Gamma' \\ \Gamma\Sigma & \Delta + \Gamma\Sigma\Gamma' \end{bmatrix} \right). \tag{1}$$

Let the random vector X have the same distribution as $W|Z \geq 0$. Then X has a closed skew-normal (CSN) distribution

$$X \sim CSN_{p,q}(\mu, \Sigma, \Gamma, \nu, \Delta).$$

[2] We plan to release this feature with Dynare 6.0.

The moment generating function (mgf) of X is

$$M_X(t) = \frac{\Phi_q(\Gamma\Sigma t; \nu, \Delta + \Gamma\Sigma\Gamma')}{\Phi_q(0; \nu, \Delta + \Gamma\Sigma\Gamma')} \exp(t'\mu + 1/2 t'\Sigma t)$$

for $t \in \mathbb{R}^p$ and $\Phi_q(\cdot; m, S)$ is the cdf of the multivariate normal distribution with expectation vector m and covariance matrix S. If the covariance matrix Σ is non-singular, then X has the probability density function

$$f_X(x; \mu, \Sigma, \Gamma, \nu, \Delta) = \frac{\Phi_q(\Gamma(x-\mu); \nu, \Delta)}{\Phi_q(0; \nu, \Delta + \Gamma\Sigma\Gamma')} \phi_p(x; \mu, \Sigma) \tag{2}$$

where ϕ_p is the pdf of a multivariate normal distribution. We do not, however, impose non-singularity in general.

Figure 1 illustrates the pdf of a univariate CSN distribution with parameters $\mu = 0$, $\Sigma = 1$, $\nu = 0$ (or $\nu = -8$), $\Delta = 1$ and different values for the shape parameter Γ. We see that, in the univariate case, the distribution is left-skewed if Γ is negative, and right-skewed if it is positive. For $\Gamma = 0$ one obtains the (symmetric) standard Gaussian distribution. Similarly, we illustrate a bivariate CSN distribution with left- and right-skewed marginals in figure 2 with the following parametrization:

$$X \sim CSN_{2,2}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix}, \Gamma, \nu, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

Note that the mean and covariance of X differ from μ and Σ unless $\Gamma = 0$ in which case the probability density of the CSN distribution reduces to the Gaussian one. Another special case is given by $CSN_{1,1}(0, 1, \gamma, 0, 1)$ which corresponds to the univariate standardized skew-normal distribution of Azzalini (1985). To summarize, μ and Σ are called the location and scale parameters of "normal dimension" p, while the dimension q is labelled "skewness dimension". Accordingly, Γ regulates skewness continuously from the normal pdf ($\Gamma = 0$) to a half normal pdf. The other skewness parameters ν and Δ are somewhat open to interpretation; however, as we outline below, they allow to establish closure of the CSN distribution under conditioning (ν), marginalization (Δ) and summation (as $\Phi_q(0; \nu, \Delta + \Gamma\Sigma\Gamma')$ is a constant).

One can see from (1) that the asymmetric deviation of the CSN distribution from the symmetric Gaussian distribution results from the covariance between W and Z; in other words, it is this correlation that adds skewness to the Gaussian distribution. Hence, the CSN distribution can be regarded as a generalization of the normal distribution and as such inherits several of its properties. In the following, we review those properties that are of special interest for the *Skewed Kalman Filter and Smoother*. Proofs can be found in González-Farías et al. (2004a) and González-Farías et al. (2004b).

Property 1 (Linear transformation, full row rank)**.**
Let $X \sim CSN_{p,q}(\mu_x, \Sigma_x, \Gamma_x, \nu_x, \Delta_x)$ and F be a real $r \times p$ matrix of rank $r \leq p$ such that $F\Sigma_x F'$ is non-

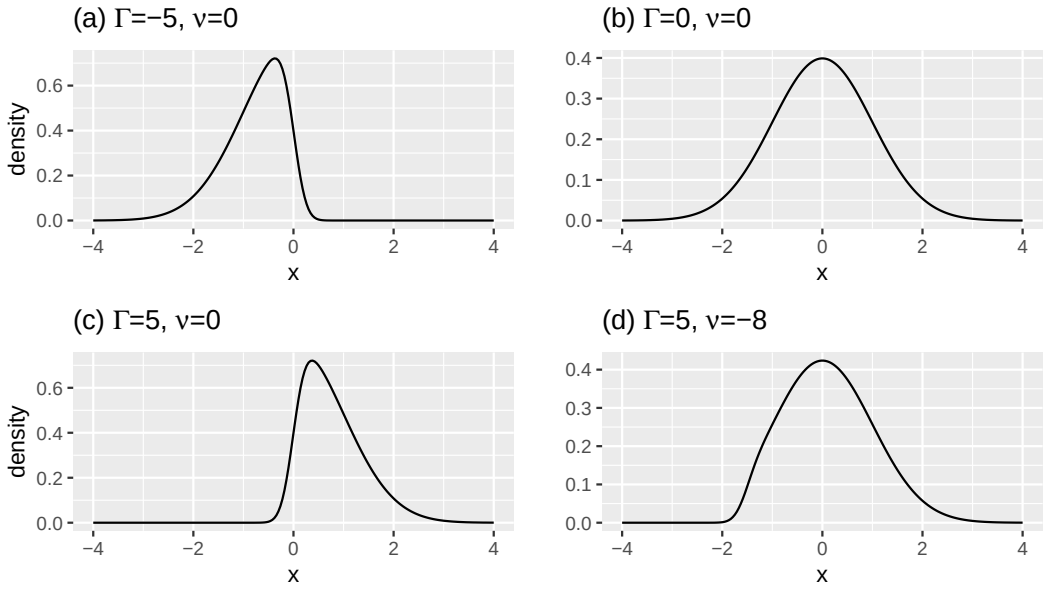

Figure 1: Density functions of univariate CSN distributions with different skewness parameters Γ and ν; other parameters are $\mu = 0$, $\Sigma = 1$ and $\Delta = 1$.

singular, then

$$Y = FX \sim CSN_{r,q}(\mu_y, \Sigma_y, \Gamma_y, \nu_y, \Delta_y)$$

with $\mu_y = F\mu_x$, $\Sigma_y = F\Sigma_x F'$, $\nu_y = \nu_x$, $\Gamma_y = \Gamma_x \Sigma_x F' \Sigma_y^{-1}$, *and* $\Delta_y = \Delta_x + \Gamma_x \Sigma_x \Gamma_x' - \Gamma_x \Sigma_x F' \Sigma_y^{-1} F \Sigma_x \Gamma_x'$.

In other words, the CSN distribution is closed under linear transformations. If F is $p \times p$ square and if both F and Σ_x have full rank p, the expressions for Γ_y and Δ_y simplify to $\Gamma_y = \Gamma_x F^{-1}$ and $\Delta_y = \Delta_x$.

Property 2 (Linear transformation, full column rank)**.**
Let $X \sim CSN_{p,q}(\mu_x, \Sigma_x, \Gamma_x, \nu_x, \Delta_x)$ *and* F *be a real* $r \times p$ *matrix with* $r > p$ *and* $rank(F) = p$, *then*

$$Y = FX \sim CSN_{r,q}(\mu_y, \Sigma_y, \Gamma_y, \nu_y, \Delta_y)$$

has a singular distribution with $\mu_y = F\mu_x$, $\Sigma_y = F\Sigma_x F'$, $\Gamma_y = \Gamma_x (F'F)^{-1} F'$, $\nu_y = \nu_x$ *and* $\Delta_y = \Delta_x$.

Property 3 (Joint distribution)**.**
Let $X \sim CSN_{p_x,q_x}(\mu_x, \Sigma_x, \Gamma_x, \nu_x, \Delta_x)$ *and* $Y \sim CSN_{p_y,q_y}(\mu_y, \Sigma_y, \Gamma_y, \nu_y, \Delta_y)$ *be independent random vectors. Then*

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \sim CSN_{p_z,q_z}(\mu_z, \Sigma_z, \Gamma_z, \nu_z, \Delta_z)$$

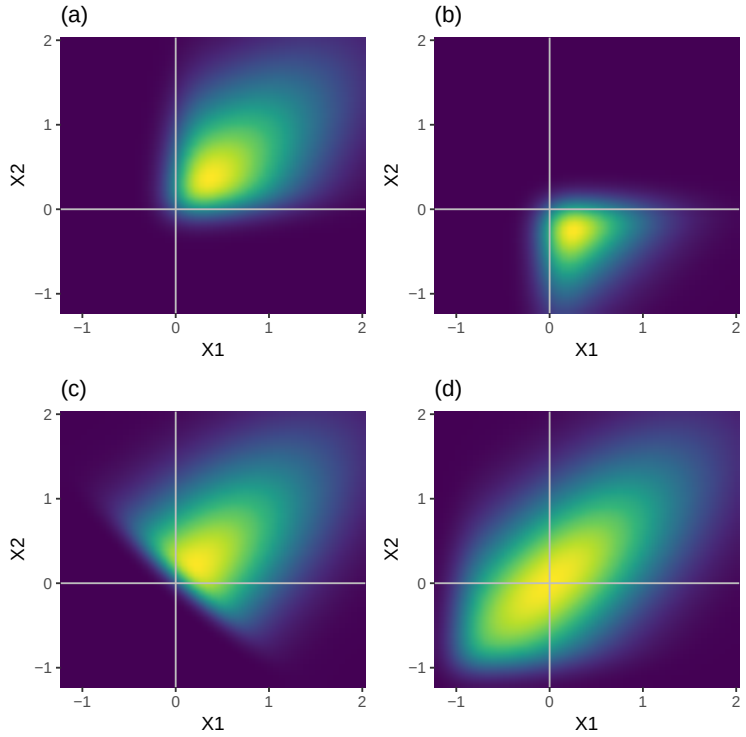

Figure 2: Density functions of bivariate CSN distributions with different skewness parameters:
(a) $\Gamma = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}$, $\nu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, (b) $\Gamma = \begin{bmatrix} 6 & 0 \\ 0 & -6 \end{bmatrix}$, $\nu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, (c) $\Gamma = \begin{bmatrix} 6 & 6 \\ 6 & 6 \end{bmatrix}$, $\nu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, (d) $\Gamma = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}$, $\nu = \begin{bmatrix} -6 \\ -6 \end{bmatrix}$.

with dimensions $p_z = p_x + p_y$, $q_z = q_x + q_y$ *and parameters*

$$\mu_z = (\mu_x', \mu_y')' \quad \Sigma_z = \begin{pmatrix} \Sigma_x & 0 \\ 0 & \Sigma_y \end{pmatrix} \quad \Gamma_z = \begin{pmatrix} \Gamma_x & 0 \\ 0 & \Gamma_y \end{pmatrix} \quad \nu_y = (\nu_x', \nu_y')' \quad \Delta_z = \begin{pmatrix} \Delta_x & 0 \\ 0 & \Delta_y \end{pmatrix}.$$

The joint distribution of independent CSN distributions is CSN again. Together with property 1 this implies that sums of independent CSN random vectors (with compatible dimensions) are CSN.

Property 4 (Summation)**.**
Let $X \sim CSN_{p,q_x}(\mu_x, \Sigma_x, \Gamma_x, \nu_x, \Delta_x)$ *and* $Y \sim CSN_{p,q_y}(\mu_y, \Sigma_y, \Gamma_y, \nu_y, \Delta_y)$ *be independent random vectors. Then*

$$Z = X + Y \sim CSN_{p,q_z}(\mu_z, \Sigma_z, \Gamma_z, \nu_z, \Delta_z)$$

with dimensions p *and* $q_z = q_x + q_y$ *and parameters*

$$\mu_z = \mu_x + \mu_y \quad \Sigma_z = \Sigma_x + \Sigma_y \quad \Gamma_z = \begin{pmatrix} \Gamma_x \Sigma_x \Sigma_z^{-1} \\ \Gamma_y \Sigma_y \Sigma_z^{-1} \end{pmatrix} \quad \nu_z = \begin{pmatrix} \nu_x \\ \nu_y \end{pmatrix} \quad \Delta_z = \begin{pmatrix} \Delta_{xx} & \Delta_{xy} \\ \Delta_{xy}' & \Delta_{yy} \end{pmatrix}$$

where

$$\Delta_{xx} = \Delta_x + \Gamma_x\Sigma_x\Gamma_x' - \Gamma_x\Sigma_x\Sigma_z^{-1}\Sigma_x\Gamma_x', \quad \Delta_{yy} = \Delta_y + \Gamma_y\Sigma_y\Gamma_y' - \Gamma_y\Sigma_y\Sigma_z^{-1}\Sigma_y\Gamma_y', \quad \Delta_{xy} = -\Gamma_x\Sigma_x\Sigma_z^{-1}\Sigma_y\Gamma_y'.$$

Note that the skewness dimension q increases when two closed skew-normal random vectors are added. While this does not matter theoretically, it turns out to be a severe numerical problem since evaluating the density function of the sum involves calculating the cdf of a higher dimensional normal distribution. For practical applications it is therefore indispensable to find a good approximation with a lower q-dimension, such as we propose in section 4.

