Pruned Skewed Kalman Filter and Smoother: With Applications to the Yield Curve and Asymmetric Monetary Policy Shocks

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

We propose a computationally efficient algorithm designed to address the *curse of increasing dimensions* found in the *Skewed Kalman Filter*. The algorithm's accuracy and efficiency are substantiated through a comprehensive simulation study encompassing both univariate and multivariate state-space models. We demonstrate applicability by estimating a multivariate dynamic Nelson-Siegel term structure model and a New Keynesian DSGE model on US data with Maximum Likelihood. In both applications, the results reveal a strong preference for a skewed error term distribution.

Keywords: state-space models, skewed Kalman filter, skewed Kalman smoother, closed skew-normal, dimension reduction, asymmetric shocks, yield curve, term structure, dynamic Nelson-Siegel, DSGE
JEL: C32, C51, E32, E43

⋆This paper was presented at the Young-Academics Mini-Symposium at the Statistische Woche 2022 Münster and the 16th annual Dynare Conference in Lancaster. The authors thank Dietmar Bauer for both critical and helpful comments as well as kindly 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

The Kalman filter is a highly effective recursive procedure for making inference about state vectors, which can be used to compute the precise Gaussian likelihood function. The filter is optimal in the sense that it minimizes the covariance matrix of one-step ahead prediction errors. Furthermore, the Kalman filter can be executed swiftly and efficiently from an applied and computational standpoint. However, non-Gaussianity, specifically skewness, characterizes many time series frequently employed for estimating linear state-space models in real data applications. As a result, it is necessary to adjust the state-space modeling framework and algorithms to accommodate skewness in the error term distribution.

In this context, the closed skew-normal (CSN) distribution proposed by González-Farías et al. (2004b) serves as an appropriate alternative, as it extends the Gaussian distribution by introducing skewness while maintaining key properties of the normal distribution, see e.g. Azzalini & Capitanio (2014) and Genton (2004) for excellent textbook introductions. Notably, this distribution encompasses both the normal distribution and the widely-used skew-normal distribution of Azzalini (1985) and Azzalini & Dalla Valle (1996) as special cases. Since the three fundamental tools for implementing the Kalman filter are closure under linear transformation, summation, and conditioning, utilizing this distribution enables the development of closed-form recursions that closely resemble the Gaussian Kalman filter (Naveau et al., 2005).

However, applications are usually limited to univariate settings and simplified model assumptions. We posit that this is primarily due to a computational challenge we refer to as the *curse of increasing skewness dimensions*, which we address in this paper. Essentially, the issue arises from the fact that the probability density function (pdf) of the CSN distribution possesses two dimensions, resulting from the multiplication of a Gaussian pdf by the ratio of two Gaussian cumulative distribution functions (cdf). While the Gaussian pdf reflects resemblance to the normal distribution, the skewness dimension originates from the Gaussian cdfs. Even though evaluating Gaussian cdfs is a well understood task, it can become numerically difficult, if not infeasibe, if the dimension of the cdfs becomes very large, a point recently echoed by Amsler et al. (2021) for the skew-normal distribution. And this manifests the core challenge intrinsic to the *Skewed Kalman Filter*, as in state-space models this dimension grows swiftly and may even explode as the recursion proceeds over many time steps. It does so, because the sum of two CSN distributed variables remains within a CSN distribution, yet the resulting skewness dimension consists of the combined sum of the individual dimensions of each variable.

To address this challenge, our primary contribution is to propose a computationally efficient method for approximating the updating distribution of the skewed Kalman filter by reducing the skewness dimension at each iteration. Our algorithm relies on the fact that a CSN distributed random variable can be represented as a conditional distribution of two normally distributed variables. Intuitively, in this representation, the correlation between the two random variables introduces asymmetry and skewness. When the correlation is high, the asymmetry of the conditional random variable, which is CSN distributed, is also large. However, when the correlation is low, the symmetry is minimally affected, and the CSN distribution closely resembles the Gaussian distribution. In the extreme case with no correlation, the conditional random variable will be identical to a normally distributed one, causing the *Skewed Kalman Filter* to morph into the *Gaussian Kalman Filter*. Our approach is hence based on a low threshold, such as 1% in absolute value, at which we discard weakly correlated elements in the skewed Kalman filtering steps, as they do not substantially distort symmetry. By doing this, we effectively decrease the overall skewness dimension by the number of pruned variables, making the *Skewed Kalman Filter* applicable for multivariate state-space models without any restrictive assumptions or constraints on the state-space system. We refer to this algorithm as the *Pruned Skewed Kalman Filter*. Our second contribution is to analytically demonstrate how skewness propagates through the system, providing motivation and derivation for the algorithm. Lastly, our third contribution is to derive the *Skewed Kalman Smoother*. To our knowledge, we are the first to provide 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 extensive Monte Carlo simulation evidence in both univariate and multivariate settings. When data exhibits skewness, the *Pruned Skewed Kalman* algorithm (i) filters and smooths the unobserved state vector more accurately than the conventional Kalman algorithm, (ii) requires only marginally more time than the Gaussian Kalman filter to evaluate the likelihood function, and (iii) offers precise maximum likelihood estimators for the shock parameters in finite samples.

We demonstrate the applicability of the *Pruned Skewed Kalman Filter and Smoother* by revisiting two well-established macroeconometric estimation exercises: the multivariate dynamic Nelson-Siegel (DNS) term structure model of Diebold et al. (2006) and the New Keynesian Dynamic Stochastic General Equilibrium (DSGE) model of Ireland (2004). The first application is motivated by the fact that it has become standard practice to analyze the term structure of interest rates through estimating DNS models with the conventional Gaussian Kalman filter. However, the resulting estimates typically reveal mild to substantial skewness in the smoothed error term distribution, which subsequently propagates through the state-space system, leading to skewed estimated latent factors. This means that skewed shocks play a crucial role in explaining level, slope and curvature changes in the yield curve, but this contradicts the assumption of either Gaussianity or linearity of the state-space system. Recent evidence also suggests that skewness of the latent factors is a significant indicator of the state of the economy, (not only but) particularly in the face of unprecedented low interest rates (Bauer & Chernov, 2021; Ruge-Murcia, 2017). Furthermore, the negative skewness of stock returns and its implications for asset pricing and investment management have been extensively documented (Neuberger, 2012). The second application is motivated by recent estimates (with the Gaussian Kalman smoother) that demonstrate significant asymmetry in structural shocks such as monetary policy, productivity, and uncertainty innovations (Lindé et al., 2016; Ludvigson et al., 2021; Ruge-Murcia, 2017). This once again challenges the validity of the Gaussian assumption when employing Kalman filtering techniques for estimation. Therefore, we re-estimate both models using the *Pruned Skewed Kalman Filter and Smoother* within a Maximum Likelihood framework on US data. In alignment with the aforementioned evidence, our findings indicate that the data clearly favors a skewed distribution for the error term distribution of all three yield curve factors in the DNS model. Similarly, in the estimated DSGE model, we discover that both productivity and monetary policy shocks exhibit substantial asymmetry.

