Pruned skewed Kalman filter and smoother with application to DSGE models

January 2025

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

[a] *Center for Quantitative Economics, University of Münster, Am Stadtgraben 9, 48143 Münster, Germany.*
[b] *Department of Economics, University of Tübingen, Mohlstr. 36, 72074 Tübingen, Germany.*
[c] *Dynare Team, CEPREMAP, 48, Boulevard Jourdan, 75014 Paris, France.*

Abstract

The *skewed Kalman filter* (SKF) extends the classical *Gaussian Kalman filter* (KF) by accommodating asymmetric (skewed) error distributions in linear state-space models. We introduce a computationally efficient method to address the *curse of increasing skewness dimensions* inherent in the SKF. Building on insights into how skewness propagates through the state-space system, we derive an algorithm that discards elements in the cumulative distribution functions which do not affect asymmetry beyond a pre-specified numerical threshold; we refer to this approach as the *pruned skewed Kalman filter* (PSKF). Through extensive simulation studies on both univariate and multivariate state-space models, we demonstrate the proposed method's accuracy and efficiency. Furthermore, we are first to derive the *skewed Kalman smoother* and implement its pruned variant. We illustrate its practical relevance by estimating a linearized New Keynesian DSGE model with U.S. data under both maximum likelihood and Bayesian MCMC frameworks. The results reveal a strong preference for skewed error distributions, especially in productivity and monetary policy shocks.

Keywords: state-space models, skewed Kalman filter, skewed Kalman smoother, closed skew-normal, dimension reduction, asymmetric shocks, DSGE

JEL: C32, C51, E32, E43

1. Introduction

The Kalman filter is a highly effective recursive procedure for making inference about state vectors, which can then be used to precisely compute the Gaussian likelihood function. The filter is optimal in the sense that it minimizes the covariance matrix of one-step ahead prediction errors. More importantly, the Kalman filter can be executed swiftly and efficiently from an applied and computational standpoint. However, non-Gaussianity—such as 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 candidate, 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 nests both the normal as well as 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 filtering steps (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 has 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 and infeasible if the dimension of a cdf becomes very large. 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. 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

1

as the recursion proceeds over many time steps, a point recently echoed by Amsler et al. (2021) for the skew-normal distribution.

To address this challenge, our primary contribution is to derive 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* (SKF) to morph into the *Gaussian Kalman filter* (KF). Our approach is hence based on a low numerical 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* (PSKF). 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 concept.

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. Importantly, in simulation studies where the SKF can also be used as a benchmark, we demonstrate that the approximation error introduced by our pruning scheme is negligible.

Finally, we illustrate the usefulness of the pruned skewed Kalman filter and smoother by estimating the linearized New Keynesian *Dynamic Stochastic General Equilibrium* (DSGE) model of Ireland (2004).[1] The reasons for choosing this model are threefold. First, it allows analysis of typical macroeconomic shocks (preference, cost-push, productivity, and monetary policy) in a stylized framework that remains representative of modern macroeconomic models while preserving clear intuition. Second, the model can be estimated via both maximum likelihood and Bayesian *Markov Chain Monte Carlo* (MCMC) methods, enabling a straightforward comparison of the PSKF across different estimation frameworks. Third, the exact same model and dataset has been used by Chib & Ramamurthy (2014) under the assumption of a multivariate Student-t distribution for the structural shocks—capturing excess kurtosis but not skewness. Our estimation results reveal substantial asymmetry in the distributions of productivity and monetary policy shocks. This finding aligns with suggestions by Curdia et al. (2014) and Lindé et al. (2016), who hypothesize that skewness might be a salient feature of the shock distribution, but do not account for it in their estimation procedures as we do here.

Our presentation and implementation of the pruned skewed Kalman filter and smoother are designed to remain highly general, mirroring the simplicity of the standard normal Kalman filtering and smoothing routines. In terms of modeling, researchers can retain their linear state-space framework while introducing additional flexibility by assuming a CSN distribution for the error terms in the state transition equation (the Gaussian and skew-normal case being nested). On the computational side, any estimation procedure—whether Bayesian or Frequentist—that employs Kalman filtering can be readily adapted by replacing the underlying Kalman filtering routine.

To facilitate widespread use, we provide model-independent implementations of the pruned skewed Kalman filter and smoother in Julia, MATLAB, Python, and R.[2] Notably, we have also developed a toolbox that integrates the pruned skewed Kalman filter and smoother into Dynare (Adjemian et al., 2024)—the leading software for estimating DSGE models—making it compatible with both maximum likelihood and Bayesian MCMC methods.[3]

---

[1] In an older working paper version (Guljanov et al., 2022) we also estimate the multivariate *Dynamic Nelson-Siegel* (DNS) term structure model of Diebold et al. (2006), demonstrating that the data strongly favor a skewed distribution for the error terms of all three yield curve factors. Likewise, in his PhD thesis, Guljanov (2024) estimates a more complex DSGE model with a richer set of features, shocks and observables—the (Smets & Wouters, 2007) model—using the PSKF and Bayesian MCMC methods.

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

[3] This feature is planned for release in Dynare 7.0; see https://git.dynare.org/wmutschl/dynare/-/tree/pskf.

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 rarely used in practice, despite its considerable potential and simplicity of 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*. 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 form, and it is also subject to the *curse of increasing skewness dimensions*. Moreover, they only provide numerical examples in univariate settings, whereas we provide simulation evidence and 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

4

practical purposes, one must either assume simplified conditions or refit the updated distribution. In this paper, we specifically choose to employ the latter strategy.

In the DSGE literature, Lindé et al. (2016) emphasize the importance of moving beyond Gaussian assumptions, noting that the shocks driving most recessions are non-Gaussian. Chib & Ramamurthy (2014) and Curdia et al. (2014) demonstrate that linearized DSGE models with Student-t-distributed shocks can outperform their Gaussian counterparts. We show that this improvement also holds when skewness is accounted for. Similar to us, Grabek et al. (2011) augment a linearized DSGE model with CSN innovations, but propose an ad-hoc two-step estimator. Our approach is to directly utilize the SKF without relying on auxiliary estimation steps that might bias the estimation results.

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 PSKF 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.

Structure

The paper is organized as follows. Section 2 provides an overview of the closed skew-normal (CSN) distribution, introducing its representation and main properties. Although no new results are presented there, it serves as a primer on the distribution and establishes the notation and concepts needed for filtering and smoothing. Section 3 presents the closed-form expressions for both the forward and backward recursions of the skewed Kalman filter and smoother. While the filtering steps follow Naveau et al. (2005) and Rezaie & Eidsvik (2014), the smoothing steps are

novel. Next, Section 4 contains our main contribution as we show how skewness propagates through the state-space system over time and use this insight to develop our pruning algorithm. Section 5 summarizes the main Monte Carlo findings (with detailed results in an online appendix), while Section 6 contains the estimation of the linearized New Keynesian DSGE model on U.S. data. Finally, Section 7 concludes with closing remarks.