A special case of property 4 is adding a CSN random vector $X \sim CSN_{p,q_x}(\mu_x, \Sigma_x, \Gamma_x, \nu_x, \Delta_x)$ to a normal random vector $Y \sim N(\mu_y, \Sigma_y) = CSN_{p,q_y}(\mu_y, \Sigma_y, 0, \nu_y, \Delta_y)$ of length p. For the normal distribution, the skewness parameter is $\Gamma_y = 0$ (and ν_y and Δ_y are irrelevant). Since all elements of the rows in Γ_z that belong to the normal distribution are zero, the q-dimension can be adjusted. The resulting formulas for the skewness parameters are: $\Gamma_z = \Gamma_x\Sigma_x\Sigma_z^{-1}$, $\nu_z = \nu_x$ and $\Delta_z = \Delta_x + \Gamma_x\Sigma_x\Gamma_x' - \Gamma_x\Sigma_x\Sigma_z^{-1}\Sigma_x\Gamma_x'$. Hence, $q_z = q_x$, i.e. the dimension does not increase when a normal distribution is added to a CSN distribution.

Property 5 (Conditioning)**.**
Let $X \sim CSN_{p,q}(\mu, \Sigma, \Gamma, \nu, \Delta)$ be partitioned into X_1 of length p_1 and X_2 of length p_2, such that $X = (X_1', X_2')'$. The parameters are partitioned accordingly,

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \Gamma_1 & \Gamma_2 \end{pmatrix}$$

Then

$$X_{1|2} = (X_1|X_2 = x_2) \sim CSN_{p_1,q}(\mu_{1|2}, \Sigma_{1|2}, \Gamma_{1|2}, \nu_{1|2}, \Delta_{1|2})$$

with $\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$, $\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, $\Gamma_{1|2} = \Gamma_1$, $\nu_{1|2} = \nu - (\Gamma_2 + \Gamma_1\Sigma_{12}\Sigma_{22}^{-1})(x_2 - \mu_2)$, and $\Delta_{1|2} = \Delta$.

This property establishes that conditioning some elements of a CSN random vector on its other elements in turn yields a CSN-distributed random variable.

To sum up, the CSN distribution has very attractive theoretical properties; however, its practical applicability is limited to cases where the skewness dimension q is small or moderate (say, $q < 25$). If q is large one has to evaluate the cdf of a high-dimensional multivariate normal distribution which is computationally very demanding.[3] For example, in the filtering algorithm (to be presented in the next section) the skewness

[3]MATLAB R2022b's `mvncdf` function requires that the number of dimensions must be less than or equal to 25. We rely instead on the Mendell & Elston (1974) method to evaluate the log cdf function which is quite fast and accurate, but also suffers from the *curse of increasing skewness dimension.*

dimension q naturally grows in each period of the observation window. This implies that the expressions cannot be numerically evaluated after a couple of periods since they involve multivariate normal distributions with possibly hundreds of dimensions. We will suggest a new approximation method to reduce the skewness dimension q in section 4, but first we outline the Kalman filtering and smoothing steps based on the CSN distribution.

3. Skewed Kalman Filter and Smoother

Linear state-space models are commonly used to describe physical and dynamical systems in economics, engineering and statistics. Since many real-world data applications exhibit skewness, we adapt the canonical linear state-space model by assuming that the innovations η_t in the transition equation of the state variables originate from the CSN distribution:

$$x_t = Gx_{t-1} + \eta_t, \qquad \eta_t \sim CSN_{p,q_\eta}(\mu_\eta, \Sigma_\eta, \Gamma_\eta, \nu_\eta, \Delta_\eta) \tag{3}$$

$$y_t = Fx_t + \varepsilon_t, \qquad \varepsilon_t \sim N(\mu_\varepsilon, \Sigma_\varepsilon) \tag{4}$$

where x_t is the vector of (unobserved) state variables and y_t the vector of observed variables at equally spaced time points $t = 1, \ldots, T$. The vector of observation errors ε_t is assumed to be normally distributed and independent of the CSN-distributed state variable shocks η_t. Moreover, we focus on a stable dynamic system, i.e. the characteristic roots of the parameter matrix G are inside the unit circle. In addition, we assume that the initial state x_0 (or its distribution) is known. These assumptions allow us to focus on the increasing dimensions problem in the Kalman recursions for the state variables. The pruning algorithm developed in section 4 could be easily extended to a more general initialization step or time-varying parameters. Likewise, CSN-distributed measurement errors can always be included as a structural innovation by adding an auxiliary state variable to equation (3). In fact, this simplified framework is the one that is most commonly used for the analysis of economic phenomena such as the one we study in section 6.

We denote the information set at time t by $\mathcal{F}_t$, i.e. it includes all observations up to time t and is therefore the σ-algebra generated by the observed variables $\mathcal{F}_t = \sigma(y_t, y_{t-1}, \ldots, y_1)$. The conditional distribution $x_{s|t}$ of the state variable vector x_s given the information set $\mathcal{F}_t$ is described by its CSN parameters which are denoted by $\mu_{s|t}$, $\Sigma_{s|t}$, $\Gamma_{s|t}$, $\nu_{s|t}$ and $\Delta_{s|t}$. Recursive expressions for these parameters can be derived in closed form. Rezaie & Eidsvik (2014) summarize the recursion steps which were originally developed – and coined the *Skewed Kalman Filter* – by Naveau et al. (2005). For the sake of completeness, we briefly review the prediction, updating and smoothing equations. An online appendix provides derivation of the smoothing step.

Prediction step:

Assume that $x_{t-1|t-1} \sim CSN_{p,q_{t-1}}(\mu_{t-1|t-1}, \Sigma_{t-1|t-1}, \Gamma_{t-1|t-1}, \nu_{t-1|t-1}, \Delta_{t-1|t-1})$ is given. The innovations $\eta_t \sim CSN_{p,q_\eta}(\mu_\eta, \Sigma_\eta, \Gamma_\eta, \nu_\eta, \Delta_\eta)$ are independent from $x_{t-1|t-1}$. The state transition equation (3) in conjunction with closure with respect to linear transformations (properties 1 and 2) and summation (property 4) yields the one-step predictive distribution:

$$x_{t|t-1} \sim CSN_{p,q_{t-1}+q_\eta}(\mu_{t|t-1}, \Sigma_{t|t-1}, \Gamma_{t|t-1}, \nu_{t|t-1}, \Delta_{t|t-1}) \tag{5}$$

where

$$\begin{aligned}
\mu_{t|t-1} &= G\mu_{t-1|t-1} + \mu_\eta \\
\Sigma_{t|t-1} &= G\Sigma_{t-1|t-1}G' + \Sigma_\eta \qquad (6) \\
\Gamma_{t|t-1} &= \begin{pmatrix} \Gamma_{t-1|t-1}\Sigma_{t-1|t-1}G'\Sigma_{t|t-1}^{-1} \\ \Gamma_\eta\Sigma_\eta\Sigma_{t|t-1}^{-1} \end{pmatrix} \qquad (7) \\
\nu_{t|t-1} &= \begin{pmatrix} \nu_{t-1|t-1} \\ \nu_\eta \end{pmatrix} \\
\Delta_{t|t-1} &= \begin{pmatrix} \Delta_{t|t-1}^{11} & \Delta_{t|t-1}^{12} \\ (\Delta_{t|t-1}^{12})' & \Delta_{t|t-1}^{22} \end{pmatrix} \qquad (8)
\end{aligned}$$

with

$$\begin{aligned}
\Delta_{t|t-1}^{11} &= \Delta_{t-1|t-1} + \Gamma_{t-1|t-1}\Sigma_{t-1|t-1}\Gamma'_{t-1|t-1} - \Gamma_{t-1|t-1}\Sigma_{t-1|t-1}G'\Sigma_{t|t-1}^{-1}G\Sigma_{t-1|t-1}\Gamma'_{t-1|t-1} \\
\Delta_{t|t-1}^{22} &= \Delta_\eta + \Gamma_\eta\Sigma_\eta\Gamma'\eta - \Gamma_\eta\Sigma_\eta\Sigma_{t|t-1}^{-1}\Sigma_\eta\Gamma'\eta, \qquad \Delta_{t|t-1}^{12} = -\Gamma_{t-1|t-1}\Sigma_{t-1|t-1}G'\Sigma_{t|t-1}^{-1}\Sigma_\eta\Gamma'\eta
\end{aligned}$$

Updating step:

From the prediction step, it is known that $x_{t|t-1}$ is CSN distributed. The measurement equation (4) implies that the conditional distribution of y_t given $\mathcal{F}_{t-1}$ is also CSN distributed since it is the sum of a linear transformation of $x_{t|t-1}$ and a normal distribution. Due to property 5 (closure with respect to conditioning), the updated distribution $x_{t|t}$ (i.e. the distribution of x_t given $\mathcal{F}_{t-1}$ and also y_t, or in short, given $\mathcal{F}_t$) is

$$x_{t|t} \sim CSN_{p,q_t}(\mu_{t|t}, \Sigma_{t|t}, \Gamma_{t|t}, \nu_{t|t}, \Delta_{t|t}) \tag{9}$$

where $q_t = q_{t-1} + q_\eta$ and

$$\begin{aligned}
\mu_{t|t} &= \mu_{t|t-1} + \Sigma_{t|t-1}F'(F\Sigma_{t|t-1}F' + \Sigma_\varepsilon)^{-1}(y_t - F\mu_{t|t-1} - \mu_\varepsilon) \\
\Sigma_{t|t} &= \Sigma_{t|t-1} - \Sigma_{t|t-1}F'(F\Sigma_{t|t-1}F' + \Sigma_\varepsilon)^{-1}F\Sigma_{t|t-1} \qquad (10)
\end{aligned}$$

$$\Gamma_{t|t} = \Gamma_{t|t-1} \tag{11}$$

$$\nu_{t|t} = \nu_{t|t-1} - \Gamma_{t|t-1}\Sigma_{t|t-1}F'(F\Sigma_{t|t-1}F' + \Sigma_\varepsilon)^{-1}(y_t - F\mu_{t|t-1} - \mu_\varepsilon)$$

$$\Delta_{t|t} = \Delta_{t|t-1}. \tag{12}$$

The updating step consists of two parts, (i) a Gaussian part which updates $\mu_{t|t}$ and $\Sigma_{t|t}$ using the *Gaussian Kalman Gain* $K_{t-1}^{Gauss} = \Sigma_{t|t-1}F'(F\Sigma_{t|t-1}F' + \Sigma_\varepsilon)^{-1}$ and (ii) a skewed part which updates the skewness parameters using the *Skewed Kalman Gain* $K_{t-1}^{Skewed} = \Gamma_{t|t-1}K_{t-1}^{Gauss}$. In our setting the only skewness parameter that is updated in the updating step is $\nu_{t|t-1}$, the parameters $\Gamma_{t|t-1}$ and $\Delta_{t|t-1}$ are not affected because the measurement errors are Gaussian. Obviously, without skewness, $\Gamma_{t|t-1} = 0$, the prediction and updating steps would be equivalent to the ones from the conventional Gaussian Kalman filter. With skewness, however, we see that the skewness dimension q_t in (5) and (9) increases in each period, because two CSN distributed random variables are added.

> *This means that the skewness dimension explodes as the recursion proceeds over many time steps. As a result the matrix dimensions grow, parameter estimation gets more complicated, sampling is harder, and so on. Thus, for practical purposes we need to assume simplified conditions (Rezaie & Eidsvik, 2014, p. 5).*

However, instead of simplifying the conditions or imposing more stringent assumptions, we suggest an approximation method to shrink the skewness dimension in section 4.