Our presentation and implementation of the *Pruned Skewed Kalman Filter and Smoother* maintain a high degree of generality, closely mirroring the simplicity found in the normal Kalman filtering and smoothing routines. As for modeling, empirical researchers can retain their linear state-space system while introducing additional flexibility by assuming a CSN distribution for the error terms in the state transition equation. In terms of computation, any estimation approach employing Kalman filtering techniques, be it Bayesian or Frequentist, can be easily and seamlessly adapted by simply replacing the Kalman filtering routine. Notably, we have already developed a preliminary implementation and interface to integrate the *Pruned Skewed Kalman Filter* into Dynare, a toolbox for estimating DSGE models using both Maximum Likelihood and Bayesian MCMC methods (Adjemian et al., 2022).[1] To highlight this versatility, we provide model-independent implementations of the *Pruned Skewed Kalman Filter and Smoother* in Julia, MATLAB, Python, and R.[2] Ultimately, our goal is to offer an accessible and intuitive tool for promoting empirical research across a wide array of fields where skewness is a crucial and inherent aspect of the research agenda.

Related Literature

On the one hand, the (closed) skew-normal distribution has been applied in various disciplines, such as property-liability insurance claims (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; Zhu et al., 2022), stock returns (Chen et al., 2003), and multivariate time series econometrics (Karlsson et al., 2023). On the other hand, the *Skewed Kalman Filter* is seldom used in practice, despite its considerable potential and simplicity of its implementation. Particularly, in economics and econometrics, the literature is very sparse, with Cabral et al. (2014) examining UK gas consumption and Emvalomatis et al. (2011) estimating dynamic efficiency measurements in agricultural economics as notable exceptions.

Naveau et al. (2005) and Cabral et al. (2014) formulate *Skewed Kalman Filters* based on the CSN distribution for linear state-space systems, but assume the CSN distribution for the initial state vector only. Interestingly, in this scenario, the skewness dimension remains constant, allowing for a straightforward derivation of the Kalman filtering steps without encountering the *curse of increasing skewness dimensions*.

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

[2] Code is available at `https://github.com/gguljanov/pruned-skewed-kalman`.

However, we demonstrate that the impact of the initial distribution and the level of skewness dissipate rapidly over time, which is not commonly observed in real data applications. Alternatively, Naveau et al. (2005) devise an extended univariate state-space model by dividing the state vector into linear and skewed components, enabling filtering without an explosion in the skewness dimension. Kim et al. (2014) later extend this approach for mixtures of skewed Kalman filters. Nonetheless, general state-space models, like the reduced-form representations of structural economic models, cannot be transformed into this extended format, and it is also subject to the *curse of increasing skewness dimensions*. Moreover, they only provide numerical examples in univariate settings, whereas we provide real data applications in multivariate frameworks. Another approach proposed by Arellano-Valle et al. (2019) is to incorporate the CSN distribution into the measurement equation, while still modeling state transition shocks as normally distributed. However, ample evidence in economics suggests that skewness primarily originates from innovations rather than measurement errors, rendering their approach unsuitable for broader contexts. Finally, Rezaie & Eidsvik (2014, 2016) develop *Skewed Unscented Kalman Filters* for nonlinear state-space systems and discuss computational aspects. They contend that, for practical purposes, one must either assume simplified conditions or refit the updated distribution. In this paper, we specifically choose to employ the latter strategy.

The structure of this paper is as follows: Section 2 provides an overview of the CSN distribution's representations and properties, which are essential for filtering and smoothing. Section 3 outlines the closed-form expressions and the forward and backward recursion steps for the *Skewed Kalman Filter and Smoother*. In Section 4, we initially demonstrate how skewness propagates through the state-space system over time and subsequently derive our *pruning algorithm*. In Section 5, we present a summary of our Monte Carlo results, with detailed results available in an online appendix. Sections 6 and 7 concentrate on our two empirical applications. Finally, we offer concluding remarks in Section 8.

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

$$W = \mu + E_1$$
$$Z = -\nu + \Gamma E_1 + E_2.$$

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

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

The moment generating function (mgf) of X is

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

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 well-known 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, with the skewness coefficient being bounded by $\pm\sqrt{2}(\pi-4)/(\pi-2)^{3/2} \approx \pm 0.995$. 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).

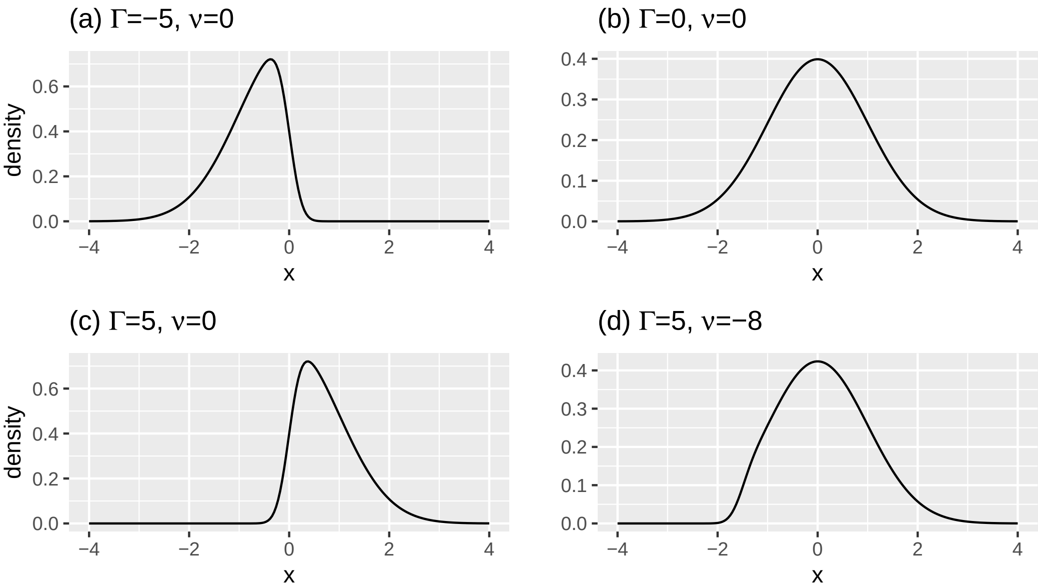

Figure 1: Density functions of univariate CSN distributions with different skewness parameters Γ and ν; other parameters are $\mu=0$, $\Sigma=1$ and $\Delta=1$.