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

$$W = \mu + E_1 \qquad \text{and} \qquad Z = -\nu + \Gamma E_1 + E_2.$$

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 CSN distribution

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

The moment generating function (mgf) of X is

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

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

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

6

Figure 1: Density functions of univariate CSN distributions with different skewness parameters

(a) $\Gamma = -5$, $\nu = 0$

(b) $\Gamma = 0$, $\nu = 0$

(c) $\Gamma = 5$, $\nu = 0$

(d) $\Gamma = 5$, $\nu = -8$
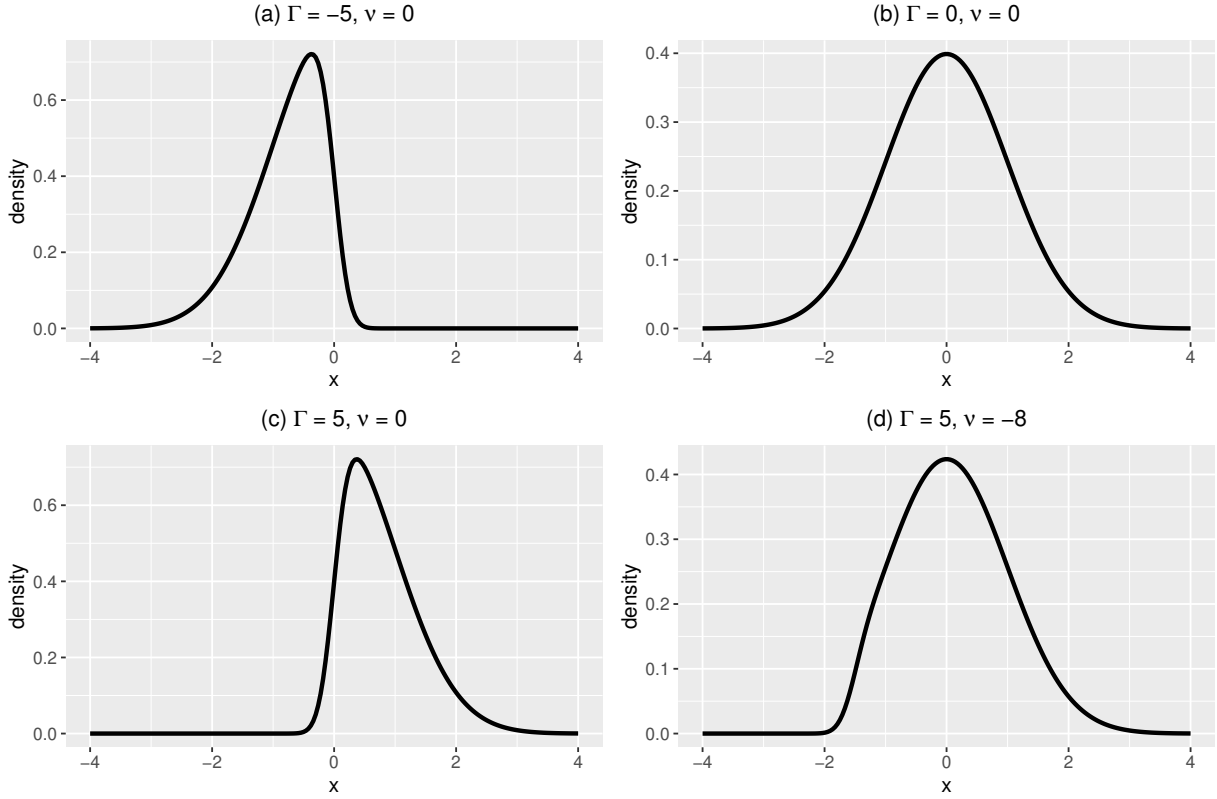
Note: Other parameters are $\mu = 0$, $\Sigma = 1$ and $\Delta = 1$.

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

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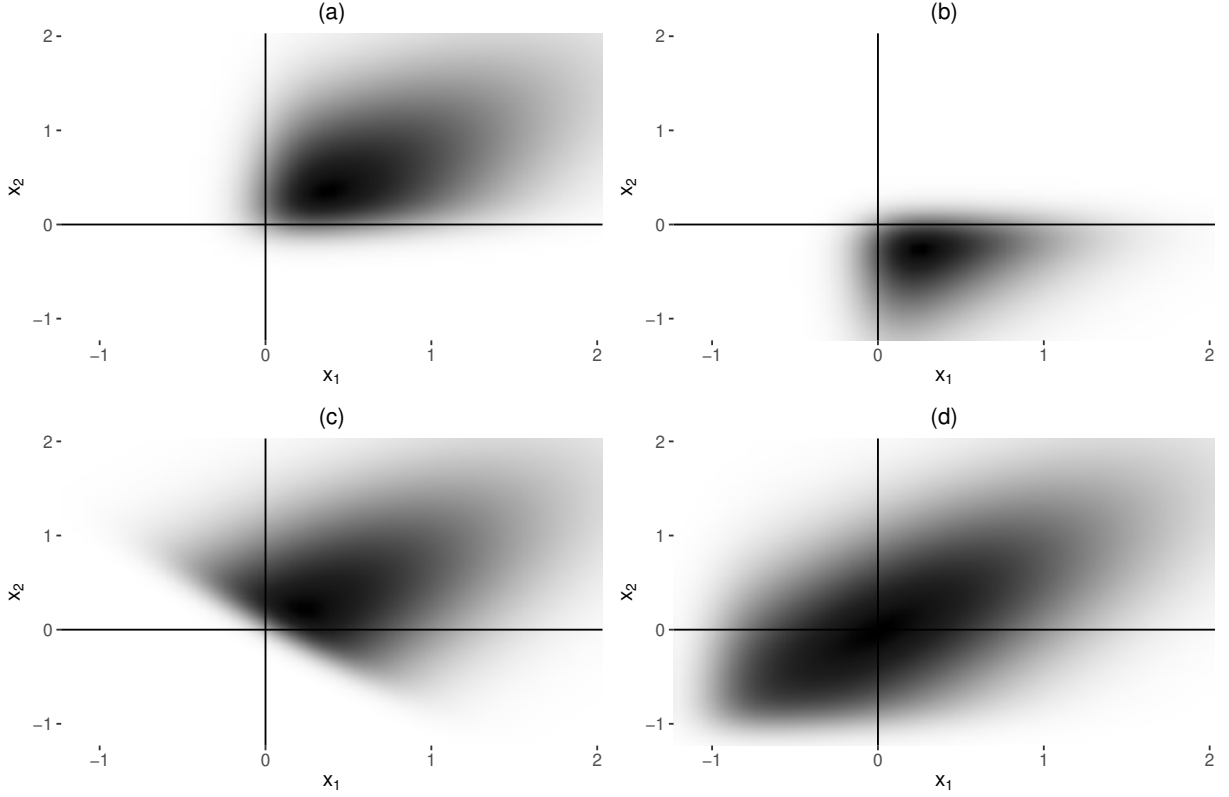
Note: (a) $\Gamma = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}$, $\nu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, (b) $\Gamma = \begin{bmatrix} 6 & 0 \\ 0 & -6 \end{bmatrix}$, $\nu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, (c) $\Gamma = \begin{bmatrix} 6 & 6 \\ 6 & 6 \end{bmatrix}$, $\nu = \begin{bmatrix}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

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).

One can see from (1) that the asymmetric deviation of the CSN distribution from the symmetric Gaussian distribution results from the covariance between W and Z; in other words, it is this correlation that adds skewness to the Gaussian distribution. Hence, the CSN distribution can be regarded as a generalization of the normal distribution and as such inherits several of its properties.

In the following, we review those properties that are of special interest for the skewed Kalman filter and smoother. Proofs can be found in González-Farías et al. (2004a) and González-Farías et al. (2004b).