Smoothing:

Often, we are not only interested in the filtered distributions ($x_{t|t}$) but also in the smoothed distributions ($x_{t|T}$), i.e. estimates of the state variables that take into consideration all available observations $y_1, \ldots, y_T$. In the last period the filtered and smoothed distributions obviously coincide. The smoothed distributions for $t = T-1, \ldots, 1$ can be calculated in a backward recursion. Chiplunkar & Huang (2021) present recursion formulas for a special case involving a non-stationary (random walk) latent variable. Adapting their approach, we present recursion formulas for the general state-space model (3) and (4) with CSN distributed innovations. As far as we know, we are the first to do so in this general setting. For ease of notation we define the following abbreviations:

$$M_t = \Sigma_{t+1|T}\Sigma_{t+1|t}^{-1}G\Sigma_{t|t}\Sigma_{t|T}^{-1}$$

$$N_t = -\Gamma_\eta G + \Gamma_\eta M_t.$$

Further, let $O_{T-1}, O_{T-2}, \ldots$ be a sequence of matrices of increasing row dimensions, such that $O_{T-1} = N_{T-1}$

and, for $t = T-2, T-3, \ldots, 1$,

$$O_t = \begin{bmatrix} N_t \\ O_{t+1} M_t \end{bmatrix}.$$

The CSN parameters of $x_t|\mathcal{F}_T \sim CSN_{p,q_T}(\mu_{t|T}, \Sigma_{t|T}, \Gamma_{t|T}, \nu_{t|T}, \Delta_{t|T})$ for $t = T-1, \ldots, 1$ are

$$\begin{aligned}
\mu_{t|T} &= \mu_{t|t} + \Sigma_{t|t} G' \Sigma_{t+1|t}^{-1} (\mu_{t+1|T} - \mu_{t+1|t}) \\
\Sigma_{t|T} &= \Sigma_{t|t} + \Sigma_{t|t} G' \Sigma_{t+1|t}^{-1} (\Sigma_{t+1|T} - \Sigma_{t+1|t}) \Sigma_{t+1|t}^{-1} G \Sigma_{t|t} \\
\Gamma_{t|T} &= \begin{pmatrix} \Gamma_{t|t} \\ O_t \end{pmatrix} \\
\nu_{t|T} &= \nu_{T|T} \\
\Delta_{t|T} &= \begin{pmatrix} \Delta_{t|t} & 0 \\ 0 & \tilde{\Delta}_t \end{pmatrix}.
\end{aligned}$$

with

$$\tilde{\Delta}_t = \begin{pmatrix} \Delta_\eta & 0 \\ 0 & \tilde{\Delta}_{t+1} \end{pmatrix} + \begin{pmatrix} \Gamma_\eta \\ O_{t+1} \end{pmatrix} (\Sigma_{t+1|T} - M_t \Sigma_{t|T} M_t') \begin{pmatrix} \Gamma_\eta \\ O_{t+1} \end{pmatrix}'$$

for $t = T-2, T-3, \ldots, 1$ and $\tilde{\Delta}_{T-1} = \Delta_\eta + \Gamma_\eta (\Sigma_{t+1|T} - M_t \Sigma_{t|T} M_t') \Gamma_\eta'$. The proof is sketched in the online appendix. Notice that the skewness dimension remains constant (at q_T) during the backward recursion. In particular, the skewness parameter $\nu_{t|T}$ is always equal to $\nu_{T|T}$ for all t. At each iteration, the row dimension of $\Gamma_{t|t}$ decreases. This decrease is offset by an increase in the row dimension of O_t. In a similar fashion, the top left block of the block-diagonal matrix $\Delta_{t|T}$ gets smaller in each iteration, while the bottom right matrix inflates such that the dimension of $\Delta_{t|T}$ does not change. Similarly to filtering, whether or not smoothing is computationally feasible, depends largely on the overall skewness dimension. Hence, a way to reduce it is also important from a smoothing perspective.

4. Pruning the skewness dimension

Our approach to reduce the skewness dimension is motivated by the characterization (1) of the CSN distribution. Evidently, if there is no correlation between W and Z, the CSN distribution is equal to a Gaussian distribution and the skewed Kalman filter morphs into the Gaussian one. Therefore if some elements of Z are only weakly correlated with the elements of W, we can prune, i.e. dispose of those elements in Z, as there is no palpable effect on the skewness behavior. Algorithm 1 outlines the pseudo-code of our pruning algorithm.

Algorithm 1 (Pruning Algorithm)**.** *The algorithm consists of the following steps, given skewness parameters* Σ*,* Γ*,* ν*,* Δ *and pruning threshold* `tol`*.*

1. *Construct and partition the covariance matrix*

$$P = \begin{pmatrix} P_1 & P_2' \\ P_2 & P_4 \end{pmatrix} = \begin{pmatrix} \Sigma & \Sigma \cdot \Gamma' \\ \Gamma \cdot \Sigma & \Delta + \Gamma \cdot \Sigma \cdot \Gamma' \end{pmatrix} \tag{13}$$

2. *Transform* P *into a correlation matrix* $R = \begin{pmatrix} R_1 & R_2' \\ R_2 & R_4 \end{pmatrix}$
3. *Find the maximum absolute value along each row of* $\texttt{abs}(R_2)$*. Save it as vector* `max_val`*.*
4. *Delete the rows of* $\begin{pmatrix} P_2 & P_4 \end{pmatrix}$ *and columns of* $\begin{pmatrix} P_2' \\ P_4 \end{pmatrix}$ *corresponding to* (`max_val` $<$ `tol`)*. Save as* $\widetilde{P}$*.*
5. *Compute pruned* ν *by removing rows corresponding to* (`max_val` $<$ `tol`)*.*
6. *Compute pruned* $\Gamma = \widetilde{P}_2 \Sigma^{-1}$*.*
7. *Compute pruned* $\Delta = \widetilde{P}_4 - \Gamma \widetilde{P}_2'$*.*
8. *Return pruned skewness parameters* Γ*,* ν*, and* Δ*.*

To illustrate the procedure numerically consider the following univariate example:

$$x_{t,t-1} \sim CSN\left(0, 1, \begin{pmatrix} 6 \\ 0.1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0.1 \\ -0.1 & 1 \end{pmatrix}\right) \tag{14}$$

with a skewness dimension of 2. Applying pruning algorithm 1 with a (rather large) pruning tolerance $\texttt{tol} = 0.1$, we get

$$R = \begin{pmatrix} 1.0000 & 0.9864 & 0.0995 \\ 0.9864 & 1.0000 & 0.0981 \\ 0.0995 & 0.0981 & 1.0000 \end{pmatrix}$$

Clearly $0.9864 > tol$, but $0.0995 < tol$, so we can reduce the skewness dimension by 1. Recomputing the new skewness parameters ($\nu = 0$, $\Gamma = 6 \cdot 1^{-1}$, $\Delta = 37 - 6 \cdot 6$), we get the approximating distribution $CSN(0, 1, 6, 0, 1)$. Figure 3 depicts the pdf and cdf of the original and the approximating distributions; the difference is hardly discernible despite the rather large pruning threshold of 0.1.

Of course, the skewness dimension can only be reduced if the correlation coefficients are sufficiently small. We now proceed to show that the added skewness dimensions induced by the prediction steps of the *Skewed Kalman Filter* will fade away over time. In other words, even though the skewness dimension grows over time, many of the dimensions will eventually be redundant and can be removed when the density function

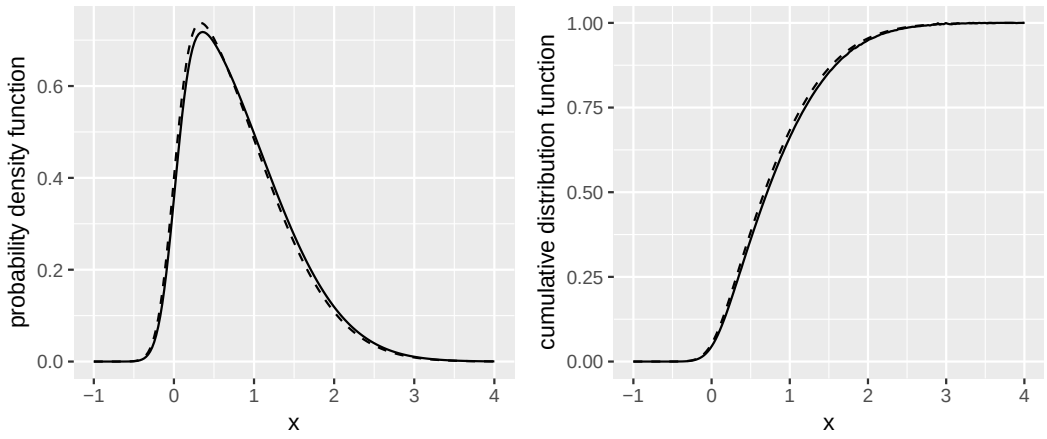

Figure 3: Probability density functions and cumulative distribution functions of a CSN distributed random variable with two skewness dimensions (skewness parameters as given in (14), solid lines) and the approximating $CSN(0, 1, 6, 0, 1)$ distribution with one skewness dimension (dashed lines).

(or the log-likelihood function) needs to be numerically evaluated. Assume that the recursion is anchored at a given initial distribution with parameters $\mu_{0|0}$, $\Sigma_{0|0}$, $\Gamma_{0|0}$, $\nu_{0|0}$, $\Delta_{0|0}$. We first focus on the recursion for the skewness parameter $\Gamma_{t|t-1}$ in (7) and (11), with $\Sigma_{t|t-1}$ as given in (6). Since $\Gamma_{t-1|t-1}$ appears in the upper row in (7), the number of rows increases at each step. For instance, in period $t = 4$ we would obtain

$$\Gamma_{4|4} = \begin{pmatrix} \Gamma_{0|0}\Sigma_{0|0}G'\Sigma_{1|0}^{-1}\Sigma_{1|1}G'\Sigma_{2|1}^{-1}\Sigma_{2|2}G'\Sigma_{3|2}^{-1}\Sigma_{3|3}G'\Sigma_{4|3}^{-1} \\ \Gamma_{\eta}\Sigma_{\eta}\Sigma_{1|0}^{-1}\Sigma_{1|1}G'\Sigma_{2|1}^{-1}\Sigma_{2|2}G'\Sigma_{3|2}^{-1}\Sigma_{3|3}G'\Sigma_{4|3}^{-1} \\ \Gamma_{\eta}\Sigma_{\eta}\Sigma_{2|1}^{-1}\Sigma_{2|2}G'\Sigma_{3|2}^{-1}\Sigma_{3|3}G'\Sigma_{4|3}^{-1} \\ \Gamma_{\eta}\Sigma_{\eta}\Sigma_{3|2}^{-1}\Sigma_{3|3}G'\Sigma_{4|3}^{-1} \\ \Gamma_{\eta}\Sigma_{\eta}\Sigma_{4|3}^{-1} \end{pmatrix}.$$

This matrix has dimension $(4q_\eta + q_0) \times p$ where p is the number of state variables, q_η is the skewness dimension of the state shocks and q_0 is the skewness dimension of the initial distribution. To find a general expression for any period t, define $L_t \equiv \Sigma_{t|t-1}^{-1}\Sigma_{t|t}G'$. Then,

$$\Gamma_{t|t} = \begin{pmatrix} \Gamma_{0|0}\Sigma_{0|0}G'\prod_{j=1}^{t-1}L_j \\ \Gamma_{\eta}\Sigma_{\eta}\prod_{j=1}^{t-1}L_j \\ \Gamma_{\eta}\Sigma_{\eta}\prod_{j=2}^{t-1}L_j \\ \vdots \\ \Gamma_{\eta}\Sigma_{\eta}\prod_{j=t}^{t-1}L_j \end{pmatrix}\Sigma_{t|t-1}^{-1} \tag{15}$$

where the empty product in the last row is defined as $\prod_{j=t}^{t-1}L_j \equiv 1$. The matrices L_t are closely related to the updating step: multiplying both sides of (10) by G from the left and by $\Sigma_{t|t-1}^{-1}$ from the right, we obtain

the transpose of L_t:

$$G\Sigma_{t|t}\Sigma_{t|t-1}^{-1} = G - G\Sigma_{t|t-1}F'(F\Sigma_{t|t-1}F' + \Sigma_\varepsilon)^{-1}F$$

As $t \to \infty$, the sequence $G\Sigma_{t|t}\Sigma_{t|t-1}^{-1}$ converges to a constant matrix with all eigenvalues inside the unit circle (Hamilton, 1994, prop. 13.1 and 13.2). The same is true for L_t as it is just the transpose of $G\Sigma_{t|t}\Sigma_{t|t-1}^{-1}$. This implies that the product terms $\prod_j L_j$ in (15) will fade away as new rows are appended at the bottom in every period. The rows at the top (i.e. those relating to older shocks) will fade away more quickly. Hence, the impact of the shocks on the skewness parameter $\Gamma_{t|t}$ (which according to (11) also equals $\Gamma_{t|t-1}$) is not persistent.