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 $r\le p$ such that $F\Sigma_x F'$ is non-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

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

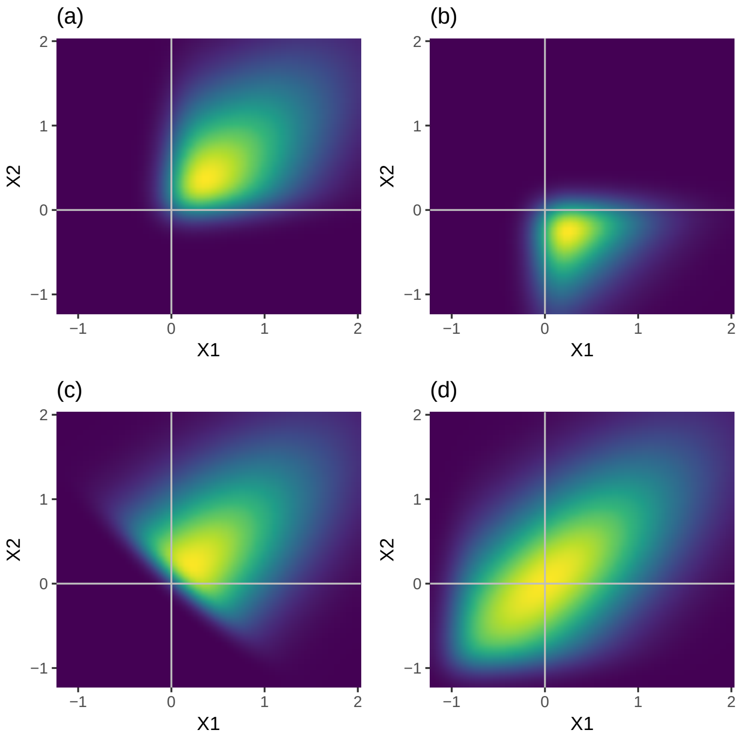

Figure 2: Density functions of bivariate CSN distributions with different skewness parameters:
(a) $\Gamma = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}$, $\nu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, (b) $\Gamma = \begin{bmatrix} 6 & 0 \\ 0 & -6 \end{bmatrix}$, $\nu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$,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

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*

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

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 $\Delta_{xx} = \Delta_x + \Gamma_x \Sigma_x \Gamma'_x - \Gamma_x \Sigma_x \Sigma_z^{-1} \Sigma_x \Gamma'_x$, $\Delta_{yy} = \Delta_y + \Gamma_y \Sigma_y \Gamma'_y - \Gamma_y \Sigma_y \Sigma_z^{-1} \Sigma_y \Gamma'_y$, *and* $\Delta_{xy} = -\Gamma_x \Sigma_x \Sigma_z^{-1} \Sigma_y \Gamma'_y$.

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 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, time-varying parameters, and even to a scale mixture class of closed skew-normal distributions as in Kim et al. (2014). 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 ones we study in sections 6 and 7.

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

[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*.

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 the derivation of the smoothing step, which is new to the literature on skewed Kalman filters.

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

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

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

$$\begin{aligned}
\mu_{t|t} &= \mu_{t|t-1} + \Sigma_{t|t-1}F'(F\Sigma_{t|t-1}F' + \Sigma_\varepsilon)^{-1}(y_t - F\mu_{t|t-1} - \mu_\varepsilon) \\
\Sigma_{t|t} &= \Sigma_{t|t-1} - \Sigma_{t|t-1}F'(F\Sigma_{t|t-1}F' + \Sigma_\varepsilon)^{-1}F\Sigma_{t|t-1} && (10) \\
\Gamma_{t|t} &= \Gamma_{t|t-1} && (11) \\
\nu_{t|t} &= \nu_{t|t-1} - \Gamma_{t|t-1}\Sigma_{t|t-1}F'(F\Sigma_{t|t-1}F' + \Sigma_\varepsilon)^{-1}(y_t - F\mu_{t|t-1} - \mu_\varepsilon) \\
\Delta_{t|t} &= \Delta_{t|t-1}. && (12)
\end{aligned}$$

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. Again we see that Γ regulates skewness continuously. Without skewness, $\Gamma_{t|t-1} = 0$ and $K_{t-1}^{Skewed} = 0$, the prediction and updating steps are 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 on the state-space system, 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* `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 `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.

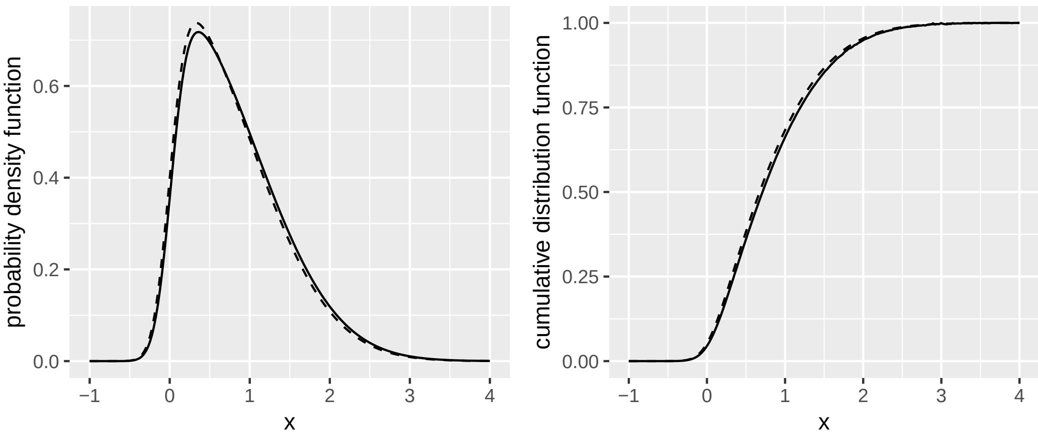

Figure 3: Probability density functions and cumulative distribution functions of a CSN distributed random variable with two skewness dimensions (skewness parameters as given in (14), solid lines) and the approximating $CSN(0, 1, 6, 0, 1)$ distribution with one skewness dimension (dashed lines).