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

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

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

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

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

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

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

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

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

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.[4] 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:

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

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

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

---

[4]MATLAB R2024b'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.*

11

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 structural innovations by adding an auxiliary state variable to equation (3). In fact, this simplified framework is the one that is most commonly used for the analysis of economic phenomena such as the one we study in Section 6.

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

3.1. Prediction

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 ahead 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

$$\mu_{t|t-1} = G\mu_{t-1|t-1} + \mu_\eta, \quad \Sigma_{t|t-1} = G\Sigma_{t-1|t-1}G' + \Sigma_\eta, \tag{6}$$

12

$$\Gamma_{t|t-1} = \begin{pmatrix} \Gamma_{t-1|t-1}\Sigma_{t-1|t-1}G'\Sigma_{t|t-1}^{-1} \\ \Gamma_\eta \Sigma_\eta \Sigma_{t|t-1}^{-1} \end{pmatrix}, \quad \nu_{t|t-1} = \begin{pmatrix} \nu_{t-1|t-1} \\ \nu_\eta \end{pmatrix}, \quad \Delta_{t|t-1} = \begin{pmatrix} \Delta_{t|t-1}^{11} & \Delta_{t|t-1}^{12} \\ (\Delta_{t|t-1}^{12})' & \Delta_{t|t-1}^{22} \end{pmatrix},$$

$$(7)$$

with $\Delta_{t|t-1}^{11} = \Delta_{t-1|t-1} + \Gamma_{t-1|t-1}\Sigma_{t-1|t-1}\Gamma'_{t-1|t-1} - \Gamma_{t-1|t-1}\Sigma_{t-1|t-1}G'\Sigma_{t|t-1}^{-1}G\Sigma_{t-1|t-1}\Gamma'_{t-1|t-1}$,
$\Delta_{t|t-1}^{22} = \Delta_\eta + \Gamma_\eta \Sigma_\eta \Gamma'_\eta - \Gamma_\eta \Sigma_\eta \Sigma_{t|t-1}^{-1}\Sigma_\eta \Gamma'_\eta$, and $\Delta_{t|t-1}^{12} = -\Gamma_{t-1|t-1}\Sigma_{t-1|t-1}G'\Sigma_{t|t-1}^{-1}\Sigma_\eta \Gamma'_\eta$.

3.2. Updating

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

$$x_{t|t} \sim CSN_{p,q_t}(\mu_{t|t}, \Sigma_{t|t}, \Gamma_{t|t}, \nu_{t|t}, \Delta_{t|t}), \tag{8}$$

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

$$\mu_{t|t} = \mu_{t|t-1} + K_{t-1}^{Gauss}(y_t - F\mu_{t|t-1} - \mu_\varepsilon),$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_{t-1}^{Gauss}F\Sigma_{t|t-1}, \tag{9}$$

$$\Gamma_{t|t} = \Gamma_{t|t-1}, \tag{10}$$

$$\nu_{t|t} = \nu_{t|t-1} - K_{t-1}^{Skewed}(y_t - F\mu_{t|t-1} - \mu_\varepsilon),$$

$$\Delta_{t|t} = \Delta_{t|t-1}. \tag{11}$$

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 (8) 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.

3.3. 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

$$\mu_{t|T} = \mu_{t|t} + \Sigma_{t|t}G'\Sigma_{t+1|t}^{-1}(\mu_{t+1|T} - \mu_{t+1|t}),$$

14

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

$$\Gamma_{t|T} = \begin{pmatrix} \Gamma_{t|t} \\ O_t \end{pmatrix}, \quad \nu_{t|T} = \nu_{T|T}, \quad \Delta_{t|T} = \begin{pmatrix} \Delta_{t|t} & 0 \\ 0 & \tilde{\Delta}_t \end{pmatrix},$$

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. Therefore, implementing reduction methods is also crucial from a smoothing perspective.

4. Pruning the skewness dimension

Our approach to reduce the skewness dimension is motivated by 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 parameters Σ, Γ, ν, Δ and a pre-specified pruning threshold* `tol`*.*

1. *Construct and partition the covariance matrix*

$$P = \begin{pmatrix} P_1 & P_2' \\ P_2 & P_4 \end{pmatrix} = \begin{pmatrix} \Sigma & \Sigma \cdot \Gamma' \\ \Gamma \cdot \Sigma & \Delta + \Gamma \cdot \Sigma \cdot \Gamma' \end{pmatrix}. \tag{12}$$

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 <u>row</u> of $\mathtt{abs}(R_2)$. Save it as vector $\mathtt{max\_val}$.*

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 $(\mathtt{max\_val} < \mathtt{tol})$.*
   Save as $\widetilde{P}$.

5. *Compute pruned ν by removing rows corresponding to $(\mathtt{max\_val} < \mathtt{tol})$.*

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:

$$x_{t,t-1} \sim CSN\left(0, 1, \begin{pmatrix} 6 \\ 0.1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0.1 \\ -0.1 & 1 \end{pmatrix}\right) \tag{13}$$

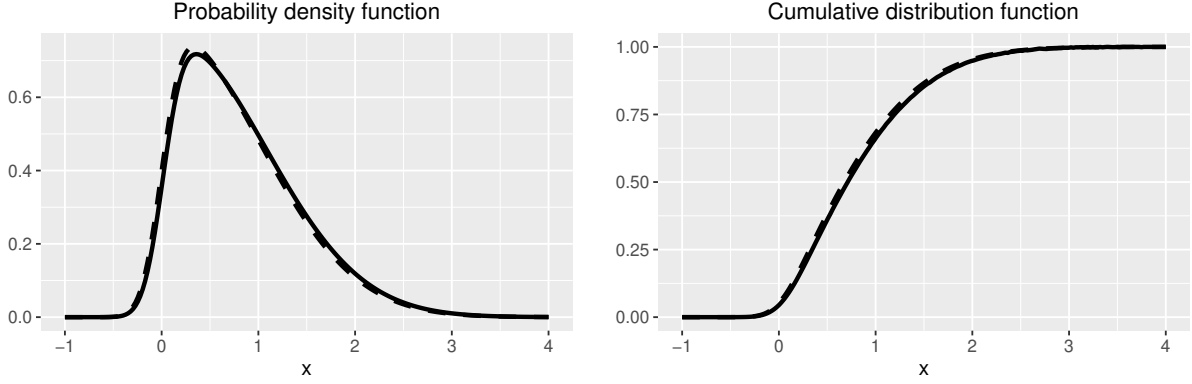with a skewness dimension of 2. Applying *Pruning Algorithm 1* with a (rather large) pruning tolerance $\mathtt{tol} = 0.1$, we get:

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

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

Of course, the skewness dimension can only be reduced if the correlation coefficients are sufficiently small. We now proceed to show that 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 (10), with $\Sigma_{t|t-1}$ as given in (6). Since $\Gamma_{t-1|t-1}$ appears in the upper row in the prediction step (7), the number of rows increases at

Figure 3: Illustration of pruned distribution

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

$$
\Gamma_{t|t} = \begin{pmatrix}
\Gamma_{0|0}\Sigma_{0|0}G'\prod_{j=1}^{t-1}L_j \\
\Gamma_\eta\Sigma_\eta\prod_{j=1}^{t-1}L_j \\
\Gamma_\eta\Sigma_\eta\prod_{j=2}^{t-1}L_j \\
\vdots \\
\Gamma_\eta\Sigma_\eta\prod_{j=t}^{t-1}L_j
\end{pmatrix}\Sigma_{t|t-1}^{-1}, \tag{14}
$$