Next, we turn to the skewness parameter $\Delta_{t|t}$, which is equal to $\Delta_{t|t-1}$ according to (12). The recursions in (8) imply that the dimension of $\Delta_{t|t}$ grows each period. The top left element of the partitioned matrix (7) shows that the matrix

$$\begin{aligned}\Gamma_{t-1|t-1}\Sigma_{t-1|t-1}\Gamma'_{t-1|t-1} - \Gamma_{t-1|t-1}\Sigma_{t-1|t-1}G'\Sigma_{t|t-1}^{-1}G\Sigma_{t-1|t-1}\Gamma'_{t-1|t-1} \\ = \Gamma_{t-1|t-1}\Sigma_{t-1|t-1}^{1/2}(I - \Sigma_{t-1|t-1}^{1/2}G'\Sigma_{t|t-1}^{-1}G\Sigma_{t-1|t-1}^{1/2})\Sigma_{t-1|t-1}^{1/2}\Gamma'_{t-1|t-1} \quad (16)\end{aligned}$$

is added to $\Delta_{t-1|t-1}$ in each iteration. To show that it is positive definite consider the matrix

$$S \equiv \begin{pmatrix} I & \Sigma_{t-1|t-1}^{1/2}G' \\ G\Sigma_{t-1|t-1}^{1/2} & \Sigma_{t|t-1} \end{pmatrix}.$$

Since both I and $\Sigma_{t|t-1} - G\Sigma_{t-1|t-1}^{1/2}I^{-1}\Sigma_{t-1|t-1}^{1/2}G' = \Sigma_\eta$ (see (6) in the prediction step) are positive definite, so is S (Horn & Johnson, 2017, theor. 7.7.7). Using Gallier (2010, prop. 2.1) we can conclude that $(I - \Sigma_{t-1|t-1}^{1/2}G'\Sigma_{t|t-1}^{-1}G\Sigma_{t-1|t-1}^{1/2})$ is also positive definite. Hence, matrix (16) is also positive definite. As positive definite matrices have strictly positive diagonal elements, the diagonal elements of $\Delta_{t|t}$ keep growing over time. Algorithm 1 reduces the skewness dimension based on the covariances in the bottom left (or top right) partition of the covariance matrix P in (13), i.e. $P_2 \equiv \Gamma_{t|t}\Sigma_{t|t}$. We focus on the (i,j)-th element P_2^{ij}, the corresponding correlation is

$$R_2^{ij} = \frac{{P_2}^{ij}}{\sqrt{\Sigma_{t|t}^{ii}}\sqrt{\Delta_{t|t}^{jj}}}.$$

As we have shown above, each element of $\Gamma_{t|t}$ matrix decreases as t increases. Further, it is a standard result of the (steady-state) Kalman filter that each element of $\Sigma_{t|t}$ converges (rather quickly) to a constant. Therefore, P_2^{ij} decreases as t increases. But, Δ^{jj} increases as time passes due to our previous calculations. All of these result in shrinking R_2^{ij} over time. To summarize, the algorithm is guaranteed to reduce the skewness dimension after sufficiently many periods.

The same line of thought can also be applied to the parameters of the prediction, i.e. to $P_2 \equiv \Gamma_{t|t-1}\Sigma_{t|t-1}$

and $R_2^{ij} = \frac{{P_2}^{ij}}{\sqrt{\Sigma^{ii}_{t|t-1}}\sqrt{\Delta^{jj}_{t|t-1}}}$.

5. A Monte Carlo Study

In this section we conduct a thorough Monte Carlo study to evaluate the performance of the *Pruned Skewed Kalman Filter and Smoother* in terms of accuracy and speed. To this end, we consider three different state-space models as data-generating processes (DGP). DGP (1) is a univariate model given by the following parametrization:

$$G = 0.8, \qquad F = 10, \qquad \mu_\varepsilon = 1, \qquad \Sigma_\varepsilon = 0.01, \qquad \mu_\eta = 0.3, \qquad \Sigma_\eta = 0.64, \qquad \nu_\eta = 0, \qquad \lambda_\eta = -0.89$$

DGP (1)

DGP (2) is a multivariate model with four state and three observable variables:

$$G = \begin{pmatrix} 0.5488 & 0.1738 & -0.2949 & 0.1534 \\ -0.2864 & 0.1060 & 0.3628 & 0.3334 \\ -0.3898 & -0.0252 & 0.5339 & 0.3163 \\ 0.2389 & 0.1958 & -0.0027 & 0.5519 \end{pmatrix} \qquad F = \begin{pmatrix} -0.7196 & 0.8221 & 0.4602 & -0.6412 \\ -2.0887 & -0.8201 & -1.2380 & 0.3937 \\ 0.6347 & -0.5109 & 0.8476 & 0.6819 \end{pmatrix}$$

$$\Sigma_\varepsilon = \begin{pmatrix} 0.0108 & -0.0276 & -0.0314 \\ -0.0276 & 0.1129 & -0.0025 \\ -0.0314 & -0.0025 & 0.2889 \end{pmatrix} \cdot 10^{-6} \qquad \mu_\varepsilon = \begin{pmatrix} 0.8565 \\ -0.3010 \\ -0.82705 \end{pmatrix}$$

$$\Sigma_\eta = \begin{pmatrix} 0.0013 & -0.0111 & 0.0116 & -0.0089 \\ -0.0111 & 0.1009 & -0.2301 & 0.1014 \\ 0.0116 & -0.2301 & 3.3198 & -1.0618 \\ -0.0089 & 0.1014 & -1.0618 & 1.0830 \end{pmatrix} \qquad \mu_\eta = \begin{pmatrix} 0.3455 \\ -1.8613 \\ 0.7765 \\ -0.5964 \end{pmatrix} \qquad \nu_\eta = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \lambda_\eta = 0.89$$

DGP (2)

DGP (3) is a multivariate model with three state variables that are all observable:

$$G = \begin{pmatrix} 0.9969 & 0.1256 & -0.4803 \\ -0.8221 & 0.0386 & 0.6687 \\ 0.5605 & 0.6397 & -0.4333 \end{pmatrix} \quad F = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \Sigma_\varepsilon = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot 10^{-4} \quad \mu_\varepsilon = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Sigma_\eta = \begin{pmatrix} 0.64 & 0 & 0 \\ 0 & 0.36 & 0 \\ 0 & 0 & 0.49 \end{pmatrix} \quad \Gamma_\eta = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -6 \end{pmatrix} \quad \Delta_\eta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mu_\eta = \begin{pmatrix} 0.3 \\ -0.1 \\ 0.2 \end{pmatrix} \quad \nu_\eta = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

DGP (3)

Note that in DGP (1) and DGP (2) we introduce the auxiliary hyperparameter $\lambda_\eta \in]-1;1[$ to re-parametrize the skewness parameters according to the following relationships:

$$\Gamma_\eta = \lambda_\eta {\Sigma_\eta}^{-1/2}, \qquad \Delta_\eta = (1-\lambda_\eta^2)\mathbf{I}_{n_\eta}$$

In this case, we can simplify $\Delta_\eta + {\Gamma_\eta}' \Sigma_\eta \Gamma_\eta$ to the identity matrix I_{n_η} such that the unconditional expectation vector and the covariance matrix of η_t can be calculated in closed-form (Flecher et al., 2009):

$$E[\eta_t] = \mu_\eta + \left(\sqrt{\frac{2}{\pi}} \lambda_\eta {\Sigma_\eta}^{1/2} \right) \mathbf{1}_{n_\eta}, \qquad V[\eta_t] = \Sigma_\eta \left(1 - \frac{2}{\pi} \lambda_\eta^2 \right) \tag{17}$$

Arellano-Valle & Azzalini (2008) and Käärik et al. (2015) provide related discussions on the usefulness of this re-parameterization for the skew-normal distribution.[4] Following a suggestion of Harvey & Phillips (1979), the initial distribution is set to a normal one with an initial covariance matrix with 10 on the diagonal: $x_{0|0} \sim CSN(0, 10I_{n_x}, 0, 0, I_{n_x}) = N(0, 10I_{n_x})$. Lastly, to compute the likelihood function, we make use of the standard predictive decomposition based on the conditional distribution of y_t given y_{t-1}:

$$y_t | y_{t-1} \sim \; CSN(\hat{y}_{t|t-1},\; \Omega_{t|t-1},\; K_{t-1}^{Skewed},\; \nu_{t|t-1},\; \Delta_{t|t-1} + (\Gamma_{t|t-1} - K_{t-1}^{Skewed} F) \Sigma_{t|t-1} \Gamma'_{t|t-1})$$

where $\hat{y}_{t|t-1} = F\mu_{t|t-1} + \mu_\epsilon$ is the predicted value and $\Omega_{t|t-1} = F\Sigma_{t|t-1}F' + \Sigma_\epsilon$ the prediction-error covariance matrix of the Gaussian Kalman filter.

5.1. State estimation

For forecasting, it is helpful to condense the filtered distribution $x_{t|t}$ into a point estimator. Since the CSN distribution is asymmetric, the expectation $E[x_{t|t}]$ is one potential, but not necessarily the best point estimator. Let $L[\tilde{x}_t, x_t]$ denote the loss function for a point estimator $\tilde{x}_t$ if the true value is x_t. Depending on the loss function, different point estimators will minimize the expected loss. Of course, if the loss function is quadratic, i.e. $L_2[\tilde{x}_t, x_t] = (\tilde{x}_t - x_t)^2$, the expected loss is minimal if $\tilde{x}_t = E[x_{t|t}]$. If the loss function is $L_1[\tilde{x}_t, x_t] = |\tilde{x}_t - x_t|$, the best point estimator is the median of $x_{t|t}$. And the asymmetric loss function

$$L_q[\tilde{x}_t, x_t] = \begin{cases} a|\tilde{x}_t - x_t| & \text{for } x_t > \tilde{x}_t \\ b|\tilde{x}_t - x_t| & \text{for } x_t \leq \tilde{x}_t \end{cases}$$

results in the $a/(a+b)$-quantile of $x_{t|t}$ as point estimator. A similar discussion applies to the smoothed states $x_{t|T}$.

We start by simulating $R = 2400$ sample paths for x_t and y_t of different length $T = \{40, 80, 110\}$ (plus a burn-in of 100 periods). The shocks η_t are drawn from the CSN distribution, whereas the measurement errors ε_t are drawn from the normal distribution. To assess how well the *Pruned Skewed Kalman Filter and Smoother* estimate the unobserved state variables in comparison to the conventional *Skewed* or *Gaussian*

[4] This is without loss of generality. We mainly use this to quickly compute $E[\eta_t]$ and $V[\eta_t]$ as we use these values as input parameters for the *Gaussian Kalman Filter*. In fact, our replication codes contain functions to compute the unconditional mean and the covariance matrix for any parameterization of the multivariate CSN distribution. Moreover, both in the last simulation exercise as well as in our empirical application we do not use this re-parametrization.

Kalman Filters and Smoothers, we compute the expected losses. That is, for each sample $r = 1, ..., R$ the loss is computed as

$$\text{Loss}^{(r)} := \sum_{t=20}^{T} L[\tilde{x}_t^{(r)}, x_t^{(r)}]$$

where in the univariate case L is any of the three loss functions L_1, L_2 and L_3 (with $a = 1$ and $b = 4$) under consideration, while in the multivariate case we focus only on L_2. Note that in order to avoid too large an impact of the initial distribution $x_{0|0}$, the losses are calculated after a burn-in phase of 20 periods. The expected loss is then estimated by averaging over all replications

$$\text{Expected Loss} = \frac{1}{R} \sum_{r=1}^{R} \text{Loss}^{(r)}.$$

Tables 1 and 2 report the *Expected Loss* and the 5th and 95th percentiles of *Loss*$^{(r)}$ of our Monte-Carlo simulation exercise for the different variants of both filters and smoothers. Three things are worth pointing out. First, the *Skewed Kalman Filter and Smoother* are superior to the *Gaussian Kalman Filter and Smoother* in all cases. Even though the better performance is rather small in the univariate case, it becomes really measurable in the multivariate case. This is not surprising, since the closed skew-normal distribution deviates only mildly from symmetry and normality (Liseo & Parisi, 2013) and the conventional Kalman filter and smoother are naturally optimal in its domain, i.e. when data is very close to normal. Nevertheless, in the more general case, the conventional Kalman filter and smoother simply neglect the skewed behavior; while the *Skewed Kalman Filter* embeds normality as a special case. Second, our pruning algorithm is very accurate and numerically almost equivalent to the conventional and un-pruned *Skewed Kalman Filter* (up to the twelfth digit in the univariate case and up to the 5th digit in the multivariate case). Third, the pruning threshold does not matter measurably in the univariate case and makes only a small numerical difference in multivariate settings. Clearly, the closer the tolerance is to 0, i.e. to the un-pruned filter and smoother, the more accurate we estimate the states. However, as we have argued above the un-pruned version of the filter and smoother is only feasible in the univariate case, while in multivariate settings we manage to deal with the numerical challenges for very small sample sizes only. Our pruning algorithm, on the other hand, is able to overcome this problem. Even with very low tolerance thresholds we are able to compute the filtering and smoothing steps without running into the curse of increasing skewness dimensionality. We conclude that overall both the *Pruned Skewed Kalman Filter and Smoother* perform well in terms of accuracy. However, there is a trade-off between accuracy and speed, which we analyze next.