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 (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

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

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 (2011, prop. 16.1) 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 these results lead to a shrinkage of R_2^{ij} over time. The same line of thought can also be applied to the parameters of the prediction step, i.e. to $P_2 \equiv \Gamma_{t|t-1}\Sigma_{t|t-1}$ and $R_2^{ij} = \frac{{P_2}^{ij}}{\sqrt{\Sigma_{t|t-1}^{ii}}\sqrt{\Delta_{t|t-1}^{jj}}}$. To summarize, the algorithm is guaranteed to reduce the skewness dimension after sufficiently many periods.

5. A Monte Carlo Study

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 both univariate as well as multivariate state-space models as data-generating processes (DGP). The Online Appendix provides a thorough description of the parameters of the different DGPs and the detailed outcomes of the Monte Carlo analysis. Overall, we find that the *Pruned Skewed Kalman Filter and Smoother* perform very well in terms of accuracy, speed and finite sample properties of Maximum Likelihood estimates of the error term parameters. In what follows we briefly summarize the key lessons.

Accuracy. We assess how accurate the filter and smoother estimate the value of the underlying state variables by considering different loss functions and corresponding optimal point estimators; namely, the expectation, the median and the quantiles of both filtered and smoothed states.[4] We simulate 2400 sample paths for x_t and y_t of different length $(40, 80, 110)$ plus a burn-in phase, where the shocks η_t are drawn from the CSN distribution and the measurement errors ε_t from the normal distribution. We compute the expected losses for both the *Gaussian* as well as *Pruned Skewed Kalman Filter and Smoother* by averaging over all replications. 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. Second, our pruning algorithm is very accurate and numerically almost equivalent to the non-pruned *Skewed Kalman Filter* (up to the twelfth digit in the univariate case and up to the 5th digit in the multivariate case). Third,

[4]Note that in the multivariate case, there is no consensus on multivariate extensions of quantiles (see e.g. Jeong (2023, footnote 3)), so there we focus only on the quadratic loss function.

the pruning threshold does not matter measurably in the univariate case and makes only a small numerical difference in multivariate settings.

Speed. We compare the time required to compute 1000 evaluations of the log-likelihood function for different sample sizes across filters and smoothers. Clearly, the *Gaussian Kalman Filter* is the speed champion: it is roughly ten times faster than our proposed algorithm, but we are on the order of *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. More importantly, while the computational time and memory requirement of the non-pruned *Skewed Kalman Filter* increases exponentially and explodes in multivariate models rather quickly, our proposed *Pruned Skewed Kalman Filter* does not suffer from this and performs very well for both univariate and multivariate settings. It is only slightly affected by a growing sample size; relatively speaking, it behaves just as the conventional Kalman filter in this regard. That is, the relative time increase between a sample size of 50 and 250 is approximately 4 both for the *Gaussian* as well as our *Pruned Skewed Kalman Filter*. Regarding the choice of pruning threshold, the average time needed to compute the likelihood once is at least twice as fast when using a pruning threshold of 10^{-2} compared to 10^{-5}. Combined with the accuracy results, we therefore suggest that a threshold of 1% seems to be a good compromise between accuracy and speed for multivariate models, in univariate models this can be easily lowered to a very tight pruning threshold of say 10^{-5}.

Maximum Likelihood Estimation Of Skewness Parameters. We simulate a multivariate DGP with three shocks (one is left-skewed, one is right-skewed and one is Gaussian) a large number of times and estimate the underlying shock parameters with Maximum likelihood. Overall the estimates using the *Pruned Skewed Kalman Filter* are convincingly good for both a very low and a rather large pruning threshold. Most mass is centered around the true values and the distribution becomes narrower with larger sample sizes. The *Pruned Skewed Kalman Filter* successfully uncovers the skewed distribution of the first two shocks, but also Gaussianity of the last shock. The Gaussian Kalman filter completely misses the skewed distribution of η_t; which is evident in biased and inflated estimates of μ_η and Σ_η (which in the Gaussian case are estimates of $E[\eta_t]$ and $V[\eta_t]$).

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.

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:

$$y_t(\tau) = L_t + S_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + C_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \tag{17}$$

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:

$$\begin{pmatrix} L_t - \mu^L \\ S_t - \mu^S \\ C_t - \mu^C \end{pmatrix} = \begin{pmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} & G_{23} \\ G_{31} & G_{32} & G_{33} \end{pmatrix} \begin{pmatrix} L_{t-1} - \mu^L \\ S_{t-1} - \mu^S \\ C_{t-1} - \mu^C \end{pmatrix} + \begin{pmatrix} \eta_t^L \\ \eta_t^S \\ \eta_t^C \end{pmatrix} \tag{18}$$

Obviously, equation (18) 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 (17):

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{pmatrix} \tag{19}$$

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)-type 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 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 assess 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 prune 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.[5] 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 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 1, 2 and 3 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 (0.008)	1.0004 (0.009)	0.0285 (0.009)	0.0253 (0.009)	−0.0222 (0.011)	−0.0218 (0.011)	8.2506 (1.086)	6.5516 (3.445)
S_t	−0.0303 (0.016)	−0.0015 (0.014)	0.9385 (0.018)	0.9767 (0.019)	0.0395 (0.021)	0.0399 (0.020)	−1.3786 (0.499)	−1.3411 (0.925)
C_t	0.0244 (0.023)	0.0085 (0.024)	0.0232 (0.026)	−0.0005 (0.027)	0.8428 (0.031)	0.8491 (0.030)	−0.3647 (0.383)	−0.3324 (0.476)

Table 1: 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 (0.008)	0.1906 (0.046)	−0.0140 (0.011)	−0.0668 (0.052)	0.0436 (0.019)	0.1648 (0.105)	0	−3.4648 (0.683)
S_t			0.3823 (0.030)	0.7546 (0.115)	0.0092 (0.034)	0.0565 (0.142)	0	−1.9895 (0.244)
C_t					0.8019 (0.081)	1.6045 (0.354)	0	1.2147 (0.225)

Table 2: 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.

[5] 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)).

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

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

Table 3: 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.

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 1) 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 1), indicating how the estimates with the conventional Kalman filter adapt to the neglected skewness in η_t. Σ_η (first six columns of table 2) 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. 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 3, 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 2). 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 $\eta_{t|T}$ 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 all error term distributions, but also some bulging dents, which are both in clear contradiction to a symmetric distribution.[6] 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 Gaussianity as a restriction ($\Gamma_\eta = 0$), we perform a likelihood ratio test and obtain a 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.