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 $\Sigma_{t|t-1}$ 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 (14) 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 (10) 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 (11). The recursions in (7) imply that the dimension of $\Delta_{t|t}$ grows each period. The top left element of the partitioned matrix (7) shows that the matrix

$$\Gamma_{t-1|t-1}\Sigma_{t-1|t-1}\Gamma'_{t-1|t-1} - \Gamma_{t-1|t-1}\Sigma_{t-1|t-1}G'\Sigma_{t|t-1}^{-1}G\Sigma_{t-1|t-1}\Gamma'_{t-1|t-1}$$
$$= \Gamma_{t-1|t-1}\Sigma_{t-1|t-1}^{1/2}(I - \Sigma_{t-1|t-1}^{1/2}G'\Sigma_{t|t-1}^{-1}G\Sigma_{t-1|t-1}^{1/2})\Sigma_{t-1|t-1}^{1/2}\Gamma'_{t-1|t-1} \quad (15)$$

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, we have shown the positive definiteness of the matrix (15). As positive definite matrices have strictly positive diagonal elements, the diagonal elements of $\Delta_{t|t}$ keep growing over time. *Pruning Algorithm* 1 reduces the skewness dimension based on the covariances in the bottom left (or top right) partition of the covariance matrix P in (12), i.e. $P_2 \equiv \Gamma_{t|t}\Sigma_{t|t}$. The corresponding correlation of the (i,j)-th element P_2^{ij} is

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

18

As we have shown above, each element of the $\Gamma_{t|t}$ 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 (Durbin & Koopman, 2012, Sec. 2.11). 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 study. 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.[5] We simulate 2400 sample paths for x_t and y_t of different lengths $(40, 80, 110)$ 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, the pruning threshold does not matter measurably in the univariate case and makes only a small numerical difference in multivariate settings.

---

[5]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.

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 hundreds or thousands. 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 in 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}.

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 PSKF 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 $V[\eta_t]$).

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.

6. Asymmetric shocks in a New Keynesian DSGE model

The model of Ireland (2004), representative of modern macroeconomic DSGE frameworks, is prototypical of how macroeconomic shocks—such as preference, cost-push, productivity, and monetary policy innovations—affect the economy in a stylized setting. Such models are usually estimated using Bayesian methods to circumvent issues like the *dilemma of absurd parameters* and *pile-up* phenomena at the boundary of the theoretically admissible parameter space (An & Schorfheide, 2007; Andreasen, 2010; Morris, 2017). However, Ireland (2004) is one of the few studies that successfully applies maximum likelihood to structurally estimate a log-linearized New Keynesian DSGE model, thereby identifying the primary drivers of aggregate fluctuations in post-war US data. We adopt the model and data to illustrate the performance of the pruned skewed Kalman filter and smoother in a real-world context—demonstrating its application under both maximum likelihood and Bayesian estimation frameworks. This dual approach highlights the filter's and smoother's versatility and ease of implementation across different estimation settings.

6.1. Model equations

The log-linearized model equations are given by:

$$\hat{x}_t = \hat{y}_t - \omega \hat{a}_t, \tag{16}$$

$$\hat{g}_t = \hat{y}_t - \hat{y}_{t-1} + \hat{z}_t, \tag{17}$$

$$\hat{x}_t = \alpha_x \hat{x}_{t-1} + (1 - \alpha_x) E_t \hat{x}_{t+1} - (\hat{r}_t - E_t \hat{\pi}_{t+1}) + (1 - \omega)(1 - \rho_a)\hat{a}_t, \tag{18}$$

$$\hat{\pi}_t = \beta \left(\alpha_\pi \hat{\pi}_{t-1} + (1 - \alpha_\pi) E_t \hat{\pi}_{t+1} \right) + \psi \hat{x}_t - \hat{e}_t, \tag{19}$$

$$\hat{r}_t - \hat{r}_{t-1} = \rho_\pi \hat{\pi}_t + \rho_x \hat{x}_t + \rho_g \hat{g}_t + \eta_{r,t}, \tag{20}$$

$$\hat{a}_t = \rho_a \hat{a}_{t-1} + \eta_{a,t}/100, \qquad \hat{e}_t = \rho_e \hat{e}_{t-1} + \eta_{e,t}/100, \qquad \hat{z}_t = \eta_{z,t}/100. \tag{21}$$

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 (16) 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 (17) defines the growth rate $\hat{g}_t$ of output subject to productivity shocks $\eta_{z,t}$. The third equation (18) 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 (19) 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; thus, $\eta_{e,t}$ is a negative cost-push shock. Finally, in equation (20) 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. Note that in light of the rather small magnitude of the innovations, we rescale all shocks by a factor of 100 directly in the model equations, see (21), to ensure numerical stability, particularly for the Bayesian MCMC sampler.

6.2. State space solution

Under rational expectations, agents know both the exact model equations and the full statistical distribution of the white noise process $\eta_t = [\eta_{a,t}, \eta_{e,t}, \eta_{z,t}, \eta_{r,t}]'$ for all t. Hence, the expectation operator E_t is conditional on the information set available in period t, which comprises the state of the economy up to period $t-1$ and the values of current shocks η_t. Given parameter restrictions on $\theta = (\beta, \psi, \omega, \alpha_x, \alpha_\pi, \rho_\pi, \rho_x, \rho_g, \rho_a, \rho_e)$ that ensure stable and unique trajectories (Blanchard & Kahn, 1980), the stochastic solution is characterized by a recursive decision rule (policy function). For a log-linearized model—equivalent to a first-order perturbation solution—this policy function takes a linear state-space form analogous to equations (3) and (4):

$$x_t = Gx_{t-1} + R\eta_t \qquad \text{and} \qquad y_t = Fx_t + \varepsilon_t$$

where $x_t = [\hat{x}_t, \hat{y}_t, \hat{g}_t, \hat{z}_t, \hat{\pi}_t, \hat{a}_t, \hat{e}_t, \hat{r}_t]'$ collects all endogenous variables and $y_t = [\hat{g}_t, \hat{\pi}_t, \hat{r}_t]'$ gathers the observable ones. While the measurement matrix F consists of zeros and ones, the reduced-form matrices G and R are nonlinear functions of θ. We compute these matrices for any given θ using Dynare's first-order perturbation solution algorithm (Villemot, 2011).

6.3. Data and estimation

We consider the same set of quarterly macroeconomic time series for the 1980Q1–2003:Q1 period as originally used in Ireland (2004): (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 $\hat{r}_t$.

Of the ten structural parameters in the model, four are calibrated rather than estimated: $\beta = 0.99$, $\psi = 0.1$, $\alpha_x = 0$, and $\alpha_\pi = 0$.[6] Hence, our interest centers around the other six model parameters plus the parameters of the distribution of η_t. Rather than assuming the conventional multivariate normality for the shocks η_t, we assume each shock follows an *independent univariate skew-normal* distribution, $\eta_{j,t} \sim CSN(\mu_{\eta_j}, \Sigma_{\eta_j}, \Gamma_{\eta_j}, 0, 1)$ for $j \in \{a, e, z, r\}$, where μ_{η_j} is automatically determined to ensure $E[\eta_j] = 0$ for any given Σ_{η_j} and Γ_{η_j}. Moreover, independence allows us to make use of closed-form univariate formulas for the standard error and skewness coefficients, therefore enabling one to directly estimate $stderr(\eta_{j,t})$ and $skew(\eta_{j,t})$ in place of Σ_{η_j} and Γ_{η_j}.