5.2. Computing time

The performance of the *Pruned Skewed Kalman Filter* should also be evaluated in relation to its computing time. Table 3 reports the time in **seconds** required to compute 1000 evaluations of the log-likelihood

			Gaussian	Pruned Skewed Kalman Filter			
DGP	T	Loss	Kalman Filter	no pruning	tol=1e-6	tol=1e-4	tol=1e-2
(1)	40	L_1	0.166535064 [0.1236;0.2142]	0.166527070 [0.1235;0.2144]	0.166527070 [0.1235;0.2144]	0.166527070 [0.1235;0.2144]	0.166527072 [0.1235;0.2144]
(1)	40	L_2	0.002080850 [0.0012;0.0033]	0.002080707 [0.0012;0.0033]	0.002080707 [0.0012;0.0033]	0.002080707 [0.0012;0.0033]	0.002080707 [0.0012;0.0033]
(1)	40	L_a	0.293173611 [0.2168;0.3852]	0.293160284 [0.2167;0.3852]	0.293160284 [0.2167;0.3852]	0.293160284 [0.2167;0.3852]	0.293160278 [0.2167;0.3852]
(1)	80	L_1	0.486149061 [0.4151;0.5639]	0.486122296 [0.4151;0.5638]	0.486122296 [0.4151;0.5638]	0.486122296 [0.4151;0.5638]	0.486122299 [0.4151;0.5638]
(1)	80	L_2	0.006087964 [0.0045;0.0080]	0.006087485 [0.0045;0.0080]	0.006087485 [0.0045;0.0080]	0.006087485 [0.0045;0.0080]	0.006087485 [0.0045;0.0080]
(1)	80	L_a	0.853449643 [0.7129;1.0017]	0.853414306 [0.7127;1.0008]	0.853414306 [0.7127;1.0008]	0.853414306 [0.7127;1.0008]	0.853414297 [0.7127;1.0008]
(1)	110	L_1	0.724620237 [0.6368;0.8186]	0.724598667 [0.6367;0.8185]	0.724598667 [0.6367;0.8185]	0.724598667 [0.6367;0.8185]	0.724598666 [0.6367;0.8185]
(1)	110	L_2	0.009073018 [0.0071;0.0113]	0.009072577 [0.0071;0.0113]	0.009072577 [0.0071;0.0113]	0.009072577 [0.0071;0.0113]	0.009072577 [0.0071;0.0113]
(1)	110	L_a	1.272405929 [1.1026;1.4533]	1.272363075 [1.1024;1.4536]	1.272363075 [1.1024;1.4536]	1.272363075 [1.1024;1.4536]	1.272363052 [1.1024;1.4536]
(2)	40	L_2	4.23932054 [2.1343;6.9488]	4.17299006 [2.1172;6.9381]	4.17299000 [2.1172;6.9381]	4.17298994 [2.1173;6.9382]	4.17450665 [2.0989;6.9805]
(2)	80	L_2	12.30937668 [8.4001;17.0700]	-	12.11085307 [8.3181;16.9039]	12.11085498 [8.3181;16.9048]	12.11665912 [8.3003;16.9054]
(2)	110	L_2	18.39547677 [13.4673;24.0186]	-	18.10271658 [13.1829;23.6744]	18.10272323 [13.1834;23.6743]	18.11208988 [13.2441;23.6814]

Table 1: Expected losses for filtered states, lower is better. 5th and 95th percentiles in square brackets.

function of univariate DGP (1) and multivariate DGP (2) for different sample sizes. We can see that as the number of observation periods grows, it takes more time to evaluate the likelihood function in all cases. The curse of increasing dimensionality inherent in the un-pruned *Skewed Kalman Filter* becomes apparent. Even though we are able to evaluate the Gaussian cdfs of increasing dimension in the univariate case, this comes at a cost as the computational time increases exponentially. In the multivariate case, the calculations are still feasible in principle for small sample sizes, but explode relatively quickly for medium to large sample sizes or require an unreasonably huge amount of computational time and memory. This becomes even more severe if we increase the dimension of the state-space system matrices which is rather likely for real data applications.

The proposed *Pruned Skewed Kalman Filter* does not suffer from this and performs convincingly well. It is only slightly affected by a growing sample size; relatively speaking, it behaves very similar to the conventional Kalman filter in this regard. That is, the relative time increase between a sample size of 50 and 250 is approximately 3.93 for the univariate *Gaussian Kalman Filter*, whereas for the *Pruned Skewed Kalman Filter* we get a factor of approximately 4.15 for a very tight pruning threshold of 1e-6, and 4.25 for a very rough tolerance of 1e-2. In absolute terms, using 1e-2 as the tolerance level is 1.5 times faster than using 1e-6 as the tolerance level. In multivariate settings, the results are similar. The average time needed

			Gaussian	Pruned Skewed Kalman Smoother			
DGP	T	Loss	Kalman Smoother	no pruning	tol=1e-6	tol=1e-4	tol=1e-2
(1)	40	L_1	0.166539615 [0.1239;0.2143]	0.166528589 [0.1236;0.2145]	0.166528589 [0.1236;0.2145]	0.166528593 [0.1236;0.2145]	0.166528593 [0.1236;0.2145]
(1)	40	L_2	0.002080868 [0.0012;0.0033]	0.002080555 [0.0012;0.0033]	0.002080555 [0.0012;0.0033]	0.002080555 [0.0012;0.0033]	0.002080555 [0.0012;0.0033]
(1)	40	L_a	0.293186237 [0.2162;0.3848]	0.293159668 [0.2163;0.3847]	0.293159668 [0.2163;0.3847]	0.293159662 [0.2163;0.3847]	0.293159662 [0.2163;0.3847]
(1)	80	L_1	0.486116200 [0.4153;0.5645]	0.486083747 [0.4156;0.5644]	0.486083747 [0.4156;0.5644]	0.486083751 [0.4156;0.5644]	0.486083751 [0.4156;0.5644]
(1)	80	L_2	0.006087517 [0.0045;0.0080]	0.006086645 [0.0045;0.0080]	0.006086645 [0.0045;0.0080]	0.006086645 [0.0045;0.0080]	0.006086645 [0.0045;0.0080]
(1)	80	L_a	0.853428859 [0.7133;1.0015]	0.853347310 [0.7125;1.0010]	0.853347310 [0.7125;1.0010]	0.853347295 [0.7125;1.0010]	0.853347295 [0.7125;1.0010]
(1)	110	L_1	0.724563642 [0.6363;0.8186]	0.724528925 [0.6357;0.8184]	0.724528925 [0.6357;0.8184]	0.724528922 [0.6357;0.8184]	0.724528922 [0.6357;0.8184]
(1)	110	L_2	0.009072162 [0.0070;0.0113]	0.009071158 [0.0070;0.0113]	0.009071158 [0.0070;0.0113]	0.009071158 [0.0070;0.0113]	0.009071158 [0.0070;0.0113]
(1)	110	L_a	1.272365795 [1.1024;1.4536]	1.272252405 [1.1017;1.4534]	1.272252405 [1.1017;1.4534]	1.272252381 [1.1017;1.4534]	1.272252381 [1.1017;1.4534]
(2)	40	L_2	0.37817718 [0.1077;1.0473]	0.37728799 [0.1100;1.0227]	0.37728800 [0.1100;1.0227]	0.37728833 [0.1100;1.0226]	0.37740658 [0.1099;1.0301]
(2)	80	L_2	0.66853761 [0.3625;1.3612]	-	0.66267846 [0.3611;1.3307]	0.66267825 [0.3611;1.3306]	0.66275740 [0.3612;1.3280]
(2)	110	L_2	0.88605162 [0.5568;1.5668]	-	0.87956632 [0.5538;1.5548]	0.87956592 [0.5538;1.5549]	0.87960797 [0.5550;1.5550]

Table 2: Expected losses for smoothed states, lower is better. 5th and 95th percentiles in square brackets.

to compute the likelihood once is at least twice as fast when using a pruning threshold of 1e-2 compared to 1e-6. Combined with the results of the previous section, we conclude that a threshold of 1e-2 to 1e-4 seems to be a good compromise between accuracy and speed for multivariate models. For univariate models, one can easily lower this to a very small threshold such as 1e-6. In a nutshell: the lower the more accurate, the higher the faster.

Nevertheless, we do not want to hide the obvious fact that the *Gaussian Kalman Filter* is clearly the speed champion: it is roughly ten times faster than our proposed algorithm, but we are on the order of (neglectable) **milliseconds** here. Other approaches to evaluate the likelihood, such as Sequential Monte Carlo, are typically much slower by a factor of several hundred or thousand. Moreover, our implementation of both the conventional as well as the *Pruned Skewed Kalman Filters* are very textbook-like to fix ideas and highlight the underlying intuition. There is still much room for performance gains in the codes by e.g. avoiding inverses, adapting a steady-state filter and using Chandrasekhar recursions. We experimented with several such changes to the code and are able to cut the computational time at least by a half.[5] The *Pruned Skewed Kalman Filter* is therefore a very attractive and comparatively fast addition to the filtering toolkit

[5]Granted our implementation of the *Gaussian Kalman Filter* can also be made faster. That's why we report the results for non-optimized, textbook-style codes.

of researchers who deal with skewed data and distributions. In the next section, we explore the finite sample properties of quasi-maximum likelihood estimators for the skewness parameters.

		Gaussian	Pruned Skewed Kalman Filter			
DGP	T	Kalman Filter	no pruning	1e-6	1e-4	1e-2
(1)	50	0.2555 [0.18;0.39]	26.4004 [19.85;42.68]	5.7352 [4.26;8.55]	5.4255 [4.15;8.35]	3.7236 [2.82;5.74]
(1)	100	0.4236 [0.34;0.63]	178.4764 [164.04;272.46]	9.3932 [8.56;14.91]	9.2007 [8.30;14.70]	6.2735 [5.61;10.14]
(1)	150	0.5988 [0.51;0.92]	764.5933 [730.60;860.85]	13.8686 [13.08;14.44]	13.5295 [12.72;14.12]	9.2433 [8.61;9.72]
(1)	200	0.7869 [0.68;1.28]	2276.7823 [2208.36;2374.31]	18.7564 [17.65;19.54]	18.3690 [17.25;19.01]	12.5310 [11.65;13.20]
(1)	250	1.0029 [0.87;1.65]	5407.6977 [5292.97;5522.95]	23.8373 [22.29;24.89]	23.4432 [21.77;24.56]	15.8381 [14.71;16.57]
(2)	50	0.8439 [0.71;1.46]	554.8769 [518.17;660.16]	16.9193 [15.82;17.94]	12.8243 [11.90;15.25]	8.3821 [7.74;9.32]
(2)	100	1.6507 [1.42;2.99]	10902.8252 [9388.04;13659.95]	36.6296 [32.88;42.71]	26.8729 [24.38;30.29]	17.0615 [15.77;18.32]
(2)	150	3.3166 [2.65;5.69]	-	56.8434 [49.07;89.69]	35.2654 [30.71;55.59]	26.1939 [22.86;41.05]
(2)	200	4.6808 [3.50;7.50]	-	80.1512 [65.44;118.95]	49.6175 [40.84;74.17]	36.6851 [30.44;54.95]
(2)	250	5.2042 [4.35;9.10]	-	90.9946 [81.35;143.74]	55.7294 [50.84;88.30]	41.4539 [37.92;64.54]

Table 3: Time in seconds to compute 1000 evaluations of the log-likelihood function on AMD EPYC 7402P (24 cores, 96GB RAM). 5th and 95th percentiles in square brackets.

5.3. Accuracy of parameter estimation

In the last simulation exercise, we generate $R = 1000$ datasets from the multivariate DGP (3) with different sample sizes $T = \{100, 150, 200\}$. We then estimate the underlying parameters of the distribution of η_t, i.e. μ_η, $log(diag(\Sigma_\eta))$ and $diag(\Gamma_\eta)$, while fixing all other parameters at their true values.[6] Note that DGP (3) implies that $\eta_{1,t}$ is right-skewed, $\eta_{2,t}$ is symmetric (Gaussian), and $\eta_{3,t}$ is left-skewed.

Inspired by Atkinson et al. (2019), we measure parameter accuracy by reporting not only the average, 5th and 95th percentile of our estimates in the Monte Carlo sample, but also the normalized root-mean square-error (NRMSE) for each estimated parameter. That is, for some parameter j and Kalman filter variant f the error is the difference between the parameter estimate $\hat{\theta}_{j,f,r}$ for dataset r and the true parameter value θ_j:

$$NRMSE_f^j = \frac{1}{\theta_j}\sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_{j,f,r} - \theta_j)^2}$$

[6] We log-transform the variance to avoid the non-negativity constraint during the estimation procedure. The reported estimates are re-transformed.