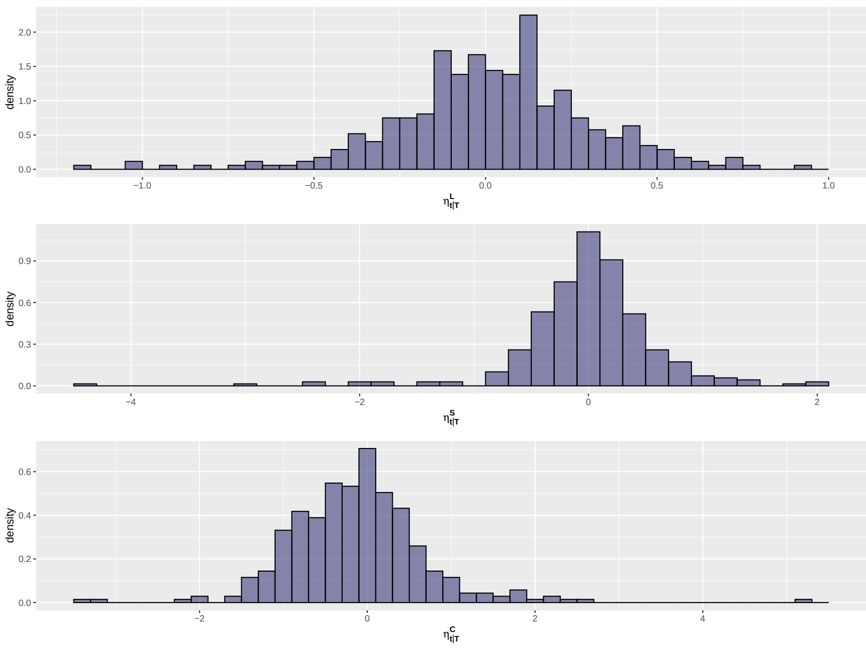

Figure 4: Histogram of smoothed innovations $\eta_{t|T}^L$, $\eta_{t|T}^S$ and $\eta_{t|T}^C$.

7. Asymmetric shocks in a New Keynesian DSGE model for US data

Estimating the structural parameters of DSGE models with Maximum Likelihood is generally not a common approach in the literature, due to the dilemma of absurd parameters and pile-up phenomena on the

[6]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.

boundary of the theoretically admissible parameter space (An & Schorfheide, 2007; Andreasen, 2010; Morris, 2017). Nevertheless, Ireland (2004) is one of the few papers that successfully employs Maximum Likelihood for structural estimation of a log-linearized New Keynesian DSGE model to asses which shocks are the major drivers of aggregate fluctuations in post-war US data. In what follows, we revisit this estimation exercise but use our proposed *Pruned Skewed Kalman Filter* for estimation. The log-linearized model equations are given by:

$$\hat{x}_t = \hat{y}_t - \omega\hat{a}_t \tag{20}$$

$$\hat{g}_t = \hat{y}_t - \hat{y}_{t-1} + \hat{z}_t \tag{21}$$

$$\hat{x}_t = \alpha_x\hat{x}_{t-1} + (1-\alpha_x)E_t\hat{x}_{t+1} - (\hat{r}_t - E_t\hat{\pi}_{t+1}) + (1-\omega)(1-\rho_a)\hat{a}_t \tag{22}$$

$$\hat{\pi}_t = \beta\left(\alpha_\pi\hat{\pi}_{t-1} + (1-\alpha_\pi)E_t\hat{\pi}_{t+1}\right) + \psi\hat{x}_t - \hat{e}_t \tag{23}$$

$$\hat{r}_t - \hat{r}_{t-1} = \rho_\pi\hat{\pi}_t + \rho_x\hat{x}_t + \rho_g\hat{g}_t + \eta_{r,t} \tag{24}$$

$$\hat{a}_t = \rho_a\hat{a}_{t-1} + \eta_{a,t}, \qquad \hat{e}_t = \rho_e\hat{e}_{t-1} + \eta_{e,t}, \qquad \hat{z}_t = \eta_{z,t}$$

where all hat variables are in log deviations from their non-stochastic steady-state. These equations are based on the optimal behavior of utility-maximizing households and profit-maximizing firms within a staggered price setting framework. Specifically, the first equation (20) defines the output gap, $\hat{x}_t$, which measures the deviation of actual output, $\hat{y}_t$, from its natural level, $\omega\hat{a}_t$, in the absence of nominal rigidities. ω is a parameter related to the Frisch elasticity of labor and $\hat{a}_t$ is an autoregressive preference shifter process with persistence parameter ρ_a and subject to preference shocks $\eta_{a,t}$. The second equation (21) defines the growth rate $\hat{g}_t$ of output subject to productivity shocks $\eta_{z,t}$. The third equation (22) describes the New Keynesian IS curve, which relates the output gap to the expectations of a future expected output gap, the ex-ante real interest rate – defined as the difference between the nominal interest rate $\hat{r}_t$ and expected inflation $E_t\hat{\pi}_{t+1}$ – and the exogenous preference shock. The parameter α_x allows for some additional flexibility for the lagged output gap to play a role in determining x_t, e.g. due to consumption habit formation. The fourth equation (23) is a forward-looking New Keynesian Phillips curve, which implies that the output gap drives the dynamics of inflation relative to expected inflation. The parameter β is the discount factor, ψ the slope of the curve (influenced by the strength of nominal rigidities) and α_π allows for a backward-looking component, e.g. due to nominal wage rigidities or indexation of prices and wages to past inflation. The equation is subject to a cost-push process $\hat{e}_t$ which evolves according to an autoregressive process with parameter ρ_e. A decrease in $\hat{e}_t$ lowers the elasticity of demand for each intermediate good and hence increases markups of the monopolistically competitive firms; hence, $\eta_{e,t}$ is a negative cost-push shock. Finally, in equation (24) monetary policy is described by a feedback rule that determines the change in the nominal interest rate, based on deviations from inflation, output gap, and output growth from their steady-state targets. ρ_π, ρ_x

and ρ_g are the sensitivity parameters of systematic monetary policy and $\eta_{r,t}$ captures any non-systematic deviation from the rule. Due to rational expectations all agents know the exact model equations and the statistical distribution of the white noise process $\eta_t = [\eta_{a,t}, \eta_{e,t}, \eta_{z,t}, \eta_{r,t}]'$ for all t. Accordingly, E_t is the expectation operator conditional on the information set in period t; namely, the state of the economy (all variables up to and including $t-1$) and the values of current shocks η_t. Subject to restrictions on the space of model parameters $\theta = (\beta, \psi, \omega, \alpha_x, \alpha_\pi, \rho_\pi, \rho_x, \rho_g, \rho_a, \rho_e)$ that yield stable and unique trajectories (Blanchard & Kahn, 1980) a stochastic solution is characterized by a recursive decision rule, so-called policy function, which for a log-linearized model (i.e. a perturbation solution at first order) resembles a linear state-space form as in equations (3) and (4):