We differentiate between two model variants. In the Gaussian variant, we set $\Gamma_{\eta_j} = 0$ for all j, foregoing the estimation of $skew(\eta_{j,t})$, while in the CSN variant we estimate all $skew(\eta_{j,t})$ parameters. For estimation, we consider two frameworks: (i) Minimizing the negative log-likelihood function and (ii) using a *Random Walk Metropolis Hastings* (RWMH) approach to draw from

---

[6]The parameters β and ψ were originally fixed by Ireland (2004) due to lack of identification, while estimates for α_x and α_π typically converge to values indistinguishable from zero, suggesting that backward-looking behavior of consumers and firms is not important in both the New Keynesian IS and Phillips curve. To avoid boundary pile-up, multi-modality, and other numerical complications, we set α_x and α_π to zero in our baseline estimations. We also estimated versions where these parameters are free; in those cases, we apply a logit transformation to place them on an unbounded domain and—following Ireland (2004)—use one-sided finite differences of the inverse Hessian to compute maximum likelihood standard errors. Bayesian posterior mode finding becomes somewhat more challenging and heavily time-intensive in that setup due to multi-modality. But as the log-likelihood and log-posterior at the mode are actually even lower with free α_x and α_π, we decide to fix these parameters in the baseline.

Table 1: Bounds and priors for model and shock parameters

Parameter	BOUNDS		PRIOR		
	Lower	Upper	Type	Mean	Std-dev
ω	0	1	Beta	0.20	0.10
ρ_π	0	1	Gamma	0.30	0.10
ρ_g	0	1	Gamma	0.30	0.10
ρ_x	0	1	Gamma	0.25	0.0625
ρ_a	0	1	Beta	0.85	0.10
ρ_e	0	1	Beta	0.85	0.10
$stderr(\sigma_a)$	0	10	InvGamma	$\sqrt{30}$	$\sqrt{30}$
$stderr(\sigma_e)$	0	10	InvGamma	$\sqrt{0.08}$	$\sqrt{1}$
$stderr(\sigma_z)$	0	10	InvGamma	$\sqrt{5}$	$\sqrt{15}$
$stderr(\sigma_r)$	0	10	InvGamma	$\sqrt{0.50}$	$\sqrt{2}$
$skew(\sigma_a)$	-0.995	0.995	GenBeta	0	0.40
$skew(\sigma_e)$	-0.995	0.995	GenBeta	0	0.40
$skew(\sigma_z)$	-0.995	0.995	GenBeta	0	0.40
$skew(\sigma_r)$	-0.995	0.995	GenBeta	0	0.40

Note: GenBeta is a Beta distribution with support on [-1,1].

the posterior distribution of the parameters. In all cases, the PSKF computes the log-likelihood function, using consistent numerical routines across model variants and frameworks. Table 1 lists the bounds used during optimization and summarizes the prior distributions employed during Bayesian estimation, which are taken from Table 2 of Chib & Ramamurthy (2014). For the skewness coefficients, we impose a *generalized Beta distribution* with shifted support from [0,1] to [-1,1], a mean of 0 (indicating Gaussianity) and standard deviation of 0.4. Our conclusions do not depend on this chosen type of prior and remain robust even when a uniform or truncated normal distribution is used instead (with a similar mean and appropriate support).

6.4. Computational remarks

First, based on our Monte Carlo evidence, we prune skewness dimensions below a 1% threshold. Second, pre-multiplying η_t by R in the state transition equation is without loss of generality due to Property 2; thus, we work with the linearly transformed distribution $R\eta_t$. Third, as the matrix G is typically singular in DSGE models, numerical issues can arise during prediction. To mitigate this, we compute, filter, and smooth the parameters of the joint distribution of $[x_t', \eta_t']'$ rather than those of x_t alone (Guljanov, 2024). Fourth, the initial distribution for the prediction-error decomposition of the likelihood is set to a normal distribution with mean zero and initial forecast-error covariance equal to the unconditional variance of the state variables (solution to the Lyapunov equation). Fifth, we penalize the likelihood (and posterior) function in several scenarios, such as when the

Blanchard & Kahn (1980) conditions are violated (i.e. a DSGE specific generalization of Eigenvalues of G being outside the unit circle), the covariance matrix of η_t is not positive semi-definite, or the skewness parameters exceed theoretical limits. Sixth, to find the mode of the likelihood, we use a sophisticated search for initial parameter values—a critical step in any (Gaussian or non-Gaussian) maximum likelihood estimation. We start with values from Ireland (2004) for model and standard error parameters, then construct an evenly spaced grid of skewness parameters for all four shocks. For each value on the grid, we compute the log-likelihood, while holding model parameters and standard errors of shocks fixed at their Gaussian estimates. By evaluating over 50,000 combinations (spanning various skewness scenarios between the shocks), we identify the best five parameter sets. These serve as starting points for further numerical optimization of both standard errors and skewness parameters, with model parameters held fixed. The resulting shock parameter estimates are merged with the Gaussian model parameter estimates to form our final initial values for the actual estimation. Equipped with these, we minimize the negative log-likelihood function over all parameters. For Bayesian estimation, a similar approach finds the posterior mode, though the Hessian at the mode can be non-positive definite.[7] To address this, we propose two solutions: (i) Run a Monte Carlo-based optimization routine to locate a high-density region for initializing the Metropolis-Hastings algorithm and to estimate the posterior covariance matrix. This method is, however, very time-consuming. Alternatively, (ii) use a (short) Slice sampler to estimate the mode and posterior covariance matrix directly, as it requires no fine-tuning and is very robust in terms of dealing with multi-modality and high-dimensional parameter spaces. Specifically, we run multiple short Slice sampler chains in parallel (e.g., 8 chains with 250 draws each or less), then use the combined draws to determine the mode and covariance matrix of the posterior distribution which then serves for the initialization of the RWMH algorithm. Our code demonstrates that both approaches yield very similar initialization matrices; therefore, we strongly recommend the

---

[7]The Metropolis-Hastings algorithm does not require starting exactly at the posterior mode; it only needs a starting point with high posterior density and an estimate of the proposal distribution's covariance matrix. Starting at the the mode is, however, beneficial for the acceptance rate and convergence speed and has become standard practice in the Bayesian estimation of DSGE models literature. Likewise, standard practice is to present Bayesian results using the RWMH algorithm, however, we would like to mention that the posterior distribution obtained by RWMH with 2,000,000 draws closely matches the one obtained via a Slice sampler with 40,000 draws. This is noteworthy, because the Slice sampler typically yields Markov chains with lower autocorrelation than RWMH (so less draws required, albeit each draw requires more function evaluations), but more importantly, it avoids both the time-consuming and often frustrating mode-finding step as well as additional fine-tuning to achieve a specific acceptance rate. Results are accessible in the replication package.