In other words, we normalize the RMSE by the true value θ_j to remove differences in the scales of the parameters.

Table 4 shows the parameter estimates by used Kalman filter variant (first column header) and for the three different sample sizes (second column header). Each cell includes the average value (first row), the 5th and 95th percentile in square brackets (second row), and the NRMSE in curly brackets (third row). Overall the estimates using the *Pruned Skewed Kalman Filter* are convincingly good both for a very low and a rather large cutting threshold. Most mass is centered around the true value and the distribution becomes narrower with larger sample sizes. The *Pruned Skewed Kalman Filter* successfully uncovers the skewed distribution of $\eta_{1,t}$ and $\eta_{3,t}$, but also Gaussianity of $\eta_{2,t}$. Note that the Gaussian Kalman filter completely misses the skewed distribution of η_t; which is evident in heavily biased values of μ_η and Σ_η. However, this bias is in fact misleading, because when using the Gaussian Kalman filter, μ_η and Σ_η are actually estimates for $E[\eta_t]$ and $V[\eta_t]$, which in our exercise are equal to $[0.9192; -0.1000; -0.3433]$ and $diag([0.2565; 0.3600; 0.1948])$, respectively. So the Gaussian Kalman filter still remains a powerful tool, if one is only concerned about estimating the mean and variance of the process. Of course those two statistics adjust to hide the underlying skewed distribution. In contrast, the *Pruned Skewed Kalman Filter* is more general as it nests Gaussianity as a special case.

6. Estimating the US yield curve using the dynamic Nelson-Siegel exponential components model

A yield curve is a graphical representation of the so-called term structure of interest rates, i.e. the relationship between the residual maturities of a homogeneous set of financial instruments and their computed interest rates. In practice, however, yield curves are not observed, but need to be estimated from observed market prices for the underlying financial instruments, typically government bonds that are traded on stock exchanges. Diebold & Rudebusch (2013) provide an excellent textbook introduction and Wahlstrøm et al. (2022) a recent discussion of the computational challenges to construct yield data.

Following the canonical contribution of Diebold & Li (2006), it has become standard practice to use the dynamic Nelson & Siegel (1987) (DNS) model to forecast yields at different maturities. Forecasting is crucial for bond portfolio management, derivatives pricing, risk management, but also for monetary policy decisions and financial stability analysis. Intuitively, the entire yield curve can be modelled by three dynamic factors, commonly labeled *Level* (L_t), *Slope* (S_t), and *Curvature* (C_t). The DNS model then achieves dimensionality reduction via a tight structure on the factor loadings. The model is not only simple and intuitive, but also parsimonious and very flexible in its ability to match changing shapes of the yield curve. Moreover, its out-of-sample forecasting performance is often second to none. So having a well estimated DNS model is of great importance.

		Pruned Skewed KF (1e-6)			*Pruned Skewed KF (1e-2)*			*Gaussian KF*		
Param	Truth	100	150	200	100	150	200	100	150	200
$[\mu_\eta]_1$	0.30	0.302 [0.19;0.43] {0.261}	0.299 [0.21;0.40] {0.193}	0.296 [0.22;0.38] {0.173}	0.302 [0.19;0.43] {0.260}	0.299 [0.21;0.40] {0.193}	0.297 [0.22;0.39] {0.173}	0.922 [0.83;1.01] {2.081}	0.924 [0.86;1.00] {2.083}	0.921 [0.86;0.98] {2.075}
$[\mu_\eta]_2$	-0.10	−0.101 [−0.20;−0.00] {−0.610}	−0.099 [−0.18;−0.02] {−0.480}	−0.100 [−0.17;−0.03] {−0.415}	−0.101 [−0.20;−0.00] {−0.615}	−0.099 [−0.18;−0.02] {−0.484}	−0.100 [−0.17;−0.03] {−0.418}	−0.101 [−0.20;−0.00] {−0.615}	−0.099 [−0.18;−0.02] {−0.484}	−0.100 [−0.17;−0.03] {−0.418}
$[\mu_\eta]_3$	0.20	0.195 [0.08;0.29] {0.321}	0.200 [0.11;0.27] {0.245}	0.198 [0.13;0.26] {0.214}	0.195 [0.08;0.29] {0.321}	0.200 [0.11;0.27] {0.245}	0.198 [0.13;0.26] {0.214}	−0.345 [−0.42;−0.28] {2.734}	−0.343 [−0.40;−0.28] {2.723}	−0.344 [−0.39;−0.29] {2.723}
$[\Sigma_\eta]_{11}$	0.64	0.654 [0.44;0.87] {0.203}	0.658 [0.50;0.84] {0.168}	0.656 [0.51;0.81] {0.148}	0.654 [0.44;0.87] {0.203}	0.658 [0.50;0.84] {0.168}	0.656 [0.51;0.81] {0.148}	0.260 [0.19;0.34] {0.597}	0.262 [0.21;0.32] {0.593}	0.261 [0.21;0.31] {0.594}
$[\Sigma_\eta]_{22}$	0.36	0.351 [0.27;0.44] {0.141}	0.353 [0.29;0.42] {0.117}	0.354 [0.30;0.41] {0.101}	0.351 [0.27;0.44] {0.141}	0.353 [0.29;0.42] {0.117}	0.354 [0.30;0.41] {0.101}	0.351 [0.27;0.44] {0.141}	0.353 [0.29;0.42] {0.117}	0.354 [0.30;0.41] {0.101}
$[\Sigma_\eta]_{33}$	0.49	0.491 [0.33;0.64] {0.197}	0.494 [0.37;0.62] {0.160}	0.489 [0.39;0.60] {0.133}	0.491 [0.33;0.64] {0.197}	0.494 [0.37;0.62] {0.160}	0.489 [0.39;0.60] {0.133}	0.193 [0.14;0.25] {0.609}	0.195 [0.15;0.24] {0.605}	0.193 [0.16;0.23] {0.607}
$[\Gamma_\eta]_{11}$	5.00	6.178 [3.31;11.47] {0.811}	5.712 [3.56;9.22] {0.513}	5.631 [3.78;8.63] {0.395}	6.173 [3.31;11.43] {0.804}	5.712 [3.56;9.21] {0.513}	5.630 [3.78;8.62] {0.394}			
$[\Gamma_\eta]_{22}$	0.00	−0.001 [−0.01;0.00] {}	−0.001 [−0.01;0.00] {}	−0.000 [−0.00;0.00] {}	−0.001 [−0.01;0.00] {}	−0.001 [−0.01;0.00] {}	−0.001 [−0.01;0.00] {}			
$[\Gamma_\eta]_{33}$	-6.00	−7.538 [−15.35;−3.76] {−0.741}	−6.849 [−11.09;−4.21] {−0.476}	−6.581 [−9.82;−4.30] {−0.376}	−7.535 [−15.36;−3.77] {−0.739}	−6.848 [−11.08;−4.21] {−0.474}	−6.581 [−9.82;−4.30] {−0.377}			

Table 4: Distribution of parameter estimates. Cells contain average, [5,95] percentiles and {NRMSE}.

Of particular interest to us is that Diebold et al. (2006) show how to formulate the DNS model as a linear state-space model which can be estimated by the Kalman filter. In more detail, let $y(\tau)$ denote the set of yields where τ denotes the maturity. The cross-section of yields at any discrete point in time $t = 1, ..., T$ is given by the DNS curve:

$$y_t(\tau) = L_t + S_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + C_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \tag{18}$$

Diebold & Li (2006) highlight the intuitiveness of the factor loadings. First, the level factor L_t is long-term as it has an identical loading of 1 at all maturities. This means that all yields are equally affected by a change in the level and there is no decay to zero in the limit $\tau \to \infty$. Second, the loading on the slope factor S_t starts at 1 and decays monotonically and quickly to zero. An increase in S_t increases short yields more than long ones; hence, it is a short-term factor and governs the slope of the yield curve. Third, the medium-term factor C_t has a loading that starts at 0 (no short term), increases at first and then decays back to 0 (no long term). An increase in C_t has little effect on very short and very long yields, but increases the medium-term yields; hence, it changes the *curvature* of the yield curve. The parameter λ governs the exponential decay rate and it determines the maturity at which the loading on the medium-term achieves its maximum (e.g. 0.0609 at exactly 30 months).

The latent factors L_t, S_t and C_t are assumed to be time-varying according to a first-order vector autoregressive process:

$$\begin{pmatrix} L_t - \mu^L \\ S_t - \mu^S \\ C_t - \mu^C \end{pmatrix} = \begin{pmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} & G_{23} \\ G_{31} & G_{32} & G_{33} \end{pmatrix} \begin{pmatrix} L_{t-1} - \mu^L \\ S_{t-1} - \mu^S \\ C_{t-1} - \mu^C \end{pmatrix} + \begin{pmatrix} \eta_t^L \\ \eta_t^S \\ \eta_t^C \end{pmatrix} \tag{19}$$

Obviously, equation (19) is a state transition equation as in (3). To get a corresponding measurement equation as in (4), we relate a set of N yields to the three latent factors according to (18):

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{pmatrix} \tag{20}$$

where $\varepsilon_t(\tau)$ is the measurement error for yield maturity τ. In a nutshell, the DNS model forms a linear state-space system with a VAR(1) transition equation for the dynamics of the latent factors.

We follow standard practice and assume a Gaussian white noise process for the vector of measurement errors with a diagonal covariance matrix Σ_ε and which is independent of the vector of state transition

disturbances $\eta_t = (\eta_t^L \ \eta_t^S \ \eta_t^C)'$. So far we have been silent on the distribution of η_t^L, η_t^S and η_t^C. Typically, as in Diebold et al. (2006), η_t is also assumed to be a Gaussian white noise process, but allowing for η_t^L, η_t^S and η_t^C to be contemporaneously correlated. However, Gaussianity of η_t implies that L_t, S_t and C_t must be also normally distributed, which is in stark contrast to the empirics. For instance, the usual proxies for the three latent factors – $(y(3) + y(24) + y(120))/3$ for the *level*, $y(3) - y(120)$ for the *slope* and $2y(24) - y(120) - y(3)$ for the *curvature factor* – typically display mild to strong skewness. In our sample (i.e. the one used by Diebold et al. (2006)) the empirical skewness coefficients are, respectively, equal to 1.14, 0.56 and 0.10. We also get similar non-symmetric coefficients for different time periods using the yield data of Liu & Wu (2021). Thus, when estimating the linear state space model with the conventional Kalman filter, we expect (and indeed find) that both the filtered and smoothed residuals are non-symmetric (see figure 4 for a preview of our estimation results). From a theoretical point of view, this is a sign of misspecification of the underlying model. Therefore, we we assume a $CSN(\mu_\eta, \Sigma_\eta, \nu_\eta, \Gamma_\eta, \Delta_\eta)$ distribution for η_t, which is flexible enough to capture both skewed as well as symmetric patterns in η_t. Due to identifiability issues, we set $\nu_\eta = 0$ and $\Delta_\eta = I$ and fix $\mu_\eta = -f(\Sigma_\eta, \Gamma_\eta)$, where $f(\cdot)$ is a correction function to make $E[\eta_t]$ equal to zero according to Domínguez-Molina et al. (2003, sec. 2.4.). While we do allow for a non-diagonal Σ_η matrix, we assume that Γ_η is diagonal, in order to focus on assessing whether it is a single innovation that drives the skewness (one nonzero diagonal element in Γ_η and a diagonal Σ_η matrix) or the combined effect of several skewed shocks (multiple nonzero diagonal elements in Γ_η and a non-diagonal Σ_η matrix).

To be close to the canonical work of Diebold et al. (2006), we use the same dataset, i.e. yields for 17 maturities (3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months) to estimate the following number of parameters: 9 parameters in the (3×3) transition matrix G; 3 level parameters μ^L, μ^S, and μ^C; 1 scalar decay rate λ that determines the measurement matrix F; 17 measurement variances in Σ_ε; $3(3+1)/2$ parameters in the scale matrix Σ_η; and 3 diagonal elements in Γ_η. In sum, 39 free parameters that we estimate by minimizing the negative log-likelihood function, which can be computed by using our proposed *Pruned Skewed Kalman Filter*. Based on our Monte Carlo evidence, we cut the skewness dimensions at a threshold level of 1%. The initial distribution for the prediction-error decomposition of the likelihood is set to a normal one with an initial covariance matrix with 10 on the diagonal. We do a sophisticated search for initial parameter values (as recently emphasized by Wahlstrøm et al. (2022)) and use a sequence and mixture of gradient-based and simulation-based optimization routines to minimize the negative log-likelihood function.[7] In more detail, we impose non-negativity on λ and the variances in Σ_ε by using a log transform during the optimization. Similarly, we focus on estimating the Cholesky factor of Σ_η instead of Σ_η directly. Moreover, the likelihood is penalized if the Eigenvalues of G are outside the

[7] Our choice of gradient-based optimizers include two different BFGS Quasi-Newton methods (`fminunc` in MATLAB R2022b and `csminwel` of Christopher Sims (1999)) and two different simulation based methods (the Nealder-Mead simplex search method of Lagarias et al. (1998) implemented as `fminsearch` in MATLAB R2022b and the covariance matrix adaptation evolution strategy (CMA-ES) of Hansen et al. (2003)).

unit circle or the covariance matrices of η_t or ϵ_t are not positive semi-definite. Asymptotic standard errors are obtained by computing the inverse of the negative log-likelihood. For the transformed parameters we compute standard errors according to the delta method and report results for the re-transformed estimates.