$$X_t = G(\theta)X_{t-1} + R(\theta)\eta_t$$
$$Y_t = FX_t$$

where $X_t = [\hat{x}_t, \hat{y}_t, \hat{g}_t, \hat{z}_t, \hat{\pi}_t, \hat{a}_t, \hat{e}_t, \hat{r}_t]'$ is the vector of all endogenous and $Y_t = [\hat{g}_t, \hat{\pi}_t, \hat{r}_t]'$ collects the observable variables. We consider the same set of quarterly macroeconomic time series for the 1980Q1-2003:Q1 period as originally used in Ireland (2004):[7] (1) Demeaned quarterly changes in seasonally adjusted real GDP, converted to per capita values by dividing by the civilian noninstitutional population aged 16 and over, are used to measure output growth $\hat{g}_t$. (2) Demeaned quarterly changes in the seasonally adjusted GDP deflator provide the measure of inflation $\hat{\pi}_t$. (3) Demeaned quarterly averages of daily values of the three-month U.S. Treasury bill rate provide the measure of the nominal interest rate $\hat{r}_t$. While F is simply a matrix of zeros and ones, the reduced-form parameters G and R are nonlinear functions of the structural model parameters θ, which we recover for any candidate θ using Dynare's first-order perturbation solution algorithm as described in Villemot (2011).

From the 10 underlying structural parameters in the model, two are held fix, $\beta = 0.99$ and $\psi = 0.1$, and are not estimated. Hence, our interest centers around the other 8 model parameters plus the parameters of the distribution of η_t, which we will estimate by minimizing the negative log-likelihood function. Common practice is to assume that the shocks η_t are distributed as multivariate normal; we, however, assume that each follows an independent univariate skew-normal distribution, i.e. $\eta_{j,t} \sim CSN(\mu_{\eta_j}, \Sigma_{\eta_j}, \Gamma_{\eta_j}, 0, 1)$ for $j \in \{a, e, z, r\}$. As all η_j are independent of each other, we make use of univariate closed-form formulas for the standard error and skewness coefficient and accordingly estimate $stderr(\eta_{j,t})$ and $skew(\eta_{j,t})$ instead of Σ_{η_j} and Γ_{η_j}. This enables us to calculate standard errors directly, equivalent robustness results with estimated Σ_{η_j} and Γ_{η_j} are available in the replication codes. Similarly, μ_{η_j} is an endogenous parameter which, given Σ_{η_j} and Γ_{η_j}, needs to be set to ensure $E[\eta_j] = 0$.

[7] The same model and dataset is used by Chib & Ramamurthy (2014) to illustrate estimating DSGE models with Student's t distributed shocks.

Some computational remarks are noteworthy. First, for DSGE models G is typically a singular matrix. To overcome the numerical issues in the prediction step, we compute, filter and smooth the parameters of the joint distribution of $[x_t', \eta_t']'$ instead of x_t. Similarly, the pre-multiplication of η_t with R in the state transition equation is without loss of generality, as we can use Property 2 and work with the linearly transformed distribution $(R\eta_t)$. Second, the initial distribution for the prediction-error decomposition of the likelihood is set to a normal one with mean zero and initial covariance matrix of the error of the forecast set equal to the unconditional variance of the state variables. The likelihood is penalized if the Blanchard & Kahn (1980) conditions are violated (i.e. a DSGE specific generalization of Eigenvalues of G being outside the unit circle) or the covariance matrix of η_t is not positive semi-definite. Third, based on our Monte Carlo evidence, we prune the skewness dimensions at a threshold level of 1%. Fourth, to impose bounds on the estimated parameters we apply a change of variables to ensure that variables are bound within their natural domain using a scaled and translated log-odds transform. For instance, if $x > 0$, we estimate $log(x)$. Similarly, if x must be between 0 and 1 we estimate $logit(x) = log\frac{x}{1-x}$. Fifth, when computing standard errors via the inverse Hessian method the transformations are reversed such that reported standard errors are for the actual model parameters. The Hessian is computed using a standard two-sided finite difference approach; however, when an estimated parameter is on the boundary, we follow Ireland (2004) and use one-sided finite differences for this parameter. Sixth, we do a sophisticated search for initial parameter values. In more detail, we use the values reported in Ireland (2004) as our starting point for the model and standard error parameters. Next, we create an evenly spaced grid of the skewness parameters for all four shocks, while keeping the variance constant to the Gaussian estimates. For each combination of variance and skewness, we recover the corresponding Σ_{η_j} and Γ_{η_j} combinations analytically and compute the negative log-likelihood of each value on the grid. In this way, we examine the likelihood surface for over 50000 combinations for all shocks having either positive, negative, high, mild, low, or no skewness in their distribution. We then take the best three combinations and use these as initial values to optimize over both standard error as well as skewness parameters (while keeping model parameters fixed) with the Pruned Skewed Kalman filter. The best estimates for the shock parameters are then combined with the Gaussian estimates for the model parameters to arrive at our chosen initial values for the actual estimation. Finally, equipped with these initial values, we run various gradient-based and simulation-based optimization routines (in parallel) to minimize the negative log-likelihood function with respect to all parameters, see Andreasen (2010) for a discussion of appropriate optimizers in the context of maximum likelihood estimation of DSGE models. Table 4 contains the final estimation results.

Overall, using the CSN distribution instead of the Gaussian one is preferred by the data as indicated by the increase in the value of the maximized log-likelihood function. To support this claim, a likelihood ratio test was conducted – as the *Skewed Kalman filter* nests Gaussianity – yielding a test statistic of 16.55 and a p-value of 0.0024. Examining the model parameters, the estimates of α_x and α_π are similar and close to

Model Parameters	*KF*	*PSKF*	*Shock Parameters*	*KF*	*PSKF*
ω	0.0581 (0.0877)	0.1536 (0.0157)	$skew(\eta_a)$	0	−0.2132 (0.0264)
α_x	0.0000 (0.0043)	0.0000 (0.0025)	$skew(\eta_e)$	0	−0.2221 (0.0273)
α_π	0.0000 (0.0025)	0.0000 (0.0020)	$skew(\eta_z)$	0	−0.9499 (0.1613)
ρ_π	0.3865 (0.1273)	0.2729 (0.0271)	$skew(\eta_r)$	0	0.8099 (0.0482)
ρ_g	0.3960 (0.0650)	0.3399 (0.0228)	$stderr(\eta_a)$	0.0302 (0.0166)	0.0249 (0.0015)
ρ_x	0.1654 (0.0615)	0.2838 (0.0057)	$stderr(\eta_e)$	0.0002 (0.0001)	0.0002 (0.0001)
ρ_a	0.9048 (0.0596)	0.9155 (0.0183)	$stderr(\eta_z)$	0.0089 (0.0015)	0.0079 (0.0012)
ρ_e	0.9907 (0.0155)	0.9816 (0.0302)	$stderr(\eta_r)$	0.0028 (0.0004)	0.0028 (0.0002)
Value of maximized Log-Likelihood function:				1207.56	1215.84

Table 4: Parameter estimates. *KF* denotes the conventional Kalman filter and *PSKF* the pruned skewed Kalman filter. Asymptotic standard errors appear in parenthesis.