Table 2: Parameter estimates

| Parameter | MAXIMUM LIKELIHOOD | | | | BAYESIAN RWMH | | | | | |
| | Gaussian | | CSN | | Gaussian | | | CSN | | |
	Mode	Std-dev	Mode	Std-dev	Mean	Mode	90%-HPD	Mean	Mode	90%-HPD
ω	0.0581	0.0685	0.1596	0.0029	0.1282	0.1231	[0.05;0.21]	0.1392	0.1377	[0.05; 0.22]
ρ_π	0.3865	0.2099	0.2810	0.0048	0.5001	0.4938	[0.35;0.66]	0.4662	0.4522	[0.31; 0.62]
ρ_g	0.3960	0.0612	0.3385	0.0058	0.3627	0.3495	[0.28;0.44]	0.3586	0.3445	[0.28; 0.43]
ρ_x	0.1654	0.0976	0.2871	0.0080	0.2050	0.1813	[0.12;0.29]	0.2227	0.2006	[0.14; 0.30]
ρ_a	0.9048	0.0579	0.9139	0.0082	0.9129	0.9189	[0.87;0.96]	0.9216	0.9336	[0.88; 0.97]
ρ_e	0.9907	0.0130	0.9805	0.0107	0.9107	0.9241	[0.85;0.97]	0.9030	0.9153	[0.85; 0.96]
$stderr(\sigma_a)$	3.0167	1.5568	2.4337	0.0434	3.2584	3.1659	[1.91;4.56]	3.2359	3.2442	[1.84; 4.59]
$stderr(\sigma_e)$	0.0248	0.0180	0.0206	0.0021	0.0602	0.0572	[0.05;0.07]	0.0603	0.0573	[0.05; 0.07]
$stderr(\sigma_z)$	0.8865	0.1245	0.7913	0.0142	0.7887	0.7648	[0.61;0.96]	0.7975	0.7546	[0.62; 0.97]
$stderr(\sigma_r)$	0.2790	0.0374	0.2853	0.0081	0.2953	0.2796	[0.24;0.35]	0.2920	0.2775	[0.24; 0.34]
$skew(\sigma_a)$	—	—	-0.1924	0.0033	—	—	—	-0.0819	-0.1925	[-0.46; 0.33]
$skew(\sigma_e)$	—	—	-0.2174	0.0039	—	—	—	-0.3584	-0.4014	[-0.73;-0.01]
$skew(\sigma_z)$	—	—	-0.9950	0.0712	—	—	—	-0.3808	-0.5166	[-0.89; 0.15]
$skew(\sigma_r)$	—	—	0.8171	0.0140	—	—	—	0.5183	0.6066	[0.23; 0.82]
Obj (mode)	1,207.56		1,215.85		1,205.11			1,211.47		

Note: Obj (mode) is value of the log-likelihood or log-posterior at the estimated mode.

Slice sampler method for it is faster and more general applicable—not just for the PSKF but for Bayesian estimation in general. Subsequently, we generate 2,000,000 draws across 8 parallel chains with the RWMH algorithm, allocating half of the samples for burn-in, and fine-tuning the proposal distribution to achieve an acceptance rate of around 30% for each chain. Table 2 presents the final estimation outcomes under both the maximum likelihood and Bayesian estimation frameworks. All estimations are performed with Dynare 7.0 and MATLAB R2024b on an Apple MacBook Pro equipped with an M2 Max chip and 64 GB RAM.

6.5. Estimation results

From the maximum likelihood estimation, the data clearly favors the CSN distribution over the Gaussian model, as evidenced by a higher maximized log-likelihood value. Given that Gaussianity is nested within the PSKF, a likelihood ratio test substantiates this by yielding a test statistic of 16.58 with a p-value of 0.0023.

Examining the model parameters, we observe differences in the maximum likelihood estimates of the policy parameters ρ_π, ρ_g, and ρ_x. When allowing for skewed monetary policy shocks, these differences suggest that the Federal Reserve's systematic policy is more responsive to movements in the output gap than to output growth. Furthermore, the rule-based inflation sensitivity parameter

is estimated to be lower than in the Gaussian model, although still within the confidence interval. While the maximum likelihood estimates for the persistence of the preference shifter, ρ_a, and the cost-push shock, ρ_e are very similar across model variants, the estimate of ω is considerably higher in the CSN model than in the Gaussian model. This finding has two implications. First, in the theoretical framework, a higher ω implies a more elastic labor supply schedule. Second, empirically, the preference shock appears to have a larger impact on the efficient level of output. We do not observe any meaningful differences in the estimated posterior distributions of the model parameters across model variants, suggesting that the chosen priors exerted a substantial influence on the Bayesian estimation. Comparing the Bayesian mode estimates to the maximum likelihood estimates reveals that the Bayesian approach more closely aligns with the CSN-based results.

Turning to the shock parameters, the estimated standard errors for the cost-push and monetary policy shocks are nearly identical, while the Gaussian filter tends to overestimate those of the preference and productivity shocks. Notably, all shocks exhibit statistically significant skewness coefficients in the maximum likelihood estimation, with particularly pronounced skewness for the productivity and monetary policy shocks—the productivity shock even reaching the theoretical bound of 0.995.

Figure 4 illustrates these differences by comparing the estimated probability density functions based on the maximum likelihood values. The CSN distribution (solid line) for the monetary policy shock has a thinner left tail and a heavier right tail compared to a normal distribution with the same standard deviation (dashed line). Combined with the estimated evidence of a less systematic monetary policy response, this suggests that large, unexpected monetary tightening events are more likely than equally large easing events. This observation aligns with the Federal Reserve's unanticipated hawkish policies during Paul Volcker's tenure as chairman.

Similarly, the productivity shock distribution features a heavier left tail and lighter right tail relative to a normal distribution with the same standard deviation. This pattern is consistent with historical episodes of rare, but large negative productivity shocks in the U.S. between 1980 and 2004—such as sharp oil price increases in the early 1980s and 1990, and the dot-com bust of 2000–2001, particularly when compounded by external disruptions like the Gulf War (Federle et al., 2024). The substantial skewness coefficient for productivity also echoes the findings of Ruge-Murcia (2017). In contrast, the preference and (negative) cost-push shocks are estimated to exhibit