Tables 5, 6 and 7 contain the estimation results. We particularly contrast the results based on the *Pruned Skewed Kalman Filter (PSKF)* with the ones using the conventional Gaussian Kalman filter (*KF*) to illustrate the usability of the CSN distribution in multivariate state-space settings.

	KF	*PSKF*	*KF*	*PSKF*	*KF*	*PSKF*	*KF*	*PSKF*
	L_{t-1}		S_{t-1}		C_{t-1}		μ	
L_t	0.9957	1.0004	0.0285	0.0253	−0.0222	−0.0218	8.2506	6.5516
	(0.008)	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(1.086)	(3.445)
S_t	−0.0303	−0.0015	0.9385	0.9767	0.0395	0.0399	−1.3786	−1.3411
	(0.016)	(0.014)	(0.018)	(0.019)	(0.021)	(0.020)	(0.499)	(0.925)
C_t	0.0244	0.0085	0.0232	−0.0005	0.8428	0.8491	−0.3647	−0.3324
	(0.023)	(0.024)	(0.026)	(0.027)	(0.031)	(0.030)	(0.383)	(0.476)

Table 5: Parameter estimates of G, μ^L, μ^S, and μ^C. Left side of a double column corresponds to estimates obtained with the conventional Kalman filter (*KF*), right side to estimates obtained with the pruned skewed Kalman filter (*PSKF*). Asymptotic standard errors appear in parenthesis.

	KF	*PSKF*	*KF*	*PSKF*	*KF*	*PSKF*	*KF*	*PSKF*
	L_t		S_t		C_t		Γ_η	
L_t	0.0948	0.1906	−0.0140	−0.0668	0.0436	0.1648	0	−3.4648
	(0.008)	(0.046)	(0.011)	(0.052)	(0.019)	(0.105)		(0.683)
S_t			0.3823	0.7546	0.0092	0.0565	0	−1.9895
			(0.030)	(0.115)	(0.034)	(0.142)		(0.244)
C_t					0.8019	1.6045	0	1.2147
					(0.081)	(0.354)		(0.225)

Table 6: Parameter estimates of Σ_η and Γ_η. Left side of a double column corresponds to estimates obtained with the conventional Kalman filter (KF), right side to estimates obtained with the pruned skewed Kalman filter (*PSKF*). Asymptotic standard errors appear in parenthesis.

Overall, the estimates of the transition matrix G (given in the columns labeled L_{t-1}, S_{t-1} and C_{t-1} of table 5) are very similar across the two filters used and in line with the results of Diebold et al. (2006). That is, first, the eigenvalues of G are inside the unit circle, so we have a stable and covariance-stationary system. Second, we see high persistence of L_t, S_t and C_t on its own lagged dynamics, whereas most of the off-diagonals appear insignificant. While the coefficient of S_{t-1} on C_t has a different sign for the *KF* compared to the *PSKF*, both coefficients are not significantly different from zero. Next, we do see different estimates of the mean factors μ (last two columns of table 5), indicating how the estimates with the conventional Kalman filter adapt to the neglected skewness in η_t. Σ_η (first six columns of table 6) is estimated with reasonable precision for both filters. There is only one marginally significant covariance term between η_t^L and η_t^C for the *KF*, whereas in the *PSKF* case Σ_η appears to be diagonal. Note that a direct comparison of Σ_η between filters is not correct, as in the *KF* case Σ_η is the covariance matrix of η, but for the *PSKF* it is just a scale matrix and the covariance is a function of the skewness parameters Σ_η, Γ_η, Δ_η and ν_η.

	Decay	Standard deviation of measurement error for maturity							
	λ	3	6	9	12	15	18	21	24
KF	0.07776	26.83	7.55	9.03	10.45	9.91	8.65	7.86	7.21
	(0.002)	(8.68)	(3.66)	(2.85)	(3.11)	(2.96)	(2.65)	(2.45)	(2.24)
PSKF	0.07783	26.54	7.35	9.11	10.48	9.93	8.65	7.85	7.19
	(0.002)	(8.51)	(3.57)	(2.85)	(3.11)	(2.96)	(2.65)	(2.45)	(2.23)

	Standard deviation of measurement error for maturity								
	30	36	48	60	72	84	96	108	120
KF	7.27	7.91	10.30	9.26	10.04	11.18	10.70	15.07	17.28
	(2.28)	(2.44)	(3.00)	(2.80)	(3.02)	(3.37)	(3.40)	(4.55)	(5.12)
PSKF	7.29	7.93	10.30	9.25	10.03	11.14	10.71	15.13	17.29
	(2.29)	(2.45)	(3.01)	(2.80)	(3.02)	(3.37)	(3.40)	(4.56)	(5.12)

Table 7: Parameter estimates of decay parameter λ and of standard deviations of measurement errors, expressed in basis points, i.e. $100\sqrt{diag(\Sigma_\epsilon)}$. *KF* denotes the conventional Kalman filter and *PSKF* the pruned skewed Kalman filter. Asymptotic standard errors appear in parenthesis.

Therefore, we also compute and compare the estimated covariance matrices:

$$\widehat{COV}[\eta_t]^{KF} = \begin{pmatrix} 0.0948 & -0.0140 & 0.0436 \\ -0.0140 & 0.3823 & 0.0092 \\ 0.0436 & 0.0092 & 0.8019 \end{pmatrix}, \quad \widehat{COV}[\eta_t]^{PSKF} = \begin{pmatrix} 0.0943 & -0.0181 & 0.0453 \\ -0.0181 & 0.3716 & 0.0223 \\ 0.0453 & 0.0223 & 0.8076 \end{pmatrix}$$

We see that the variances of η_t^L and η_t^S are estimated slightly lower with the *PSKF*, but the differences are negligible. The overall estimation is quite accurate according to table 7, as the standard deviations of the measurement errors are very small (reported in basis points) and do not differ across the filters. The same holds true for the estimate of the decay parameter λ, which would imply the loading on the curvature factor to be maximized at a maturity of 23.06 months for the *KF* and 23.04 months for the *PSKF*. Finally, we turn towards the estimates of the diagonal elements in Γ_η (last two columns of table 6). The estimation with the *PSKF* indeed reveals significant left-skewness in the underlying distributions of η_t^L and η_t^S, whereas η_t^C is right-skewed. We use the proposed *Pruned Skewed Kalman Smoother* to compute the smoothed values for $\eta_{t|T}$ and contrast the histograms of the values with the ones from the conventional *Gaussian Kalman Smoother* in figure 4. As far as we know, we are the first to actually report smoothed (and not filtered) innovations using the CSN distribution in a multivariate state-space setting. There is a clear skewed pattern for both filters, but also some bulging dents, which are both in clear contradiction to a symmetric distribution.[8] Theoretically speaking, this indicates a misspecification of the linear state-space model when using the Gaussian assumption for η_t, whereas the CSN distribution is flexible enough to incorporate the skewed shapes in the estimation. Accordingly, since the *Skewed Kalman Filter* nests

[8] A negative sign of Γ_η indicates left-skewness, while a positive sign indicates right-skewness. Note, however, that the magnitude of the estimates of Γ_η do not directly translate into the same magnitude of the empirical skewness coefficient, as it is a function of both Γ_η and Σ_η.

Gaussianity as a restriction ($\Gamma_\eta = 0$), we perform a likelihood ratio test and obtain a very high test statistic of 28.86. In summary, on the basis of our estimation results, the data strongly favors a skewed error term distribution for η_t.

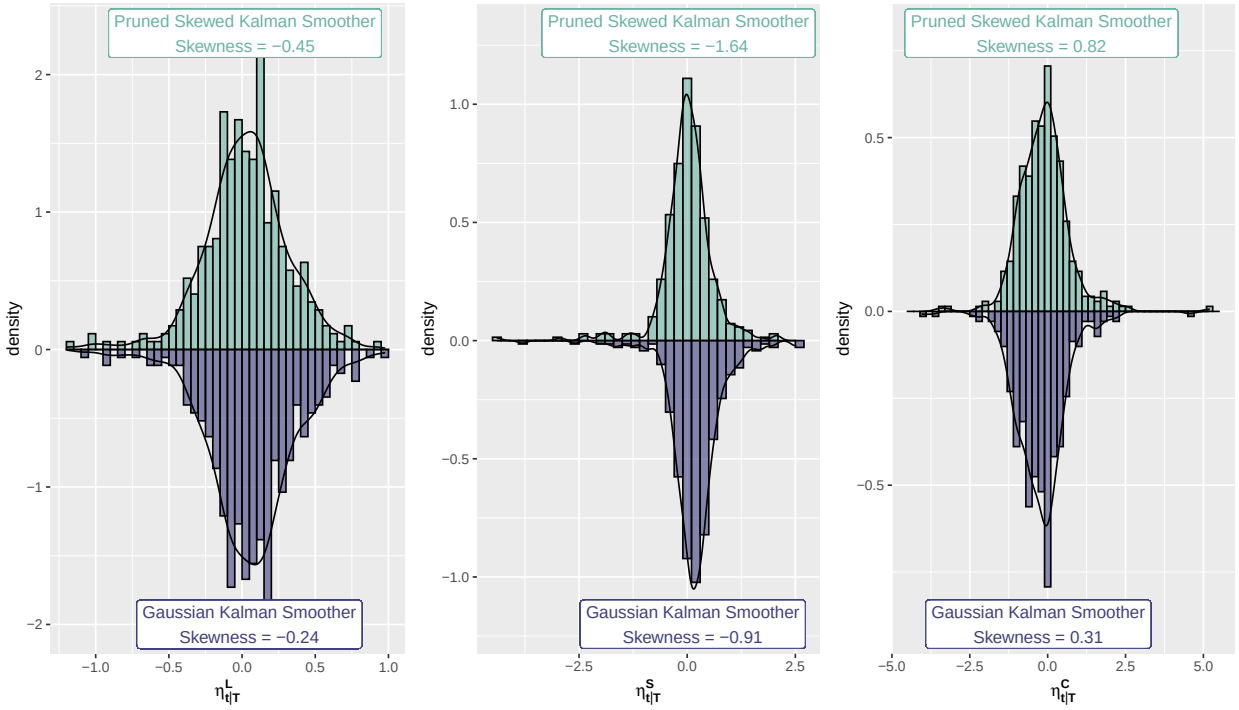

Figure 4: Mirror histogram and kernel density estimate of smoothed innovations $\eta^L_{t|T}$, $\eta^S_{t|T}$ and $\eta^C_{t|T}$.

7. Conclusion

The skewed Kalman filter is an analytical recursive procedure for inference about the state vector in linear state-space systems and can be exploited to compute the exact likelihood function when the innovations stem from the CSN distribution. An intriguing feature of the Skewed Kalman filter is that it nests both Gaussianity and the skew-normal distribution as special cases. Previously, however, applying this filter to data required extensive computational capabilities or was even impossible for multivariate models, because it involves the evaluations of high-dimensional multivariate normal cdfs of growing dimensions. We propose a fast and simple pruning algorithm to the updating step of the filter that overcomes this curse of increasing dimensions. We theoretically demonstrate that it is valid for any dataset and set of parameter values. Our *Pruned Skewed Kalman Filter and Smoother* operate effectively and efficiently in practice in spite of – or perhaps precisely because of – its simplicity as shown in our extensive Monte Carlo study and real data application. Therefore, we believe that the applicability of this filter extends across many disciplines whenever the skew-normal distribution and its numerous variants are used to identify modest departures from

symmetry, including economics, econometrics, engineering, medical science, psychology, risk management, and statistics.

There are various areas to conduct further research. From a statistical point of view, other variants of the closed skew-normal distribution and their application for Kalman filtering are worth investigating. The unified skew-normal distribution of Arellano-Valle & Azzalini (2006) may be a good alternative to alleviate some of the CSN distribution's identification issues. From an economic point of view, our empirical application demonstrated that yield data favors a dynamic Nelson-Siegel yield curve model with skewed innovations. Naturally, the next step on our research agenda is to examine the forecasting performance, especially in conjunction with other macroeconomic factors. We are also exploring the usefulness of the *Pruned Skewed Kalman Filter* for estimating linearized Dynamic Stochastic General Equilibrium (DSGE) models. Preliminary results on the workhorse model of Smets & Wouters (2007) provides evidence for asymmetric monetary policy shocks.