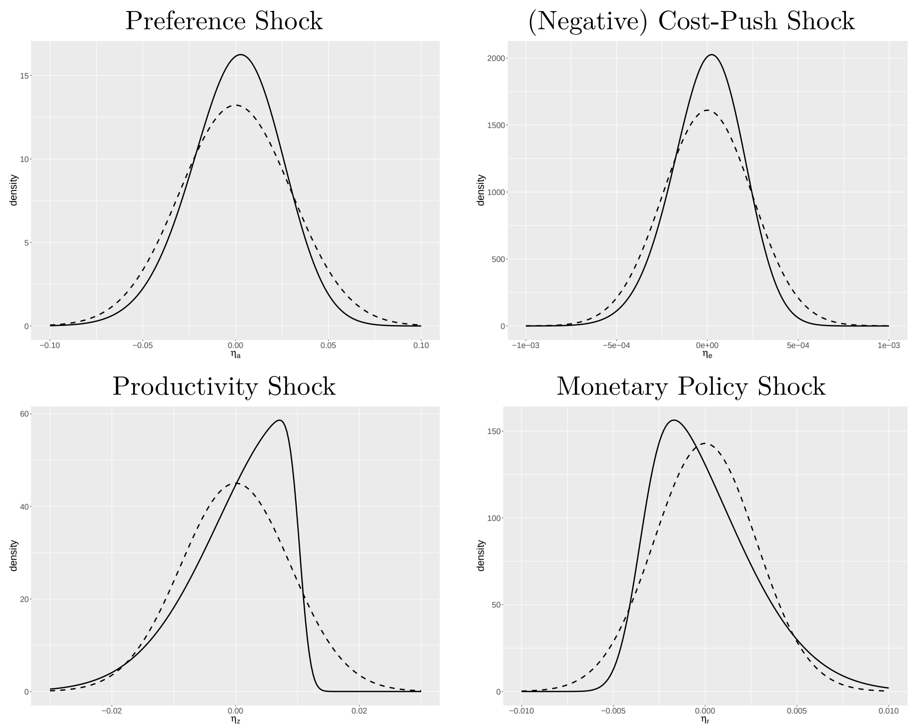

Figure 5: Estimated Probability Density Functions

Notes: solid lines are estimated CSN distributions and dashed lines are estimated Gaussian distributions.

zero, suggesting that backward-looking behavior of consumers and firms is not important in both the New Keynesian IS and Phillips curve. Notably, the policy parameters ρ_π, ρ_g, and ρ_x differ, indicating that the Federal Reserve's *systematic policy* is more responsive to movements in output gap than output growth.

Furthermore, the rule-based inflation sensitivity parameter is smaller than the Gaussian estimate, though still within the confidence band. While estimates of the persistence parameters ρ_a and ρ_e of the preference and cost-push shocks are not different across filters, the estimate of ω is. This has two implications: First, in the underlying theoretical model, a higher value of ω implies a more elastic labor supply schedule. Second, in the empirical model, the impact of the preference shock on the efficient level of output is estimated to be larger.

Turning towards the shock parameters, the estimates for the standard errors of the cost-push and monetary policy shocks are nearly identical, while the Gaussian filter slightly overestimates the standard errors of the preference and productivity shocks. Interestingly, statistically significant skewness coefficients are found for all shocks, with particularly strong skewness observed for the productivity and monetary policy shocks. To illustrate these differences, we depict the estimated probability density functions in figure 5. The CSN distribution (solid line) of the monetary policy shock has less mass in the left tail and more mass in the right tail than the estimated normal distribution with the same standard deviation (dashed line). This reflects the Federal Reserve's unanticipated hawkish policies during Paul Volcker's tenure as chairman. Accordingly, combined with the evidence of slightly less systematic monetary policy, suggests that large surprises of monetary tightening are more likely than large monetary easing ones. Analogously, the distribution of the productivity shock has more mass in the left tail and less mass in the right tail than the estimated normal distribution with (almost) the same standard deviation (dashed line). This pattern for productivity is consistent with the estimates of Ruge-Murcia (2017) and captures events like the dot-com bubble in the sample. The preference and (negative) cost-push shocks are estimated to have mild negative skewness.

Finally, we investigate the consequences of a one-time monetary policy shock on the model variables as depicted in Figure 6. Solid lines represent estimates using the Pruned Skewed Kalman filter (column *PSKF* in Table 4), while dashed lines correspond to Gaussian Kalman filter estimates (column *KF* in Table 4). It is essential to differentiate between positive and negative shocks in the presence of asymmetry. Consequently, we adhere to the standard practice, defining the size of a *typical* shock as ± 1 standard deviation. This aligns with the *empirical rule*, which equates to the 16th and 84th percentiles of the normal distribution. Thus, we employ the 16th percentiles of the estimated normal and CSN distributions as typical negative shocks, while the 84th percentiles are used for typical positive shocks. Even though the systematic parameters are estimated slightly different, the shape of the impulse-response function is qualitatively the same, so we refer to Ireland (2004) for a discussion of the economic transmission channels. However, we underscore that, even in a linear model, *size and direction* matter for conducting monetary policy, as the transmission channels of typical monetary easing versus monetary tightening shocks are asymmetric. In a broader context, the presence of skewed shocks leads to the propagation of asymmetry through differently amplified transmission channels, ultimately resulting in asymmetric business cycles.