Figure 4: Estimated probability density functions of shocks
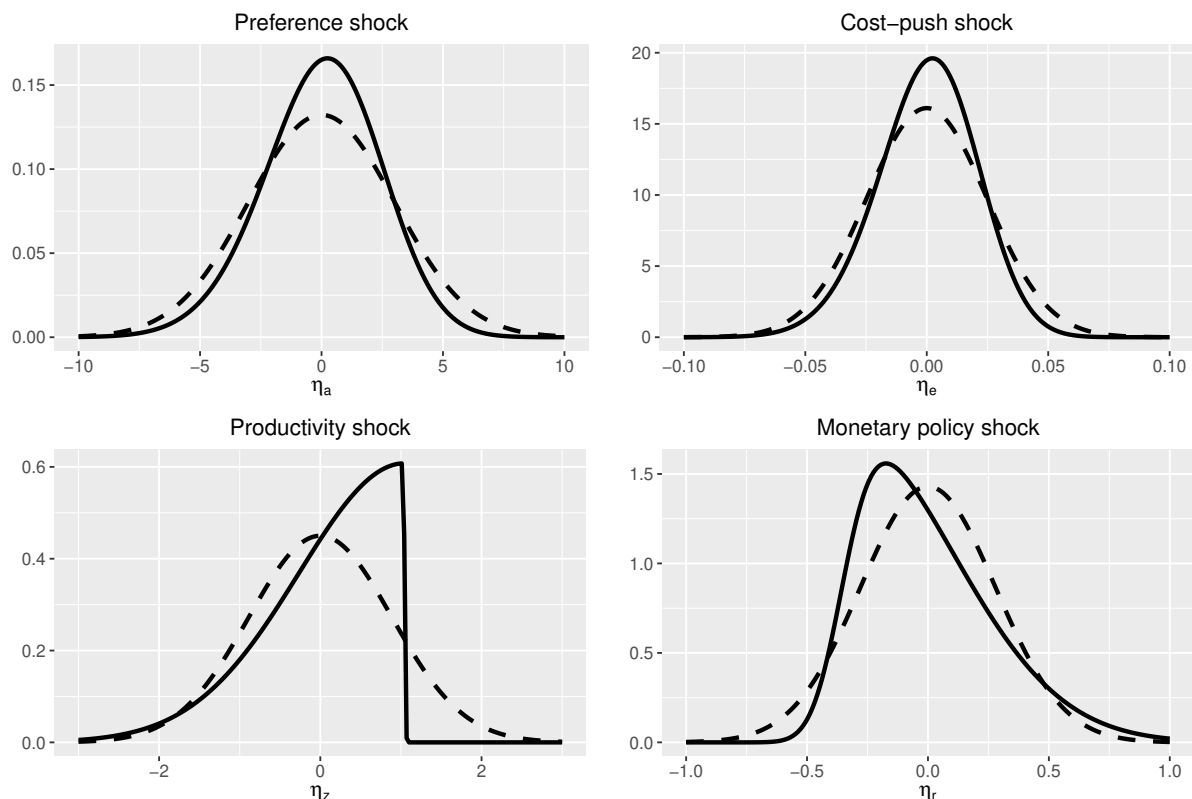
Note: Solid lines represent the estimated CSN distributions, while dashed lines denote the estimated Gaussian distributions based on the maximum likelihood estimates presented in Table 2.

only low levels of negative skewness.

Because our Bayesian prior is centered on zero skewness (i.e. Gaussianity), the posterior estimates exhibit less pronounced skewness than the maximum likelihood estimates, while still following the same pattern in relative magnitude and direction. However, given the historical context and related studies, it would be reasonable to adjust the prior accordingly.

6.6. Impulse responses

We explore the differences between model variants by analyzing the effects of a one-time monetary policy shock on the model variables, as shown in Figure 5. Solid lines represent the maximum likelihood estimates from the *CSN* column in Table 2, while dashed lines correspond to the *Gaussian* column. In the presence of asymmetry, it is crucial to distinguish between positive and negative shocks. Following standard practice—as e.g. in Ruge-Murcia (2017)—we define the 16th percentiles of the estimated distributions as *typical negative* shocks (light gray lines) and the 84th

Figure 5: Impulse response functions to a monetary policy shock
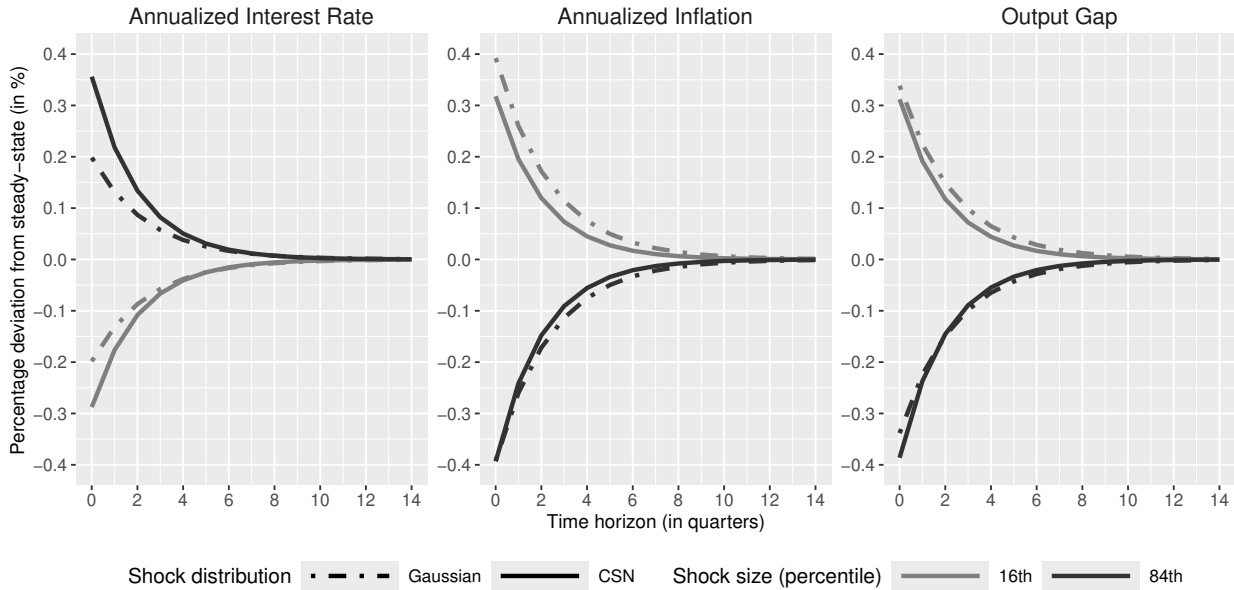
Note: The solid lines represent responses from maximum likelihood estimates of the CSN model, while the dashed lines show responses from the Gaussian model with same model parameters and standard errors of shocks. The light gray lines indicate the 16th percentile (negative shock size) of each estimated distribution (CSN or Gaussian), while the dark gray lines indicate the 84th percentile (positive shock size).

percentiles as *typical positive* shocks (dark gray lines). The responses to a *typical* monetary tightening shock are more pronounced than those to a *typical* easing shock, with the output gap and inflation decreasing more sharply and the interest rate rising more steeply. Since the systematic parameters are estimated slightly differently, the dynamics of the impulse-response function are slightly different as well. It is important to emphasize that with non-Gaussian shocks, even in a linear model, the *size and direction* of shocks are significant for conducting monetary policy, as the transmission channels of typical monetary easing versus monetary tightening shocks are asymmetric.

6.7. Economic implications

Our primary contribution is methodological; therefore, the estimated model is deliberately stylized and primarily intended for illustrative purposes. Even so, the estimates yield valuable insights into the nature of business cycles and the role of monetary policy.

The right-skewed monetary policy shocks imply that large, unexpected tightening episodes—such as sharp increases in the policy rate—are more likely than equally large easing episodes. From

Table 3: Statistics on recessions in the model

	All		Severe		Mild	
	Gaussian	CSN	Gaussian	CSN	Gaussian	CSN
Number recessions ($\hat{g}_t \leq -0.5\%$)	42564	39536	14188	13179	14188	13179
Frequency recessions (in %)	8.51	7.91	2.84	2.64	2.84	2.64
Mean duration (peak to trough, quarters)	3.13	3.07	4.44	4.29	2.23	2.23
Mean output loss (peak to trough, in %)	-2.29	-2.41	-3.68	-3.93	-1.18	-1.21

Note: Statistics are based on a 500,000 time period simulation of model with either Gaussian or CSN distributed shocks. Calibration is based on column CSN in Table 2. Model and stderr parameters are the same in both model variants, Gaussian shocks have skewness coefficients set to zero. A severe (mild) recession is a recession associated with a peak-to-trough output loss in the top (bottom) three deciles of the distribution.

a policy perspective, this suggests that unexpected central bank interventions tend to exhibit an *asymmetry* in response, which can amplify downside risks and make recessions deeper if they coincide with contractionary shocks in other sectors. In historical terms, this is consistent with episodes like the early 1980s *Volcker shock*, where unanticipated, rapid policy tightening occurred in response to high inflation.