In conclusion, the *Pruned Skewed Kalman Filter* offers an elegant, fast and appealing tool to explicitly model and estimate departures from symmetry.

References

Adjemian, S., Bastani, H., Juillard, M., Karamé, F., Mihoubi, F., Mutschler, W., Pfeifer, J., Ratto, M., Villemot, S., & Rion, N. (2022). *Dynare: Reference Manual Version 5*. Technical Report 72 CEPREMAP. URL: `https://ideas.repec.org/p/cpm/dynare/072.html`.

Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable Growth. *American Economic Review*, *109*, 1263–1289. URL: `https://pubs.aeaweb.org/doi/10.1257/aer.20161923`. doi:`10.1257/aer.20161923`.

Amsler, C., Papadopoulos, A., & Schmidt, P. (2021). Evaluating the cdf of the Skew Normal distribution. *Empirical Economics*, *60*, 3171–3202. doi:`10.1007/s00181-020-01868-6`.

Arellano-Valle, R. B., & Azzalini, A. (2006). On the Unification of Families of Skew-normal Distributions. *Scandinavian Journal of Statistics*, *33*, 561–574. URL: `https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9469.2006.00503.x`. doi:`10.1111/j.1467-9469.2006.00503.x`.

Arellano-Valle, R. B., & Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis*, *99*, 1362–1382. doi:`10.1016/j.jmva.2008.01.020`.

Arellano-Valle, R. B., Contreras-Reyes, J. E., Quintero, F. O. L., & Valdebenito, A. (2019). A skew-normal dynamic linear model and Bayesian forecasting. *Computational Statistics*, *34*, 1055–1085. doi:`10.1007/s00180-018-0848-1`.

Atkinson, T., Richter, A., & Throckmorton, N. (2019). The zero lower bound and estimation accuracy. *Journal of Monetary Economics*, *115*, 249–264. doi:`10.1016/j.jmoneco.2019.06.007`.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, *12*, 171–178. URL: `https://www.jstor.org/stable/4615982`.

Azzalini, A., & Capitanio, A. (2014). *The Skew-Normal and Related Families*. Number 3 in Institute of Mathematical Statistics Monographs. Cambridge: Cambridge University Press.

Azzalini, A., & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, *83*, 715–726. doi:`10.1093/biomet/83.4.715`.

Bauer, M., & Chernov, M. (2021). *Interest Rate Skewness and Biased Beliefs*. Technical Report w28954 National Bureau of Economic Research Cambridge, MA. URL: http://www.nber.org/papers/w28954.pdf.

Cabral, C. R. B., Da-Silva, C. Q., & Migon, H. S. (2014). A Dynamic Linear Model with Extended Skew-normal for the Initial Distribution of the State Parameter. *Computational Statistics & Data Analysis*, *74*, 64–80. doi:10.1016/j.csda.2013.12.008.

Chen, J. T., Gupta, A. K., & Troskie, C. G. (2003). The Distribution of Stock Returns When the Market Is Up. *Communications in Statistics - Theory and Methods*, *32*, 1541–1558. doi:10.1081/STA-120022244.

Chen, Y.-Y., Schmidt, P., & Wang, H.-J. (2014). Consistent estimation of the fixed effects stochastic frontier model. *Journal of Econometrics*, *181*, 65–76. doi:10.1016/j.jeconom.2013.05.009.

Chiplunkar, R., & Huang, B. (2021). Latent variable modeling and state estimation of non-stationary processes driven by monotonic trends. *Journal of Process Control*, *108*, 40–54. doi:10.1016/j.jprocont.2021.10.010.

Christopher Sims (1999). Matlab Optimization Software. Quantitative Macroeconomics & Real Business Cycles.

Counsell, N., Cortina-Borja, M., Lehtonen, A., & Stein, A. (2011). Modelling Psychiatric Measures Using Skew-Normal Distributions. *European Psychiatry*, *26*, 112–114. doi:10.1016/j.eurpsy.2010.08.006.

de Roon, F., & Karehnke, P. (2017). A Simple Skewed Distribution with Asset Pricing Applications. *Review of Finance*, *21*, 2169–2197. doi:10.1093/rof/rfw040.

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, *130*, 337–364. doi:10.1016/j.jeconom.2005.03.005.

Diebold, F. X., & Rudebusch, G. D. (2013). *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*. The Econometric and Tinbergen Institutes Lectures. Princeton: Princeton University Press.

Diebold, F. X., Rudebusch, G. D., & Aruoba, B. S. (2006). The macroeconomy and the yield curve: A dynamic latent factor approach. *Journal of Econometrics*, *131*, 309–338. doi:10.1016/j.jeconom.2005.01.011.

Domínguez-Molina, J., González-Farías, G., & Gupta, A. K. (2003). *The Multivariate Closed Skew Normal Distribution*. Technical Report 03-12 Bowling Green State University.

Eling, M. (2012). Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics*, *51*, 239–248. doi:10.1016/j.insmatheco.2012.04.001.

Emvalomatis, G., Stefanou, S. E., & Lansink, A. O. (2011). A Reduced-Form Model for Dynamic Efficiency Measurement: Application to Dairy Farms in Germany and The Netherlands. *American Journal of Agricultural Economics*, *93*, 161–174. doi:10.1093/ajae/aaq125.

Fagiolo, G., Napoletano, M., & Roventini, A. (2008). Are output growth-rate distributions fat-tailed? Some evidence from OECD countries. *Journal of Applied Econometrics*, *23*, 639–669. doi:10.1002/jae.1003.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., & Schorfheide, F. (2016). Solution and Estimation Methods for DSGE Models. In J. B. Taylor, & H. Uhlig (Eds.), *Handbook of Macroeconomics* (pp. 527–724). Elsevier North-Holland volume A. URL: https://doi.org/10.1016/bs.hesmac.2016.03.006.

Flecher, C., Naveau, P., & Allard, D. (2009). Estimating the closed skew-normal distribution parameters using weighted moments. *Statistics & Probability Letters*, *79*, 1977–1984. doi:10.1016/j.spl.2009.06.004.

Gallier, J. H. (2010). *Notes on the Schur Complement*. Departmental Papers (CIS) 12-10-2010 University of Pennsylvania Pennsylvania. URL: https://repository.upenn.edu/cis_papers/601/.

Genton, M. G. (2004). *Skew-Elliptical Distributions and Their Applications - A Journey Beyond Normality*. S.l.: CRC PRESS.

González-Farías, G., Domínguez-Molina, A., & Gupta, A. K. (2004a). Additive properties of skew normal random vectors. *Journal of Statistical Planning and Inference*, *126*, 521–534. doi:10.1016/j.jspi.2003.09.008.

González-Farías, G., Domínguez-Molina, A., & Gupta, A. K. (2004b). The closed skew-normal distribution. In M. G. Genton (Ed.), *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality* (pp. 25–42). London: Chapman

& Hall/CRC. URL: https://doi.org/10.1201/9780203492000.

Grabek, G., Kłos, B., & Koloch, G. (2011). *Skew-Normal Shocks in the Linear State Space Form DSGE Model*. Working Paper 101 National Bank of Poland Warszawa. URL: https://www.nbp.pl/publikacje/materialy_i_studia/101_en.pdf.

Gürkaynak, R. S., & Wright, J. H. (2012). Macroeconomics and the Term Structure. *Journal of Economic Literature*, *50*, 331–367. doi:10.1257/jel.50.2.331.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, N.J: Princeton University Press.

Hansen, N., Müller, S. D., & Koumoutsakos, P. (2003). Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). *Evolutionary Computation*, *11*, 1–18. doi:10.1162/106365603321828970.

Harvey, A. C., & Phillips, G. D. A. (1979). Maximum Likelihood Estimation of Regression Models with Autoregressive- Moving Average Disturbances. *Biometrika*, *66*, 49. doi:10.2307/2335241.

Horn, R. A., & Johnson, C. R. (2017). *Matrix Analysis*. (Second edition, corrected reprint ed.). New York, NY: Cambridge University Press.

Käärik, M., Selart, A., & Käärik, E. (2015). On Parametrization of Multivariate Skew-Normal Distribution. *Communications in Statistics - Theory and Methods*, *44*, 1869–1885. doi:10.1080/03610926.2012.760277.

Karlsson, S., Mazur, S., & Nguyen, H. (2021). Vector autoregression models with skewness and heavy tails, . doi:10.48550/ARXIV.2105.11182.

Kilian, L., & Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*. Themes in Modern Econometrics. Cambridge: Cambridge University Press. URL: https://doi.org/10.1017/9781108164818.

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions. *SIAM Journal on Optimization*, *9*, 112–147. doi:10.1137/S1052623496303470.

Lindé, J., Smets, F., & Wouters, R. (2016). Challenges for Central Banks' Macro Models. In J. B. Taylor, & H. Uhlig (Eds.), *Handbook of Macroeconomics* (pp. 527–724). Elsevier North-Holland volume B. URL: https://doi.org/10.1016/bs.hesmac.2016.04.009.

Liseo, B., & Parisi, A. (2013). Bayesian inference for the multivariate skew-normal model: A population Monte Carlo approach. *Computational Statistics & Data Analysis*, *63*, 125–138. doi:10.1016/j.csda.2013.02.007.

Liu, Y., & Wu, J. C. (2021). Reconstructing the yield curve. *Journal of Financial Economics*, *142*, 1395–1425. doi:10.1016/j.jfineco.2021.05.059.

Ludvigson, S. C., Ma, S., & Ng, S. (2021). Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response? *American Economic Journal: Macroeconomics*, *13*, 369–410. doi:10.1257/mac.20190171.

Manouchehri, T., & Nematollahi, A. R. (2019). Periodic autoregressive models with closed skew-normal innovations. *Computational Statistics*, *34*, 1183–1213. doi:10.1007/s00180-019-00893-z.

Mendell, N. R., & Elston, R. C. (1974). Multifactorial Qualitative Traits: Genetic Analysis and Prediction of Recurrence Risks. *Biometrics*, *30*, 41. doi:10.2307/2529616.

Naveau, P., Genton, M. G., & Shen, X. (2005). A skewed Kalman filter. *Journal of Multivariate Analysis*, *94*, 382–400. doi:10.1016/j.jmva.2004.06.002.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious Modeling of Yield Curves. *The Journal of Business*, *60*, 473–489. URL: http://www.jstor.org/stable/2352957.

Neuberger, A. (2012). Realized Skewness. *Review of Financial Studies*, *25*, 3423–3455. doi:10.1093/rfs/hhs101.

Nurminen, H., Ardeshiri, T., Piché, R., & Gustafsson, F. (2018). Skew-t Filter and Smoother With Improved Covariance Matrix Approximation. *IEEE Transactions on Signal Processing*, *66*, 5618–5633. doi:10.1109/TSP.2018.2865434.

Pescheny, J. V., Gunn, L. H., Pappas, Y., & Randhawa, G. (2021). The impact of the Luton social prescribing programme on mental well-being: A quantitative before-and-after study. *Journal of Public Health*, *43*, e69–e76. doi:10.1093/pubmed/fdz155.

Rezaie, J., & Eidsvik, J. (2014). Kalman filter variants in the closed skew normal setting. *Computational Statistics & Data*

Analysis, *75*, 1–14. doi:10.1016/j.csda.2014.01.014.

Rezaie, J., & Eidsvik, J. (2016). A skewed unscented Kalman filter. *International Journal of Control*, *89*, 2572–2583. doi:10.1080/00207179.2016.1171912.

Ruge-Murcia, F. (2017). Skewness Risk and Bond Prices. *Journal of Applied Econometrics*, *32*, 379–400. doi:10.1002/jae.2528.

Smets, F., & Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, *97*, 586–606. doi:10.1257/aer.97.3.586.

Vernic, R. (2006). Multivariate skew-normal distributions with applications in insurance. *Insurance: Mathematics and Economics*, *38*, 413–426. doi:10.1016/j.insmatheco.2005.11.001.

Wahlstrøm, R. R., Paraschiv, F., & Schürle, M. (2022). A Comparative Analysis of Parsimonious Yield Curve Models with Focus on the Nelson-Siegel, Svensson and Bliss Versions. *Computational Economics*, *59*, 967–1004. doi:10.1007/s10614-021-10113-w.

Wei, Z., Zhu, X., & Wang, T. (2021). The extended skew-normal-based stochastic frontier model with a solution to 'wrong skewness' problem. *Statistics*, *55*, 1387–1406. doi:10.1080/02331888.2021.2004142.

Wolf, E. (2022). Estimating growth at risk with skewed stochastic volatility models, . doi:10.17169/REFUBIUM-33629.