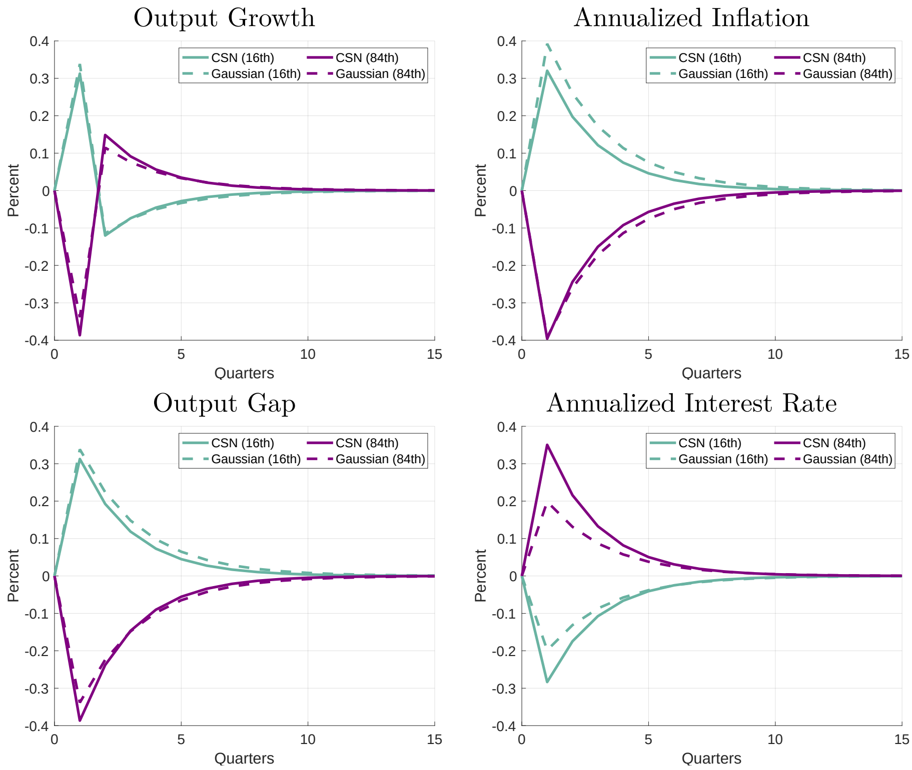

Figure 6: Impulse Responses: Monetary Policy Shock

8. Conclusion

The Skewed Kalman Filter is an analytical recursive method for inferring the state vector in linear state-space systems and can be used to compute the exact likelihood function when innovations originate from the CSN distribution. Intriguingly, the Skewed Kalman Filter encompasses both Gaussianity and the skew-normal distribution as special cases. Applying this filter to data demands substantial computational resources or is even unfeasible for multivariate models or large sample sizes because it involves the evaluations of high-dimensional multivariate normal cdfs of growing dimensions. We introduce a fast and intuitive pruning algorithm for the filter's updating step, overcoming this *curse of increasing dimensions.* We provide theoretical evidence for its validity across any dataset and parameter values. Our *Pruned Skewed Kalman Filter and Smoother* operate effectively and efficiently in practice, as demonstrated in our comprehensive Monte Carlo study and two multivariate real data applications.

Naturally, there are several other methods and algorithms for statistical inference of time series with asymmetric distributions. For example, sequential Monte Carlo methods can be easily adapted to skewed distributions, although the computational complexity and runtime of these filters increase rapidly with the state dimension. Skewness can also be modeled using a mixture of normal distributions, for which numerous filtering algorithms exist. However, as recently noted by Nurminen et al. (2018), Gaussian mixtures have exponentially decaying tails and can be overly sensitive to outlier measurements, while the computational cost of a mixture reduction algorithm is substantial. Bayesian methods are often tailored to specific modeling frameworks and assumptions, enabling fine-tuning of certain sampling algorithms, such as combining a Gibbs sampler with Metropolis-Hastings stages, as exemplified in Karlsson et al. (2023) for Vectorautoregressive models. We do not assert that the *Pruned Skewed Kalman Filter* inherently outperforms these approaches, but we contend that its ease of use and compatibility with existing toolboxes and standard estimation methods will promote its adoption across various disciplines.

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. 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.

An, S., & Schorfheide, F. (2007). Bayesian Analysis of DSGE Models. *Econometric Reviews*, *26*, 113–172. doi:10.1080/07474930701220071.

Andreasen, M. (2010). How to Maximize the Likelihood Function for a DSGE Model. *Computational Economics*, *35*, 127–154. doi:10.1007/s10614-009-9182-6.

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.

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. arXiv: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.

Blanchard, O. J., & Kahn, C. M. (1980). The Solution of Linear Difference Models under Rational Expectations. *Econometrica*, *48*, 1305–1311. doi:10.2307/1912186.

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.

Chib, S., & Ramamurthy, S. (2014). DSGE Models with Student-t Errors. *Econometric Reviews*, *33*, 152–171. doi:10.1080/07474938.2013.807152.

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.

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.

Gallier, J. (2011). Schur Complements and Applications. In *Geometric Methods and Applications* (pp. 431–437). New York, NY: Springer New York volume 38. URL: 10.1007/978-1-4419-9961-0_16.

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.

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.

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

Ireland, P. N. (2004). Technology Shocks in the New Keynesian Model. *Review of Economics and Statistics*, *86*, 923–936. doi:10.1162/0034653043125158.

Jeong, M. (2023). A numerical method to obtain exact confidence intervals for likelihood-based parameter estimators. *Journal of Statistical Planning and Inference*, *226*, 20–29. doi:10.1016/j.jspi.2022.12.006.

Karlsson, S., Mazur, S., & Nguyen, H. (2023). Vector autoregression models with skewness and heavy tails. *Journal of Economic Dynamics and Control*, *146*, 104580. doi:10.1016/j.jedc.2022.104580.

Kim, H.-M., Ryu, D., Mallick, B. K., & Genton, M. G. (2014). Mixtures of skewed Kalman filters. *Journal of Multivariate Analysis*, *123*, 228–251. URL: https://linkinghub.elsevier.com/retrieve/pii/S0047259X13001942. doi:10.1016/j.jmva.

2013.09.002.

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.

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.

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

Morris, S. D. (2017). DSGE pileups. *Journal of Economic Dynamics and Control*, *74*, 56–86. doi:10.1016/j.jedc.2016.11.002.

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. arXiv: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.

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.

Villemot, S. (2011). *Solving Rational Expectations Models at First Order: What Dynare Does*. Dynare Working Papers 2 CEPREMAP.

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*. Discussion Paper Freien Universität Berlin Berlin. URL: http://dx.doi.org/10.17169/refubium-33629.

Zhu, X., Wei, Z., & Wang, T. (2022). Multivariate Skew Normal-Based Stochastic Frontier Models. *Journal of Statistical Theory and Practice*, *16*, 20. doi:10.1007/s42519-022-00249-9.