Meanwhile, the left-skewed productivity shocks indicate a greater likelihood of large adverse technology or supply-side disturbances than large positive ones. This asymmetry can help explain why certain recessionary episodes (e.g., those triggered by oil price spikes or abrupt technological disruptions like the dot-com bust or wars) can be more severe than expansions. From a broader perspective on the business cycle, such skewness implies that while expansions tend to unfold gradually, recessions may be sharper and more abrupt.

To illustrate this, we follow Boissay et al. (2016) and simulate two 500,000-period time series of the model, one with Gaussian and the other one with CSN distributed shocks. The model is calibrated using the CSN-based maximum likelihood estimates from Table 2. In the Gaussian variant, we set the skewness coefficients of the shocks to zero, but keep the model parameters and standard errors of shocks at their CSN estimates. We define a recession as a period where output growth $\hat{g}_t$ falls below -0.5% per quarter and stays negative for at least two quarters. Severe and mild recessions are determined by the top and bottom deciles of the implied distribution of peak-to-trough output losses. As Table 3 shows, the skew-normal (CSN) model yields slightly fewer total recessions (7.91% vs. 8.51%), which are also a bit shorter on average (3.07 vs. 3.13 quarters). However, their mean output loss is higher (-2.41% vs. -2.29%), driven mainly by deeper severe recessions (-3.93% vs. -3.68%). These patterns align with the model's assumptions. The

strong left-skewed productivity shock (and to some extent the right-skewed monetary policy shock) occasionally trigger sharper downturns, while the right-skewed monetary policy shock delivers a higher likelihood of easing surprises, preventing or shortening some recessions and ultimately resulting in slightly fewer but somewhat deeper recessions overall. Mild recessions behave almost the same in both models, reflecting that *smaller* negative shocks are common in both distributions and don't trigger outsized policy responses or large output drops.[8]

Taken together, these findings offer fresh insights into the dynamics of cyclical fluctuations, productivity and monetary policy. They highlight the importance of modeling the asymmetric nature of shocks rather than defaulting to symmetrical, Gaussian assumptions. Policymakers can use this perspective to better evaluate the risks posed by abrupt tightening or severe negative supply shocks. The observed skewness patterns suggest that central banks and regulators should design policies robust to asymmetric risks. For example, in light of the elevated probability of sharp monetary tightening, financial regulators might consider counter-cyclical buffers or stress tests that account for such scenarios. Similarly, fiscal and monetary interventions could be tailored to mitigate the impact of rare but severe productivity downturns (Galí, 2020).

7. 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 skewness dimensions*. We provide theoretical evidence for its

---

[8]We acknowledge that the differences we find are inherently smaller than those in Boissay et al. (2016). Their study employs a richer, fully nonlinear model with more frictions—particularly mechanisms for banking busts—while our approach is deliberately stylized and fully linear. They also calibrate towards annual data, whereas our model is quarterly, and we have much lower estimated standard errors (scaled by 1/100). Thus, it is unsurprising that our simpler setup yields more modest differences compared to their large effects. Nevertheless, our primary objective is methodological. Even in this pared-down framework, we do observe notable distinctions between a Gaussian version and one featuring skew-normal shocks. Incorporating these shock asymmetries into models with, for instance, financial frictions or larger shocks should generate more pronounced dynamics.

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 a multivariate real data applications. In related work, we already demonstrate the scalability and applicability of the PSKF to high-dimensional state-space models, namely the Smets & Wouters (2007) model with about 30 state variables, embedded in a standard Bayesian MCMC estimation framework (Guljanov, 2024). Furthermore, the extension of the *pruned skewed Kalman filter* to the *pruned skewed Student's-t filter* is an area for future research as there is evidence that both skewness as well as heavy tails in the innovations are important drivers of business cycle models. The pruning algorithm can be applied in a similar manner to achieve this extension.

Lastly, the evidence of skewness in key macroeconomic shocks opens promising avenues for further research. It challenges the standard framework that assumes symmetric shocks and prompts investigation into the sources of asymmetry—whether market frictions, policy errors, or structural factors. This could lead to refined theoretical models that better capture how shocks are distributed and how they propagate through the economy. Importantly, these models can still operate within a linear setting, provided they account for non-normal shock distributions. Once skewness is identified, it becomes possible to explore how asymmetric shocks interact with financial frictions or nonlinear policy rules, potentially yielding richer explanations of macroeconomic volatility and more robust policy responses.

Acknowledgments

Declaration of generative AI and AI-assisted technologies in the writing process

Generative AI and AI-assisted technologies were employed to enhance the manuscript's readability and language. These tools were used under human supervision and all authors thoroughly reviewed the final text taking full responsibility for its published content.

References

Adjemian, S., Juillard, M., Karamé, F., Mutschler, W., Pfeifer, J., Ratto, M., Villemot, S., & Rion, N. (2024). *Dynare: Reference Manual Version 6*. Dynare Working Papers 80 CEPREMAP.

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.

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.

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

Boissay, F., Collard, F., & Smets, F. (2016). Booms and Banking Crises. *Journal of Political Economy*, *124*, 489–538. doi:10.1086/685475.

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.

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.

Curdia, V., Del Negro, M., & Greenwald, D. L. (2014). Rare Shocks, Great Recessions. *Journal of Applied Econometrics*, *29*, 1031–1052. doi:10.1002/jae.2395.

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.

Durbin, J., & Koopman, S. J. (2012). *Time Series Analysis by State Space Methods: Second Edition*. Oxford University Press. doi:10.1093/acprof:oso/9780199641178.001.0001.

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.

Federle, J., Meier, A., Müller, G., Mutschler, W., & Schularick, M. (2024). *The Price of War*. CEPR Discussion Papers 18834 Centre for Economic Policy Research.

Galí, J. (2020). The effects of a money-financed fiscal stimulus. *Journal of Monetary Economics*, *115*, 1–19. doi:10.1016/j.jmoneco.2019.08.002.

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.

Guljanov, G. (2024). *Estimation of DSGE Models: Skewness Matters*. Ph.D. thesis Westfälische Wilhelmsuniversität Münster Münster. URL: https://miami.uni-muenster.de/Record/8dd8e05e-e4d6-452d-80af-fbb4684610a1. doi:10.17879/35918450119.

Guljanov, G., Mutschler, W., & Trede, M. (2022). *Pruned Skewed Kalman Filter and Smoother: With Applications to the Yield Curve*. CQE Working Papers 101 Center for Quantitative

Economics (CQE), University of Münster. URL: `https://www.wiwi.uni-muenster.de/cqe/sites/cqe/files/CQE_Paper/cqe_wp_101_2022.pdf`.

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

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. doi:`10.1016/j.jmva.2013.09.002`.

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`.

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. (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`.

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

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

Rezaie, J., & Eidsvik, J. (2014). Kalman filter variants in the closed skew normal setting. *Computational Statistics & Data Analysis*, *75*, 1–14. doi:10.1016/j.csda.2014.01.014.

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

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

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

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

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

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.